SEKOLAH PASCASARJANA
UNIVERSITAS SAHID JAKARTA

SURAT TUGAS

Nomor : 1176/USJ-17/N-50/2022

Pertimbangan : 1. Bahwa dalam rangka pelaksanaan Tri Dharma Perguruan Tinggi untuk itu Dosen Tetap Sekolah Pascasarjana Program Studi Doktor Ilmu Komunikasi Universitas Sahid Jakarta tersebut dibawah ini untuk **Melakukan Pengembangan Materi Bahan Kuliah**

2. Untuk itu perlu diterbitkan Surat Tugas

Dasar : Kepentingan Dinas

MENUGASKAN

Kepada : **Morissan. SH., MA., Ph.D**

Untuk : 1. **Melakukan Pengembangan Materi Bahan Kuliah**

2. Selesai melaksanakan tugas agar memberikan laporan kepada Sekolah Pascasarjana Usahid Jakarta

Selesai

Jakarta, 01 September 2022
Sekolah Pascasarjana Usahid Jakarta
Direktur

Dr. Marlinda Irwanti. P. M.Si

MODUL KULIAH 1

STATISTIK SOSIAL TERAPAN

POPULASI & SAMPEL

Morissan, PhD

1.1 Pengertian populasi dan sampel

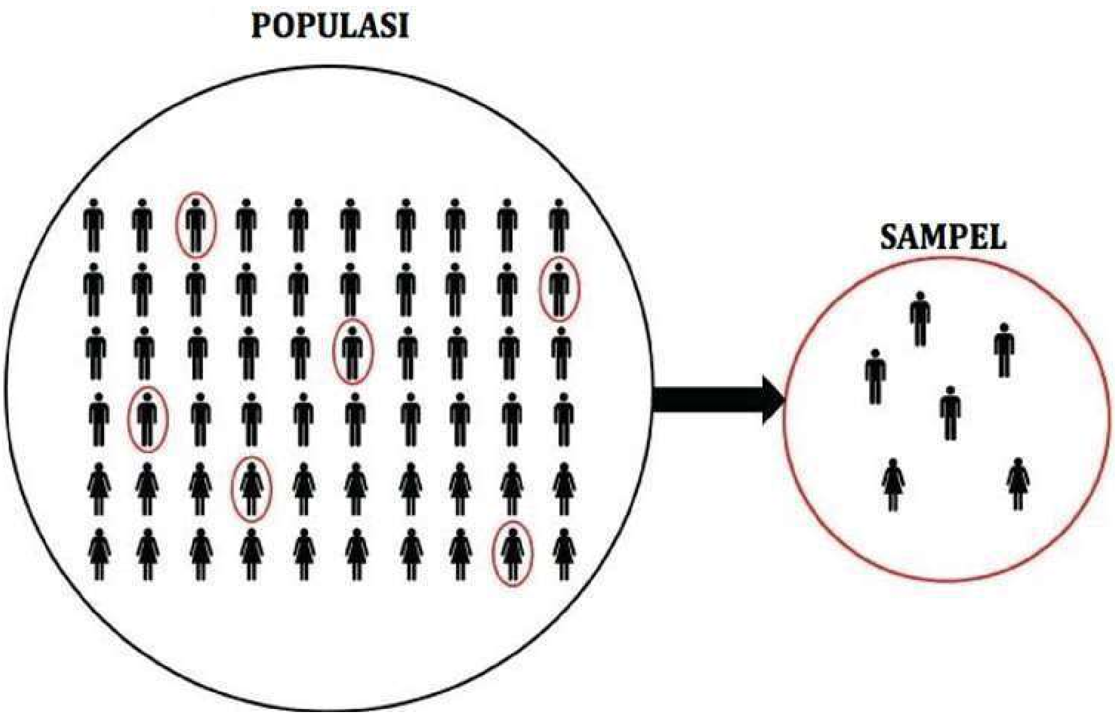
Setiap penelitian selalu diawali dengan pertanyaan mengenai satu atau beberapa kelompok individu atau objek tertentu. Salah satu tujuan penelitian adalah menjelaskan sifat populasi. Secara bahasa, populasi diartikan sebagai sejumlah orang atau hewan yang tinggal di suatu tempat (Merriam-Webster). Pengertian bahasa ini menunjukkan bahwa populasi dibatasi oleh suatu wilayah geografis tertentu, mulai dari suatu wilayah yang sangat luas, misalnya dunia, suatu benua atau satu negara hingga desa atau kampung.

Gravetter dan Wallnau (2013) mendefinisikan populasi sebagai *the set of all the individuals of interest in a particular study*. Hal ini berarti populasi adalah seluruh individu yang hendak diteliti. Namun kata 'individu' pada definisi tersebut tidak boleh hanya diartikan sebagai manusia saja. Anggota populasi dapat berupa manusia (individu, subjek), misalnya populasi mahasiswa di suatu perguruan tinggi; atau bukan manusia (objek), misalnya, populasi tikus, populasi perusahaan, hingga populasi komponen otomotif yang dihasilkan suatu pabrik.

Karena populasi cenderung berjumlah sangat besar maka sangat tidak mudah bagi peneliti untuk meneliti setiap individu satu per satu. Kita dapat meneliti setiap anggota populasi untuk mengetahui sifat populasi bersangkutan. Proses meneliti setiap anggota populasi ini dinamakan **sensus**, dan hasilnya disebut dengan **parameter**. Namun demikian tidaklah mudah melakukan sensus karena membutuhkan waktu dan biaya besar. Karena alasan ini peneliti harus mengambil sampel. **Sampel** adalah sejumlah individu yang dipilih dari populasi, dan dimaksudkan untuk mewakili populasi.

Sampel merupakan bagian dari populasi yang mewakili keseluruhan anggota populasi. Sampel yang baik memiliki sifat representatif terhadap populasi. Suatu sampel yang tidak representatif terhadap setiap anggota populasi, berapapun ukuran sampel itu, tidak dapat digeneralisir terhadap populasi.

Proses pemilihan sampel dapat dijelaskan dengan menggunakan dua lingkaran besar dan lingkaran kecil sebagaimana terlihat pada gambar 1.1. Suatu populasi diwakili oleh lingkaran yang lebih besar. Jika sensus berfungsi menguji atau mengukur setiap elemen pada populasi, maka sampel mengukur atau menguji bagian dari populasi. Pada gambar 1.1, sampel dipilih dari setiap bagian populasi.



Gambar 1.1: Penarikan sampel dari populasi

Jika suatu sampel dipilih berdasarkan suatu panduan tertentu sehingga bersifat representatif terhadap populasi maka data yang diperoleh dari sampel tersebut dapat digeneralisir terhadap populasi. Data yang diperoleh dari sampel disebut dengan **statistik**. Namun demikian, generalisasi data yang diperoleh dari sampel harus dilakukan dengan cermat dan hati-hati karena adanya kesalahan (*error*) yang melekat pada setiap penelitian.

Pertanyaan yang sering kali muncul adalah, “Bagaimana mungkin sampel yang terdiri dari beberapa ratus orang bisa mewakili pendapat jutaan orang?”. Ilmu statistik dapat membuktikan bahwa jika pemilihan sampel dilakukan dengan benar, maka ukuran sampel kecil sudah cukup mewakili pandangan jutaan orang. Bagian paling penting dari setiap prosedur penarikan sampel adalah menghindari bias. Artinya, setiap anggota populasi harus memiliki peluang yang sama untuk dapat terpilih sebagai sampel.

1.2 Metode Penarikan Sampel

Pemilihan sampel merupakan bagian yang sangat penting dari semua penelitian, namun kesalahpahaman sering kali terjadi mengenai sampel ini, khususnya diantara peneliti pemula, atau mereka yang tidak tahu menahu mengenai penelitian. Pertanyaan yang sering kali muncul adalah bagaimana mungkin suatu sampel yang terdiri sejumlah kecil responden bisa mewakili pendapat masyarakat yang berjumlah jutaan orang. Jika anda seorang peneliti pemula, maka camkan ini: jika pemilihan sampel dilakukan dengan benar, maka walaupun jumlah sampel anda hanya terdiri atas beberapa ratus orang, maka jumlah itu sudah cukup mewakili pandangan populasi yang terdiri atas jutaan orang.

Bagian paling penting dari setiap prosedur penarikan sampel adalah menghindari bias, apapun jenisnya. Artinya, setiap responden harus memiliki peluang yang sama untuk dapat terpilih sebagai sampel. Dengan kata lain, rancangan sampel kita harus bebas dari bias.

Metode penarikan sampel, atau disebut juga dengan prosedur sampling (*sampling procedures*) pada umumnya terbagi atas dua bagian besar yaitu: sampel probabilitas dan sampel non-probabilitas [1]

1) **Sampel probabilitas** atau sampling probabilitas (*probability sampling*). Teknik penarikan sampel probabilitas dilakukan dengan menggunakan panduan matematis berdasarkan teori kemungkinan (*probability theory*) dimana peluang setiap unit untuk terpilih sebagai sampel telah dapat diketahui. Teknik penarikan sampel probabilitas dilakukan dengan cara memilih atau menarik sampel secara acak (*random*) dari suatu daftar yang berisi seluruh nama anggota populasi yang tengah diambil sampelnya.

[1] Lihat Earl Babbie, *The Basic of Social Research*, 4th Edition, Wadsworth, 2008, hal 203. Lihat juga Wimmer D, Roger dan Joseph R. Dominick, *Mass Media Research,* Wadsworth, 2011, hal 89.

Contoh sederhana teknik acak atau *random* ini adalah kocok arisan. Peserta arisan yang ingin menentukan siapa yang akan menerima uang arisan biasanya menuliskan nama setiap anggota arisan di secarik kertas kecil, menggulungnya, memasukkannya di gelas yang ditutup kertas bolong, dan mengocoknya dan kemudian megeluarkan satu gulungan kertas kecil untuk menentukan penerima uang arisan. Anda dapat menggunakan teknik lain dengan prinsip yang sama yaitu memberikan setiap anggota peluang yang sama untuk terpilih. Berdasarkan teori probabilitas atau teori kemungkinan maka peluang setiap anggota populasi untuk terpilih sebagai anggota sampel dapat diketahui. Teknik sampel probabilitas dinilai sebagai metode yang paling unggul dalam memilih sampel karena sifatnya yang mewakili populasi (representatif), dan hasil penelitian dapat digeneralisir terhadap seluruh populasi.

2) **Sampel nonprobabilitas.** Sampel probabilitas sering kali sulit dilaksanakan atau tidak sesuai untuk situasi penelitian tertentu karena, misal, tidak tersedianya daftar nama seluruh anggota populasi, atau jika daftar nama tersedia, tetapi tidak tersedia cara untuk dapat menghubungi sampel yang terpilih. Misal, jika tidak tersedia alamat atau nomer telepon yang dapat dihubungi, atau sebagian besar responden tidak tinggal pada daftar alamat yang tersedia.

Sampel nonprobabilitas atau sampling nonprobabilitas (*non-probability sampling*) merupakan teknik penarikan sampel yang tidak mengikuti panduan probabilitas matematis. Namun demikian, karakteristik paling penting yang membedakan kedua tipe sampel adalah bahwa: a) sampling probabilitas menuntut adanya suatu daftar nama seluruh anggota populasi; b) sampling probabilitas memungkinkan peneliti untuk menghitung jumlah kesalahan sampling (*sampling error*) pada suatu penelitian, sedangkan sampling nonprobabilitas tidak.

1.3 Sampel probabilitas

Sampling probabilitas mengacu pada pemilihan sampel dari suatu populasi yang didasarkan pada prinsip pengacakan yang berarti pemilihan dilakukan secara acak atau kebetulan dari suatu daftar nama anggota populasi yang hendak ditarik sampelnya. Sampling probabilitas adalah teknik di mana peneliti memilih sampel dari populasi yang lebih besar dengan menggunakan metode berdasarkan teori probabilitas. Agar seorang peserta dianggap sebagai sampel probabilitas, ia harus dipilih menggunakan pemilihan acak. Metode statistik ini digunakan untuk memilih sampel dari suatu populasi sedemikian rupa sehingga setiap anggota populasi memiliki peluang terpilih yang diketahui dan bukan nol. Persyaratan paling kritis dari sampling probabilitas adalah bahwa setiap orang dalam populasi memiliki peluang yang diketahui dan sama untuk terpilih. Sampel probabilitas memiliki setidaknya tujuh teknik yaitu:

1) Sampel acak sederhana (Simple random sampling)
2) Sampel acak sistematis (Systematic random sampling)
3) Sampel stratifikasi (Stratified sampling)
4) Sampel klaster multitahap (Multistage cluster sampling)
5) Sampel klaster stratifikasi multitahap (Stratified multistage cluster sampling)
6) PPS Sampling
7) Bobot Sampling

1.3.1 Sampel Acak Sederhana

Tipe sampling probabilitas yang paling dasar adalah sampel random sederhana (simple random sample) dimana setiap individu (subjek), elemen, peristiwa, atau unit dalam populasi memiliki peluang yang sama untuk terpilih sebagai anggota sampel. Untuk menghasilkan suatu sampel random sederhana, peneliti dapat menggunakan dua teknik berikut:

a) Menggunakan tabel angka acak.
b) Menggunakan random number generator (aplikasi)

a) Menggunakan tabel angka acak.

Peneliti seringkali menggunakan suatu tabel nilai acak (*random*) untuk menghasilkan suatu sampel random sederhana (Lihat Gambar 1). Tabel angka acak adalah rangkaian angka (0 sampai 9) yang disusun secara acak yang biasanya terdiri dari lima digit, disusun dalam baris dan kolom, agar mudah dibaca. Tabel angka acak telah digunakan dalam statistik untuk memilih sampel secara acak. Cara ini dipandang jauh lebih efektif daripada memilih sampel acak secara manual (misalnya mengocok angka, memilih suatu angka dengan menutup mata, mengocok dadu dll).

Misal, Asumsikan Anda ingin menarik sampel sebanyak lima orang (n = 5) dari suatu populasi yang terdiri dari 200 siswa. Setiap siswa diberi nomor dari 1 sampai 200. Karena ukuran populasi adalah 200 yang berarti memiliki tiga digit maka kita akan menggunakan tiga digit pertama dari angka yang tercantum dalam tabel angka acak. Tanpa melihat ke tabel, arahkan pensil atau pena anda ke suatu titik awal di tabel angka acak. Asumsikan kita mendarat di 46941 (kolom ke-3, entri ke-2).

Lokasi ini memberikan tiga digit pertama yaitu 469. Pilihan ini terlalu besar (>200), jadi kita memilih angka berikutnya (misalnya kita begrgerak ke bawah) di kolom tersebut. Ingatlah bahwa kita mencari bilangan yang tiga digit pertamanya adalah dari 001 hingga 200 (mewakili siswa).

Pilihan kedua memberikan tiga angka pertama menjadi 486, juga terlalu besar. Lanjutkan ke bawah kolom hingga Anda menemukan 5 angka yang tiga digit pertamanya kurang dari atau sama dengan 200.

Dalam contoh ini, kita bergerak ke bawah dan jika sudah *mentok* pindah ke kolom sebelah kanan dan bergerak ke atas maka kita mendapatkan 154 (15445), 18 (01887), 159 (15934), 161 (16194), 187 (18792).

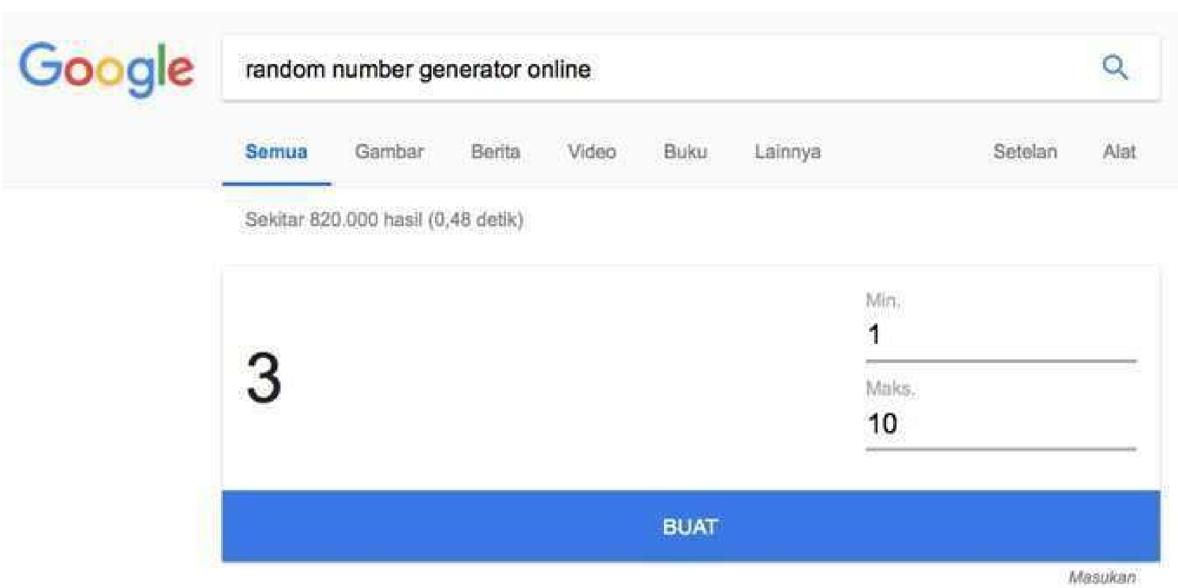
a) Menggunakan generator angka acak (aplikasi)

Saat ini, tabel angka acak sudah jarang digunakan dan telah digantikan oleh generator angka acak yang lebih praktis. Google menyediakan fasilitas ini secara gratis. Silahkan ketik "random number generator online" Kita tinggal memasukkan angka satu pada bagian minimal dan jumlah keseluruhan anggota populasi pada bagian maksimal (Lihat Gambar 2). Jika anda ingin menarik sampel sebanyak 100 dari populasi sebanyak

1000 maka isikan angka 1 pada bagian Min. dan angka 1000 pada bagian Maks. dan klik 'buat' sebanyak 100 kali.

