



METODE STATISTIKA



Andi Daniah Pahrany, Ramdhan Fazrianto Suwarman,
Lita Wulandari Aeli, Mukhammad Solikhin.

METODE STATISTIKA

Andi Daniah Pahrany, Ramdhan Fazrianto Suwarman,
Lita Wulandari Aeli, Mukhammad Solikhin.



METODE STATISTIKA

Penulis:

Andi Daniah Pahrany, Ramdhan Fazrianto Suwarman,
Lita Wulandari Aeli, Mukhammad Solikhin.

Editor:

Erik Santoso

Layouter :

Tim Kreatif PRCI

Cover:

Rusli

Cetakan Pertama : Oktober 2023

Hak Cipta 2023, pada Penulis. Diterbitkan pertama kali oleh:

Perkumpulan Rumah Cemerlang Indonesia

ANGGOTA IKAPI JAWA BARAT

Pondok Karisma Residence Jalan Raflesia VI D.151
Panglayungan, Cipedes Tasikmalaya – 085223186009

Website : www.rcipress.rcipublisher.org

E-mail : rumahcemerlangindonesia@gmail.com

Copyright © 2023 by Perkumpulan Rumah Cemerlang Indonesia

All Right Reserved

- Cet. I – : Perkumpulan Rumah Cemerlang Indonesia, 2023

; 18 x 25 cm

ISBN : 978-623-448-685-8

Hak cipta dilindungi undang-undang

Dilarang memperbanyak buku ini dalam bentuk dan dengan
cara apapun tanpa izin tertulis dari penulis dan penerbit

Undang-undang No.19 Tahun 2002 Tentang

Hak Cipta Pasal 72

Undang-undang No.19 Tahun 2002 Tentang Hak Cipta
Pasal 72

Barang siapa dengan sengaja melanggar dan tanpa hak melakukan perbuatan sebagaimana dimaksud dalam pasal ayat (1) atau pasal 49 ayat (1) dan ayat (2) dipidana dengan pidana penjara masing-masing paling sedikit 1 (satu) bulan dan/atau denda paling sedikit Rp.1.000.000,00 (satu juta rupiah), atau pidana penjara paling lama 7 (tujuh) tahun dan/atau denda paling banyak Rp.5.000.000.000,00 (lima miliar rupiah).

Barang siapa dengan sengaja menyiarkan, memamerkan, mengedarkan, atau menjual kepada umum suatu ciptaan atau barang hasil pelanggaran hak cipta terkait sebagai dimaksud pada ayat (1) dipidana dengan pidana penjara paling lama 5 (lima) tahun dan/atau denda paling banyak Rp.500.000.000,00 (lima ratus juta rupiah)

KATA PENGANTAR

Statistika merupakan salah satu ilmu yang berperan penting dalam dunia pendidikan. Seiring berjalannya waktu, penggunaan data serta analisis statistika menjadi suatu hal yang memegang peran penting dalam proses pengambilan keputusan. Terutama pada institusi pendidikan, mulai dari jenjang sekolah dasar hingga perguruan tinggi. Di era informasi dan teknologi yang sangat berkembang ini, pemahaman terkait statistika menjadi penting dalam memecahkan masalah-masalah kompleks dalam kehidupan.

Dalam pengaturan akademis, baik mahasiswa maupun dosen harus memiliki pemahaman yang kuat tentang statistika. Dosen memerlukan pengetahuan yang mendalam tentang metode statistika untuk penelitian dan penilaian akademik, sementara mahasiswa diharapkan mampu menerapkan konsep statistika dalam penelitian mereka dan sebagai landasan bagi pengambilan keputusan yang tepat.

Namun, pengajaran statistika pada jenjang perkuliahan seringkali dihadapi dengan beberapa tantangan. Materi yang kompleks dan terminologi teknis sering membuat mahasiswa dan bahkan beberapa dosen merasa cemas. Oleh karena itu, ada kebutuhan yang mendalam untuk sumber daya pendidikan yang dapat menyajikan statistika dengan cara yang mudah dipahami dan relevan dengan konteks pendidikan tinggi.

Buku ini dibuat dari kesadaran akan kebutuhan ini. Kami memahami bahwa untuk berhasil di dunia pendidikan tinggi, baik dosen maupun mahasiswa perlu memiliki pemahaman yang kuat tentang statistika. Oleh karena itu, buku ini dirancang sebagai panduan komprehensif yang mencakup berbagai aspek statistika yang relevan dalam konteks pendidikan tinggi.

Melalui buku ini, kami bertujuan untuk memberikan bantuan dalam merinci konsep-konsep statistika dengan jelas, menawarkan contoh praktis yang relevan dengan dunia pendidikan tinggi, dan memberikan panduan yang langkah demi langkah untuk menerapkan metode statistika dalam penelitian dan pengambilan keputusan di lingkungan perguruan tinggi.

Kami berharap buku ini akan menjadi sumber daya berharga bagi dosen dan mahasiswa di seluruh dunia yang menghadapi tantangan statistika dalam pendidikan tinggi. Dengan pengetahuan dan keterampilan yang diperoleh dari buku ini, kami berharap bahwa pembaca akan merasa lebih percaya diri dan siap untuk menghadapi tantangan statistika dalam pendidikan tinggi.

