

SEPTIN PUJI ASTUTI



STATISTIKA



Fakultas Ekonomi dan Bisnis Islam
Institut Agama Islam Negeri Surakarta

SEPTIN PUJI ASTUTI



STATISTIKA



Fakultas Ekonomi dan Bisnis Islam
Institut Agama Islam Negeri Surakarta

STATISTIKA

Septin Puji Astuti © Penulis 2020,

Hak cipta dilindungi undang-undang. Sebagian atau seluruh isi buku ini dilarang diperbanyak dalam bentuk apa pun tanpa izin tertulis dari penerbit.

STATISTIKA / Septin Puji Astuti

– cet.1. – Yogtakarta: gerbang media aksara, 2020

xvi + 182 hlm., 15,5 x 23 cm.

ISBN : 978-602-6248-88-6

CV Gerbang Media Aksara Bekerjasama dengan Fakultas Ekonomi dan Bisnis Islam IAIN Surakarta.

Alamat. Jl. Sampangan No. 58A, Rt 01 Banguntapan, Bantul, Yogyakarta Telp. (0174)4353651

Kata Pengantar

Assamualaukum Wr.wb

Alhamdulillah puji syukur kita panjatkan kehadiran Allah SWT yang telah melimpahkan rahmat dan hidayahnya, sehingga penyusunan Buku Daras Fakultas Ekonomi dan Bisnis Islam IAIN Surakarta tahun Anggaran 2020 dapat terselesaikan dengan baik.

Penyusunan Buku Daras ini memiliki maksud adalah sebagai acuan dan refrensi bagi seluruh civitas Akademika di lingkungan Fakultas Ekonomi dan Bisnis Islam khususnya bagi Mahasiswa yang sedang menyelesaikan mata kuliah yang tertera dalam judul buku daras ini. Selain itu buku daras ini digunakan sebagai panduan bagi Dosen dalam melaksanakan pembelajaran dan pengajaran kepada mahasiswa.

Buku Daras ini merupakan salah satu program tahunan dilaksanakan di Fakultas Ekonomi dan Bisnis Islam yang bersumber dari anggaran PNBPN yang tertera dalam daftar isian pelaksanaan anggaran tahun 2020

Besar harapan kami buku daras ini memberikan manfaat kepada Dosen, Mahasiswa dan seluruh civitas akademika yang sedang melakukan proses pembelajaran dan perkuliahan pada mata kuliah yang tertera dalam buku ini

Dengan mengucapkan Alhamdulillah dengan hadirnya buku daras ini semoga memberikan manfaat yang sangat besar kepada peningkatan kualitas dan mutu pembelajaran di lingkungan Fakultas Ekonomi dan Bisnis Islam.

Akhirnya kami ucapkan terima kasih kepada seluruh pihak yang membantu dalam proses penyelesaian buku dasar ini khususnya kepada para penulis.

Wassalamualaikum wr.wb

Dekan
Fakultas Ekonomi dan Bisnis Islam

Dr. Rahmawan Arifin, SE., M.Si

Pengantar Penulis

Kebutuhan akan analisis data menggunakan statistika dalam semua bidang, membuat Mata Kuliah Statistika menjadi mata kuliah yang diajarkan di hampir semua program studi. Mempelajari Statistika menjadi penting karena dianggap mampu membantu menyelesaikan berbagai masalah dan dalam pengambilan keputusan.

Statistika pada dasarnya tidak hanya bicara mengolah data. Akan tetapi juga berbicara bagaimana mendapatkan data yang baik, menampilkannya dalam bentuk yang mudah dipahami, mengolahnya, hingga menyimpulkan informasi dari data yang ada hingga diperoleh informasi adalah bagian dari statistika. Setiap tahap dalam statistika dari awal hingga mengambil kesimpulan adalah penting dan harus dipelajari dengan benar. Kesalahan dalam tahap awal, yaitu proses pengambilan data akan menyebabkan kesalahan dalam pengambilan data, walaupun analisis yang dilakukan sudah sesuai dengan kaidah statistika.

Buku ini hadir sebagai buku ajar Statistika yang diharapkan akan berguna dalam pengajaran Mata Kuliah Statistika Dasar. Pengertian Statistika dan bagaimana sejarah statistika serta dasar-dasar dari Statistika dibahas dalam Bab 1. Teknik-teknik pengambilan sampel juga dibahas pada Bab 1. Selanjutnya, ukuran-ukuran data yang menjelaskan tipe data dan skala data dibahas lebih rinci pada Bab 2. Bab

ini diperlukan sebagai dasar pengetahuan ilmu data agar dalam pengolahannya tepat dan memberi informasi yang bermanfaat. Bab 3 membahas tentang dasar-dasar probabilitas yang merupakan dasar dari Statistika. Ilmu probabilitas ini yang akan membantu pengolahan data saat terdapat ketidakpastian. Dalam Statistika, setelah mendapatkan data, bagaimana menampilkannya dan meringkasnya akan dituntaskan untuk dibahas pada Bab 4. Selanjutnya masuk pada analisis inferensia yaitu uji hipotesis yang dibahas pada Bab 5. Bab terakhir adalah Bab 6 akan membahas bagaimana analisis inferensia lebih lanjut yaitu analisis korelasi dan analisis regresi.

Buku telah disusun sedemikian rupa sehingga mudah dipahami oleh mahasiswa dan pengguna statistika pemula. Contoh-contoh soal dibuat beserta penyelesaiannya agar pembaca mampu menyelesaikan dengan baik. Penjelasan beserta filosofi dari analisis juga dijelaskan supaya pembaca memahami mengapa suatu analisis dilakukan.

Pada akhirnya, saya sebagai penulis mengucapkan terima kasih kepada Fakultas Ekonomi dan Bisnis Islam beserta jajaran pimpinan yang telah memberikan kesempatan kepada saya untuk menyelesaikan buku ajar ini. Juga, kepada staf akademik yang banyak membantu dalam penyelesaian kegiatan administrasi penerbitan buku ini. Akhirul kata, semoga buku ini memberi manfaat kepada semua pembaca. Apabila ada kritik dan saran mohon disampaikan melalui email saya septin.astuti@iain-surakarta.ac.id. Terima kasih.

Penulis

Septin Puji Astuti, S.Si., M.T., Ph.D

DAFTAR ISI

Kata Pengantar	iii
Pengantar Penulis	v
Daftar Isi	vii
Daftar Gambar	xi
Daftar Tabel	xv
BAB 1. PENDAHULUAN	1
1.1. Pengantar.....	2
1.2. Populasi dan sampel.....	7
1.3. Data dan Pengambilan Keputusan dalam Statistika	18
1.4. Istilah-istilah dalam Statistika	21
BAB 2. PENGUKURAN DATA	23
2.1. Level Pengukuran Data.....	23
2.2. Tipe Data	28
BAB 3. TEORI PROBABILITAS	33
3.1. Elemen-elemen dalam Probabilitas	33
3.1.1. Ruang Sampel (<i>Sample Space</i>).....	34
3.1.2. Kejadian (<i>event</i>)	35
3.2. Menghitung titik sampel: Permutasi dan Kombinasi.....	41
3.3. Probabilitas	45

BAB 4.	MERINGKAS DAN MENAMPILKAN	
	DATA	52
4.1.	Meringkas data dalam Tabel dan Grafik	52
4.1.1.	Tabel Frekuensi	53
4.1.2.	Grafik Data	56
4.2.	Mengukur Pemusatan dan Penyebaran	
	Data	65
4.2.1.	Rata-rata (<i>average</i>).....	65
4.2.2.	Median.....	69
4.2.3.	Modus	71
4.2.4.	Kuartil	72
4.2.5.	<i>Range</i>	74
4.2.6.	Varians	75
4.2.7.	Standar Deviasi.....	77
4.2.8.	Skewness	78
4.2.9.	Penerapan ukuran pemusatan dan penyebaran data	79
4.3.	Distribusi Normal	82
BAB 5.	UJI HIPOTESIS	88
5.1.	Pengantar Uji Hipotesis	89
5.2.	Uji Hipotesis untuk Satu Sampel	94
5.2.1.	Uji <i>t</i> satu sampel.....	95
5.2.2.	Uji proporsi satu sampel.....	100
5.3.	Uji Hipotesis Dua Sampel	102
5.3.1.	Uji rata-rata dua sampel berpasangan.....	103
5.3.2.	Uji rata-rata dua sampel yang independen	107
5.3.3.	Uji proporsi dua sampel	115
5.4.	Uji Hipotesis lebih dari Dua Sampel: <i>Analysis of Variance</i>	118

BAB 6.	ANALISIS KORELASI DAN REGRESI....	124
6.1.	Analisis Korelasi	124
6.1.1.	Korelasi <i>Product Moment Parson</i>	125
6.1.2.	Korelasi <i>Rank Spearman</i>	130
6.2.	Dasar-dasar Analisis Regresi.....	132
6.2.1.	Apa Analisis Regresi?.....	132
6.2.2.	Linearitas dalam Model Regresi	134
6.3.	Metode <i>Ordinary Least Square</i> untuk Menaksir Model Regresi	135
6.4.	Koefisien Determinasi (R^2).....	140
6.5.	Penarikan Kesimpulan dalam Model.....	143
6.5.1.	Taksiran interval dan uji hipotesis parameter dalam model....	143
6.5.2.	Analysis of Variance dari Model Regresi	147
6.6.	Asumsi dalam Model Regresi	149
6.6.1.	Asumsi 1: Model Regresi Linear.....	149
6.6.2.	Asumsi 2: Nilai x adalah tetap dalam sampling yang berulang	152
6.6.3.	Asumsi 3: Rata-rata residual adalah nol.....	152
6.6.4.	Asumsi 4: Homoskedastisitas atau varians dari residual adalah sama	153
6.6.5.	Asumsi 5: Tidak ada autokorelasi antar residual.....	156
6.6.6.	Asumsi 6: Tidak ada kovarians antar residual dan x	158
6.6.7.	Asumsi 7: Jumlah observasi harus lebih besar dari parameter yang akan ditaksir	158
6.6.8.	Asumsi 8: Variabilitas dalam x	159
6.6.9.	Asumsi 9: Model regresi harus ditentukan secara tepat	159

6.6.10. Asumsi 10: Tidak ada multikolinearitas	159
6.7. Asumsi Residual Berdistribusi Normal	161
Daftar Pustaka	165
Lampiran	
Lampiran 1: Tabel Distribusi z	167
Lampiran 2: Tabel Distribusi t	169
Lampiran 3: Distribusi F ($\alpha = 0,1$).....	171
Lampiran 4: Distribusi F ($\alpha = 0,05$).....	173
Lampiran 5: Tabel Durbin-Watson Statistika d pada $\alpha = 0,05$	175
Lampiran 6: Tabel Durbin-Watson Statistika d pada $\alpha = 0,01$	179

Daftar Gambar

Gambar 1.1:	Hubungan antara statistika inferensia dan probabilitas	6
Gambar 1.2:	Populasi dan sampel	7
Gambar 1.3:	Contoh pengambilan sampel acak sederhana tanpa pengembalian.	11
Gambar 1.4:	Contoh pengambilan sampel acak sederhana dengan pengembalian.	11
Gambar 1.5:	Ilustrasi pengambilan sampel dengan teknik sistematik.....	13
Gambar 1.6:	Ilustrasi pengambilan sampel dengan teknik stratifikasi.	14
Gambar 1.7:	Ilustrasi pengambilan sampel dengan teknik <i>cluster</i>	15
Gambar 1.8:	Ilustrasi pengambilan sampel dengan teknik <i>multistage</i>	16
Gambar 1.9:	Peranan statistika dalam sistem manajemen pengambilan keputusan (Wegner, 2013)	20
Gambar 2.1:	Data inflasi bulanan Indonesia sejak Mei 2017 hingga April 2020	29
Gambar 2.2:	Data temperatur udara rata-rata harian di Kota Semarang pada 8 Mei 2019 hingga 5 Mei 2020.....	30
Gambar 3.1:	Irisan A dan B.....	37

Gambar 3.2: Gabungan A dan B	38
Gambar 3.3: Himpunan komplementer	39
Gambar 3.4: Kejadian A dan C adalah mutually exclusive.....	40
Gambar 4.1: Dotplot nilai pelatihan karyawan	57
Gambar 4.2: Diagram batang dan daun dari data pada Tabel 4.1.	58
Gambar 4.3: Diagram batang penjualan handphone pada Januari 2020	59
Gambar 4.4: <i>Histogram</i> nilai pelatihan karyawan dari Tabel 4.1.	60
Gambar 4.5: <i>Pie chart</i> atau diagram lingkaran.....	62
Gambar 4.6: <i>Scatterplot</i> sebaran data pada Tabel 4.6 ..	63
Gambar 4.7: <i>Time series plot</i> inflasi Indonesia pada Mei 2017 hingga April 2020	64
Gambar 4.8: Lokasi rata-rata pada sebaran data.....	67
Gambar 4.9: Lokasi rata-rata relatif terhadap data lainnya.....	68
Gambar 4.10: Lokasi median relatif terhadap data lainnya.....	70
Gambar 4.11: Lokasi Q_1 , Q_2 , dan Q_3	72
Gambar 4.12: Lokasi Kuantil pada keseluruhan data ..	74
Gambar 4.13: Beberapa derajat <i>skewness</i> suatu data	78
Gambar 4.14: Kurva normal	82
Gambar 4.15: Distribusi kasus dalam kurva normal	83
Gambar 4.16: Berbagai kurva normal dengan rata-rata dan standar deviasi yang bervariasi (a) $\mu_1 < \mu_2$ dan $\sigma_1 = \sigma_2$, (b) $\mu_1 = \mu_2$ dan $\sigma_1 < \sigma_2$, dan (c) $\mu_1 < \mu_2$ dan $\sigma_1 < \sigma_2$	85
Gambar 4.17: Contoh data berdistribusi normal dengan berbagai rata-rata dan standar deviasi.	85
Gambar 4.18: Daerah yang diarsir adalah $P(x_1 < x < x_2)$	86

Gambar 4.19: Luas $P(-0,5 < x < 1,2)$	87
Gambar 5.1: Lokasi daerah kritis pada distribusi t jika menggunakan uji dua arah.....	91
Gambar 5.2: Penentuan uji hipotesis.....	94
Gambar 5.3: Uji parametrik dan non-parametrik untuk satu sampel	95
Gambar 5.4: Mencari t dari Tabel distribusi t di Lampiran 2.....	98
Gambar 5.5: Daerah penolakan untuk Contoh 5.1.....	99
Gambar 5.6: Mencari t dari Tabel distribusi z di Lampiran 1.....	102
Gambar 5.7: Uji-uji parametrik dan non-parametrik untuk dua sampel.....	102
Gambar 5.8: Lokasi $t_{\alpha/2,29}$ pada Tabel distribusi t di Lampiran 2.....	106
Gambar 5.9: Daerah penolakan pada Contoh 5.3	107
Gambar 5.10: Cara mencari $z_{0,025}$ dalam tabel distribusi z	111
Gambar 5.11: Mencari $z_{0,01}$ pada tabel distribusi z	117
Gambar 5.12: Uji parametrik dan non-parametrik pada lebih dari dua sampel.....	118
Gambar 5.13: Tabel distribusi F di Lampiran 4 untuk mencari $F_{5\%,3,24}$	123
Gambar 6.1: Pola hubungan antar dua variabel.	125
Gambar 6.2: Pola korelasi diadopsi dari Henri Theil (1978) dalam bukunya <i>Introduction to Econometrics</i> yang dikutip oleh (Gujarati, 2003).....	127
Gambar 6.3: <i>Scatterplot</i> hubungan tinggi badan dan berat badan yang korelasinya 0,76.....	129
Gambar 6.4: Gambar regresi dalam diagram <i>scatterplot</i>	134

Gambar 6.5: Pola hubungan linear dalam parameter namun, non linear dalam variabel.....	135
Gambar 6.6: Kriteria <i>least square</i>	136
Gambar 6.7: Lokasi \bar{x} dan \bar{y} pada garis regresi	137
Gambar 6.8: Model regresi dinyatakan dalam grafik.	139
Gambar 6.9: Taksiran hubungan antar pendapatan dan pengeluaran keluarga dengan model linear.....	150
Gambar 6.10: Taksiran hubungan antar pendapatan dan pengeluaran keluarga dengan model kuadratik	151
Gambar 6.11: Taksiran hubungan antar pendapatan dan pengeluaran keluarga dengan model kubik	151
Gambar 6.12: Model regresi yang residualnya homoskedastis.....	154
Gambar 6.13: Model regresi yang residualnya heteroskedastis.....	154
Gambar 6.14: <i>Scatterplot</i> antara ε^2 dengan \hat{y}	155
Gambar 6.15: <i>Scatterplot</i> antara ε^2 dengan \hat{y} dari model regresi pada Contoh 6.4.....	156
Gambar 6.16: Histogram residual.....	162
Gambar 6.17: Uji distribusi normal residual dengan <i>Kolmogorov-Smirnov test</i>	163
Gambar 6.18: Uji distribusi normal residual dengan <i>Anderson-Darling test</i>	163

Daftar Tabel

Tabel 2.1:	Ciri setiap skala data.....	28
Tabel 2.2:	Data kependudukan Kota Surakarta tahun 2016 (BPS Kota Surakarta, 2017)	31
Tabel 2.3:	Data kependudukan Kota Surakarta tahun 2016 dan 2017 (BPS Kota Surakarta, 2017, 2018)	31
Tabel 3.1:	Tabel susunan pasangan karyawan yang mungkin mendapat tiga penghargaan	42
Tabel 3.2:	Susunan karyawan yang mungkin akan dikirim ke kantor pusat.....	44
Tabel 4.1:	Nilai hasil pelatihan karyawan	53
Tabel 4.2:	Tabel Frekuensi model pertama yang diringkas dari Tabel 4.1	54
Tabel 4.3:	Tabel Frekuensi model kedua yang diringkas dari Tabel 4.1	55
Tabel 4.4:	Tabel Frekuensi dari data pada Tabel 4.1 yang mengikuti tahapan Salkind (2008)	56
Tabel 4.5:	Data penjualan handphone pada Januari 2020.....	59
Tabel 4.6:	Data berat badan dan tinggi orang.....	63
Tabel 4.7:	Inflasi di Indonesia sejak Mei 2017 hingga April 2020.....	64
Tabel 4.8:	Data dari Tabel 4.2 yang akan digunakan untuk ilustrasi.....	70

Tabel 4.9: Penggunaan ukuran-ukuran statistik sesuai dengan skala datanya	80
Tabel 4.10: Data pendidikan konsumen produk X...	81
Tabel 4.11: Konversi data pada Tabel 4.10	81
Tabel 5.1: Tipe kesalahan dalam mengambil kesimpulan dalam uji hipotesis	93
Tabel 5.2: Rekap pengunjung yang datang pada hari Minggu selama tahun 2019	97
Tabel 5.3: Penilaian pelanggan terhadap kemasan lama dan baru	105
Tabel 5.4: Penilaian pengetahuan lingkungan masyarakat di daerah A dan B	110
Tabel 5.5: Data pengunjung toko di hari Minggu dan hari kerja	112
Tabel 5.6: Data pengunjung toko di hari Sabtu.....	114
Tabel 5.7: Struktur data untuk ANOVA.....	119
Tabel 5.8: Tabel ANOVA	120
Tabel 5.9: Hasil pengujian TDS yang dihasilkan oleh tiap mesin.....	121
Tabel 5.10: Tabel ANOVA hasil oleh data.....	123
Tabel 6.1: Kekuatan korelasi menurut Salkind (2008)	127
Tabel 6.2: Data tinggi badan dan berat badan.	128
Tabel 6.3: Data tinggi badan dan berat badan.	131
Tabel 6.4: Model-model regresi berdasarkan linearitas parameter dan variabel	135
Tabel 6.5: Pendapatan dan pengeluaran kepala keluarga di kampung ABC	138
Tabel 6.6: Tabel ANOVA Regresi	147
Tabel 6.7: Hasil ANOVA latihan soal	148
Tabel 6.8: Residual dari model regresi pada Contoh 6.4.....	153
Tabel 6.9: Aturan keputusan untuk uji <i>Durbin-Watson</i>	157

BAB 1.

PENDAHULUAN

STATISTIKA sudah menjadi ilmu yang digunakan oleh berbagai cabang ilmu. Dalam penerapannya yang sangat luas, statistika telah terbukti mampu membantu menyelesaikan berbagai permasalahan yang ada. Pemahaman tentang statistika, pemahaman mengenai perkembangannya di masa lalu hingga sekarang, pemahaman menggunakan statistika dengan baik harus dilakukan supaya statistika dapat digunakan untuk menyelesaikan masalah dengan tepat. Pada Bab Pendahuluan ini akan dibahas bagaimana sejarah singkat statistika, apa itu statistika, bagaimana statistika dapat digunakan untuk mengambil keputusan dan penyelesaian masalah, dan bagaimana pengambil data dengan teknik-teknik pengambilan sampel. Materi tersebut disampaikan pada Bab Pendahuluan ini dengan harapan di akhir bab ini pembaca:

- Mampu mengetahui bagaimana statistika dan mengetahui bagaimana penerapannya.
- Mampu mengetahui definisi dari statistika dan apa bedanya dengan statistik, serta mampu menerapkannya dalam praktik.
- Mampu mengetahui data yang merupakan elemen penting dari statistika yang nantinya dapat digunakan sebagai penyusun informasi untuk mendukung pengambilan keputusan.

- Mampu mengetahui dan membedakan elemen penting dalam pengambilan data yaitu populasi dan sampel.
- Mampu mengetahui dan membedakan apa statistika deskriptif dan inferensia.

1.1. Pengantar

1.1.1. Sejarah Singkat Statistika

Statistika digunakan oleh para ilmuwan berabad-abad tahun yang lalu. Pada awalnya, sebagian besar statistik diterapkan untuk menyelesaikan masalah sains dan kesehatan. Dalam paragraf berikut adalah ringkasan perkembangan statistik dirangkum secara sangat singkat dari beberapa sumber.

Desrosières (1998) merangkum beberapa sejarah perkembangan statistika dalam bukunya berjudul *The Politics of Large Numbers: A History of Statistical Reasoning*. Statistika berkembang pertama kalinya di Eropa, terutama di Jerman dan Inggris. Jerman adalah negara yang memformalkan statistika pertama kalinya. Statistik diinspirasi dari logika Aristoteles yang kemudian pada tahun 1660 dikodifikasi oleh Conring (1606-1681). Kemudian, baru dilanjutkan pada abad ke-18 di mana University of Göttingen memiliki Program Studi Statistika yang digawangi oleh Achenwall (1719-1772). Achenwall adalah penemu istilah statistik. Schlözer (1735-1809) yang adalah ketua Program Studi Statistika di University of Göttingen yang juga penulis buku *Treatise on Statistics*. Pepatah dari Schlözer yang cukup terkenal adalah, “*Statistics is history without motion. History is statistics in motion*”.

Selain Jerman, Inggris adalah negara yang juga menggunakan statistik sejak abad ke-17. Pekerjaan Graunt (1620-1674) yang membuat *bills of mortality* menisiasi Petty (1623-1687) melakukan hal sama. *Bills of mortality* melaporkan

kematian mingguan yang dirancang untuk merekam pemakaman pada tahun 1592 sampai 1595 di London.

Pada tahun 1713, Jacques Bernoulli sudah memformulasikan teorema *The Law of Large Numbers*. Teorema ini adalah dasar dari ilmu peluang dan statistik yang membuktikan hubungan yang kuat antara probabilitas dan frekuensi. Selanjutnya statistika digunakan untuk menaksir konstanta fisik dan astronomi dilakukan sekitar tahun 1750. Quetelet (1830an-1840an) mengenalkan hubungan antara teori peluang dan pengamatan secara statistik. Dia mengenalkan penggunaan rata-rata. Namun, pada tahun 1830a, mulai muncul kontroversi dari penggunaan rata-rata yang digunakan di dunia medis.

Francis Galton (1822-1911) yang merupakan sepupu dari Charles Darwin (Salkind, 2008), pada tahun 1886 mengenalkan analisis regresi dalam papernya *Family Likeliness in Stature* yang diterbitkan di *Proceeding of Royal Society Volume 40* (Gujarati, 2003). Pada waktu itu, Galton menemukan bahwa ada kecenderungan bahwa tinggi badan anak laki-laki diturunkan dari orangtua laki-laki. Selanjutnya, penerapan statistika dalam industri sempat membawa Jepang sukses dalam menjalankan industrinya. Di saat industri-industri di dunia gagal, Jepang sukses menerapkan pola pikir statistika dalam meningkatkan kualitas pada produk-produk industrinya (Walpole et al., 2013).

Hingga berabad-abad tahun kemudian statistika telah berkembang dengan sangat cepat dengan banyak ilmuwan yang telah melahirkannya. Penerapannya juga dilakukan dalam berbagai bidang. Ekonometrika merupakan salah satu penerapan statistika dalam bidang ekonomi. Penggunaan istilah ekonometrika digunakan pertama kali oleh ekonom Polandia, Powel Ciompa, pada tahun 1910. Namun perkembangannya baru mulai pesat sejak tahun 1930-

1940an. Namun, jauh sebelumnya ada biostatistika, sekitar 1900an. Gregor Mendel menggunakan statistika untuk melakukan eksperimen dalam penelitian genetiknya. Selanjutnya, peneliti lain seperti Francis Galton dan Ronald Fisher menggunakan statistika dalam melakukan penelitiannya.

Kini statistika modern banyak dikembangkan seiring dengan perkembangan teknologi. Penerapannya juga telah menggunakan berbagai alat-alat teknologi yang dapat membantu mendapatkan data sehingga dapat diambil suatu informasi yang lebih komprehensif. Beberapa di antaranya adalah diterapkan dalam manajemen pemasaran yang memanfaatkan teknologi informasi dengan mengidentifikasi karakteristik konsumen, iklan dapat dipromosikan kepada calon konsumen yang tepat. Dalam bidang politik, statistika dan ilmu data sangat membantu mengidentifikasi calon pemilih atau orang-orang potensial supaya mau memilih calon yang dijagokan. Juga, statistika dapat mengidentifikasi isu yang tepat yang diwacanakan ke masyarakat dan mampu mempengaruhi pemilih.

1.1.2. Definisi Statistika

Ilmu Statistika adalah bagian dari ilmu Matematika. Namun, statistika dan matematika harus dibedakan. Patil dan Rao (1994) mengatakan bahwa matematika tidak membutuhkan data dan cenderung filosofis, sementara statistika membutuhkan data yang harus diolah untuk menjadi informasi.

Statistika bisa dianggap sebagai sekumpulan kegiatan administratif yang dibutuhkan untuk menggambarkan suatu keadaan dan populasi (Desrosières, 1998). Statistika juga bisa diartikan sebagai sekumpulan alat dan teknik-teknik yang digunakan untuk menggambarkan, mengelola, dan menginterpretasikan informasi atau data (Salkind,

2008). Wegner (2013) mendefinisikan statistika sebagai sekumpulan alat-alat berdasarkan matematika dan teknik-teknik mengubah data mentah ke dalam ukuran-ukuran yang ringkas yang menggambarkan pola data, menguji hubungan antar data yang berguna untuk memberi informasi dalam mengambil keputusan. Patil dan Rao (1994) mengatakan bahwa sebagai statistisi harus menguasai kemampuan mengumpulkan data, mengelolanya, menyimpan, dan mengambil data yang akan digunakan untuk melihat pola, keteraturan, dan penyederhanaan dari sekumpulan data untuk mendapatkan statistik atau model yang digunakan menginformasikan sekumpulan data yang telah dianalisis. Metode-metode statistika dibuat untuk membantu suatu pernyataan ilmiah berdasarkan data-data yang telah diperoleh di mana dalam data tersebut terdapat ketidakpastian dan da variasi (Walpole et al., 2013).

Dari berbagai pendapat tersebut, bisa diambil suatu kesimpulan bahwa statistika adalah sekumpulan **konsep** dan **metode** untuk mengumpulkan, mengelola, menyajikan, menganalisis data, dan mengambil kesimpulan berdasarkan hasil analisis data dalam situasi data yang memiliki **ketidakpastian** dan **variasi**. Ketidakpastian tersebut melahirkan variabilitas yang mana data akan menjadi heterogen atau tidak seragam atau tidak monoton. Kondisi yang penuh ketidakpastian dan ada variabilitas itulah yang mendorong metode-metode statistika lahir. Walpole et al. (2013) bahkan mengatakan, jika terdapat data yang selalu sama, yang berarti tidak memiliki variabilitas, maka statistika tidak dibutuhkan lagi.

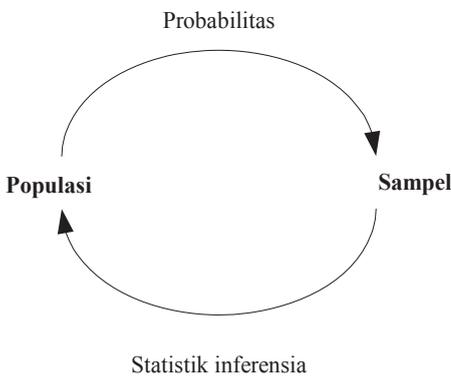
1.1.3. Teknik-teknik Statistika: Deskriptif vs Inferensia

Dalam analisis statistika terdapat dua teknik yaitu statistika deskriptif dan statistika inferensia. Dalam teknik

statistika deskriptif sesuai dengan namanya yaitu hanya bersifat menggambarkan data. Kegiatan yang dilakukan dalam statistika deskriptif adalah mengelola data dan menjelaskan karakteristik dari sekumpulan data (Salkind, 2008). Beberapa kegiatan mengelola data seperti meringkas dan menyajikan data juga merupakan statistika deskriptif. Menjelaskan atau mendeskripsikan data bisa dilakukan dalam bentuk tabel dan grafik yang dapat juga dilakukan dengan penjelasan rinci dalam bentuk kalimat. Seringkali statistika deskriptif disebut juga sebagai statistika deduktif menunjukkan hasil secara umum dari data saja tanpa merinci.

Sementara, dalam **statistika inferensia**, selain meringkas dan menyajikan, juga bisa melakukan analisa hingga mengambil suatu kesimpulan berdasarkan informasi dari hasil analisis. Sehingga, dengan analisis inferensia ini mampu didapatkan informasi yang lebih komprehensif beserta alasan-alasannya. Dalam statistika inferensia, probabilitas memiliki peran yang sangat penting. Ini karena inferensia atau kesimpulan diambil dari analisis data yang diambil dengan menggunakan teknik sampling probabilistik. Kemampuan statistika membuat kesimpulan berdasarkan sampel dari populasi inilah yang menjadi kekuatan statistika.

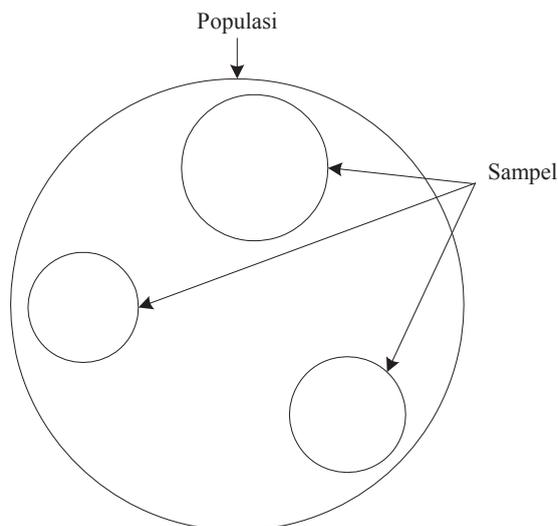
Secara ringkas Walpole et al. (2013) ringkas hubungan antara statistika inferensia dan probabilitas seperti pada Gambar 1.1.



Gambar 1.1: Hubungan antara statistika inferensia dan probabilitas

1.2. Populasi dan sampel

Dalam melakukan penelitian, seringkali berhadapan dengan jumlah data yang banyak, atau penentuan waktu kapan yang tepat dalam mengambil data. Mengambil data keseluruhan, yang dalam statistika dinamakan sebagai sensus, seringkali menjadi masalah jika dana dan waktu tidak mencukupi. Oleh karena itu, diambil sebagian data dari keseluruhan data. Ini yang kemudian dalam statistika dikenal dengan populasi dan sampel. **Populasi** adalah seluruh obyek yang menjadi target dalam pengamatan. **Sampel** adalah bagian dari populasi. Dalam hal ini, karakteristik sampel dan populasi adalah sama. Pembedanya hanya pada jumlah di mana jumlah populasi lebih besar dari sampel. Dalam bentuk sederhana, populasi digambarkan seperti dalam Gambar 1.2.



Gambar 1.2: Populasi dan sampel

Sebagai contoh dari populasi dan sampel adalah sebagai berikut. Akan dilakukan studi terhadap pelaku kon-

sumsi produk ramah lingkungan pada masyarakat di Kota A yang pendidikannya minimal adalah sarjana. Dalam studi ini, populasinya adalah semua penduduk di Kota A yang memiliki pendidikan sarjana atau lebih tinggi dari sarjana. Di Kota A terdapat 6.754 penduduk yang memiliki pendidikan sarjana atau lebih tinggi dari sarjana. Mereka semua adalah populasi dari studi yang akan dilakukan. Dari 6.754 penduduk tersebut akan diambil sampel sejumlah 650. Ke-650 sampel ini memiliki sifat yang sama dengan populasi, yaitu penduduk Kota A yang pendidikannya sarjana atau lebih tinggi dari sarjana.

Contoh lainnya, apabila ingin dilakukan penelitian terhadap perempuan yang sarjana atau lebih tinggi dari sarjana, maka populasinya adalah semua perempuan yang pendidikannya minimal sarjana di Kota A adalah sebagai populasinya. Misalnya, ada 2.450 perempuan di Kota A yang menempuh pendidikan sarjana atau lebih. Maka itulah populasinya, bukan lagi 6.754 seperti contoh sebelumnya. Dari populasi 2.450 tersebut kemudian diambil sampel.

Pengambilan sampel dilakukan karena beberapa alasan yang mendasari. Cochran (1991) menyebutkan ada empat alasan melakukan pengambilan sampel yaitu untuk mengurangi biaya, waktu pengambilan data lebih cepat, cakupan informasinya bisa lebih besar, dan tingkat ketelitiannya lebih besar. Pengambilan sampel mampu mengurangi biaya karena biaya pengambilan data yang dalam populasi yang seharusnya besar dapat dikurangi jumlahnya. Jumlah data yang lebih kecil berarti biaya yang dibutuhkan untuk mengambil data lebih kecil. Dalam hal waktu, mengambil data sampel, mengambil data sampel sejumlah satu juta tentu akan lebih singkat waktunya dibandingkan mengambil data populasi yang berjumlah 200 juta. Meski jumlah data yang diambil sebagai sampel lebih sedikit, dengan teknik peng-

ambilan sampel yang benar, data sampel terbukti telah mampu memberi informasi hampir sama dengan populasi dengan tingkat ketelitian yang besar pula. Sebagai contoh di sini adalah metode *quick count*, yang merupakan contoh dari teknik pengambilan sampel, mampu memprediksi peluang seseorang terpilih dalam suatu kontestasi politik.

Namun, untuk mendapatkan informasi dari sampel agar mampu menggambarkan populasi, dibutuhkan teknik-teknik sampel yang tepat supaya data yang diambil mampu menggambarkan kondisi kenyataan di lapangan. Data yang berkualitas adalah data yang mampu menggambarkan kenyataan di lapangan. Bukan data yang sesuai dengan keinginan peneliti. Untuk itu, perlu mengetahui teknik-teknik pengambilan data yang benar.

Teknik pengambilan sampel ada dua yaitu teknik pengambilan sampel non probabilitas dan probabilitas. Dalam statistik, teknik pengambilan sampel yang akan digunakan adalah teknik pengambilan sampel probabilitas. Maka, dalam buku ini hanya akan dijelaskan empat jenis teknik pengambilan sampel probabilitas.

Teknik probabilitas dijelaskan di sini karena data-data yang menggunakan analisis teknik-teknik ini yang akan banyak digunakan dalam statistika. Selain itu, ada keterkaitannya dengan *sampling error* dan bias dalam sampling. Bias dalam sampling yang terjadi karena keberpihakan peneliti dalam mengambil kelompok sampel tertentu harus dihindari dalam analisis statistika. Bias dalam pengambilan sampel misalnya, ketika survey politik hanya dipilih kelompok tertentu yang cenderung mendukungnya untuk dijadikan sampel. Tujuannya keberpihakan dalam pengambilan sampel cara ini adalah untuk mendapatkan data yang sesuai kemauannya yang akan digunakan untuk mempengaruhi orang dalam memilih. Pengambilan sampel

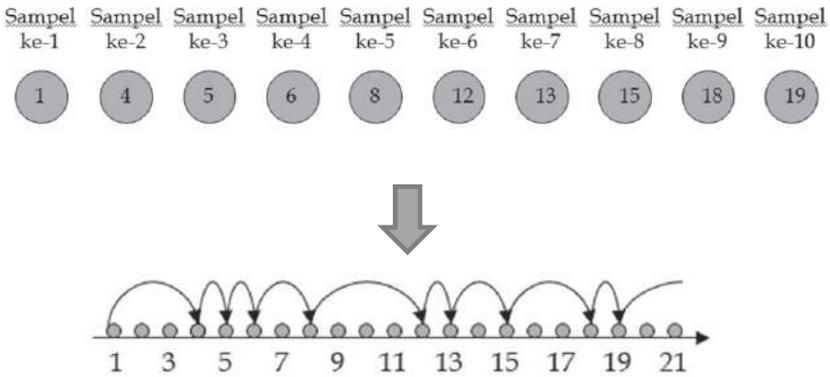
yang menyebabkan hasil yang bias seperti ini juga sering dilakukan untuk tujuan promosi produk. Kelompok tertentu diambil sebagai sampel, namun dalam kesimpulan dinyatakan sebagai keterwakilan semua kelompok. Inilah yang menyebabkan bias dalam pengambilan sampel.

Dalam statistika, *random sampling* atau acak adalah cara untuk menghindari bias dalam pengambilan sampel (Leekley, 2010). *Random sampling* ini adalah bagian dari teknik pengambilan sampel probabilistik. Oleh karenanya, hanya teknik-teknik pengambilan sampel dengan cara *random* atau acak yang mengikuti konsep probabilitas akan dibahas dalam buku ini.

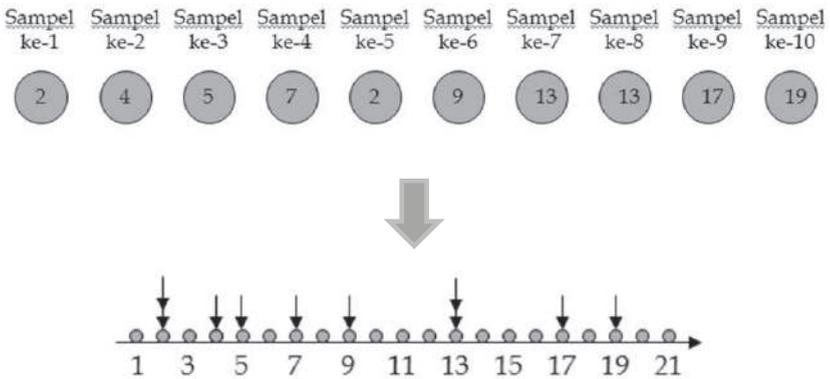
1.2.1. Teknik Pengambilan Sampel Acak Sederhana

Simple random sampling (pengambilan sampel acak sederhana) adalah teknik pengambilan sampel yang mendasar. Dalam teknik ini, setiap elemen dalam sampel mempunyai kesempatan yang sama untuk dipilih menjadi sampel. Teknik pengambilan sampel secara acak dilakukan dengan mengasumsikan bahwa karakteristik populasi adalah sama.

Dalam teknik ini terdapat dua jenis, tanpa pengembalian dan dengan pengembalian. Teknik pengambilan sampling acak dengan pengembalian adalah, setelah data sampel tersebut terpilih, dikembalikan lagi untuk dapat dipilih lagi menjadi sampel. Jadi data tersebut dapat terpilih beberapa kali menjadi sampel. Sementara, teknik pengambilan sampel acak tanpa pengembalian adalah setelah data terambil menjadi sampel tidak dikembalikan lagi sehingga hanya memiliki sekali kesempatan untuk menjadi sampel. Sebagai ilustrasi teknik pengambilan sampel acak sederhana tanpa pengembalian ditunjukkan pada Gambar 1.3 dan ilustrasi teknik pengambilan sampel acak sederhana dengan pengembalian ditunjukkan pada Gambar 1.4.



Gambar 1.3: Contoh pengambilan sampel acak sederhana tanpa pengembalian.



Gambar 1.4: Contoh pengambilan sampel acak sederhana dengan pengembalian.

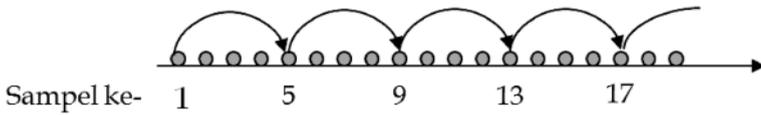
Dalam contoh sampel dengan pengembalian pada Gambar 1.3, data yang terambil adalah data ke-1, ke-4, ke-5, ke-6, ke-8, ke-12, ke-13, ke-15, ke-18, dan ke-19. Berbeda dengan teknik pengambilan sampel acak sederhana dengan pengembalian. Sampel yang pernah terpilih, dikembalikan lagi sehingga ada kemungkinan diambil kembali untuk menjadi sampel kembali. Sedangkan, dalam teknik pengambilan sampel dengan pengembalian seperti pada Gambar 1.4 nampak ada dua angka yang terambil dua kali yaitu angka

2 terambil pada sampel ke-1 dan ke-5, dan angka 13 yang terambil saat pengambilan sampel ke-7 dan ke-8. Bisa saja, dalam waktu yang berbeda menghasilkan angka acak yang berbeda. Tentu saja, akan menghasilkan hasil yang berbeda dari hasil acak sebelumnya. Hal tersebut bukan masalah karena sifatnya acak yang berarti tidak ada keberpihakan. Akan jadi masalah jika dalam pengambilan data adalah tidak acak karena dimungkinkan ada 'keberpihakan' yang tidak menggambarkan kenyataan dari populasi.

Memperoleh bilangan acak dapat dilakukan dengan beberapa cara. Cara sederhana untuk membangkitkan data acak adalah dengan menggunakan lotre yang diberi nomor dan kemudian diacak. Angka yang keluar itulah yang akan dijadikan sebagai sampel. Selain itu, membangkitkan angka acak dapat dilakukan dengan menggunakan tabel acak. Cara selanjutnya dalam membangkitkan angka acak adalah dengan bantuan *software* komputer.

1.2.2. Teknik Pengambilan Sampel Sistematis

Systematic sampling (pengambilan sampel sistematis) adalah teknik mengambil sampel dengan cara memilih elemen sampel dari interval tertentu. Teknik ini dipilih dengan asumsi karakteristik populasi adalah homogen. Teknik pengambilan sampel ini adalah alternatif dari teknik pengambilan sampel acak sederhana. Biasanya, teknik ini dilakukan dalam penelitian pada penduduk di lokasi tertentu, pelanggan di suatu perusahaan jasa yang sedang mengantre, dan pemeriksaan pada produk-produk di pabrik. Teknik dalam mengambil sampel diilustrasikan dalam Gambar 1.5. Dalam gambar tersebut, terlihat pengambilan sampel dilakukan pada interval 4 secara berurutan. Karena berurutan dengan interval tertentu inilah maka dinamakan pengambilan sampel sistematis.



Gambar 1.5: Ilustrasi pengambilan sampel dengan teknik sistematis

Contoh dari teknik pengambilan sampel sistematis misalnya kita akan mengambil data perilaku konsumsi ramah lingkungan pada ibu-ibu di Desa Makmur Selalu. Jumlah ibu rumah tangga di Desa Makmur selalu adalah 1000 ibu rumah tangga. Dari sejumlah ibu rumah tangga tersebut, akan diambil sampel sebanyak 250 ibu rumah tangga.

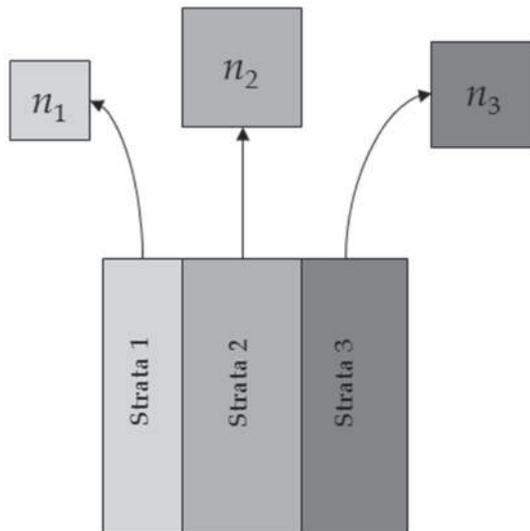
Pada kasus tersebut, jumlah populasi adalah 1000. Jika akan melakukan pengambilan sampel sejumlah 250 ibu rumah tangga, maka intervalnya dihitung dari jumlah populasi dibagi dengan jumlah sampel yang diinginkan, sehingga $1000/250 = 4$. Cara mengambil sampelnya adalah diawali dari rumah tangga paling ujung, dilakukan pengambilan sampel pada setiap interval 4 rumah. Misal, rumah pertama dijadikan sampel. Sampel kedua adalah rumah kelima, dan seterusnya seperti yang ditunjukkan dalam Gambar 1.5 sampai didapatkan 250 sampel.

1.2.3. Teknik Pengambilan Sampel Stratifikasi

Dalam suatu populasi, sering ditemukan kelompok yang memungkinkan akan menghasilkan hasil yang berbeda disebabkan karena perbedaan strata. Dalam hal ini, mengambil sampel secara random dan menganggap populasi heterogen tentu tidaklah tepat. *Stratified sampling* (pengambilan sampel stratifikasi) dilakukan dengan mengelompokkan populasi ke dalam sub kelompok berdasarkan stratanya. Setelah mendapatkan kelompok berdasarkan

strata, kemudian ditentukan jumlah sampel di setiap strata. Penentuan jumlah sampel bisa proporsional berdasarkan jumlah populasi dalam strata, atau dibuat jumlah sampel setiap strata adalah sama. Teknik ini bertujuan untuk membuat agar karakteristik populasi lebih homogen dan semua strata terwakili.

Sebagai contoh, jika akan melakukan penelitian terhadap konsumen produk kendaraan roda dua yang bertenaga listrik. Konsumen dikelompokkan berdasarkan usianya: remaja, dewasa, dan tua. Dari kelompok ini kemudian diambil sampel dengan jumlah yang sesuai proporsinya atau bisa dibuat sama antar strata. Sebut saja pada strata 1 jumlah sampelnya adalah n_1 , n_2 untuk jumlah sampel strata ke-2, dan n_3 untuk jumlah sampel strata ke-3. Total dari sampel di tiap strata adalah sama dengan total sampel yang diambil. Pengambilan sampel pada setiap strata boleh dilakukan secara acak karena pada strata tersebut sudah homogen. Ilustrasinya ditunjukkan pada Gambar 1.6.

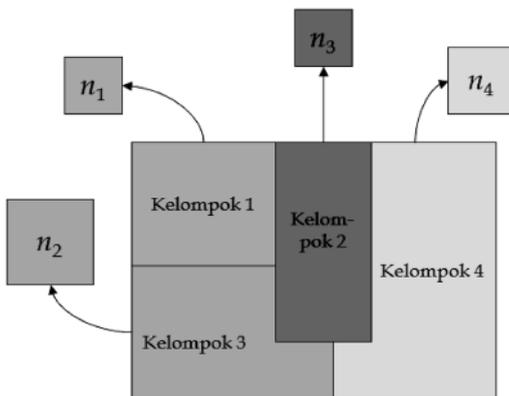


Gambar 1.6: Ilustrasi pengambilan sampel dengan teknik stratifikasi.

1.2.4. Teknik Pengambilan Sampel Kelompok

Cluster sampling (pengambilan sampel kelompok), jika dalam teknik pengambilan sampel stratifikasi, pengelompokan berdasarkan strata, dalam teknik pengambilan sampel kelompok dibedakan berdasarkan kelompoknya. Kelompok ini bisa berupa wilayah, suku, agama, kelompok politik dan lain-lain. Tujuan mengelompokkan di sini adalah sama dengan teknik pengambilan sampel stratifikasi yaitu agar semua kelompok dapat terwakili dalam sampel sehingga data menggambarkan populasi yang sebenarnya.