13962	70992	65172	28053	02190	83634	66012	70305	66761	88344
43905	46941	72300	11641	43548	30455	07686	31840	03261	89139
00504	48658	38051	59408	16508	82979	92002	63606	41078	86326
61274	57238	47267	35303	29066	02140	60867	39847	50968	96719
43753	21159	16239	50595	62509	61207	86816	29902	23395	72640
83503	51662	21636	68192	84294	38754	84755	34053	94582	29215
36807	71420	35804	44862	23577	79551	42003	58684	09271	68396
19110	55680	18792	41487	16614	83053	00812	16749	45347	88199
82615	86984	93290	87971	60022	35415	20852	02909	99476	45568
05621	26584	36493	63013	68181	57702	49510	75304	38724	15712
06936	37293	55875	71213	83025	46063	74665	12178	10741	58362
84981	60458	16194	92403	80951	80068	47076	23310	74899	87929
66354	88441	96191	04794	14714	64749	43097	83976	83281	72038
49602	94109	36460	62353	00721	66980	82554	90270	12312	56299
78430	72391	96973	70437	97803	78683	04670	70667	58912	21883
33331	51803	15934	75807	46561	80188	78984	29317	27971	16440
62843	84445	56652	91797	45284	25842	96246	73504	21631	81223
19528	15445	77764	33446	41204	70067	33354	70680	66664	75486
16737	01887	50934	43306	75190	86997	56561	79018	34273	25196

Gambar 1: Tabel angka acak



Gambar 2: Generator angka acak

2.3.2 Sampel Acak Sistematis

Jenis atau tipe sampling probabilitas kedua dinamakan dengan sampling acak sederhana (*systematic random sampling*). Penarikan sampel dengan cara ini adalah dengan menentukan suatu bilangan atau angka ke-*n* dimana setiap subjek atau individu ke-*n* pada populasi terpilih sebagai sampel. Cara menghitung *n* adalah dengan membagi jumlah anggota populasi dengan jumlah anggota sampel yang diinginkan. Nilai *n* disebut juga dengan interval sampling (*sampling interval*). Jika suatu populasi memiliki 10.000 anggota, dan peneliti menginginkan sampel yang terdiri dari 1000 anggota maka ia akan memilih setiap elemen ke-10 (10.000/1000) dari daftar populasi untuk menjadi anggota sampel. Untuk memastikan tidak munculnya bias pada diri peneliti dalam menggunakan metode ini, sampel pertama harus dipilih secara acak. Dalam contoh sebelumnya peneliti harus memilih secara acak satu nomer antara satu hingga 10 yang akan menjadi nomer sampel pertama.

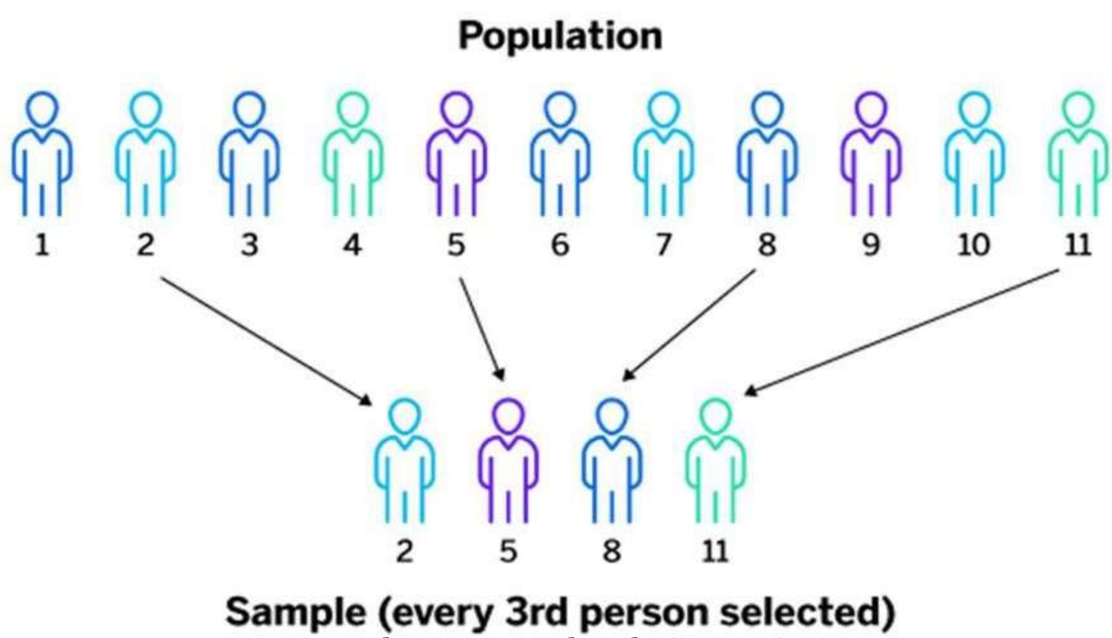
Dengan demikian terdapat dua konsep penting yang sering digunakan dalam penarikan sampel acak sistematis yaitu **interval sampling** yaitu jarak standar atau jarak baku antara elemen yang dipilih pada sampel, dan **sampling rasio** (*sampling ratio*) yang merupakan proporsi dari elemen pada populasi yang dipilih sebagai sampel.

$$\text{Interval sampling} = \frac{\text{Jumlah anggota populasi}}{\text{Jumlah anggota sampel}}$$

$$\text{Rasio sampling} = \frac{\text{Jumlah anggota sampel}}{\text{Jumlah anggota populasi}}$$

Misal, untuk memperoleh suatu sampel yang terdiri dari 20 orang dari suatu populasi sebanyak 100 orang, peneliti secara acak memilih suatu interval sampling (*sampling interval*) dan suatu nilai awal. Nilai interval yang diperoleh adalah: 100/20 = 5. Rasio sampling pada contoh ini adalah 20/100 = 1/5. Nilai awal ditentukan dengan memilih secara acak antara nomer 1 hingga 5. Jika misalnya angka 1 terpilih sebagai nilai awal maka sampel akan mencakup sampel nomer 6, 11, 16, 21 dan seterusnya. Contoh lain, jika suatu nilai interval adalah 3 dengan nilai awal 2 maka setiap kelipatan tiga pada nomer daftar populasi terpilih sebagai sampel yaitu nomer 2, 5, 8, 11, dan seterusnya (Lihat Gambar 3).

Penggunaan sampel acak sistematis harus dilakukan dengan cermat. Pola susunan tertentu dari daftar anggota populasi yang disebut dengan periodisitas (*periodicity*) dapat membuat penggunaan sampling sistematis menjadi tidak bijaksana. Periodisitas terjadi jika daftar anggota populasi diatur dalam suatu pola siklus yang kebetulan memiliki nilai sama dengan nilai interval sampling. Misal, peneliti menarik sampel dari suatu populasi prajurit militer yang namanya tercantum dalam daftar nama prajurit (roster). Setiap prajurit kesepuluh yang tercantum pada daftar nama dipilih sebagai sampel. Namun daftar nama tersebut disusun mengikuti struktur organisasi militer dimana terdapat kesatuan-kesatuan yang masing-masing terdiri dari 10 orang, dan dipimpin seorang komandan berpangkat sersan. Pada daftar nama prajurit itu, komandan selalu berada pada urutan nomer satu pada kesatuannya, dibawahnya adalah kopral dan prajurit. Sebagai hasilnya, setiap prajurit ke-10 pada daftar nama adalah komandan kesatuan. Hasil penelitian sebagaimana kasus ini sudah pasti menjadi sangat bias sehingga tidak dapat digunakan untuk menggambarkan keseluruhan prajurit militer.



Gambar 3: Sampel acak sistematis
Sumber: Qualtrics (2023). The complete guide to systematic random sampling.

Contoh lain adalah dalam hal penarikan sampel pada suatu apartemen pada suatu gedung. Setiap apartemen biasanya diberi nomer yang menunjukkan tingkat atau nomer apartemen tertentu (misalnya, apartemen nomer 101, 102, 103, 201, 202 dan seterusnya). Terdapat kemungkinan nilai interval sampling yang ditentukan peneliti secara kebetulan sama dengan nomer apartemen tertentu dan kelipatannya. Hal ini dapat menyebabkan sampel yang terpilih hanyalah apartemen yang semuanya berada pada posisi yang sama, misal semua apartemen yang menghadap ke barat atau ke timur, atau semua apartemen yang terletak di dekat tangga atau *lift*. Jika apartemen yang terpilih sebagai sampel memiliki karakteristik lain yang sama, misal tarif sewa yang sama-sama lebih mahal, atau sama-sama dilengkapi fasilitas televisi kabel, sementara apartemen lainnya tidak, maka hal ini akan menghasilkan bias penelitian.

Peneliti harus cermat dalam menggunakan sampel sistematis yang diambil dari suatu daftar. Peneliti harus memahami karakteristik atau sifat suatu daftar. Jika anggota populasi disusun dalam suatu urutan tertentu maka peneliti harus memperhitungkan apakah susunan tersebut akan dapat menimbulkan bias terhadap sampel yang dipilih, jika ternyata terdapat potensi bias maka peneliti harus melakukan langkah pencegahan. Misalnya dengan menggunakan sampel acak sederhana. Namun biasanya sampel acak sistematis dinilai lebih baik dibandingkan dengan sampel acak sederhana dengan beberapa alasan sebagai berikut: 1) pemilihan sampel lebih mudah dilakukan dan lebih akurat dan; 2) lebih murah.

2.3.3 Sampel Stratifikasi

Dua metode penarikan sampel yang telah kita pelajari sebelumnya yaitu sampel acak sederhana dan sampel sistematis memiliki derajat keterwakilan terhadap populasi, dan memungkinkan peneliti untuk memperkirakan tingkat kesalahan yang terjadi. Pada bagian ini kita akan mempelajari metode penarikan sampel yang berbeda dengan sebelumnya yang dinamakan sampel stratifikasi (*stratified sampling*). Metode ini digunakan untuk memperoleh suatu derajat keterwakilan yang lebih besar dengan cara mengurangi kesalahan sampel probabilitas. Untuk memahami metode ini, kita harus kembali sejenak ke teori dasar mengenai distribusi sampel.

Harap dipahami bahwa kesalahan sampling dapat dikurangi karena adanya dua faktor dalam suatu rancangan sampel. Pertama, suatu sampel besar akan mampu menghasilkan suatu kesalahan sampling yang lebih kecil. Kedua, suatu populasi yang homogen akan menghasilkan sampel dengan tingkat kesalahan sampling yang lebih kecil dibandingkan dengan suatu populasi yang heterogen. Jika 99 persen populasi setuju dengan suatu pernyataan maka sangatlah tidak mungkin sampel probabilitas yang dihasilkannya tidak mewakili tingkat dukungan populasi. Jika populasi terbagi dua 50-50 antara yang menerima dan menolak suatu pernyataan, maka tingkat kesalahan sampling akan menjadi jauh lebih besar. Untuk mengatasi kelemahan ini, peneliti menggunakan sampel stratifikasi.

Metode ini digunakan berdasarkan prinsip menarik sampel dari bagian populasi yang homogen, dan bukan memilihnya dari total populasi yang heterogen. Jika peneliti ingin mengetahui sikap mahasiswa pada suatu universitas maka ia harus membagi-bagi atau menstratifikasi populasi mahasiswa di perguruan tinggi tersebut menjadi beberapa kelas, kelompok atau bagian populasi. Misalnya mahasiswa tahun pertama, kedua, ketiga dan seterusnya. Teknik pengelompokkan lainnya adalah membagi-bagi populasi mahasiswa berdasarkan jenis kelamin, atau mengelompokan mahasiswa berdasarkan indeks prestasi, misalnya kelompok mahasiswa dengan IP 0-1, 1-2, 2-3 dan seterusnya. Pengelompokkan berdasarkan tingkat ekonomi, misalnya kelompok mahasiswa dengan latar belakang tingkat ekonomi rendah, menengah atau atas. Pengelompokkan berdasarkan agama (Islam, Kristen, Hindu dll) atau suku bangsa dan seterusnya.

Fungsi terpenting dari stratifikasi sampel adalah membagi-bagi populasi yang heterogen ke dalam kelompok-kelompok yang bersifat lebih homogen dan memilih jumlah anggota yang akan menjadi sampel mewakili setiap kelompok homogen. Kesalahan sampling atau *sampling error* pada sampel yang distratifikasi berdasarkan kelas atau kelompok homogen semacam ini dapat dikurangi hingga nol. Hal ini tidak terjadi pada teknik sampel yang tidak distratifikasi sebagaimana teknik sampel acak sederhana dan acak sistematis yang telah kita pelajari sebelumnya.

Pilihan terhadap variabel stratifikasi akan tergantung pada variabel apa yang tersedia. Jika kita memiliki suatu daftar nama, kita akan dengan mudah memperoleh variabel jenis kelamin berdasarkan nama-nama yang tercantum. Susi pastilah nama wanita, Joni pastilah nama pria. Suatu kebalikan nama Susi ternyata pria sangat jarang terjadi walaupun mungkin saja ada. Daftar nama mahasiswa di perguruan tinggi biasanya disusun berdasarkan angkatan atau jurusan, daftar nama karyawan suatu perusahaan biasanya dibagi berdasarkan departemen atau bagian dimana mereka bekerja. Dokumen pemerintahan biasanya dikelompokkan berdasarkan wilayah geografis. Daftar pemilih pada pemilihan umum biasanya disusun berdasarkan wilayah-wilayah tertentu pada suatu kota atau kabupaten.

Hal yang harus diperhatikan peneliti dalam memilih variabel stratifikasi dari daftar yang tersedia adalah menentukan variabel mana yang paling penting atau paling berpengaruh terhadap variabel penelitian yang hendak diuji atau diteliti. Karena jenis kelamin memiliki hubungan erat dengan banyak variabel penelitian maka variabel stratifikasi ini sering kali digunakan. Untuk meneliti variabel penelitian seperti variabel pilihan politik, pilihan terhadap isi media (TV, surat kabar, majalah dll), pembelian produk dan sebagainya sangat ditentukan pada jenis kelamin seseorang. Pendidikan sebagai salah satu variabel stratifikasi juga memiliki hubungan erat dengan berbagai variabel penelitian. Begitu pula lokasi geografis dalam suatu kota, provinsi, negara akan berhubungan dengan banyak hal. Variabel stratifikasi berdasarkan lokasi geografis pada suatu kota akan menentukan variabel penelitian seperti kelas sosial atau latar belakang etnis seseorang. Misal, warga Jakarta yang tinggal di kawasan Menteng atau Pondok Indah tentu menunjukkan kelas sosial mereka. Kawasan suatu kota juga dapat menunjukkan latar belakang etnis seseorang, misal kawasan Glodok di Jakarta yang didominasi etnis Cina. Bagaimana dengan kawasan-kawasan tertentu di kota anda dimana banyak tinggal masyarakat dari etnis atau suku tertentu?

Terdapat dua metode stratifikasi yang paling banyak digunakan peneliti dalam mengolah daftar seluruh anggota populasi. Pertama, anggota populasi dibagi-bagi ke dalam kelompok-kelompok yang berbeda-beda berdasarkan variabel stratifikasi yang diinginkan. Peneliti kemudian memilih, secara acak atau sistematis, beberapa anggota dari masing-masing kelompok secara proporsional yaitu berdasarkan jumlah atau proporsi relatif anggota kelompok terhadap jumlah keseluruhan anggota populasi. Misal, jika dari seluruh total populasi mahasiswa di suatu universitas, terdapat satu persen mahasiswa tingkat pertama yang memiliki indeks prestasi 4.0, sementara peneliti menginginkan sampel sebanyak 1000 orang, maka peneliti hanya boleh memilih 10 mahasiswa yang mewakili kelompok bersangkutan yaitu kelompok mahasiswa tingkat pertama dengan IP 4.0.

Metode lain adalah mengelompokkan mahasiswa sebagaimana yang telah dijelaskan sebelumnya, dan meletakkan kelompok-kelompok tersebut secara bersama-sama dalam suatu daftar panjang mahasiswa, dimulai dari, misalnya, seluruh mahasiswa pria tingkat pertama dengan IP 4.0 dan berakhir dengan seluruh mahasiswa wanita tingkat akhir dengan IP kurang dari 1.0. Peneliti kemudian menarik sampel sistematis dengan memilih secara acak satu nomer awal dari daftar yang tersedia, dan kemudian menentukan nilai interval yang diinginkan sebagaimana prosedur yang telah kita bahas sebelumnya pada sampel acak sistematis.