Malang, Oktober 2023

Penulis

DAFTAR ISI

KATA PENGANTAR	I
DAFTAR ISI	III
BAB 1 PENDAHULUAN, SAMPLING, DAN PENGUKURAN	1
A. Pendahuluan tentang Metodologi Statistik	1
B. Statistik Deskriptif dan Statistik Inferensial	7
C. Peran Komputer dan Perangkat Lunak dalam Statistik	10
D. Variabel dan Pengukurannya	12
E. Sampling	18
F. Variabelitas dan Potensi Bias	23
G. Metode Sampel Probabilitas Lainnya	29
H. Ringkasan Bab	34
BAB 2 STATISTIKA DESKRIPTIF	36
A. Mendeskripsikan Data dengan Tabel dan Grafik	36
B. Mendeskripsikan Pusat Data	44
C. Menjelaskan Variabilitas Data	54
E. Statistik Deskriptif Bivariat	68
F. Statistik Sampel dan Parameter Populasi	74
G. Ringkasan Bab	75
BAB 3 DISTRIBUSI PELUANG DAN ESTIMASI	78
A. Pendahuluan Probabilitas	78
B. Distribusi Probabilitas dengan Variabel Diskrit dan Kontinu	79
C. Beberapa Distribusi Probabilitas Diskrit	81
D. Beberapa Distribusi Peluang Kontinu	84
E. Estimasi Titik dan Interval	87
F. Interval Kepercayaan untuk Proporsi	91
G. Interval Kepercayaan untuk Mean	102
H. Pemilihan Ukuran Sampel	114
I. Metode Estimasi: Metode Maksimum Likelihood dan Bootstrap*	123
J. Ringkasan Bab	129
BAB 4 STATISTIKA INFERENSIAL: UJI SIGNIFIKANSI	132
A. Pendahuluan	132
B. Uji Signifikansi untuk Mean	134

C. Uji Signifikansi Proporsi	136
D. Keputusan dan Jenis Kesalahan dalam Tes	137
E. Keterbatasan Uji Signifikansi	139
F. Menentukan P (Kesalahan Tipe II)	141
BAB 5 MEMBANDINGKAN DUA GRUP	143
A. Pendahuluan	143
B. Data Kategorik: Membandingkan Dua Proporsi	146
C. Data Kuantitatif: Membandingkan Mean	147
K. Membandingkan Mean dengan Sampel Terikat (Dependen)	148
BAB 6 ANALISIS HUBUNGAN ANTARA VARIABEL KATEGORIK	150
A. Tabel Kontingensi (<i>Contingency Table</i>)	151
B. Uji Independensi Chi-Kuadrat	155
C. Residu: Mendeteksi Pola Asosiasi	164
D. Mengukur Asosiasi pada Tabel Kontingensi	167
E. Ringkasan Bab	171
BAB 7 REGRESI LINEAR DAN KORELASI	172
A. Hubungan Linear	172
B. Persamaan Prediksi Kuadrat Terkecil	174
C. Model Regresi Linear	178
D. Mengukur Asosiasi Linear: Korelasi	179
E. Inferensi Slope dan Korelasi	182
F. Model Asumsi dan Pelanggaran	183
BAB 8 REGRESI BERGANDA DAN KORELASI	186
A. Model Regresi Berganda	186
B. Korelasi Berganda dan R^2	200
C. Inferensi untuk Parameter Regresi Berganda	206
D. Ringkasan Bab	211
BAB 9 REGRESI DENGAN PREDIKTOR KATEGORIK: ANOVA	213
A. Pemodelan Regresi dengan Variabel <i>Dummy</i> untuk Kategori	214
B. Perbandingan Berganda Rata-Rata	219
C. Membandingkan Beberapa Rerata: Analisis Variansi	224
D. Anova Dua Arah dan Pemodelan Regresi	228
E. Ringkasan Bab	238
DAFTAR PUSTAKA	240
TENTANG PENULIS	241

BAB 1

PENDAHULUAN, SAMPLING, DAN PENGUKURAN

A. Pendahuluan tentang Metodologi Statistik

Beberapa tahun terakhir telah menyaksikan peningkatan dramatis dalam penggunaan metode statistik oleh ilmuwan sosial, baik mereka bekerja di akademisi, pemerintah, atau sektor swasta. Ilmuwan sosial mempelajari topik-topik yang diminati, seperti menganalisis seberapa baik suatu program bekerja atau menyelidiki faktor-faktor yang terkait dengan keyakinan dan opini dari jenis tertentu, dengan menganalisis bukti kuantitatif yang disediakan oleh data. Pertumbuhan Internet dan daya komputasi telah menghasilkan peningkatan dalam jumlah informasi kuantitatif yang tersedia dengan mudah. Pada saat yang sama, evolusi metodologi statistik dan perangkat lunak baru membuat metode-metode baru tersedia yang dapat lebih realistis mengatasi pertanyaan-pertanyaan yang ingin dijawab oleh ilmuwan sosial.