Ilustrasi teknik pengambilan sampel kelompok (*cluster*) ditunjukkan pada Gambar 1.7. Misalnya, akan dilakukan penelitian tentang kepuasan masyarakat atas layanan pemerintah kelompok berdasarkan organisasi kemasyarakatan. Sebut saja Kelompok 1, 2, 3, dan 4. Setiap kelompok memiliki sejumlah anggota yang berbeda jumlahnya. Untuk mengambil data agar mampu mewakili aspirasi semua kelompok, setiap kelompok dibedakan dan diambil sampel dengan jumlah sampel yang proporsional, sebut saja n_1 untuk jumlah sampel *cluster* ke-1, n_2 jumlah sampel *cluster* ke-2, n_3 jumlah sampel *cluster* ke-3, dan n_4 jumlah sampel *cluster* ke-4. Jumlah sampel di setiap *cluster* bisa menggunakan jumlah sampel yang sama sama untuk semua *cluster* atau

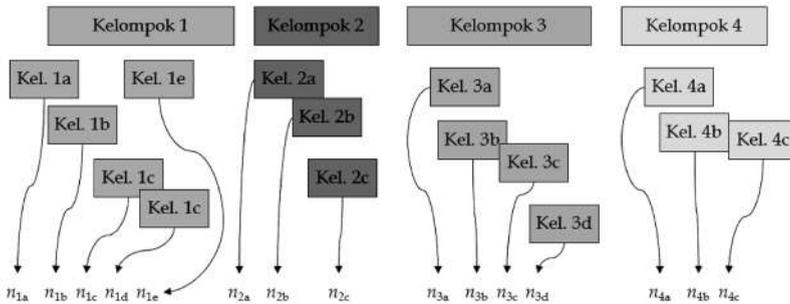


bisa menggunakan prinsip proporsional yang berdasarkan jumlah populasi di setiap *cluster*.

Gambar 1.7: Ilustrasi pengambilan sampel dengan teknik *cluster*.

1.2.5. Teknik Pengambilan Sampel *Multistage*

Teknik pengambilan *multistage* atau multi tahap adalah pengembangan dari teknik pengambilan sampel kelompok. Jika dalam teknik pengambilan sampel kelompok, pengelompokan hanya dilakukan sekali, dalam teknik pengambilan sampel *multistage* pengelompokan dilakukan secara bertahap. Sebagai ilustrasi dari teknik pengambilan sampel *multistage*, disediakan Gambar 1.8. Pada gambar tersebut terlihat ada empat kelompok besar. Masing-masing kelompok memiliki kelompok kecil yang bisa sama dan bisa berbeda. Dalam ilustrasi, tiap kelompok memiliki sub-grup yang berbeda. Misalnya, Kelompok 1, memiliki 5 sub-kelompok, sementara Kelompok 4 memiliki 4 sub-kelompok.



Gambar 1.8: Ilustrasi pengambilan sampel dengan teknik *multistage*.

Setelah populasi dikelompokkan dan dibuat sub-kelompok yang lebih homogen, maka ditentukan jumlah sampel yang akan diambil di setiap sub-kelompok tersebut. Jumlah sampel di setiap sub-kelompok bisa sama dengan sub kelompok lain, atau bisa menggunakan prinsip proporsional dimana jumlah sampel adalah relative terhadap jumlah populasi pada sub-kelompok tersebut. Setelah mendapatkan jumlah sampel, pengambilan sampelnya boleh dilakukan

secara acak atau sistematis karena pada sub-kelompok karakteristik populasinya sudah homogen.

1.2.6. Bagaimana Menggunakan Teknik Pengambilan

Kelima teknik tersebut tidak harus berdiri sendiri, namun dapat dikombinasikan. Misalnya, dalam kasus penelitian untuk mendapatkan aspirasi warga masyarakat. Pengambilan sampel bisa dilakukan dengan teknik *multistage* yang aturannya menggunakan kombinasi *cluster-stratified-random sampling* atau *cluster-stratified-systematic sampling*. Ini dilakukan untuk mendapatkan aspirasi dari kelompok berdasarkan strata dan kelompok yang dilakukan seacak mungkin agar mampu menggambarkan populasi yang sebenarnya.

Setiap pengambilan sampel dimungkinkan terdapat kesalahan. Kesalahan yang dimaksud di sini bukan kesalahan personal yang dilakukan oleh pengambil data, tetapi kesalahan dalam pengambilan sampel yang mengakibatkan perbedaan dalam sampel (Leekley, 2010). Oleh karena itu perlu diukur *sampling error* untuk melihat seberapa bagus suatu sampel mendekati karakteristik dari populasi. Salkind (2008) mengatakan *sampling error* adalah perbedaan antara nilai statistik sampel dan parameter populasi. Ini berarti, jika *sampling error*-nya besar maka bisa dikatakan bahwa kondisi hasil prediksi dengan menggunakan sampel tersebut jauh dari kondisi populasi sebenarnya.

Teknik pengambil sampel sudah dilakukan oleh para ilmuwan sejak dulu. Desrosières (1998) mencatat pada awal abad ke-19 metode pengambilan sampel acak diterapkan. Dia juga mencatat pada akhir abad ke-18, Laplace sudah menggunakan teknik pengambilan sampel. Namun, selama lebih dari seabad hingga akhir abad ke-19 teknik survey baru digunakan lagi secara informal dengan meng-

gunakan sampel probabilitas. Penggunaan teknik pengambilan sampling menggunakan teknik stratifikasi baru digunakan secara formal oleh Neyman pada tahun 1934. Sampai sekarang, penggunaan data dengan sampel sudah diterapkan di banyak sektor dan telah terbukti dapat digunakan menyelesaikan permasalahan-permasalahan yang ada.

1.3. Data dan Pengambilan Keputusan dalam Statistika

Data adalah elemen penting dalam statistika. Data yang akan digunakan dalam statistika dapat diperoleh melalui percobaan, pengamatan, survey, atau bisa diperoleh melalui data simulasi yang merupakan hasil pemodelan dari kejadian nyata. Data-data tersebut juga bisa didapat dari sampel yang merupakan bagian dari populasi atau boleh menggunakan data keseluruhan atau populasi. Namun, yang perlu dicatat di sini adalah unsur variabilitas dalam data adalah penentu penggunaan statistika.

Data monoton, tentu mudah dianalisis. Lain halnya jika data yang didapat bervariasi. Misalnya, mengapa orang pendidikannya tinggi ternyata ada yang miskin, biasa saja, dan ada yang kaya. Sebaliknya, ada orang yang tidak berpendidikan ternyata kaya bahkan melebihi yang pendidikannya tinggi. Di saat dunia mendorong peningkatan pendidikan untuk meningkatkan kualitas ekonomi masyarakat karena pendidikan dianggap mampu mengentaskan kemiskinan, ketidakpastian ekonomi seseorang pasca mendapat menyelesaikan pendidikannya menjadi teka-teki tersendiri yang tentu menjadi bagian dalam analisis sosial-ekonomi masyarakat.

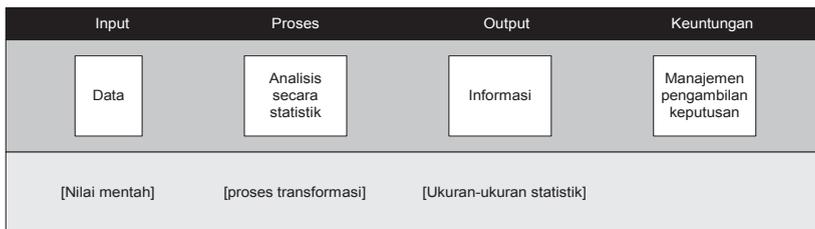
Data bisa dalam berbagai bentuk yang nanti akan dibahas lebih mendalam pada Bab 2. Dalam hal ini, data tidak

harus selalu berupa angka, namun juga bisa berupa huruf atau kata atau kalimat. Contoh data adalah penjualan produk, data jumlah pasien yang terjangkit penyakit Covid-19, data nilai tukar mata uang dari waktu ke waktu, dan seterusnya. Sementara contoh data yang bukan berupa angka adalah nama kabupaten yang laporan keuangannya Wajar Tanpa Pengecualian, nama orang yang lolos seleksi masuk perguruan tinggi, merk produk *handphone* yang beredar di masyarakat, dan lain-lain. Data-data tersebut hanya berupa data yang tidak akan memberi informasi yang lebih banyak jika tidak diolah dan dianalisis. Misalnya, data jumlah kematian pasien Covid-19 dari hari ke hari. Tanpa dilakukan pengolahan lebih jauh, data tersebut hanya memberi informasi sederhana kenaikan dan penurunan jumlah pasien Covid-19 dari hari ke hari. Tanpa didukung data tambahan seperti daftar riwayat penyakit pasien atau tingkat keparahan Covid-19, analisis penyebab kematian pasien Covid-19 atau analisis mengapa terjadi kenaikan dan penurunan dari hari ke hari tidak akan bisa dilakukan.

Data-data tersebut juga harus dibuat seakurat mungkin. Pengukuran data sangat menentukan kualitas data. Sebagai contoh, mengukur kedalaman sungai tentu akan lebih akurat jika menggunakan alat ukur standar yang satuannya bisa menggunakan centi meter (cm) atau meter (m). Menggunakan istilah, tidak dalam, sedang-sedang saja, dan dalam pada pengukuran kedalaman sungai tentu sangat subyektif. Persepsi tidak dalam bagi seseorang akan berbeda dengan persepsi tidak dalam bagi orang lain. Contoh lain yang biasanya digunakan dalam mengukur variabel-variabel yang sifatnya subyektif atau non latent, seperti perilaku konsumsi seseorang. Untuk mendapatkan ukuran yang sesuai untuk perilaku konsumsi tentu harus mengacu pada teori yang sudah ada. Dari teori tersebut kemudian dijabarkan

menjadi suatu indikator-indikator perilaku konsumsi. Indikator itulah yang dipakai untuk mengukur perilaku konsumsi.

Data yang berkualitas yang didukung oleh data lain akan sangat membantu dalam mendapatkan informasi yang lebih komprehensif. Tentu, sebelumnya harus dilakukan analisis statistika yang tepat yang sesuai dengan karakteristik data. Analisis data yang tidak sesuai dengan skala dan tipe data, akan menghasilkan informasi yang salah. Namun, jika analisis data yang dilakukan sudah tepat, maka akan menghasilkan informasi yang sangat berguna dan dapat menyelesaikan masalah. Dalam hal ini berarti informasi yang baik (*output*) tidak hanya didapat dari data yang akurat, valid, dan berkualitas (*input*), tetapi juga dipengaruhi cara menganalisis data (proses) yang harus tepat pula. Informasi yang diperoleh tersebut dapat dijadikan masukan untuk pengambil keputusan. Alur peran statistika yang dapat dimanfaatkan untuk dalam sistem pengambilan keputusan ditunjukkan pada Gambar 1.9.



Gambar 1.9: Peranan statistika dalam sistem manajemen pengambilan keputusan (Wegner, 2013)

Beberapa kegiatan yang membutuhkan statistika untuk sistem pengambilan keputusan dalam dunia bisnis dan industri antara lain adalah penentuan metode pemasaran yang tepat yang mampu memberi nilai yang tinggi terhadap perusahaan, menentukan lokasi toko yang diminati

oleh pelanggan, apa penyebab suatu produk diminati oleh pelanggan, dan mengapa terjadi produk cacat dalam produksi. Sedangkan kegiatan dalam bidang pendidikan yang membutuhkan statistika dalam mengambil keputusan seperti menentukan metode pembelajaran yang efektif bagi mahasiswa dan melakukan evaluasi atas pendidikan yang telah dilakukan. Penerapan statistika dalam bidang kesehatan di masyarakat adalah mencari solusi atas metode efektif agar cuci tangan pakai sabun menjadi kebiasaan pada masyarakat.

1.4. Istilah-istilah dalam Statistika

Sebelum melangkah menuju bab selanjutnya, di akhir Bab Pendahuluan ini akan dibuatkan beberapa istilah-istilah penting yang sering digunakan dalam statistika. Beberapa istilah tersebut adalah sebagai berikut:

- **Variabel acak** (*random variable*) adalah segala atribut dimana data akan dikumpulkan dan dianalisis. Contoh variabel acak adalah jarak perjalanan karyawan dari rumah ke kantor (dalam kilometer atau dalam waktu), durasi waktu yang diperlukan staf untuk menyelesaikan satu pekerjaan (dalam bentuk menit atau jam), kedalaman banjir di beberapa titik di suatu wilayah (dalam centimeter atau meter), nama merk HP (data: Apple, Samsung, Oppo)
- **Data** adalah nilai aktual atau hasil dari variabel.
- **Unit sampel** adalah obyek yang akan diukur, dihitung, atau diamati berkaitan dengan variabel acak yang akan dipelajari..
- **Populasi** adalah sekumpulan semua nilai data yang mungkin yang ada dalam variabel acak yang akan dipelajari.

- **Sampel** adalah bagian dari populasi.
- **Parameter populasi** adalah ukuran yang menggambarkan karakteristik dari populasi. Misal, rata-rata populasi.
- **Statistik sampel** adalah ukuran yang menggambarkan karakteristik sampel. Misalnya, rata-rata sampel.

BAB 2.

PENGUKURAN DATA

DATA adalah elemen penting dalam statistika. Data inilah yang akan diolah dengan teknik-teknik analisis statistik yang tepat sesuai dengan karakteristik data untuk mendapatkan informasi yang tepat. Analisis data mengikuti karakteristik data. Oleh karena itu, memahami data dan macam-macamnya sangat diperlukan supaya analisis yang dilakukan tidak salah. Dalam bab ini akan dijelaskan jenis-jenis pengukuran data, tipe data, dan cara menampilkan data secara sederhana agar mudah dibaca. Dari materi pengukuran, tipe dan menampilkan data ini, diharapkan di akhir bab ini pembaca:

- Mampu mengetahui, membedakan, dan mempraktekkan level pengukuran data dalam statistika.
- Mampu mengetahui, membedakan, dan mempraktekkan tipe-tipe data dalam statistika.

2.1. Level Pengukuran Data

Pengukuran data menjadi hal yang sangat penting dalam statistika. Pengukuran data yang tepat akan menentukan hasil dari analisis. Sebaliknya, jika pengukurannya tidak akurat, tentu akan dihasilkan output yang tidak akan menghasilkan apapun. Sebagai contoh, ketika akan mengukur data kedalaman banjir. Orang bisa melakukan

pengukuran kedalaman sungai dengan menggunakan standar tubuh manusia seperti: se-mata kaki, selutut, se-pinggang, sebahu dan seterusnya. Pengukuran dengan cara ini tentu akan sangat bias karena tinggi manusia berbeda-beda. Selutut anak balita berarti lebih pendek dari selutut manusia dewasa. Akan berbeda jika alat ukur yang digunakan adalah alat ukur standar. Misalnya dengan menggunakan meteran. Jadi, kedalamannya bisa jelas terukur seperti setengah meter, semeter, dua meter, dan seterusnya. Data ukuran se-mata kaki, selutut, sepinggang dan seterusnya berbeda dengan ukuran satu meter, dua meter, tiga meter dan seterusnya. Dalam sub-bab ini akan dijelaskan beberapa skala data.

Secara umum, skala data dikelompokkan menjadi empat yaitu skala data nominal, skala data ordinal, skala data interval, dan skala data rasio. Penjelasan masing-masing skala data adalah sebagai berikut:

- **Skala data nominal**

Data nominal adalah variabel data yang dilabelkan dalam bentuk nama. Data ini hanya bisa dibedakan saja berdasarkan label atau namanya. Contoh dari data ini adalah sebagai berikut:

Nama kota : Lampung, Jakarta, Bandung, Yogyakarta, Surakarta, Sragen, Surabaya, Jember, Banjarmasin, Denpasar

Nama orang : Indah, Diah, Dani, Sumardi, Sujali, Zahra, Denandra, Jumali, Paidi, Bandi

Nama Universitas : Universitas Indonesia, Universitas Gadjah Mada, Institut Pertanian Bogor, Institut Sepuluh Nopember,

Universitas Islam Negeri Syarif
Hidayatullah, Sekolah Tinggi
Agama Islam Negeri Surakarta,
Universitas Islam Indonesia

Warna : Merah, kuning, biru, hijau, oranye,
ungu, abu-abu

Macam-macam : Handphone, tablet, komputer,
gadget

Nama platform : Facebook, Twitter, Instagram,
media sosial Whatsapp

Jenis mata : Rupiah, Dolar, Poundsterling, Yen
uang

Skala ini jika akan diolah datanya akan diubah ke kode numerik. Misalnya, untuk jenis mata uang, Rupiah diberi kode 1, Dolar diberi kode 2, Poundsterling diberi kode 3, dan seterusnya. Karena tidak ada urutan, kode tersebut bisa diganti angkanya. Namun, seiring perkembangan *software* statistika, skala ini tidak perlu diberi kode numerik.

- **Skala data ordinal**

Skala ordinal juga dalam bentuk label atau nama seperti halnya skala data nominal. Namun dalam skala ordinal terdapat urutan tingkatan mulai dari terkecil sampai terbesar atau sebaliknya dari besar ke kecil. Contohnya adalah sebagai berikut:

Tingkat : Taman Kanak-kanak, Sekolah
sekolahan Dasar, Sekolah Menengah Pertama,
Sekolah Menengah Atas, Diploma,
Sarjana, Magister, Doktor

Pangkat dalam militer	: Prajurit, Kopral, Sersan, Letnan, Kapten, Mayor, Kolonel, Jendral
Tingkatan Pemerintahan	: Rukun Tetangga, Rukun Warga, Dusun, Desa, Kecamatan, Kabupaten/Kotamadya, Propinsi, Negara
Jenis SIM	: SIM A, SIM B, SIM, C
Jenjang usia	: Bayi, anak, remaja, dewasa, orangtua
Juara	: Juara 1, Juara 2, Juara 3, Harapan 1, Harapan 2, Harapan 3

Seperti halnya skala data nominal, dalam pengolahan datanya, skala ini juga diberi kode numerik. Namun karena sifatnya berurutan, untuk memudahkan dalam analisa data, kode numeriknya dibuat berurutan. Misal, SIM A diberi kode 1, SIM B diberi kode 2, dan seterusnya atau menggunakan kebalikannya yaitu SIM C diberi kode 1, SIM B diberi kode 2, dan seterusnya.

- **Skala data interval**

Lain halnya dengan skala data nominal dan ordinal yang bersifat kualitatif, skala data interval adalah skala kuantitatif. Dalam skala ini, setiap perbedaan antar data memiliki interval yang sama. Contoh data berskala interval adalah sebagai berikut:

Suhu udara	: -10 °C, -5 °C, 5 °C, 10 °C
Garis lintang	: 0° LS, 15° LS, 20° LU, 45° LU

Pada suhu udara, jarak dari -10°C ke -8°C adalah sama dengan jarak antara 1°C ke 3°C. Namun, angka 0°C di sini bukan berarti tidak ada suhu. Ada suhu udara namun skalanya adalah 0°C.

- **Skala data rasio**

Skala data rasio adalah skala data yang sifatnya paling lengkap yaitu, antar data bisa dibedakan, memiliki jenjang urutan, memiliki interval yang sama dan jika ada nilai 0 berarti menyatakan benar-benar nol yang tidak ada nilainya. Contoh dari data rasio adalah sebagai berikut:

Jumlah tabungan : Rp 100.000,-; Rp 500.000,-; Rp
nasabah/bulan 1.000.000,-;

Jumlah penduduk : 10.000 jiwa, 50.000 jiwa, 60.000
di desa jiwa, 100.000 jiwa

Nilai Matematika : 70, 75, 80, 81, 72, 93, 85

Jumlah mahasiswa : 35 anak, 37 anak, 41 anak, 40
per kelas anak, 33 anak

Usia : 1 bulan, 12 bulan, 1 tahun, 5
tahun, 25 tahun

Dalam memudahkan membedakan keempat skala tersebut, Tabel 2.1 dapat digunakan sebagai petunjuk. Dalam tabel tersebut, nampak jelas cara membedakan empat skala data: nominal, ordinal, interval, dan rasio berdasarkan apakah data dapat dibedakan, memiliki urutan, memiliki interval yang sama, dan mempunyai nilai nol mutlak. Seperti yang telah disebutkan sebelumnya, skala data rasio memiliki nilai nol mutlak yang artinya jika terdapat angka nol pada variabel yang menggunakan skala data rasio maka nilai nol tersebut adalah nol yang sebenarnya yang berarti kosong. Berbeda dengan skala data interval. Pada contoh di atas adalah suhu udara, garis bujur dan garis lintang. Nilai 0° C dan 0° LU bukan berarti tidak ada suhu dan tidak ada garis lintang utara, tapi itu hanya lambang untuk menyatakan suatu skala suhu dan garis lintang.

Tabel 2.1: Ciri setiap skala data

	Dapat dibedakan	Mempunyai urutan	Interval sama	Mempunyai nilai nol mutlak
Nominal	√			
Ordinal	√	√		
Interval	√	√	√	
Rasio	√	√	√	√

Skala data nominal dan skala data ordinal seringkali berupa data yang berbentuk kata atau beberapa kata. Data ini kemudian dibuat kode data dalam bentuk angka supaya bisa diolah. Sebagai contoh, untuk data nama orang seperti Indah, Diah, Dani, Sumardi, Sujali, Zahra diberi kode 1, 2, 3, 4, 5 dan 6. Kode angka dalam data nominal ini tidak bisa dijumlahkan antara satu dengan lainnya. Sebagai ilustrasi, Nana yang diberi kode 1 dan Diah yang diberi kode 2. Kedua data ini tidak bisa dijumlahkan menjadi $Nana + Diah = 1 + 2 = 3$ yang notabene kode 3 adalah kode dari Yani. Oleh karena kode angka pada skala nominal ini tidak bisa dijumlahkan, maka data ini disebut sebagai data yang bersifat kualitatif. Data yang bersifat kualitatif ini juga berlaku pada data data ordinal.

Lain halnya dengan skala data interval dan rasio. Data ini berupa angka, maka angka tersebut adalah angka yang bersifat kuantitatif. Penjumlahan dari angka ini bermakna sama dengan jumlahan angka tersebut. Misalnya, tabungan Rp 100.000,- ditambahkan dengan tabungan Rp 500.000,- sama dengan Rp 600.000,-. Nilai Rp 600.000,- adalah nilai tabungan sebenarnya.

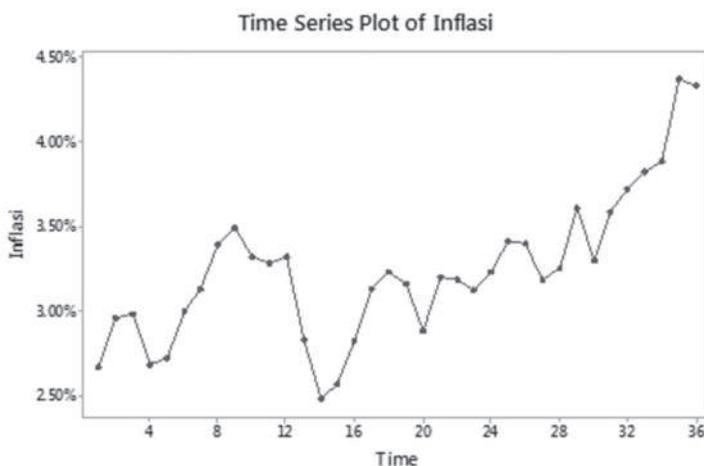
2.2. Tipe Data

Jika sub-bab sebelumnya adalah skala data. Pada sub-bab ini dibahas tipe data. Dalam Gujarati (2003) tipe data

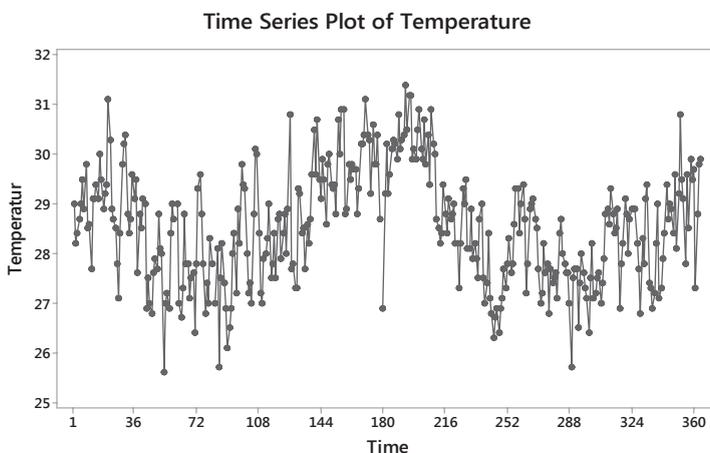
ada tiga yaitu data *cross section*, data *time series*, data *pooled*, dan data panel. Penjelasan dari ketiga tipe data tersebut adalah sebagai berikut:

2.2.1. Data Time Series

Data *time series* adalah data hasil pengamatan yang nilainya diambil pada waktu berbeda yang rentang waktunya adalah sama. Bisa dikatakan, data *time series* adalah sejarah suatu kasus yang direkam dalam bentuk data yang direkam dari waktu ke waktu. Rentang waktu pembiliran data bisa dalam satuan detik, menit, jam, hari, minggu, bulanan, tahun, dan seterusnya. Bisa juga menggunakan rentang interval waktu seperti setiap 12 jam, setiap 3 hari, setiap 2 minggu, dan seterusnya. Contoh-contoh dari data *time series* adalah data laporan cuaca tiap hari, inflasi bulanan, tingkat pengangguran per bulan, pengamatan pertumbuhan tumbuhan atau binatang, catatan jumlah cacat produk per hari.



Gambar 2.1: Data inflasi bulanan Indonesia sejak Mei 2017 hingga April 2020



Gambar 2.2: Data temperatur udara rata-rata harian di Kota Semarang pada 8 Mei 2019 hingga 5 Mei 2020.

Dalam buku ini, contoh gambar data *time series* tersebut ditunjukkan pada Gambar 2.1 dan 2.2. Data tersebut adalah inflasi bulanan di Indonesia pada bulan Mei 2017 hingga April 2020 (Gambar 2.1) dan data suhu udara rata-rata harian di Kota Semarang pada 8 Mei 2019 hingga 5 Mei 2020 (Gambar 2.2). Data *time series* memiliki ciri data yang terpengaruh oleh periode sebelumnya. Bisa dikatakan, data *time series* adalah data yang dependan terhadap periode waktu sebelumnya.

2.2.2. Data Cross Section

Data *cross section* adalah data satu atau lebih variabel yang dikumpulkan dalam suatu waktu (Bryman, 2012; Gujarati, 2003). Sebagai contoh, data survey penduduk pada Tahun 2020. Dalam survey tersebut ditanyakan beberapa variabel seperti jumlah keluarga, pendidikan keluarga, status tempat tinggal, pekerjaan. Contoh data *cross section* yang diambil dari Laporan Badan Pusat Statistik (BPS) Kota Surakarta ditunjukkan pada Tabel 2.2. Dalam tabel tersebut

adalah hasil survey BPS Kota Surakarta pada tahun 2016. Data yang diambil contoh pada buku ini adalah data jumlah penduduk, kepadatan penduduk, angka kelahiran dan angka kematian pada tahun 2016.

Tabel 2.2: Data kependudukan Kota Surakarta tahun 2016 (BPS Kota Surakarta, 2017)

Kecamatan	Jumlah penduduk	Kepadatan penduduk per km ²	Angka kelahiran	Angka kematian
Laweyan	88.614	10.257,91	1369	853
Serengan	44.950	14.073,26	670	539
Pasar Kliwon	76.474	15.881,79	721	206
Jebres	142.152	11.298,22	1927	1141
Banjarsari	161.981	10,936,53	1756	831

2.2.3. Data Pooled

Data *pooled* adalah gabungan antara data *cross section* dan *time series*. Dalam hal ini berarti, data *cross section* yang diambil secara periodik, misalnya harian, mingguan, bulanan, tahunan, adalah data *pooled*. Contoh datanya diambil dari BPS Kota Surakarta yaitu, data jumlah penduduk dan angka kematian yang terekam pada tahun 2016 dan 2017. Data tersebut diambil dari Kota Surakarta dalam angka 2016 (BPS Kota Surakarta, 2017) dan Kota Surakarta dalam Angka 2017 (BPS Kota Surakarta, 2018).

Tabel 2.3: Data kependudukan Kota Surakarta tahun 2016 dan 2017 (BPS Kota Surakarta, 2017, 2018)

Kecamatan	Jumlah penduduk		Angka kematian	
	2016	2017	2016	2017
Laweyan	88.614	88.947	853	800
Serengan	44.950	45.119	539	404

Kecamatan	Jumlah penduduk		Angka kematian	
	2016	2017	2016	2017
Pasar Kliwon	76.474	76.760	206	678
Jebres	142.152	142.686	1141	1007
Banjarsari	161.981	162.590	831	1331

BAB 3.

TEORI PROBABILITAS

TEORI probabilitas adalah dasar dari statistika. Teori ini sangat penting untuk disampaikan karena basis dari statistika adalah adanya ketidakpastian yang dinyatakan dalam teori-teori probabilitas. Pada bab ini akan diberikan elemen-elemen teori probabilitas (peluang), bagaimana menghitung titik sampel, dan juga diberikan beberapa distribusi probabilitas dari variabel diskrit dan kontinyu yang banyak digunakan dalam analisis-analisis statistika. Melalui bab ini diharapkan pembaca mampu:

- Mengetahui dan memahami elemen-elemen probabilitas seperti ruang sampel dan kejadian.
- Mengetahui, membedakan, menghitung, dan menerapkan kejadian sederhana dan majemuk seperti irisan, gabungan, komplementer, *mutually exclusive*, dan *collectively exchausive*.
- Mengetahui, membedakan, menghitung, dan mempraktikkan perhitungan permutasi dan kombinasi.
- Mengetahui membedakan, menghitung, dan mempraktikkan bagaimana menghitung probabilitas.

3.1. Elemen-elemen dalam Probabilitas

Kejadian-kejadian yang ada di sekitar kita dan kita alami sehari-hari bisa dikatakan adalah kejadian yang *random*

(acak). Seringkali, kejadian tersebut tidak dapat diprediksi meski beberapa dapat diprediksi. Kejadian-kejadian yang dapat diprediksi maupun tidak dapat diprediksi tersebut bisa dikatakan sebagai *outcome* atau luaran dari suatu eksperimen atau pengamatan (Dekking et al. 2005). Seberapa mungkin suatu kejadian akan terjadi yang dinyatakan dengan angka numerik disebut sebagai probabilitas.

Elemen-elemen dalam teori peluang ada tiga yaitu ruang sampel, kejadian, dan ukuran probabilitas. Ruang sampel dan kejadian akan dibahas dalam sub-bab ini, sedangkan ukuran akan dibahas dalam Sub-bab 3.3.

3.1.1. Ruang Sampel (*Sample Space*)

Ruang sampel adalah semua kemungkinan yang mungkin akan terjadi dalam suatu pengamatan atau eksperimen. Ruang sampel bisa dikatakan sebagai hasil (*outcome*) dari suatu eksperimen yang diamati. Ruang sampel dinyatakan dengan notasi Ω .

Contoh 3.1:

Seseorang akan datang ke pasar pada hari apa. Orang tersebut akan datang pada ke tujuh hari yaitu, Senin, Selasa, Rabu, Kamis, Jumat, Sabtu, dan Minggu. Bagaimana ruang sampelnya?

Penyelesaian Contoh 3.1:

Ruang sampelnya adalah: $\Omega = \{\text{Senin, Selasa, Rabu, Kamis, Jumat, Sabtu, Minggu}\}$

Contoh 3.2:

Jumlah pasien yang akan mendatangi Dr X di Poli Kandungan. Bagaimana ruang sampelnya?

Penyelesaian Contoh 3.2:

Dalam kasus ini, ruang sampelnya yang mungkin adalah adalah

$$\Omega = \{0,1,2,3,4,5,\dots\}$$

Angka 0, 1, 2, 3, dan seterusnya adalah bilangan bulat (I) yang menunjukkan jumlah pasien yang akan mendatangi Dr X di Poli Kandungan yang jumlah maksimalnya tidak bisa ditentukan. Ruang sampel di atas bisa ditulis dalam bentuk lain yaitu $\Omega = \{x \mid x \in I\}$, dimana $I = \{\text{Bilangan bulat}\}$.

Contoh 3.3:

Dalam suatu toko hanya menjual masker dengan harga per box Rp 40.000,-. Pembeli hanya dibolehkan membeli maksimal sepuluh box. Bagaimana ruang sampelnya:

Penyelesaian Contoh 3.3:

Besaran transaksi yang akan dilakukan oleh pembeli di toko tersebut ruang sampelnya adalah

$$\Omega = \{\text{Rp } 0, \text{Rp } 40.000, \text{Rp } 80.000, \text{Rp } 120.000, \dots, \text{Rp } 400.000\}$$

Pembeli yang masuk ke toko tersebut bisa saja memutuskan tidak membeli, sehingga kemungkinan tidak terjadi transaksi apapun yang berarti nilai transaksinya adalah Rp 0,-. Karena dibatasi hanya boleh membeli 10 box, maka transaksi maksimal untuk membeli masker adalah Rp 400.000,. Oleh karenanya, transaksi dalam ruang sampel tersebut transaksi yang mungkin akan terjadi berkisar antara Rp 0,- sampai Rp 400.000,- dengan transaksi yang dilakukan adalah kelipatan dari Rp 40.000.

3.1.2. Kejadian (*event*)

Suatu kejadian adalah himpunan bagian dari ruang sampel. Dimungkinkan juga, suatu kejadian adalah sama

dengan ruang sampel. Contoh dari kejadian adalah sebagai berikut:

Contoh 3.4:

Dr X praktik di hari Senin hingga Jumat. Selama dua minggu, jumlah pasien yang mengunjungi Dr X di poli kandungan adalah sebagai berikut:

$$\Omega = \{15, 9, 10, 8, 13, 12, 6, 7, 11, 16\}$$

A adalah kejadian di mana jumlah pasien yang mengunjungi Dr X di Poli Kandungan jumlahnya di atas 10. Bagaimana kejadian A?

Penyelesaian Contoh 3.4:

$$A = \{15, 13, 12, 11, 16\}.$$

Suatu kejadian bisa sederhana dan bisa majemuk. Kejadian sederhana terjadi jika hanya mengandung satu unsur ruang sampel. Sementara, kejadian majemuk bisa terjadi karena gabungan dari beberapa kejadian. A adalah contoh kejadian sederhana. Contoh lain, B adalah kejadian dimana jumlah pengunjung dokter X adalah jumlahnya mengikuti bilangan prima. Dari sini berarti,

$$B = \{13, 7, 11\}$$

Sementara, C adalah banyaknya pengunjung Dr X yang jumlahnya 10 atau kurang. Ini berarti,

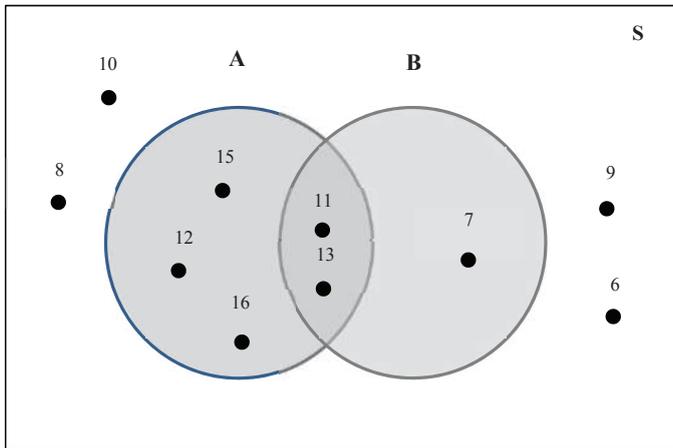
$$C = \{9, 10, 8, 6, 7\}$$

A, B, dan C jika berdiri sendiri adalah kejadian sederhana. Namun, akan menjadi kejadian majemuk jika ada kejadian dimana jumlah pengunjung dr X melebihi 10 orang dan jumlahnya mengikuti bilangan prima. Di sini berarti anggota himpunannya adalah 11 dan 13. Kejadian ini bisa

dinamakan sebagai **irisan** atau perpotongan. Selain irisan, dalam kejadian majemuk terdapat **gabungan**. Berikut ini akan dijelaskan secara detail mengenai beberapa kejadian dalam suatu ilmu peluang.

- **Irisan (*intersection*)**

Irisan A dan B adalah jika A dan B terjadi bersamaan. Irisan antara A dan B dinotasikan dengan $A \cap B$. Ilustrasinya ditunjukkan dalam Gambar 3.1.



Gambar 3.1: Irisan A dan B

Saat $A = \{15, 13, 12, 11, 16\}$ dan $B = \{13, 7, 11\}$. Maka $A \cap B = \{11, 13\}$. Contoh irisan dalam kasus real ditunjukkan pada Contoh 3.5.

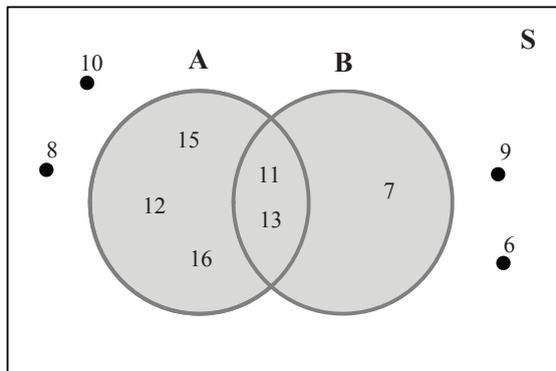
Contoh 3.5:

Pengusaha yang berdomisili di Solo Raya yang sekaligus adalah muslim. Kejadian pertama adalah pengusaha yang berdomisili di Solo Raya, kejadian kedua adalah muslim yang berdomisili di Solo Raya. Dua kejadian ini sangat dimungkinkan dijalankan bersama sehingga ada di antara pengusaha tersebut yang muslim.

Karena tidak semua pengusaha adalah muslim dan tidak semua muslim adalah pengusaha, jumlah anggota yang masuk dalam irisan dari pengusaha yang sekaligus muslim bisa jadi lebih kecil dari jumlah pengusaha atau muslim. Bisa juga, semua pengusaha adalah muslim, atau semua muslim adalah pengusaha.

- **Gabungan**

Gabungan A dan B berarti terjadi kejadian A atau B atau keduanya. Gabungan dinyatakan dengan $A \cup B$. Dalam bentuk diagram venn, gabungan antara A dan B digambarkan dalam Gambar 3.2. Jadi, $A \cup B = \{7, 11, 12, 13, 15, 16\}$. Contoh realnya ditunjukkan pada Contoh 3.6.



Gambar 3.2: Gabungan A dan B

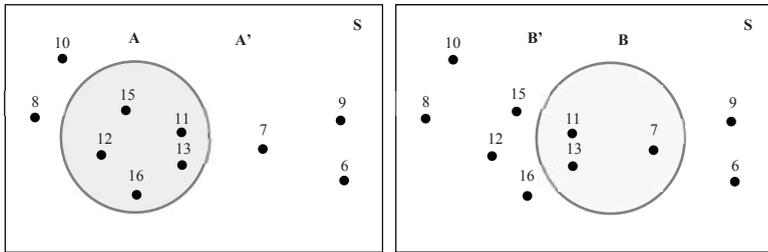
Contoh 3.6:

Kejadian pertama adalah konsumen mesin cuci di semua toko elektronika di Kota Surakarta. Kejadian kedua adalah pembeli alat-alat elektronika di toko Maju Jaya yang berada di Surakarta. Gabungan dari kejadian ini adalah orang yang membeli mesin cuci di toko elektroika di kota Surakarta atau konsumen alat-alat eletronika di toko Maju Jaya. Jumlah total gabungan

dua kejadian ini tentu lebih besar dari jumlah masing-masing kejadian.

- **Komplementer (A')**

Komplementer adalah semua kejadian A dalam ruang sampel (Ω) yang bukan termasuk A. Dalam contoh kasus di sini, kejadian $A' = \{9, 8, 6, 7\}$, $B' = \{15, 9, 10, 8, 12, 6, 16\}$. Gambar 3.3 adalah dua diagram *venn* untuk A' dan B' . Contoh realnya ditunjukkan pada Contoh 3.7.



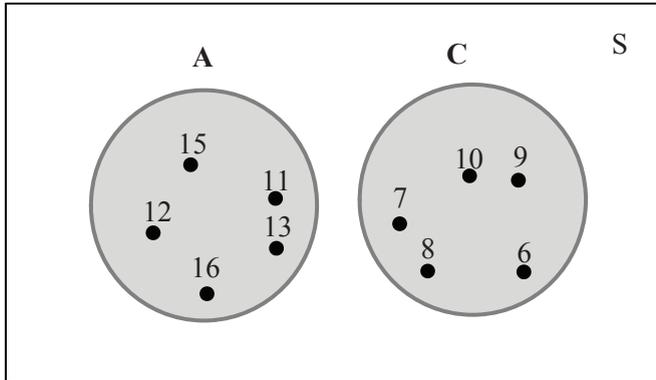
Gambar 3.3: Himpunan komplementer

Contoh 3.7:

Kejadian komplementer ini adalah kejadian orang-orang terkena penyakit Covid 19 di Kota Surakarta. Komplementer dari kejadian ini adalah orang-orang yang tidak terkena penyakit Covid 19.

- **Mutually exclusive**

Kejadian A dan B adalah *mutually exclusive* jika kejadian yang satu tidak memungkinkan kejadian yang lain. Ini bisa dinyatakan bahwa $A \cap B = \emptyset$. Kejadian *mutually exclusive* disebut juga *disjoint* dimana dua kejadian benar-benar terpisah atau tidak memiliki irisan. Pada kasus ini akan diambil contoh A dan C.



Gambar 3.4: Kejadian A dan C adalah mutually exclusive

Contoh 3.8:

Orang-orang pada hari Senin tanggal 20 April 2020 berada di Kota Surakarta selama 24 jam. Kejadian kedua adalah kejadian di mana orang-orang pada hari Senin tanggal 20 April 2020 berada di Jakarta selama 24 jam. Pada kasus ini, tidak mungkin satu orang berada dalam dua lokasi pada saat yang sama. Maka, ini adalah contoh *mutually exclusive*.

- **Collectively exhaustive**

Dua kejadian adalah *mutually exclusive* jika gabungan antara dua kejadian membentuk ruang sampel atau $A \cup B = \Omega$. Kejadian A dan C pada contoh kasus di buku ini adalah contoh *collectively exhaustive* karena $A \cup C = \Omega$.

Contoh 3.9:

Kejadian di mana murid SMPN 1 Sukakaya adalah perempuan. Kejadian berikutnya adalah kejadian di mana murid SMPN 1 Sukakaya adalah laki-laki. Apabila dua kejadian ini digabungkan akan menjadi ruang sampel yaitu murid SMPN 1 Sukakaya.

3.2. Menghitung titik sampel: Permutasi dan Kombinasi

Dalam suatu ruang sampel kadangkala diperoleh beberapa kejadian. Kejadian-kejadian tersebut bisa kejadian yang sederhana atau kejadian kompleks. Misalnya, dalam suatu tempat terdapat dari beberapa anggota. Dari anggota tersebut diambil sebagian. Namun, susunan dan urutannya dari anggota yang terpilih tersebut tidak dapat dipertukarkan. Kejadian ini menggunakan teknik **permutasi** dimana susunan dari semua atau sebagian dari sekumpulan obyek yang urutan susunannya tidak bisa dipertukarkan. Permutasi dari n obyek yang diambil sebanyak r adalah:

$$P_r^n = \frac{n!}{(n-r)!} \quad (3.1)$$

dimana $n =$ banyaknya obyek

$r =$ banyaknya komponen dalam n yang akan diambil

Contoh 3.10:

Dalam perusahaan terdapat 5 karyawan pemasaran yaitu A, B, C, D, dan E. Setiap bulan perusahaan memberikan tiga jenis penghargaan kepada karyawan pemasaran berdasarkan penjualan tertinggi, terinovatif dalam memasarkan, dan penambahan jaringan pasar terbanyak. Setiap karyawan hanya boleh mendapatkan satu penghargaan. Bagaimana kemungkinan dan berapa susunan karyawan yang akan menerima penghargaan?

Penyelesaian Contoh 3.10:

Maka susunan karyawan yang akan menerima penghargaan ditunjukkan pada Tabel 3.1. Total susunan tiga karyawan yang mendapat penghargaan ada 60 pasang. Penerima yang terdiri dari susunan ABE, tidak sama dengan BAE,

BEA, EBA, atau EAB. Itu karena jenis penghargaannya berbeda. ABE berarti A mendapatkan penghargaan karena penjualannya tertinggi, B mendapat penghargaan terinovatif, E mendapatkan penghargaan penambahan jaringan pasar terbanyak. Pada susunan lain, misalnya BEA maka B mendapatkan penghargaan atas penjualan tertinggi, E mendapatkan penghargaan terinovatif, dan A mendapatkan penghargaan karena mendapatkan penambahan jaringan pasar terbanyak. Berapa banyak susunan karyawan yang mungkin akan mendapatkan tiga penghargaan itu dinamakan permutasi.

Tabel 3.1: Tabel susunan pasangan karyawan yang mungkin mendapat tiga penghargaan

No.	Susunan	No.	Susunan	No.	Susunan	No.	Susunan
1.	ABC	16.	BCA	31.	CDA	46.	DEA
2.	ABD	17.	BCD	32.	CDB	47.	DEB
3.	ABE	18.	BCE	33.	CDE	48.	DEC
4.	ACB	19.	BDA	34.	CEA	49.	EAB
5.	ACD	20.	BDC	35.	CEB	50.	EAC
6.	ACE	21.	BDE	36.	CED	51.	EAD
7.	ADB	22.	BEA	37.	DAB	52.	EBA
8.	ADC	23.	BEC	38.	DAC	53.	EBC
9.	ADE	24.	BED	39.	DAE	54.	EBD
10.	AEB	25.	CAB	40.	DBA	55.	ECA
11.	AEC	26.	CAD	41.	DBC	56.	ECB
12.	AED	27.	CAE	42.	DBE	57.	ECD
13.	BAC	28.	CBA	43.	DCA	58.	EDA
14.	BAD	29.	CBD	44.	DCB	59.	EDB
15.	BAE	30.	CBE	45.	DCE	60.	EDC

Jika menggunakan Persamaan 3.1, banyaknya susunan dari karyawan yang mendapatkan penghargaan ada:

$$\begin{aligned}
 P_r^n &= \frac{n!}{(n-r)!} \\
 &= \frac{5!}{(5-3)!} \\
 &= \frac{5!}{2!} \\
 &= \frac{(5)(4)(3)2!}{2!} \\
 &= 60
 \end{aligned}$$

Enampuluh susunan ini adalah ruang sampel yang mungkin akan terjadi.

Dalam suatu kejadian, kadang bisa saja dalam sekelompok susunan dianggap sama saja. A berada di urutan pertama, dianggap sama dengan A jika berada di urutan kedua atau ketiga dan seterusnya. Kejadian ini dinamakan **kombinasi** dimana persamaannya mengikuti persamaan berikut:

$$C_r^n = \frac{n!}{(n-r)!r!} \tag{3.2}$$

dimana n = banyaknya obyek

r = banyaknya komponen dalam n yang akan diambil

Contoh 3.11.

Misalnya pada Contoh 3.10, karyawan pemasaran yang terdiri dari 5 anggota yaitu A, B, C, D, dan E. Akan dikirimkan tiga orang untuk dikirim ke kantor pusat untuk mendapatkan pelatihan pemasaran. Bagaimana dan berapa susunan tim yang mungkin akan dikirim?

Penyelesaian Contoh 3.11:

Maka, kemungkinan anggota tim yang akan dikirim ditunjukkan pada Tabel 3.2.

Tabel 3.2: Susunan karyawan yang mungkin akan dikirim ke kantor pusat

No.	Susunan	No.	Susunan
1.	ABC	6.	ADE
2.	ABD	7.	BCD
3.	ABE	8.	BCE
4.	ACD	9.	BDE
5.	ACE	10.	CDE

Terdapat 10 pasang tim karyawan yang akan dikirim untuk pelatihan di kantor pusat. Dengan menggunakan rumus pada Persamaan 3.2, penyelesaian kombinasi pada permasalahan karyawan pemasaran di atas adalah sebagai berikut:

$$\begin{aligned} C_r^n &= \frac{n!}{(n-r)!r!} \\ &= \frac{5!}{(5-3)!3!} \\ &= \frac{5!}{2!3!} \\ &= \frac{(5)(4)(3!)}{2!3!} \\ &= \frac{(5)(4)}{(2)(1)} \\ &= 10 \end{aligned}$$

Di sini, ABC dianggap sama dengan BCA, CBA, BAC bagaimanapun urutannya, tiga orang ABC itulah yang akan berangkat.

3.3. Probabilitas

Probabilitas atau peluang adalah berapa kemungkinan suatu kejadian terjadi. Fungsi probabilitas P dalam ruang sampel terbatas Ω yang ditugaskan pada setiap kejadian A dalam Ω sejumlah $P(A)$ yang berada pada rentang nilai numerik $0 \leq P(A) \leq 1$. Probabilitas dari ruang sampel adalah 1, atau dinotasikan dengan $P(\Omega) = 1$ dan probabilitas dari himpunan kosong adalah 0, atau $P(\emptyset) = 0$.