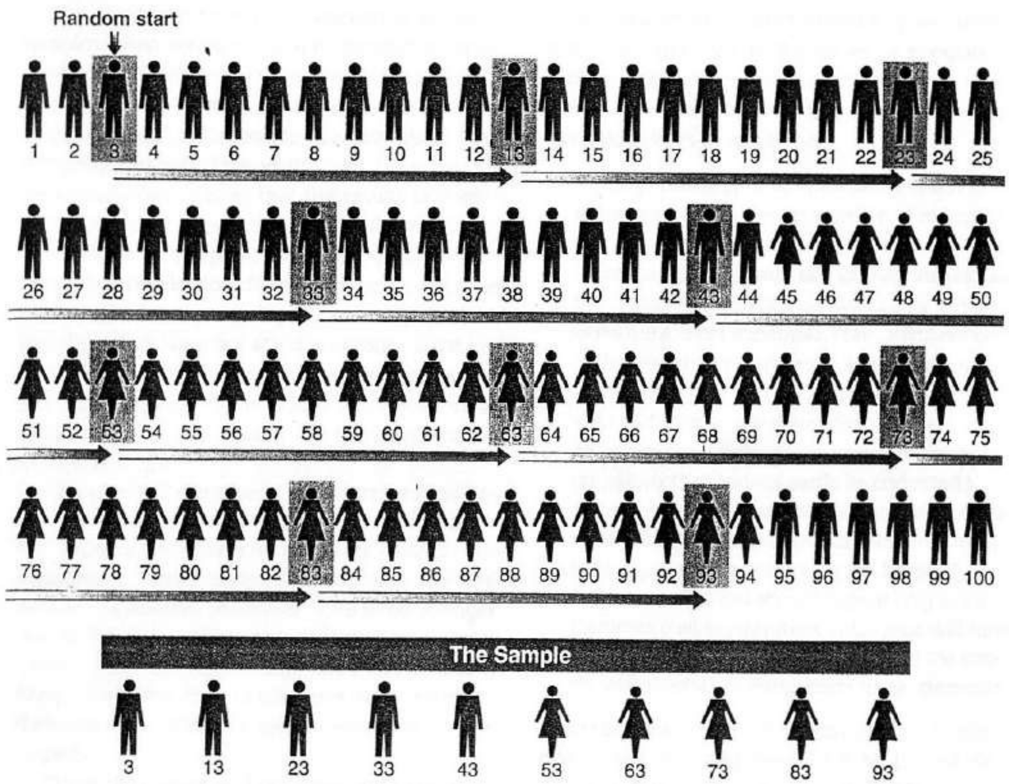
Gambar 5 menunjukkan suatu ilustrasi grafis dari sampling sistematis stratifikasi. Sebagaimana yang dapat kita lihat, seluruh anggota populasi yang berjumlah 100 orang disusun berbaris memanjang berdasarkan jenis kelamin. Dari populasi ini, peneliti menginginkan suatu sampel yang terdiri dari 10 orang, sehingga diperoleh interval sampling 10 (100/10), dan rasio sampling 10/100 = 1/10. Nomer sampel pertama ditentukan dengan cara memilih secara acak salah satu nomer dari angka 1 hingga 10. Jika nomer yang terpilih adalah "3" maka penghitungan awal (*random start*) dimulai dari anggota populasi nomer 3, dan nilai interval 10, maka peneliti menarik setiap mahasiswa ke-10 sehingga menghasilkan daftar anggota sampel dengan nomer 3, 13, 23, 33.., dan seterusnya hingga nomer 93.

Sekali lagi, sampel stratifikasi menghasilkan representasi variabel stratifikasi, hal ini pada gilirannya mendorong representasi variabel lain yang berhubungan. Dengan demikian, secara umum, sampel stratifikasi memberikan peluang lebih besar untuk menghasilkan sampel yang lebih representatif pada beberapa variabel dibandingkan dengan sampel acak sederhana. Walaupun sampel acak sederhana sering kali dipandang lebih sakral, namun dari contoh yang dikemukakan jelaslah bagi kita bahwa sampel stratifikasi seringkali lebih baik.

2.3.4 Sampel Klaster Multi-Tahap

Sejauh ini kita telah mempelajari teknik penarikan sampel yang dilakukan berdasarkan daftar anggota populasi yang telah kita miliki. Namun demikian situasi ideal semacam ini dimana kita memiliki daftar nama dan alamat seluruh anggota populasi tidaklah selalu terjadi, dan terkadang tidak mudah dilakukan. Banyak riset sosial yang menarik tidak dapat dilakukan karena tidak tersedianya daftar seluruh anggota populasi.

Misalnya: daftar anggota populasi penduduk suatu kota, provinsi, dan negara; atau daftar anggota populasi seluruh mahasiswa di Indonesia, seluruh anggota jemaah masjid atau gereja di seluruh Indonesia dan seterusnya.

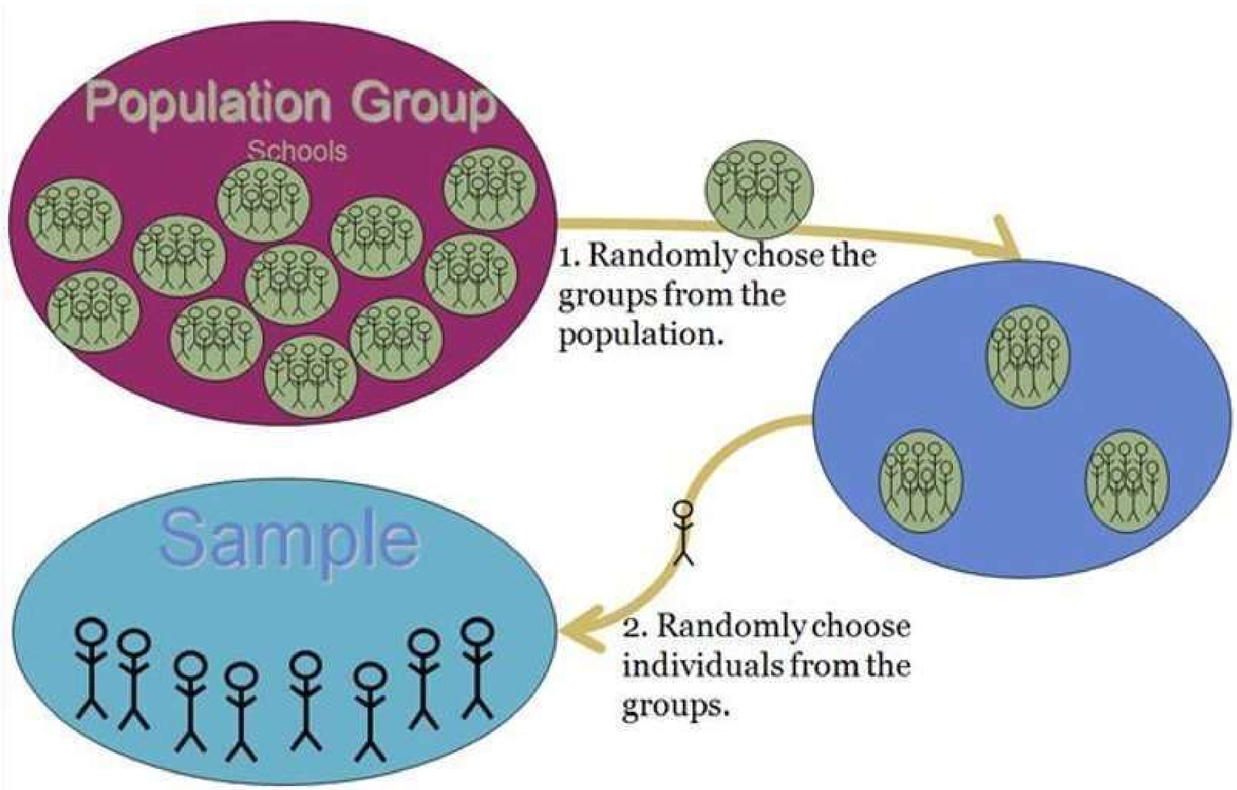


Gambar 4 *:* Sampel Stratifikasi Sistematis

Sumber: Earl Babbie, *The Basic of Social Research* (2010), 12th Edition, Wadsworth, hal 229

Dalam penelitian dimana tidak tersedia daftar nama, dan alamat seluruh anggota populasi dibutuhkan suatu rancangan sampel (*sample design*) yang sedikit lebih kompleks. Rancangan sampel yang disiapkan adalah, pertama, melakukan sampling awal terhadap kelompok-kelompok anggota populasi atau disebut dengan klaster (*cluster*) diikuti dengan pemilihan anggota yang berada pada setiap klaster. Misal, peneliti memilih suatu sampel dari populasi yang terdiri dari seluruh sekolah lanjutan tingkat atas (SLTA) di suatu kota yang tercantum pada daftar SLTA di kota bersangkutan. Selanjutnya, peneliti mencari daftar nama seluruh siswa SLTA terpilih, dan kemudian menarik sampel kedua, yaitu nama-nama siswa dari setiap sekolah terpilih (Lihat Gambar 5).

Sampling klaster semacam ini dapat digunakan jika tidak mungkin atau tidak praktis melakukan pengumpulan data seluruh anggota populasi, misalnya daftar seluruh penduduk suatu kota. Walaupun daftar tunggal seluruh penduduk kota tidak tersedia namun biasanya penduduk tinggal di berbagai blok pemukiman, atau komplek perumahan di suatu kota. Dalam hal ini, peneliti harus melakukan pemilihan sampel tahap pertama yaitu memilih suatu sampel yang terdiri dari blok pemukiman atau komplek perumahan, dan mengumpulkan daftar nama orang-orang yang tinggal pada blok, atau komplek yang terpilih sebagai sampel. Selanjutnya peneliti melakukan penarikan sampel tahap kedua berdasarkan daftar seluruh warga yang tinggal pada blok atau komplek yang terpilih sebagai sampel.



Gambar 5: Sampel klaster multitahap. Gambar menunjukkan peneliti memilih suatu sampel dari populasi (*population group*) yang terdiri dari seluruh sekolah di suatu kota yang tercantum pada daftar sekolah di kota bersangkutan. Selanjutnya, peneliti mencari daftar nama seluruh siswa sekolah terpilih, dan kemudian menarik sampel kedua, yaitu nama-nama siswa dari setiap sekolah terpilih.

Pada suatu rancangan sampel yang lebih kompleks, pemilihan sampel dapat dilakukan lebih dari dua tahap dengan tahapan sebagai berikut: (1) memilih sampel dari daftar blok atau komplek pemukiman; (2) membuat daftar rumah tangga dari setiap blok atau komplek yang terpilih sebagai sampel; (3) memilih sampel rumah tangga; (4) membuat daftar nama seluruh anggota rumah tangga, dan (5) memilih sampel anggota rumah tangga terpilih. Rancangan sampel multi-tahap semacam ini berakhir pada pemilihan sampel individu tetapi tidak memerlukan daftar seluruh individu penduduk kota.

Dengan demikian sampel klaster multi-tahap terdiri dari dua tahapan pekerjaan yaitu: membuat daftar anggota (*listing*) dan memilih sampel. Pertama, daftar unit atau anggota sampling utama dikumpulkan dan kemudian distratifikasi untuk dilakukan sampling. Anggota sampling utama yang terpilih dikumpulkan dan kemudian distratifikasi lagi. Daftar anggota sampling kedua kemudian dipilih lagi untuk menjadi sampel berikutnya, begitu seterusnya.

2.3.5 Sampel klaster stratifikasi multitahap

Sejauh ini kita menggunakan sampling klaster multi-tahap dengan menggunakan sampel acak sederhana yang digunakan pada setiap tahap penarikan sampel. Cara lain yang dapat kita gunakan adalah dengan teknik stratifikasi sebagai upaya peneliti untuk memperbaiki sampel yang diperoleh. Teknik stratifikasi ini mirip dengan sampling satu tahap dimana anggota sampel dipilih dari daftar seluruh anggota populasi. Dari contoh sampel klaster multitahap sebelumnya maka pada peneliti terlebih dahulu

Misal, dalam memilih sampel kelompok pengajian di suatu daerah maka penarikan sampel dengan menggunakan stratifikasi sampel klaster multi-tahap peneliti pertama-tama menstratifikasi daftar populasi sekolah berdasarkan, misalnya, sekolah umum dan sekolah agama (madrasah). Contoh lain menstratifikasi kelompok pengajian berdasarkan beberapa kategori tertentu: wilayah geografis, kelas sosial, suku bangsa, lokasi (kota atau desa), ajaran (NU, Muhammadiyah, Ahmadiyah dll).

Jika populasi kelompok pengajian telah dapat dibagi-bagi atau dikelompokkan berdasarkan kategori relevan tertentu maka peneliti dapat mengambil sampel dari populasi yang telah distratifikasi tersebut, baik dengan menggunakan teknik sampel acak sederhana atau teknik sampel sistematis. Dalam hal ini, peneliti dapat mengambil jumlah tertentu dari anggota masing-masing kelompok, atau stratum, atau menyusun kelompok-kelompok yang telah distratifikasi ke dalam daftar panjang (*list*), dan mengambil sampel dari daftar tersebur secara sistematis. Tujuan utama stratifikasi --mengelompokkan anggota populasi berdasarkan kesamaan karakteristiknya-- adalah untuk menciptakan homogenitas. Stratikasi tentu saja dapat dilakukan pada setiap tahap sampling. Daftar anggota dari setiap klaster terpilih dapat distratifikasi sebelum tahap sampling berikutnya. Jika klaster-klaster dapat dikelompokkan berdasarkan kesamaan karakteristiknya maka kesalahan sampling dapat dikurangi.

2.3.6 PPS Sampling

Pada bagian ini kita akan membahas salah satu teknik dalam sampling klaster yang sering dipraktekkan di negara-negara maju dalam penelitian survei skala luas yang mencakup populasi dalam jumlah besar. Sebelumnya, kita telah membahas mengenai pemilihan sampel klaster secara acak atau sistematis, dan dilanjutkan dengan pemilihan sampel secara acak atau sistematis di dalam masing-masing klaster terpilih. Harap perhatikan bahwa metode seperti ini menghasilkan suatu skema sampling keseluruhan dimana setiap anggota atau elemen populasi memiliki peluang yang sama untuk terpilih sebagai sampel.

Umpamakan kita hendak mengambil sampel rumah tangga di kota A. Jika terdapat 1000 blok pemukiman atau komplek perumahan di kota A dan kita menginginkan suatu sampel yang terdiri dari 100 rumah tangga maka masing-masing blok atau komplek memiliki peluang sebesar 100/1000 atau 1/10 untuk terpilih sebagai anggota sampel. Jika selanjutnya kita memilih satu dari 10 rumah tangga yang terdapat pada masing-masing blok maka masing-masing rumah tangga memiliki peluang sebesar 1/10 untuk terpilih sebagai sampel mewakili blok. Untuk menghitung probabilitas yaitu kemungkinan keseluruhan suatu rumah tangga terpilih, kita hanya mengalikan probablitas dari setiap tahapannya. Masing-masing rumah memiliki peluang 1/10 bagi bloknya untuk terpilih sebagai sampel, dan jika blok suatu rumah tangga terpilih maka terdapat peluang 1/10 bagi rumah tangga di blok bersangkutan terpilih sebagai sampel. Dalam kasus ini, masing-masing rumah tangga memiliki peluang 1/10 x 1/10 = 1/100 untuk terpilih sebagai sampel.

Metode pemilihan sampel semacam ini berpotensi memiliki kesalahan yang disebabkan jumlah rumah tangga pada setiap blok pemukiman dan komplek perumahan tidaklah sama. Misalnya, setengah jumlah penduduk kota A tinggal di pusat kota yang mencakup lahan seluas 100 blok yang penuh dengan gedung pemukiman (apartemen, flat, rumah susun dsb), dan misalnya setengah lagi penduduk kota tinggal menyebar di 900 blok sisanya. Ketika kita memilih sampel tahap pertama sebanyak 100 blok dimana setiap blok memiliki peluang 1/10 maka sangatlah mungkin kita tidak akan mendapatkan seluruh 100 blok yang tinggal di pemukiman padat dimana bermukim setengah penduduk kota. Tidak peduli apapun yang terjadi pada penarikan sampel tahap kedua, sampel akhir rumah tangga yang kita miliki menjadi sangat tidak representatif terhadap penduduk kota.

Manakala sampel klaster menunjukkan perbedaan jumlah yang sangat besar maka rancangan sampling perlu dimodifikasi disesuaikan dengan ukuran masing-masing klaster. Metode penarikan sampel semacam ini dinamakan *probability proportionate to size* (PPS) yang berfungsi mencegah terjadinya masalah semacam ini, dan mampu menghasilkan sampel final yang mana setiap elemen memiliki peluang sama untuk terpilih. Sebagaimana namanya, peluang setiap klaster untuk terpilih ditentukan oleh ukuran besar atau kecilnya klaster bersangkutan. Jadi, suatu blok pemukiman yang memiliki 200 rumah tangga memiliki peluang dua kali lebih besar untuk terpilih sebagai sampel dibandingkan dengan suatu blok yang hanya memiliki 100 rumah tangga. Namun demikian pada setiap klaster ditentukan satu jumlah tetap rumah tangga, misalnya, lima rumah tangga untuk setiap blok yang akan dipilih sebagai sampel. Prosedur ini menyebabkan setiap rumah tangga memiliki peluang yang sama untuk terpilih sebagai sampel secara keseluruhan.