Bab ini memperkenalkan "statistik" sebagai ilmu yang berurusan dengan mendeskripsikan data dan membuat prediksi yang jangkauannya jauh lebih luas daripada sekadar merangkum data yang dikumpulkan. Jadi, mengapa pengetahuan tentang ilmu statistik penting bagi seorang mahasiswa yang sedang belajar untuk menjadi ilmuwan sosial? ¹

MENGAPA MEMPELAJARI ILMU STATISTIK?

Peningkatan penggunaan metode statistik terlihat dalam perubahan konten artikel yang diterbitkan di jurnal penelitian ilmu sosial dan laporan yang disiapkan di pemerintahan dan industri. Melalui sekilas cepat pada isu-isu terbaru dari jurnal seperti *American Political Science Review* dan *American*

¹ Alan Agresti, *Statistical Methods for the Social Sciences*, 5. Auflage (Boston: Pearson, 2018).

Sociological Review, terlihat peran mendasar metodologi statistik dalam penelitian. Contohnya, untuk memahami faktor-faktor yang memiliki dampak terbesar pada kinerja siswa di sekolah atau menyelidiki apa yang memengaruhi keyakinan politik orang atau kualitas perawatan kesehatan mereka atau keputusan mereka tentang pekerjaan dan kehidupan di rumah, para peneliti mengumpulkan informasi dan menganalisanya dengan menggunakan metode statistik. Karena itu, semakin banyak departemen akademis yang mengharuskan mahasiswanya mengambil kursus statistik.

Saat ini, ilmuwan sosial bekerja di berbagai bidang yang menggunakan metode statistik, seperti lembaga pemerintahan, organisasi bisnis, dan fasilitas perawatan kesehatan. Ilmuwan sosial di Lembaga pemerintah yang berurusan dengan kesejahteraan manusia atau isu lingkungan atau kebijakan kesehatan masyarakat umumnya perlu membaca laporan yang berisi argumen statistik, dan mungkin juga menggunakan metode statistik sendiri dalam menyusun laporan tersebut. Beberapa ilmuwan sosial membantu manajer mengevaluasi kinerja karyawan dengan menggunakan patokan kuantitatif dan menentukan faktor-faktor yang membantu memprediksi penjualan produk. Sosiolog medis dan dokter sering harus mengevaluasi rekomendasi dari studi yang berisi evaluasi statistik tentang terapi baru atau cara baru merawat orang tua. Bahkan, sebuah edisi terbaru dari The Journal of the American Medical Association menyatakan bahwa Ujian Masuk Kedokteran telah direvisi untuk memerlukan lebih banyak statistik, karena dokter semakin perlu mampu mengevaluasi secara kuantitatif faktor-faktor yang memengaruhi kesehatan manusia.

Faktanya, semakin banyak pekerjaan bagi ilmuwan sosial yang memerlukan pengetahuan tentang metode statistik sebagai alat kerja dasar. Seperti joke yang beredar, "Apa yang dikatakan sosiolog yang lulus statistik kepada sosiolog yang tidak lulus? 'Saya akan pesan Big Mac, kentang goreng, dan Coca-Cola.' "

Namun, pemahaman tentang ilmu statistik penting bahkan jika Anda tidak pernah menggunakan metode statistik dalam karier Anda sendiri. Seringkali Anda terpapar komunikasi yang mengandung argumen statistik, seperti dalam iklan, laporan berita, kampanye politik, dan survei tentang pendapat tentang isu-isu kontroversial. Ilmu statistik membantu Anda memahami informasi ini dan mengevaluasi argumen mana yang valid dan mana yang tidak valid. Konsep-konsep dari teks ini akan membantu Anda dalam menilai informasi

yang Anda temui dalam kehidupan sehari-hari. Lihat www.youtube.com/user/ThisisStats untuk melihat beberapa testimonial singkat dengan tema bahwa "Statistik bukan hanya tentang analisis data atau angka; ini tentang memahami dunia di sekitar kita. Keberagaman statistik berarti Anda dapat menggunakan pendidikan statistik Anda dan menerapkannya pada hampir semua area yang Anda minati, seperti lingkungan, perawatan kesehatan, hak asasi manusia, olahraga...."

Kami menyadari Anda tidak membaca buku ini dengan harapan menjadi seorang ahli statistik. Selain itu, Anda mungkin mengalami rasa takut matematika dan merasa takut terhadap apa yang akan datang. Harap dipastikan bahwa Anda dapat membaca buku ini dan memahami konsep-konsep utama dan metode-metode statistik dengan pengetahuan matematika yang sedikit. Hanya karena Anda mungkin memiliki kesulitan dalam mata kuliah matematika sebelumnya tidak berarti Anda akan mengalami kerugian di sini. Untuk memahami buku ini, pemikiran logis dan ketekunan lebih penting daripada matematika. Berdasarkan pengalaman kami, faktor paling penting dalam seberapa baik Anda dalam kursus statistik adalah seberapa banyak waktu yang Anda habiskan di kursus tersebut—menghadiri kelas, mengerjakan pekerjaan rumah, membaca dan membaca kembali teks ini, mempelajari catatan kuliah Anda, bekerja bersama dengan teman sekelas Anda, dan meminta bantuan dari profesor atau asisten pengajaran Anda—bukan pengetahuan matematika Anda atau jenis kelamin Anda atau ras Anda atau apakah Anda sekarang merasa takut terhadap statistik.