Di sub bab sebelumnya dibahas bahwa kejadian itu bisa terjadi secara individual (kejadian sederhana) dan bisa terjadi dari beberapa kejadian (majemuk). Kejadian majemuk meungkinan terjadi irisan diantaranya, gabungan di antaranya, kejadian komplementer dan seterusnya. Probabilitas beberapa kejadian majemuk tersebut mengikuti beberapa aturan berikut ini:

- Dilakukan pengamatan terhadap N hasil. Jika suatu kejadian menghasilkan n hasil, maka probabilitas dari A adalah

$$P(A) = \frac{n}{N} \quad (3.3)$$

- Penjumlahan dalam probabilitas

$$P(A \cap B) = P(A) + P(B) \quad (3.4)$$

- Perkalian dalam probabilitas

$$P(A \cup B) = P(A) \times P(B) \quad (3.5)$$

- Jika A dan B adalah kejadian sembarang, peluang akan terjadi kejadian A atau B adalah:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (3.6)$$

- Jika ada tiga kejadian, sebut saja A , B , dan C adalah kejadian sembarang, peluang akan terjadi kejadian A atau B atau C adalah:

$$\begin{aligned}
 P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) \\
 &\quad - P(B \cap C) \\
 &\quad + P(A \cap B \cap C)
 \end{aligned} \tag{3.7}$$

- Jika A dan B kejadian yang *mutually exclusive*, maka

$$P(A \cup B) = P(A) + P(B) \tag{3.8}$$

- Jika A_1, A_2, \dots, A_n *mutually exclusive*, maka

$$P(A_1 \cup A_2 \cup \dots \cup A_n) = P(A_1) + P(A_2) + \dots + P(A_n) \tag{3.9}$$

- Kejadian A dan A' adalah kejadian yang komplementer maka

$$P(A') = 1 - P(A) \tag{3.10}$$

Berikut adalah contoh soal untuk penerapan teori-teori probabilitas di atas.

Contoh 3.12:

Dari hasil studi, diketahui terdapat 124 mahasiswa di program studi X, 82 adalah mahasiswi dan 42 adalah mahasiswa. Mahasiswa yang aktif di himpunan program studi X ada sebanyak 46 mahasiswa. Mereka terdiri dari 18 mahasiswa perempuan dan sisanya adalah mahasiswa laki-laki. Selain itu, terdapat 36 mahasiswa dari program studi X yang aktif di organisasi mahasiswa yang berhubungan dengan masalah lingkungan. Diantara mereka, 14 mahasiswa adalah perempuan dan 22 mahasiswa adalah laki-laki. Dari mahasiswa yang aktif di organisasi mahasiswa yang berhubungan dengan lingkungan, 16 diantaranya adalah juga anggota himpunan program studi X.

- Probabilitas bahwa mahasiswa di program studi X adalah wanita?
- Probabilitas bahwa mahasiswa di program studi X adalah laki-laki?

- c. Probabilitas bahwa mahasiswa di program studi X aktif di himpunan program studi?
- d. Probabilitas bahwa mahasiswa di program studi X aktif organisasi lingkungan?
- e. Probabilitas bahwa mahasiswa di program studi X aktif di himpunan program studi dan dia adalah wanita?
- f. Dari sini dapat diketahui probabilitas mahasiswa program studi X yang aktif di himpunan program studi atau wanita?
- g. Probabilitas mahasiswa program studi X yang aktif di dua organisasi, himpunan program studi dan organisasi lingkungan?
- h. Jika ingin diketahui berapa probabilitas mahasiswa program studi X yang aktif di organisasi himpunan mahasiswa program studi atau organisasi lingkungan?

Penyelesaian Contoh 3.12:

Dari soal tersebut didapat informasi bahwa:

- S adalah ruang sampel yang jumlahnya $(n(S)) = 124$
- W adalah kejadian terdapat mahasiswi di program studi yang jumlahnya $(n(W)) = 82$
- L adalah kejadian terdapat mahasiswa di program studi yang jumlahnya $(n(L)) = 42$
- A adalah kejadian mahasiswa yang aktif di himpunan program studi, jumlahnya $(n(A)) = 46$
- Jumlah mahasiswa yang aktif di himpunan dan wanita $(n(A \cap W)) = 18$
- Jumlah mahasiswa yang aktif di himpunan dan laki-laki $(n(A \cap L)) = 28$
- B adalah kejadian mahasiswa di program studi X yang aktif di organisasi lingkungan tingkat institut yang jumlahnya $(n(B)) = 36$

- Jumlah mahasiswa yang aktif di organisasi lingkungan yang wanita ($n(B \cap W) = 14$)
- Jumlah mahasiswa yang aktif di organisasi lingkungan yang laki-laki ($n(B \cap L) = 22$)
- Jumlah mahasiswa yang aktif di himpunan sekaligus aktif di organisasi lingkungan ($n(A \cap B) = 16$)

Maka,

- a. Probabilitas bahwa mahasiswa di program studi X adalah wanita dengan menggunakan Persamaan 3.3 adalah:

$$P(W) = n(W)/n(S) = 82/124 = 0,661$$

- b. Probabilitas bahwa mahasiswa di program studi X adalah laki-laki (dengan Persamaan 3.3):

$$P(L) = n(L)/n(S) = 42/124 = 0,339.$$

- c. Probabilitas bahwa mahasiswa di program studi X aktif di himpunan program studi (dengan Persamaan 3.3):

$$P(A) = n(A)/n(S) = 54/124 = 0,435$$

- d. Probabilitas bahwa mahasiswa di program studi X aktif organisasi lingkungan (dengan Persamaan 3.3):

$$P(B) = n(B)/n(S) = 36/124 = 0,290$$

- e. Probabilitas bahwa mahasiswa di program studi X aktif di himpunan program studi dan dia adalah wanita adalah:

$$P(A \cap W) = n(A \cap W)/n(S) = 18/124 = 0,145$$

- f. Dari sini dapat diketahui probabilitas mahasiswa program studi X yang aktif di himpunan program studi atau wanita menggunakan Persamaan 3.6 adalah:

$$\begin{aligned} P(A \cup W) &= P(A) + P(W) - P(A \cap W) \\ &= 54/124 + 82/124 - 18/124 \\ &= 0,952 \end{aligned}$$

- g. Probabilitas mahasiswa program studi X yang aktif di dua organisasi, himpunan program studi dan organisasi lingkungan adalah:

$$P(A \cap B) = n(A \cap B) / n(S) = 16 / 124 = 0,129$$

- h. Jika ingin diketahui berapa probabilitas mahasiswa program studi X yang aktif di organisasi himpunan mahasiswa program studi atau organisasi lingkungan, dengan menggunakan Persamaan 3.4 maka:

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 54/124 + 36/124 - 16/124 \\ &= 0,597 \end{aligned}$$

Contoh 3.13.

Pada contoh ini diambil dari Contoh 3.10. Terdapat 5 karyawan pemasaran, sebut saja A, B, C, D, dan E. Setiap bulan akan diberi 3 penghargaan. Dengan menggunakan prinsip permutasi, diperoleh susunan karyawan pemasaran yang memungkinkan mendapat tiga penghargaan setiap bulan adalah sebanyak 60 susunan. Di waktu yang lain, akan ada pelatihan di kantor pusat yang mengharuskan mengirim tiga orang karyawan pemasaran. Dari sini didapat kemungkinan 10 pasang karyawan yang akan dikirim ke kantor pusat. Dari sini ada beberapa permasalahan:

- Pada kasus pemberian penghargaan kepada karyawan, berapa probabilitas D akan mendapat penghargaan terinovatif?
- Masih pada kasus pemberian penghargaan kepada karyawan, berapa kemungkinan A, C dan D mendapat penghargaan?
- Pada masalah pengiriman karyawan ke kantor pusat untuk pelatihan, ada berapa kemungkinan D dan E akan bersama-sama dikirim pelatihan?

Penyelesaian Contoh 3.13:

Untuk menyelesaikan ketiga permasalahan di atas, yang harus diketahui adalah pada kasus pemberian penghargaan kepada karyawan yang menggunakan konsep permutasi. Jumlah ruang sampel adalah (N_p) 60. Sementara pada kasus pengiriman karyawan ke kantor pusat yang menggunakan konsep kombinasi, jumlah ruang sampelnya adalah (N_c) 10.

- a. Dicari probabilitas (peluang) D akan mendapat penghargaan terinovatif. Di sini berarti posisi D adalah di urutan ke-2. Dari Tabel 3.1, diketahui D berada di urutan kedua pada sampel ke-7 (ADB), ke-8 (ADC), ke-9 (ADE), ke-19 (BDA), ke-20 (BDC), ke-21 (BDE), ke-31 (CDA), ke-32 (CDB), ke-33 (CDE), ke-58 (EDA), ke-59 (EDB), dan ke-60 (EDC). Terdapat 12 kemungkinan. Maka, probabilitas D akan mendapat penghargaan terinovatif dalam pemasaran adalah $P(D_2) = 12/60 = 0,20$.
- b. Dicari probabilitas A, C, dan D mendapat penghargaan bersama. Di sini diasumsikan mendapat penghargaan apapun, asalkan dalam suatu kesempatan mereka bertiga mendapat penghargaan bersama-sama. Dari Tabel 3.1, sampel yang terdapat A, C, dan D adalah sampel ke-5 (ACD), ke-8 (ADC), ke-26 (CAD), ke-31 (CDA), ke-38 (DAC), ke-44 (DCA). Berarti ada 6 pasang. Maka, probabilitas A, C, D akan mendapat penghargaan secara bersama-sama adalah $P(ACD) = 6/60 = 0,10$
- c. Dicari probabilitas D dan E bersama-sama dikirim ke pelatihan di kantor pusat. Dari Tabel 3.2, didapat sampel ke-6 (ADE), sampel ke-9 (BDE), dan sampel ke-10 (CDE) didapat D dan E bersama-sama. Ada tiga sampel dari 10 sampel. Maka, probabilitas D dan E bersama-sama dikirim ke pelatihan di kantor pusat adalah sebesar $3/10 = 0,30$.

Pada suatu kejadian bisa bergantung pada terjadinya kejadian yang lain. Ini dinamakan **kejadian bersyarat** (*conditional probability*). Probabilitas bersyarat A dengan mengasumsikan B telah terjadi dinyatakan dengan $P(A|B)$ yang dinyatakan:

$$P(A|B) = \frac{P(A \cap B)}{P(B)} \text{ dimana } P(B) > 0 \quad (3.11)$$

Atau

$$P(A \cap B) = P(A | B)P(B) \quad (3.12)$$

Untuk kasus $P(A \cap B \cap C)$, maka

$$P(A \cap B \cap C) = P(A | B \cap C)P(B | C)P(C) \quad (3.13)$$

atau

$$P(A \cap B \cap C) = P(A \cap B | C)P(C) \quad (3.14)$$

Pada kasus dimana dua kejadian saling bebas, maka Persamaan 3.11 menjadi

$$P(A | B) = P(A) \quad (3.15)$$

Contoh 3.14

Digunakan Contoh 3.12 di atas. Berapa probabilitas mahasiswa yang aktif di organisasi lingkungan dengan syarat dia adalah laki-laki?

Penyelesaian Contoh 3.14:

Permasalahan ini dinotasikan dengan $P(B | L)$ yang diperoleh dengan menggunakan Persamaan 3.11:

$$\begin{aligned} P(B|L) &= \frac{P(B \cap L)}{P(L)} = \frac{n(B \cap L)/n(S)}{n(L)/n(S)} \\ &= \frac{22/124}{42/124} \\ &= 0,524 \end{aligned}$$

BAB 4.

MERINGKAS DAN MENAMPILKAN DATA

SETELAH mendapatkan data, data perlu diringkas untuk dapat memberikan informasi yang informatif. Meringkas data dapat membantu pembaca dalam membaca data. Teknik-teknik dalam meringkas data adalah bagian dari statistika. Dalam Bab ini akan dibahas meringkas data dalam bentuk tabel, grafik dan dalam ukuran-ukuran statistik yang dilambangkan dengan notasi yang lebih ringkas. Oleh karena itu, dari bab ini pembaca diharapkan mampu:

- Membedakan, memahami, dan menerapkan untuk memilih yang terbaik dan tepat dalam meringkas data dalam bentuk tabel dan atau grafik.
- Membedakan, memahami, dan menerapkan meringkas data dengan teknik-teknik statistik menggunakan ukuran pemusatan data.
- Membedakan, memahami, dan menerapkan meringkas data dengan teknik-teknik statistik menggunakan ukuran penyebaran data.

4.1. Meringkas data dalam Tabel dan Grafik

Menurut Clements (1991) dalam menyelesaikan masalah ada tiga tahap yang harus dilakukan yaitu sederhanakan masalah, rangkum, dan visualisasikan. Memvisualisasikan data adalah bagian dari statistika. Membaca data satu per satu seringkali membingungkan pembacanya. Apalagi jika

pembacanya adalah orang awam. Oleh karena itu perlu menampilkan data supaya lebih mudah dibaca. Data statistika dapat ditampilkan dalam bentuk tabel dan grafik. Apalagi jika gambar dianggap mampu menyampaikan informasi melebihi dari ribuan kata (Salkind, 2008). Beberapa jenis tabel dan grafik untuk menampilkan data akan dijelaskan pada sub-bab ini.

4.1.1. Tabel Frekuensi

Tabel frekuensi adalah tabel yang paling populer digunakan dalam statistika. Tabel ini menggambarkan frekuensi data pada kelompok tertentu. Dalam tabel tersebut, terdapat dua kolom yang kolom pertama adalah kelompok data, sedangkan kolom kedua adalah frekuensi dari kelompok data tersebut. Ada beberapa contoh tabel frekuensi yang sering digunakan. Dalam sub-sub bab ini akan diberikan beberapa contoh tabel tersebut.

Contoh 4.1:

Ilustrasi pertama ditunjukkan pada Tabel 4.1. Tabel 4.1 adalah hasil tes akhir pada pelatihan kerja 50 karyawan di sebuah perusahaan. Nilai hasil penilaian pelatihan semua peserta ditunjukkan pada Tabel 4.1. Dari nilai tersebut kemudian diubah dalam Tabel Frekuensi. Dalam membuat Tabel Frekuensi, dapat digunakan *tally* untuk membantu menghitung frekuensi suatu nilai. Hasil akhir dari Tabel Frekuensi ditunjukkan pada Tabel 4.2.

Tabel 4.1: Nilai hasil pelatihan karyawan

No.	Nilai								
1.	81	11.	76	21.	79	31.	71	41.	79
2.	79	12.	81	22.	85	32.	71	42.	76
3.	85	13.	73	23.	73	33.	79	43.	71

No.	Nilai								
4.	76	14.	79	24.	76	34.	85	44.	81
5.	89	15.	81	25.	76	35.	76	45.	81
6.	76	16.	91	26.	71	36.	71	46.	73
7.	73	17.	76	27.	89	37.	79	47.	79
8.	79	18.	81	28.	79	38.	81	48.	71
9.	81	19.	76	29.	85	39.	76	49.	81
10.	76	20.	81	20.	81	40.	71	50.	76

Jika dalam Tabel 4.1, sulit sekali melihat gambaran data secara keseluruhan, namun dalam Tabel Frekuensi di Tabel 4.2 kita bisa melihat jelas ringkasan datanya. Dalam Tabel Frekuensi tersebut nampak jelas bahwa nilai peserta terendah adalah 71 dan nilai tertinggi adalah 91. Peserta yang mendapat nilai tinggi, misalnya 89 dan 91, hanya sedikit. Sementara, peserta yang mendapat nilai terendah, yaitu 71, ada 7 karyawan.

Tabel 4.2: Tabel Frekuensi model pertama yang diringkas dari Tabel 4.1

Nilai setelah pelatihan kerja	Tally	Jumlah peserta
91		1
89		2
85		4
81		11
79		9
76		12
73		4
71		7
Total peserta		50

Dalam beberapa tabel frekuensi, kategori atau kelompok data bisa saja dilakukan dalam rentang kelompok.

Misalkan dalam kasus pelatihan karyawan yang nilainya ditampilkan pada Tabel 4.1, kategori nilainya dibuat rentang. Contohnya ditunjukkan pada Tabel 4.3. Dalam tabel tersebut, rentang nilai dibuat dengan interval 5, sehingga ada 5 kategori nilai yaitu 70-74 hingga 90-94. Karena kelompoknya berbeda dengan Tabel 4.3 yang pengelompokan nilainya dibuat individual, maka frekuensinya juga berbeda. Namun, jumlah total frekuensi tetap sama yaitu 50 peserta.

Tabel 4.3: Tabel Frekuensi model kedua yang diringkas dari Tabel 4.1

Nilai setelah pelatihan kerja	Jumlah peserta
70-74	11
75-79	21
80-84	11
85-89	6
90-94	1
Total peserta	50

Namun, secara teoritis, membuat Tabel Frekuensi memiliki aturan-aturan yang harus diikuti. Salkind (2008) memberikan panduannya sebagai berikut, tahapan ini sekaligus mempraktikkan data membuat tabel frekuensi pada data di Tabel 4.1

- (1) Pilih rentang interval kelas, misalnya kita tentukan 5.
- (2) Pilih interval kelas sehingga mampu mencakup semua data. Untuk menghitung ini adalah, hitung rentang nilai tertinggi dengan nilai terendahnya, kemudian bagi dengan jumlah kelas yang diinginkan. Dalam contoh pada Tabel 4.1, nilai tertinggi adalah 91, sedangkan nilai terendah adalah 71. Rentangnya adalah $range = 91 - 71 = 20$. Nilai 20 ini kemudian dibagi dengan banyaknya

kelas yang diinginkan, misalnya 10 kelas. Sehingga $20/10 = 2,1$. Rentang 2,1 dibulatkan ke angka 3.

- (3) Buat interval kelas diawali dari nilai yang paling rendah. Dalam kasus Tabel 4.1, nilai terendah adalah 71. Sehingga kelas pertama adalah 71-73, kelas kedua adalah 74-76, dan seterusnya.
- (4) Hitung frekuensinya.

Hasil Tabel frekuensi yang mengikuti tahapan di atas ditunjukkan pada Tabel 4.4. Tabel Frekuensi pada Tabel 4.3 dan 5.4 sedikit berbeda, sehingga frekuensinya berbeda.

Tabel 4.4: Tabel Frekuensi dari data pada Tabel 4.1 yang mengikuti tahapan Salkind (2008)

Nilai setelah pelatihan kerja	Jumlah peserta
71-73	11
74-76	12
77-79	9
80-82	11
83-85	4
86-88	2
89-91	1
Total peserta	50

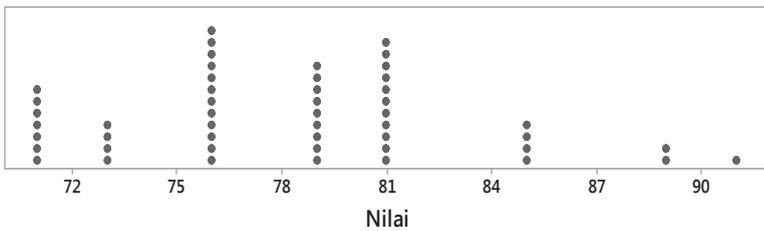
4.1.2. Grafik Data

Dalam menampilkan data, grafik adalah visualisasi data yang mudah dikenali selain tabel. Ada banyak bentuk grafik untuk menampilkan data. Beberapa diantaranya akan dibahas dalam paragraf berikutnya.

- **Dotplot**

Dotplot sudah digunakan sejak 100 tahun yang lalu (Wilkinson, 1999). *Dotplot* adalah bentuk grafik suatu data yang menampilkan frekuensi data pada kelompok

data tertentu yang disimbolkan dalam bentuk *dot*. Kelompok data ditampilkan dalam garis horisontal, sedangkan frekuensi data ditampilkan dalam garis vertikal. Dengan diagram ini, banyaknya *dot* pada kelompok data menunjukkan besaran frekuensi suatu kelompok data tersebut. Dalam *dotplot*, kelompok data bisa berupa data diskrit (kualitatif) atau kontinyu. Sebagai contoh digunakan data pada Tabel 4.1. Hasil pembuatan *dotplot* ditunjukkan dalam Gambar 4.1.



Gambar 4.1: *Dotplot* nilai pelatihan karyawan

Titik-titik dalam *dotplot* pada Gambar 4.1 menggambarkan frekuensi kelompok nilai. Misal, Nilai pertama, dalam hal ini adalah 71, memiliki jumlah dot tersusun ke atas sejumlah 7. Pada nilai 81, jumlah dot yang tersusun ke atas ada 11. Dengan gambar seperti ini, pembaca data akan lebih bisa memahami distribusi data secara visual. Nilai pelatihan yang banyak didapat oleh karyawan dapat dilihat dengan mudah, begitu juga nilai yang tidak banyak didapat oleh karyawan selama pelatihan.

- **Diagram Batang dan Daun (*Steam and leaf diagram*)**
 Diagram batang dan daun adalah kombinasi antara tabel dan grafik. Dalam buku ini akan digunakan contoh pada Tabel 4.1 untuk membuat diagram batang dan daun.

Cara membuat diagram batang dan daun adalah membagi kolom dalam dua kolom yang berisi batang (*steam*) dan daun (*leaf*). Kolom pertama yaitu batang (*steam*) berisi angka desimal pertama, dan kolom kedua (*leaf*) adalah angka desimal kedua. Kemudian masukkan angka ke dalam kelompoknya dan diurutkan. Misalnya, angka pertama adalah 81. Masukkan angka 8 pada kolom batang dan 1 pada kolom daun. Selanjutnya, untuk angka 79, angka 7 dimasukkan ke kolom batang dan 9 ke kolom daun. Begitu seterusnya. Kemudian menghitung jumlah angka dalam daun dan dimasukkan ke kolom frekuensi. Hasilnya dapat dilihat pada Gambar 4.2.

Steam	Leaf	Frekuensi
7	1111111333336666666666666999999999	32
8	111111111111555599	17
9	1	1

Gambar 4.2: Diagram batang dan daun dari data pada Tabel 4.1.

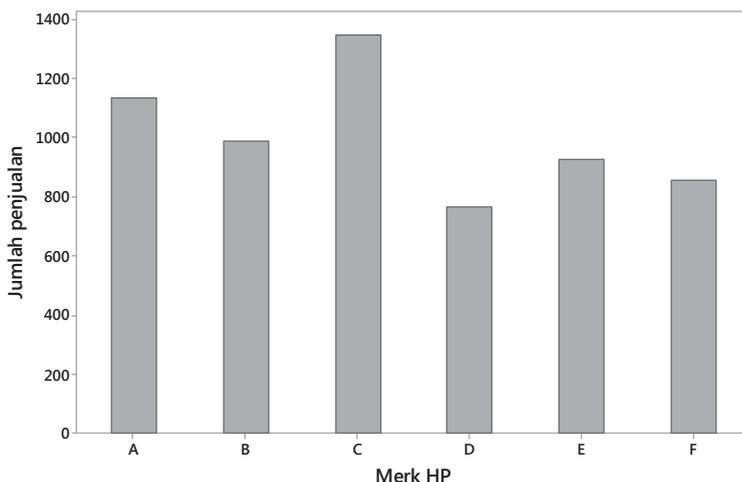
- **Diagram Batang (*barchart*)**

Diagram batang adalah bentuk lain dari *dotplot*. Perbedaannya adalah pada pernyataan frekuensi yang dinyatakan dalam bentuk batang balok dan data kelompok adalah data yang bersifat diskrit atau data kualitatif yang berskala nominal atau ordinal. Lebar balok dalam diagram batang (*barchart*) pada setiap kategori, harus sama (Wegner, 2013). Contoh dari diagram batang dapat dilihat di Gambar 4.3 yang datanya diambil dari Tabel 4.5.

Tabel 4.5: Data penjualan handphone pada Januari 2020

Merk handphone	Penjualan
A	1,132
B	987
C	1,345
D	765
E	924
F	856

Pada Gambar 4.3, garis vertikal menyatakan merk HP dan garis horisontal menyatakan jumlah penjualan. Lebar batang tiap merk adalah sama. Selain itu, batang untuk merk A, B, C, hingga F jaraknya terpisah. Ini untuk menunjukkan bahwa antar kategori memang sifatnya terpisah. Dalam skala data yang berskala nominal atau ordinal, antar kategori tidak sifatnya berbeda dan tidak saling tumpang tindih.

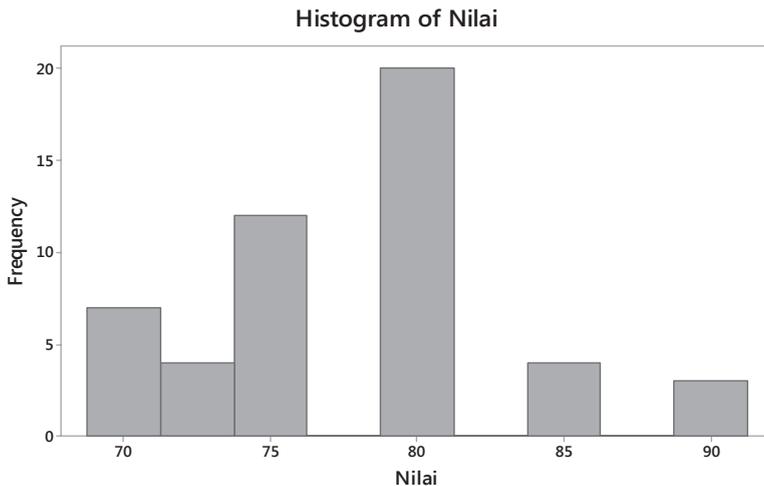


Gambar 4.3: Diagram batang penjualan handphone pada Januari 2020

Dengan menggunakan diagram batang ini, penjualan HP berbagai merk pada bulan Januari 2020 dapat dievaluasi lebih mudah. Mana HP yang banyak terjual atau HP yang penjualannya di bawah merk lainnya dapat dilihat lebih mudah. Misalnya, HP merk C adalah HP yang penjualannya lebih tinggi dibanding merk HP lainnya, sedangkan HP merk D penjualannya paling kecil dibandingkan merk HP lainnya.

- **Histogram**

Histogram juga merupakan bentuk lain dari dotplot. Namun, frekuensinya tidak digambarkan dalam *dot* akan tetapi dalam sebuah batang balok. Jika pada diagram batang atau *barchart* kelompok data adalah bersifat kualitatif, maka pada histogram, kelompok data adalah data kontinyu. Artinya, histogram hanya digunakan ketika skala data bersifat rasio atau interval.



Gambar 4.4: *Histogram* nilai pelatihan karyawan dari Tabel 4.1.

Sebagai contoh untuk membuat histogram, dibuat histogram nilai pelatihan yang terdapat pada Tabel 4.1.

Histogramnya ditunjukkan pada Gambar 4.4. Dalam histogram tersebut, terdapat sembilan kelompok: Kelompok 1: 68,75 s/d 71, 25; Kelompok 2: 71,25 s/d 73,75; Kelompok 3: 73,75 s/d 76,25; Kelompok 4: 76,25 s/d 78,75; Kelompok 5: 78,75 s/d 81,25; Kelompok 6: 81,25 s/d 83,75; Kelompok 7: 83,75 s/d 86,25; Kelompok 8: 86,25 s/d 88,75; dan Kelompok 9: 88,75 s/d 91,25. Kelompok nilai yang frekuensinya kosong adalah Kelompok 4, 6, dan 8. Sementara, kelompok yang memiliki balok (*bar*) tertinggi adalah Kelompok 5.

Histogram ini menggambarkan distribusi data yang telah didapat. Sehingga, jika ingin memprediksi distribusi suatu data, dapat dilihat dari histogramnya. Suatu gambar dari distribusi data adalah bagian penting dari ringkasan statistik dan dianggap sebagai alat statistik yang kuat (Clements, 1991).

- **Diagram Lingkaran (*pie chart*)**

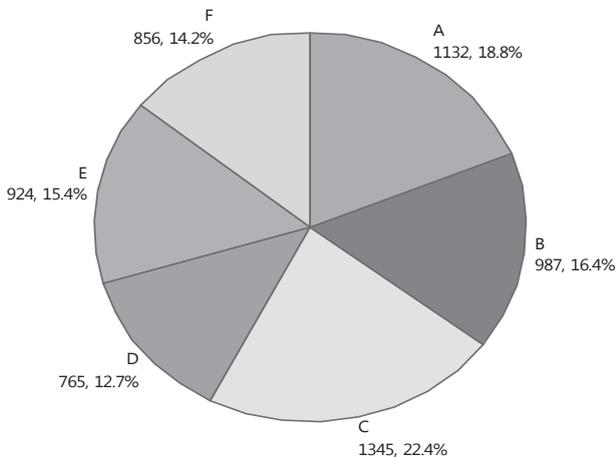
Berbeda dengan grafik-grafik sebelumnya yang menggunakan frekuensi untuk merekam banyaknya data, dalam diagram lingkaran atau *pie chart* bisa menggunakan persentase selain frekuensi. Lingkaran penuh dianggap sebagai total dari pengamatan. Jika dalam persentase berarti itu 100%.

Gambar 4.5 adalah diagram lingkaran dari penjualan HP yang dinyatakan dalam frekuensi dan persentase. Penjualan berbagai merk HP digambarkan dalam bagian-bagian lingkaran yang besarnya sesuai. Untuk menggambar diagram lingkaran akan lebih mudah jika dalam persentase dari pada frekuensi karena bentuk lingkaran lebih mudah diinterpretasikan dalam persentase daripada frekuensi.

Bagian yang memiliki luasan terbesar berarti yang dominan. Dalam diagram lingkaran pada Gambar 4.5,

merk C lebih banyak penjualannya yang dalam hal ini merk C mendominasi 22,4%. Sementara merk HP yang penjualannya terbesar kedua adalah merk A. Jika dilihat kasat mata, dengan menggunakan diagram lingkaran ini, kelihatannya penjualan HP merk A dan B memang hampir sama. Sama halnya dengan penjualan HP merk B dan E yang besaran bagiannya hampir sama. Tanpa menunjukkan angkanya, cukup dengan gambar saja, diagram lingkaran sudah mampu menggambarkan data.

Pie Chart of Penjualan HP



Gambar 4.5: *Pie chart* atau diagram lingkaran

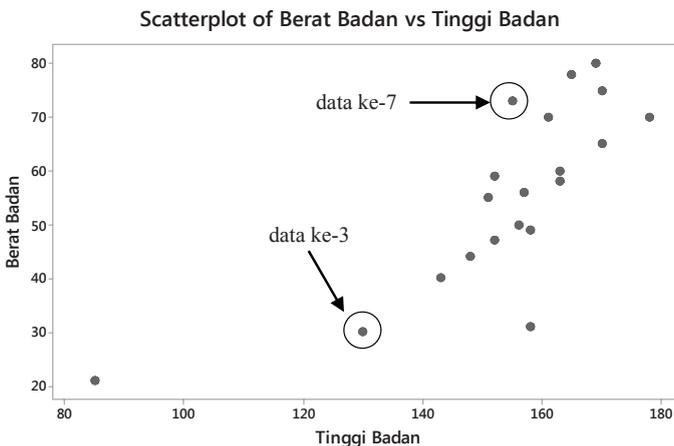
- **Scatterplot**

Scatterplot atau diagram pencar dibentuk dari dua garis yaitu garis vertikal (variabel x) dan horizontal (variabel y). Data-data yang didapat diletakkan sesuai dengan posisinya di garis vertikal dan horisontal. *Scatterplot* dapat menggambarkan hubungan antar variable x dan y . Contoh dari scatterplot ditunjukkan pada Gambar 4.6 yang datanya diambil dari Tabel 4.6.

Tabel 4.6: Data berat badan dan tinggi orang.

No.	Tinggi Badan (x) dalam cm	Berat badan (y) dalam kg	No.	Tinggi Badan (x) dalam cm	Berat badan (y) dan kg
1.	178	70	11.	156	50
2.	170	65	12.	158	49
3.	130	30	13.	163	58
4.	152	47	14.	157	56
5.	143	40	15.	161	70
6.	151	55	16.	155	73
7.	165	78	17.	85	21
8.	169	80	18.	152	59
9.	163	60	19.	148	44
10.	170	75	20.	158	31

Tabel 4.6 adalah variabel tinggi badan dan berat badan yang hubungannya digambarkan dalam *scatterplot* pada Gambar 4.5. Data ke-3, tinggi badan 130 cm dengan berat badan 30 kg, dan data ke-7 tinggi badan 169 cm dengan berat badan 80 kg. *Scatterplot* juga dapat digunakan untuk melihat hubungan antar dua variabel. Pembahasan lebih jelas mengenai ini akan dilakukan pada Bab 7.



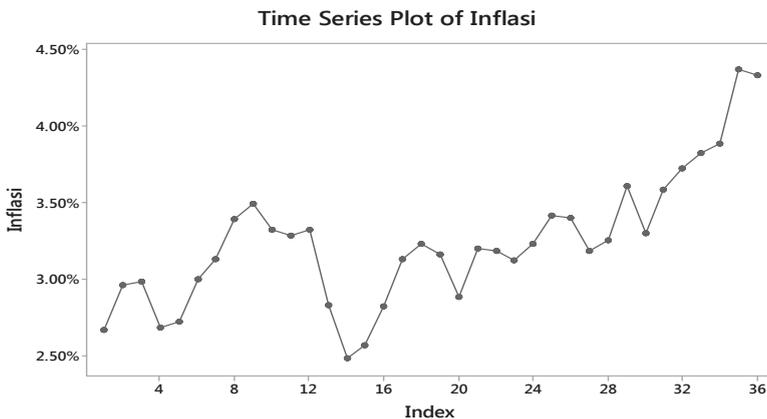
Gambar 4.6: *Scatterplot* sebaran data pada Tabel 4.6

- **Time Series Plot**

Time series plot atau plot deret waktu menggambarkan trend sata dari waktu ke waktu. Plot ini hanya digunakan pada data yang tipenya *time series*. Contoh dari *time series* plot ditunjukkan pada Gambar 4.7 yang merupakan plot dari data pada Tabel 4.7.

Tabel 4.7: Inflasi di Indonesia sejak Mei 2017 hingga April 2020

Bulan	Inflasi	Bulan	Inflasi	Bulan	Inflasi
April 2020	2.67%	April 2019	2.83%	April 2018	3.41%
March 2020	2.96%	March 2019	2.48%	March 2018	3.40%
February 2020	2.98%	February 2019	2.57%	February 2018	3.18%
January 2020	2.68%	January 2019	2.82%	January 2018	3.25%
December 2019	2.72%	December 2018	3.13%	December 2017	3.61%
November 2019	3.00%	November 2018	3.23%	November 2017	3.30%
October 2019	3.13%	October 2018	3.16%	October 2017	3.58%
September 2019	3.39%	September 2018	2.88%	September 2017	3.72%
August 2019	3.49%	August 2018	3.20%	August 2017	3.82%
July 2019	3.32%	July 2018	3.18%	July 2017	3.88%
June 2019	3.28%	June 2018	3.12%	June 2017	4.37%
May 2019	3.32%	May 2018	3.23%	May 2017	4.33%



Gambar 4.7: *Time series plot* inflasi Indonesia pada Mei 2017 hingga April 2020

Dari *time series plot* pada Gambar 4.7 fluktuasi inflasi secara bulanan dapat dilihat lebih jelas polanya. Pada tahun 2020, sejak inflasi lebih tinggi dibandingkan dengan tahun-tahun sebelumnya. Meski gambar data ini masih belum bisa memberikan informasi banyak, tapi tren naiknya inflasi pada bulan Januari 2020 hingga Maret 2020 bisa menjadi informasi awal bahwa pada waktu itu terjadi kenaikan inflasi yang harus diperhatikan oleh pemerintah dan harus dilakukan tindakan.

4.2. Mengukur Pemusatan dan Penyebaran Data

Clements (1991) dalam bukunya *Handbook of Statistical Methods in Manufacturing* menyebutkan membuat data dalam bentuk ukuran pemusatan data dan dalam bentuk ukuran penyebaran adalah cara-cara melaporkan laporan Statistika. Ukuran pemusatan data adalah bagaimana menempatkan pusat dari sekelompok data dalam suatu pengamatan. Contoh dari ukuran pemusatan data adalah rata-rata, median atau nilai tengah, dan modus. Sementara ukuran penyebaran data adalah bagaimana menggambarkan penyebaran sekelompok data. Penyebaran data diukur dengan menggunakan beberapa statistik seperti *range*, varians, dan standar deviasi. Dengan melihat di mana pusat data dan bagaimana penyebarannya, data akan dapat dianalisis lebih mendalam.

4.2.1. Rata-rata (*average*)

Rata-rata adalah ukuran pemusatan data yang sangat populer dan banyak digunakan dalam analisis Statistika. Rata-rata didefinisikan sebagai hasil dari jumlahan semua data dalam sampel yang dibagi dengan banyaknya sampel yang diambil. Dalam sebuah rumus, rata-rata dinyatakan

sebagai berikut:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} \quad (4.1)$$

di mana \bar{x} = rata-rata, x_i = observasi i
 n = banyaknya data
 $i = 1, 2, \dots, n$

Contoh 4.2:

Data berikut adalah hasil observasi dari suatu penelitian tinggi tumbuhan dalam centimeter (cm):

Data ke-	1	2	3	4	5	6	7
Hasil pengamatan	32	45	34	48	38	41	40

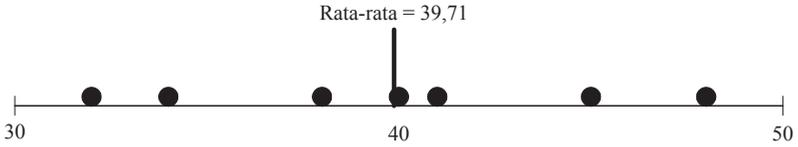
Berapa rata-ratanya?

Penyelesaian Contoh 4.2:

Data tersebut $n = 7$ dengan observasi $x_1 = 32$, $x_2 = 45$, dan seterusnya. Dengan menggunakan rumus rata-rata pada Persamaan 4.1, rata-rata dari data tersebut adalah:

$$\begin{aligned} \bar{x} &= \frac{\sum_{i=1}^n x_i}{n} \\ &= \frac{x_1 + x_2 + x_3 + x_4 + x_5 + x_6 + x_7}{n} \\ &= \frac{32 + 45 + 34 + 48 + 38 + 41 + 40}{7} \\ &= \frac{278}{7} \\ &= 39,71 \text{ cm} \end{aligned}$$

Rata-rata pertumbuhan tumbuhan adalah 39,7 cm. Posisi rata-rata pada semua data ditunjukkan pada Gambar 4.8. Dalam contoh ini, posisi rata-rata berada di tengah data.



Gambar 4.8: Lokasi rata-rata pada sebaran data

Dalam suatu kasus, kadang ditemui data yang tidak ditampilkan satu per satu, tetapi dalam bentuk ringkasan data seperti yang ditunjukkan pada Tabel frekuensi pada Tabel 4.2. Data dalam bentuk tabel frekuensi tersebut ini dapat dihitung rata-ratanya yaitu dengan menggunakan rumus:

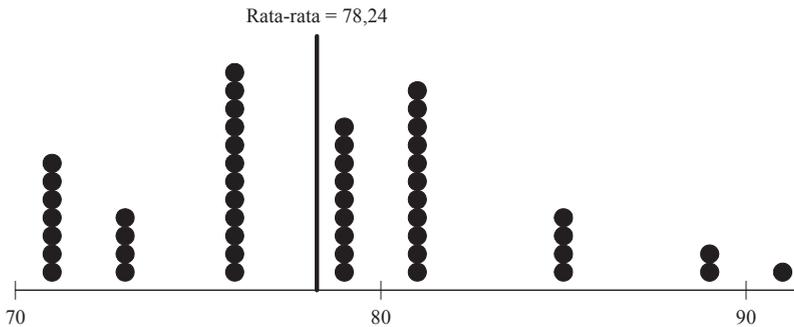
$$\bar{x} = \frac{\sum_{i=1}^k (x_i f_i)}{\sum_{i=1}^k (f_i)} \quad (4.2)$$

di mana \bar{x} = rata-rata,
 x_i = nilai observasi pada kelompok ke- i
 f_i = frekuensi data pada kelompok ke- i
 i = 1, 2, ..., k

Dalam hal ini $k = 8$. Rata-rata nilai dari peserta dengan menggunakan Persamaan 4.2 adalah:

$$\begin{aligned} \bar{x} &= \frac{\sum_{i=1}^8 (x_i f_i)}{\sum_{i=1}^8 (f_i)} \\ &= \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + x_4 f_4 + x_5 f_5 + x_6 f_6 + x_7 f_7 + x_8 f_8}{f_1 + f_2 + f_3 + f_4 + f_5 + f_6 + f_7} \\ &= \frac{91(1) + 89(2) + 85(4) + 81(11) + 79(9) + 76(12) + 73(4) + 71(7)}{1 + 2 + 4 + 11 + 9 + 12 + 4 + 7} \\ &= \frac{3912}{50} \\ &= 78,24 \end{aligned}$$

Ternyata rata-rata nilai training peserta adalah 78,24. Lokasi rata-rata dalam sekelompok data digambarkan pada Gambar 4.9. Pada kasus ini letak rata-rata data cenderung ke kiri jika dilihat dari penyebaran data. Posisinya tidak persis di tengah.



Gambar 4.9: Lokasi rata-rata relatif terhadap data lainnya

Rata-rata dari sekelompok data tidak bisa diartikan bahwa semua data adalah sama nilainya berada dalam rata-rata tersebut. Rata-rata adalah nilai kompromi dari bervariasi data yang telah di dapat. Pada contoh di atas, rata-rata nilai hasil pelatihan karyawan adalah 78,24. Bukan berarti, ke-50 karyawan itu nilainya sama yaitu 78,24 sehingga tidak perlu ada pernyataan: “Karyawan yang saya temui nilainya tidak 78,24, akan tetapi 76 saja”. Nilai rata-rata 78,24 berarti karyawan lain ada yang bernilai lebih dari 78,24 atau kurang dari 78,24 atau bahkan ada yang sama persis bernilai 78,24. Jika ada seorang karyawan yang nilainya berada di bawah rata-rata berarti kemampuannya di bawah nilai umum relative terhadap teman-temannya yang ikut pelatihan. Sebaliknya, jika ada karyawan yang nilainya di atas rata-rata berarti dia unggul dibandingkan teman-temannya yang ikut pelatihan.

4.2.2. Median

Median adalah nilai tengah dari suatu data. Dalam menghitung median, data diurutkan terlebih dahulu dari terkecil hingga terbesar atau sebaliknya dari besar ke terkecil. Kemudian dicari nilai tengahnya. Jika jumlah sampel sifatnya genap, maka median atau nilai tengahnya diperoleh dengan menghitung rata-rata dari dua data yang berlokasi di tengah tersebut. Rumus menentukan lokasi median jika jumlah datanya ganjil adalah

$$k = \frac{n + 1}{2} \tag{4.3}$$

$$\text{Median} = x_k \tag{4.4}$$

dimana n = banyaknya data

k = lokasi data yang dicari mediannya

Sementara, untuk data yang bersifat genap menggunakan rumus

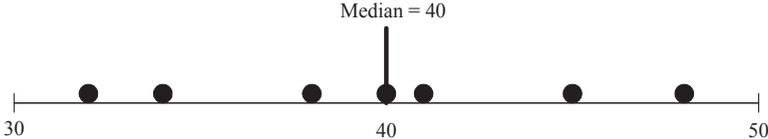
$$k = \frac{n}{2} \tag{4.5}$$

$$\text{Median} = \frac{x_k + x_{k+1}}{2} \tag{4.6}$$

Untuk memudahkan memahami, berikut adalah data sebagai ilustrasinya dengan menggunakan data pada Contoh 4.2. Pada data tersebut jumlah data adalah ganjil. Data tersebut diurutkan terlebih dahulu, sehingga menjadi:

Urutan	1	2	3	4	5	6	7
Data ke-	1	3	5	7	6	2	4
Hasil pengamatan	32	34	38	40	41	45	48

Dengan menggunakan rumus pada Persamaan 4.5, didapatkan $k=(n+1)/2=8/2=4$. Berarti median $x_k = x_4$ yaitu bernilai 40. Lokasi median relatif terhadap data lainnya ditunjukkan pada Gambar 4.10.



Gambar 4.10: Lokasi median relatif terhadap data lainnya

Jika menggunakan data pada Tabel 4.6, maka nilai mediannya adalah $k=n/2=50/2=25$. Berarti $x_k = x_{25}$ dan $x_{k+1} = x_{26}$. Data pada Tabel 4.6 diurutkan berdasarkan angka terkecil ke terbesar. Kemudian dihitung frekuensi kumulatifnya. Hasilnya ditunjukkan pada Tabel 4.8. Pada nilai tes 71, frekuensi kumulatifnya 7. Ini bisa berarti nilai tes 71 berada pada urutan 1 sampai 7. Nilai 73 memiliki nilai frekuensi kumulatif 11. Ini bisa dikatakan data 73 berada di urutan 8 sampai dengan 11. Begitu seterusnya. Data ke-25

Tabel 4.8: Data dari Tabel 4.2 yang akan digunakan untuk ilustrasi

Nilai setelah pelatihan kerja	Frekuensi	Frekuensi kumulatif	Urutan data
71	7	7	1 - 7
73	4	11	8 - 11
76	12	23	12 - 23
79	9	32	24 - 32
81	11	43	33 - 43
85	4	47	44 - 47
89	2	49	48 - 49
91	1	50	50
Total peserta	50		

atau x_{25} dalam hal ini adalah nilai 79 dan x_{26} juga 79. Dengan menggunakan rumus median pada Persamaan 4.6 didapat median nilai hasil tes pelatihan kerja adalah $(79 + 79)/2 = 79$.

Nilai tengah sebagai alternatif dari rata-rata dalam mengukur pemusatan data. Interpretasi nilai tengah tentu berbeda dengan nilai rata-rata. Nilai tengah didapat dengan membuat urutan data terlebih dulu, sedangkan rata-rata tidak membutuhkan urutan data terlebih dahulu. Nilai tengah melihat pemusatan data dari urutan data.

4.2.3. Modus

Modus adalah nilai yang paling sering muncul atau data yang frekuensinya paling banyak. Perhitungan modus ini yang paling mudah dilakukan dibandingkan ukuran pemusatan data lainnya. Modus digunakan karena pada umumnya, distribusi data yang paling banyak terjadi biasanya berlokasi di pusat data. Ilustrasi dalam menghitung modus di sini digunakan contoh sebagai berikut:

Contoh 4.3:

Berapa modus dari data berikut:

Data ke-	1	2	3	4	5	6	7	8
Hasil pengamatan	32	45	34	48	38	41	40	37

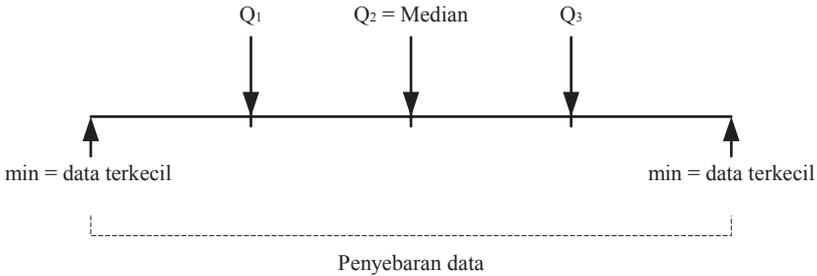
Penyelesaian Contoh 4.4:

Pada data tersebut, nilai 32, 45 ... hingga 37 frekuensinya adalah sama, yaitu 1. Di sini, berarti frekuensi tertinggi yaitu 1 sehingga semua data bisa dianggap sebagai modus. Sementara, jika menggunakan contoh data pada Tabel 4.2, maka nilai yang frekuensinya paling tinggi adalah 76 dengan frekuensi 12 kali. Sehingga, modusnya adalah 76.

Modus bisa saja berada pada nilai yang terkecil, bisa juga pada nilai yang terbesar. Modus tidak memperhatikan urutan data. Tetapi melihat dari frekuensi kemunculan data. Jika menggunakan modus sebagai ukuran pemusatan data, harus berhati-hati dalam menafsirkan karena modus tidak memperhatikan posisinya dibandingkan data lainnya. Modus juga tidak memperhatikan urutan data.

4.2.4. Kuartil

Kuartil (*quartile*) adalah membagi penyebaran data ke dalam empat lokasi, sehingga didapatkan tiga titik lokasi yaitu data seperempat (25%), setengah (50%), dan sepertiga perempat (75%). Data ke-25% dinamakan kuartil pertama atau Q_1 , data ke-50% dinamakan kuartil ke dua atau Q_2 , dan data ke-75% disebut kuartil ketiga atau Q_3 . Q_2 disebut juga median. Lebih detilnya bisa dilihat di Gambar 4.11.



Gambar 4.11: Lokasi Q_1 , Q_2 , dan Q_3

Dalam menghitung Kuartil pertama (Q_1) jika jumlah datanya adalah ganjil, maka

$$q_1 = \frac{0,5n + 1}{2} \tag{4.7}$$

Dimana q_1 = kuartil pertama, n adalah banyaknya data. Jika jumlah datanya genap, maka

$$q_1 = \frac{0,5n}{2} \quad (4.8)$$

Dimana q_1 adalah lokasi data kuartil pertama.

Untuk menghitung data kuartil kedua, sama dengan menghitung median. Bisa menggunakan Persamaan 4.3 hingga Persamaan 4.6. Sehingga di sini tidak akan diberikan lagi. Sementara itu, untuk menghitung kuartil ketiga, jika jumlah datanya adalah ganjil menggunakan persamaan sebagai berikut:

$$q_3 = n - \frac{(0,5n + 1)}{2} \quad (4.9)$$

Dimana q_3 = kuartil ketiga. Jika jumlah datanyanya genap, maka

$$q_3 = n - \frac{0,5n}{2} \quad (4.10)$$

Dimana q_3 adalah lokasi data kuartil ketiga.