Mari kita lihat pada contoh dua blok pemukiman, misalnya blok A dan blok B, pada suatu kota. Blok A terdiri dari 100 rumah tangga, sedangkan blok B memiliki hanya 10 rumah tangga. Dengan menggunakan metode sampling PPS, kita dapat memberikan peluang 10 kali lebih besar kepada blok A untuk terpilih sebagai sampel dari pada blok B. Dengan demikian, dalam rancangan sampel keseluruhan, blok A memiliki peluang 1/20 untuk terpilih, dan blok B memiliki peluang hanya 1/200.

Jika blok A terpilih, dan kita menentukan untuk mengambil lima rumah tangga pada setiap blok terpilih maka setiap rumah tangga di blok A memiliki peluang 5/100 untuk terpilih sebagai sampel blok. Karena kita dapat mengalikan probabilitas dalam kasus ini maka kita melihat setiap rumah di blok A memiliki peluang keseluruhan untuk terpilih sebesar 1/20 x 5/100 = 5/2000 = 1/400.

Sebaliknya jika blok B yang terpilih maka rumah tangga blok B memiliki peluang lebih baik untuk terpilih yaitu 5/10. Jika nilai probabilitas ini dikalikan dengan probabilitas blok B untuk terpilih sebagai sampel blok yaitu 1/200 maka blok B pada akhirnya memiliki peluang yang sama dengan blok A yaitu 1/200 x 5/10 = 5/2000 = 1/400. Hal ini menunjukkan bahwa PPS merupakan metode yang sangat efisien dan efektif digunakan untuk memilih sampel klaster besar.

2.3.7 Bobot Sampling

Suatu sampel probabilitas disebut mewakili suatu populasi (representatif) jika seluruh elemen pada populasi memiliki peluang yang sama untuk terpilih sebagai sampel. Pada pembahasan sebelumnya mengenai berbagai macam metode atau prosedur sampling kita melihat adanya peluang yang sama bagi setiap elemen populasi untuk terpilih sebagai sampel, walaupun pada akhirnya pengambilan sampel probabilitas merupakan produk yang tidak sepenuhnya berdasarkan probabilitas.

Secara lebih umum, disebut sampel probabilitas jika setiap elemen pada populasi memiliki kemungkinan tidak nol untuk terpilih walaupun elemen atau anggota populasi yang berbeda memiliki kemungkinan yang berbeda. Jika prosedur probabilitas terkontrol telah digunakan maka sampel yang dihasilkan akan representatif terhadap populasi dari mana sampel ditarik jika setiap elemen sampel diberikan bobot yang sama yaitu 1 berarti setiap elemen populasi memiliki peluang yang sama untuk terpilih (*self-weighting sample*).

Terkadang peneliti perlu memberikan bobot berbeda terhadap elemen atau anggota sampel tertentu. Proses memberikan bobot tertentu terhadap elemen sampel tertentu dinamakan dengan pembobotan (*weighting*) yang dilakukan dengan dua cara. Pertama, peneliti mengambil sebagian elemen populasi sebagai sampel secara tidak proporsional untuk memastikan jumlah kasus yang cukup dari setiap sampel untuk dianalisa. Misal, suatu kota memiliki wilayah pinggiran yang memiliki jumlah penduduk seperempat dari total penduduk kota. Dalam hal ini, peneliti tertarik untuk melakukan analisa secara khusus mengenai rumah tangga masyarakat pinggiran ini, dan ia merasa jumlah anggota sampel yang mewakili wilayah tersebut terlalu sedikit (hanya seperempat dari total sampel). Peneliti kemudian meningkatkan jumlah anggota sampel yang mewakili wilayah tersebut menjadi sama dengan anggota sampel dari wilayah lainnya. Rumah tangga yang berada di wilayah pinggiran diberikan peluang lebih besar secara tidak proporsional dari pada rumah tangga di wilayah lainnya.

Kita tidak perlu mencemaskan soal perbedaan sampling ini sepanjang peneliti menganalisa kedua wilayah sampel tersebut secara terpisah, dan bertujuan untuk membandingkan (komparatif). Jika peneliti berkeinginan menggabungkan dua sampel untuk menghasilkan gambaran umum keseluruhan kota maka peneliti harus melakukan sampling secara tidak proporsional ini. Jika n adalah jumlah rumah tangga terpilih dari masing-masing wilayah, maka rumah tangga di wilayah pinggiran memiliki peluang terpilih, yaitu n dibagi dengan seperempat dari total populasi kota. Karena keseluruhan populasi kota, dan ukuran sampel adalah sama untuk kedua wilayah maka rumah tangga wilayah pinggiran kota diberi bobot 1/4n, dan rumah tangga di wilayah lainnya diberi bobot 3/4n. Dengan kata lain, rumah tangga di wilayah pinggiran kota memperoleh bobot 1, dan rumah tangga wilayah lainnya memperoleh bobot 3. Prosedur semacam ini menghasilkan representasi secara proporsional kepada setiap elemen sampel yang diambil.

Berikut ini adalah contoh masalah yang dapat muncul jika sampling non-proporsional tidak disertai dengan penentuan bobot sampel sebagaimana contoh pada majalah bisnis terkemuka Harvard Business Review.[2] Majalah ini pernah melakukan survei terhadap para pelanggan mereka mengenai isu pelecehan seks di tempat kerja. Majalah memiliki pelanggan pria sebanyak 93% dan 7% wanita. Karena jumlah pelanggan pria jauh lebih banyak dari pada wanita maka dapatlah diterima jika majalah tersebut mengirimkan kuesioner dengan proporsi lebih besar kepada pelanggan wanita untuk dipilih sebagai sampel dengan perbandingan 68% (pria) dn 32% (wanita). Respon yang diberikan pelanggan dalam bentuk pengembalian kuesioner adalag 52% (pria) dan 44% (wanita).

Hasil penelitian menyatakan bahwa 73% responden mendukung peraturan perusahaan yang melarang segala bentuk pelecehan seksual. Angka 73% ini tentu saja terlalu tinggi karena secara tidak proporsional pengambilan sampel telah memberikan

[2] Earl Babbie, *OpCit*, hal 237

peluang lebih besar kepada wanita yang kemungkinan lebih mendukung kebijakan tersebut untuk terpilih sebagai sampel. Hasil penelitian juga menyatakan bahwa manajemen puncak adalah pihak yang paling merasakan bahwa isu pelecehan seks di tempat kerja adalah sesuatu yang dilebih-lebihkan dibandingkan dengan karyawan level menengah, dan level bawah. Kesimpulan ini juga patut dicurigai karena data menunjukkan bahwa sebagian besar sampel wanita secara tidak proporsional mewakili karyawan level bawah. Hal ini menunjukkan adanya perbedaan pandangan antara level karyawan mengenai isu bersangkutan. Singkatnya, majalah tersebut gagal memisahkan hasil temuannya berdasarkan jenis kelamin. Solusi bagi survei semacam ini adalah memberikan bobot berdasarkan jenis kelamin terhadap jawaban yang diberikan sampel sebagaimana yang telah dijelaskan sebelumnya.

Sampel nonprobabilitas

Jika teknik penarikan sampel probabilitas tidak memungkinkan untuk dilakukan, maka peneliti dapat menggunakan teknik sampel nonprobabilitas yang terdiri dari empat tipe penarikan sampel yaitu:

1) Sampel tersedia (*available sampling*), atau disebut juga sampel kebetulan (*accidental sampling*)
2) Sampel terpilih (*purposive sampling*), atau disebut juga *judgemental sampling*);
3) Sampel bola salju (*snowball sampling*)
4) Sampel kuota (*quota sampling*)

2.4.1 Accidental sampling

Salah satu cara yang banyak digunakan peneliti pemula dalam memilih sampel adalah menggunakan apa yang disebut dengan sampel tersedia (*available sample*) yang banyak menerima kritik dalam hal efektivitasnya. Sampel tersedia atau sering disebut dengan sampel kenyamanan (*convenience sample*) adalah kumpulan individu, elemen atau peristiwa yang sudah langsung tersedia, dan dapat langsung digunakan untuk penelitian, seperti pengunjung pusat perbelanjaan, atau kelompok mahasiswa yang mendaftar pada suatu mata kuliah tertentu dan sebagainya. Walaupun sampel tersedia dalam kasus tertentu dapat membantu peneliti dalam mengumpulkan informasi eksploratif, dan boleh jadi dapat menghasilkan data yang berguna namun demikian sampel semacam ini dapat menimbulkan masalah karena memiliki tingkat kesalahan yang tidak diketahui. Peneliti perlu mempertimbangkan aspek positif dan negatif dari sampel tersedia sebelum menggunakannya dalam suatu penelitian.

Penggunaan sampel tersedia telah menimbulkan perdebatan di kalangan peneliti. Mereka yang menolak penggunaannya berpandangan bahwa hasil yang diperoleh tidak dapat digunakan mewakili populasi dan karenanya tidak memiliki validitas eksternal. Responden terpilih sebagai sampel semata-mata karena ia kebetulan ada atau tersedia saat itu. Misal, pemilihan responden diantara para pengunjung pusat perbelanjaan banyak dikritik karena hanya mereka yang berada di pusat perbelanjaan yang akan terpilih sementara yang berada di luar pusat perbelanjaan tidak mungkin terpilih. Namun para pendukung penggunaan sampel tersedia menyatakan bahwa jika suatu fenomena, karakteristik, sifat itu memang ada maka kesemua hal tersebut harus ada pada setiap sampel.

Penggunaan sampel tersedia dalam penelitian sebaiknya dihindari karena dapat menimbulkan bias yang disebabkan kedekatan responden terhadap situasi penelitian, namun demikian penggunaan sampel tersedia masih dapat memberikan manfaat dalam suatu penelitian pendahuluan (*pilot study*) atau untuk menguji suatu kuesioner. Penggunaan sampel tersedia sering kali membantu peneliti dalam mengatasi hambatan dalam hal prosedur penelitian, pengujian (*testing*), dan metodologi sebelum penelitian yang sesungguhnya dilaksanakan dengan menggunakan sampel yang dipilih secara lebih baik.

2.4.2 Purposive sampling

Sampel nonprobabilitas lainnya adalah sampel terpilih atau *purposive sample* yang mencakup responden, subjek atau elemen yang dipilih karena karakteristik atau kualitas tertentu, dan mengabaikan mereka yang tidak memenuhi kriteria yang ditentukan. Melalui teknik *purposive sample* ini, sampel dipilih berdasarkan pengetahuan yang telah dimiliki sebelumnya mengenai populasi, yaitu pengetahuan mengenai elemen-elemen yang terdapat pada populasi, dan tujuan penelitian yang hendak dilakukan.

Sampel terpilih (sering pula disebut dengan *judgmental sampling*) dapat didefiniskan sebagai tipe penarikan sampel nonprobabilitas yang mana unit yang hendak diamati atau diteliti dipilih berdasarkan pertimbangan peneliti dalam hal unit yang mana dianggap paling bermanfaat dan representatif.[3] Dengan demikian pada sampel *purposive*, responden atau anggota sampel dengan sengaja dipilih tidak secara acak. Penentuan sampel terpilih dilakukan dengan pengetahuan bahwa sampel bersangkutan tidaklah representatif terhadap populasi. Dengan kata lain sampel *purposive* adalah sampel yang dipilih berdasarkan suatu panduan tertentu.

Panduan sampel yang digunakan akan menentukan batasan jumlah, atau kategori responden yang boleh dipilih, dan diundang sebagai anggota sampel. Misal, jika manajemen suatu stasion radio ingin melakukan penelitian terhadap target audien mereka yaitu pria berumur 25-44 tahun untuk mengetahui tanggapan mereka terhadap program radio bersangkutan, maka penelitian tersebut hanya ditujukan kepada siapa saja pria berusia 25-44 tahun. Calon responden yang memenuhi kriteria tersebut kemudian dihubungi, dan diundang untuk bersedia menjadi responden penelitian. Dalam penelitian semacam ini biasanya tidak tersedia daftar lengkap nama-nama pria berusia 25-44 tahun, sehingga tidak dapat menggunakan panduan matematis, dan kalaupun ada belum tentu tersedia daftar alamat atau telepon mereka. Namun demikian kriteria atau panduan terhadap responden telah ditentukan, pria berusia 25-44 tahun, sebagai sampel yang memenuhi kriteria.

2.4.3 Quota sampling

Pada sampel kuota (*quota sample*) individu atau responden dipilih untuk memenuhi suatu prosentase yang sudah diketahui atau sudah ditentukan sebelumnya. Sampel kuota dapat didefiniskan sebagai suatu tipe penarikan sampel nonprobabilitas dimana unit sampel (responden) dipilih sebagai sampel berdasarkan karakteristik yang telah ditentukan sebelumnya, sedemikian rupa sehingga total sampel akan memiliki distribusi dengan karakteristik yang sama sebagaimana yang diperkirakan terdapat dalam populasi yang tengah diteliti.[4]

Untuk melakukan penarikan sampel dengan menggunakan sampel kuota, peneliti harus mengawalinya dengan membuat suatu matrik atau tabel yang menjelaskan karakteristik dari populasi yang akan diteliti (lihat Tabel 1).

---

[3] Earl Babbie, *The Basic of Social Research*, 4th Edition, Thomson Wadsworth, 2008, hal 204

[4] Earl Babbie, *The Basic of Social Research, Ibid*, hal 205

Tergantung pada tujuan riset yang ingin dicapai, peneliti terlebih dulu harus mengetahui, misalnya, berapa jumlah laki-laki dan perempuan yang terdapat pada suatu populasi, dan dari masing-masing kelompok laki-laki dan perempuan tersebut, berapa jumlah anak-anak, remaja, pemuda, dewasa, dan orang tua; berapa jumlah yang berpendidikan sarjana, sekolah menengah (SMP/SMU), atau hanya sekolah dasar. Begitu pula, berapa jumlah orang dengan latar belakang etnis atau suku bangsa tertentu (suku Jawa, Sunda, Batak dll) yang terdapat dalam suatu populasi. Pada sampel kuota, setiap kelompok masyarakat tersebut harus memiliki wakilnya masing-masing dalam jumlah yang proporsional. Pada tingkat nasional, penarikan sampel kuota terkadang harus pula mempertimbangkan sampel yang mewakili wilayah perkotaan, pedesaan, Jawa atau luar Jawa, kelas menengah, pribumi atau keturunan dan lain-lain.

Ketika matrik atau tabel yang tersusun dari sejumlah sel yang mewakili kelompok-kelompok dalam masyarakat berdasarkan karakteristiknya masing-masing tersebut telah dapat disusun, dan jumlah anggota masing-masing kelompok tersebut telah dapat diketahui, maka peneliti dapat menentukan jumlah responden yang akan mewakili masing-masing sel tersebut secara proporsinal. Selanjutnya, peneliti dapat mulai melakukan pengumpulan data dari orang-orang yang mewakili masing-masing sel dalam jumlah yang ditentukan dulu sebelumnya secara proporsional. Jika semua data telah dapat diperoleh dari sampel secara proporsional maka

Tabel 1: Karakteristik populasi

	Laki-laki	Perempuan
Usia	Anak-anak	Anak-anak
	Ramaja	Ramaja
	Muda	Muda
	Dewasa	Dewasa
	Tua	Tua
Pendidikan	Sarjana	Sarjana
	SMP/SMU	SMP/SMU
	SD	SD
Agama	Islam	Islam
	Kristen	Kristen
	Hindu	Hindu
	Budha	Budha
Etnis/suku	Jawa	Jawa
	Sunda	Sunda
	Batak	Batak
	Minang	Minang

kita dapat mengatakan bahwa data yang kita peroleh adalah representatif terhadap populasi.[5]

Teknik penarikan sampel kuota ini mirip dengan sampel probabilitas, namun jika tidak dilakukan dengan cermat penarikan sampel kuota memiliki potensi bermasalah. Pertama, jumlah anggota masing-masing sel (kelompok) haruslah akurat, namun sering kali peneliti dalam menyusun matrik atau tabel menggunakan data lama yang tidak menggambarkan perkembangan masyarakat terbaru. Hal ini yang terjadi pada lembaga survei terkenal di AS, Gallup, ketika pada tahun 1948 mengumumkan hasil survei yang menyatakan pemenang pemilu presiden AS tahun itu adalah gubernur negara bagian New York, Thomas Dewey, yang mengalahkan Presiden Harry Truman. Hasil survei ini keliru karena ternyata Trumanlah yang menang. Gallup menggunakan teknik sampel kuota yang menuntut peneliti mengetahui secara pasti mengenai data kependudukan (data pemilih) yang biasanya diperoleh dari data sensus penduduk. Dalam hal ini, Gallup menggunakan data sensus penduduk tahun 1940. Namun sejak Perang Dunia ke-2 meletus hingga tahun 1948 banyak penduduk pedesaan (*country*) di AS pindah ke kota sehingga karakter penduduk AS berubah secara signifikan, dari sebelumnya kebanyakan tinggal di kawasan pertanian di desa menjadi lebih banyak tinggal di perkotaan. Warga kota yang dinamis cenderung memilih Truman yang didukung Partai Demokrat yang dipandang progresif sedangkan warga desa lebih suka Partai Republik yang konservatif.