Jangan merasa frustrasi jika pembelajaran berjalan lambat dan Anda perlu membaca sebuah bab beberapa kali sebelum menjadi masuk akal. Sama seperti Anda tidak akan berharap bisa mengambil kursus tunggal dalam bahasa asing dan mampu berbicara dalam bahasa tersebut dengan lancar, hal yang sama berlaku dengan bahasa ilmu statistik. Setelah Anda menyelesaikan bahkan sebagian dari teks ini, Anda akan lebih memahami cara membuat sense dari informasi statistik.

DATA

Pengumpulan informasi adalah inti dari semua ilmu pengetahuan, menyediakan observasi yang digunakan dalam analisis statistik. Observasi yang dikumpulkan tentang karakteristik yang diminati secara kolektif disebut data.

Sebagai contoh, sebuah studi mungkin melakukan survei terhadap 1000 orang untuk mengamati karakteristik seperti pendapat tentang legalisasi pernikahan sesama jenis, afiliasi partai politik, seberapa sering menghadiri layanan keagamaan, jumlah tahun pendidikan, pendapatan tahunan, status pernikahan, ras, dan jenis kelamin. Data untuk seseorang akan terdiri dari observasi seperti (pendapat = tidak mendukung legalisasi, partai = Republik, ketaatan beragama = sekali seminggu, pendidikan = 14 tahun, pendapatan tahunan dalam kisaran 40-60 ribu dolar, status pernikahan = menikah, ras = putih, jenis kelamin = perempuan). Melihat data dengan cara yang tepat membantu kita memahami bagaimana karakteristik tersebut berhubungan. Kita kemudian bisa menjawab pertanyaan seperti "Apakah orang yang lebih sering menghadiri gereja cenderung lebih tidak mendukung pernikahan sesama jenis?"

Untuk menghasilkan data, ilmuwan sosial menggunakan berbagai metode, termasuk survei menggunakan kuesioner, eksperimen, dan pengamatan langsung terhadap perilaku dalam pengaturan alamiah. Selain itu, ilmuwan sosial sering kali menganalisis data yang sudah direkam untuk tujuan lain, seperti catatan kepolisian, materi sensus, dan berkas rumah sakit. Koleksi data yang sudah ada dan diarsipkan disebut sebagai basis data. Banyak basis data sekarang tersedia di Internet. Basis data penting bagi ilmuwan sosial berisi hasil-hasil sejak 1972 dari General Social Survey.

Contoh 1.1 General Social Survey Setiap dua tahun sekali, National Opinion Research Center di University of Chicago melakukan General Social Survey (GSS). Survei ini terhadap sekitar 2000 orang dewasa menyediakan data tentang opini dan perilaku masyarakat Amerika. Ilmuwan sosial menggunakannya untuk menyelidiki bagaimana orang Amerika dewasa menjawab berbagai pertanyaan, seperti "Apakah Anda percaya ada kehidupan setelah kematian?" "Apakah Anda bersedia membayar harga lebih tinggi untuk melindungi lingkungan?" dan "Apakah Anda pikir seorang anak prasekolah mungkin menderita jika ibunya bekerja?" Survei serupa terjadi di negara-negara lain, seperti General Social Survey yang diadministrasikan oleh Statistics Canada, British Social Attitudes Survey, dan survei Eurobarometer dan European Social Survey untuk negara-negara di Uni Eropa.

Mudah untuk mendapatkan ringkasan data dari basis data GSS. Kami akan menunjukkan, menggunakan pertanyaan yang diajukan dalam satu survei, "Berapa banyak teman baik yang Anda miliki?"

- Buka situs web sda.berkeley.edu/GSS/ di Survey Documentation and Analysis di University of California, Berkeley.
- Klik pada GSS—with NO WEIGHT VARIABLES predefined. Anda akan melihat daftar "variable selection" di sisi kiri yang berurutan dengan isu-isu yang diatasi selama bertahun-tahun, dan menu di sisi kanan untuk memilih karakteristik tertentu yang menarik.
- Nama GSS untuk pertanyaan tentang jumlah teman baik adalah NUMFRIEND. Ketik NUMFRIEND dalam kotak Baris. Klik Run the table.

Situs GSS kemudian akan menghasilkan tabel yang menunjukkan nilai-nilai yang mungkin untuk "jumlah teman baik" dan jumlah orang serta persentase yang memberikan setiap respons yang mungkin. Respons yang paling umum adalah 2 dan 3 (sekitar 16% memberikan masing-masing respons ini).

APA ITU ILMU STATISTIKA?

Anda sudah memiliki gambaran tentang apa yang dimaksud dengan kata statistik. Anda mendengar statistik dikutip tentang acara olahraga (seperti jumlah poin yang dicetak oleh setiap pemain di tim bola basket), statistik tentang ekonomi (seperti pendapatan median atau tingkat pengangguran), dan statistik tentang opini, keyakinan, dan perilaku (seperti persentase siswa yang melakukan minum berlebihan). Dalam arti ini, statistik hanyalah angka yang dihitung dari data. Tetapi buku ini menggunakan statistik dalam arti yang jauh lebih luas—sebagai ilmu yang memberikan kita cara untuk mendapatkan dan menganalisis data.