Contoh 4.4:

Sebagai ilustrasi, akan digunakan data dari Tabel 4.8 yang akan dihitung Q_1 , Q_2 , dan Q_3 . Terdapat data sebanyak 50 yang berarti datanya adalah genap. Dengan menggunakan Persamaan 4.7, maka lokasi kuartil pertama berada pada urutan ke:

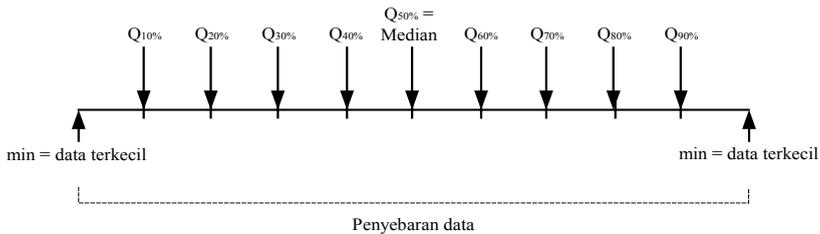
$$q_1 = \frac{0,5n}{2} = \frac{0,5(50)}{2} = 12,5$$

$Q_1 = x_{12,5}$ nilai hasil pelatihan kerja karyawan yang posisinya berada pada urutan 12,5 adalah nilai 76. Selanjutnya menghitung kuartil ke-3 (Q_3) yang akan menggunakan Persamaan 4.10. Berarti,

$$q_3 = n - \frac{0,5n}{2} = 50 - \frac{0,5(50)}{2} = 37,5$$

Maka $Q_3 = x_{37,5}$. Nilai pada Tabel 4.7 yang berada pada urutan ke 37,5 adalah 81. Q_2 tidak dihitung lagi karena sama dengan menghitung median.

Selain kuartil, ada juga kuantil. Jika kuartil membagi keseluruhan data yang telah diurutkan menjadi 4 bagian, maka kuantil membagi keseluruhan data yang telah diurutkan menjadi 10 bagian. Jadi, ada $Q_{10\%}$, $Q_{20\%}$, sampai dengan $Q_{90\%}$. Lokasi kuantil ditunjukkan pada Gambar 4.12. Jika Penggunaan kuartil biasanya digunakan dalam bidang ekonomi untuk mengidentifikasi penduduk miskin. Biasanya menggunakan istilah *income decil*, dimana desil pertama adalah kelompok ekonomi masyarakat yang berada pada 10% ekonomi atau ekonominya terendah 10%. Desil selanjutnya adalah tingkat ekonomi lebih tinggi.



Gambar 4.12: Lokasi Kuantil pada keseluruhan data

4.2.5. Range

Range adalah jarak antara nilai terendah dan tertinggi. Rentang dapat dihitung dengan mengurangi nilai tertinggi dengan nilai terendah. Dalam bentuk notasi adalah sebagai berikut:

$$\text{range} = x_{\min} - x_{\max} \quad (4.11)$$

x_{min} adalah data terkecil, sedangkan x_{max} adalah data terbesar. Contohnya, jika menggunakan data pada Tabel 4.8, maka $range = 91 - 71 = 20$. *Range* menggambarkan penyebaran data. Semakin besar *range*-nya berarti semakin heterogen, sebaliknya jika semakin kecil *range*-nya maka semakin homogen. Bisa dikatakan, *range* adalah ukuran penyebaran data.

4.2.6. Varians

Selain *range*, ukuran penyebaran data yang banyak digunakan adalah varians (*variance*). Dalam varians, penyebaran data dilihat dari penyimpangannya terhadap rata-rata. Rata-rata adalah prediksi nilai pusat suatu data. Jika data menyimpang dari pusat, dianggap sebagai penyimpangan. Penyimpangan bisa bernilai positif atau negatif. Misalnya, rata-ratanya adalah 7, sementara nilai amatan adalah 5. Jarak dari 5 ke 7 adalah 2. Namun, dalam matematika $5 - 7 = -2$ yang bernilai negatif. Data pengamatan lainnya, misalnya 8, akan menghasilkan $8 - 7 = 1$ yang bernilai positif. Untuk mendapatkan nilai satuan, data-data penyimpangan tersebut dikuadratkan. Semakin besar penyimpangannya, meskipun bernilai negatif, akan menghasilkan kuadrat yang besar. Misalnya, penyimpangan 2 dan -2 jika dikuadratkan maka akan bernilai sama besar yaitu 4. Kuadrat dari penyimpangan inilah yang dijumlahkan untuk mendapatkan angka varians. Semakin besar hasil penjumlahannya maka penyimpangannya akan semakin besar. Oleh karena itu, varians didefinisikan sebagai kuadrat dari penyimpangan data dari rata-rata yang dibagi dengan banyaknya data dikurangi satu. Dalam persamaan dinyatakan sebagai berikut:

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \tag{4.12}$$

Dimana $s^2 =$ varians

\bar{x} = rata-rata yang dapat diperoleh dengan menggunakan Persamaan 4.1

x_i = data pengamatan

n = jumlah sampel data.

Contoh menghitung varians di sini akan menggunakan data pengamatan tinggi pertumbuhan tumbuhan yang telah digunakan untuk menghitung rata-rata pada Contoh 4.2. Dari data tersebut diketahui rata-ratanya pertumbuhan tumbuhan adalah 39,71 cm. Untuk memudahkan menghitung, akan dibuat tabel sebagai berikut:

								Total
i	1	2	3	4	5	6	7	
x_i	32	45	34	48	38	41	40	
$x_i - \bar{x}$	-7,71	5,29	-5,71	8.29	-1.71	1.29	0.29	
$(x_i - \bar{x})^2$	50.4441	27.9841	32.6041	68.7241	2.9241	1.6641	0.0841	184,4287

Jika hitungan tersebut dimasukkan ke dalam Persamaan 4.12,

$$\begin{aligned}
 s^2 &= \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1} \\
 &= \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + (x_4 - \bar{x})^2 + (x_5 - \bar{x})^2 + (x_6 - \bar{x})^2 + (x_7 - \bar{x})^2}{7 - 1} \\
 &= \frac{50,4441 + 27,9841 + 32,6941 + 68,7241 + 2,941 + 1,6641 + 0,0841}{6} \\
 &= \frac{184,4287}{6} \\
 &= 30,7381
 \end{aligned}$$

Varians dari pertumbuhan tumbuhan adalah 30,7381. Dengan nilai tersebut, masih belum bisa dikatakan penyebarannya besar atau kecil karena tidak ada nilai pembandingnya. Tentu, dengan menggunakan ukuran yang sama. Misalnya dengan membandingkan varians pertumbuhan tumbuhan jika pada saat eksperimen diberi intervensi tertentu.

4.2.7. Standar Deviasi

Standar deviasi adalah bentuk lain dari varians. Standar deviasi adalah akar dari varians. Oleh karena itu, standar deviasi dinotasikan dengan huruf s yang merupakan akar dari s^2 . Rumus dari standar deviasi adalah sebagai berikut:

$$s = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n - 1}} \quad (4.13)$$

Dimana s adalah standar deviasi.

Dengan menggunakan data yang sama dengan menghitung varians, didapatkan standar deviasi dari pertumbuhan tumbuhan adalah $\sqrt{s^2} = \sqrt{30,7381} = 5,4898$.

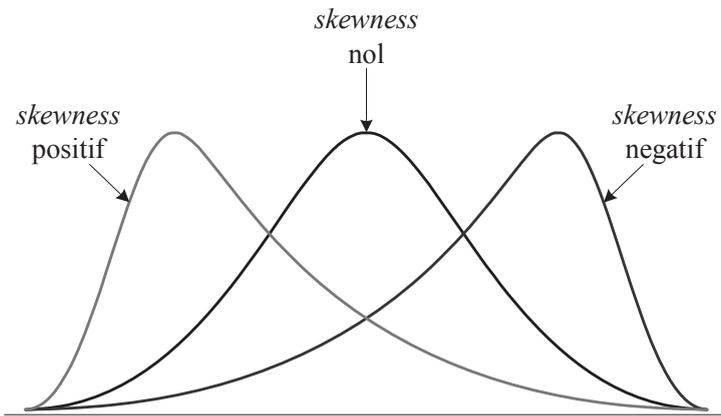
Standar deviasi adalah bentuk lain dari varians. Keduanya menyatakan penyimpangan data relatif terhadap rata-rata. Besar atau kecilnya standar deviasi atau varians tidak bisa langsung dilihat dari satu data, tetapi dibandingkan dengan sampel lain untuk kasus yang sama. Misalnya dengan menggunakan sampel yang berbeda diperoleh standar deviasi sebesar 3,45. Dari sini, dibandingkan sampel sebelumnya, sampel ke dua ini memiliki homogenitas lebih rendah.

4.2.8. Skewness

Ukuran penyebaran lainnya adalah *skewness* yang melihat penyimpangan distribusi data dari kesimetrisan distribusinya. Untuk menghitung *skewness* atau kemiringan distribusi, melibatkan standar deviasi, rata-rata, dan modus. Sehingga, persamaan untuk menghitung *skewness* adalah sebagai berikut:

$$skewness = \frac{\bar{x} - \text{modus}}{s} \quad (4.13)$$

Dengan syarat s tidak sama dengan 0. \bar{x} diperoleh dengan menggunakan Persamaan 4.1, dan modus diperoleh dengan cara seperti yang dijelaskan pada Bab 4.2.3. Nilai *skewness* ini bisa bernilai negatif atau positif. Jika nilai rata-rata lebih besar dari modus, maka akan menghasilkan nilai positif. Jika *skewness* bernilai positif, berarti 'ekor' dari distribusi banyak yang di kanan. Lain halnya Jika nilai rata-rata lebih rendah dari modus, maka akan dihasilkan nilai negatif. Ini berarti 'ekor' dari penyebaran datanya ada di kiri. Bisa juga *skewness* bernilai 0 jika rata-rata sama dengan modus.



Gambar 4.13: Beberapa derajat *skewness* suatu data

Pada saat *skewness* bernilai nol, maka bisa dikatakan pola datanya simetris. Nilai mutlak *skewness* semakin besar menunjukkan kemiringannya semakin besar. Gambaran *skewness* ditunjukkan pada Gambar 4.13.

Sebagai ilustrasi, menggunakan data pada Tabel 4.8. Dari data tersebut diperoleh rata-rata adalah 78,24, standar deviasi sebesar 4,955, dan modus adalah 76. Dengan nilai ini, diperoleh *skewness* sebesar

$$\begin{aligned} skewness &= \frac{\bar{x} - \text{modus}}{s} \\ &= \frac{78,24 - 76}{4,955} \\ &= 0,452 \end{aligned}$$

Skewness bernilai 0,452 yang berarti 'ekor' dari distribusi data berada di kanan karena bernilai positif. Jika dilihat pada Gambar 4.9 yang menunjukkan distribusi data dari Tabel 4.8 juga nampak bahwa data cenderung ke kiri dengan ekor yang berada di bagian kanan. Distribusi datanya tidak simetris.

4.2.9. Penerapan ukuran pemusatan dan penyebaran data

Berbagai ukuran pemusatan data dan penyebaran data sudah dibahas pada sub-sub bab sebelumnya. Bagaimana penggunaannya dalam praktik tentu harus memenuhi aturan-aturan tertentu. Tidak semua ukuran tersebut dapat diterapkan pada semua skala data. Pada Tabel 4.9 ditunjukkan kapan ukuran-ukuran statistik yang telah disebutkan di atas dapat dipakai. Ukuran-ukuran pemusatan data dan skala data yang dapat menggunakan semua jenis skala data adalah modus.

Skala data ordinal dan nominal tidak dapat dihitung rata-rata, median, *range*, varians dan standar deviasi karena ukuran-ukuran ini hanya menghitung data yang bersifat kontinyu. Sementara, data ordinal dan nominal sifatnya diskrit atau disebut juga data kategori. Sebagai ilustrasi, berikut akan diberikan penjelasannya.

Tabel 4.9: Penggunaan ukuran-ukuran statistik sesuai dengan skala datanya

Statistik \ Skala	Nominal	Ordinal	Interval	Rasio
Rata-rata	Tidak	Tidak	Ya	Ya
Median	Tidak	Tidak	Ya	Ya
Modus	Ya	Ya	Ya	Ya
Range	Tidak	Tidak	Ya	Ya
Varians	Tidak	Tidak	Ya	Ya
Standar deviasi	Tidak	Tidak	Ya	Ya
<i>Skewness</i>	Tidak	Tidak	Ya	Ya

Dalam suatu penelitian, diperoleh data jumlah konsumen produk X sesuai dengan jenjang pendidikannya. Datanya ditunjukkan pada Tabel 4.10. Data pendidikan adalah berskala nominal dengan urutan Sekolah Dasar, Sekolah Menengah Pertama, Sekolah Menengah Atas, dan Perguruan Tinggi. Jika menggunakan kode, level pendidikan tersebut dapat dibuat kode 1 untuk Sekolah Dasar, 2 untuk Sekolah Menengah Pertama, 3 untuk Sekolah Menengah Atas, dan 4 untuk Perguruan Tinggi. Tabel 4.10 kemudian dikonversi ke Tabel 4.11.

Tabel 4.10: Data pendidikan konsumen produk X

Pendidikan	Frekuensi
Sekolah Dasar	23
Sekolah Menengah Pertama	41
Sekolah Menengah Atas	54
Perguruan Tinggi	27

Tabel 4.11: Konversi data pada Tabel 4.10

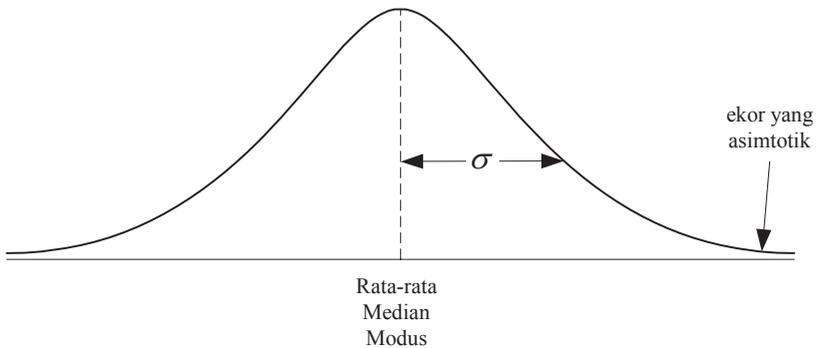
Pendidikan	Kode	Frekuensi	Kode x Frekuensi
Sekolah Dasar	1	23	23
Sekolah Menengah Pertama	2	41	82
Sekolah Menengah Atas	3	54	162
Perguruan Tinggi	4	27	108
	total	145	375

Dari pada Tabel 4.11 dihitung rata-rata dengan menggunakan Persamaan 4.2, maka diperoleh $\bar{x} = 375/145 = 2,586$. Di sini didapatkan rata-rata pendidikan konsumen produk X adalah 2,586. Angka 2,586 jika dikembalikan lagi ke dalam data pendidikan, SD, SMP, SMA atau Perguruan Tinggi, tentu akan menginterpretasikannya. Itu karena data pendidikan di sini hanya diberi kode 1, 2, 3, dan 4. Sementara 2,586 berada di antara SMA dan Perguruan Tinggi, yang berarti bukan SMA dan bukan Perguruan Tinggi. Inilah yang membuat data berskala ordinal dan nominal tidak dapat menggunakan rata-rata untuk menilai ukuran pemu-satannya. Menggunakan median untuk data nominal juga tidak bisa dilakukan karena jika ditemukan angka pecahan akan sulit menentukan di mana posisi median sebenarnya. Sementara, jika menggunakan modus lebih masuk akal. Dalam contoh kasus ini modulusnya adalah SMA yang jumlah konsumennya ada 54 orang. Ini bisa dikatakan,

sebagian besar konsumen produk X dikonsumsi oleh orang berpendidikan SMA.

4.3. Distribusi Normal

Pada saat membuat histogram, data akan mengikuti pola tertentu yang menggambarkan distribusi data. Distribusi data adalah gambaran penyebaran semua data dalam sampel. Dalam statistika, distribusi data dibedakan dalam distribusi diskrit dan kontinu. Contoh distribusi diskrit adalah distribusi Binomial, distribusi Hipergeometrik, dan distribusi Poisson. Contoh distribusi kontinu adalah distribusi Normal, distribusi Gamma, dan distribusi Eksponensial.



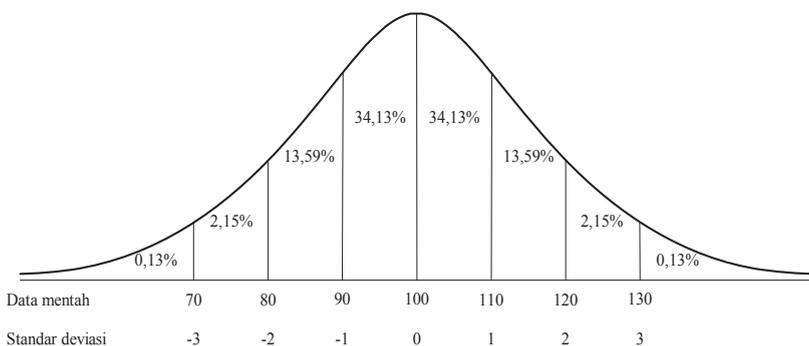
Gambar 4.14: Kurva normal

Dari berbagai macam distribusi tersebut, distribusi normal adalah distribusi yang sangat penting dalam statistika. Distribusi Normal seringkali disebut sebagai data berdistribusi Gauss. Data ini ditemukan oleh Carl Friedich Gauss. Kurva distribusi Normal bentuknya seperti lonceng, oleh karenanya kurva distribusi normal ini dinamakan *bell-shaped curve*. Bentuknya dari kurva normal ditunjukkan pada Gambar 4.14.

Teori yang mendasari distribusi ini adalah *Limit Central Theorema* yang mengatakan bahwa jika \bar{x} adalah rata-rata sampel acak (*random*) dengan ukuran sebesar n sampel yang sangat besar yang diambil dari populasi dengan rata-rata μ dan varians adalah σ , maka bentuk distribusinya akan mengikuti,

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \tag{4.14}$$

Dengan mengombinasikan pernyataan Salkind (2008) dan Walpole et al., (2013), distribusi normal memiliki tiga karakteristik penting. Pertama, kurva distribusi normal memiliki bentuk yang simetris. Bentuk antar kiri dan kanan adalah sama. Kedua, dalam distribusi normal, lokasi rata-rata, median, dan modus berada dalam satu titik yang sama. Nilai modus, dimana kurva mengalami titik tertinggi pada y berada pada satu lokasi dengan modus. Jika ketiga nilai statistik tersebut tidak berada dalam satu lokasi, maka distribusinya akan tidak seimbang atau *skewness*-nya akan lebih besar dari nol. Ketiga, kurva normal memiliki titik belok pada $x = \mu \pm \sigma$. Keempat, ekor dari distribusi normal adalah asimtotik. Maksudnya adalah, garis pada ekor kiri dan kanan semakin mendekati garis horisontal namun tidak



Gambar 4.15: Distribusi kasus dalam kurva normal

pernah menyentuh garis horisontal. Kelima, luas total daerah dari kurva normal adalah sama dengan 1. Ini bisa dilihat dari Gambar 4.15.

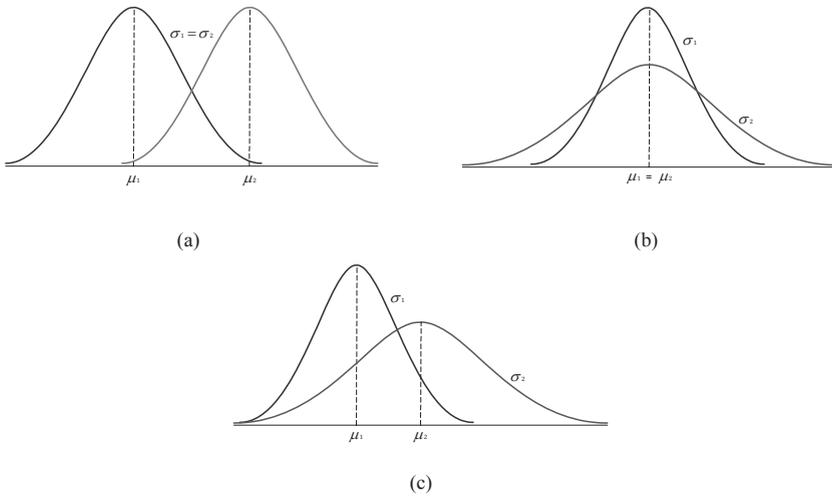
Distribusi Normal memiliki parameter rata-rata (μ) dan varians (σ). *Probability density function* (pdf) atau fungsi kepadatan probabilitas dari distribusi normal dinyatakan dengan:

$$n(x; \mu, \sigma) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}, \quad -\infty < x < \infty \quad (4.15)$$

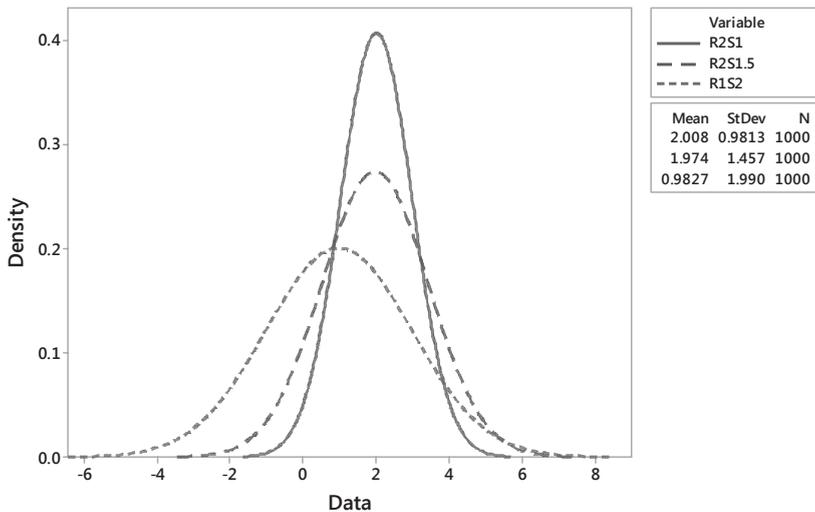
dimana $\pi = 3,14159$ dan $e = 2,71828$.

Gambar 4.16 menunjukkan perbandingan bentuk kurva normal jika μ dan σ berbeda. Lokasi rata-rata (μ) semakin ke kanan maka berarti nilainya semakin besar seperti yang ditunjukkan pada Gambar 4.16(a) dan (c). Pada Gambar 4.16(a) rata-ratanya berbeda namun memiliki varians yang berbeda. Sementara, standar deviasi semakin besar, maka bentuk kurva normal akan semakin melebar seperti yang terlihat pada Gambar 4.16(b) dan (c). Semakin sempit penyebarannya (standar deviasi), maka semakin homogen datanya. Begitu juga sebaliknya, semakin besar standar deviasinya, semakin heterogen datanya. Pada Gambar 4.16(b) rata-ratanya sama, namun variansnya berbeda. Sedangkan pada Gambar 4.16(c) baik rata-rata maupun varians semuanya berbeda.

Gambar 4.17 adalah contoh distrusi normal dari data simulasi dengan berbagai variasi rata-rata dan standar deviasi. Data RS1 adalah data dengan $\mu = 2,00$ dan $\sigma = 0,98$. Bentuknya, lebih ramping dengan tingkat kepadatan (garis vertikal, *density*) yang lebih tinggi. Sedangkan data RS1,5 adalah data dengan $\mu = 1,97$ dan $\sigma = 1,46$. Sementara itu, data RS2 adalah data dengan $\mu = 0,98$ dan $\sigma = 1,99$.



Gambar 4.16: Berbagai kurva normal dengan rata-rata dan standar deviasi yang bervariasi (a) $\mu_1 < \mu_2$ dan $\sigma_1 = \sigma_2$, (b) $\mu_1 = \mu_2$ dan $\sigma_1 < \sigma_2$, dan (c) $\mu_1 < \mu_2$ dan $\sigma_1 < \sigma_2$

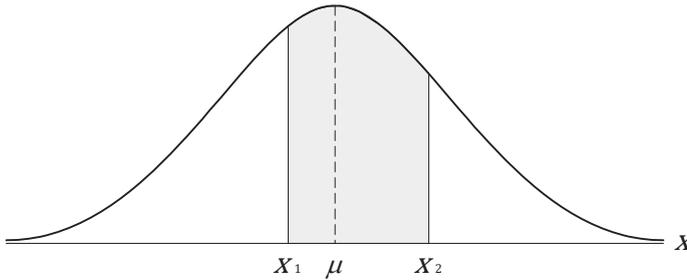


Gambar 4.17: Contoh data berdistribusi normal dengan berbagai rata-rata dan standar deviasi.

Bagaimana luasan di dalam kurva normal? Luasan kurva normal dinyatakan dalam persamaan berikut:

$$P(x_1 < x < x_2) = \int_{x_1}^{x_2} n(x; \mu, \sigma) dx = \frac{1}{\sqrt{2\pi\sigma}} \int_{x_1}^{x_2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \quad (4.16)$$

Persamaan ini dalam bentuk grafik dinyatakan sebagai berikut:



Gambar 4.18: Daerah yang diarsir adalah $P(x_1 < x < x_2)$

Jika terdapat kurva normal lain dengan rata-rata dan standar deviasi yang berbeda, maka luasan $P(x_1 < x < x_2)$ juga akan berbeda. Jika suatu variabel, sebut saja z , memiliki rata-rata sama dengan 0 dan varians adalah 1, maka transformasi dari distribusi normalnya adalah:

$$z = \frac{x - \mu}{\sigma} \quad (4.17)$$

$$\begin{aligned} P(x_1 < x < x_2) &= \frac{1}{\sqrt{2\pi\sigma}} \int_{x_1}^{x_2} e^{-\frac{1}{2\sigma^2}(x-\mu)^2} dx \\ &= \frac{1}{\sqrt{2\pi\sigma}} \int_{z_1}^{z_2} e^{-\frac{1}{2}z^2} dz \\ &= \int_{z_1}^{z_2} n(z; 0,1) dz \\ &= P(z_1 < z < z_2) \end{aligned} \quad (4.18)$$

Dimana z adalah variabel acak normal dengan rata-rata 0 dan varians 1.

Berikut adalah contoh yang diambil dari Walpole et al. (2013).

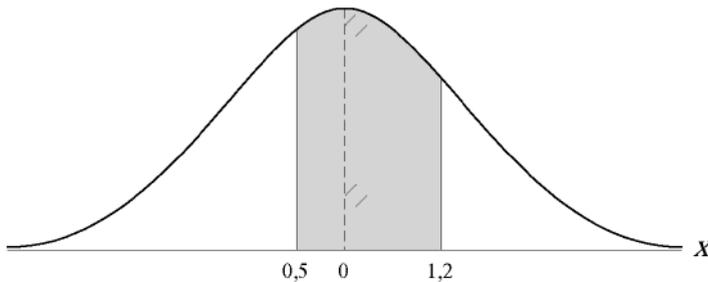
Terdapat variabel acak x yang berdistribusi normal dengan $\mu = 50$ dan $\sigma = 10$, cari probabilitas bahwa x bernilai antara 45 dan 62. Penyelesaiannya adalah sebagai berikut. Dengan menggunakan Persamaan 4.17, z_1 dan z_2 adalah

$$z_1 = \frac{45 - 50}{10} = -0,5 \quad \text{dan} \quad z_2 = \frac{62 - 50}{10} = 1,2$$

Sehingga

$$P(45 < x < 62) = P(-0,5 < z < 1,2)$$

Dalam bentuk grafik, area $P(-0,5 < x < 1,2)$ ditunjukkan pada gambar di bawah ini.



Gambar 4.19: Luas $P(-0,5 < x < 1,2)$

Dengan menggunakan Tabel z pada Lampiran 1:

$$\begin{aligned} P(45 < x < 62) &= P(-0,5 < z < 1,2) \\ &= P(z < 1,2) - P(z < -0,5) \\ &= 0,8849 - 0,3085 \\ &= 0,5764 \end{aligned}$$

Maka, luasannya adalah 0,5764.

BAB 5.

UJI HIPOTESIS

Uji hipotesis adalah statistika inferensia yang paling sederhana. Dalam uji hipotesis berdasarkan data yang akan diuji, terdapat dua jenis yaitu uji hipotesis parametrik dan non-parametrik. Uji hipotesis parametrik digunakan jika data telah memenuhi distribusi normal. Variabel yang skalanya datanya interval atau rasio seringkali, dalam jumlah yang sangat banyak akan mengikuti *central limit theorem*. Sehingga pada data-data yang berskala interval dan rasio seringkali diasumsikan mengikuti distribusi normal. Sementara, uji hipotesis non-parametrik diberlakukan pada data yang berskala ordinal atau nominal. Selain itu, jika jumlah sampelnya kecil, biasanya kurang dari 30, uji non-parametrik sering diterapkan untuk menguji variabel yang bersifat interval atau rasio. Asumsinya, dalam skala kecil, distribusi normal variabel akan sulit dipenuhi.

Pada Bab ini hanya membahas uji hipotesis parametrik. Beberapa yang akan dibahas adalah uji hipotesis satu arah dan uji hipotesis dua arah. Oleh karena itu, melalui pembahasan buku ini diharapkan pembaca akan:

- Mampu memahami elemen-elemen dalam uji hipotesis
- Mampu memahami dan menerapkan tahapan dalam uji hipotesis.
- Mampu memahami, menghitung, dan menerapkan uji hipotesis satu arah seperti uji satu sampel dan uji proporsi satu sampel

- Mampu memahami, menghitung, dan menerapkan uji hipotesis dua sampel, baik uji dua sampel independen, uji dua sampel berpasangan, dan uji proporsi dua sampel.
- Mampu memahami, menghitung, dan menerapkan *analysis of variance*.

5.1. Pengantar Uji Hipotesis

Uji hipotesis adalah statistika inferensia untuk menguji apakah data-data sampel yang kita ambil sudah sesuai dengan dugaan. Hipotesis statistik adalah pernyataan mengenai parameter-parameter yang akan diuji. Di Sub-Bab 4.2 telah dibahas bagaimana menaksir titik pada ukuran pemusatan dan penyebaran data. Ukuran-ukuran tersebut seperti rata-rata, median, modus, dan varians, yang akan diuji hipotesisnya. Uji hipotesis juga bisa digunakan untuk mengidentifikasi adanya perbedaan hasil antar dua kelompok data sampel yang kita amati. Dengan mengetahui perbedaan hasil, alat analisis ini juga mampu digunakan untuk melihat pengaruh dari suatu perlakuan.

Dalam uji hipotesis statistik terdapat elemen-elemen penting yang harus ada yaitu hipotesis awal dan hipotesis alternatif, uji statistik, dan daerah kritis. Dalam bab ini akan dijelaskan masing-masing elemen tersebut.

- **Hipotesa awal (H_0) dan Hipotesis alternatif (H_1)**

Uji hipotesis menggunakan istilah hipotesis awal (H_0) dan hipotesis alternatif (H_1). Hipotesis awal adalah negasi dari hipotesis alternatif. Ini berarti pernyataan dalam hipotesis awal selalu kontradiktif dengan pernyataan dalam hipotesis alternatif. Jika pernyataan dalam hipotesis awal ditolak, maka berarti pernyataan pada hipotesis alternatif tidak ditolak.

Hipotesis alternatif (H_1) adalah pernyataan yang sesuai dengan teori yang akan kita uji. Sebaliknya, Hipotesis awal (H_0) adalah pernyataan yang kontradiktif dengan teori. Inilah mengapa hipotesis awal dalam Bahasa Inggris disebut sebagai *null hypothesis* karena H_0 membatalkan (*nullifies*) pernyataan pada hipotesis alternatif (H_1). Pernyataan hipotesis ini kemudian dibuat keputusan yang akan jatuh pada **tolak H_0** atau **gagal tolak H_0** karena tidak cukup bukti dalam data yang dianalisis. Jika tolak H_0 , maka kesimpulannya adalah menganggap pernyataan H_0 adalah benar. Sebaliknya, jika gagal tolak H_0 maka pernyataan pada H_0 adalah benar. Contoh pernyataan hipotesis adalah:

$H_0 : \mu = 0$ (Hipotesis awal adalah rata-rata sampel sama dengan nol)

$H_1 : \mu \neq 0$ (Hipotesis awal adalah rata-rata sampel tidak sama dengan nol)

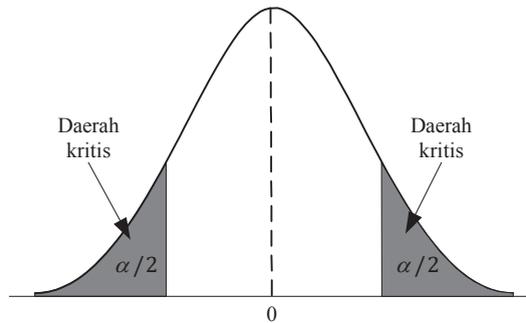
- **Uji statistik**

Uji statistik adalah menghitung statistik dari data. Statistik apa yang akan diuji harus ditentukan terlebih dulu di awal. Misalnya, akan menguji proporsi apakah sama dengan nilai tertentu atau sama dengan proporsi sampel lainnya, maka statistik uji untuk menguji proporsi dengan nilai tertentu atau statistik untuk menguji proporsi dua sampel. Sama halnya ketika akan melakukan uji rata-rata sampel, varians, dan lain-lain. Contoh dari statistik uji adalah t dan z yang ini akan dijelaskan lebih rinci di sub-bab berikutnya.

- **Daerah kritis atau daerah penolakan**

Daerah kritis menyatakan sekumpulan daerah yang menentukan apakah hasil statistik yang telah dihitung berada pada daerah kritis. Jika nilai statistik masuk

pada daerah kritis maka hipotesis awal (H_0). Gambaran daerah kritis ditunjukkan pada Gambar 5.1.



Gambar 5.1: Lokasi daerah kritis pada distribusi t jika menggunakan uji dua arah

Lokasi daerah kritis adalah di daerah yang diberi warna. Jika statistik dari data jatuh pada daerah tersebut, maka kesimpulannya adalah tolak H_0 .

Elemen-elemen tersebut kemudian digunakan dalam uji hipotesis statistik. Tahapan-tahapan dalam uji hipotesis statistik adalah sebagai berikut:

1. **Tentukan parameter** yang akan diuji. Parameter-parameter yang digunakan untuk uji hipotesis antara lain rata-rata (μ), standar deviasi (σ), varians (σ^2), dan proporsi (p).
2. **Menentukan hipotesis.** Hipotesis awal (H_0) dan hipotesis alternatif (H_1) ditentukan dengan panduan bahwa hipotesis awal adalah negasi dari hipotesis alternatif. Panduan penentuan telah dibahas sebelumnya. Pernyataan hipotesis ini bisa menggunakan uji hipotesis satu arah atau dua arah. Uji hipotesis satu arah, contoh pernyataannya adalah:

$$H_0 : \mu \leq 0$$

$$H_1 : \mu > 0$$

atau

$$H_0 : \mu \geq 0$$

$$H_1 : \mu < 0$$

Jika uji hipotesisnya dua arah maka

$$H_0 : \mu = 0$$

$$H_1 : \mu \neq 0$$

3. **Hitung level signifikansi (α).** Level signifikansi bernilai antara 1 sampai dengan 100, atau $0 \leq \alpha \leq 1$ or $0\% \leq \alpha \leq 100\%$. Semakin kecil nilainya maka semakin kecil kesalahan yang diharapkan terjadi. Biasanya digunakan α sebesar 1%, 5%, atau 10%. Level $\alpha = 5\%$ adalah yang paling sering digunakan.
4. **Menghitung uji statistik test atau *p-value*.** Nilai ini dihitung berdasarkan jenis hipotesis yang akan digunakan yang akan dibahas di sub-bab berikutnya.
5. **Analisis** dengan membandingkan level signifikansi pada tahap 3 dengan nilai statistik pada tahap 4.
6. **Membuat keputusan.** Hasil analisis pada tahap 5 menentukan keputusan. Jika nilai statistik uji lebih dari sama dengan nilai statistik dalam tabel, maka keputusannya adalah tolak H_0 . Pengambilan keputusan juga bisa menggunakan *p-value*, dimana jika *p-value* kurang dari α maka keputusannya adalah tolak H_0 . Dalam membuat keputusan, juga bergantung pada apakah menggunakan satu arah atau dua arah. Penjelasan lebih detil mengenai ini akan dijelaskan di bab selanjutnya dan akan dipraktikkan dalam contoh-contoh soal.

Dalam mengambil keputusan, karena kesalahan pengambilan sampel, bisa terjadi kekeliruan yang tidak diharapkan. Beberapa kemungkinan dalam membuat keputusan ditampilkan dalam kuadran seperti dalam Tabel 5.1 Kesalahan pengambilan keputusan tersebut adalah ke-

salahan tipe I (α) dan kesalahan tipe II (β). Peluang melakukan **kesalahan tipe I** yang diotasikan dengan α adalah peluang melakukan kesalahan menolak hipotesis awal padahal hipotesis tersebut benar. Dalam Tabel 5.1, kesalahan tipe I berada pada kuadran II. Kuadran III yang besarnya adalah $1 - \alpha$ adalah selang kepercayaan yang didefinisikan sebagai peluang membuat keputusan menerima hipotesis awal yang hipotesis tersebut benar. Sementara pada kuadran IV adalah **kesalahan tipe II** yang peluang melakukan kesalahannya dinotasikan dengan β . Peluang melakukan kesalahan tipe II adalah peluang gagal tolak hipotesis awal padahal hipotesis tersebut adalah salah.

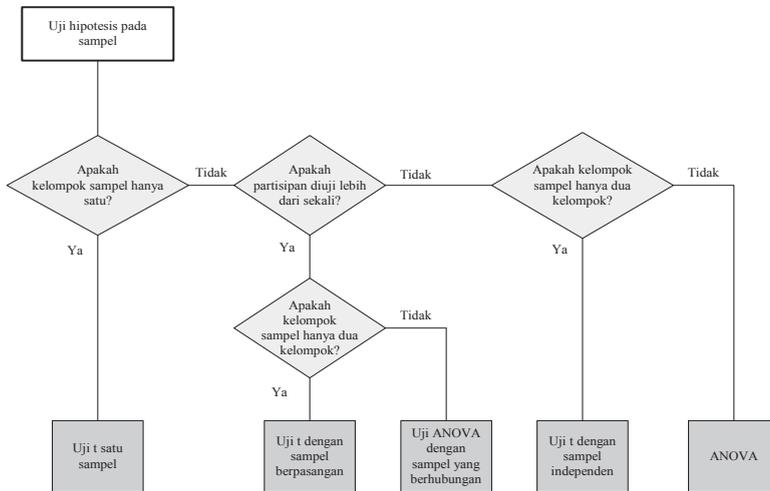
Tabel 5.1: Tipe kesalahan dalam mengambil kesimpulan dalam uji hipotesis

Keputusan	H_0 benar	H_1 benar
Tolak H_0	(Kuadran II) Tipe kesalahan I (α)	(Kuadran I) Keputusan yang benar ($1-\beta$)
Gagal tolak H_0	(Kuadran III) Keputusan yang benar ($1-\alpha$)	(Kuadran IV) Tipe kesalahan II (β)

Dalam uji hipotesis terdapat beberapa uji yang bisa dilakukan. Seperti yang dijelaskan di pendahuluan di bab ini, terdapat uji hipotesis parametrik dan uji hipotesis non-parametrik. Uji hipotesis parametrik digunakan pada variabel-variabel yang mengikuti skala kontinyu atau seringkali mengikuti distribusi normal. Variabel kontinyu di sini seperti variabel yang berskala interval atau rasio. Sedangkan uji hipotesis non-parametrik digunakan ketika variabelnya adalah diskrit. Variabel diskrit misalnya data yang bersifat nominal atau ordinal. Variabel ini seringkali disebut sebagai variabel kategori. Uji hipotesis non-parametrik juga digunakan pada data yang dalam jumlah yang

relatif kecil sehingga dimungkinkan variabel tersebut tidak mengikuti distribusi normal. Jika sampel kurang 30 seringkali dijadikan patokan untuk menggunakan uji hipotesis non parametrik, meskipun variabelnya bersifat kontinyu.

Dalam buku ini akan dibahas uji hipotesis parametrik seperti uji rata-rata untuk satu sampel, uji rata-rata untuk dua sampel, dan uji *analysis of variance*. Sebelum memasuki masing-masing uji hipotesis tersebut, bagan pada Gambar 5.2 dapat dijadikan acuan sekali untuk menentukan jenis uji hipotesis yang akan digunakan.

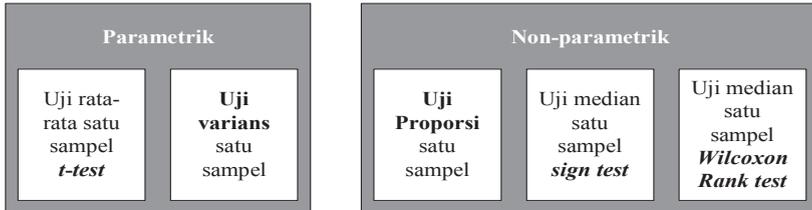


Gambar 5.2: Penentuan uji hipotesis

5.2. Uji Hipotesis untuk Satu Sampel

Uji hipotesis satu sampel digunakan untuk menguji apakah parameter pada sampel tertentu sama dengan yang seperti kita duga. Dalam Gambar 5.3 ditampilkan beberapa alat uji statistik untuk satu sampel, baik yang parametrik maupun non parametrik. Dalam uji parametrik untuk satu sampel terdapat uji rata-rata *t* satu sampel dan uji varians satu sampel. Pada uji non parametrik untuk satu sampel

terdapat uji proporsi satu sampel, uji median satu sampel dengan *sign test* atau *Wilcoxon Rank test*. Namun dalam buku ini hanya dibahas uji hipotesis satu sampel yang parametrik dengan menggunakan uji *t* dan uji non-parametrik dengan uji proporsi.



Gambar 5.3: Uji parametrik dan non-parametrik untuk satu sampel

5.2.1. Uji *t* satu sampel

Uji *t* satu sampel adalah alat uji hipotesis untuk melihat apakah rata-rata suatu populasi sama dengan dugaan kita. Di sini berarti yang akan diuji adalah rata-rata sampel. Pernyataan hipotesis dinyatakan dalam pernyataan satu arah atau dua arah. Pernyataan hipotesis satu arah adalah:

$$H_0: \mu \leq \mu_0 \text{ (rata-rata sampel kurang dari sama dengan } \mu_0 \text{)}$$

$$H_1: \mu > \mu_0 \text{ (rata-rata sampel lebih dari } \mu_0 \text{)}$$

atau

$$H_0: \mu \geq \mu_0 \text{ (rata-rata sampel lebih dari sama dengan } \mu_0 \text{)}$$

$$H_1: \mu < \mu_0 \text{ (rata-rata sampel kurang dari } \mu_0 \text{)}$$

Dinamakan satu arah karena pernyataannya hanya menyatakan satu sisi saja, lebih besar atau lebih kecil. Sedangkan jika menggunakan dua sisi dinamakan dua arah yang pernyataan hipotesisnya adalah:

$$H_0: \mu \geq \mu_0 \text{ (rata-rata sampel sama dengan } \mu_0 \text{)}$$

$$H_1: \mu < \mu_0 \text{ (rata-rata sampel tidak sama dengan } \mu_0 \text{)}$$

Dimana μ_0 adalah nilai rata-rata yang akan diuji. Dengan menggunakan hipotesis satu arah, maka daerah penolakannya adalah:

$$t > t_{\alpha} \text{ untuk } H_0: \mu \leq \mu_0$$

$$t < -t_{\alpha} \text{ untuk } H_0: \mu \geq \mu_0$$

Untuk hipotesis dua arah, daerah penolakannya adalah:

$$|t| > t_{\alpha/2} \text{ untuk } H_0: \mu = \mu_0.$$

Menghitung statistik t untuk rata-rata satu sampel bisa dilakukan dengan dua cara yaitu jika varians diketahui dan jika varians tidak diketahui. Jika varians diketahui, menggunakan Persamaan 5.1 sebagai berikut:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \tag{5.1}$$

dimana σ = standar deviasi yang sudah diketahui

\bar{x} = rata-rata sampel

n = banyaknya data dalam sampel

Nilai σ diperoleh dari informasi yang ada. Nilai μ adalah rata-rata yang akan diujikan. Sedangkan nilai \bar{x} diperoleh dari data.

Sementara, uji rata-rata satu sampel jika varians tidak diketahui menggunakan Persamaan 5.2, yaitu:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \tag{5.2}$$

dimana \bar{x} dan s adalah nilai rata-rata dan standar deviasi yang dihitung dari data sampel. Sedangkan n adalah banyaknya data dalam sampel. Nilai statistik ini yang kemudian digunakan untuk mengambil kesimpulan.

Contoh 5.1

Seorang manajer operasional perusahaan ingin melakukan evaluasi atas pelayanan yang diberikan. Ini karena ada komplain dari pelanggan terkait banyaknya kasir yang tidak cukup banyak. Terutama pada hari Minggu. Dia punya dugaan bahwa selama hari Minggu rata-rata jumlah pembeli yang datang ke tokonya tidak lebih dari 75 pengunjung. Jika melebihi itu maka manajer operasional akan menambah kasir untuk melayani pelanggan. Untuk itu, manajer mengumpulkan data-data pengunjung pada hari Minggu selama setahun. Datanya ditunjukkan pada Tabel 5.2:

Penyelesaian Contoh 5.1:

Hipotesisnya adalah:

$H_0: \mu \leq 75$ (rata-rata pengunjung kurang dari sama dengan 75)

$H_1: \mu > 75$ (rata-rata sampel lebih dari 75)

Tabel 5.2: Rekap pengunjung yang datang pada hari Minggu selama tahun 2019

Waktu	Jumlah pengunjung						
Jan-I	60	Apr-I	107	Jul-I	64	Okt-I	78
Jan-II	75	Apr-II	63	Jul-II	80	Okt-II	78
Jan-III	114	Apr-III	78	Jul-III	72	Okt-III	75
Jan-IV	82	Apr-IV	79	Jul-IV	83	Okt-IV	61
Feb-I	80	Mei-I	75	Aug-I	91	Nov-I	109
Feb-II	85	Mei-II	81	Aug-II	78	Nov-II	77
Feb-III	89	Mei-III	81	Aug-III	74	Nov-III	84
Feb-IV	85	Mei-IV	88	Aug-IV	69	Nov-IV	83
Mar-I	86	Jun-I	94	Sep-I	62	Des-I	59
Mar-II	81	Jun-II	77	Sep-II	78	Des-II	67
Mar-III	97	Jun-III	65	Sep-III	91	Des-III	70
Mar-IV	99	Jun-IV	67	Sep-IV	84	Des-IV	88

Menghitung rata-rata \bar{x} dari Persamaan 4.1:

$$\begin{aligned}\bar{x} &= \frac{\sum_{i=1}^{48} x_i}{n} = \frac{x_1 + x_2 + \dots + x_{48}}{n} \\ &= \frac{60 + 75 + \dots + 88}{48} = \frac{3843}{48} = 80,06\end{aligned}$$

Menghitung s dari Persamaan 4.13:

$$\begin{aligned}s &= \sqrt{\frac{\sum_{n=1}^{48} (x_i - \bar{x})^2}{n - 1}} = \sqrt{\frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_{48} - \bar{x})^2}{n - 1}} \\ &= \sqrt{\frac{(60 - 80,06)^2 + (75 - 80,06)^2 + \dots + (88 - 80,06)^2}{48 - 1}} = \sqrt{\frac{7234,81}{47}} = 12,41\end{aligned}$$

Dari nilai \bar{x} dan s maka dapat dihitung statistik uji dengan menggunakan Persamaan 5.2, yaitu:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{80,06 - 75}{12,41/\sqrt{48}} = 2,8269$$

Dengan $\alpha = 5\%$ pada $df = 48 - 1 = 47$. Dalam Tabel t di Lampiran 2, $df = 47$ ($t_{0,05;47}$) tidak diperoleh dengan jelas nilai t -nya. Namun, ada nilai $t_{0,05;40}$ dan $t_{0,05;60}$, lihat Gambar 5.4.

df	α		
	0,1	0,05	0,025
...
40	1,303	1,684	2,041
60	1,296	1,671	2,021
...

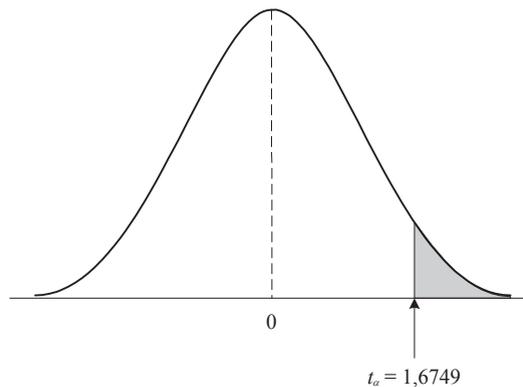
Gambar 5.4: Mencari t dari Tabel distribusi t di Lampiran 2.