2.4.4 Snowball sampling

Metode sampel nonprobabilitas lainnya disebut dengan sampel bola salju (*snowball sampling*) dimana peneliti secara acak menghubungi beberapa responden yang memenuhi kriteria (*qualified volunteer sample*) dan kemudin meminta responden bersangkutan untuk merekomendasikan teman, keluarga, atau kenalan yang mereka ketahui yang memenuhi kriteria untuk dijadikan sebagai responden penelitian. Peneliti

---

[5] Misal, seorang peneliti tertarik untuk mengetahui perbedaan penggunaan televisi antara mereka yang memiliki *DVD player* dengan mereka yang tidak memiliki *DVD player*. Jika peneliti mengetahui 40% penduduk dari suatu populasi tertentu memiliki DVD maka sampel yang harus dipilih peneliti harus terdiri dari 40% pemilik DVD dan 60% mewakili mereka yang tidak memiliki DVD agar dapat mencerminkan karakteristik populasi.

kemudian menghubungi orang dimaksud untuk menentukan apakah mereka memenuhi kriteria sebagai responden.

Istilah "bola salju" mengacu pada proses pengumpulan sampel dengan meminta responden yang diketahui keberadaannya untuk menunjukkan calon responden lainnya. Dengan demikian sampel bola salju dapat didefiniskan sebagai suatu metode penarikan sampel nonprobabilitas dimana setiap orang yang diwawancarai kemudian ditanyakan sarannya mengenai orang lain yang dapat diwawancarai.[6]

Prosedur sampel ini dapat digunakan dalam hal anggota populasi yang hendak diteliti sulit diketahui keberadaannya sehingga tidak mudah untuk ditemui, misalnya, para pekerja migran, tuna wisama, pekerka seks komersial, atau tenaga kerja illegal. Misal, suatu penelitian dilakukan untuk mengetahui pola-pola rekrutmen anggota baru suatu organisasi. Peneliti mewawancarai beberapa orang yang diketahuinya sebagai anggota baru suatu organisasi, dan menanyakan kepada mereka siapa yang memperkenalkan atau mengajak mereka bergabung ke dalam organisasi bersangkutan. Peneliti kemudian menghubungi orang-orang yang disebut namanya dan menanyakan pertanyaan yang sama begitu seterusnya.

Walaupun prosedur pengambilan sampel semacam ini sepertinya valid, namun kurang memenuhi syarat bagi suatu penelitian ilmiah (*legitimate*) karena sampel yang dihasilkan bisa menjadi sangat bias. Peneliti bisa jadi menemukan sampel yang terdiri dari organisasi atau kelompok tertentu saja. Prosedur semacam ini dapat menghasilkan sampel yang diragukan keterwakilannya, karena itu sampel semacam ini biasanya digunakan untuk penelitian eksploratif.

----

[6] Earl Babbie, *The Basic of Social Research, OpCit,* hal 205.

STATISTIK SOSIAL TERAPAN

MODUL KULIAH

Karakteristik Data

Salah satu fungsi statistik adalah mengelola dan meringkaskan data yaitu kumpulan nilai atau skor pengukuran atau pengamatan yang diperoleh peneliti dari populasi atau sampel. Statistik yang digunakan untuk mengelola dan meringkaskan data ini disebut dengan statistik deskriptif. Hal paling penting dalam statistik deskriptif adalah memahami karakteristik data yaitu ke mana sebagian besar skor atau nilai tersebut berkumpul, dan cara yang paling umum digunakan untuk memahami karakteristik data adalah dengan menentukan suatu nilai rata-rata yang menjadi indikator ke mana skor lebih banyak berkumpul. Nilai rata-rata merupakan nilai tunggal yang dapat mewakili (representatif) keseluruhan distribusi skor. Dalam statistik, konsep nilai rata-rata atau nilai representatif ini dinamakan dengan kecenderungan menuju ke pusat atau disebut juga dengan 'kecenderungan memusat' (*central tendency*).

Selain kecenderungan untuk memusat, data juga memiliki karakteristik untuk berbeda atau menyebar dari nilai rata-ratanya. Ukuran seberapa jauh jarak rata-rata setiap skor dari nilai rata-rata distribusi skor dinamakan variabilitas. Dalam statistik, karakteristik data yang memusat namun sekaligus menyebar ini merupakan sesuatu yang sangat penting untuk dipahami. Data yang diperoleh dari populasi atau sampel adalah barat dua sisi mata uang koin, pada satu sisi memiliki sifat memusat namun pada sisi lainnya menyebar. Dengan demikian terdapat dua karakteristik penting dari suatu data statistik yaitu: 1) kecenderungan memusat (*central tendency*) dan; 2) penyebaran atau pemencaran (variabilitas), atau disebut juga dengan dispersi. Kita akan membahas kedua hal tersebut pada bagian ini.

Kecenderungan Memusat

Data dengan kecenderungan memusat (*measures of central tendency*) menjelaskan apa yang sedang terjadi pada beberapa kelompok sampel atau populasi secara rata-rata.

Anda mungkin pernah mendengar orang berkata: curah hujan “rata-rata”, tingkat pemahaman komunikasi “rata-rata”, jumlah “rata-rata” jam orang menonton televisi. Tetapi bagi seorang peneliti, pengertian rata-rata memiliki maksud berbeda dengan orang awam.

Statistik dengan kecenderungan memusat berfungsi menjawab pertanyaan, “Apa ciri khas yang dimiliki suatu data?” Data statistik biasanya memberikan informasi mengenai suatu skor atau nilai diantara sejumlah nilai lainnya yang menjadi ciri khas, sifat atau karakteristik dari distribusi bersangkutan. Penentuan nilai yang menjadi ciri khas suatu distribusi tergantung pada tingkat pengukuran dan tujuan penggunaan data.

Walaupun nilai rata-rata merupakan nilai yang berperan penting dalam statistik namun nilai tersebut sering terganggu dengan adanya nilai ekstrim. Karena itu, metode kecenderungan memusat lainnya seperti median dan modus dapat kita gunakan. Pada setiap distribusi dengan kecenderungan memusat terdapat tiga karakteristik yang harus kita kenali sebagaimana yang akan kita bahas pada bagian berikut ini yaitu: modus, median dan *mean*.

Tabel 3.1

Skor	*f*
70	2
35 – 69	0
34	1
33	1
32	1
31	1
30	1
29	1
28	1
27	1
26	1

Modus

Suatu modus (Inggris *mode*) adalah suatu angka atau nilai yang paling sering muncul dalam suatu distribusi. Penentuan modus tidak berdasarkan suatu perhitungan atau kalkulasi tertentu tetapi berdasarkan pengamatan atau inspeksi secara manual terhadap suatu distribusi. Umpamakan kita memiliki sampel yang terdiri dari delapan mahasiswa yang mengambil mata kuliah statistik. Mereka ditanya seberapa suka mereka dengan mata kuliah statistik dengan cara memberikan nilai atau poin 1 sampai 10 (poin 1 menyatakan sangat tidak suka dan poin 10 menyatakan sangat menyukai). Mereka kemudian memberikan jawaban sebagai berikut:

3 4 5 6 6 7 8 9

Nilai yang paling sering muncul dari data tersebut adalah 6.

Walaupun penentuan modus tampaknya mudah dilakukan namun terkadang memiliki kelemahan karena terlalu terfokus pada satu nilai saja. Penentuan modus dalam statistik deskriptif dapat menutupi atau menyembunyikan sejumlah fakta penting terkait dengan data yang terdapat pada suatu distribusi. Hal ini dapat dilihat pada Tabel 3.1 yang menunjukkan angka 70 sebagai nilai modus, namun ciri penting dari distribusi tersebut adalah data nilai kebanyakan terkonsentrasi di sekitar 30. Kelemahan serius lainnya, suatu distribusi terkadang memiliki lebih dari satu modus. Jika hal ini yang terjadi maka modus tidak akan efektif dalam melakukan analisa data.

Soal Latihan

1. Selama bulan Oktober, seorang guru mencatat jumlah ketidakhadiran setiap pelajar di kelasnya yang memiliki siswa n = 20 dan ia memiliki nilai distribusi ketidakhadiran sbb:

Jumlah Ketidahadiran	f
5	1
4	2
3	7
2	5
1	3
0	2

a. Tentukan nilai rata-rata ketidakhadiran pelajar di kelas tersebut?

b. Dengan menggunakan median, tentukan nilai rata-rata ketidakhadiran pelajar?

c. Dengan menggunakan modus, tentukan nilai rata-rata ketidakhadiran pelajar?

Jawaban

a. Nilai rata-rata = $\frac{5+4+4+3+3+3+3+3+3+3+2+2+2+2+2+1+1+1+0+0}{20}$

$= \frac{47}{20} = 2.35$

b. Median = $\frac{3+2}{2} = 2.5$

c. Modus adalah 3

Median

Suatu distribusi memiliki ciri lain yaitu adanya nilai median (Mdn) yaitu nilai tengah yang berarti setengah dari keseluruhan nilai berada di atas nilai tersebut dan setengah lainnya berada di bawah nilai tersebut. Dengan kata lain median adalah nilai yang muncul ditengah-tengah dari suatu daftar nilai yang teratur. Karena letaknya ditengah-tengah maka median berfungsi memisahkan setengah bagian data dengan setengah bagian lainnya.

Jika jumlah unit suatu distribusi adalah ganjil maka nilai median terletak persis ditengah-tengah distribusi. Jika jumlah unit distribusi adalah genap maka nilai median ditentukan berdasarkan penjumlahan dua nilai paling tengah dan kemudian dibagi dua. Untuk menentukan median, keseluruhan nilai disusun mulai dari yang terkecil hingga nilai terbesar, nilai tengahnya ditentukan berdasarkan pengamatan.

Data yang kita miliki sebelumnya memiliki delapan nilai yang disusun dari nilai terendah hingga tertinggi. Sehingga tidak ada nilai yang terletak di tengah. Apa yang harus kita lakukan? Untuk mengetahui median, kita "bagi dua perbedaan" antara kedua nilai yang terletak ditengah. Pada data mengenai kesukaan mahasiswa terhadap kuliah statistik, dua nilai terletak di tengah-tengah yaitu dua nilai 6, dengan demikian nilai media adalah 6 ($\frac{6+6)}{2}$). Kembali ke tabel 5.1 dimananilai mediannya adalah 14.

Contoh, median dari suatu distribusi yang terdiri dari sembilan angka berikut ini adalah 6 karena terdapat empat nilai di atas dan dibawah nilai 6.

0 2 2 5 **6** 17 18 19 67

Perhatikan suatu distribusi yang terdiri atas 10 angka berikut ini:

$$0 \quad 2 \quad 2 \quad 5 \quad 6 \quad \underset{\underset{11.5}{\uparrow}}{} \quad 17 \quad 18 \quad 19 \quad 67 \quad 75$$

Pada contoh distribusi yang terdiri atas 10 angka ini, tidak ada satu nilai yang persis berada ditengah. Untuk menentukan median, dua nilai yang terletak di tengah harus ditambahkan dan dibagi dengan dua:

$$\text{Mdn} = \frac{6 + 17}{2} = 11.5$$

Soal Latihan

2. Tentukan median untuk setiap distribusi skor berikut ini:
   a. 3, 4, 6, 7, 9, 10, 11
   b. 8, 10, 11, 12, 14, 15
3. Jika anda memiliki skor 52 poin pada suatu ujian dengan skor tertinggi 80 poin, maka nilai anda sudah pasti di atas median

Jawaban

1. a. Nilai median adalah $x = 7$
   b. Nilai median adalah $x = \frac{11+12}{2} = 11.5$
2. Salah. Nilai median akan tergantung pada dimana seluruh skor berada.

Mean

Karakteristik sentral ketiga dari suatu data statistik adalah nilai rata-rata-rata atau *mean* (baca 'min') yang didefiniskan sebagai penjumlahan seluruh nilai dibagi dengan *n* yaitu jumlah nilai seluruhnya. Karena nilai rata-rata digunakan secara luas baik pada statistik deskriptif dan statistik inferensial, maka nilai rata-rata akan kita bahas secara cukup mendalam pada bagian ini.

Terdapat beberapa cara untuk menjelaskan kecenderungan memusat dari suatu data. Nilai rata-rata aritmatika (*arithmetic mean*) merupakan nilai yang sering disebut dengan nilai rata-rata. Nilai ini diperoleh dengan menjumlahkan seluruh nilai dan membaginya dengan banyaknya nilai yang ada. Nilai rata-rata populasi dihitung dengan menggunakan rumus berikut:

$$\mu = \frac{\Sigma x}{N}$$

Simbol μ mewakili nilai rata-rata populasi. Huruf N besar mewakili jumlah keseluruhan anggota populasi atau jumlah seluruh kejadian dalam populasi. Tanda Σ dinamakan "sigma" yang merupakan simbol matematika yang memerintahkan kita untuk menjumlahkan atau menambahkan nilai apa saja yang terdapat setelah simbol itu.

Untuk menghitung nilai rata-rata sampel kita menggunakan rumus berikut:

$$\bar{x} = \frac{\Sigma x}{n}$$

Nilai rata-rata sampel diwakili oleh simbol x, dibaca "x bar" dan huruf n digunakan untuk menunjukkan jumlah seluruh anggota sampel atau jumlah seluruh kejadian pada sampel. Jika kita masukkan nilai dari sampel yang kita miliki sebelumnya mengenai kesukaan mahasiswa terhadap mata kuliah statistik maka kita menemukan nilai rata-rata adalah 6. Rumus untuk mencari nilai rata-rata sampel dan nilai rata-rata populasi adalah sama. Nilai rata-rata sampel dikatakan sebagai **estimator tidak bias** (*unbiased estimator*) yang berarti bahwa nilai rata-rata sampel bersifat representatif mendekati nilai-rata populasi. Jika kita tidak memiliki nilai rata-rata populasi untuk dimasukkan ke dalam rumus maka nilai rata-rata sampel dapat menggantikannya.

Persamaan tersebut menunjukkan bahwa nilai rata-rata (x) sama dengan penjumlahan seluruh nilai (Σx) dibagi dengan total nilai keseluruhan. Berdasarkan data pada tabel 3.1 maka nilai rata-ratanya adalah:

$$\bar{x} = \frac{293}{20} = 14.65$$

Jika suatu data dimasukkan ke dalam suatu tabel distribusi frekuensi, kita menggunakan suatu rumus yang sedikit agak berbeda untuk menghitung nilai rata-rata sebagai berikut:

$$\bar{x} = \frac{\sum fx}{n}$$

Dalam kasus ini x mewakili nilai tengah dari setiap interval, dan f adalah frekuensi interval tersebut. Tabel 3.2 menggunakan formula ini untuk menghitung nilai rata-rata distribusi frekuensi dari contoh sebelumnya pada Tabel 2.8.

Tidak seperti modus dan median, nilai rata-rata dihasilkan dengan memperhitungkan seluruh nilai yang terdapat pada suatu distribusi. Hal ini menyebabkan nilai rata-rata sensitif terhadap nilai ekstrim (*outlier*). Keberadaan nilai ekstrim pada suatu distribusi memberikan pengaruh pada nilai rata-rata secara keseluruhan. Jika, misal, tabel 3.1 memiliki satu nilai ekstrim yaitu responden U yang menyatakan menggunakan televisi selama 100 jam maka nilai rata-rata baru kemudian menjadi 18.71, suatu kenaikan sekitar 28% hanya karena penambahan satu nilai ekstrim.

Sejauh ini kita telah mempelajari tiga nilai yang menjadi ciri suatu distribusi yaitu: nilai modus, median dan *mean* atau nilai rata-rata. Lantas, dari ketiga kecenderungan memusat atau karakteristik sentral tersebut yang manakah yang sebaiknya kita gunakan dalam suatu laporan analisa data. Dua faktor harus dipertimbangkan dalam memutuskan yang mana dari ketiga pengukuran karakteristik sentral tersebut yang akan digunakan. Pertama, level pengukuran yang digunakan memberikan pengaruh terhadap pilihan. Jika data merupakan data nominal maka nilai modus yang memberikan makna; jika data adalah ordinal maka nilai modus dan median lebih tepat digunakan. Ketiga tipe pengukuran tersebut –modus, median dan *mean*- dapat digunakan bagi data interval dan data rasio

Tabel 3.2: menghitung nilai rata-rata distribusi

Jam	Frekuensi	fx
6	1	6
8	2	16
9	1	9
11	3	33
12	2	24
14	2	28
15	1	15
16	1	16
18	1	18
19	2	38
21	2	42
23	1	23
25	1	25
	$n = 20$	$\sum fX = 293$
	$\bar{X} = \frac{293}{20} = 14.65$	

.