Statistika

Statistika terdiri dari sekelompok metode untuk mendapatkan dan menganalisis data.

Secara khusus, ilmu statistik menyediakan metode untuk

- Rancangan: Merencanakan bagaimana mengumpulkan data untuk studi penelitian untuk menyelidiki pertanyaan-pertanyaan yang menarik bagi kita.
- Deskripsi: Merangkum data yang diperoleh dalam studi tersebut.

- Inferensi: Membuat prediksi berdasarkan data, untuk membantu kita mengatasi ketidakpastian secara objektif.

Rancangan mengacu pada perencanaan studi agar data yang dihasilkan informatif. Sebagai contoh, untuk survei, rancangan menentukan bagaimana memilih orang-orang untuk diwawancara dan membangun kuesioner yang akan diberikan kepada orang-orang tersebut.

Deskripsi mengacu pada merangkum data, untuk membantu memahami informasi yang disediakan oleh data. Sebagai contoh, analisis jumlah teman baik berdasarkan data GSS mungkin dimulai dengan daftar jumlah yang dilaporkan untuk setiap orang yang disurvei. Data mentah ini kemudian merupakan daftar lengkap observasi, orang per orang. Namun, data ini tidak mudah dipahami. Kita bingung dengan angka-angka tersebut. Untuk presentasi hasil, daripada mencantumkan semua observasi, kita bisa merangkum data dengan grafik atau tabel yang menunjukkan persentase yang melaporkan 1 teman baik, 2 teman baik, 3 teman baik, dan seterusnya. Atau kita bisa melaporkan rata-rata jumlah teman baik, yang sekitar 5, atau respons paling umum, yang adalah 2. Grafik, tabel, dan ringkasan numerik seperti rata-rata dan persentase disebut statistik deskriptif. Kita menggunakan statistik deskriptif untuk mengurangi data menjadi bentuk yang lebih sederhana dan lebih dapat dimengerti tanpa merusak atau kehilangan banyak informasi.

Inferensi mengacu pada penggunaan data untuk membuat prediksi. Misalnya, untuk data GSS tentang jumlah teman baik yang dilaporkan, 6,1% melaporkan hanya memiliki 1 teman baik. Bisakah kita menggunakan informasi ini untuk memprediksi persentase dari 250 juta orang dewasa di Amerika Serikat yang hanya memiliki 1 teman baik? Metode yang disajikan dalam buku ini memungkinkan kita untuk dengan percaya diri memprediksi bahwa persentase tersebut tidak lebih besar dari 8%. Prediksi yang dibuat menggunakan data disebut inferensi statistik.

Deskripsi dan inferensi adalah dua jenis cara menganalisis data. Ilmuwan sosial menggunakan statistik deskriptif dan inferensial untuk menjawab pertanyaan tentang fenomena sosial. Misalnya, "Apakah tersedianya hukuman mati untuk hukuman terkait dengan pengurangan kejahatan kekerasan?" "Apakah kinerja siswa di sekolah bergantung pada jumlah uang yang dihabiskan per siswa, ukuran kelas, atau gaji guru?"

B. Statistik Deskriptif dan Statistik Inferensial

Isi Bagian 1.1 menjelaskan bahwa ilmu statistik terdiri dari metode untuk merancang studi dan menganalisis data yang dikumpulkan dalam studi tersebut. Analisis statistik diklasifikasikan sebagai deskriptif atau inferensial, tergantung pada apakah tujuan utamanya adalah untuk menggambarkan data atau membuat prediksi. Untuk menjelaskan perbedaan ini lebih lanjut, kita kemudian mendefinisikan populasi dan sampel.

POPULASI DAN SAMPEL

Entitas di mana sebuah studi membuat pengamatan disebut subjek sampel untuk studi tersebut. Biasanya subjek adalah orang, seperti dalam General Social Survey, tetapi tidak harus begitu. Sebagai contoh, subjek dalam penelitian sosial mungkin keluarga, sekolah, atau kota. Meskipun kita memperoleh data untuk subjek sampel, minat utama kita adalah pada populasi yang diwakili oleh sampel tersebut.

Populasi adalah himpunan total subjek yang diminati dalam sebuah studi. Sampel adalah subset dari populasi di mana studi mengumpulkan data.

Pada General Social Survey tahun 2014, sampel adalah 2538 warga Amerika dewasa yang berpartisipasi dalam survei tersebut. Populasi adalah semua warga Amerika dewasa pada saat itu—sekitar 250 juta orang. Satu orang diambil sampel untuk sekitar setiap 100.000 orang dalam populasi.

Tujuan dari setiap studi adalah untuk mempelajari tentang populasi. Tetapi hampir selalu diperlukan, dan lebih praktis, untuk mengamati hanya sampel dari populasi tersebut. Sebagai contoh, organisasi survei seperti Gallup Poll biasanya memilih sampel sekitar 1000–2000 orang Amerika untuk mengumpulkan informasi tentang opini dan keyakinan populasi seluruh orang Amerika.