Cara mencarinya adalah dengan pendekatan interpolasi linear sebagai berikut:

$$\frac{x - x_1}{x_2 - x_1} = \frac{y - y_1}{y_2 - y_1} \quad (5.3)$$

dimana x adalah nilai yang akan dicari. Sementara x_1 , x_2 , y , y_1 , dan y_2 sudah diketahui. Dalam hal ini akan dicari $t_{0,05;47}$. Diketahui $t_{0,05;40} = 1,684$; $t_{0,05;60} = 1,671$; maka,

$$\begin{aligned} \frac{t_{0,05;47} - t_{0,05;40}}{t_{0,05;60} - t_{0,05;40}} &= \frac{df(47) - df(40)}{df(60) - df(40)} \\ \Leftrightarrow \frac{t_{0,05;47} - 1,684}{1,671 - 1,684} &= \frac{47 - 40}{60 - 40} \\ \Leftrightarrow t_{0,05;47} &= \frac{7}{10}(1,671 - 1,684) + 1,684 \\ \Leftrightarrow t_{0,05;47} &= 1,6749 \end{aligned}$$



Gambar 5.5: Daerah penolakan untuk Contoh 5.1

H_0 ditolak apabila $t > t_\alpha$ untuk $H_0: \mu \leq \mu_0$. Pada persoalan ini didapat t sebesar 2,8269 dan $t_{0,05;47}$ sebesar 1,6749. Ini berarti $t > t_{0,05;47}$. Dalam Gambar 5.5, lokasi t berada di sebelah

kanan t_α yang berarti jatuh di dalam daerah penolakan. Dari sini berarti kesimpulannya adalah tolak H_0 yang mengatakan rata-rata pengunjung kurang dari 75 orang. Oleh karena tidak cukup data untuk mengatakan bahwa jumlah pengunjung tidak lebih dari 75 orang, maka manajer perusahaan harus menambah kasir untuk melayani pelanggan.

5.2.2. Uji proporsi satu sampel

Uji proporsi adalah salah satu uji yang penerapannya banyak dilakukan. Misalnya, ingin mengetahui berapa proporsi orang yang akan membeli produk baru. Bisa juga digunakan dalam politik yaitu untuk menaksir berapa persen pemilih suatu calon. Jika dalam uji rata-rata parameter-nya adalah rata-rata (μ), maka dalam uji proporsi ini parameter-nya adalah proporsi (p).

Pernyataan hipotesis dalam hipotesis satu arah dan dua arah untuk uji proporsi satu sampel ini adalah:

$$H_0: p \leq p_0 \text{ vs } H_1: p > p_0$$

atau

$$H_0: p \geq p_0 \text{ vs } H_1: p < p_0$$

Untuk hipotesis dua arah, dinyatakan dengan:

$$H_0: p = p_0 \text{ vs } H_1: p \neq p_0$$

Dimana p_0 adalah proporsi yang akan diuji. Daerah penolakannya untuk masing-masing hipotesis tersebut adalah $t > t_\alpha$ untuk $H_0: p \leq p_0$ atau $t < -t_\alpha$ untuk $H_0: p \geq p_0$. Untuk hipotesis dua arah, daerah penolakannya adalah $|t| > t_{\alpha/2}$ untuk $H_0: p = p_0$.

Menghitung statistik uji proporsi satu sampel bisa dilakukan dengan menggunakan Persamaan 5.3 berikut:

$$z = \frac{p - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} \quad (5.3)$$

Dimana p adalah proporsi dari sampel dan p_0 adalah nilai proporsi yang akan diujikan. n adalah jumlah sampel.

Contoh 5.2:

Suatu toko alat elektronik yang menjual alat-alat elektroik bertenaga energi matahari ingin mencapai target kepuasan pelanggannya minimal 80% di tahun ini. Manajer toko mendapat informasi bahwa 7 dari 82 pelanggan yang telah membeli panel surya komplain bahwa pemasangan panel surya di rumahnya tidak memuaskan. Dengan menggunakan $\alpha = 5\%$, apakah keinginan mendapatkan kepuasan 80% sudah tercapai?

Penyelesaian Contoh 5.2:

Hipotesisnya adalah

$$H_0: p \leq 0,8 \text{ vs } H_1: p > 0,8$$

Berdasarkan informasi tersebut, diketahui yang tidak puas ada 7 orang dari 82 pelanggannya. Berarti terdapat 75 pelanggan yang puas. Ini berarti $p = 75/82 = 0,9146$. Maka statistik ujinya adalah:

$$z = \frac{p - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{0,9146 - 0,8}{\sqrt{\frac{0,8(1 - 0,8)}{82}}} = 2,5951$$

Dengan $\alpha = 5\%$, $t_{\alpha;81}$ di dalam tabel distribusi t diperoleh dari Lampiran 2.

df	α		
	...	0,05	...
...
60	...	1,671	...
120	...	1,658	...

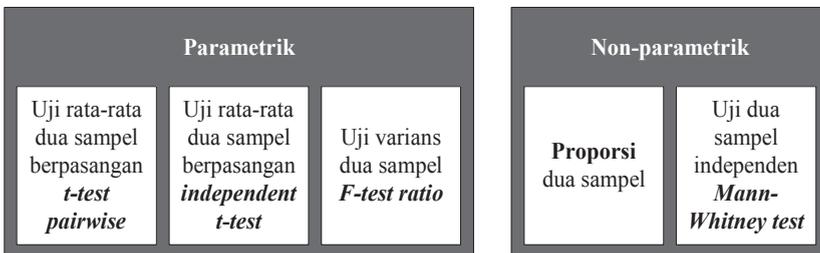
Gambar 5.6: Mencari t dari Tabel distribusi z di Lampiran 1.

Dari tabel distribusi t pada Lampiran 2 didapat nilai $t_{\alpha;60}$ dan $t_{\alpha;120}$ (lihat Gambar 5.6). Dari nilai ini, dengan cara interpolasi dengan menggunakan rumus 5.3 didapat $t_{\alpha;81}$ sebesar 1,993.

Daerah penolakannya adalah ketika $t > t_{\alpha}$. Dari hasil ini berarti t lebih dari t_{α} , sehingga bisa dikatakan bahwa ada cukup data mengatakan bahwa proporsi pelanggan yang puas melebihi 80%. Jika dilihat proporsi pelanggan yang puas diperoleh angka 91,46% yang ini sudah melebihi target yang diharapkan.

5.3. Uji Hipotesis Dua Sampel

Dalam uji rata-rata dua sampel, terdapat beberapa uji parametrik yaitu uji t untuk rata-rata dua sampel yang berpasangan, uji rata-rata dua sampel yang independen,



Gambar 5.7: Uji-uji parametrik dan non-parametrik untuk dua sampel

dan juga bisa dilakukan dengan uji varians dengan menggunakan rasio F. Selain uji parametrik, terdapat uji non-parametrik seperti uji proporsi untuk dua sampel dan uji dua sampel independen dengan menggunakan *Mann-Whitey test*. Uji-uji tersebut digambarkan dalam Gambar 5.7.

5.3.1. Uji rata-rata dua sampel berpasangan

Uji rata-rata dua sampel berpasangan adalah uji untuk dua sampel yang partisipannya (responden) diuji lebih dari sekali, seperti yang ditunjukkan pada Gambar 5.2. Dalam hal ini, jumlah sampel, adalah sama karena obyeknya sama. Uji ini biasanya dilakukan untuk melihat adanya perbedaan ketika ada intervensi. Misalnya, untuk melihat efektifitas suatu pelatihan karyawan. Evaluasinya dilakukan dengan mengukur kinerja karyawan sebelum diberi intervensi pelatihan dan mengukur kinerjanya lagi setelah karyawan diberi pelatihan. Untuk melakukan uji hipotesis dua sampel yang berpasangan yang satu arah, hipotesisnya adalah:

$H_0: \mu_d \leq 0$ (rata-rata selisih antar sampel kurang dari sama dengan 0)

$H_1: \mu_d > 0$ (rata-rata selisih antar sampel lebih dari 0)

atau

$H_0: \mu_d \geq 0$ (rata-rata selisih antar sampel lebih dari sama dengan 0)

$H_1: \mu_d < 0$ (rata-rata selisih antar sampel kurang dari 0)

Pada hipotesis ini, pernyataan jika μ_d nilaiya positif berarti rata-rata sampel pertama lebih besar dari rata-rata sampel kedua. Sebaliknya, jika μ_d bernilai negatif berarti rata-rata sampel pertama lebih kecil daripada rata-rata sampel kedua. Untuk hipotesis dua arah adalah:

$H_0: \mu_d = 0$ (rata-rata selisih antar sampel sama dengan 0)

$H_1: \mu_d \neq 0$ (rata-rata selisih antar sampel tidak sama dengan 0)

Pada pernyataan H_0 arti dari $\mu_d = 0$ adalah, rata-rata sampel pertama sama dengan rata-rata sampel kedua. Jika tidak, maka dalam H_1 dimana rata-rata sampel pertama bisa lebih besar atau lebih kecil dari sampel kedua. Statistik uji untuk ini adalah menggunakan statistik sebagai berikut:

$$z = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}} \quad (5.4)$$

dimana, $\bar{d} = \bar{x}_1 - \bar{x}_2$ atau dapat juga diperoleh dengan cara $\bar{d} = d_i/n$. Nilai d_i diperoleh dari $d_i = x_{1i} - x_{2i}$.

$$s_d = \sqrt{\frac{\sum (d_i - \bar{d})^2}{n - 1}}$$

dengan $df = n - 1$.

Daerah penolakannya adalah:

- $t > t_{\alpha}$ jika hipotesisnya $H_0: \mu_d \leq 0$
- $t < -t_{\alpha}$ jika hipotesisnya adalah $H_0: \mu_d \geq 0$
- $|t| > t_{\alpha/2}$ jika hipotesisnya adalah $H_0: \mu_d = 0$

Contoh 5.3:

Seorang manajer pemasaran akan menguji suatu produk yang merupakan perbaikan dari produk lama. Produk tersebut awalnya menggunakan kemasan yang tidak ramah lingkungan. Namun karena mengikuti permintaan pasar yang semakin peduli lingkungan, bagian pengembangan produk mencoba melakukan rekayasa pengemasan produk ramah lingkungan. Sebelum benar-benar diluncurkan ke pasar, dilakukan riset pasar untuk melihat efektifitas perubahan kemasan. Dalam riset tersebut, pertama, pelanggan diminta

menilai kemasan lama. Kemudian dilanjutkan menilai kemasan baru. Hasilnya ditunjukkan pada Tabel 5.3. Dugaan awal dari riset ini adalah, ada perbedaan penilaian oleh pelanggan atas kemasan.

Tabel 5.3: Penilaian pelanggan terhadap kemasan lama dan baru

No. Resp.	Nilai kemasan		No. Resp.	Nilai kemasan		No. Resp.	Nilai kemasan	
	Lama	Baru		Lama	Baru		Lama	Baru
1.	68	74	11.	69	91	21.	70	65
2.	81	46	12.	78	75	22.	71	69
3.	76	81	13.	64	71	23.	71	78
4.	64	62	14.	67	74	24.	76	67
5.	53	86	15.	74	68	25.	82	84
6.	71	57	16.	81	71	26.	69	85
7.	65	71	17.	75	77	27.	78	80
8.	71	81	18.	58	61	28.	81	68
9.	72	71	19.	74	68	29.	74	83
10.	58	89	20.	76	84	30.	75	64

Penyelesaian Contoh 5.3:

Dalam masalah ini digunakan dua arah, maka hipotesisnya adalah:

$H_0: \mu_d = 0$ (rata-rata selisih penilaian kemasan lama dan baru adalah sama dengan 0)

$H_1: \mu_d \neq 0$ (rata-rata selisih penilaian kemasan lama dan baru adalah tidak sama dengan 0)

Statistik ujinya adalah menggunakan Persamaan 5.4. Untuk menghitung statistik tersebut maka perlu dihitung \bar{d} , maka:

$$\bar{x}_1 = \frac{\sum_{i=1}^n x_{1i}}{n} = \frac{x_{11} + x_{12} + \dots + x_{130}}{n} = \frac{68 + 81 + \dots + 75}{30} = \frac{2142}{30} = 71,40$$

$$\bar{x}_2 = \frac{\sum_{i=1}^n x_{2i}}{n} = \frac{x_{21} + x_{22} + \dots + x_{230}}{n} = \frac{74 + 46 + \dots + 64}{30} = \frac{2201}{30} = 73,37$$

$$\bar{d} = \bar{x}_1 - \bar{x}_2 = 71,40 - 73,37 = -1,9667$$

Selanjutnya dicari s_d yaitu:

$$\begin{aligned} s_d &= \sqrt{\frac{\sum (d_i - \bar{d})^2}{n-1}} = \sqrt{\frac{(d_1 - \bar{d})^2 + (d_2 - \bar{d})^2 + \dots + (d_{30} - \bar{d})^2}{n-1}} \\ &= \sqrt{\frac{((x_{11} - x_{21}) - \bar{d})^2 + ((x_{12} - x_{22}) - \bar{d})^2 + \dots + ((x_{130} - x_{230}) - \bar{d})^2}{n-1}} \\ &= \sqrt{\frac{(((68 - 74) - (-1,97))^2 + ((81 - 46) - (-1,97))^2 + \dots + ((75 - 64) - (-1,97))^2}{30-1}} \\ &= \sqrt{\frac{5190,97}{29}} = 13,3791 \end{aligned}$$

Dengan menggunakan Persamaan 5.4, maka

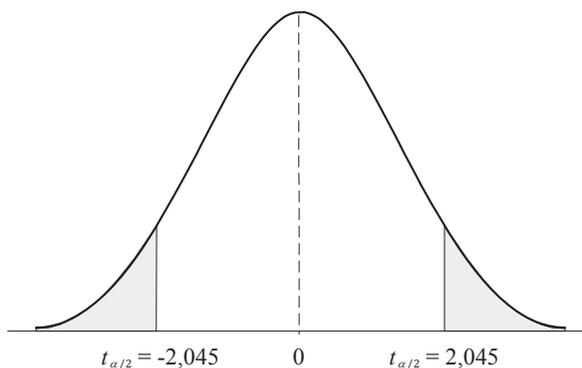
$$z = \frac{\bar{d} - \mu_d}{s_d/\sqrt{n}} = \frac{-1,9667}{13,3791/\sqrt{30}} = -0,8051$$

Dengan menggunakan $\alpha = 5\%$ dan $df = n - 1$, pada Tabel distribusi t pada Lampiran 2, didapat $t_{\alpha/2;29}$ adalah 2,045. Lihat Gambar 5.8 untuk mencari nilai $t_{\alpha/2;29}$ dalam Lampiran 2.

df	α		
	...	0,025	...
...
28	...	2,048	...
29	...	2,045	...
30

Gambar 5.8: Lokasi $t_{\alpha/2;29}$ pada Tabel distribusi t di Lampiran 2

Di sini digunakan $\alpha/2$ karena hipotesisnya menggunakan dua arah. Berdasarkan hasil ini diperoleh $|t| = 0.8051$ dan $t_{\alpha/2;29} = 2,045$ yang artinya $|t| < t_{\alpha/2;29}$. Dari Gambar 5.9 juga bisa dilihat posisi $|t|$ dibandingkan $t_{\alpha/2;29}$ yang berada di luar daerah penolakan (daerah yang diarsir). Oleh karena $|t|$ tidak lebih dari $t_{\alpha/2;29}$, maka H_0 gagal ditolak, yang artinya pelanggan menilai kemasan lama dan baru tidak ada beda secara signifikan.



Gambar 5.9:Daerah penolakan pada Contoh 5.3

5.3.2. Uji rata-rata dua sampel yang independen

Berbeda dengan uji sampel yang berpasangan, uji dua sampel independen digunakan pada partisipan yang berbeda, seperti yang ditunjukkan pada Gambar 5.2. Pada uji ini dimungkinkan variansnya diketahui dan tidak diketahui. Juga dimungkinkan variansnya sama atau tidak sama. Sehingga ada empat variasi yaitu varians diketahui dan tidak sama, varians diketahui dan sama, varians tidak diketahui dan varians tidak sama, dan terakhir adalah varians tidak diketahui dan varians sama. Hipotesis untuk uji rata-rata dua sampel independen untuk uji satu arah adalah:

$$H_0: \mu_1 \leq \mu_2 \text{ vs } H_1: \mu_1 > \mu_2$$

atau $H_0: \mu_1 \geq \mu_2 \text{ vs } H_1: \mu_1 < \mu_2$

Uji hipotesis untuk dua arah adalah:

$$H_0: \mu_1 = \mu_2 \text{ vs } H_1: \mu_1 \neq \mu_2$$

Dengan hipotesis seperti ini, maka daerah penolakannya adalah $t > t_\alpha$ untuk $H_0: \mu_1 \leq \mu_2$, $t < -t_\alpha$ untuk $H_0: \mu_1 \geq \mu_2$, dan $|t| > t_{\alpha/2}$ untuk $H_0: \mu_1 = \mu_2$.

Perbedaan uji rata-rata dua sampel independen menyebabkan rumus penyelesaiannya juga berbeda. Berikut adalah pembahasannya:

- a. **Varians diketahui.** Uji rata-rata pada dua sampel independen yang varians populasinya diketahui menggunakan persamaannya sebagai berikut:

$$z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{5.5}$$

Dimana \bar{x}_1 dan \bar{x}_2 ditaksir dari data, σ_1^2 dan σ_2^2 sudah diketahui. Dalam hal ini, df adalah $v=n_1+n_2-2$.

- b. **Varians tidak diketahui dengan $\sigma_1^2 = \sigma_2^2$.** Varians populasi seringkali tidak diketahui dan varians dua populasi adalah sama, maka menggunakan persamaan sebagai berikut:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} \tag{5.6}$$

dimana

$$S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} \quad (5.7)$$

dengan df adalah $v=n_1+n_2-2$.

- c. **Varians tidak diketahui** dengan $\sigma_1^2 \neq \sigma_2^2$. Pada uji dua sampel ini, statistik ujinya adalah:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\left(\frac{s_1^2}{n_1}\right) + \left(\frac{s_2^2}{n_2}\right)}} \quad (5.8)$$

dengan df adalah:

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{[(s_1^2/n_1)^2/(n_1 - 1)] + [(s_2^2/n_2)^2/(n_2 - 1)]} \quad (5.9)$$

Uji-uji tersebut digunakan pada kasus yang berbeda-beda. Pada kasus dimana varians populasi diketahui, juga dimungkinkan varians sampel pertama dan kedua adalah sama. Juga dimungkinkan jumlah sampel pertama sama dengan sampel kedua. Berikut ini akan diberikan contoh-contoh soal untuk memperjelas penerapan uji-uji tersebut. Ada tiga contoh yaitu Contoh 5.4, 5.5, dan 5.6.

Contoh 5.4

Diketahui bahwa pengetahuan seseorang akan lingkungan mempengaruhinya untuk memutuskan membeli produk-produk ramah lingkungan. Seorang pemasar akan melakukan uji pengetahuan lingkungan di dua daerah untuk membuat keputusan dalam memasarkan produk ramah lingkungannya. Diketahui, varians dari pengetahuan masyarakat terhadap lingkungan adalah 9,56 di daerah A, dan 11,21 di daerah B. Hasil penilaian pengetahuan masyarakat di daerah A dan B ditunjukkan pada Tabel 5.4 berikut:

Tabel 5.4: Penilaian pengetahuan lingkungan masyarakat di daerah A dan B

No	Daerah A	Daerah B	No	Daerah A	Daerah B	No	Daerah A	Daerah B
1.	64	66	14.	58	79	27.	60	77
2.	69	61	15.	42	55	28.	61	54
3.	62	98	16.	61	87	29.	55	63
4.	67	44	17.	61	86	30.	72	67
5.	54	65	18.	73	85	31.	53	90
6.	62	49	19.	47	80	32.	62	59
7.	59	67	20.	65	91	33.		81
8.	39	78	21.	66	78	34.		74
9.	62	87	22.	47	81	35.		57
10.	44	62	23.	56	62	36.		64
11.	52	75	24.	52	53	37.		76
12.	71	54	25.	55	56	38.		56
13.	61	72	26.	78	84	39.		70

Buktikan bahwa pernyataan manajer bahwa pengetahuan masyarakat di Daerah A sama dengan di Daerah B dengan menggunakan $\alpha = 5\%$.

Penyelesaian Contoh 5.4.

Hipotesis pada permasalahan ini adalah:

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Dari permasalahan ini diketahui $\sigma_1 = 9,56$, $\sigma_2 = 11,21$, $n_1 = 32$, dan $n_2 = 39$. Selanjutnya dicari \bar{x}_1 dan \bar{x}_2 adalah:

$$\bar{x}_1 = \frac{\sum x_{1i}}{n} = \frac{x_{11} + x_{12} + \dots + x_{132}}{n} = \frac{64 + 69 + \dots + 62}{32} = \frac{1890}{32} = 59,06$$

$$\bar{x}_2 = \frac{\sum x_{2i}}{n} = \frac{x_{21} + x_{22} + \dots + x_{239}}{n} = \frac{66 + 61 + \dots + 64}{39} = \frac{2743}{39} = 70,33$$

Dengan menggunakan Persamaan 5.5, diperoleh nilai statistik:

$$z = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} = \frac{(59,06 - 70,33)}{\sqrt{\frac{5,96}{32} + \frac{11,21}{39}}} = -14,912$$

df = 32 + 39 - 2 = 69, dengan $\alpha = 5\%$. Nilai $z_{0,025}$ didapat dari Tabel Distribusi z Lampiran 1. Cara mencari nilai ini adalah, di dalam tabel, cari angka yang berilai 0,025 atau yang mendekati. Lihat Gambar 5.10. Dalam Tabel tersebut ditemukan angka 0,0250 yang berada pada -1,9 dan 0,06. Maka berarti $z_{0,025} = -1,96$.

<i>z</i>	...	0,05	0,06	0,07
...	↑	...
-2,0	...	0,0202	0,0197	...
-1,9	←	0,0256	0,0250	...
-1,8	...	0,0322	0,0314	...

Gambar 5.10: Cara mencari $z_{0,025}$ dalam tabel distribusi z

Daerah penolakannya adalah jika $|z| > z_{0,025}$. Karena $|z| = -14,9$ dan $z_{0,025} = -1,96$, maka keputusannya adalah tolak H_0 yang berarti bahwa terdapat cukup data untuk mengatakan bahwa rata-rata pengetahuan lingkungan masyarakat di Daerah A dan B adalah berbeda.

Contoh 5.5

Dari Contoh 5.2 diperoleh informasi bahwa pengunjung pada hari Minggu adalah lebih dari 75 orang per hari. Manajer ingin membandingkan jumlah pengunjung ke toko pada hari kerja. Ini untuk memperkuat keputusan untuk menambah staf atau tidak. Untuk itu diambil data sebanyak

50 yang diambil sampel data secara acak selama satu tahun. Datanya terdapat pada Tabel 5.5. Dari data tersebut, ingin dibuktikan apakah jumlah pengunjung pada hari Minggu lebih banyak dari pengunjung pada hari kerja.

Tabel 5.5: Data pengunjung toko di hari Minggu dan hari kerja

No	Jumlah Pengunjung		No	Jumlah Pengunjung		No	Jumlah Pengunjung	
	Minggu	Hari kerja		Minggu	Hari kerja		Minggu	Hari kerja
1.	60	55	21.	94	62	41.	109	71
2.	75	62	22.	77	67	42.	77	52
3.	114	49	23.	65	40	43.	84	54
4.	82	74	24.	67	79	44.	83	55
5.	80	72	25.	64	77	45.	59	47
6.	85	66	26.	80	82	46.	67	64
7.	89	95	27.	72	62	47.	70	75
8.	85	61	28.	83	60	48.	88	47
9.	86	70	29.	91	62	49.		71
10.	81	64	30.	78	59	50.		51
11.	97	79	31.	74	61	51.		60
12.	99	65	32.	69	67	52.		52
13.	107	60	33.	62	62	53.		80
14.	63	73	34.	78	52	54.		49
15.	78	34	35.	91	41	55.		39
16.	79	51	36.	84	65	56.		41
17.	75	77	37.	78	62	57.		81
18.	81	48	38.	78	47	58.		69
19.	81	68	39.	75	57	59.		79
20.	88	47	40.	61	67	60.		55

Dari informasi, diketahui bahwa standar deviasi pengunjung pada hari Minggu sama dengan standar deviasi pada hari kerja. Di sini akan digunakan $\alpha = 5\%$.

Penyelesaian Contoh 5.5

Hipotesis untuk menguji permasalahan ini adalah:

$H_0 : \mu_1 \leq \mu_2$ (pengunjung pada hari Minggu lebih kecil atau sama dengan hari kerja)

$H_1 : \mu_1 > \mu_2$ (pengunjung pada hari Minggu lebih besar dari hari kerja)

Dari data diketahui $n_1 = 48$ dan $n_2 = 60$. Kemudian diperoleh, $\bar{x}_1 = 80,06$, $s_1^2 = 153,93$ dan $\bar{x}_2 = 61,50$, $s_2^2 = 159,98$. Selanjutnya dicari S_p^2 dengan Persamaan 5.7,

$$S_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{(48 - 1)153,93 + (60 - 1)159,98}{48 + 60 - 2} = 157,3001$$

Dari nilai-nilai ini kemudian digunakan untuk mendapatkan statistik t ,

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{S_p^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{80,06 - 61,50}{\sqrt{157,30 \left(\frac{1}{48} + \frac{1}{60} \right)}} = 7,6429$$

$df = n_1 + n_2 - 2 = 48 + 60 - 2 = 106$, dengan $\alpha = 5\%$. Dari tabel distribusi t pada Lampiran 2 dicari $t_{5\%;106}$. Dengan cara interpolasi $t_{5\%;106}$ dengan cara interpolasi seperti pada Persamaan 5.3 dan pencarian $t_{5\%;106}$ seperti yang ditunjukkan pada Gambar 5.6, maka didapatkan $t_{5\%;106}$ sebesar 1.6610. Daerah penolakannya adalah $t > t_\alpha$. Maka dalam hal ini berarti H_0 ditolak. Artinya, data ini sudah cukup membuktikan pernyataan bahwa rata-rata pengunjung pada hari Minggu lebih banyak daripada hari kerja.

Contoh 5.6

Dari Contoh 5.2 juga ingin dibuktikan bahwa rata-rata pengunjung pada hari Sabtu sama dengan rata-rata pengunjung pada hari Minggu. Data pengunjung hari Sabtu selama setahun ditunjukkan pada Tabel 5.6. Di sini diketa-

hui dari data populasi bahwa standar deviasi pengunjung di hari Minggu tidak sama dengan hari Sabtu. Buktikan bahwa rata-rata pengunjung di hari Minggu sama dengan hari Sabtu dengan selang kepercayaan 95%.

Penyelesaian Contoh 5.6.

Hipotesisnya adalah:

$H_0: \mu_1 = \mu_2$ (pengunjung pada hari Minggu sama dengan hari Sabtu)

$H_1: \mu_1 \neq \mu_2$ (pengunjung pada hari Minggu tidak sama dengan hari Sabtu)

Tabel 5.6: Data pengunjung toko di hari Sabtu

Waktu	Jumlah pengunjung						
1.	82	13.	84	25.	89	37.	92
2.	81	14.	92	26.	72	38.	82
3.	65	15.	66	27.	59	39.	80
4.	64	16.	76	28.	73	40.	93
5.	82	17.	86	29.	69	41.	87
6.	66	18.	68	30.	54	42.	74
7.	77	19.	69	31.	69	43.	52
8.	47	20.	94	32.	69	44.	96
9.	97	21.	72	33.	72	45.	61
10.	72	22.	61	34.	80	46.	86
11.	71	23.	77	35.	84	47.	70
12.	93	24.	74	36.	72	48.	87

Dari permasalahan ini diketahui jumlah data pada sampel pertama dan sampel kedua adalah sama yaitu $n_1 = 48$ dan $n_2 = 48$. Kemudian dari Contoh 5.2 diperoleh, $\bar{x}_1 = 80,06$, dan dari Contoh 5.6 $s_1^2 = 153,93$. Dari data dicari \bar{x}_2 dan s_2^2 adalah:

$$\bar{x}_2 = \frac{\sum x_2}{n_2} = 75,79,$$

$$s_2^2 = \frac{\sum (x_2 - \bar{x}_2)^2}{n_2 - 1} = \frac{6681,92}{48 - 1} = 142,17$$

Dengan Persamaan 5.8, didapatkan nilai t sebesar:

$$t = \frac{(\bar{x}_1 - \bar{x}_2)}{\sqrt{\left(\frac{s_1^2}{n_1}\right) + \left(\frac{s_2^2}{n_2}\right)}} = \frac{(80,06 - 75,79)}{\sqrt{\left(\frac{153,93}{48}\right) + \left(\frac{142,17}{48}\right)}} = 1,7195$$

dengan df adalah:

$$v = \frac{(s_1^2/n_1 + s_2^2/n_2)^2}{\left[\frac{(s_1^2/n_1)^2}{(n_1 - 1)}\right] + \left[\frac{(s_2^2/n_2)^2}{(n_2 - 1)}\right]}$$

$$= \frac{\left(\frac{153,93}{48} + \frac{142,17}{48}\right)^2}{\left[\frac{(153,93/48)^2}{48 - 1} + \frac{(142,17/48)^2}{48 - 1}\right]} = 93,85 \approx 94$$

H_0 ditolak jika $|t| > t_{\frac{\alpha}{2}, df}$. Pada $\alpha = 5\%$, dan $df = 94$, dengan cara interpolasi, diperoleh $t_{0,025;94}$ sebesar 1,989. Dari sini berarti nilai $|t| < t_{0,025;94}$. Artinya, H_0 gagal ditolak sehingga bisa dikatakan bahwa tidak ada cukup data untuk menolak pernyataan bahwa rata-rata pengunjung pada hari Minggu sama dengan rata-rata pengunjung di hari Sabtu.

5.3.3. Uji proporsi dua sampel

Dalam Bab 5.2.2 dibahas uji proporsi untuk satu sampel. Bisa saja dalam suatu masalah akan dilakukan perbandingan proporsi dengan sampel lainnya. Dengan kata lain, ada dua sampel yang akan dibandingkan. Dalam bab ini uji proporsi dua sampel akan dibahas. Hipotesis dalam uji proporsi dua sampel adalah sebagai berikut:

$$H_0: p_1 \leq p_2 \text{ vs } H_1: p_1 > p_2$$

atau

$$H_0: p_1 \geq p_2 \text{ vs } H_1: p_1 < p_2$$

Untuk hipotesis dua arah, dinyatakan dengan:

$$H_0: p_1 = p_2 \text{ vs } H_1: p_1 \neq p_2$$

Di sini p_1 adalah proporsi sampel pertama, sedangkan p_2 adalah proporsi sampel kedua. Kadangkala, hipotesis ini dinyatakan dengan $H_0: p_1 - p_2 \leq 0$, $H_0: p_1 - p_2 \geq 0$, atau $H_0: p_1 - p_2 = 0$.

Daerah penolakannya untuk $H_0: p_1 \leq p_2$ adalah $z > z_{\alpha}$, untuk $H_0: p_1 \geq p_2$ adalah $z < -z_{\alpha}$, dan untuk $H_0: p_1 = p_2$ adalah $|z| > z_{\alpha/2}$. Statistik ujinya dihitung dengan persamaan sebagai berikut:

$$z = \frac{(p_1 - p_2) - (p_{10} - p_{20})}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}} \quad (5.9)$$

Contoh 5.7

Manajer operasi akan melihat hasil uji kualitas air minum dalam botol yang diproduksi malam hari dan siang hari. Dari 500 botol yang diproduksi pada siang hari ternyata terdapat 12 botol yang kualitasnya buruk. Sedangkan pada malam hari didapat 15 botol yang kualitasnya buruk dari 450 botol yang diproduksi. Dengan menggunakan $\alpha = 1\%$, apakah kualitas air minum yang diproduksi di siang hari lebih baik dari kualitas air minum yang diproduksi di malam hari?

Penyelesaian Contoh 5.7

p_1 adalah proporsi air berkualitas di siang hari, yaitu:

$$\text{Total botol air - air yang tidak berkualitas} = 500 - 12 = 488$$

$$p_1 = 488/500 = 0,976$$

p_2 adalah proporsi air berkualitas di malam hari, yaitu:

Total botol air - air yang tidak berkualitas = $450 - 15 = 435$

$$p_2 = 435/450 = 0,967$$

Hipotesisnya adalah $H_0: p_1 \leq p_2$ vs $H_1: p_1 > p_2$

$$z = \frac{(p_1 - p_2) - (p_{10} - p_{20})}{\sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}}}$$

$$= \frac{0,976 - 0,967}{\sqrt{\frac{0,976(1-0,976)}{500} + \frac{0,967(1-0,967)}{450}}}$$

$$= 0,8575$$

$$df = 500 + 450 - 2 = 948$$

Nilai $z_\alpha = z_{0,01}$, diperoleh dari tabel distribusi z di Lampiran 1. Dari tabel tersebut didapat angka 0,0051 yang berada di -2,5 dan 0,07 dan 0,049 yang berada di -2,5 dan 0,08 (lihat Gambar 5.11). Dengan cara interpolasi dengan menggunakan $z_{0,0102}$ dan $z_{0,0099}$ didapat $z_{0,01} = -2,327$.

z	...	0,02	0,03	0,04
...
-2,4	...	0,0078	0,0075	...
-2,3	...	0,0102	0,0099	...
-2,2

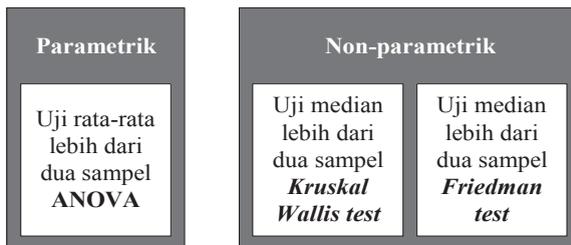
Gambar 5.11: Mencari $z_{0,01}$ pada tabel distribusi z

Dari masalah ini diperoleh $z = 0,8575$ yang ternyata lebih besar dari $z_{0,01}$. Daerah penolakan adalah $z > z_\alpha$. Karena z lebih besar dari z_α maka dapat dikatakan bahwa terdapat cukup bukti untuk menolak pernyataan H_0 yang mengatakan bahwa proporsi air berkualitas yang diproduksi di siang

hari kurang dari atau sama dengan proporsi air berkualitas yang diproduksi di malam hari. Dari nilai rata-rata juga terlihat bahwa proporsi air berkualitas yang diproduksi pada siang hari lebih besar dari yang diproduksi malam hari.

5.4. Uji Hipotesis lebih dari Dua Sampel: *Analysis of Variance*

Seringkali, sampel yang akan dibandingkan lebih dari dua sampel. Statistik uji untuk dua sampel yang telah dibahas pada Bab 5.3 di atas tidak bisa mengatasi jika sampelnya lebih dari dua. Jika mau tetap menggunakan maka harus dilakukan berpasangan. Misalnya ada tiga sampel yang akan diuji, sebut saja A, B, dan C. Jika akan menggunakan statistik uji untuk dua sampel maka dilakukan tiga kali uji yaitu uji sampel A dan B, uji sampel B dan C, dan uji sampel A dan C. Jika sampelnya lebih banyak, maka dilakukan uji lebih banyak lagi.



Gambar 5.12: Uji parametrik dan non-parametrik pada lebih dari dua sampel

Analysis of Variance adalah alat uji statistik parametrik yang dapat digunakan untuk menguji rata-rata lebih dari dua sampel secara bersamaan. Sedangkan, jika datanya adalah non-parametrik, uji *Kruskal Wallis* dan uji *Friedman*. Dalam buku ini hanya akan dibahas *Analysis of Variance* (ANOVA).

ANOVA sering disebut juga dengan ANAVA yang merupakan kepanjangan dari Analisis Varians. ANOVA dapat digunakan untuk uji beberapa sampel secara serempak. Uji ini dapat digunakan untuk mendeteksi adanya efek dari suatu faktor atau perlakuan terhadap variabel respons.

Andaikan terdapat a sampel yang dalam konteks ANOVA sampel-sampel tersebut disebut sebagai **perlakuan**. Biasanya, ANOVA digunakan untuk menganalisis eksperimen yang mengguakan beberapa perlakuan. Masing-masing perlakuan tersebut terdiri dari n pengamatan. Jumlah pegamatan pada perlakuan bisa sama, bisa juga berbeda. Struktur datanya ditunjukkan pada Tabel 5.7. Dalam struktur data tersebut, hasil pengamatan tiap perlakuan dijumlahkan dan hasilnya dimasukkan dalam dalam kolom total. Juga, setiap perlakuan dihitung rata-ratanya menggunakan Persamaan 4.1.

Tabel 5.7: Struktur data untuk ANOVA

Perlakuan	Pengamatan				Total	Rata-rata
	1	2	..	n		
1	y_{11}	y_{12}	...	y_{1n}	$y_{1.}$	$\bar{y}_1.$
2	y_{21}	y_{22}	...	y_{2n}	$y_{2.}$	$\bar{y}_2.$
...
a	y_{a1}	y_{a2}	...	y_{2n}	$y_{a.}$	$\bar{y}_a.$
total					$y_{..}$	$\bar{y}_{..}$

Hipotesis uji ANOVA ini adalah:

$$H_0: \mu_1 = \mu_2 = \dots = \mu_a = 0$$

(rata-rata semua perlakuan adalah sama atau bisa dikatakan tidak pengaruh perlakuan terhadap respons)

$$H_1: \mu_i \neq 0$$

(minimal ada satu perlakuan yang rata-ratanya tidak sama atau dengan kata lain ada pengaruh perubahan perlakuan terhadap respons)

Pada H_1 , jika hanya ada satu perlakuan yang rata-ratanya tidak sama dengan perlakuan lainnya, maka dianggap tidak sama. Ini karena pada H_0 dinyatakan semua perlakuan rata-ratanya adalah sama.

Dari struktur data tersebut kemudian data diolah dan menggunakan Tabel ANOVA seperti yang ditunjukkan pada Tabel 5.8.

Tabel 5.8: Tabel ANOVA

Sumber variasi	Derajat kebebasan (df)	Jumlah kuadrat (SS)	Rata-rata kuadrat (MS)	F
Antar perlakuan	$a - 1$	$SS_{\text{Perlakuan}}$	$MS_{\text{Perlakuan}}$	$\frac{MS_{\text{Perlakuan}}}{MSE}$
Error (dalam perlakuan)	$a(n - 1)$	SSE	MSE	
Total	$an - 1$	SST		

Penyelesaian Tabel ANOVA tersebut menggunakan persamaan-persamaan sebagai berikut:

$$SS_{\text{Perlakuan}} = \sum_{i=1}^a \frac{y_{i.}^2}{n} - \frac{y_{..}^2}{an} \quad (5.10)$$

$$SST = \sum_{i=1}^a \sum_{j=1}^n y_{ij}^2 - \frac{y_{..}^2}{an} \quad (5.11)$$

$$SSE = SST - SS_{\text{Perlakuan}} \quad (5.12)$$

Daerah penolakannya adalah $F > F_{\alpha; a-1, a(n-1)}$. Nilai $F_{\alpha; a-1, a(n-1)}$ diperoleh dari Tabel distribusi F pada Lampiran 3 dan 4.

Contoh 5.8

Manajer operasional perusahaan air mineral akan menguji 4 jenis mesin yang akan mengolah air minum. Keempat mesin tersebut menghasilkan air dengan jumlah *Total Dissolve Solid* (TDS) yang berbeda. Standar TDS yang diperbolehkan adalah di bawah 500 mg/l. Namun, perusahaan air mineral menargetkan TDS di bawah 400 mg/l. Mesin pengolahan air yang mampu menghasilkan TDS di bawah 400 akan menjadi prioritas. Hasil pengukuran TDS dari masing-masing mesin ditunjukkan di Tabel 5.9.

Tabel 5.9: Hasil pengujian TDS yang dihasilkan oleh tiap mesin

Replikasi	Total Dissolve Solid (TDS)			
	Mesin A	Mesin B	Mesin C	Mesin D
1	391.10	408.92	410.50	389.91
2	391.22	388.59	400.80	381.90
3	401.50	392.76	403.77	384.67
4	370.02	393.63	420.01	390.33
5	402.14	372.71	407.34	380.20
6	397.74	394.84	415.55	388.47
7	384.82	404.34	372.92	400.21

Dengan menggunakan $\alpha = 5\%$, manajer operasional ingin mengetahui apakah hasil TDS masing-masing mesin pengolahan air berbeda.

Penyelesaian Contoh 5.9

Uji hipotesis pada permasalahan ini adalah:

$$H_0: \mu_1 = \mu_2 = \mu_3 = \mu_4 = 0$$

$$H_1: \mu_i \neq 0 \text{ di mana } i = 1, 2, 3, \text{ dan } 4.$$

$\mu_1 \neq 0$ bisa berarti:

$$\mu_1 \neq \mu_2 = \mu_3 = \mu_4 = 0$$

$$\mu_1 = \mu_2 \neq \mu_3 = \mu_4 = 0$$

$$\mu_1 = \mu_2 = \mu_3 \neq \mu_4 = 0$$

$$\mu_1 \neq \mu_2 \neq \mu_3 = \mu_4 = 0$$

$$\mu_1 = \mu_2 \neq \mu_3 \neq \mu_4 = 0$$

$$\mu_1 \neq \mu_2 = \mu_3 \neq \mu_4 = 0$$

$$\mu_1 \neq \mu_2 \neq \mu_3 \neq \mu_4 = 0$$

Data pada Tabel 5.9 kemudian diolah ke dalam Tabel 5.7

Per-lakuan	Pengamatan							Total	Rata-rata
	1	2	3	4	5	6	7		
A	391,10	391,22	401,50	370,02	402,14	397,74	384,82	2738,53	391,22
B	408,92	388,59	392,76	393,63	372,71	394,84	404,34	2755,79	393,68
C	410,50	400,80	403,77	420,01	407,34	415,55	372,92	2830,90	404,41
D	389,91	381,90	384,67	390,33	380,20	388,47	400,21	2715,68	387,95
Total								11040,90	394,32

Selanjutkan menghitung persamaan 5.10 sampai 5.12, di mana

$$\begin{aligned}
 SS_{\text{perlakuan}} &= \sum_{i=1}^a \frac{y_i \cdot^2}{n} - \frac{y_{..} \cdot^2}{an} = \frac{(y_{1.} \cdot^2 + y_{2.} \cdot^2 + y_{3.} \cdot^2 + y_{4.} \cdot^2)}{n} - \frac{y_{..} \cdot^2}{an} \\
 &= \frac{(2738,53^2 + 2755,79^2 + 2830,90^2 + 2715,68^2)}{7} - \frac{11040^2}{4(7)} \\
 &= 1066,96
 \end{aligned}$$

$$\begin{aligned}
 SST &= \sum_{i=1}^a \sum_{j=1}^n y_{ij}^2 - \frac{y_{..} \cdot^2}{an} = (y_{12}^2 + y_{12}^2 + \dots + y_{44}^2) - \frac{y_{..} \cdot^2}{an} \\
 &= (391,10^2 + 391,22^2 + \dots + 400,21^2) - \frac{11040,90^2}{4(7)} \\
 &= 4324,68
 \end{aligned}$$

$$SSE = SST - SS_{\text{perlakuan}} = 4324,68 - 1066,96 = 3257,73$$

Tabel 5.10: Tabel ANOVA hasil oleh data

Sumber variasi	Derajat kebebasan (df)	Jumlah kuadrat (SS)	Rata-rata kuadrat (MS)	F
Antar perlakuan	3	1066,96	355,65	2,62
Error (dalam perlakuan)	24	3257,73	135,74	
Total	27	4325,68		

Daerah penolakannya adalah $F > F_{\alpha; a-1, a(n-1)}$. Nilai $F_{5\%; 3, 24}$ diperoleh dari Tabel distribusi F pada Lampiran 3. Caranya mencarinya seperti ditunjukkan pada Gambar 5.13.

$v_2 \backslash v_1$	Derajat bebas (df) pembilang			
	...	2	3	...
...	↓	...
23	↓	...
24	→	...	3,01	...
25

Gambar 5.13: Tabel distribusi F di Lampiran 4 untuk mencari $F_{5\%; 3, 24}$

Dari tabel diperoleh $F_{5\%; 3, 24} = 3,01$, sedangkan F dari ANOVA diperoleh nilai 2,62. Di sini berarti $F < F_{5\%; 3, 24}$. Ini berarti keputusannya adalah gagal tolak H_0 . Artinya, dari data dibuktikan bahwa rata-rata mesin A sampai D menghasilkan TDS yang bisa dibilang relatif sama rata-ratanya. Namun, karena manajer operasional memprioritaskan mesin yang mampu menghasilkan TDS di bawah 300, maka mesin C tidak jadi prioritas untuk dipilih.

BAB 6.

ANALISIS KORELASI DAN REGRESI

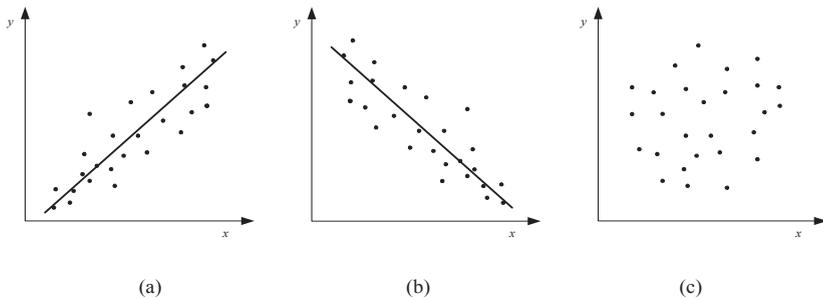
DALAM statistika inferensia, data sampel diambil secara reperiensatif untuk dianalisis dan diambil kesimpulan berdasarkan data tersebut. Salah satu analisis yang sering digunakan di dalam statistika adalah analisis untuk mencari pola hubungan antar variabel. Francis Galton dan Karl Pearson adalah ilmuwan-ilmuwan yang menggunakan regresi untuk pertama kalinya (Gujarati, 2003). Galton menggunakan regresi untuk menyimpulkan bahwa tinggi badan anak laki-laki cenderung diturunkan dari ayahnya. Pearson kemudian menguatkan dengan menggunakan sampel lebih banyak. Dalam bab ini akan dibahas analisis korelasi dan analisis regresi. Melalui bab ini, diharapkan pembaca:

- Mampu memahami dan menerapkan analisis korelasi
- Mampu memahami, membedakan, dan menerapkan analisis korelasi *Product Moment Pearson* dan *Rank Spearman*.
- Mampu memahami dan menerapkan analisis regresi linear sederhana

6.1. Analisis Korelasi

Galton dan Pearson, dua ahli rekayasa genetika dari Inggris pada tahun 1890an menggunakan analisis korelasi dan regresi. Analisis digunakan untuk mencari hubungan

asosiasi antar dua variabel, bukan hubungan kausal (Herkenhoff & Fogli, 2013). Lebih mudahnya, pola hubungan antar dua variabel dapat dideteksi dengan menggunakan *scatterplot* yang telah dibahas pada Bab 4.1.2.



Gambar 6.1: Pola hubungan antar dua variabel.

Gambar 6.1 menggambarkan hubungan antar dua variabel, sebut saja variabel x pada garis horizontal dan variabel y pada garis vertikal. Pada Gambar 6.1(a) x bertambah diikuti y yang bertambah. Kemiringannya hubungannya ke kanan. Ini dinamakan hubungan positif yang artinya jika x bertambah naik, y juga akan ikut bertambah naik. Lain halnya dengan Gambar 6.1(b) dimana jika x bertambah naik, y malah sebaliknya turun nilainya. Ini yang dinamakan hubungan negatif. Gambar 6.1(a) dan (b) arah hubungannya nampak jelas. Pada Gambar 6.1(a) dan (b) ini pola hubungannya adalah linear. Sementara, Gambar 6.1(c) tidak jelas arahnya. Ini adalah contoh pola hubungan antara variabel x dan y yang tidak ada hubungannya.