Kedua, penting untuk mengetahui tujuan statistik yang hendak kita gunakan. Jika tujuannya adalah untuk menjelaskan data maka kita dapat menggunakan pengukuran yang paling khusus bagi pengukuran distribusi. Misal, nilai ujian statistik mahasiswa adalah sebagai berikut: 100, 100, 100, 100, 0, dan 0. Mengatakan bahwa nilai rata-ratanya adalah 67 tentunya kurang dapat menggambarkan distribusi nilai secara tepat, namun penjelasan terhadap karakteristik distribusi akan lebih tepat jika kita menggunakan nilai modus.

Rangkuman

1. Kecenderungan memusat, atau pemusatan (*central tendency*), adalah suatu cara untuk menentukan suatu nilai tunggal yang berfungsi untuk mengidentifikasi pusat distribusi yang juga menjadi representasi terbaik bagi keseluruhan skor atau nilai distribusi. Tiga ukuran standar kecenderungan memusat adalah modus, median dan *mean*.

2. *Mean* (baca 'min') merupakan nilai rata-rata aritmatika yang dihitung dengan cara menjumlahkan seluruh skor dan kemudian membaginya dengan jumlah skor. Nilai rata-rata diperoleh dengan membagi total nilai (Σx dengan banyaknya skor (N atau n). Nilai rata-rata dapat pula didefinisikan sebagai nilai keseimbangan dari suatu distribusi. Jarak di atas nilai rata-rata akan diimbangi secara tepat dengan jarak yang terletak di bawahnya. Walaupun cara perhitungannya sama bagi nilai rata-rata untuk populasi atau sampel tetapi berbeda dalam hal notasi yang digunakan. Nilai rata-rata untuk populasi menggunakan simbol μ (baca 'miu') dan nilai rata-rata untuk sampel menggunakan simbol $\bar{x}$ (baca x bar). Nilai rata-rata merupakan alat ukur yang paling sering digunakan untuk menentukan nilai pemusatan atau kecenderungan memusat. Nilai rata-rata lebih sering digunakan pada data yang diperoleh dari hasil pengukuran yang menggunakan skala interval dan rasio.

3. Mengubah skor pada suatu distribusi akan menyebabkan perubahan nilai rata-rata. Jika suatu nilai ditambahkan atau dikurangkan dari setiap skor distribusi, maka nilai rata-rata akan bertambah atau berkurang pula dengan nilai yang sama. Jika suatu nilai dikalikan dengan setiap skor pada distribusi maka nilai rata-rata juga akan bertambah sebesar nilai pengalinya.

4. Median adalah nilai tengah (*midpoint*) suatu distribusi skor. Penggunaan media lebih disukai ketika suatu distribusi memiliki bebera nilai ekstrim (*outlier*). Median juga digunakan untuk distribusi terbuka khususnya bila terdapat nilai yang tidak terbatas (*infinite score*) sehingga tidak mungkin menghitung nilai rata-rata. Median

lebih sering digunakan pada data dari hasil pengukuran dengan menggunakan skala ordinal.

5. Modus adalah nilai yang paling sering muncul dalam suatu distribusi. Modus lebih tepat digunakan untuk data yang menggunakan skala nominal. Suatu distribusi data dapat memiliki lebih dari satu modus.

3.1.5 Soal Latihan

1. Tentukan nilai rata-rata sampel berikut ini yang memiliki n = 5 skor dengan distribusi : 1, 8, 7, 5, 9
2. Suatu sampel dengan n = 6 skor memiliki nilai rata-rata (x) = 8. Berapa nilai Zx sampel?
3. Suatu sampel pertama dengan n = 5 skor memiliki nilai rata-rata x= 4. Suatu sampel kedua memiliki n = 3 skor dan nilai rata-rata $\bar{x}$= 10. Jika kedua sampel digabungkan, berapa nilai rata-rata sampel gabungan?
4. Suatu sampel n = 6 skor memiliki nilai rata-rata x= 40. Satu skor baru ditambahkan ke dalam sampel dan nilai rata-ratanya berubah menjadi x= 35. Apa yang dapat disimpulkan mengenai skor baru tersebut?
   a. skor baru lebih besar dari 40.
   b. skor baru kurang dari 40.
5. Tentukan nilai bagi n, Σx, dan $\bar{x}$ dari tabel distribusi frekuensi berikut ini

x	f
5	1
4	2
3	3
2	5
1	1

6. Menambahkan satu skor baru ke dalam suatu distribusi akan selalu mengubah nilai rata-ratanya? (Benar atau salah?)
7. Mengubah nilai suatu skor pada suatu distribusi nilai akan selalu mengubah nilai rata-ratanya (Benar atau salah?)
8. Suatu populasi memiliki nilai rata-rata $\mu = 40$
   a. Jika setiap skor ditambah nilainya dengan 5, berapa nilai rata-ratanya sekarang?
   b. Jika setiap skor dikalikan nilainya dengan 3, berapa nilai rata-ratanya sekarang?
9. Suatu sampel memiliki $n = 4$ skor, dan nilai rata-rata 9. Jika satu orang yang memiliki skor 3 dikeluarkan dari sampel, berapa nilai rata-rata sampel sekarang?
10. Nilai apakah yang paling mudah terpengaruh jika suatu skor ekstrim ditambahkan ke dalam suatu distribusi. Apakah nilai rata-rata, median ataukah modus?

Jawaban

1. Nilai rata-rata $\bar{x} = \frac{\Sigma x}{n} = \frac{1+8+7+5+9}{5} = \frac{30}{5} = 6$
2. Jika $\bar{x} = \frac{\Sigma x}{n}$ maka $\Sigma x = x.\ n = 6 . 8 = 48$
3. Nilai n gabungan = 5 + 3 = 8 skor, dan gabungan $\Sigma x = (x_1 . n_1) + (\bar{x}_2 . n_2) = (5 . 4) + (3 . 10) = 20 + 30 = 50$. Nilai rata-rata $\bar{x} = \frac{50}{n_1 + n_2} = \frac{50}{5+3} = \frac{50}{8} = 6.25$
4. b
5. Distribusi data tabel adalah sbb: 5, 4, 4, 3, 3, 3, 2, 2, 2, 2, 2, 1 sehingga $n = 12$, $\Sigma X = 5 + 4 + 4 + 3 + 3 + 3 + 2 + 2 + 2 + 2 + 2 + 1 = 33$, dan $x = \frac{33}{12} = 2.75$
6. Salah. Jika skor baru sama dengan nilai rata-rata maka nilai rata-rata tidak akan berubah.
7. Benar
8. a. Nilai rata-rata (μ) baru menjadi = 40 + 5 = 45
   b. Nilai rata-rata (μ) baru menjadi = 40 . 3 = 120

9. Sampel awal memiliki n = 4 skor sehingga $\Sigma x = x.\ n$ = 9 . 4 = 36. Sampel baru memiliki n = 4 – 1 = 3 sehingga Σx = 8 . 3 = 36. Sampel baru memiliki n = 3 skor sehingga Σx =36 – 3 = 33. Nilai rata-rata baru menjadi = $\frac{33}{3}$ = 11
10. Nilai rata-rata

D I S P E R S I

Jenis statistik deskriptif kedua adalah dispersi yang digunakan untuk mengukur penyebaran data. Dispersi sering juga disebut dengan varian. Berbagai pengukuran data dengan kecenderungan memusat akan menentukan nilai khas suatu distribusi; pengukuran dispersi menjelaskan bagaimana nilai tersebar dari titik pusatnya. Pengukuran dispersi sangat bermanfaat untuk membandingkan berbagai distribusi yang berbeda. Misal, nilai kuliah metode riset dari dua kelas berbeda menunjukkan nilai rata-rata yang sama; namun demikian salah satu kelas memiliki beberapa mahasiswa dengan nilai yang sangat baik sedangkan sisanya memiliki nilai sangat buruk. Sedangkan mahasiswa pada kelas lain semuanya menunjukkan nilai pada kisaran rata-rata. Dalam hal ini, kita dapat menggunakan pengukuran dispersi untuk menunjukkan perbedaan. Dalam banyak kasus, data mengenai nilai mahasiswa tersebut dapat dijelaskan dengan menggunakan pengukuran dengan kecenderungan memusat dan indeks dispersi. Kita dapat menggunakan tiga pengukuran dispersi yaitu jarak, varian dan deviasi standar.

J a r a k

Jarak atau *Range* (R) merupakan bentuk pengukuran yang paling sederhana. Jarak adalah perbedaan antara nilai tertinggi dan nilai terendah dalam suatu distribusi nilai. Rumus yang digunakan untuk menghitung jarak adalah

$$R = X_{hi} - X_{lo}$$

Dimana X_{hi} adalah nilai tertinggi dan X_{lo} adalah nilai terendah. Karena jarak hanya menggunakan dua nilai dari keseluruhan nilai yang terdapat pada suatu distribusi maka jarak tidak dapat mewakili atau menjelaskan data yang ada. Nilai jarak sering kali menjadi lebi besar seiring dengan peningkatan jumlah sampel karena sampel yang lebih besar cenderung menerima nilai ekstrim. Karena alasan ini, jarak jarang digunakan sebagai satu-satunya pengukuran dispersi.

Dalam penelitian, kita terkadang harus harus menentukan dispersi yaitu seberapa dekat atau seberapa jauh data yang diperoleh dengan nilai rata-rata. Misal, suatu tim olah raga bowling membutuhkan seorang pemain cadangan untuk menggantikan pemain utama yang cidera. Tim memiliki dua pilihan pemain pengganti, pemain A dan pemain B, untuk menggantikan pemain yang sakit. Catatan prestasi A pada tiga pertandingan terakhir mencetak skor (pin) 155, 175 dan 210 pin. Skor yang dicetak B pada tiga pertandingan terakhir adalah 172, 180, 188. Siapa yang harus dipilih? Kedua pemain cadangan tersebut memiliki nilai rata-rata yang sama.

Sebagaimana telah disinggung, metode pengukuran dispersi atau variabilitas menjelaskan seberapa jauh suatu data dari nilai rata-ratanya, dan jarak atau *Range* adalah interval dari nilai tertinggi hingga terendah. Pada contoh ini, data skor yang dimiliki pemain A memiliki jarak 55 (210 – 155), sedangkan pemain B adalah 16 (188 – 172).[1] Jarak sering digunakan untuk merangkum data, tetapi karena jarak tidak berhubungan dengan penyebaran nilai atau skor di sekitar nilai yang memiliki kecenderungan memusat maka peneliti tidak mampu membandingkan jarak secara berguna. Selain itu, jarak sangat dipengaruhi oleh nilai ekstrim.

Varian

Pengukuran kedua adalah varian yang merupakan indek matematis dari derajat penyimpangan suatu nilai dari nilai rata-ratanya. Suatu varian adalah ukuran dari

[1] Terkadang peneliti membagi suatu jarak ke dalam empat bagian untuk menjelaskan pembagian tambahan. Lihat Reinard C, John. *Introduction to Communication Research*, 4th Edition, McGraw-Hill. Inc, 2007, hal 460

pangkat dua deviasi nilai yang berbeda dari nilai rata-rata dengan simbol s^2 yang mewakili varian suatu sampel. Sedangkan simbol σ^2 (sigma kecil pangkat dua) mewakili varian populasi. Tetapi sekedar definisi belumlah cukup untuk menjelaskan betapa pentingnya kedua ukuran variabilitas ini. Mari kita gunakan beberapa logika. Kita ingin mengetahui suatu nilai yang menjelaskan jarak data rata-rata dari nilai rata-ratanya. Misal, kita memiliki data dari suatu sampel sebagai berikut: 1, 5, 9. Nilai rata-rata data tersebut adalah 5. Untuk menghitung dispersi data maka pertanyaannya adalah, seberapa jauh dari nilai rata-rata nilai 5 dengan dua nilai lainnya? Jika anda mengatakan 4 maka itulah deviasi standar sampel tersebut. Seluruh pengukuran deviasi standar pada prinsipnya adalah sesederhana ini. Namun persoalannya tidak sesederhana ini jika jumlah skor yang terlibat cukup banyak atau bahkan sangat banyak.

Suatu nilai varian yang kecil menunjukkan bahwa sebagian besar nilai yang terdapat pada distribusi terletak cukup dekat dengan nilai rata-ratanya. Dengan kata lain, kebanyakan nilai adalah sama; suatu nilai varian yang besar menunjukkan nilai-nilai pada distribusi tersebar meluas. Karena itu, varian secara langsung bersifat proporsional terhadap derajat dispersi atau berbeda diantara kelompok nilai.

Perhatikan contoh penelitian berikut untuk mendapatkan pemahaman lebih baik mengenai varian. Suatu penelitian dilakukan untuk mengetahui tingkat kesejahteraan pendukung suatu partai politik (parpol), misalnya antara parpol A dan parpol B guna membuktikan apakah benar klaim tidak resmi selama ini yang menyatakan pendukung parpol A lebih sejahtera dibandingkan pendukung parpol B. Peneliti memilih suatu sampel yang terdiri dari sejumlah individu, dan menanyakan kepada mereka dua pertanyaan: (1) apakah mereka mengidentifikasi diri mereka sebagai pendukung parpol A atau B?; (2) Berapa pendapatan mereka tahun lalu?

Data yang diperoleh dari sampel menunjukkan informasi bahwa pendukung parpol A memiliki tingkat penghasilan tahun lalu rata-rata sebesar Rp 21 juta, sedangkan tingkat penghasilan rata-rata pendukung parpol B adalah Rp 19 juta (lihat Gambar 3.1). Data secara jelas menunjukkan pendukung parpol A lebih sejahtera, namun apakah perbedaan ini signifikan? Apakah kita juga akan menemukan perbedaan tingkat penghasilan sebesar Rp 2 juta tersebut jika kita membuat dua kelompok sampel yang berbeda yang dipilih secara acak.

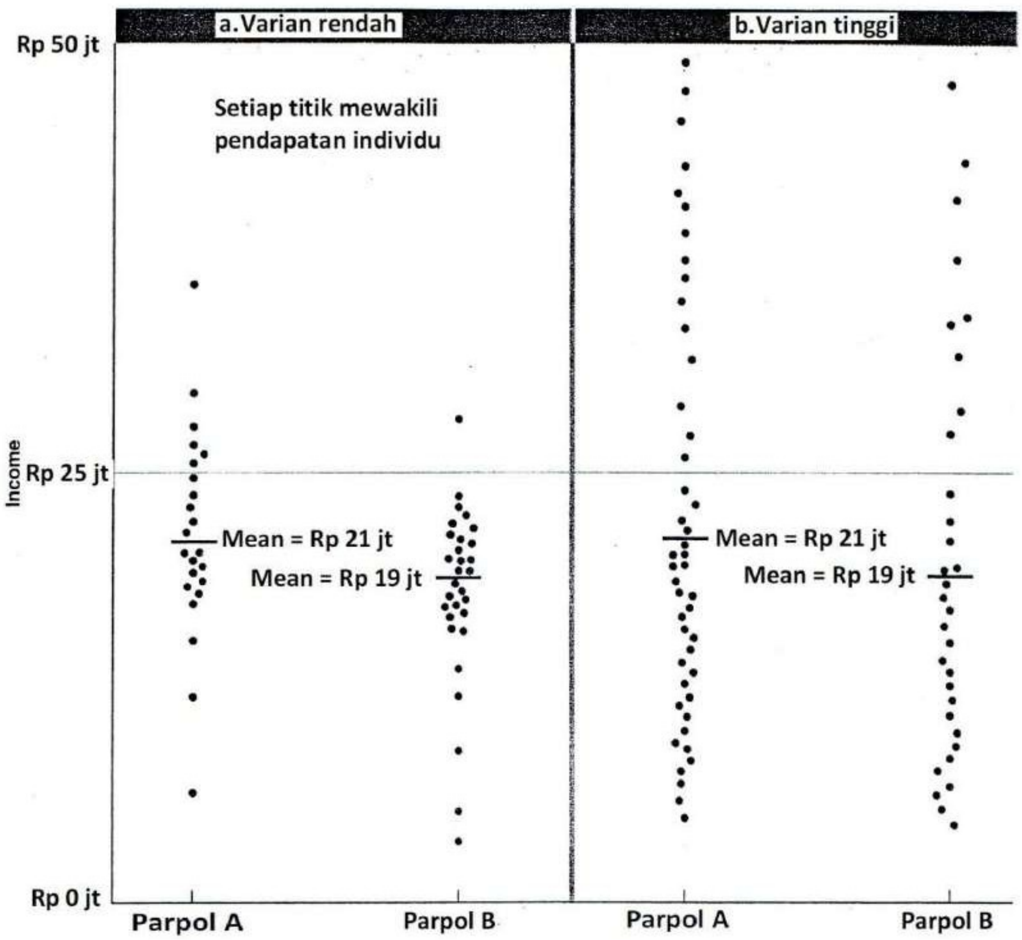
Secara sederhana, varian dari suatu distribusi seperti pendapatan tersebut adalah suatu pengukuran untuk mengetahui seberapa jauh sejumlah atau seperangkat nilai berkumpul mendekati nilai rata-rata atau jarak nilai-nilai tersebut dari nilai rata-ratanya. Gambar 3.1 menunjukkan kedua kemungkinan tersebut. Perhatikan Gambar 3.1a (varian rendah) dari distribusi tersebut, kebanyakan pendukung parpol A memiliki tingkat penghasilan rata-rata yang mendekati nilai rata-ratanya Rp 21 juta, sedangkan kebanyakan pendukung parpol B menunjukkan tingkat penghasilan yang juga mendekati nilai rata-ratanya yaitu Rp 19 juta.