Statistik deskriptif merangkum informasi dalam kumpulan data.

Statistik deskriptif terdiri dari grafik, tabel, dan angka seperti rata-rata dan persentase. Statistik deskriptif mengurangi data menjadi bentuk yang lebih sederhana dan lebih dapat dimengerti tanpa merusak atau kehilangan banyak informasi.

Meskipun data biasanya hanya tersedia untuk sampel, statistik deskriptif juga berguna ketika data tersedia untuk seluruh populasi, seperti dalam sensus. Sebagai perbandingan, statistik inferensial digunakan ketika data hanya tersedia untuk sampel, tetapi kita ingin membuat prediksi tentang seluruh populasi.

Statistika inferensial memberikan prediksi tentang suatu populasi, berdasarkan data dari sampel populasi tersebut.

Contoh 1.2 Berapa Banyak Orang yang Percaya Surga? Dalam tiga surveinya, General Social Survey bertanya, "Apakah Anda percaya akan surga?" Populasi yang diminati adalah kumpulan semua orang dewasa di Amerika Serikat. Dalam survei terbaru di mana pertanyaan ini diajukan, 85% dari 1326 subjek sampel menjawab ya. Ini adalah statistik deskriptif. Namun, kita tidak hanya tertarik pada 1326 orang itu, tetapi pada seluruh populasi semua orang dewasa di Amerika Serikat.

Statistik inferensial menggunakan data sampel untuk membuat prediksi tentang seluruh populasi. Metode inferensial yang disajikan dalam Bab 5 memprediksi bahwa persentase populasi yang percaya akan surga berada antara 83% dan 87%. Artinya, nilai sampel 85% memiliki "margin of error" sebesar 2%. Meskipun ukuran sampel sangat kecil dibandingkan dengan ukuran populasi, kita dapat menyimpulkan bahwa sebagian besar populasi percaya akan surga.

Analisis statistik inferensial dapat memprediksi karakteristik populasi dengan baik dengan memilih sampel yang relatif kecil dibandingkan dengan ukuran populasi. Itulah mengapa banyak jajak pendapat hanya mengambil sampel sekitar seribu orang, bahkan jika populasi memiliki jutaan orang. Dalam buku ini, kita akan melihat mengapa hal ini berhasil.

Dalam beberapa tahun terakhir, ilmuwan sosial semakin mengakui kekuatan metode statistik inferensial. Penyajian metode-metode ini membentuk sebagian besar dari buku teks ini pada beberapa berikutnya.

PARAMETER DAN STATISTIK

Statistik deskriptif adalah ringkasan numerik dari data sampel. Ringkasan numerik yang sesuai untuk populasi disebut parameter.

Parameter adalah ringkasan numerik dari populasi.

Contoh 1.2 memperkirakan persentase orang Amerika yang percaya akan surga. Parameter adalah persentase populasi yang percaya akan surga. Nilainya tidak diketahui. Inferensi tentang parameter ini didasarkan pada statistik deskriptif—persentase dari 1326 subjek yang diwawancarai dalam survei yang menjawab ya, yaitu 85%.

Dalam praktiknya, minat utama kita adalah pada nilai-nilai parameter, bukan hanya nilai-nilai statistik untuk sampel tertentu yang dipilih. Sebagai contoh, ketika melihat hasil jajak pendapat sebelum pemilihan, kita lebih tertarik pada persentase populasi yang mendukung berbagai kandidat daripada persentase sampel untuk orang yang diwawancarai. Sampel dan statistik yang menggambarkannya penting hanya sejauh mereka membantu kita membuat inferensi tentang parameter populasi yang tidak diketahui.

Aspek penting dari inferensi statistik melibatkan pelaporan ketepatan yang mungkin dari statistik sampel yang memperkirakan parameter populasi. Untuk Contoh 1.2 tentang kepercayaan kepada surga, metode statistik inferensial memprediksi seberapa dekat nilai sampel 85% kemungkinan akan terhadap persentase populasi yang tidak diketahui yang percaya akan surga. Margin of error yang dilaporkan adalah 2%.

Ketika data ada untuk seluruh populasi, seperti dalam sensus, mungkin untuk menemukan nilai-nilai parameter yang diminati. Kemudian, tidak perlu menggunakan metode statistik inferensial.

MENDEFINISIKAN POPULASI: SEBENARNYA DAN KONSEPTUAL

Biasanya populasi yang di mana inferensi diterapkan adalah himpunan subjek yang sebenarnya, seperti semua penduduk dewasa di Amerika Serikat. Terkadang, meskipun, generalisasi merujuk pada populasi konseptual—yang sebenarnya tidak ada tetapi bersifat hipotetis.

Sebagai contoh, misalkan sebuah tim peneliti medis menyelidiki obat yang baru diusulkan untuk mengobati kanker paru-paru dengan melakukan studi di beberapa pusat medis. Penelitian medis seperti itu disebut uji klinis. Kondisi yang dibandingkan dalam uji klinis atau eksperimen lain disebut perlakuan. Statistik deskriptif dasar membandingkan pasien kanker paru-paru yang diberi

perlakuan baru dengan pasien kanker paru-paru lainnya yang malah menerima perlakuan standar, dengan menggunakan persentase yang merespons positif terhadap kedua perlakuan tersebut. Dalam menerapkan metode statistik inferensial, para peneliti idealnya ingin inferensi merujuk pada populasi konseptual dari semua orang yang menderita kanker paru-paru sekarang atau suatu saat di masa depan.