6.1.1. Korelasi *Product Moment Parson*

Koefisien korelasi *product moment Pearson* adalah korelasi yang seringkali digunakan dalam analisis korelasi. Data kontinyu seperti data berskala rasio dan interval adalah syarat dalam menggunakan analisis *product moment Pearson*

ini. Menghitung korelasi ini menggunakan persamaan sebagai berikut:

$$\rho_{xy} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \quad (6.1)$$

dimana ρ_{xy} = korelasi *product moment Pearson*
 x_i = data ke- i dari variabel x
 y_i = data ke- i dari variabel y
 \bar{x} = rata-rata dari variabel x
 \bar{y} = rata-rata dari variabel y
 n = banyaknya data

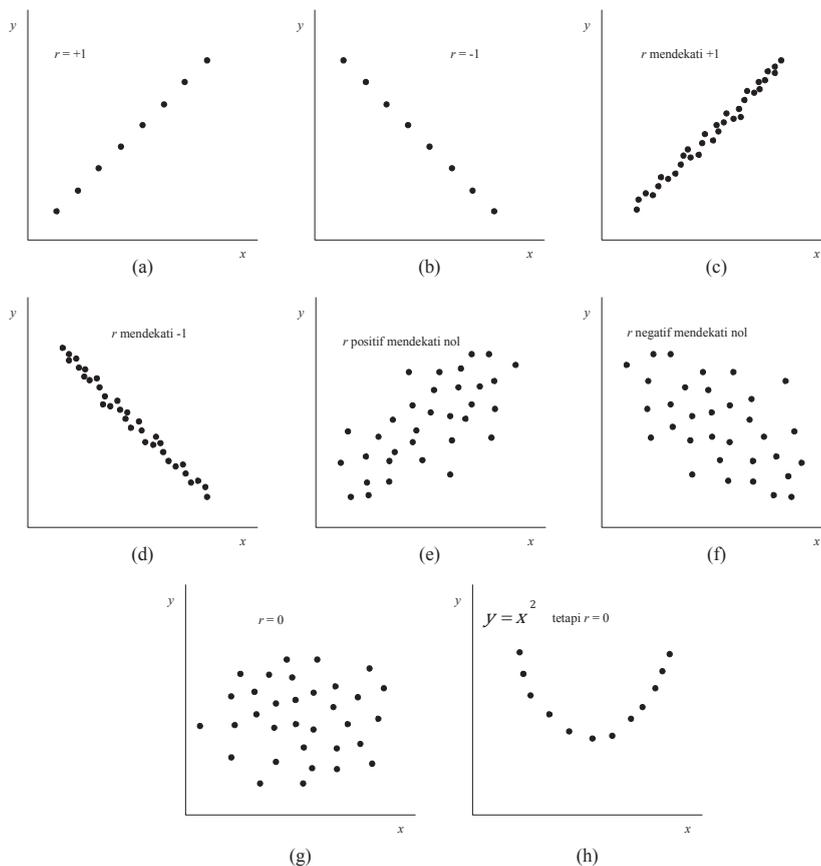
Persamaan 6.1 jika dijabarkan akan berbentuk seperti berikut:

$$\rho_{xy} = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{\sqrt{n \sum_{i=1}^n x_i^2 - (\sum_{i=1}^n x_i)^2} \sqrt{n \sum_{i=1}^n y_i^2 - (\sum_{i=1}^n y_i)^2}} \quad (6.2)$$

Nilai korelasi berkisar antara - 1 sampai dengan 1 atau $-1 \leq \rho \leq 1$. Nilai korelasi yang nilai korelainya (ρ) bernilai 0 berarti tidak ada korelasi antar dua variabel. Namun, apabila nilainya negatif berarti memiliki hubungan negatif dimana jika variabel satu bertambah nilainya, variabel lainnya akan berkurang. Jika korelasi bernilai positif maka hubungan antar dua variabel tersebut adalah positif. Hubungan positif artinya, apabila satu variabel bertambah maka variabel lain ikut bertambah. Sebaliknya, apabila satu variabel bernilai negatif, maka variabel lainnya bernilai negatif. Semakin korelasi mendekati nilai 1 atau -1 maka dikatakan hubungannya semakin kuat. Salkind (2008) mengusulkan kekuatan korelasi seperti ditunjukkan dalam Tabel 6.1. Gambar 6.2 memberikan ilustrasi besaran korelasi jika dalam bentuk grafis.

Tabel 6.1: Kekuatan korelasi menurut Salkind (2008)

Ukuran korelasi	Interpretasi dari koefisien korelasi
0,8 sampai 1,0	Hubungan sangat kuat
0,6 sampai 0,8	Hubungan kuat
0,4 sampai 0,6	Hubungan moderat
0,2 sampai 0,4	Hubungan lemah
0,0 sampai 0,2	Tidak ada hubungan atau hubungan lemah



Gambar 6.2: Pola korelasi diadopsi dari Henri Theil (1978) dalam bukunya *Introduction to Econometrics* yang dikutip oleh (Gujarati, 2003)

Contoh 6.1:

Dalam Tabel 4.6 adalah data hasil pengukuran berat badan (y) dan tinggi badan (x) pada 20 orang. Data tersebut ditampilkan lagi di Tabel 6.2. Akan dicari korelasi antara tinggi badan dan berat badan.

Tabel 6.2: Data tinggi badan dan berat badan.

No	Tinggi Badan (cm)	Berat Badan (kg)	No	Tinggi Badan (cm)	Berat Badan (kg)	No	Tinggi Badan (cm)	Berat Badan (kg)	No	Tinggi Badan (cm)	Berat Badan (kg)
1.	178	70	6.	151	55	11.	156	50	16.	155	73
2.	170	65	7.	165	78	12.	158	49	17.	85	21
3.	130	30	8.	169	80	13.	163	58	18.	152	59
4.	152	47	9.	163	60	14.	157	56	19.	148	44
5.	143	40	10.	170	75	15.	161	70	20.	158	31

Penyelesaian Contoh 6.1:

Dalam menyelesaikan masalah ini, menghitung korelasi menggunakan rumus pada Persamaan 6.1, Untuk menyelesaikan dengan menggunakan Persamaan 6.1 perlu dihitung terlebih dahulu rata-rata tinggi badan (\bar{x}) dan rata-rata berat badan (\bar{y}). Maka,

Rata-rata tinggi badan (\bar{x}) adalah:

$$\begin{aligned}\bar{x} &= \frac{\sum_{i=1}^n x_i}{n} \\ &= \frac{x_1 + x_2 + \dots + x_{20}}{20} \\ &= \frac{178 + 170 + \dots + 158}{20} \\ &= \frac{3084}{20} \\ &= 154,20\end{aligned}$$

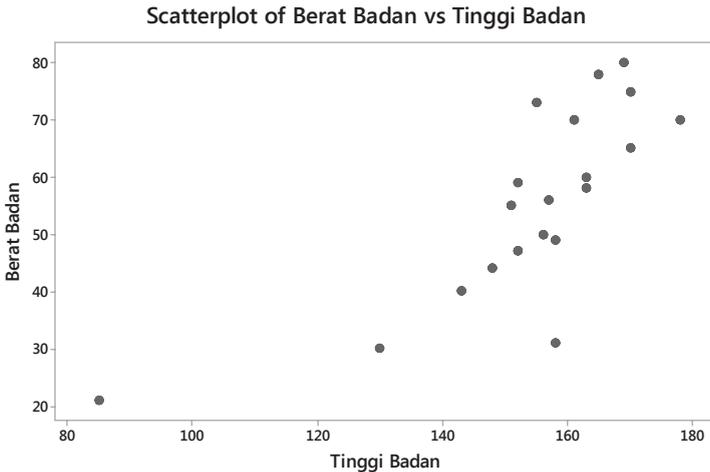
Rata-rata berat badan (\bar{y}) adalah:

$$\begin{aligned}\bar{y} &= \frac{\sum_{i=1}^n y_i}{n} \\ &= \frac{y_1 + y_2 + \dots + y_{20}}{20} \\ &= \frac{70 + 65 + \dots + 31}{20} \\ &= \frac{1111}{20} \\ &= 55,55\end{aligned}$$

Berdasarkan hasil ini, maka

$$\begin{aligned}
 \rho_{xy} &= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}} \\
 &= \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_{20} - \bar{x})(y_{20} - \bar{y})}{\sqrt{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_{20} - \bar{x})^2} \sqrt{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \dots + (y_{20} - \bar{y})^2}} \\
 &= \frac{(178 - 154,20)(70 - 55,55) + \dots + (158 - 154,20)(31 - 55,55)}{\sqrt{(178 - 154,20)^2 + \dots + (158 - 154,20)^2} \sqrt{(70 - 55,55)^2 + \dots + (31 - 55,55)^2}} \\
 &= \frac{(23,80)(14,45) + \dots + (3,80)(-24,55)}{\sqrt{(23,80)^2 + \dots + (3,80)^2} \sqrt{(14,45)^2 + \dots + (-24,55)^2}} \\
 &= \frac{4718,8}{\sqrt{7201,20} \sqrt{5300,95}} \\
 &= 0,7638
 \end{aligned}$$

Korelasi tinggi badan dan berat badan dalam kasus ini adalah 0,76. Jika menggunakan pendapat Salkind (2008), korelasi 0,76 ini dikatakan memiliki hubungan asosiasi yang kuat. Dalam bentuk grafis, *scatterplot* hubungan antara tinggi badan dan berat badan ditunjukkan pada Gambar 6.3.



Gambar 6.3: *Scatterplot* hubungan tinggi badan dan berat badan yang korelasinya 0,76.

6.1.2. Korelasi *Rank Spearman*

Jika korelasi *Product Moment Pearson* digunakan untuk mendeteksi hubungan antar dua variabel kontinyu, korelasi *Rank Spearman* digunakan untuk mendeteksi hubungan asosiasi dua variabel yang datanya berskala kontinyu atau ordinal. Korelasi *Rank Spearman* ini adalah versi nonparametrik dari analisis korelasi. *Rank Spearman* mampu mendeteksi hubungan monotonik, yaitu jika satu variabel naik maka variabel lainnya juga naik, atau sebaliknya jika satu variabel naik maka variabel lainnya akan turun. Persamaan untuk menghitung korelasi *Rank Spearman* adalah sebagai berikut:

$$\rho_{xy} = 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \quad (6.3)$$

dimana d = selisih dari pasangan *rank* ke- i
 n = banyaknya data

Dalam *Rank Spearman* ini, jika data yang akan dianalisis adalah data kontinyu, maka perlu mengurutkan data terlebih dahulu atau mengubahnya dalam data ordinal. Setelah itu, dihitung selisih ranking data pada satu pengamatan. Lebih detil mengenai penerapan korelasi *Rank Spearman* akan diberikan contoh menghitungnya di Contoh 6.2.

Contoh 6.2.

Data yang digunakan adalah data pada Contoh 6.1. Untuk keperluan dengan menghitung korelasi *Rank Spearman*, data pada Tabel 6.2 perlu dibuat ranking dari tertinggi sampai terendah untuk masing-masing variabel. Hasilnya ditunjukkan pada Tabel 6.3.

Tabel 6.3: Data tinggi badan dan berat badan.

No	Tinggi Badan (cm)	Ranking Tinggi Badan	Berat Badan (kg)	Ranking Berat Badan	d
1.	178	1	70	5	-4
2.	170	2	65	6	-4
3.	130	15	30	18	-3
4.	152	11	47	14	-3
5.	143	14	40	16	-2
6.	151	12	55	11	1
7.	165	4	78	2	2
8.	169	3	80	1	2
9.	163	5	60	7	-2
10.	170	2	75	3	-1
11.	156	9	50	12	-3
12.	158	7	49	13	-5
13.	163	5	58	9	-4
14.	157	8	56	10	-2
15.	161	6	70	5	1
16.	155	10	73	4	6
17.	85	16	21	19	-3
18.	152	11	59	8	3
19.	148	13	44	15	-2
20.	158	7	31	17	-10

Penyelesaian Contoh 6.2:

Dengan menggunakan Persamaan 6.3, maka

$$\begin{aligned} \rho_{xy} &= 1 - \frac{6 \sum_{i=1}^n d_i^2}{n(n^2 - 1)} \\ &= 1 - \frac{6(d_1^2 + d_2^2 + \dots + d_{10}^2)}{n(n^2 - 1)} \end{aligned}$$

$$\begin{aligned}
&= 1 - \frac{6((-4)^2 + (-4)^2 + \dots + (-10)^2)}{20(20^2 - 1)} \\
&= 1 - \frac{6(281)}{20(400 - 1)} \\
&= 0,7887
\end{aligned}$$

Korelasi *Rank Spearman* antara tinggi badan dan berat badan adalah sebesar 0,7887. Nilai ini lebih besar dari korelasi *Product Moment Pearson*. Dalam beberapa literatur dikatakan, korelasi *Rank Spearman* lebih sensitif. Terutama jika terdapat hubungan monotonik.

6.2. Dasar-dasar Analisis Regresi

6.2.1. Apa Analisis Regresi?

Dalam statistika inferensia, salah satu analisis yang diinginkan adalah mencari hubungan antara suatu variabel atau lebih terhadap variabel lainnya. Analisis korelasi hanya dapat mendeteksi hubungan asosiasi antar dua variabel. Selain itu, terdapat analisis **regresi** yang dapat digunakan untuk mendeteksi hubungan antar dua variabel atau lebih. Namun, analisis regresi memiliki kelebihan dibanding analisis korelasi. Analisis regresi berkaitan dengan studi satu variabel sebagai variabel dependen dan satu variabel sebagai variabel independen untuk menaksir atau memprediksi nilai rata-rata (populasi) dari nilai-nilai tetap dalam pengambilan sampel yang diulang (Gujarati, 2003). Gujarati (2003) mengatakan bahwa meski analisis regresi menyatakan keterkaitan antara variabel bebas pada variabel lain, tetapi tidak bisa dikatakan sebagai hubungan kausalitas. Ini karena menurut Kendall dan Stuart, “Hubungan statistika, meskipun kuat dan sugestif, tidak dapat menyatakan

hubungan kausal. Ide kausalitas berasal dari luar statistik yaitu dari beberapa teori atau lainnya". Tanpa ada teori pendukung dan logika berfikir yang masuk akal, walaupun analisis regresi menghasilkan nilai hubungan yang besar, tidak akan memberi arti apapun.

Dalam analisis regresi terdapat variabel independen dan variabel dependen. Variabel **independen** atau variabel **bebas** disebut juga sebagai variabel **prediktor**. Variabel ini seringkali dinyatakan sebagai variabel x . Sedangkan variabel **dependen** atau variabel **tidak bebas** juga disebut sebagai variabel **respon**. Variabel ini biasanya dinotasikan dengan y . Variabel dependen maupun variabel independen bisa lebih dari satu. Model regresi yang paling sederhana adalah model regresi bivariat atau model regresi dua variabel atau disebut juga dengan model regresi linear sederhana. Dalam model ini hanya terdapat satu variabel independen (x) dan satu variabel dependen (y). Model ini dinyatakan dalam persamaan sebagai berikut:

$$y = \beta_0 + \beta_1 x + \varepsilon \quad (6.4)$$

dimana y = variabel respon

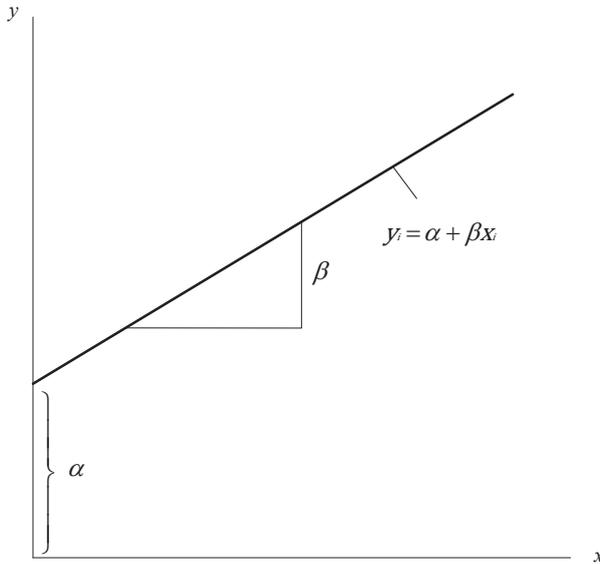
x = variabel prediktor

β_0 = konstanta atau *intercept*

β_1 = parameter dari variabel prediktor (x) atau *slope*

ε = residual atau *error* dari model

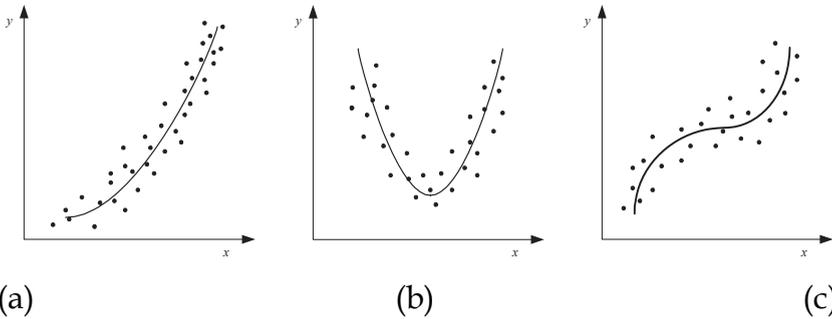
Parameter α seringkali juga dinotasikan dengan notasi lain yaitu β_0 dan parameter β dinotasikan dengan β_1 . Parameter α adalah konstanta yang menyatakan besaran saat x bernilai nol. Sedangkan parameter β adalah *slope* atau kemiringan dari model yang menyatakan efek. Dalam grafik, persamaan model regresi dinyatakan dalam Gambar 6.4. Parameter β ini yang akan ditaksir nilainya.



Gambar 6.4: Gambar regresi dalam diagram *scatterplot*

6.2.2. Linearitas dalam Model Regresi

Dari parameter-parameter pada Persamaan 6.4, yaitu α dan β terlihat bahwa model regresi tersebut adalah model regresi linear (MRL). Model tersebut juga linear dalam variabelnya, yaitu x dan y . Jika divisualisasikan dalam diagram *scatterplot* pola hubungannya seperti ditunjukkan pada Gambar 6.1(a) dan Gambar 6.1(b). Namun, ada model regresi yang non linear (MRNL). Gambar 6.5 adalah contoh pola hubungannya antar variabel yang parameternya linear, namun non linear dalam variabel. Gambar 6.5 (a) adalah pola hubungan dua variabel yang eksponensial dengan persamaannya adalah $y = e^{\beta_0 + \beta_1 x}$. Gambar 6.5(b) adalah pola hubungan variabel yang kuadrat yang persamaannya adalah $y = \beta_0 + \beta_1 x + \beta_2 x^2$, dan Gambar 6.5(c) adalah pola hubungan variabel yang kubik yang persamaannya $y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3$.



Gambar 6.5: Pola hubungan linear dalam parameter namun, non linear dalam variabel.

Gujarati (2003) memberikan panduan bagaimana model regresi linear dan model regresi non linear dalam bentuk tabel di Tabel 6.4. Dalam tabel tersebut, selama parameter-nya adalah linear, maka model tersebut dikatakan **model regresi linear** meskipun variabelnya tidak linear. Namun, jika parameter modelnya non linear, maka model regresi dinamakan **model regresi non linear** walaupun variabelnya linear.

Tabel 6.4: Model-model regresi berdasarkan linearitas parameter dan variabel

Model linear dalam parameter?	Model linear dalam variabel?	
	Iya	Tidak
Iya	Model regresi linear	Model regresi linear
Tidak	Model regresi non linear	Model regresi non linear

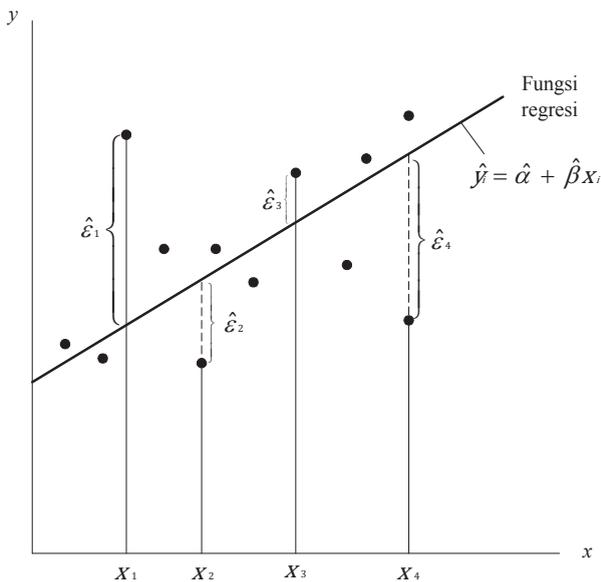
6.3. Metode *Ordinary Least Square* untuk Menaksir Model Regresi

Dalam Persamaan 6.4, variabel x dan y adalah sumber data yang akan digunakan untuk membuat modelnya. Sementara β_0 dan β_1 adalah parameter-parameter yang akan

ditaksir. Alasan mengapa menggunakan *Ordinary Least Square* (OLS) untuk menaksir parameter regresi adalah karena OLS itu *best linear unbiased estimation* (BLUE) untuk menaksir β_1 .

1. *Linear*. Fungsi linear dari variabel acak seperti variabel y dalam model regresi
2. *Unbiased*, dimana rata-rata atau nilai taksiran $E(\hat{\beta})$ adalah sama dengan nilai sebenarnya, β .
3. Penaksir yang efisien, karena memiliki varians yang minimum.

Dengan alasan ini, OLS dianggap sebagai metode terbaik dalam menaksir model regresi. *Least square* atau kuadrat terkecil yang dimaksud dalam OLS adalah mendapatkan nilai kuadrat terkecil dari residual (ϵ). Dimana residual atau *error* ini diperoleh dari selisih antara variabel respon (y) dengan variabel respon taksiran (\hat{y}). Gambarnya ditunjukkan pada Gambar 6.6.



Gambar 6.6: Kriteria *least square*

Pada Gambar 6.7, $\hat{\varepsilon}_1$, $\hat{\varepsilon}_2$, dan seterusnya diperoleh dari,

$$\hat{\varepsilon}_i = y_i - \hat{y}_i \quad (6.5)$$

dimana \hat{y}_i adalah

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i \quad (6.6)$$

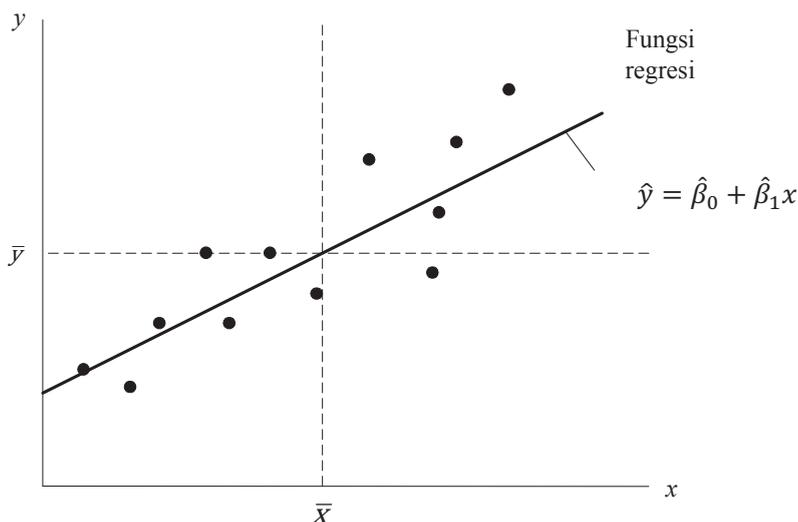
$\hat{\beta}_0$ dan $\hat{\beta}_1$ akan ditaksir dengan menggunakan OLS dengan pada Persamaan 6.4, yaitu:

$$\hat{\beta}_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \quad (6.7)$$

Sementara, untuk menaksir $\hat{\beta}_0$ adalah

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad (6.8)$$

Garis regresi yang telah diperoleh tersebut, garis regresi akan berada pada rata-rata sampel dari variabel x dan y . Gambar 6.7. menggambarkan lokasi \bar{x} dan \bar{y} pada garis regresi.



Gambar 6.7: Lokasi \bar{x} dan \bar{y} pada garis regresi

Contoh 6.3

Diperoleh hasil survey pendapatan dan pengeluaran ke 30 keluarga di suatu kampung. Hasil dari survey tersebut ditunjukkan pada Tabel 6.5

Tabel 6.5: Pendapatan dan pengeluaran kepala keluarga di kampung ABC

No.	Pendapatan (Juta Rupiah)	Pengeluaran (Juta Rupiah)	No.	Pendapatan (Juta Rupiah)	Pengeluaran (Juta Rupiah)	No.	Pendapatan (Juta Rupiah)	Pengeluaran (Juta Rupiah)
1.	9,91	7,37	11.	9,99	6,04	21.	10,94	8,74
2.	10,30	8,67	12.	9,14	7,83	22.	9,79	7,40
3.	12,83	9,84	13.	7,61	5,63	23.	7,05	5,52
4.	10,55	6,12	14.	8,78	6,91	24.	8,31	5,58
5.	9,39	6,66	15.	9,12	6,75	25.	9,62	7,14
6.	10,37	8,98	16.	8,57	6,87	26.	8,72	6,03
7.	9,38	7,34	17.	8,73	7,18	27.	10,31	8,38
8.	8,43	7,41	18.	9,46	7,44	28.	8,97	6,07
9.	9,82	7,32	19.	8,07	6,01	29.	8,62	6,27
10.	9,77	7,96	20.	10,45	8,50	30.	7,64	5,30

Dari data tersebut akan dicari apakah pendapatan keluarga mempengaruhi pengeluarannya.

Penyelesaian Contoh 6.3:

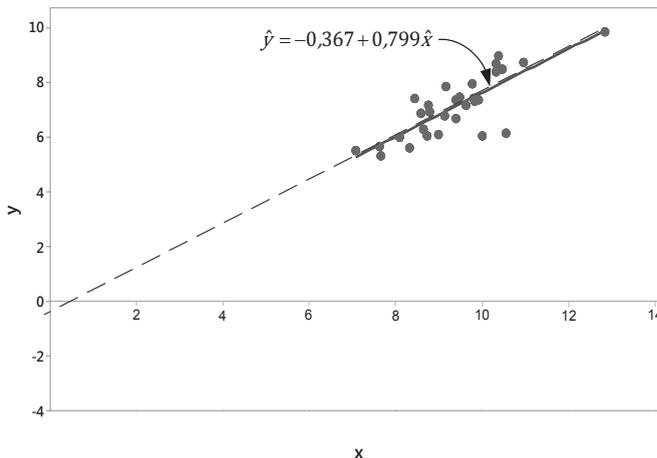
Di sini akan menggunakan analisis regresi. Dalam hal ini pendapatan keluarga adalah variabel independen (x) dan pengeluaran keluarga adalah variabel dependen (y). Model regresinya adalah $y = \beta_0 + \beta_1 x$. Dengan menggunakan data x dan y , maka taksiran β_0 dan β_1 dengan menggunakan Persamaan 6.7 dan 6.8 adalah:

$$\hat{\beta}_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \text{ dan } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

maka dicari dulu rata-rata pendapatan keluarga (\bar{x}) dan rata-rata pengeluarannya (\bar{y}) adalah:

$$\begin{aligned} \hat{\beta}_1 &= \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} \\ &= \frac{n(x_1 y_1 + x_2 y_2 + \dots + x_{30} y_{30}) - (x_1 + x_2 + \dots + x_{30})(y_1 + y_2 + \dots + y_{30})}{n(x_1^2 + x_2^2 + \dots + x_{30}^2) - (x_1 + x_2 + \dots + x_{30})^2} \\ &= \frac{30((9,91)(7,37) + \dots + (7,64)(5,30)) - (9,91 + \dots + 7,64)(7,37 + \dots + 5,30)}{30(9,91^2 + \dots + 7,64^2) - (7,37 + \dots + 5,30)^2} \\ &= \frac{30(2026,02) - (280,64)(213,26)}{30(2664,14) - (280,64)^2} \\ &= \frac{60780,73 - 59849,29}{79924,34 - 78758,81} \\ &= 0,799 \end{aligned}$$

$$\begin{aligned} \hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\ &= 7,11 - (0,799)(9,35) \\ &= -0,367 \end{aligned}$$



Gambar 6.8: Model regresi dinyatakan dalam grafik

Jadi, persamaan regresi dari data tersebut adalah $y = -0,367 + 0,799x$. Dalam persamaan ini, jika x bernilai nol atau pendapatannya adalah sekitar $-0,367$ atau sekitar minus Rp 367.000,-. Jika pendapatan x bertambah dalam satu satuan (dalam hal ini satu satuannya adalah Rp 1.000.000), maka pengeluaran akan bertambah 0,799 atau sekitar Rp 799.000,-. Grafik *scatterplot* beserta garis regresi dari model ini ditunjukkan pada Gambar 6.8. Dalam grafik tersebut, nampak garis regresi jika diperpanjang hingga $x = 0$ maka posisinya akan berada pada nilai minus 0,367.

6.4. Koefisien Determinasi (R^2)

Setelah melakukan estimasi parameter model, selanjutnya adalah menguji apakah garis regresi yang telah dibuat sudah sesuai dengan data. Ini yang kemudian dinamakan dengan *Goodness of Fit*. Suatu model atau garis regresi dikatakan bagus apabila hasil prediksinya memiliki residual yang sekecil mungkin atau data-datanya berada di sekitar garis regresi. Untuk mengukur ketepatan model atau mengukur *goodness of fit* digunakan koefisien determinasi R^2 . R^2 dihitung dengan menggunakan rumus pada Persamaan 6.9.

$$R^2 = \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} \quad (6.9)$$

R^2 akan bernilai antara 0 sampai 1 atau dinyatakan dengan persentasi yaitu dari 0 hingga 100%. Semakin besar R^2 berarti semakin besar variabel independen mampu menjelaskan hubungannya dengan variabel dependen. Jika tidak, berarti ada faktor lain selain variabel independen yang tidak dimasukkan dalam model yang memberi kontribusi terhadap variabel dependen.

Contoh 6.4.

Dari data pada Contoh 6.4, dengan menggunakan rumus 7.11, maka perlu menghitung \hat{y} . Dengan menggunakan model $y = -0,367 + 0,799x$, maka \hat{y} adalah:

No.	Y	\hat{y}	No.	y	\hat{y}	No.	y	\hat{y}
1.	7,37	7,55	11.	6,04	7,44	21.	8,74	7,98
2.	8,67	7,86	12.	7,83	7,62	22.	7,40	8,38
3.	9,84	9,89	13.	5,63	6,94	23.	5,52	7,46
4.	6,12	8,06	14.	6,91	5,71	24.	5,58	5,27
5.	6,66	7,14	15.	6,75	6,65	25.	7,14	6,27
6.	8,98	7,92	16.	6,87	6,92	26.	6,03	7,32
7.	7,34	7,13	17.	7,18	6,48	27.	8,38	6,60
8.	7,41	6,37	18.	7,44	6,61	28.	6,07	7,87
9.	7,32	7,48	19.	6,01	7,19	29.	6,27	6,80
10.	7,96	7,55	20.	8,50	6,08	30.	5,30	6,52

Berapa koefisien determinasinya?

Penyelesaian Contoh 6.4:

Koefisien determinasinya (R^2) adalah:

$$\begin{aligned} R^2 &= \frac{\sum(\hat{y}_i - \bar{y})^2}{\sum(y_i - \bar{y})^2} \\ &= \frac{(\hat{y}_1 - \bar{y})^2 + (\hat{y}_2 - \bar{y})^2 + \dots + (\hat{y}_{30} - \bar{y})^2}{(y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \dots + (y_{30} - \bar{y})^2} \\ &= \frac{((7,55 - 7,11)^2 + (7,86 - 7,11)^2 + \dots + (6,52 - 7,11)^2)}{((7,37 - 7,11)^2 + (8,67 - 7,11)^2 + \dots + (5,30 - 7,11)^2)} \\ &= \left(\frac{24,81}{38,33} \right) \\ &= 0,6473 \end{aligned}$$

Koefisien determinasi dari model regresi $y = -0,367 + 0,799x$ adalah 64,73%. Ini artinya, hubungan bahwa pendapatan

keluarga mempengaruhi pengeluarannya dapat dijelaskan sebesar 64,73% saja, dari 100%. Sisanya, yaitu 35,27% yang menentukan pengeluaran dijelaskan oleh variabel lainnya.

Selain R^2 , ada **R^2 adjusted** yang biasanya dituliskan dengan R^2 adj. Istilah *adjusted* berkaitan dengan derajat bebas (*degree of freedom*) dalam model yang melibatkan k parameter. Rumus dari R^2 adj adalah sebagai berikut:

$$R^2 adj = 1 - (1 - R^2) \frac{n - 1}{n - k} \quad (6.10)$$

Dimana R^2 = koefisien determinasi

n = banyaknya data

k = banyaknya parameter

Dalam R^2 adj melibatkan jumlah parameter (k) dalam model. Koefisien dapat digunakan untuk membandingkan koefisien determinasi antar model yang memiliki jumlah parameter berbeda-beda. Semakin besar R^2 adj semakin bagus modelnya. R^2 adj ini dianggap lebih baik dibandingkan dengan R^2 karena R^2 adj mampu menangkap pertambahan variabel dalam model. Umumnya, jika variabelnya bertambah R^2 akan semakin besar. Namun, dengan R^2 adj, akan lebih sensitif menangkap pertambahan variabel.

Sebagai contoh untuk menghitung R^2 adj, akan digunakan data pada Contoh 6.4 yang sudah didapat R^2 nya. Maka,

$$\begin{aligned} R^2 adj &= 1 - (1 - R^2) \frac{n - 1}{n - k} \\ &= 1 - (1 - 0,6473) \frac{30 - 1}{30 - 2} \\ &= 1 - 0,3527 \frac{29}{28} \\ &= 0,6347 \end{aligned}$$

Koefisien R^2 adj di sini bernilai 63,47%

6.5. Penarikan Kesimpulan dalam Model

Model regresi didapat dari hasil prediksi berdasarkan data. Model statistika tidak selalu dinyatakan sebagai model prediksi yang benar-benar tepat dapat memprediksi sesuai kenyataan dalam populasi. Kesalahan dalam prediksi dimungkinkan terjadi karena data diambil dari sampel. Seberapa handal model tersebut, perlu dilakukan pengujian. Dalam sub-bab ini akan dibahas taksiran titik parameter model, taksiran interval parameter, selang kepercayaan atau *confidence interval* dan uji hipotesis model regresi.

6.5.1. Taksiran interval dan uji hipotesis parameter dalam model

Dalam penaksiran statistik, angka yang didapat belum tentu tepat sesuai dengan populasinya. Ini karena data yang diambil adalah data sampel yang sanga memungkinkan terjadi ketidaktepatan dalam penaksiran. Estimasi parameter-parameter model regresi pada Persamaan 6.7 dan 6.8 adalah taksiran titik parameter-parameter dalam model regresi. Di sini akan dibahas taksiran interval dari parameter-parameter model regresi. Taksiran interval parameter untuk β_0 dan β_1 adalah sebagai berikut,

$$\hat{\sigma}_{\beta_1}^2 = \frac{\sigma^2}{\sum(x_i - \bar{x})^2} \quad (6.11)$$

$$\hat{\sigma}_{\beta_0}^2 = \frac{\sum x_i^2}{n \sum(x_i - \bar{x})^2} \sigma^2 \quad (6.12)$$

Dimana taksiran tidak bias dari σ^2 diperoleh dari,

$$\hat{\sigma}^2 = \frac{\sum(y_i - \hat{y}_i)^2}{n - k} \quad (6.13)$$

$\hat{\sigma}^2$ disebut juga *Mean Square Error (MSE)*.

Confidence interval (CI) atau selang kepercayaan untuk parameter β_1 adalah

$$\hat{\beta}_1 - t_{\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{\sum(x_i - \bar{x}_i)^2}} < \beta_1 < \hat{\beta}_1 + t_{\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{\sum(x_i - \bar{x}_i)^2}} \quad (6.14)$$

Uji hipotesis untuk menguji β_1 adalah:

$$H_0: \beta_1 = \beta_{10}$$

Untuk menguji ini, menggunakan distribusi t dengan derajat bebas (df) adalah $n - k$ untuk menentukan daerah kritis berikut:

$$t = \frac{\hat{\beta}_1 - \beta_{10}}{\hat{\sigma} / \sqrt{\sum(x_i - \bar{x}_i)^2}} \quad (6.15)$$

Sedangkan *confidence interval* untuk parameter β_0 adalah

$$\hat{\beta}_0 - t_{\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n \sum(x_i - \bar{x}_i)^2}} \sqrt{\sum x_i^2} < \beta_0 < \hat{\beta}_0 + t_{\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{n \sum(x_i - \bar{x}_i)^2}} \sqrt{\sum x_i^2} \quad (6.16)$$

Uji hipotesis untuk menguji β_0 adalah:

$$H_0: \beta_0 = \beta_{00}$$

Untuk menguji ini, menggunakan distribusi t dengan derajat bebas (df) adalah $n - k$ untuk menentukan daerah kritis berikut:

$$t = \frac{\hat{\beta}_0 - \beta_{00}}{\hat{\sigma} \sqrt{\sum x_i^2 / (n \sum(x_i - \bar{x}_i)^2)}} \quad (6.17)$$

Contoh 6.5

Dalam latihan ini akan dihitung *confidence interval* dan uji parameter untuk $\hat{\beta}_1$ dan $\hat{\beta}_0$ pada Contoh 6.4.

Penyelesaian Contoh 6.5:

Untuk menyelesaikannya perlu menghitung $\hat{\sigma}$ dengan Persamaan 6.13:

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{\sum(y_i - \hat{y}_i)^2}{n - k} \\
&= \frac{(y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \dots + (y_{30} - \hat{y}_{30})^2}{n - k} \\
&= \frac{(7,37 - 7,55)^2 + (8,67 - 7,86)^2 + \dots + (5,30 - 6,52)^2}{30 - 2} \\
&= 0,4829
\end{aligned}$$

Jadi, $\hat{\sigma}^2$ atau MSE adalah sebesar 0,4829. Selanjutnya akan dicari selang kepercayaan atau *confidence interval* untuk $\hat{\beta}_1$ dan $\hat{\beta}_0$ dengan menggunakan Persamaan 6.14 dan 6.16. Perhitungannya adalah sebagai berikut:

$$\hat{\beta}_1 - t_{\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{\sum(x_i - \bar{x}_i)^2}} < \beta_1 < \hat{\beta}_1 + t_{\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{\sum(x_i - \bar{x}_i)^2}}$$

Tabel distribusi t ditunjukkan pada Lampiran 1. Dalam Lampiran tersebut, dengan menggunakan $df = n - k = 30 - 2 = 28$, dengan $\alpha = 5\%$ didapatkan $t_{\alpha/2, df}$ adalah 2,048. Maka,

$$t_{\frac{\alpha}{2}} \frac{\hat{\sigma}}{\sqrt{\sum(x_i - \bar{x}_i)^2}} = 2,048 \frac{\sqrt{0,4829}}{\sqrt{38,85}} = 0,228$$

confidence interval untuk $\hat{\beta}_1$ adalah:

$$0,799 - 0,228 < \beta_1 < 0,799 + 0,228$$

$$0,571 < \beta_1 < 1,028$$

Dengan tingkat kepercayaan 95%, taksiran β_1 berkisar antara 0,571 hingga 1,028.

Uji hipotesis dari parameter β_1 adalah $H_0: \hat{\beta}_1 = 0$. Dengan menggunakan persamaan 6.15, didapat:

$$\begin{aligned}
 t &= \frac{\hat{\beta}_1 - \beta_{10}}{\hat{\sigma} / \sqrt{\sum (x_i - \bar{x}_i)^2}} \\
 &= \frac{0,799 - 0}{\sqrt{0,4829} / \sqrt{38,85}} \\
 &= 7,1681
 \end{aligned}$$

Nilai kritis ini kemudian dibandingkan dengan $t_{\frac{\alpha}{2}, df}$ dimana $df = 28$, dan $\alpha = 5\%$, maka diperoleh $t_{0,025;28} = 2,048$. Jika $t > t_{\text{tabel}}$ maka tolak H_0 yang artinya, pernyataan bahwa $\beta_1 = 0$ tidak dapat dibuktikan dengan data yang cukup. Sehingga, taksiran β_1 yang bernilai 0,799 adalah signifikan dalam model. Selanjutnya, uji parameter β_0 dengan menggunakan Persamaan 6.17. Dalam hal ini $H_0: \beta_0 = 0$

$$\begin{aligned}
 t &= \frac{\hat{\beta}_0 - \beta_{00}}{\hat{\sigma} \sqrt{\sum x_i^2 / (n \sum (x_i - \bar{x})^2)}} \\
 &= \frac{\hat{\beta}_0 - \beta_{00}}{\hat{\sigma} \sqrt{(x_1^2 + x_2^2 + \dots + x_{30}^2) / (n((x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_{30} - \bar{x})^2)}} \\
 &= \frac{-0,367}{\sqrt{0,4829} \sqrt{(9,91^2 + \dots + 7,64^2) / (30((9,91 - 7,11)^2 + \dots + (7,64 - 7,11)^2)}} \\
 &= -0,3495
 \end{aligned}$$

Nilai kritis (t) untuk $\hat{\beta}_0$ ini kemudian dibandingkan dengan $t_{\frac{\alpha}{2}, df}$ dimana $df = 28$, dan $\alpha = 5\%$, maka diperoleh $t_{0,025;28} = 2,048$. Jika $|t| > t_{\text{tabel}}$ maka tolak H_0 . Dari sini t lebih besar dari t_{tabel} yang artinya pernyataan $\beta_0 = 0$ adalah gagal diterima. Ini berarti parameter $\hat{\beta}_0$ adalah signifikan berada dalam model.

6.5.2. Analysis of Variance dari Model Regresi

Analysis of Variance dalam regresi digunakan untuk menguji kualitas model regresi yang telah dibuat. Berikut adalah tahapan-tahapan dalam membuat ANOVA

$$\hat{\sigma}^2 = \frac{SSE}{n - k - 1} \quad (6.18)$$

dimana,

$$SSE = \sum \varepsilon_i^2 = \sum (y_i - \hat{y}_i)^2 \quad (6.19)$$

$$SST = SSR + SSE \quad (6.20)$$

dimana *SST* adalah *total sum of square* dan *SSR* adalah *regression sum of square* yang persamannya adalah:

$$SST = \sum (y_i - \bar{y})^2 \quad (6.21)$$

$$SSR = \sum (\hat{y}_i - \bar{y})^2 \quad (6.22)$$

$$MSR = \frac{SSR}{k} \quad (6.23)$$

$$MSE = \frac{SSE}{n - (k + 1)} \quad (6.24)$$

MSR adalah *mean square regression* dan *MSE* adalah *mean square error*.

Tabel 6.6: Tabel ANOVA Regresi

Sumber	Degree of freedom	Sum of square (SS)	Mean Square (MS)	F
Regresi	k	SSR	$MSR = SSR/k$	$F = MSR/MSE$
Error	$n - (k + 1)$	SSE	$MSE = SSE/(n-(k+1))$	
Total	$n - 1$	SST		

Hipotesis yang akan digunakan untuk menguji model adalah:

$$H_0: \beta_0 = \beta_1 = \dots = \beta_k = 0$$

Contoh 6.6

Menggunakan data pada Contoh 6.4. Untuk mengisi ANOVA seperti Tabel 6.6, maka akan menggunakan Persamaan 6.19 sampai 6.21.

$$\begin{aligned} SSE &= \sum (y_i - \hat{y}_i)^2 \\ &= (y_1 - \hat{y}_1)^2 + (y_2 - \hat{y}_2)^2 + \dots + (y_{30} - \hat{y}_{30})^2 \\ &= (7,37 - 7,55)^2 + (8,67 - 7,86)^2 + \dots + (5,30 - 6,52)^2 \\ &= 13,52 \end{aligned}$$

$$\begin{aligned} SST &= \sum (y_i - \bar{y})^2 \\ &= (y_1 - \bar{y})^2 + (y_2 - \bar{y})^2 + \dots + (y_{30} - \bar{y})^2 \\ &= (7,37 - 7,11)^2 + (8,67 - 7,11)^2 + \dots + (5,30 - 7,11)^2 \\ &= 38,3333 \end{aligned}$$

$$SST = SSR + SSE$$

$$SSR = SST - SSE = 38,3333 - 13,5211 = 24,8122$$

$$MSR = SSR/k = 24,81/1 = 24,8122$$

$$MSE = SSE/(n - (k + 1)) = 13,5211/(30 - (1 + 1)) = 0,4829$$

$$F = MSR/MSE = 51,3819$$

Tabel 6.7: Hasil ANOVA latihan soal

Sumber	Sum of square	Degree of freedom	Mean Square	F
Regresi	24,8122	1	24,8122	51,3819
Error	13,5211	28	0,4829	
Total	38,3333	29		

6.6. Asumsi dalam Model Regresi

Dalam penerapan model regresi linear, terdapat asumsi klasik dalam model regresi (*classical linear regression model*). Umumnya, hanya empat yang dikenal yaitu model linear, homoskedastisitas, tidak ada kolinearitas, dan tidak ada autokorelasi residual. Namun di buku ini akan dijelaskan 10 asumsi klasik dalam model regresi. Kesepuluh asumsi tersebut dibahas dalam sub-bab ini.

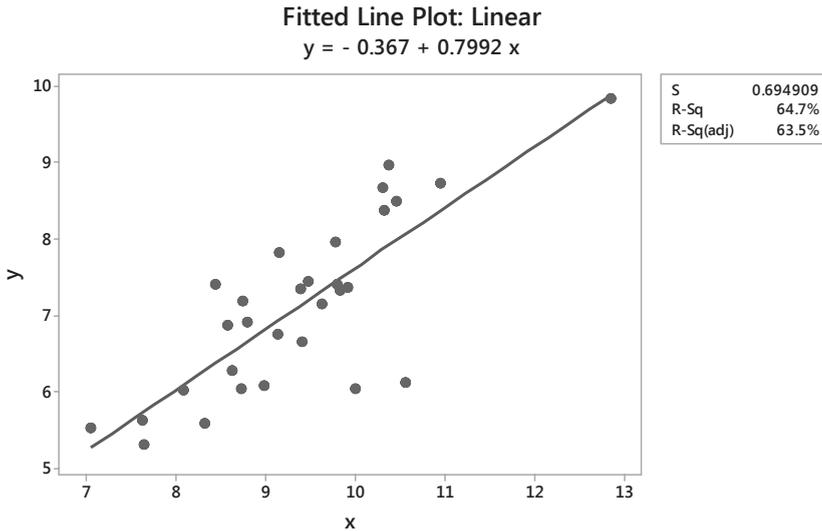
6.6.1. Asumsi 1: Model Regresi Linear

Model regresi pada Persamaan 6.4 adalah model regresi linear. Seperti yang dijelaskan sebelumnya di Bab 6.2.2 ada kemungkinan hubungan antar dua variabel adalah non-linear. Namun, seperti yang dijelaskan pada Tabel 6.2, selama parameternya masih linear maka model regresi adalah linear. Asumsi linearitas antar variabel harus dipenuhi karena model yang dibuat adalah model linear. Apabila asumsi linearitas tidak dapat dipenuhi, maka akan ada bias dalam penaksiran parameter-parameternya. Ini akan berimbas pada masalah kesesuaian model.

Selanjutnya, asumsi linearitas ini kemudian berhubungan dengan asumsi bahwa residual dari model regresi adalah berdistribusi normal. Ini berasal dari *property* dari distribusi normal yang mengatakan bahwa **fungsi linear dari variabel yang berdistribusi normal adalah berdistribusi normal** (Gujarati, 2003). Juga, penaksir OLS diturunkan dari distribusi normal.

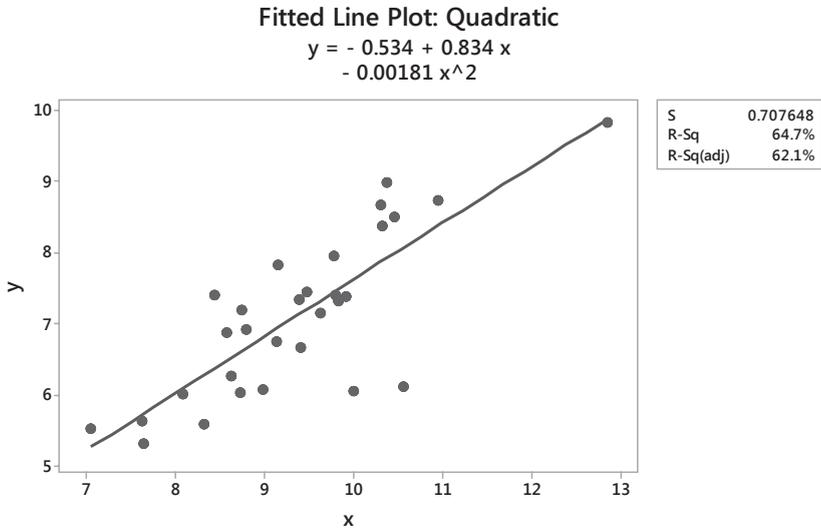
Linearitas hubungan antar variabel dapat dideteksi sebelum membuat model. Cara yang sederhana untuk mendeteksi bahwa antar variabel adalah linear atau tidak adalah dengan cara grafis menggunakan grafik *scatterplot*. Contoh pola hubungan linear ditunjukkan pada Gambar 6.1, dan

contoh pola hubungan non-linear dalam variabel ditunjukkan pada Gambar 6.5. Dalam Contoh 6.4, nampak jelas pola *scatterplot*-nya adalah positif linear seperti yang ditunjukkan pada Gambar 6.9, 6.10, dan 6.11.