Gambar 3.1b (varian tinggi) menunjukkan gambaran yang cukup berbeda. Walaupun tingkat penghasilan rata-rata kelompok menunjukkan nilai yang sama sebagaimana 3.1a, namun kedua kelompok pendukung menunjukkan perbedaan tingkat pendapatan yang mencolok mulai dari yang sangat tinggi hingga sangat rendah. Dengan kata lain, terdapat derajat varian yang lebih tinggi pada Gambar 3.1b dibandingkan Gambar 3.1a. Pada Gambar 3.1b, kebanyakan tingkat penghasilan pendukung kedua parpol tidak berkumpul mendekati nilai rata-rata. Kita lebih cenderung menyatakan Gambar 3.1a menunjukkan informasi yang lebih bisa diterima mengenai data tingkat pendapatan pendukung parpol A dan B dibandingkan Gambar 3.1b yang tampaknya menunjukkan tingkat kesalahan sampling yang lebih besar.

Mari kita gunakan contoh lain. Umpamakan kita memiliki seperangkat data yang terdiri dari sejumlah skor, dan kita ingin mengetahui seberapa jauh letak data tersebut dari nilai rata-ratanya. Untuk menjawab pertanyaan ini kita harus menemukan seberapa jauh jarak setiap poin terhadap nilai rata-ratanya. Kita mulai dengan mengurangkan nilai rata-rata (μ) dengan setiap skor (x). Tabel 3.3 menunjukkan data mengenai minat atau kesukaan mahasiswa terhadap mata kuliah statistik dengan cara memberikan nilai atau poin 1 sampai 10 (nilai 1 sangat tidak suka, nilai 10 sangat suka).

Gambar 3.1

Sumber: diadaptasi dari Earl Babbie, *The Practice of Social Research*, 11th Edition, Wadsworth, hal 477



Sekarang coba anda hitung nilai rata-rata setiap skor dari (x-$\bar{x}$). Dapatkah anda melakukannya? Ternyata tidak. Ketika kita menambahkan keseluruhan nilai positif dan negatif ternyata hasilnya adalah 0. Kita perlu menghilangkan skor negatif untuk sementara waktu. Caranya, setiap nilai dari (x-x) dipangkatkan dengan dua. Kemudian kita dapat mencari nilai rata-ratanya dan mengakhirnya dengan menghilangkan pangkat dua nilai rata-rata tersebut sehingga kita bisa kembali ke jenis bilangan kardinal yang kita mulai sebelumnya. Perhatikan contoh pada Tabel 3.4

Tabel 3.4: Minat mahasiswa terhadap mata kuliah statistik

x	$\bar{x}$	$(x - \bar{x})$	$(x - \bar{x})^2$
3	6	-3	9
4	6	-2	4
5	6	-1	1
6	6	0	0
6	6	0	0
7	6	1	1
8	6	2	4
9	6	3	9
			Σ = 28

Penjumlahan tersebut dinamakan dengan "penjumlahan pangkat dua dari selisih nilai rata-rata" atau disingkat "jumlah pangkat dua". Sekarang kita dapat menentukan nilai rata-rata dari penjumlahan ini yaitu 3.5 (28:8). Nilai ini disebut dengan varian. Jadi, suatu varian adalah pangkat dua dari selisih nilai dari nilai rata-rata. Dengan demikian rumus varian (untuk populasi) dapat dituliskan sebagai berikut:

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N}$$

Rumus ini dibaca sebagai berikut, "sigma pangkat dua (varian populasi) adalah sama dengan jumlah pangkat dua dari setiap nilai dikurangi nilai rata-rata populasi, dan dibagi dengan jumlah nilai populasi." Rumus tersebut menyatakan bahwa seolah-olah kita telah mengetahui kejadian pada keseluruhan populasi padahal sebenarnya tidak demikian. Pada kenyataannya peneliti bekerja berdasarkan sampel untuk mencari tahu dan mendapatkan gambaran mengenai keseluruhan populasi. Sebagaimana telah disinggung sebelumnya bahwa nilai rata-rata sampel (x) merupakan estimator tidak bias dari rata-rata populasi (μ). Sesuai dengan namanya 'estimator' maka fungsinya hanya untuk membuat estimasi atau perkiraan saja terhadap nilai rata-rata populasi. Sebagai suatu perkiraan maka nilai yang dihasilkan akan mendekati saja tetapi tidak sepenuhnya tepat.

Pada contoh sebelumnya kita mengetahui bahwa setiap skor pada sampel dikurangi dengan nilai rata-rata sampel (x-x) dan kemudian dipangkatkan dua untuk mendapatkan penjumlahan pangkat dua yaitu 28. Harap diperhatikan bahwa nilai x adalah tetap yaitu 6. Jika kita memiliki keseluruhan data populasi, misalnya melalui sensus, maka kita dapat memastikan tidak ada kesalahan dalam penghitungan nilai rata-rata populasi, tetapi dalam suatu sampel hal itu tidaklah mungkin. Setiap kita menghitung varian sampel maka kita akan kehilangan satu derajat kebebasan (*degrees of freedom*) yang setara dengan jumlah nilai rata-rata yang kita hitung untuk membuat perbandingan. Para ahli statistik mengatakan setiap saat kita melakukan estimasi terhadap sampel maka satu derajat kebebasan akan hilang. Untuk menggunakan nilai rata-rata sampel dalam rumus mencari varian kita harus mengubah cara kita menghitung nilai varian yaitu dengan membagi jumlah pangkat dua selisih skor dari nilai rata-rata sampel dengan jumlah kejadian dikurangi satu untuk nilai rata-rata yang kita perkirakan dari sampel.

Dengan demikian, untuk menghitung varian sampel, setiap nilai pada distribusi harus dikurangi dengan nilai rata-ratanya sehingga menghasilkan nilai deviasi yang kemudian dipangkatkan dengan dua. Seluruh nilai deviasi yang telah dipangkatkan dua kemudian dijumlahkan dan dibagi dengan n - 1. Simbol untuk varian adalah s^2, sehingga rumus untuk mencari varian adalah:

$$s^2 = \frac{\sum(x \; - \; x)^2}{n - \; 1}$$

Pada contoh sebelumnya, varian sampel dalam kasus ini akan sedikit berbeda karena pembagi dalam rumus ini adalah n - 1 (8 - 1). Jadi, jika kita bagi 28 dengan 7, kita akan mendapatkan varian (s^2) yaitu 4.

Harap diperhatikan rumus ini digunakan untuk menghitung varian dari suatu distribusi data yang berstatus sebagai sampel yang akan digunakan untuk memperkirakan varian populasi. Tabel 3.5 menunjukkan cara penggunaan rumus ini. Jika peneliti bekerja dengan data populasi maka pembaginya adalah N bukan n-1 sebagaimana rumus yang telah dikemukakan sebelumnya yaitu:

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N}$$

Sebagaimana telah dijelaskan sebelumnya bahwa varian adalah rata-rata pangkat dua dari selisih setiap skor dari nilai rata-rata, tapi nilai ini sulit diinterpretasikan sehingga kita perlu menghilangkan efek dari selisih pangkat dua rata-rata. Bagian ini mengarahkan kita pada bagain akhir cara kita mengukur variabilitas yaitu deviasi standar.

Tabel 3.5: Menghitung Varian

X	X^2	$X - \bar{X}$	$(X - \bar{X})^2$
6	14.65	-8.65	74.8
8	14.65	-6.65	44.2
8	14.65	-6.65	44.2
9	14.65	-5.65	31.9
11	14.65	-3.65	13.3
11	14.65	-3.65	13.3
11	14.65	-3.65	13.3
12	14.65	-2.65	7.0
12	14.65	-2.65	7.0
14	14.65	-0.65	0.4
14	14.65	-0.65	0.4
15	14.65	0.35	0.1
16	14.65	1.35	1.8
18	14.65	3.35	11.2
19	14.65	4.35	18.9
19	14.65	4.35	18.9
21	14.65	6.35	40.3
21	14.65	6.35	40.3
23	14.65	8.35	69.7
25	14.65	10.35	107.1
		Σ	558.6

$$S^2 = \frac{\Sigma(X - \bar{X})^2}{n - 1} = \frac{558.6}{19} = 29.4$$

Deviasi Standar

Rumus varian telah secara luas digunakan peneliti dan berperan penting untuk menghitung dispersi. Rumus ini juga sering digunakan untuk melakukan analisis varian yang umum digunakan dalam statistik inferensial. Namun demikian rumus varian memiliki sedikit ketidaknyamanan yaitu bahwa varian dinyatakan dalam deviasi yang dipangkatkan dari nilai rata-rata dan bukan berdasarkan ukurannya semula. Untuk memperoleh suatu ukuran dispersi yang telah disesuaikan sebagaimana data awalnya maka perlulah kita menggunakan rumus *akar pangkat dua varian* atau disebut juga dengan rumus deviasi standar yang merupakan rumus ketiga dalam mengukur dispersi. Rumus deviasi standar mampu memberikan pengertian yang lebih baik karena dinyatakan dalam unit yang sama sebagaimana pengukuran yang digunakan untuk menghitungnya.

Misal, suatu penelitian terhadap pendapatan rumah tangga menghasilkan varian Rp 9.000.000 dan diinterpretasikan sebagai "9.000.000 rupiah dipangkatkan dua" karena konsep "rupiah dipangkatkan dua" tidak dapat diterima maka digunakan deviasi standar: 3000 rupiah ($\sqrt{9.000.000}$=3.000). Deviasi standar untuk populasi biasanya direpresentasikan dengan simbol σ (sigma kecil) sedangkan deviasi standar untuk sampel menggunakan simbol huruf '*s*' dan dalam hal ini terdapat dua persamaan untuk mencari deviasi standar untuk populasi sebagai berikut:

$$\sigma^2 = \sqrt{\frac{\sum(x-\mu)^2}{N}} \quad \text{atau} \quad \sigma^2 = \sqrt{\frac{\sum x^2}{N} - \mu^2}$$

sedangkan rumus deviasi standar untuk sampel adalah sebagai berikut:

$$s = \sqrt{\frac{\sum(x-x)^2}{n-1}} \quad \text{atau} \quad s = \sqrt{\frac{\sum x^2}{n-1} - \bar{x}^2}$$

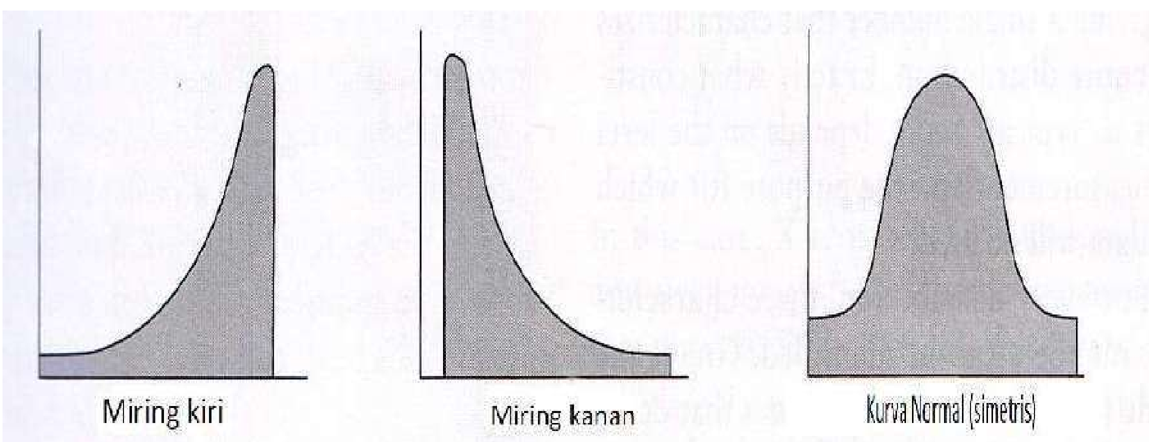
Harap diperhatikan bahwa kedua persamaan tersebut berkaitan dengan dua rumus mengenai varian yang telah kita bahas sebelumnya. Deviasi standar berfungsi menggambarkan jarak suatu nilai dari nilai rata-rata pada suatu distribusi. Dengan kata lain deviasi standar adalah perbedaan rata-rata setiap elemen dari nilai rata-rata yang tersedia pada distribusi.

Deviasi standar terasa sekali manfaatnya dalam menjelaskan hasil tes berdasarkan standar seperti tes kecerdasan (IQ tes) yang memiliki *mean* atau nilai rata-rata 100 dengan deviasi standar 15. Seseorang dengan IQ 115 berarti memiliki deviasi standar 1 di atas *mean* dan seseorang yang memiliki skor IQ 85 berarti memiliki deviasi standar 1 di bawah *mean*.

Distribusi Data

Pembahasan sebelumnya mengenai kurva frekuensi memiliki hubungan erat dengan kurva normal, yaitu suatu kurva yang memiliki bentuk seperti lonceng yang memiliki sejumlah karakteristik yang akan kita bahas lebih mendalam pada bagian selanjutnya di buku ini. Grafik 4 (paling kanan) merupakan contoh kurva normal, namun terkadang terjadi deviasi atau penyimpangan frekuensi sehingga menghasilkan bentuk kurva sebagaimana terlihat pada Grafik 4 bagian kiri dan tengah. Pola-pola penyimpangan semacam ini disebut dengan 'kemiringan' (*skewness*). Dengan demikian, terdapat dua jenis distribusi: distribusi normal dan distribusi tidak normal.

Grafik 4: Berbagai bentuk kurva



Distribusi Normal

Salah satu instrumen terpenting yang digunakan peneliti adalah kurva normal standar yang menjadi dasar bagi banyak distribusi dan statistik lainnya. Kurva normal standar atau disebut juga dengan kurva Gaussian (*The Gaussian Curve*) berdasarkan nama penemunya Karl Friedrich Gauss (1777-1855) adalah suatu distribusi probabilitas yang menjelaskan nilai yang diharapkan yang akan diperoleh melalui penarikan sampel (sampling) secara acak. Gauss dan Marquis de Laplace yang mengembangkan kurva normal standar ini menemukan bahwa umumnya data cenderung mengumpul di sekitar nilai rata-rata dengan semakin sedikit nilai ekstrim yang menjauh dari nilai rata-rata.

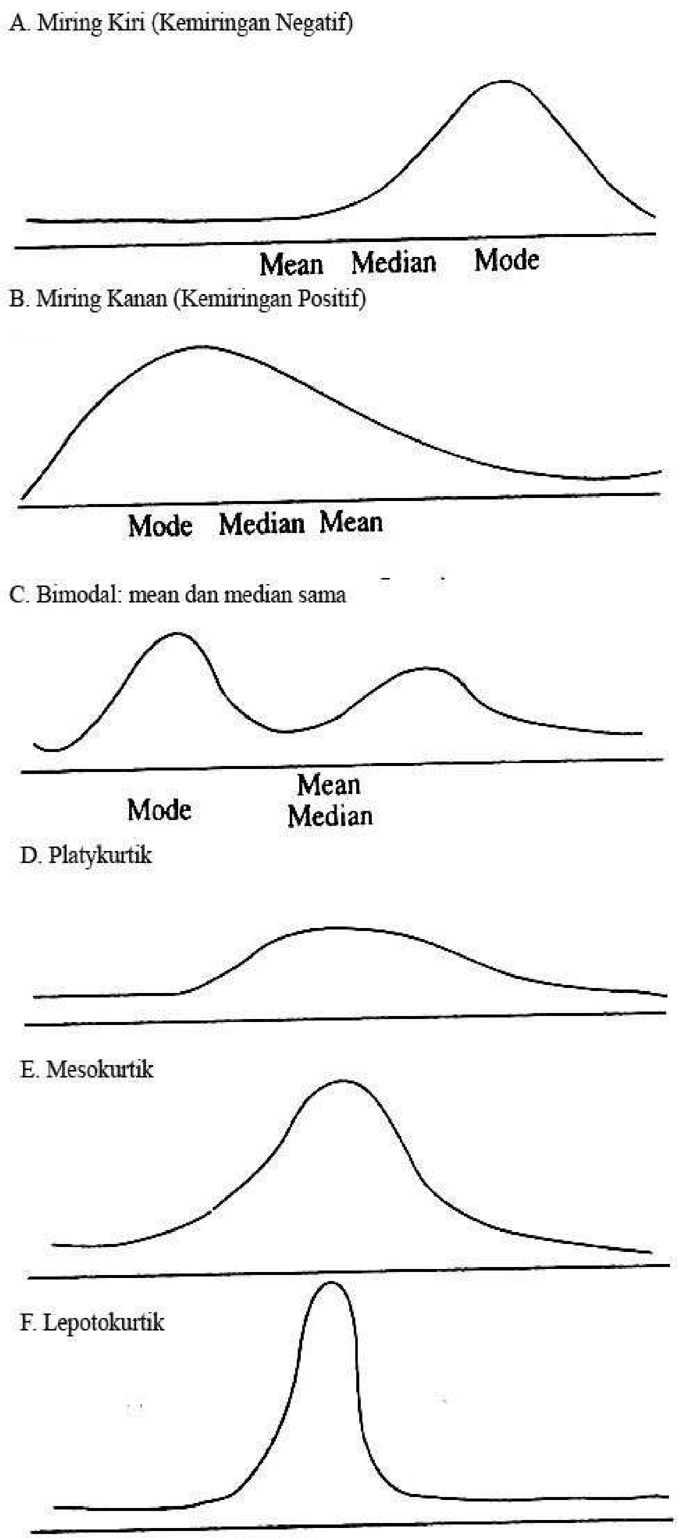
Data yang kita kumpulkan (distribusi data) dapat membentuk kurva normal yang memiliki bentuk simetris, terpusat dan memuncak secara sempurna. Median, *mean* dan modus berada di satu titik pada distribusi dengan kemiringan 0 karena distribusi terpusat secara sempurna. Kurtosis dari kurva normal standar adalah 3. Kita juga dapat melihat bagian ekor dari kurva normal standar tidak pernah menyentuh garis dasar.