C. Peran Komputer dan Perangkat Lunak dalam Statistik

Seiring waktu, perangkat lunak yang kuat dan mudah digunakan telah dikembangkan untuk mengimplementasikan metode statistik. Perangkat lunak ini memberikan manfaat besar bagi penggunaan statistik.

PERANGKAT LUNAK STATISTIK

Paket perangkat lunak statistik meliputi R, SPSS,¹ SAS,² dan Stata. Lampiran A menjelaskan cara menggunakannya, yang diatur berdasarkan bab. Anda dapat merujuk ke Lampiran A untuk perangkat lunak yang digunakan dalam kursus Anda saat Anda membaca setiap bab, untuk mempelajari bagaimana mengimplementasikan analisis dari bab tersebut. Jauh lebih mudah menerapkan metode statistik menggunakan perangkat lunak daripada melakukan perhitungan manual. Selain itu, banyak metode yang disajikan dalam teks ini terlalu kompleks untuk dilakukan secara manual atau dengan kalkulator manual. Perangkat lunak meringankan kita dari tugas komputasi dan membantu kita fokus pada aplikasi yang tepat dan interpretasi dari metode statistik tersebut.

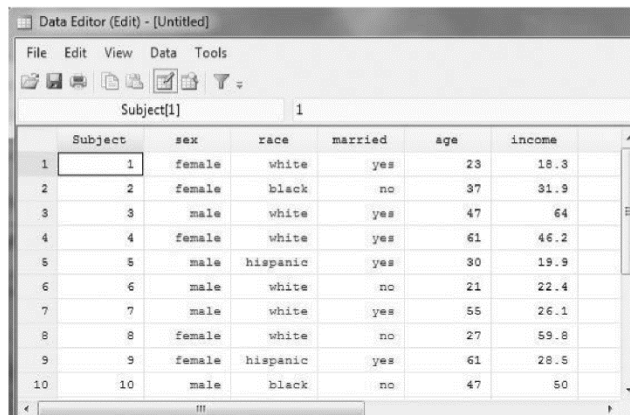
Banyak contoh dalam teks ini juga menunjukkan output yang diperoleh dengan menggunakan perangkat lunak statistik. Salah satu tujuan dari buku teks ini adalah untuk mengajari Anda apa yang harus dicari dalam output dan bagaimana cara menginterpretasikannya. Pengetahuan tentang pemrograman komputer tidak diperlukan untuk menggunakan perangkat lunak statistik.

BERKAS DATA

Perangkat lunak statistik menganalisis data yang diorganisir dalam bentuk lembar kerja berupa berkas data:

- Setiap baris berisi pengamatan untuk subjek tertentu (misalnya, orang) dalam sampel.
- Setiap kolom berisi pengamatan untuk karakteristik tertentu.

Gambar 1.1 menunjukkan contoh berkas data, dalam bentuk jendela untuk mengedit data menggunakan perangkat lunak Stata. Ini menampilkan data untuk 10 subjek pertama dalam sampel, untuk karakteristik jenis kelamin, kelompok ras, status pernikahan, usia, dan pendapatan tahunan (dalam ribuan dolar). Beberapa data bersifat numerik, dan beberapa terdiri dari label. Bab 2 memperkenalkan jenis data untuk berkas data.



	Subject	sex	race	married	age	income
1	1	female	white	yes	23	18.3
2	2	female	black	no	37	31.9
3	3	male	white	yes	47	64
4	4	female	white	yes	61	46.2
5	5	male	hispanic	yes	30	19.9
6	6	male	white	no	21	22.4
7	7	male	white	yes	55	26.1
8	8	female	white	no	27	59.8
9	9	female	hispanic	yes	61	28.5
10	10	male	black	no	47	50

GAMBAR 1.1 CONTOH SEBAGIAN BERKAS DATA STATA

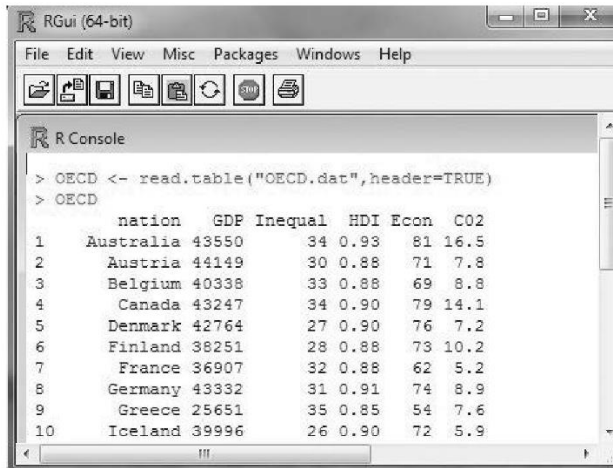
R adalah paket perangkat lunak yang semakin populer, sebagian karena dapat diunduh secara gratis di www.r-project.org. Gambar 1.2 menunjukkan sebagian dari sesi R untuk memuat berkas data yang disebut OECD.dat dari direktori PC dan menampilkannya.