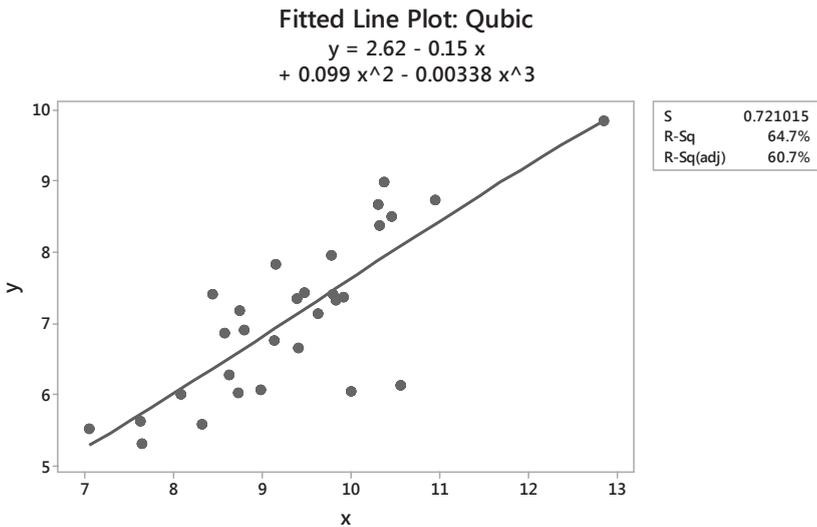


Gambar 6.9: Taksiran hubungan antar pendapatan dan pengeluaran keluarga dengan model linear

Model kuadratik pada Gambar 6.10 dan model kubik pada Gambar 6.11 adalah alternatif dari model non linear dalam variabel. Namun, jika dilihat dari pola data yang terbentuk, model linear pada Gambar 6.9 lebih sesuai dibandingkan model kudratik dan model kubik. Selain itu, untuk mencari model terbaik dapat dilihat dari σ terkecil, R^2 terbesar, dan R^2 adj terbesar. Besaran σ dapat dilihat dari s yang merupakan estimasi dari varians residual. Semakin kecil varians residual, maka semakin bagus modelnya. Dari ketiga model tersebut, model yang memiliki σ terkecil, R^2 terbesar, dan R^2 adj terbesar adalah model linear pada Gambar 6.10.



Gambar 6.10: Taksiran hubungan antar pendapatan dan pengeluaran keluarga dengan model kuadratik



Gambar 6.11: Taksiran hubungan antar pendapatan dan pengeluaran keluarga dengan model kubik

6.6.2. Asumsi 2: Nilai x adalah tetap dalam sampling yang berulang

Dalam analisis regresi yang akan diprediksi nilainya adalah variabel respons atau variabel dependen (y). Variabel ini akan diprediksi dari variabel independen (x). Kedua variabel tersebut diambil dari sampel secara acak. Pada nilai variabel independen (x) tertentu bisa saja menghasilkan variabel respons yang bermacam-macam. Sebagai contoh, untuk hubungan antara pendapatan per bulan (variabel independen) dan konsumsi per bulan (variabel dependen). Bisa saja, orang yang sama-sama berpendapatan Rp 2.000.000,- akan mengeluarkan konsumsi per bulan yang berbeda, sebut saja Rp 1.250.000,- dan Rp 1.500.000,-. Di sini berarti, dengan variabel independen yang sama, jika diulang akan menghasilkan variabel respons yang berbeda.

6.6.3. Asumsi 3: Rata-rata residual adalah nol

Residual (ε) semakin kecil, maka taksiran parameter akan semakin baik. Semakin kecil residual, berarti tingkat kesalahannya semakin minim. Ini berarti model yang dibuat sangat mendekati kenyataan. Dengan kata lain, model yang dibuat ketepatannya tinggi. Sebagai contoh, akan digunakan data pada Contoh 6.4. Rumus untuk menghitung residual menggunakan Persamaan 6.5. Hasilnya di tunjukkan pada Tabel 6.8.

Tabel 6.8: Residual dari model regresi pada Contoh 6.4

No.	ε	No.	ε	No.	ε
1.	-0.18	11.	0.52	21.	0.52
2.	0.81	12.	-1.58	22.	0.36
3.	-0.05	13.	0.89	23.	-0.06
4.	-1.94	14.	-0.08	24.	0.25
5.	-0.48	15.	0.26	25.	-0.69
6.	1.06	16.	-0.17	26.	-0.18
7.	0.21	17.	0.39	27.	-0.57
8.	1.04	18.	0.57	28.	0.51
9.	-0.16	19.	0.25	29.	-0.73
10.	-0.18	20.	-0.07	30.	-0.25

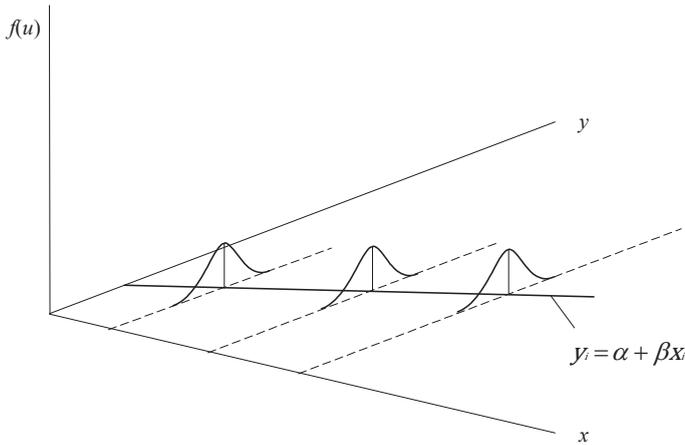
$$\begin{aligned}
 \bar{\varepsilon} &= \frac{\sum \varepsilon_i}{n} \\
 &= \frac{\varepsilon_1 + \varepsilon_2 + \dots + \varepsilon_{30}}{30} \\
 &= \frac{-0,18 + 0,81 + \dots \pm 0,25}{30} \\
 &= 0
 \end{aligned}$$

Rata-rata residual pada contoh ini adalah nol.

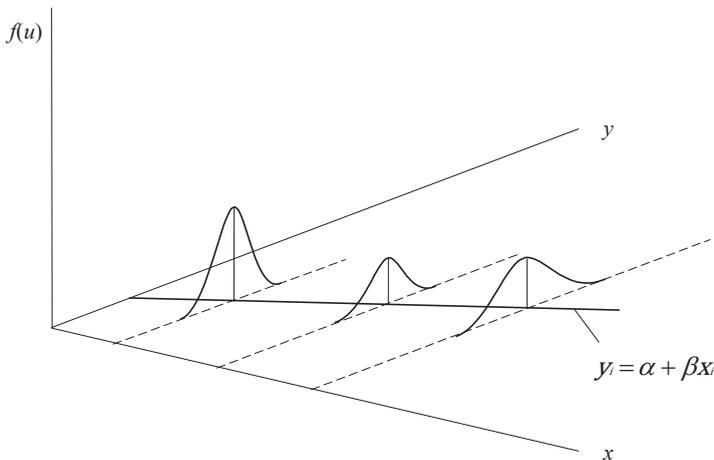
6.6.4. Asumsi 4: Homoskedastisitas atau varians dari residual adalah sama

Kata homoskedastis berasal dari kata *skedanime* yang berasal Bahasa Yunani yang artinya menyebar. Homo-skedastis berarti penyebarannya sama, sedangkan heteroskedastis adalah penyebarannya tidak sama. Jika residual terdapat homoskedastisitas berarti penyebaran residual adalah sama. Asumsi homoskedastisitas ini ada karena suatu kondisi dimana ada ketidaksamaan dalam penyebaran atau varians.

Maka homoskedastisitas residual adalah ketika varians residualnya adalah sama. Gambaran data yang homoskedastis dan heteroskedastis ditunjukkan pada Gambar 6.12 dan Gambar 6.13.



Gambar 6.12: Model regresi yang residualnya homoskedastis

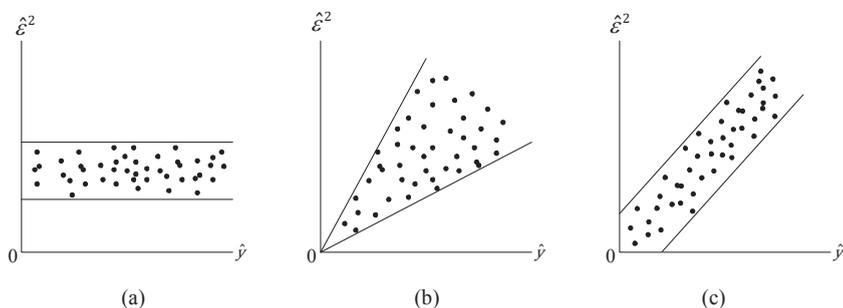


Gambar 6.13: Model regresi yang residualnya heteroskedastis

Uji homoskedastisitas dapat dilakukan dengan berbagai cara baik yang informal dan formal. Cara informal adalah dengan menggunakan grafis, sedangkan cara formal

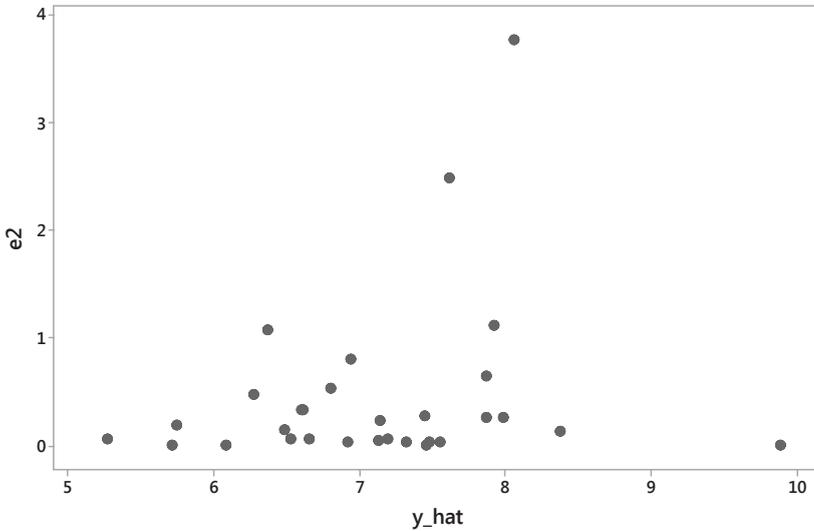
adalah menggunakan beberapa metode seperti uji *Park*, uji *Glejser*, uji *Spearman's Rank*, uji *Goldfeld-Quandt*, uji *Breusch-Pagan-Godfrey*, uji *White*, dan lain-lain. Di dalam buku ini akan dibahas uji informal. Sedangkan uji formal dapat dilihat pada Gujarati (2003). Cara grafis adalah dengan membuat scatterplot antara $\hat{\varepsilon}^2$ dengan \hat{y} atau $\hat{\varepsilon}^2$ dengan x . Kentuannya ditunjukkan pada Gambar 6.14.

Pada Gambar 6.14(a) variansnya homogen, sama halnya dengan Gambar 6.14(c). Bisa dikatakan bahwa pada dua gambar tersebut residualnya (ε) homoskedastis. Sedangkan pada Gambar 6.14(b) nampak jelas ada ketidakseragaman varians atau terdapat heteroskedastisitas.



Gambar 6.14: Scatterplot antara ε^2 dengan \hat{y}

Sebagai bahan untuk latihan, akan digunakan Contoh 6.4 untuk mendeteksi heteroskedastisitas pada residual model regresi. Hasil residual ditampilkan pada Tabel 6.8 yang kemudian dibuat *scatterplot*. Hasilnya ditunjukkan pada Gambar 6.15. Dari gambar tersebut terdapat beberapa titik yang melenceng jauh dari titik lainnya. Dari grafik ini nampak ada ketidakhomogenan varians residual. Namun, uji formal seperti menggunakan uji *Park*, uji *Glejser*, uji *Spearman's Rank*, uji *Goldfeld-Quandt*, uji *Breusch-Pagan-Godfrey*, uji *White*, dan lain-lain perlu dilakukan supaya didapatkan kesimpulannya lebih baik dan kuat.



Gambar 6.15: Scatterplot antara ε^2 dengan \hat{y} dari model regresi pada Contoh 6.4.

6.6.5. Asumsi 5: Tidak ada autokorelasi antar residual

Antar residual (ε) tidak boleh ada autokorelasi atau serial korelasi. Autokorelasi adalah adanya korelasi antar anggota deret data. Dalam analisis regresi, adanya asumsi tidak ada korelasi dalam residual karena diharapkan antar residual tidak saling mempengaruhi. Ini maksudnya adalah jika ε_1 adalah residual, maka ε_1 tidak boleh mempengaruhi ε_2 , ε_2 tidak boleh mempengaruhi ε_3 , dan seterusnya.

Deteksi adanya autokorelasi pada residual uji *Durbin Watson*. Persamaan dalam uji *Durbin Watson* adalah sebagai berikut:

$$d = \frac{\sum_{t=2}^n (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^n \hat{\varepsilon}_t^2} \quad (6.25)$$

dimana d = statistik *Durbin-Watson*

Uji d adalah, dengan level signifikan sebesar α ,

- $H_0: \rho = 0$ vs $H_1: \rho > 0$. Tolak H_0 pada level α jika $d < d_{U'}$ yaitu statistik signifikan ada autokorelasi positif.
- $H_0: \rho = 0$ vs $H_1: \rho > 0$. Tolak H_0 pada level α jika $(4 - d) < d_{U'}$ yaitu statistik signifikan ada autokorelasi negatif.
- $H_0: \rho = 0$ vs $H_1: \rho \neq 0$. Tolak H_0 pada level 2α jika $d < d_U$ atau $(4 - d) < d_{U'}$ yaitu statistik signifikan ada autokorelasi positif atau negatif.

Aturannya secara detil dapat dilihat pada Tabel 6.9.

Tabel 6.9: Aturan keputusan untuk uji *Durbin-Watson*

Hipotesis awal	Keputusan	Jika
Tidak ada autokorelasi positif	Tolak	$0 < d < d_L$
Tidak ada autokorelasi positif	Tidak ada keputusan	$d_L \leq d \leq d_U$
Tidak ada autokorelasi negatif	Tolak	$4 - d_L < d < 4$
Tidak ada autokorelasi negatif	Tidak ada keputusan	$4 - d_U \leq d \leq 4 - d_L$
Tidak ada autokorelasi positif atau negatif	Tidak tolak	$d_U < d < 4 - d_L$

Contoh penggunaan uji *Durbin-Watson* ini akan menggunakan data pada Contoh 6.4. Data residual terdapat di Tabel 6.8. Dengan menggunakan Persamaan 6.25, maka didapat d ,

$$\begin{aligned}
 d &= \frac{\sum_{t=2}^n (\hat{\varepsilon}_t - \hat{\varepsilon}_{t-1})^2}{\sum_{t=1}^n \hat{\varepsilon}_t^2} \\
 &= \frac{(\hat{\varepsilon}_2 - \hat{\varepsilon}_1)^2 + (\hat{\varepsilon}_3 - \hat{\varepsilon}_2)^2 + \dots + (\hat{\varepsilon}_{30} - \hat{\varepsilon}_{29})^2}{\hat{\varepsilon}_2^2 + \hat{\varepsilon}_3^2 + \dots + \hat{\varepsilon}_{30}^2} \\
 &= \frac{(0,81 - (-0,18))^2 + ((-0,05) - 0,81)^2 + \dots + ((-0,25) - (-0,73))^2}{(0,81)^2 + (-0,05)^2 + \dots + (-0,25)^2} \\
 &= \frac{30,3539}{13,4879} \\
 &= 2,2505
 \end{aligned}$$

Diperoleh d sebesar 2,2505. Kemudian dicari nilai d_U dan d_L di Lampiran 5 yang menggunakan $\alpha = 5\%$. Pada k (banyaknya parameter) adalah 2 dan $n = 30$, diperoleh $d_L = 1,284$ dan $d_U = 1,567$. Nilai $4 - d_L = 4 - 1,284 = 2,716$. Dalam hal ini d berada di antara d_U dan $4 - d_L$. Artinya, tidak ada autokorelasi positif maupun autokorelasi negatif antar residual.

6.6.6. Asumsi 6: Tidak ada kovarians antar residual dan x

Asumsi keenam adalah tidak ada korelasi antara variabel x dengan residual. Sebut saja jika x dan ε berkorelasi. Maka, jika x naik, maka ε ikut naik atau sebaliknya akan turun. Sementara, x diprediksi memiliki hubungan baik positif maupun negatif dengan y . Jika ε memiliki korelasi dengan x , siapa yang memiliki pengaruh ke y akan sulit ditentukan karena *redundant*.

6.6.7. Asumsi 7: Jumlah observasi harus lebih besar dari parameter yang akan ditaksir

Banyaknya observasi atau pengamatan atau sampel harus lebih banyak dari parameter yang akan ditaksir. Ini berlaku untuk analisis regresi berganda yang seringkali menggunakan banyak variabel. Asumsi ini diberlakukan karena berkaitan dengan menaksir parameter dalam model. Jika ada dua variabel, sebut saja x dan y , maka minimal ada dua data pengamatan pada masing-masing variabel tersebut agar bisa dibuat model regresi. Ini berkaitan dengan penentuan koefisien determinasi dan uji kualitas model yang melibatkan banyaknya parameter dalam model. Jika jumlah data (n) kurang atau sama dengan parameter (k) maka pembagi dalam menghitung menghitung taksiran varians model (Persamaan 6.18) tidak dapat diperoleh karena akan menghasilkan nilai nol. Ini berimbas pada penilaian kualitas model yang tidak bisa dilakukan karena akan menghasilkan

MSE (Persamaan 6.24) pada Tabel ANOVA model regresi (Tabel 6.6) tidak bisa dilakukan karena MSE menghasilkan angka tak hingga (∞).

6.6.8. Asumsi 8: Variabilitas dalam x .

Dalam model regresi, x harus bervariasi dan tidak bisa semuanya sama. Jika x tidak bervariasi, maka akan didapatkan $x = \bar{x}$ dengan varians x sama dengan nol karena tidak ada variasi data. Maka untuk mendapatkan parameter β_0 dan β_1 tidak dimungkinkan karena akan melibatkan \bar{x} dan $\text{var}(x)$ dalam menghitungnya. Sebagai penjelasan, Persamaan 6.7 dapat diubah menjadi,

$$\hat{\beta}_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} \quad (6.7)$$

6.6.9. Asumsi 9: Model regresi harus ditentukan secara tepat

Penentuan model regresi harus didasarkan pada teori yang sudah ada. Jika tidak ada teori pendukung, filosofi dan logika berfikir dari model yang dibuat harus kuat. Ini harus menjadi pegangan bagi para pembuat model.

Selain itu, pola hubungan antar variabel harus diidentifikasi secara benar. Misalnya, pola hubungan linear antar variabel seperti yang telah dibahas pada 6.6.1. Hubungan antar variabel bisa linear atau tidak ada hubungan seperti yang digambarkan pada Gambar 6.1 dan Gambar 6.2, atau hubungannya non linear seperti dalam Gambar 6.5 harus diidentifikasi dengan benar. Apabila tidak dilakukan, maka akan menghasilkan model yang bias.

6.6.10. Asumsi 10: Tidak ada multikolinearitas

Asumsi ini berlaku pada model regresi dengan banyak variabel independen atau regresi berganda. Sebagai contoh,

akan dideteksi hubungan antara x_1 dan x_2 terhadap y . Dari sini x_1 akan memiliki hubungan dengan y . Begitu juga dengan x_2 yang juga akan memiliki hubungan dengan y . Jika variabel x_1 dan x_2 memiliki hubungan yang kuat juga, maka di sini berarti ada multikolinearitas dalam model.

Multikolinearitas dalam model akan memberikan konsekuensi. Menurut Gujarati (2003) konsekuensinya adalah sebagai berikut:

1. Meskipun OLS akan menghasilkan taksiran yang BLUE, OLS memiliki varians dan kovarians yang besar yang akan menyulitkan dalam membuat taksiran parameter yang tepat. Jadi, meskipun analisis regresi akan menghasilkan nilai taksiran parameter, parameter tersebut bisa jadi tidak tepat.
2. Karena taksirannya tidak tepat, maka taksiran intervalnya akan lebih besar yang cenderung akan menerima hipotesis null (H_0). Artinya, parameter yang sudah ditaksir tersebut cenderung dianggap tidak bermakna.
3. Konsekuensi dari taksiran tidak tepat juga berdampak pada rasio t yang cenderung membuat keputusan bahwa parameternya tidak signifikan.
4. Jika ada perubahan data yang kecil, penaksir dengan OLS dan standar deviasinya akan lebih sensitif.

Untuk menghindari kesalahan-kesalahan seperti di atas, maka multikolinearitas variabel harus diatasi di dalam model agar tidak diambil kesimpulan yang salah.

Karena multikolinearitas sangat penting dalam membuat model regresi, maka perlu dilakukan deteksi. Cara mendeteksi multikolinearitas secara formal maupun informal bisa dilakukan dengan beberapa cara. Pertama, dengan melihat hasil analisis regresi dengan melihat R^2 dan t . Jika didapatkan R^2 yang tinggi namun terdapat nilai t yang signifikansinya kecil atau bahkan tidak signifikan. Kedua,

dengan membuat korelasi berpasangan antar variabel independen. Ketiga, dengan menggunakan *Vector Inflation Factor* (VIF) yang persamaannya adalah sebagai berikut:

$$VIF = \frac{1}{(1 - \rho_{x_1, x_2}^2)} \quad (6.xx)$$

6.7. Asumsi Residual Berdistribusi Normal

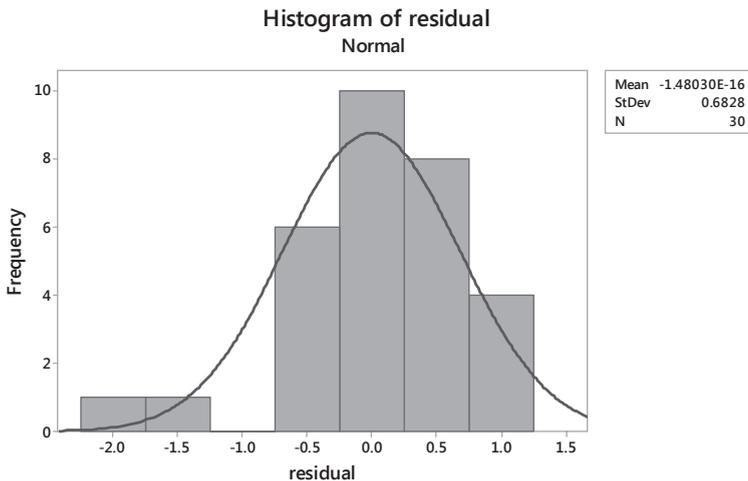
Parameter model regresi ditaksir dari *Ordinary Least Square* (OLS). Model regresi sendiri adalah linear. Karena modelnya adalah linear, jika variabel-variabel pembentuk dari model ini adalah berdistribusi normal, maka residual dari model regresi akan berdistribusi normal juga. Lebih jelasnya, taksiran parameter β_0 dan β_1 dengan OLS adalah fungsi linear dari ε . Maka, jika ε berdistribusi normal, maka β_0 dan β_1 akan berdistribusi normal juga.

Mengapa harus berdistribusi normal? Pertama, distribusi normal adalah distribusi yang sangat sederhana. Sebagian besar fenomena yang ada mengikuti distribusi normal. Kedua, diasumsikan, dalam membuat model regresi akan digunakan data yang cukup banyak atau dalam skala besar. Sementara itu, Limit Central Theorema mengatakan bahwa jika terdapat variabel acak yang jumlah datanya sangat banyak, maka distribusi variabel tersebut akan cenderung membentuk distribusi normal. Inilah yang mendasari mengapa residual model regresi harus berdistribusi normal.

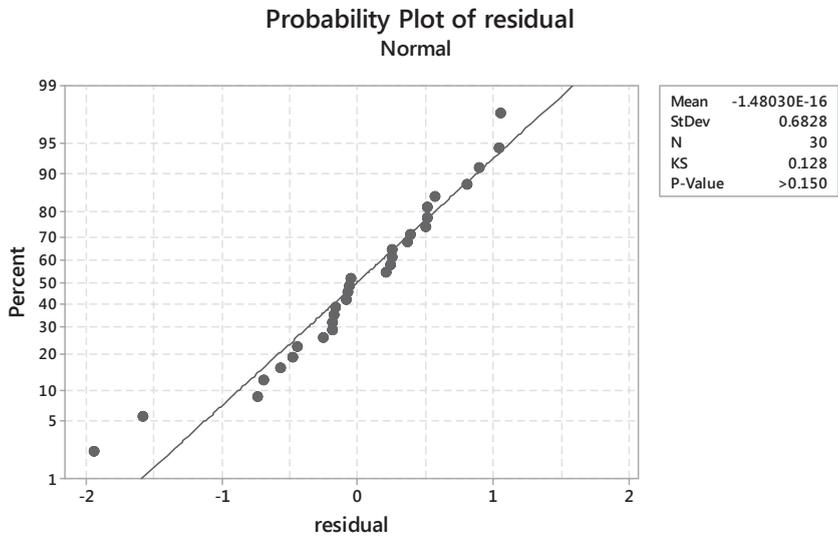
Untuk menguji apakah residual (ε) berdistribusi normal, dapat dilakukan dengan cara informal dan cara formal. Cara informal dapat dilakukan dengan menggunakan grafis distribusi variabelnya. Dalam hal ini histogram dapat digunakan untuk mendeteksi suatu variabel berdistribusi

normal, seperti yang telah dipaparkan dalam 4.1.2. Sedangkan cara formal mendeteksi suatu variabel berdistribusi normal adalah dengan menggunakan beberapa uji seperti uji *Kolmogorov-Smirnov*, tes *Anderson-Darling*, dan uji *Jarque-Bera*. Uji-uji ini dapat dilakukan dengan menggunakan bantuan software statistika seperti SPSS, Minitab, Eviews, dan SAS.

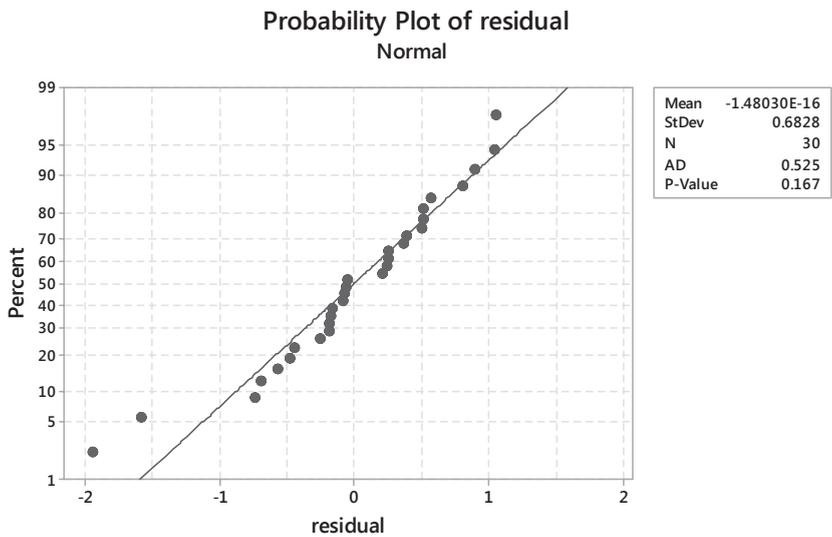
Dalam buku ini akan diberikan contoh deteksi distribusi normal residual secara grafis dan uji formal dengan bantuan software statistik. Contoh 6.4 digunakan sebagai latihan dengan residual yang ditampilkan pada Tabel 6.6. Histogram dari residual ditunjukkan pada Gambar 6.16. Hasilnya, histogram hampir mengikuti garis normal. Namun, ada asimetris karena terdapat ekor yang lebih panjang di bagian kiri. Dengan gambar ini masih belum dipastikan apakah residual berdistribusi normal. Oleh karenanya perlu uji formal. Dengan bantuan *software* Minitab, residual akan diuji dengan metode *Kolmogorov-Smirnov* dan *Anderson-Darling*. Hasilnya ditunjukkan pada Gambar 6.17 dan 6.18.



Gambar 6.16: Histogram residual



Gambar 6.17: Uji distribusi normal residual dengan *Kolmogorov-Smirnov test*



Gambar 6.18: Uji distribusi normal residual dengan *Anderson-Darling test*

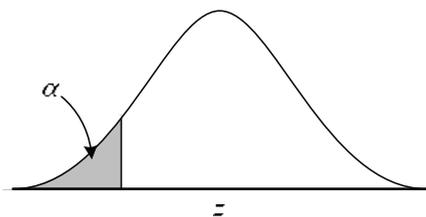
Hasil uji dapat dilihat dari nilai statistik *Kolmogorov-Smirnov* dan *Anderson-Darling*. Lebih mudahnya, dapat dilihat dari *p-value* pada kedua uji tersebut. Hipotesis awalnya (H_0), residual berdistribusi normal dengan H_1 residual tidak berdistribusi normal. Apabila nilai *p-value* lebih kecil dari nilai α maka keputusannya tolak H_0 . Pada Gambar 6.17 dan Gambar 6.18 diperoleh nilai *p-value* sebesar 0,150 dengan uji *Kolmogorov-Smirnov* dan 0,167 dengan uji *Anderson-Darling*. Dengan menggunakan α sebesar 5%, maka bisa dikatakan bahwa H_0 gagal ditolak karena *p-value* untuk kedua uji tersebut lebih besar dari α . Dari sini bisa dikatakan bahwa residual sudah mengikuti distribusi normal.

DAFTAR PUSTAKA

- BPS Kota Surakarta. (2017). *Kota Surakarta Dalam Angka 2017*. Surakarta: Badan Pusat Statistik Kota Surakarta.
- BPS Kota Surakarta. (2018). *Kota Surakarta dalam Angka 2018*. Surakarta: Badan Pusat Statistik Kota Surakarta.
- Bryman, A. (2012). *Social Research Methods* (Fourth Edi). Oxford: Oxford University Press.
- Clements, Ri. B. (1991). *Handbook of Statistical Methods in Manufacturing*. New Jersey: Prentice-Hall, Inc.
- Cochran, W. G. (1991). *Teknik Penarikan Sampel (terjemahan)*. (Rudiansyah & E. R.Osman, Eds.) (Edisi keti). Jakarta: UI Press.
- Dekking, F. M., Kraaikamp, D., Lopuhaa, H. P., & Meester, L. E. (2005). *A Modern Introduction to Probability and Statistics. Understanding Why and How*. London: Springer-Verlag.
- Desrosières, A. (1998). *The Politics of Large Numbers: A History of Statistical Reasoning*. Massachusetts: Harvard University Press.
- Gujarati, D. N. (2003). *Basic Econometrics* (Fourth Edi). New York: McGraw-Hill.
- Herkenhoff, L., & Fogli, J. (2013). *Applied Statistics for Business and Management using Microsoft Excel*. New York: Springer Science+Business Media.
- Leekley, R. M. (2010). *Applied Statistics for Business and Economics*. Boca Raton: CRS Press.

- Patil, G. P., & Rao, C. R. (1994). *Handbook of Statistics 12: Environmental Statistics*. Amsterdam: Elsevier Science B.V.
- Salkind, N. J. (2008). *Statistics for People Who Think They Hate Statistics* (3rd Editio). London: SAGE Publications.
- Walpole, R. E., Myers, R. H., Myers, S. L., & Ye, K. (2013). *Essentials of Probability and Statistics for Engineers and Scientist*. Boston: Pearson Education Ltd.
- Wegner, T. (2013). *Applied Business Statistics: Methods and Excel-based Applications* (3rd Editio). Claremont: Juta and Company.
- Wilkinson, L. (1999). Dot Plots. *The American Statistician*, 53(3), 276–281.

Lampiran 1: Tabel Distribusi z

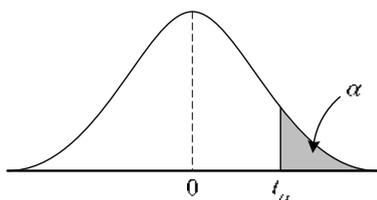


z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
-3,4	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0003	0,0002
-3,3	0,0005	0,0005	0,0005	0,0004	0,0004	0,0004	0,0004	0,0004	0,0004	0,0003
-3,2	0,0007	0,0007	0,0006	0,0006	0,0006	0,0006	0,0006	0,0005	0,0005	0,0005
-3,1	0,0010	0,0009	0,0009	0,0009	0,0008	0,0008	0,0008	0,0008	0,0007	0,0007
-3,0	0,0013	0,0013	0,0013	0,0012	0,0012	0,0011	0,0011	0,0011	0,0010	0,0010
-2,9	0,0019	0,0018	0,0018	0,0017	0,0016	0,0016	0,0015	0,0015	0,0014	0,0014
-2,8	0,0026	0,0025	0,0024	0,0023	0,0023	0,0022	0,0021	0,0021	0,0020	0,0020
-2,7	0,0035	0,0034	0,0033	0,0032	0,0031	0,0030	0,0029	0,0028	0,0027	0,0026
-2,6	0,0047	0,0045	0,0044	0,0043	0,0041	0,0040	0,0039	0,0038	0,0037	0,0036
-2,5	0,0062	0,0060	0,0059	0,0057	0,0055	0,0054	0,0052	0,0051	0,0049	0,0048
-2,4	0,0082	0,0080	0,0078	0,0075	0,0073	0,0071	0,0069	0,0068	0,0066	0,0064
-2,3	0,0107	0,0104	0,0102	0,0099	0,0096	0,0094	0,0091	0,0089	0,0087	0,0084
-2,2	0,0139	0,0136	0,0132	0,0129	0,0125	0,0122	0,0119	0,0116	0,0113	0,0110
-2,1	0,0179	0,0174	0,0170	0,166	0,0162	0,0158	0,0154	0,0150	0,0146	0,0143
-2,0	0,0287	0,0222	0,0217	0,212	0,0207	0,0202	0,0197	0,0192	0,0188	0,0183
-1,9	0,0287	0,0281	0,0274	0,0268	0,0262	0,0256	0,0250	0,0244	0,0239	0,0233
-1,8	0,0359	0,0351	0,0344	0,0336	0,0329	0,0322	0,0314	0,0307	0,0301	0,0294
-1,7	0,0446	0,0436	0,0427	0,0418	0,0409	0,0401	0,0392	0,0384	0,0375	0,0367
-1,6	0,0548	0,0537	0,0526	0,0516	0,0505	0,0495	0,0485	0,0475	0,0465	0,0455
-1,5	0,0668	0,0655	0,0643	0,0630	0,0618	0,0606	0,0594	0,0582	0,0571	0,0559
-1,4	0,0808	0,0793	0,0778	0,0764	0,0749	0,0735	0,0721	0,0708	0,0694	0,0681
-1,3	0,0968	0,0951	0,0934	0,0918	0,0901	0,0885	0,0869	0,0853	0,0838	0,0823
-1,2	0,1151	0,1131	0,1112	0,01093	0,1075	0,1056	0,1038	0,1020	0,1003	0,0985
-1,1	0,1357	0,1335	0,1314	0,1292	0,1271	0,1251	0,1230	0,1210	0,1190	0,1173
-1,0	0,1587	0,1562	0,1539	0,1515	0,1492	0,1469	0,1446	0,1423	0,1401	0,1379
-0,9	0,1841	0,1814	0,1788	0,1762	0,1736	0,1711	0,1685	0,1660	0,1635	0,1611
-0,8	0,2119	0,2090	0,2061	0,2033	0,2005	0,1977	0,1949	0,1922	0,18904	0,1867
-0,7	0,2420	0,2389	0,2358	0,2327	0,2296	0,2266	0,2236	0,2206	0,2177	0,2148
-0,6	0,2743	0,2709	0,2676	0,2643	0,2611	0,2578	0,2546	0,2514	0,2483	0,2451
-0,5	0,3085	0,3050	0,3015	0,2981	0,2946	0,2912	0,2877	0,2843	0,2810	0,2776
-0,4	0,3446	0,3409	0,3372	0,3336	0,3300	0,3264	0,3228	0,3192	0,3156	0,3121
-0,3	0,3821	0,37983	0,3745	0,3707	0,3669	0,3632	0,3594	0,3557	0,3520	0,3483
-0,2	0,4207	0,4168	0,4129	0,4090	0,4052	0,4013	0,3974	0,3936	0,3897	0,3859
-0,1	0,4602	0,4562	0,4522	0,4483	0,4443	0,4404	0,4364	0,4325	0,4286	0,4247
-0,0	0,5000	0,4960	0,4920	0,4880	0,4840	0,4801	0,4761	0,4721	0,4681	0,464

Lanjutan Lampiran 1 ...

z	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,0	0,5000	0,5040	0,5080	0,5120	0,5160	0,5199	0,5239	0,5279	0,5319	0,5359
0,1	0,5398	0,5468	0,5478	0,5517	0,5557	0,5596	0,5636	0,5672	0,5714	0,5753
0,2	0,5793	0,5832	0,5871	0,5910	0,5948	0,5987	0,6026	0,6064	0,6103	0,6141
0,3	0,6179	0,6217	0,6255	0,6293	0,6331	0,6368	0,6406	0,6443	0,6480	0,6517
0,4	0,6554	0,6591	0,6628	0,6664	0,6700	0,6736	0,6772	0,6808	0,6844	0,6879
0,5	0,6915	0,6950	0,6985	0,7019	0,7054	0,7088	0,7123	0,7157	0,7190	0,7224
0,6	0,7257	0,7291	0,7324	0,7357	0,7389	0,7422	0,7454	0,7486	0,7517	0,7549
0,7	0,7560	0,7611	0,7642	0,7673	0,7704	0,7734	0,7764	0,7794	0,7823	0,7852
0,8	0,7881	0,7910	0,7939	0,7967	0,7995	0,9023	0,8051	0,8078	0,8106	0,8133
0,9	0,8159	0,8186	0,8218	0,8238	0,8264	0,8289	0,8315	0,8340	0,8365	0,8389
1,0	0,8413	0,8438	0,8461	0,8485	0,8508	0,8531	0,8554	0,8577	0,8599	0,8621
1,1	0,8643	0,8665	0,8686	0,8708	0,8729	0,8749	0,770	0,8790	0,8810	0,8830
1,2	0,8819	0,8869	0,8888	0,8907	0,8925	0,8944	0,8962	0,8980	0,8997	0,9015
1,3	0,9032	0,9049	0,9066	0,9082	0,9099	0,9115	0,9131	0,9147	0,9162	0,9177
1,4	0,9192	0,9207	0,9222	0,9236	0,9251	0,9264	0,9279	0,9292	0,9306	0,9319
1,5	0,9332	0,9345	0,9357	0,9370	0,9382	0,9394	0,9406	0,9418	0,9429	0,9441
1,6	0,9452	0,9463	0,9474	0,9484	0,9495	0,9505	0,9515	0,9525	0,9535	0,9545
1,7	0,9554	0,9564	0,9573	0,9582	0,9591	0,9599	0,9608	0,9616	0,9625	0,9633
1,8	0,9641	0,9646	0,9656	0,9664	0,9671	0,9678	0,9686	0,9693	0,9699	0,9706
1,9	0,9713	0,9719	0,9726	0,9732	0,9738	0,9744	0,9750	0,9756	0,9761	0,9767
2,0	0,9772	0,9778	0,9783	0,9788	0,9793	0,9798	0,9803	0,9808	0,9812	0,9817
2,1	0,9821	0,9826	0,9830	0,9834	0,9838	0,9842	0,9846	0,9850	0,9854	0,9857
2,2	0,9861	0,9864	0,9868	0,9871	0,9875	0,9878	0,9881	0,9884	0,9887	0,9890
2,3	0,9893	0,9896	0,9898	0,9901	0,9904	0,9906	0,9909	0,9911	0,9913	0,9916
2,4	0,9918	0,9920	0,9922	0,9925	0,9927	0,9929	0,9931	0,9932	0,9934	0,9936
2,5	0,9938	0,9940	0,9941	0,9943	0,9945	0,9946	0,9948	0,9949	0,9951	0,9952
2,6	0,9953	0,9955	0,9956	0,9957	0,9959	0,9960	0,9961	0,9962	0,9963	0,9964
2,7	0,9965	0,9966	0,9967	0,9968	0,9969	0,9970	0,9971	0,9972	0,9973	0,9974
2,8	0,9974	0,9975	0,9976	0,9977	0,9977	0,9978	0,9979	0,9979	0,9980	0,9981
2,9	0,9981	0,9982	0,9982	0,9983	0,9984	0,9984	0,9985	0,9985	0,9986	0,9986
3,0	0,9987	0,9987	0,9987	0,9988	0,9988	0,9989	0,9989	0,9989	0,9990	0,9990
3,1	0,9990	0,9991	0,9991	0,9991	0,9992	0,9992	0,9992	0,9992	0,9993	0,9993
3,2	0,9993	0,9993	0,9994	0,9994	0,9994	0,9994	0,9994	0,9995	0,9995	0,9995
3,3	0,9995	0,9995	0,9995	0,9996	0,9996	0,9996	0,9996	0,9996	0,9996	0,9997
3,4	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9997	0,9998

Lampiran 2: Tabel Distribusi t

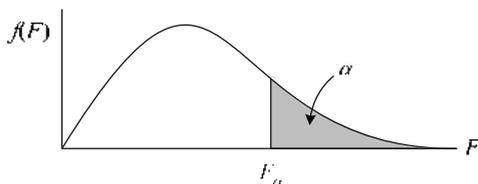


df	α					
	0,25	0,1	0,05	0,025	0,01	0,005
1	1,000	3,078	6,314	12,706	31,821	63,657
2	0,816	1,886	2,920	4,303	6,965	9,925
3	0,765	1,638	2,353	3,182	4,541	5,841
4	0,741	1,533	2,132	2,776	3,747	4,604
5	0,727	1,476	2,015	2,571	3,365	4,032
6	0,718	1,440	1,943	2,447	3,143	3,707
7	0,711	1,415	1,895	2,365	2,998	3,499
8	0,706	1,397	1,860	2,306	2,896	3,355
9	0,703	1,383	1,833	2,262	2,821	3,250
10	0,700	1,372	1,812	2,228	2,764	3,169
11	0,697	1,363	1,796	2,201	2,718	3,106
12	0,695	1,356	1,782	2,179	2,681	3,055
13	0,694	1,350	1,771	2,160	2,650	3,012
14	0,692	1,345	1,761	2,145	2,624	2,977
15	0,691	1,341	1,753	2,131	2,602	2,947
16	0,690	1,337	1,746	2,120	2,583	2,921
17	0,689	1,333	1,740	2,110	2,567	2,898
18	0,688	1,330	1,734	2,101	2,552	2,878
19	0,688	1,328	1,729	2,093	2,539	2,861
20	0,687	1,325	1,725	2,086	2,528	2,845
21	0,686	1,323	1,721	2,080	2,518	2,831
22	0,686	1,321	1,717	2,074	2,508	2,819
23	0,685	1,319	1,714	2,069	2,500	2,807
24	0,685	1,318	1,711	2,064	2,492	2,797
25	0,684	1,316	1,708	2,060	2,485	2,787
26	0,684	1,315	1,706	2,056	2,479	2,779
27	0,684	1,314	1,703	2,052	2,473	2,771
28	0,683	1,313	1,701	2,048	2,467	2,763
29	0,683	1,311	1,669	2,045	2,462	2,756
30	0,683	1,310	1,697	2,042	2,457	2,750
40	0,681	1,303	1,684	2,021	2,423	2,704
60	0,679	1,296	1,671	2,000	2,390	2,660
120	0,677	1,289	1,658	1,980	2,358	2,617
∞	0,674	1,282	1,645	1,960	2,326	2,576

Lanjutan Lampiran 2 ...

df	α								
	0,0025	0,0010	0,0005	0,00025	0,00010	0,00005	0,00025	0,00001	
1	127,321	318,309	636,619	1.273,239	3.183,099	6.366,198	12.732,395	31.830,989	
2	14,089	22,327	31,598	44,705	70,700	99,992	141,416	233,603	
3	7,453	10,214	12,924	16,326	22,204	28,000	35,298	47,928	
4	5,598	7,173	8,610	10,306	13,034	15,544	18,522	23,332	
5	4,773	5,893	6,869	7,976	9,678	11,178	12,893	15,547	
6	4,317	5,208	5,959	6,788	8,025	9,082	10,261	12,032	
7	4,029	4,785	5,408	6,082	7,063	7,885	8,782	10,103	
8	3,833	4,501	5,041	5,618	6,442	7,120	7,851	8,907	
9	3,690	4,297	4,781	5,291	6,010	6,594	7,215	8,102	
10	3,581	4,144	4,587	5,049	5,694	6,211	6,757	7,527	
11	3,497	4,025	4,437	4,863	5,453	5,921	6,412	7,098	
12	3,428	3,930	4,318	4,716	5,263	5,694	6,143	6,756	
13	3,372	3,852	4,221	4,597	5,111	5,513	5,928	6,501	
14	3,326	3,787	4,141	4,499	4,985	5,363	5,753	6,287	
15	3,286	3,733	4,073	4,417	4,880	5,239	5,607	6,109	
16	3,252	3,686	4,015	4,346	4,791	5,134	5,484	5,960	
17	3,223	3,646	3,965	4,286	4,714	5,044	5,379	5,832	
18	3,197	3,610	3,922	4,233	4,648	4,966	5,288	5,722	
19	3,174	3,579	3,883	4,187	4,590	4,897	5,209	5,627	
20	3,153	3,552	3,850	4,146	4,539	4,837	5,139	5,543	
21	3,135	3,527	3,819	4,110	4,493	4,784	5,077	5,469	
22	3,119	3,505	3,792	4,077	4,452	4,736	5,022	5,402	
23	3,104	3,485	3,768	4,048	4,415	4,693	4,972	5,343	
24	3,090	3,467	3,745	4,021	4,382	4,654	4,927	5,290	
25	3,078	3,450	3,725	3,997	4,352	4,619	4,887	5,241	
26	3,067	3,435	3,707	3,974	4,324	4,587	4,850	5,197	
27	3,057	3,421	3,690	3,954	4,299	4,558	4,816	5,157	
28	3,047	3,408	3,674	3,935	4,275	4,530	4,784	5,120	
29	3,038	3,396	3,659	3,918	4,254	4,506	4,756	5,086	
30	3,030	3,385	3,646	3,902	4,234	4,482	4,729	5,054	
40	2,971	3,307	3,551	3,788	4,094	4,321	4,544	4,835	
50	2,937	3,261	3,496	3,723	4,014	4,228	4,438	4,711	
60	2,915	3,232	3,460	3,681	3,962	4,169	4,370	4,631	
70	2,899	3,211	3,435	3,651	3,926	4,127	4,323	4,576	
80	2,887	3,195	3,416	3,629	3,899	4,096	4,288	4,535	
90	2,878	3,183	3,402	3,612	3,878	4,072	4,261	4,503	
100	2,871	3,174	3,390	3,598	3,862	4,053	4,240	4,478	
∞	2,807	3,090	3,291	3,481	3,719	3,891	4,056	4,265	

Lampiran 3: Distribusi F ($\alpha=0,1$)

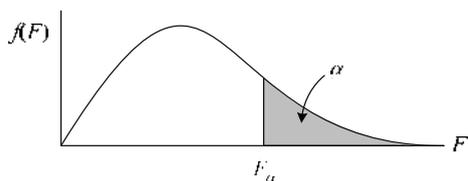


		Derajat bebas (df) pembilang (v_1)								
		1	2	3	4	5	6	7	8	9
Derajat bebas (df) penyebut (v_2)	1	39,86	49,50	53,59	55,83	57,24	58,20	58,91	59,44	59,86
	2	8,53	9,00	9,16	9,24	9,29	9,33	9,35	9,37	9,38
	3	5,54	5,46	5,39	5,34	5,31	5,28	5,27	5,25	5,24
	4	4,54	4,32	4,19	4,11	4,05	4,01	3,98	3,95	3,94
	5	4,06	3,78	3,62	3,52	3,45	3,40	3,37	3,34	3,32
	6	3,78	3,46	3,29	3,18	3,11	3,05	3,01	2,98	2,96
	7	3,59	3,26	3,07	2,96	2,88	2,83	2,78	2,75	2,72
	8	3,46	3,11	2,92	2,81	2,73	2,67	2,62	2,59	2,56
	9	3,36	3,01	2,81	2,69	2,61	2,55	2,51	2,47	2,44
	10	3,29	2,92	2,73	2,61	2,52	2,46	2,41	2,38	2,35
	11	3,23	2,86	2,66	2,54	2,45	2,39	2,34	2,30	2,27
	12	3,18	2,81	2,61	2,48	2,39	2,33	2,28	2,24	2,21
	13	3,14	2,76	2,56	2,43	2,35	2,28	2,23	2,20	2,16
	14	3,10	2,73	2,52	2,39	2,31	2,24	2,19	2,15	2,12
	15	3,07	2,70	2,49	2,36	2,27	2,21	2,16	2,12	2,09
	16	3,05	2,67	2,46	2,33	2,24	2,18	2,13	2,09	2,06
	17	3,03	2,64	2,44	2,31	2,22	2,15	2,10	2,06	2,03
	18	3,01	2,62	2,42	2,29	2,20	2,13	2,08	2,04	2,00
	19	2,99	2,61	2,40	2,27	2,18	2,11	2,06	2,02	1,98
	20	2,97	2,59	2,38	2,25	2,16	2,09	2,04	2,00	1,96
	21	2,96	2,57	2,36	2,23	2,14	2,08	2,02	1,98	1,95
	22	2,95	2,56	2,35	2,22	2,13	2,06	2,01	1,97	1,93
	23	2,94	2,55	2,34	2,21	2,11	2,05	1,99	1,95	1,92
	24	2,93	2,54	2,33	2,19	2,10	2,04	1,98	1,94	1,91
	25	2,92	2,53	2,32	2,18	2,09	2,02	1,97	1,93	1,89
	26	2,91	2,52	2,31	2,17	2,08	2,01	1,96	1,92	1,88
	27	2,90	2,51	2,30	2,17	2,07	2,00	1,95	1,91	1,87
	28	2,89	2,50	2,29	2,16	2,06	2,00	1,94	1,90	1,87
	29	2,89	2,50	2,28	2,15	2,06	1,99	1,93	1,89	1,86
	30	2,88	2,49	2,28	2,14	2,05	1,98	1,93	1,88	1,85
40	2,84	2,44	2,23	2,09	2,00	1,93	1,87	1,83	1,79	
60	2,79	2,39	2,18	2,04	1,95	1,87	1,82	1,77	1,74	
120	2,75	2,35	2,13	1,99	1,90	1,82	1,77	1,72	1,68	
∞	2,71	2,30	2,08	1,94	1,85	1,77	1,72	1,67	1,63	

Lanjutan Lampiran 3 ...

$v_2 \backslash v_1$	Derajat bebas (df) pembilang (v_1)										
	10	12	15	20	24	30	40	60	120	∞	
Derajat bebas (df) penyebut (v_2)	1	60,19	60,71	61,22	61,74	62,00	62,26	62,53	62,79	63,06	63,33
	2	9,39	9,41	9,42	9,44	9,45	9,46	9,47	9,47	9,48	9,49
	3	5,23	5,22	5,20	5,18	5,18	5,14	5,16	5,15	5,14	5,13
	4	3,92	3,90	3,87	3,84	3,83	3,82	3,80	3,79	3,78	3,76
	5	3,30	2,70	3,24	3,21	3,19	3,17	3,16	3,14	3,12	3,10
	6	2,94	2,90	2,87	2,84	2,82	2,80	2,78	2,76	2,74	2,72
	7	2,70	2,67	2,63	2,59	2,58	2,56	2,54	2,51	2,49	2,47
	8	2,54	2,50	2,46	2,42	2,40	2,38	2,36	2,34	2,32	2,29
	9	2,42	2,38	2,34	2,30	2,28	2,25	2,23	2,21	2,18	2,16
	10	2,32	2,28	2,24	2,20	2,18	2,16	2,13	2,11	2,08	2,06
	11	2,25	2,21	2,17	2,12	2,10	2,08	2,05	2,03	2,00	1,97
	12	2,19	2,15	2,10	2,06	2,04	2,01	1,99	1,96	1,93	1,90
	13	2,14	2,10	2,05	2,01	1,98	1,96	1,93	1,90	1,88	1,85
	14	2,10	2,05	2,01	1,96	1,94	1,91	1,89	1,86	1,83	1,80
	15	2,06	2,02	1,97	1,92	1,90	1,87	1,85	1,82	1,79	1,76
	16	2,03	1,99	1,94	1,89	1,87	1,84	1,81	1,78	1,75	1,72
	17	2,00	1,96	1,91	1,86	1,84	1,81	1,78	1,75	1,72	1,69
	18	1,98	1,93	1,89	1,84	1,81	1,78	1,75	1,72	1,69	1,66
	19	1,96	1,91	1,86	1,81	1,79	1,76	1,73	1,70	1,67	1,63
	20	1,94	1,89	1,84	1,79	1,77	1,74	1,71	1,68	1,64	1,61
	21	1,92	1,87	1,83	1,78	1,75	1,72	1,69	1,66	1,62	1,59
	22	1,90	1,86	1,81	1,76	1,73	1,70	1,67	1,64	1,60	1,57
	23	1,89	1,84	1,80	1,74	1,72	1,69	1,66	1,62	1,59	1,55
	24	1,88	1,83	1,78	1,73	1,70	1,67	1,64	1,61	1,57	1,53
	25	1,87	1,82	1,77	1,72	1,69	1,66	1,63	1,59	1,56	1,52
	26	1,86	1,81	1,76	1,71	1,68	1,65	1,61	1,58	1,54	1,50
	27	1,85	1,80	1,75	1,70	1,67	1,64	1,60	1,57	1,53	1,49
	28	1,84	1,79	1,74	1,69	1,66	1,63	1,59	1,56	1,52	1,48
	29	1,83	1,78	1,73	1,68	1,65	1,62	1,58	1,55	1,51	1,47
	30	1,82	1,77	1,72	1,67	1,64	1,61	1,57	1,54	1,50	1,46
40	1,76	1,71	1,66	1,61	1,57	1,54	1,151	1,47	1,42	1,38	
60	1,71	1,66	1,60	1,54	1,51	1,48	1,44	1,40	1,35	1,29	
120	1,65	1,60	1,55	1,48	1,45	1,41	1,37	1,32	1,26	1,19	
∞	1,60	1,55	1,49	1,42	1,38	1,34	1,30	1,24	1,17	1,00	

Lampiran 4: Distribusi F ($\alpha = 0,05$)



v_2	v_1	Derajat bebas (df) pembilang (v_1)								
		1	2	3	4	5	6	7	8	9
Derajat bebas (df) penyebut (v_2)	1	161,4	199,5	215,7	224,6	230,2	234,0	236,8	238,9	240,5
	2	18,51	19,0	19,16	19,25	19,30	19,33	19,35	19,37	19,38
	3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81
	4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00
	5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77
	6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10
	7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68
	8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39
	9	2,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18
	10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02
	11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90
	12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80
	13	4,67	3,81	3,41	3,18	3,03	2,92	2,93	2,77	2,71
	14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65
	15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59
	16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54
	17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49
	18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46
	19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42
	20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39
	21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37
	22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34
	23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32
	24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30
	25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28
	26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27
	27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25
	28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24
	29	4,18	3,33	2,93	2,70	2,55	2,43	2,35	2,28	2,22
	30	4,17	3,32	2,92	2,69	2,53	2,42	2,33	2,27	2,21
40	4,08	3,23	2,84	2,61	2,45	2,34	2,25	2,18	2,12	
60	4,00	3,15	2,76	2,53	2,37	2,25	2,17	2,10	2,04	
120	3,92	3,07	2,68	2,45	2,29	2,17	2,09	2,02	1,96	
∞	3,84	3,00	2,60	2,37	2,21	2,10	2,01	1,94	1,88	

Lanjutan Lampiran 4 ...

$v_1 \backslash v_2$	Derajat bebas (df) pembilang (v_1)								
	10	12	15	20	30	40	60	120	∞
1	241,9	243,9	245,9	248,0	250,1	251,1	252,2	253,3	254,3
2	19,40	19,41	19,43	19,45	19,46	19,47	19,48	19,49	19,50
3	8,79	8,74	8,70	8,66	8,62	8,59	8,57	8,55	8,53
4	5,96	5,91	5,86	5,80	5,75	5,72	5,69	5,66	5,63
5	4,74	4,68	4,62	4,56	4,50	4,46	4,43	4,40	4,36
6	4,06	4,00	3,94	3,87	3,81	3,77	3,74	3,70	3,67
7	3,64	3,57	3,51	3,44	3,38	3,34	3,30	3,27	3,23
8	3,35	3,28	3,22	3,15	3,08	3,04	3,01	2,97	2,93
9	3,14	3,07	3,01	2,94	2,86	2,83	2,79	2,75	2,71
10	2,98	2,91	2,85	2,77	2,70	2,66	2,62	2,58	2,54
11	2,85	2,79	2,72	2,65	2,57	2,53	2,49	2,45	2,40
12	2,75	2,69	2,62	2,54	2,47	2,43	2,38	2,34	2,30
13	2,67	2,60	2,53	2,46	2,38	2,34	2,30	2,25	2,21
14	2,60	2,53	2,46	2,39	2,31	2,27	2,22	2,18	2,13
15	2,54	2,48	2,40	2,33	2,25	2,20	2,16	2,11	2,07
16	2,49	2,42	2,35	2,28	2,19	2,15	2,11	2,06	2,01
17	2,45	2,38	2,31	2,23	2,15	2,10	2,06	2,01	1,96
18	2,41	2,34	2,27	2,19	2,11	2,06	2,02	1,97	1,92
19	2,38	2,31	2,23	2,16	2,07	2,03	1,98	1,93	1,88
20	2,35	2,28	2,20	2,12	2,04	1,99	1,95	1,90	1,84
21	2,32	2,25	2,18	2,10	2,01	1,96	1,92	1,87	1,81
22	2,30	2,23	2,15	2,07	1,98	1,94	1,89	1,84	1,78
23	2,27	2,20	2,13	2,05	1,96	1,91	1,86	1,81	1,76
24	2,25	2,18	2,11	2,03	1,94	1,89	1,84	1,79	1,73
25	2,24	2,16	2,09	2,01	1,92	1,87	1,82	1,77	1,71
26	2,22	2,15	2,07	1,99	1,90	1,85	1,80	1,75	1,69
27	2,20	2,13	2,06	1,97	1,88	1,84	1,79	1,73	1,67
28	2,19	2,12	2,04	1,96	1,87	1,82	1,77	1,71	1,65
29	2,18	2,10	2,03	1,94	1,85	1,81	1,75	1,70	1,64
30	2,16	2,09	2,01	1,93	1,84	1,79	1,74	1,68	1,62
40	2,08	2,00	1,92	1,84	1,74	1,69	1,64	1,58	1,51
60	1,99	1,92	1,84	1,75	1,65	1,59	1,53	1,47	1,39
120	1,91	1,83	1,75	1,66	1,55	1,50	1,43	1,35	1,25
∞	1,83	1,75	1,67	1,57	1,46	1,39	1,32	1,22	1,00

Lampiran 5: Tabel Durbin-Watson Statistika d pada $\alpha = 0,05$

n	$K' = 1$		$K' = 2$		$K' = 3$		$K' = 4$		$K' = 5$		$K' = 6$		$K' = 7$		$K' = 8$		$K' = 9$		$K' = 10$			
	d_t	d_u	d_t	d_u																		
6	0,610	1,400																				
7	0,700	1,356	0,467	1,896																		
8	0,763	1,332	0,559	1,777	0,367	2,287																
9	0,824	1,320	0,629	1,699	0,455	2,128	0,296	2,588														
10	0,879	1,320	0,697	1,641	0,525	2,016	0,376	2,414	0,243	2,822												
11	0,927	1,324	0,758	1,604	0,595	1,928	0,444	2,283	0,315	2,645	0,203	3,004										
12	0,971	1,331	0,812	1,579	0,658	1,864	0,512	2,177	0,380	2,506	0,268	2,832	0,171	3,149								
13	1,001	1,340	0,861	1,562	0,715	1,816	0,574	2,094	0,444	2,390	0,328	2,692	0,230	2,985	0,147	3,266						
14	1,045	1,350	0,905	1,551	0,767	1,779	0,632	2,030	0,505	2,296	0,389	2,572	0,286	2,848	0,200	3,111	0,127	3,36				
15	1,077	1,361	0,946	1,543	0,814	1,75	0,685	1,977	0,562	2,220	0,447	2,471	0,343	2,727	0,251	2,979	0,175	3,216	0,111	3,438		
16	1,106	1,371	0,982	1,539	0,857	1,728	0,734	1,935	0,615	2,157	0,502	2,388	0,398	2,624	0,304	2,860	0,222	3,090	0,155	3,304		
17	1,133	1,381	1,015	1,536	0,897	1,710	0,779	1,900	0,664	2,104	0,554	2,318	0,451	2,537	0,356	2,757	0,272	2,975	0,198	3,184		
18	1,158	1,391	1,046	1,535	0,933	1,696	0,820	1,872	0,710	2,060	0,603	2,258	0,502	2,461	0,407	2,668	0,321	2,873	0,244	3,073		
19	1,180	1,401	1,074	1,536	0,967	1,685	0,859	1,848	0,752	2,023	0,649	2,206	0,549	2,396	0,456	2,589	0,369	2,783	0,29	2,974		
20	1,201	1,411	1,100	1,537	0,998	1,676	0,894	1,828	0,792	1,991	0,691	2,162	0,595	2,339	0,502	2,521	0,416	2,704	0,336	2,885		
21	1,221	1,420	1,125	1,538	1,026	1,669	0,927	1,812	0,829	1,964	0,731	2,124	0,637	2,290	0,546	2,461	0,461	2,633	0,380	2,806		
22	1,239	1,429	1,147	1,541	1,053	1,664	0,958	1,797	0,863	1,940	0,769	2,090	0,677	2,246	0,588	2,407	0,504	2,571	0,424	2,735		
23	1,257	1,437	1,168	1,543	1,078	1,66	0,986	1,785	0,895	1,920	0,804	2,061	0,715	2,208	0,628	2,36	0,545	2,514	0,465	2,670		
24	1,273	1,446	1,188	1,546	1,101	1,656	1,013	1,775	0,925	1,902	0,837	2,035	0,750	2,174	0,666	2,318	0,584	2,464	0,506	2,613		
25	1,288	1,454	1,206	1,550	1,123	1,654	1,038	1,767	0,953	1,886	0,868	2,013	0,784	2,144	0,702	2,280	0,621	2,419	0,544	2,560		
26	1,302	1,461	1,224	1,553	1,143	1,652	1,062	1,759	0,979	1,873	0,897	1,992	0,816	2,117	0,735	2,246	0,657	2,379	0,581	2,513		
27	1,316	1,469	1,24	1,556	1,162	1,651	1,084	1,753	1,004	1,861	0,925	1,974	0,845	2,093	0,767	2,216	0,691	2,342	0,616	2,470		
28	1,328	1,476	1,255	1,560	1,181	1,650	1,104	1,747	1,028	1,850	0,951	1,959	0,874	2,071	0,798	2,188	0,723	2,309	0,649	2,431		
29	1,341	1,483	1,270	1,563	1,198	1,650	1,124	1,743	1,050	1,841	0,975	1,944	0,900	2,052	0,826	2,164	0,753	2,278	0,681	2,396		

n	K' = 1		K' = 2		K' = 3		K' = 4		K' = 5		K' = 6		K' = 7		K' = 8		K' = 9		K' = 10	
	d _l	d _u																		
30	1,352	1,489	1,284	1,567	1,214	1,650	1,143	1,739	1,071	1,833	0,998	1,931	0,926	2,034	0,854	2,141	0,782	2,251	0,712	2,363
31	1,363	1,496	1,297	1,570	1,229	1,650	1,160	1,735	1,090	1,825	1,020	1,920	0,950	2,018	0,879	2,120	0,810	2,226	0,741	2,333
32	1,373	1,502	1,309	1,574	1,244	1,650	1,177	1,732	1,109	1,819	1,041	1,909	0,972	2,004	0,904	2,102	0,836	2,203	0,769	2,306
33	1,383	1,508	1,321	1,577	1,258	1,651	1,193	1,730	1,127	1,813	1,061	1,900	0,994	1,991	0,927	2,085	0,861	2,181	0,796	2,281
34	1,393	1,514	1,333	1,580	1,271	1,652	1,208	1,728	1,144	1,808	1,079	1,891	1,015	1,978	0,950	2,069	0,885	2,162	0,821	2,257
35	1,402	1,519	1,343	1,584	1,283	1,653	1,222	1,726	1,160	1,803	1,097	1,884	1,034	1,967	0,971	2,054	0,908	2,144	0,845	2,236
36	1,411	1,525	1,354	1,587	1,295	1,654	1,236	1,724	1,175	1,799	1,114	1,876	1,053	1,957	0,991	2,041	0,930	2,127	0,868	2,216
37	1,419	1,530	1,364	1,590	1,307	1,655	1,249	1,723	1,190	1,795	1,131	1,870	1,071	1,948	1,011	2,029	0,951	2,112	0,891	2,197
38	1,427	1,535	1,373	1,594	1,318	1,656	1,261	1,722	1,204	1,792	1,146	1,864	1,088	1,939	1,029	2,017	0,970	2,098	0,912	2,180
39	1,435	1,540	1,382	1,597	1,328	1,658	1,273	1,722	1,218	1,789	1,161	1,859	1,104	1,932	1,047	2,007	0,990	2,085	0,932	2,164
40	1,442	1,544	1,391	1,600	1,338	1,659	1,285	1,721	1,23	1,786	1,175	1,854	1,120	1,924	1,064	1,997	1,008	2,072	0,952	2,149
45	1,475	1,566	1,430	1,615	1,383	1,666	1,336	1,72	1,287	1,776	1,238	1,835	1,189	1,895	1,139	1,958	1,089	2,022	1,038	2,088
50	1,503	1,585	1,462	1,628	1,421	1,674	1,378	1,721	1,335	1,771	1,291	1,822	1,246	1,875	1,201	1,930	1,156	1,986	1,110	2,044
55	1,528	1,601	1,490	1,641	1,452	1,681	1,414	1,724	1,374	1,768	1,334	1,814	1,294	1,861	1,253	1,909	1,212	1,959	1,170	2,010
60	1,549	1,616	1,514	1,652	1,48	1,689	1,444	1,727	1,408	1,767	1,372	1,808	1,335	1,85	1,298	1,894	1,260	1,939	1,222	1,984
65	1,567	1,629	1,536	1,662	1,503	1,696	1,471	1,731	1,438	1,767	1,404	1,805	1,370	1,843	1,336	1,882	1,301	1,923	1,266	1,964
70	1,583	1,641	1,554	1,672	1,525	1,703	1,494	1,735	1,464	1,768	1,433	1,802	1,401	1,838	1,369	1,874	1,337	1,910	1,305	1,948
75	1,598	1,652	1,571	1,680	1,543	1,709	1,515	1,739	1,487	1,770	1,458	1,801	1,428	1,834	1,399	1,867	1,369	1,901	1,339	1,935
80	1,611	1,662	1,586	1,688	1,56	1,715	1,534	1,743	1,507	1,772	1,480	1,801	1,453	1,831	1,425	1,861	1,397	1,893	1,369	1,925
85	1,624	1,671	1,600	1,696	1,575	1,721	1,55	1,747	1,525	1,774	1,500	1,801	1,474	1,829	1,448	1,857	1,422	1,886	1,396	1,916
90	1,635	1,679	1,612	1,703	1,589	1,726	1,566	1,751	1,542	1,776	1,518	1,801	1,494	1,827	1,469	1,854	1,445	1,881	1,420	1,909
95	1,645	1,687	1,623	1,709	1,602	1,732	1,579	1,755	1,557	1,778	1,535	1,802	1,512	1,827	1,489	1,852	1,465	1,877	1,442	1,903
100	1,654	1,694	1,634	1,715	1,613	1,736	1,592	1,758	1,571	1,780	1,55	1,803	1,528	1,826	1,506	1,850	1,484	1,874	1,462	1,898
150	1,720	1,747	1,706	1,760	1,693	1,774	1,679	1,788	1,665	1,802	1,651	1,817	1,637	1,832	1,622	1,846	1,608	1,862	1,593	1,877
200	1,758	1,779	1,748	1,789	1,738	1,799	1,728	1,809	1,718	1,820	1,707	1,831	1,697	1,841	1,686	1,852	1,675	1,863	1,665	1,874

Lanjutan Lampiran 5 ...

n	K' = 1		K' = 2		K' = 3		K' = 4		K' = 5		K' = 6		K' = 7		K' = 8		K' = 9		K' = 10		
	d _t	d ₀																			
6	0,610	1,400																			
7	0,700	1,356	0,467	1,896																	
8	0,763	1,332	0,559	1,777	0,367	2,287															
9	0,824	1,320	0,629	1,699	0,455	2,128	0,296	2,588													
10	0,879	1,320	0,697	1,641	0,525	2,016	0,376	2,414	0,243	2,822											
11	0,927	1,324	0,758	1,604	0,595	1,928	0,444	2,283	0,315	2,645	0,203	3,004									
12	0,971	1,331	0,812	1,579	0,658	1,864	0,512	2,177	0,380	2,506	0,268	2,832	0,171	3,149							
13	1,010	1,340	0,861	1,562	0,715	1,816	0,574	2,094	0,444	2,390	0,328	2,692	0,230	2,985	0,147	3,266					
14	1,045	1,350	0,905	1,551	0,767	1,779	0,632	2,030	0,505	2,296	0,389	2,572	0,286	2,848	0,200	3,111	0,127	3,36			
15	1,077	1,361	0,946	1,543	0,814	1,750	0,685	1,977	0,562	2,220	0,447	2,471	0,343	2,727	0,251	2,979	0,175	3,216	0,111	3,438	
16	1,106	1,371	0,982	1,539	0,857	1,728	0,734	1,935	0,615	2,157	0,502	2,388	0,398	2,624	0,304	2,860	0,222	3,09	0,155	3,304	
17	1,133	1,381	1,015	1,536	0,897	1,710	0,779	1,900	0,664	2,104	0,554	2,318	0,451	2,537	0,356	2,757	0,272	2,975	0,198	3,184	
18	1,158	1,391	1,046	1,535	0,933	1,696	0,820	1,872	0,710	2,060	0,603	2,258	0,502	2,461	0,407	2,668	0,321	2,873	0,244	3,073	
19	1,18	1,401	1,074	1,536	0,967	1,685	0,859	1,848	0,752	2,023	0,649	2,206	0,549	2,396	0,456	2,589	0,369	2,783	0,29	2,974	
20	1,201	1,411	1,100	1,537	0,998	1,676	0,894	1,828	0,792	1,991	0,691	2,162	0,595	2,339	0,502	2,521	0,416	2,704	0,336	2,885	
21	1,221	1,420	1,125	1,538	1,026	1,669	0,927	1,812	0,829	1,964	0,731	2,124	0,637	2,290	0,546	2,461	0,461	2,633	0,38	2,806	
22	1,239	1,429	1,147	1,541	1,053	1,664	0,958	1,797	0,863	1,940	0,769	2,090	0,677	2,246	0,588	2,407	0,504	2,571	0,424	2,735	
23	1,257	1,437	1,168	1,543	1,078	1,660	0,986	1,785	0,895	1,920	0,804	2,061	0,715	2,208	0,628	2,360	0,545	2,514	0,465	2,670	
24	1,273	1,446	1,188	1,546	1,101	1,656	1,013	1,775	0,925	1,902	0,837	2,035	0,750	2,174	0,666	2,318	0,584	2,464	0,506	2,613	
25	1,288	1,454	1,206	1,550	1,123	1,654	1,038	1,767	0,953	1,886	0,868	2,013	0,784	2,144	0,702	2,280	0,621	2,419	0,544	2,56	
26	1,302	1,461	1,224	1,553	1,143	1,652	1,062	1,759	0,979	1,873	0,897	1,992	0,816	2,117	0,735	2,246	0,657	2,379	0,581	2,513	
27	1,316	1,469	1,240	1,556	1,162	1,651	1,084	1,753	1,004	1,861	0,925	1,974	0,845	2,093	0,767	2,216	0,691	2,342	0,616	2,470	
28	1,328	1,476	1,255	1,560	1,181	1,650	1,104	1,747	1,028	1,850	0,951	1,959	0,874	2,071	0,798	2,188	0,723	2,309	0,649	2,431	
29	1,341	1,483	1,270	1,563	1,198	1,650	1,124	1,743	1,050	1,841	0,975	1,944	0,900	2,052	0,826	2,164	0,753	2,278	0,681	2,396	

n	K' = 1		K' = 2		K' = 3		K' = 4		K' = 5		K' = 6		K' = 7		K' = 8		K' = 9		K' = 10	
	d _t	d _o																		
30	1,352	1,489	1,284	1,567	1,214	1,650	1,143	1,739	1,071	1,833	0,998	1,931	0,926	2,034	0,854	2,141	0,782	2,251	0,712	2,363
31	1,363	1,496	1,297	1,570	1,229	1,650	1,160	1,735	1,090	1,825	1,020	1,920	0,950	2,018	0,879	2,120	0,810	2,226	0,741	2,333
32	1,373	1,502	1,309	1,574	1,244	1,650	1,177	1,732	1,109	1,819	1,041	1,909	0,972	2,004	0,904	2,102	0,836	2,203	0,769	2,306
33	1,383	1,508	1,321	1,577	1,258	1,651	1,193	1,730	1,127	1,813	1,061	1,900	0,994	1,991	0,927	2,085	0,861	2,181	0,796	2,281
34	1,393	1,514	1,333	1,580	1,271	1,652	1,208	1,728	1,144	1,808	1,079	1,891	1,015	1,978	0,95	2,069	0,885	2,162	0,821	2,257
35	1,402	1,519	1,343	1,584	1,283	1,653	1,222	1,726	1,160	1,803	1,097	1,884	1,034	1,967	0,971	2,054	0,908	2,144	0,845	2,236
36	1,411	1,525	1,354	1,587	1,295	1,654	1,236	1,724	1,175	1,799	1,114	1,876	1,053	1,957	0,991	2,041	0,930	2,127	0,868	2,216
37	1,419	1,530	1,364	1,590	1,307	1,655	1,249	1,723	1,190	1,795	1,131	1,870	1,071	1,948	1,011	2,029	0,951	2,112	0,891	2,197
38	1,427	1,535	1,373	1,594	1,318	1,656	1,261	1,722	1,204	1,792	1,146	1,864	1,088	1,939	1,029	2,017	0,970	2,098	0,912	2,180
39	1,435	1,540	1,382	1,597	1,328	1,658	1,273	1,722	1,218	1,789	1,161	1,859	1,104	1,932	1,047	2,007	0,990	2,085	0,932	2,164
40	1,442	1,544	1,391	1,600	1,338	1,659	1,285	1,721	1,230	1,786	1,175	1,854	1,120	1,924	1,064	1,997	1,008	2,072	0,952	2,149
45	1,475	1,566	1,430	1,615	1,383	1,666	1,336	1,720	1,287	1,776	1,238	1,835	1,189	1,895	1,139	1,958	1,089	2,022	1,038	2,088
50	1,503	1,585	1,462	1,628	1,421	1,674	1,378	1,721	1,335	1,771	1,291	1,822	1,246	1,875	1,201	1,930	1,156	1,986	1,110	2,044
55	1,528	1,601	1,490	1,641	1,452	1,681	1,414	1,724	1,374	1,768	1,334	1,814	1,294	1,861	1,253	1,909	1,212	1,959	1,170	2,010
60	1,549	1,616	1,514	1,652	1,480	1,689	1,444	1,727	1,408	1,767	1,372	1,808	1,335	1,850	1,298	1,894	1,260	1,939	1,222	1,984
65	1,567	1,629	1,536	1,662	1,503	1,696	1,471	1,731	1,438	1,767	1,404	1,805	1,370	1,843	1,336	1,882	1,301	1,923	1,266	1,964
70	1,583	1,641	1,554	1,672	1,525	1,703	1,494	1,735	1,464	1,768	1,433	1,802	1,401	1,838	1,369	1,874	1,337	1,910	1,305	1,948
75	1,598	1,652	1,571	1,680	1,543	1,709	1,515	1,739	1,487	1,770	1,458	1,801	1,428	1,834	1,399	1,867	1,369	1,901	1,339	1,935
80	1,611	1,662	1,586	1,688	1,560	1,715	1,534	1,743	1,507	1,772	1,480	1,801	1,453	1,831	1,425	1,861	1,397	1,893	1,369	1,925
85	1,624	1,671	1,600	1,696	1,575	1,721	1,550	1,747	1,525	1,774	1,500	1,801	1,474	1,829	1,448	1,857	1,422	1,886	1,396	1,916
90	1,635	1,679	1,612	1,703	1,589	1,726	1,566	1,751	1,542	1,776	1,518	1,801	1,494	1,827	1,469	1,854	1,445	1,881	1,420	1,909
95	1,645	1,687	1,623	1,709	1,602	1,732	1,579	1,755	1,557	1,778	1,535	1,802	1,512	1,827	1,489	1,852	1,465	1,877	1,442	1,903
100	1,654	1,694	1,634	1,715	1,613	1,736	1,592	1,758	1,571	1,780	1,550	1,803	1,528	1,826	1,506	1,850	1,484	1,874	1,462	1,898
150	1,720	1,747	1,706	1,760	1,693	1,774	1,679	1,788	1,665	1,802	1,651	1,817	1,637	1,832	1,622	1,846	1,608	1,862	1,593	1,877
200	1,758	1,779	1,748	1,789	1,738	1,799	1,728	1,809	1,718	1,820	1,707	1,831	1,697	1,841	1,686	1,852	1,675	1,863	1,665	1,874

Lampiran 6: Tabel Durbin-Watson Statistika d pada $\alpha = 0,01$

n	$k' = 1$		$k' = 2$		$k' = 3$		$k' = 4$		$k' = 5$		$k' = 6$		$k' = 7$		$k' = 8$		$k' = 9$		$k' = 10$			
	d_L	d_U	d_L	d_U																		
6	0,390	1,142																				
7	0,435	1,036	0,294	1,676																		
8	0,497	1,003	0,345	1,489	0,229	2,102																
9	0,554	0,998	0,408	1,389	0,279	1,875	0,183	2,433														
10	0,604	1,001	0,466	1,333	0,340	1,733	0,230	2,193	0,150	2,690												
11	0,653	1,010	0,519	1,297	0,396	1,640	0,286	2,030	0,193	2,453	0,124	2,892										
12	0,697	1,023	0,569	1,274	0,449	1,575	0,339	1,913	0,244	2,280	0,164	2,665	0,105	3,053								
13	0,738	1,038	0,616	1,261	0,499	1,526	0,391	1,826	0,294	2,150	0,211	2,490	0,140	2,838	0,090	3,182						
14	0,776	1,054	0,660	1,254	0,547	1,490	0,441	1,757	0,343	2,049	0,257	2,354	0,183	2,667	0,122	2,981	0,078	3,287				
15	0,811	1,070	0,700	1,252	0,591	1,465	0,487	1,705	0,390	1,967	0,303	2,244	0,226	2,530	0,161	2,817	0,107	3,101	0,068	3,374		
16	0,844	1,086	0,738	1,253	0,633	1,447	0,532	1,664	0,437	1,901	0,349	2,153	0,269	2,416	0,200	2,681	0,142	2,944	0,094	3,201		
17	0,873	1,102	0,773	1,255	0,672	1,432	0,574	1,631	0,481	1,847	0,393	2,078	0,313	2,319	0,241	2,566	0,179	2,811	0,127	3,053		
18	0,902	1,118	0,805	1,259	0,708	1,422	0,614	1,604	0,522	1,803	0,435	2,015	0,355	2,238	0,282	2,467	0,216	2,697	0,160	2,925		
19	0,928	1,133	0,835	1,264	0,742	1,416	0,650	1,583	0,561	1,767	0,476	1,963	0,396	2,169	0,322	2,381	0,255	2,597	0,196	2,813		
20	0,952	1,147	0,862	1,270	0,774	1,410	0,684	1,567	0,598	1,736	0,515	1,918	0,436	2,110	0,362	2,308	0,294	2,510	0,232	2,174		
21	0,975	1,161	0,889	1,276	0,803	1,408	0,718	1,554	0,634	1,712	0,552	1,881	0,474	2,059	0,400	2,244	0,331	2,434	0,268	2,625		
22	0,997	1,174	0,915	1,284	0,832	1,407	0,748	1,543	0,666	1,691	0,587	1,849	0,510	2,015	0,437	2,188	0,368	2,367	0,304	2,548		
23	1,017	1,186	0,938	1,290	0,858	1,407	0,777	1,535	0,699	1,674	0,620	1,821	0,545	1,977	0,473	2,140	0,404	2,308	0,340	2,479		
24	1,037	1,199	0,959	1,298	0,881	1,407	0,805	1,527	0,728	1,659	0,652	1,797	0,578	1,944	0,507	2,097	0,439	2,255	0,375	2,417		
25	1,055	1,210	0,981	1,305	0,906	1,408	0,832	1,521	0,756	1,645	0,682	1,776	0,610	1,915	0,540	2,059	0,473	2,209	0,409	2,362		
26	1,072	1,222	1,000	1,311	0,928	1,410	0,855	1,517	0,782	1,635	0,711	1,759	0,640	1,889	0,572	2,026	0,505	2,168	0,441	2,313		
27	1,088	1,232	1,019	1,318	0,948	1,413	0,878	1,514	0,808	1,625	0,738	1,743	0,669	1,867	0,602	1,997	0,536	2,131	0,473	2,269		
28	1,104	1,244	1,036	1,325	0,969	1,414	0,901	1,512	0,832	1,618	0,764	1,729	0,696	1,847	0,630	1,970	0,566	2,098	0,504	2,229		
29	1,119	1,254	1,053	1,332	0,988	1,418	0,921	1,511	0,855	1,611	0,788	1,718	0,723	1,830	0,658	1,947	0,595	2,068	0,533	2,193		

n	K' = 1		K' = 2		K' = 3		K' = 4		K' = 5		K' = 6		K' = 7		K' = 8		K' = 9		K' = 10	
	d _t	d _u																		
30	1,134	1,264	1,070	1,339	1,006	1,421	0,941	1,510	0,877	1,606	0,812	1,707	0,748	1,814	0,684	1,925	0,622	2,041	0,562	2,160
31	1,147	1,274	1,085	1,345	1,022	1,425	0,960	1,509	0,897	1,601	0,834	1,698	0,772	1,800	0,710	1,906	0,649	2,017	0,589	2,131
32	1,160	1,283	1,100	1,351	1,039	1,428	0,978	1,509	0,917	1,597	0,856	1,690	0,794	1,788	0,734	1,889	0,674	1,995	0,615	2,104
33	1,171	1,291	1,114	1,358	1,055	1,432	0,995	1,510	0,935	1,594	0,876	1,683	0,816	1,776	0,757	1,874	0,698	1,975	0,641	2,080
34	1,184	1,298	1,128	1,364	1,070	1,436	1,012	1,511	0,954	1,591	0,896	1,677	0,837	1,766	0,779	1,860	0,722	1,957	0,665	2,057
35	1,195	1,307	1,141	1,370	1,085	1,439	1,028	1,512	0,971	1,589	0,914	1,671	0,857	1,757	0,800	1,847	0,744	1,940	0,689	2,037
36	1,205	1,315	1,153	1,376	1,098	1,442	1,043	1,513	0,987	1,587	0,932	1,666	0,877	1,749	0,821	1,836	0,766	1,925	0,711	2,018
37	1,217	1,322	1,164	1,383	1,112	1,446	1,058	1,514	1,004	1,585	0,950	1,662	0,895	1,742	0,841	1,825	0,787	1,911	0,733	2,001
38	1,227	1,330	1,176	1,388	1,124	1,449	1,072	1,515	1,019	1,584	0,966	1,658	0,913	1,735	0,860	1,816	0,807	1,899	0,754	1,985
39	1,237	1,337	1,187	1,392	1,137	1,452	1,085	1,517	1,033	1,583	0,982	1,655	0,930	1,729	0,878	1,807	0,826	1,887	0,774	1,970
40	1,246	1,344	1,197	1,398	1,149	1,456	1,098	1,518	1,047	1,583	0,997	1,652	0,946	1,724	0,895	1,799	0,844	1,876	0,749	1,956
45	1,288	1,376	1,245	1,424	1,201	1,474	1,156	1,528	1,111	1,583	1,065	1,643	1,019	1,704	0,974	1,768	0,927	1,834	0,881	1,902
50	1,324	1,403	1,285	1,445	1,245	1,491	1,206	1,537	1,164	1,587	1,123	1,639	1,081	1,692	1,039	1,748	0,997	1,805	0,955	1,864
55	1,356	1,428	1,320	1,466	1,284	1,505	1,246	1,548	1,209	1,592	1,172	1,638	1,134	1,685	1,095	1,734	1,057	1,785	1,018	1,837
60	1,382	1,449	1,351	1,484	1,317	1,520	1,283	1,559	1,248	1,598	1,214	1,639	1,179	1,682	1,144	1,726	1,108	1,771	1,072	1,817
65	1,407	1,467	1,377	1,500	1,346	1,534	1,314	1,568	1,283	1,604	1,251	1,642	1,218	1,680	1,186	1,720	1,153	1,761	1,120	1,802
70	1,429	1,485	1,400	1,514	1,372	1,546	1,343	1,577	1,313	1,611	1,283	1,645	1,253	1,680	1,223	1,716	1,192	1,754	1,162	1,792
75	1,448	1,501	1,422	1,529	1,395	1,557	1,368	1,586	1,340	1,617	1,313	1,649	1,284	1,682	1,256	1,714	1,227	1,748	1,199	1,783
80	1,465	1,514	1,440	1,541	1,416	1,568	1,390	1,595	1,364	1,624	1,338	1,653	1,312	1,683	1,285	1,714	1,259	1,745	1,232	1,777
85	1,481	1,529	1,458	1,553	1,434	1,577	1,411	1,603	1,386	1,630	1,362	1,657	1,337	1,685	1,312	1,714	1,287	1,743	1,262	1,773
90	1,496	1,541	1,474	1,563	1,452	1,587	1,429	1,611	1,406	1,636	1,383	1,661	1,360	1,687	1,336	1,714	1,312	1,741	1,288	1,769
95	1,510	1,552	1,489	1,573	1,468	1,596	1,446	1,618	1,425	1,641	1,403	1,666	1,381	1,690	1,358	1,715	1,336	1,741	1,313	1,767
100	1,522	1,562	1,502	1,582	1,482	1,604	1,461	1,625	1,441	1,647	1,421	1,670	1,400	1,693	1,378	1,717	1,357	1,741	1,335	1,765
150	1,611	1,637	1,598	1,651	1,584	1,665	1,571	1,679	1,557	1,693	1,543	1,708	1,530	1,722	1,515	1,737	1,501	1,752	1,486	1,767
200	1,664	1,684	1,653	1,693	1,643	1,704	1,633	1,715	1,623	1,725	1,613	1,735	1,603	1,746	1,592	1,757	1,582	1,768	1,571	1,779

Lanjutan Lampiran 6 ...

n	K' = 1		K' = 2		K' = 3		K' = 4		K' = 5		K' = 6		K' = 7		K' = 8		K' = 9		K' = 10		
	d _t	d ₀																			
6	0,610	1,400																			
7	0,700	1,356	0,467	1,896																	
8	0,763	1,332	0,559	1,777	0,367	2,287															
9	0,824	1,320	0,629	1,699	0,455	2,128	0,296	2,588													
10	0,879	1,320	0,697	1,641	0,525	2,016	0,376	2,414	0,243	2,822											
11	0,927	1,324	0,758	1,604	0,595	1,928	0,444	2,283	0,315	2,645	0,203	3,004									
12	0,971	1,331	0,812	1,579	0,658	1,864	0,512	2,177	0,380	2,506	0,268	2,832	0,171	3,149							
13	1,010	1,340	0,861	1,562	0,715	1,816	0,574	2,094	0,444	2,390	0,328	2,692	0,230	2,985	0,147	3,266					
14	1,045	1,350	0,905	1,551	0,767	1,779	0,632	2,030	0,505	2,296	0,389	2,572	0,286	2,848	0,200	3,111	0,127	3,360			
15	1,077	1,361	0,946	1,543	0,814	1,750	0,685	1,977	0,562	2,220	0,447	2,471	0,343	2,727	0,251	2,979	0,175	3,216	0,111	3,438	
16	1,106	1,371	0,982	1,539	0,857	1,728	0,734	1,935	0,615	2,157	0,502	2,388	0,398	2,624	0,304	2,860	0,222	3,090	0,155	3,304	
17	1,133	1,381	1,015	1,536	0,897	1,710	0,779	1,900	0,664	2,104	0,554	2,318	0,451	2,537	0,356	2,757	0,272	2,975	0,198	3,184	
18	1,158	1,391	1,046	1,535	0,933	1,696	0,820	1,872	0,710	2,060	0,603	2,258	0,502	2,461	0,407	2,668	0,321	2,873	0,244	3,073	
19	1,180	1,401	1,074	1,536	0,967	1,685	0,859	1,848	0,752	2,023	0,649	2,206	0,549	2,396	0,456	2,589	0,369	2,783	0,29	2,974	
20	1,201	1,411	1,100	1,537	0,998	1,676	0,894	1,828	0,792	1,991	0,691	2,162	0,595	2,339	0,502	2,521	0,416	2,704	0,336	2,885	
21	1,221	1,420	1,125	1,538	1,026	1,669	0,927	1,812	0,829	1,964	0,731	2,124	0,637	2,290	0,546	2,461	0,461	2,633	0,38	2,806	
22	1,239	1,429	1,147	1,541	1,053	1,664	0,958	1,797	0,863	1,940	0,769	2,090	0,677	2,246	0,588	2,407	0,504	2,571	0,424	2,735	
23	1,257	1,437	1,168	1,543	1,078	1,660	0,986	1,785	0,895	1,920	0,804	2,061	0,715	2,208	0,628	2,360	0,545	2,514	0,465	2,670	
24	1,273	1,446	1,188	1,546	1,101	1,656	1,013	1,775	0,925	1,902	0,837	2,035	0,75	2,174	0,666	2,318	0,584	2,464	0,506	2,613	
25	1,288	1,454	1,206	1,550	1,123	1,654	1,038	1,767	0,953	1,886	0,868	2,013	0,784	2,144	0,702	2,280	0,621	2,419	0,544	2,560	
26	1,302	1,461	1,224	1,553	1,143	1,652	1,062	1,759	0,979	1,873	0,897	1,992	0,816	2,117	0,735	2,246	0,657	2,379	0,581	2,513	
27	1,316	1,469	1,240	1,556	1,162	1,651	1,084	1,753	1,004	1,861	0,925	1,974	0,845	2,093	0,767	2,216	0,691	2,342	0,616	2,470	
28	1,328	1,476	1,255	1,560	1,181	1,650	1,104	1,747	1,028	1,850	0,951	1,959	0,874	2,071	0,798	2,188	0,723	2,309	0,649	2,431	
29	1,341	1,483	1,270	1,563	1,198	1,650	1,124	1,743	1,050	1,841	0,975	1,944	0,900	2,052	0,826	2,164	0,753	2,278	0,681	2,396	

n	K' = 1		K' = 2		K' = 3		K' = 4		K' = 5		K' = 6		K' = 7		K' = 8		K' = 9		K' = 10	
	d _t	d _o																		
30	1,352	1,489	1,284	1,567	1,214	1,650	1,143	1,739	1,071	1,833	0,998	1,931	0,926	2,034	0,854	2,141	0,782	2,251	0,712	2,363
31	1,363	1,496	1,297	1,570	1,229	1,650	1,160	1,735	1,090	1,825	1,020	1,920	0,950	2,018	0,879	2,120	0,810	2,226	0,741	2,333
32	1,373	1,502	1,309	1,574	1,244	1,650	1,177	1,732	1,109	1,819	1,041	1,909	0,972	2,004	0,904	2,102	0,836	2,203	0,769	2,306
33	1,383	1,508	1,321	1,577	1,258	1,651	1,193	1,730	1,127	1,813	1,061	1,900	0,994	1,991	0,927	2,085	0,861	2,181	0,796	2,281
34	1,393	1,514	1,333	1,580	1,271	1,652	1,208	1,728	1,144	1,808	1,079	1,891	1,015	1,978	0,950	2,069	0,885	2,162	0,821	2,257
35	1,402	1,519	1,343	1,584	1,283	1,653	1,222	1,726	1,160	1,803	1,097	1,884	1,034	1,967	0,971	2,054	0,908	2,144	0,845	2,236
36	1,411	1,525	1,354	1,587	1,295	1,654	1,236	1,724	1,175	1,799	1,114	1,876	1,053	1,957	0,991	2,041	0,930	2,127	0,868	2,216
37	1,419	1,530	1,364	1,590	1,307	1,655	1,249	1,723	1,190	1,795	1,131	1,870	1,071	1,948	1,011	2,029	0,951	2,112	0,891	2,197
38	1,427	1,535	1,373	1,594	1,318	1,656	1,261	1,722	1,204	1,792	1,146	1,864	1,088	1,939	1,029	2,017	0,970	2,098	0,912	2,180
39	1,435	1,540	1,382	1,597	1,328	1,658	1,273	1,722	1,218	1,789	1,161	1,859	1,104	1,932	1,047	2,007	0,990	2,085	0,932	2,164
40	1,442	1,544	1,391	1,600	1,338	1,659	1,285	1,721	1,230	1,786	1,175	1,854	1,120	1,924	1,064	1,997	1,008	2,072	0,952	2,149
45	1,475	1,566	1,430	1,615	1,383	1,666	1,336	1,720	1,287	1,776	1,238	1,835	1,189	1,895	1,139	1,958	1,089	2,022	1,038	2,088
50	1,503	1,585	1,462	1,628	1,421	1,674	1,378	1,721	1,335	1,771	1,291	1,822	1,246	1,875	1,201	1,930	1,156	1,986	1,110	2,044
55	1,528	1,601	1,490	1,641	1,452	1,681	1,414	1,724	1,374	1,768	1,334	1,814	1,294	1,861	1,253	1,909	1,212	1,959	1,170	2,010
60	1,549	1,616	1,514	1,652	1,480	1,689	1,444	1,727	1,408	1,767	1,372	1,808	1,335	1,850	1,298	1,894	1,260	1,939	1,222	1,984
65	1,567	1,629	1,536	1,662	1,503	1,696	1,471	1,731	1,438	1,767	1,404	1,805	1,370	1,843	1,336	1,882	1,301	1,923	1,266	1,964
70	1,583	1,641	1,554	1,672	1,525	1,703	1,494	1,735	1,464	1,768	1,433	1,802	1,401	1,838	1,369	1,874	1,337	1,910	1,305	1,948
75	1,598	1,652	1,571	1,680	1,543	1,709	1,515	1,739	1,487	1,770	1,458	1,801	1,428	1,834	1,399	1,867	1,369	1,901	1,339	1,935
80	1,611	1,662	1,586	1,688	1,560	1,715	1,534	1,743	1,507	1,772	1,48	1,801	1,453	1,831	1,425	1,861	1,397	1,893	1,369	1,925
85	1,624	1,671	1,600	1,696	1,575	1,721	1,550	1,747	1,525	1,774	1,500	1,801	1,474	1,829	1,448	1,857	1,422	1,886	1,396	1,916
90	1,635	1,679	1,612	1,703	1,589	1,726	1,566	1,751	1,542	1,776	1,518	1,801	1,494	1,827	1,469	1,854	1,445	1,881	1,420	1,909
95	1,645	1,687	1,623	1,709	1,602	1,732	1,579	1,755	1,557	1,778	1,535	1,802	1,512	1,827	1,489	1,852	1,465	1,877	1,442	1,903
100	1,654	1,694	1,634	1,715	1,613	1,736	1,592	1,758	1,571	1,780	1,550	1,803	1,528	1,826	1,506	1,850	1,484	1,874	1,462	1,898
150	1,720	1,747	1,706	1,760	1,693	1,774	1,679	1,788	1,665	1,802	1,651	1,817	1,637	1,832	1,622	1,846	1,608	1,862	1,593	1,877
200	1,758	1,779	1,748	1,789	1,738	1,799	1,728	1,809	1,718	1,820	1,707	1,831	1,697	1,841	1,686	1,852	1,675	1,863	1,665	1,874



STATISTIKA

Kebutuhan akan analisis data menggunakan statistika dalam semua bidang, membuat Mata Kuliah Statistika menjadi mata kuliah yang diajarkan di hampir semua program studi. Statistika tidak hanya bicara mengolah data, tapi juga bagaimana mendapatkan data yang baik, menampilkannya dalam bentuk yang mudah dipahami, mengolahnya, hingga menyimpulkan informasi dari data yang ada hingga diperoleh informasi adalah bagian dari statistika.

Buku ini hadir sebagai buku ajar Statistika yang diharapkan akan berguna dalam pengajaran Mata Kuliah Statistika. Pengertian Statistika dan bagaimana sejarah statistika serta dasar-dasar dari Statistika dibahas dalam Bab 1. Teknik-teknik pengambilan sampel juga dibahas pada Bab 1. Selanjutnya, ukuran-ukuran data yang menjelaskan tipe data dan skala data dibahas lebih rinci pada Bab 2. Bab ini diperlukan sebagai dasar pengetahuan ilmu data agar dalam pengolahannya tepat dan memberi informasi yang bermanfaat. Bab 3 membahas tentang dasar-dasar probabilitas yang merupakan dasar dari Statistika. Ilmu probabilitas ini yang akan membantu pengolahan data saat terdapat ketidakpastian. Dalam Statistika, setelah mendapatkan data, bagaimana menampilkannya dan meringkasnya akan dituntaskan untuk dibahas pada Bab 4. Selanjutnya masuk pada analisis inferensia yaitu uji hipotesis yang dibahas pada Bab 5. Bab terakhir adalah Bab 6 akan membahas bagaimana analisis inferensia lebih lanjut yaitu analisis korelasi dan analisis regresi.

Buku telah disusun sedemikian rupa sehingga mudah dipahami. Contoh-contoh soal dibuat beserta penyelesaiannya agar mahasiswa mampu menyelesaikan dengan baik. Penjelasan beserta filosofi dari analisis juga dijelaskan supaya pembaca memahami mengapa suatu analisis dilakukan.



ISBN: 978-602-6248-88-6



9 786026 248886