PENGUNAAN DAN PENGGUNAAN TIDAK BENAR PERANGKAT LUNAK STATISTIK

Catatan penting: Kemudahan akses ke metode statistik menggunakan perangkat lunak memiliki risiko serta manfaat. Sangat mudah menerapkan metode yang tidak tepat. Komputer melakukan analisis yang diminta tanpa memperhatikan apakah asumsi-asumsi yang diperlukan untuk penggunaan yang benar dipenuhi.

Analisis yang tidak benar terjadi ketika peneliti tidak cukup waktu untuk memahami metode statistik, asumsi-asumsi yang diperlukan untuk penggunaannya, atau kelayakan metode tersebut untuk masalah tertentu. Penting untuk memahami metodenya sebelum menggunakannya. Hanya dengan mengetahui cara menggunakan perangkat lunak statistik tidak menjamin analisis yang benar. Anda akan memerlukan latar belakang yang baik dalam statistik untuk memahami metode mana yang harus dipilih, opsi

mana yang harus dipilih dalam metode tersebut, dan bagaimana membuat kesimpulan yang valid dari outputnya. Tujuan dari teks ini adalah memberi Anda latar belakang ini.



GAMBAR 1.2 CONTOH SEBAGIAN SESI R UNTUK MEMUAT DAN MENAMPILKAN DATA.

D. Variabel dan Pengukurannya

Metode statistik membantu kita menentukan faktor-faktor yang menjelaskan variasi di antara subjek-subjek. Sebagai contoh, variasi terjadi dari satu mahasiswa ke mahasiswa lain dalam indeks prestasi kumulatif (IPK) mereka di perguruan tinggi. Apa yang bertanggung jawab atas variasi tersebut? Cara mahasiswa-mahasiswa tersebut bervariasi dalam seberapa banyak mereka belajar per minggu? Seberapa sering mereka menonton TV per hari? IQ mereka? Skor ujian masuk perguruan tinggi mereka? IPK SMA mereka?

VARIABEL

Setiap karakteristik yang dapat kita ukur untuk setiap subjek disebut variabel. Nama tersebut mencerminkan bahwa nilai-nilai karakteristik tersebut bervariasi di antara subjek-subjek.

Variabel adalah karakteristik yang dapat bervariasi dalam nilai di antara subjek-subjek dalam sampel atau populasi.

Subjek-subjek yang berbeda mungkin memiliki nilai-nilai yang berbeda dari suatu variabel. Contoh variabel adalah pendapatan tahun lalu, jumlah saudara kandung, apakah bekerja, dan jenis kelamin. Nilai-nilai yang dapat diambil oleh variabel membentuk skala pengukuran. Misalnya, untuk jenis kelamin, skala pengukurannya terdiri dari dua label, yaitu (perempuan, laki-laki). Untuk jumlah saudara kandung, skala pengukurannya adalah (0, 1, 2, 3, 4, dan seterusnya).

Metode statistik yang valid untuk suatu variabel bergantung pada skala pengukurannya. Kami memperlakukan variabel yang bernilai numerik seperti pendapatan tahunan dengan cara yang berbeda dari variabel dengan skala pengukuran yang terdiri dari kategori, seperti (ya, tidak) untuk apakah bekerja. Selanjutnya, kami akan menyajikan cara untuk mengklasifikasikan variabel. Tipe pertama mengacu pada apakah skala pengukuran tersebut terdiri dari kategori atau angka. Tipe lainnya mengacu pada jumlah tingkat dalam skala tersebut.

VARIABEL KUANTITATIF DAN VARIABEL KATEGORIKAL

Variabel disebut kuantitatif ketika skala pengukurannya memiliki nilai numerik yang mewakili magnitudo yang berbeda dari variabel tersebut. Contoh variabel kuantitatif adalah pendapatan tahunan subjek, jumlah saudara kandung, usia, dan jumlah tahun pendidikan yang diselesaikan.

Variabel disebut kategorikal ketika skala pengukurannya adalah himpunan kategori. Sebagai contoh, status pernikahan, dengan kategori (lajang, menikah, bercerai, duda/janda), adalah kategorikal. Untuk penduduk Kanada, provinsi tempat tinggal adalah kategorikal, dengan kategori (Alberta, British Columbia, dan seterusnya). Variabel kategorikal lainnya adalah apakah bekerja (ya, tidak), tujuan berbelanja pakaian utama (mall lokal, pusat kota lokal, internet, lainnya), jenis musik favorit (klasik, country, folk, jazz, rock), afiliasi agama (Kekristenan, Islam, Hinduisme, Buddhisme, Yahudi, lainnya, tidak ada), dan preferensi partai politik.

Untuk variabel kategorikal, kategori-kategori yang berbeda memiliki kualitas yang berbeda, bukan magnitudo numerik. Variabel kategorikal sering disebut kualitatif. Kami membedakan antara variabel kategorikal dan kuantitatif karena metode statistik yang berbeda berlaku untuk setiap tipe.