Dalam pengamatannya terhadap kurva normal standar, Gauss menemukan sejumlah titik dimana kemiringan kurva mengalami perubahan arah. Ia juga menemukan adanya interval yang identik satu sama lain ketika ia mengikuti kurva baik yang berada di atas atau di bawah nilai rata-rata (*mean*). Gauss menamakan titik-titik ini dengan sebutan deviasi kemiringan kurva normal standar atau disingkat 'deviasi standar' saja.

Nilai rata-rata (*mean*), median dan modus disimbolkan dengan huruf Yunani μ dibaca myu. Deviasi standar disimbolkan dengan huruf Yunani lainnya yaitu σ atau dibaca sigma kecil. Para ahli statistik menyebut pusat distribusi sebagai titik nol (*ground zero*) atau nol saja. Masing-masing deviasi standar ditentukan memiliki lebar sebanyak 1 unit (*1 unit wide*), karenanya σ ditentukan setara dengan 1. Deviasi standar di bawah titik nol memiliki tanda negatif didepannya, dan deviasi standar di atas titik nol memiliki nilai positif di depannya. Peneliti sering kali mengumpulkan data yang memiliki distribusi normal. Tetapi jika distribusi tersebut tidak memiliki nilai rata-rata 0, deviasi standar 1, kemiringan 0, kurtosis 3 dan ekor yang memanjang tak terbatas maka data tersebut bukanlah kurva normal standar. Kita dapat membandingkan data yang kita peroleh dengan kurva normal standar agar dapat membuat keputusan yang benar.

Distribusi tidak normal

Lokasi dimana median, nilai rata-rata dan modus berada pada data dapat menunjukkan sifat dari berbagai pola data. Berbagai pola yang berbeda dapat ditemukan dalam data, dan setiap pola penyebaran data memberikan informasi tertentu. Dalam hal ini, peneliti harus mengetahui bagaimana memahami berbagai bentuk distribusi data tersebut.



Distribusi Miring. Suatu distribusi data yang menjauh dari pusatnya memiliki bentuk miring atau tidak simetris (*skewed*). Kemiringan tidak menjelaskan dimana data berada tetapi menjelaskan seberapa jauh letak ekor dari puncak bukit grafik dimana sebagian besar data berada. Kemiringan mengacu pada konsentrasi skor atau nilai di sekitar nilai tertentu pada sumbu *x*. Jika konsentrasi ini berada pada posisi yang menuju ke arah nilai-nilai rendah, dengan ekor kurva memanjang ke kanan maka kurva semacam ini disebut dengan kurva yang *miring kanan*. Sebaliknya, jika ekor kurva memanjang ke kiri, dan konsentrasi skor pada posisi yang menuju ke arah nilai-nilai tinggi maka kurva semacam ini disebut kurva *miring kiri*.

Jika distribusi data terpusat secara sempurna sehingga menghasilkan kurva simetris maka tidak ada kemiringan sama sekali (koefisien kemiringan 0), atau dengan kata lain, jika bagian kiri dan kanan kurva berbentuk sama (identik) maka ini disebut kurva simetris atau normal. Dengan demikian, masing-masing ekor akan memiliki panjang yang sama.[2]

Jika kemiringan memiliki nilai negatif maka panjang ekor akan berada di bawah nilai rata-rata atau disebut dasar nol (*ground zero*) sehingga kita dapat mengatakan bahwa suatu distribusi ‘miring ke kiri’ atau memiliki ‘kemiringan negatif sebagaimana terlihat pada Grafik 5A. Jika kemiringan memiliki nilai positif maka panjang ekor akan berada di atas nilai rata-rata dasar nol. Grafik 5B menunjukkan kemiringan positif yang berarti distribusi miring ke kanan. Suatu distribusi dapat pula menampilkan kemiringan yang berlebihan karena memiliki dua puncak di dua lokasi berbeda. Distribusi semacam ini disebut distribusi bimodal sebagaimana yang ditunjukkan grafik 5C.

[2] Suatu distribusi data normal tidak miring ke kiri atau kanan. Jika suatu data menunjukkan suatu penyimpangan yang sangat signifikan dari kurva normal maka data tersebut harus diubah sedemikian rupa (akan kita bahas kemudian) agar dapat mencapai suatu distribusi yang lebih normal.

Puncak Distribusi. Selain pengukuran keterpusatan (*centeredness*), kita juga dapat menentukan seberapa tinggi puncak kurva distribusi. Pengukuran puncak distribusi dinamakan kurtosis. Pada distribusi normal sempurna maka distribusi memiliki ketinggian maksimal tiga deviasi standar. Jika puncak distribusi lebih tinggi dari tiga deviasi standar maka kurtosis lebih besar dari tiga. Jika puncak distribusi lebih rendah dari tiga deviasi standar maka kurtosis lebih rendah dari tiga. Grafik 5D menunjukkan suatu distribusi yang memiliki kurtosis lebih rendah dari tiga sehingga sangat datar. Distribusi semacam ini disebut dengan platykurtik. Grafik 5E menunjukkan suatu distribusi dengan puncak yang tidak terlalu rendah tetapi juga tidak terlalu tinggi (kurtosis 3), distribusi semacam ini disebut mesokurtik. Grafik 5F menunjukkan distribusi dengan pucak yang sangat ekstrim. Model semacam ini disebut leptokurtik. Dengan cara melihat pada kemiringan dan kurtosis ini, peneliti dapat memperoleh gambaran yang baik terhadap data yang diperolehnya.

4.2 <u>Rangkuman</u>

1. Perhitungan dispersi bertujuan untuk mengukur dan menjelaskan seberapa jauh setiap skor pada distribusi data menyebar dari pusatnya. Menghitung dispersi dilakukan dengan tiga cara yaitu dengan menggunakan jarak (*range*), varian dan deviasi standar. Jarak menunjukkan seberapa jauh perbedaan antara nilai atau skor tertinggi dengan skor terendah dalam suatu distribusi data. Penggunaan jarak dinilai cara yang paling lemah untuk menunjukkan penyebaran data dari pusatnya. Deviasi standar dan varian adalah dua cara yang paling lazim digunakan untuk mengukur dispersi. Ide dasar kedua pengukuran ini merupakan penjelasan bahwa setiap skor memiliki deviasi atau jarak tertentu dengan nilai rata-rata suatu distribusi nilai. Varian adalah nilai rata-rata dari deviasi pangkat dua. Deviasi standar adalah akar dari suatu varian yang memberikan ukuran jarak standar dari nilai rata-rata.

2. Untuk menghitung varian dan standar deviasi dari suatu populasi atau sampel, jumlah deviasi pangkat dua harus ditentukan terlebih dahulu. Cara untuk menghitung deviasi pangkat dua antara populasi dan sampel adalah sama, perbedaannya hanya terletak pada notasi bagi nilai rata-rata (untuk populasi 'μ' dan untuk sampel $\bar{x}$) . Cara menghitung deviasi pangkat dua populasi adalah dengan melakukan langkah-langkah sbb:
   a. Tentukan deviasi $(x - \mu)$ untuk setiap skor pada distribusi.
   b. Setiap deviasi dipangkatkan dua
   c. Jumlahkan deviasi pangkat dua.

   Proses ini dapat diringkaskan dalam formula deviasi pangkat dua = $\Sigma(x - \mu)^2$

3. Varian adalah nilai rata-rata deviasi pangkat dua yang diperoleh dengan cara menjumlahkan deviasi pangkat dua dan kemudian membaginya dengan jumlah skor (N). Rumus untuk menghitung varian populasi

$$\sigma^2 = \frac{\sum (x - \mu)^2}{N}$$

   sedangkan rumus untuk menghitung varian sampel adalah:

$$S^2 = \frac{\sum(x - x)}{n - 1}$$

4. Deviasi standar adalah akar dari varian. Rumus untuk deviasi standar populasi adalah:

$$\sigma = \sqrt{\frac{\sum(x - \mu)^2}{N}}$$

   sedangkan rumus untuk menghitung varian sampel adalah:

$$S = \sqrt{\frac{\sum(x - x)^2}{n - 1}}$$

5. Menambahkan suatu nilai untuk setiap skor dalam suatu distribusi tidak akan mengubah deviasi standar. Mengalikan setiap skor dengan suatu nilai tertentu akan menyebabkan nilai deviasi standar juga meningkat sebesar nilai deviasi standar dikalikan dengan nilai tertentu itu.

6. Data statistik dapat mengalami penyimpangan (bias) yang berarti nilai rata-rata data statistik sampel tidak mewakili secara tepat parameter populasi. Data yang bias berarti data bersangkutan lebih tinggi atau lebih rendah dari parameter populasi. Data statistik yang tidak bias berarti nilai rata-rata data statistik sampel merupakan representasi yang tepat dari parameter populasi.

4.3 Soal Latihan

1. Jelaskan secara singkat apa yang diukur oleh deviasi standar dan apa yang diukur oleh varian?
2. Pada suatu populasi yang terdiri dari $N = 4$, tiga individu pertama menunjukkan nilai deviasi masing-masing +2, +4, dan -1. Tentukan nilai deviasi individu ke-empat?
3. Berapakah deviasi standar dari seperangkat nilai $N = 5$ skor berikut: 10, 10, 10, 10, 10? (Jawab pertanyaan hanya berdasarkan definisi deviasi standar, tanpa melakukan kalkulasi).
4. Hitung varian untuk populasi berikut $N = 5$, skor: 4, 0, 7, 1, 3.
5. Suatu sampel yang memiliki $n = 7$ skor diambil dari populasi. Skor tersebut adalah 1, 6, 4, 3, 8, 7, 6. Carilah varian dan standar deviasinya.
6. Jika suatu populasi memiliki nilai rata-rata $\mu = 50$ dan deviasi standar $\sigma = 10$, apakah suatu nilai $x = 58$ meruapakan nilai ekstrim? Bagaimana jika deviasi standar $\sigma = 3$
7. Suatu populasi memiliki nilai rata-rata $\mu = 70$ dan deviasi standar $\sigma = 5$.
   a. Jika setiap skor pada distribusi ditambahkan nilainya sebanyak 10 poin, berapa sekarang nilai rata-rata populasi dan deviasi standarnya?

b. Jika setiap skor pada populasi dikalikan dengan dua, berapa sekarang nilai rata-rata populasi dan deviasi standarnya?

Jawaban

1. Deviasi standar mengukur jarak standar dari nilai rata-rata sedangkan varian mengukur jarak rata-rata pangkat dua dari nilai rata-rata.
2. Penjumlahan keseluruhan nilai deviasi pada populasi haruslah nol. Penjumlahan dari tiga nilai deviasi pertama adalah 5 (+2 +4 -1) maka nilai deviasi keempat haruslah -5.
3. Karena tidak terdapat variasi pada nilai (semua skor 10) maka deviasi standar adalah nol.
4. Perhatikan tabel berikut untuk menghitung varian

Skor	*mean*	Deviasi	Dev. Pangkat dua
x	μ	$x - \mu$	$(x - \mu)^2$
4	3	1	1
0	3	-3	9
7	3	4	16
1	3	-2	4
3	3	0	0
N = 5	Σ	0	30

$$\text{Varian} = \sigma^2 = \frac{\Sigma\,(x - \mu)^2}{n} = \frac{30}{5} = 6$$

5. Perhatikan tabel berikut ini untuk mengolah data tersebut:

Skor	mean	Deviasi	Dev. Pangkat dua
x	μ	$x - \mu$	$(x - \mu)^2$
1	5	-4	16
6	5	1	1
4	5	-1	1
3	5	-2	4
8	5	3	9
7	5	2	4
6	5	1	1
n = 7	Σ	0	36

Varian sampel, $S^2 = \frac{\Sigma(x - x)}{n - 1} = \frac{36}{7-1} = 6$

Standar deviasi, S = $\sqrt{\frac{\Sigma(x - \bar{x})^2}{n - 1}} = \sqrt{6} = 2.45$

6. Jika σ = 10 maka suatu skor x = 58 akan masuk atau terletak pada wilayah pusat (sentral) distribusi nilai tersebut. Jika σ = 3 maka skor x = 58 merupakan nilai ekstrim karena nilai tersebut berada jauh di luar batas atas nilai rata-rata 53.

7. a. Nilai rata-rata populasi berubah menjadi μ = 80, tetapi deviasi standar akan tetap σ = 5
   b. Nilai rata-rata populasi berubah menjadi μ = 140, dan deviasi standar menjadi σ = 10

MODUL KULIAH 3
STATISTIK SOSIAL TERAPAN
KURVA NORMAL & SKOR-Z

Morissan, PhD

Salah satu instrumen penting yang sering digunakan peneliti dalam kalkulasi statistik adalah kurva normal (*the normal curve*). Grafik 4.2 menunjukkan suatu contoh kurva normal dan bagian-bagiannya. Perhatikan bentuk kurva yang sangat simetris dengan titik ketinggian maksium berada pada *mean* yang sekaligus menjadi median dan modus. Perhatikan pula bahwa Grafik 4.2 telah disesuaikan dengan unit skor standar. Kurva yang dinyatakan dengan cara ini disebut dengan kurva normal standar. Para ahli statistik telah mempelajari kurva normal ini secara cermat untuk dapat menjelaskan segala atribut atau kelengkapan yang dimilikinya.

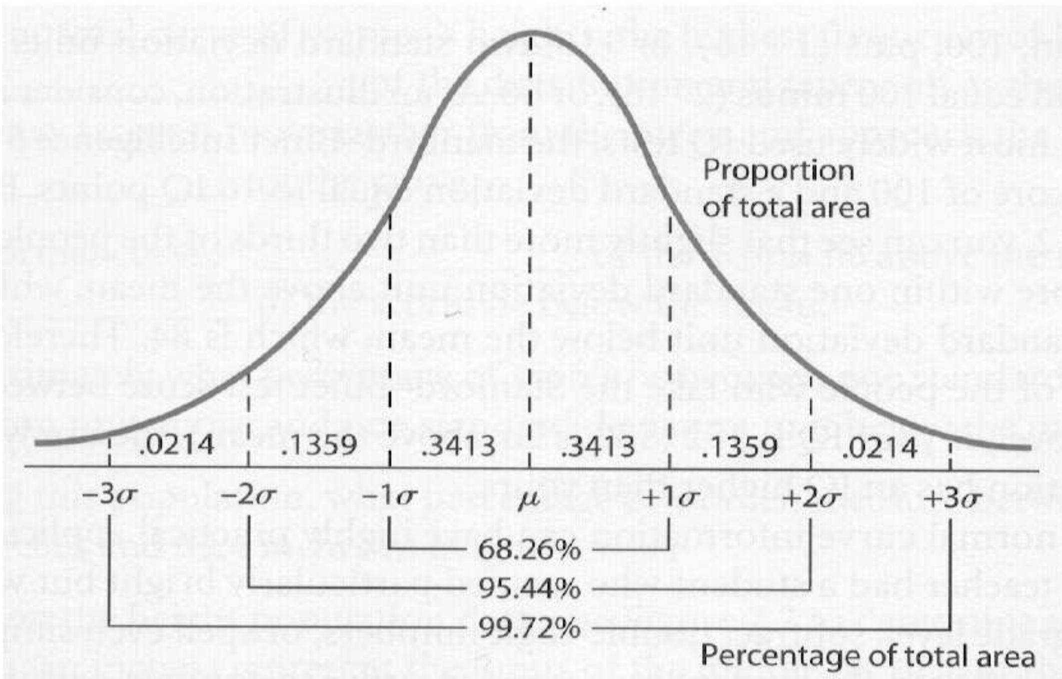
6.1 Wilayah dibawah Kurva

Kelengkapan paling penting yang dimiliki kurva normal adalah suatu wilayah tetap yang berada di bawah kurva yang terletak diantara *mean* dengan setiap titik atau nilai deviasi standar. Wilayah yang berada di bawah suatu bagian tertentu dari kurva mewakili frekuensi dari skor yang berada di bagian tersebut. Grafik 4.2 menunjukkan bagian-bagian wilayah (*proportion of total area*) yang memiliki skor atau nilai tertentu yang berada di bawah kurva normal. Berdasarkan grafik tersebut maka .3413 (atau 34.13%) + .3413 (34.13%) = 68.26% total wilayah (skor) berada diantara deviasi standar (σ) yaitu -1σ dan +1σ dari *mean* dan sekitar 95% atau tepatnya 95.44% (13.59%+34.13%+34.13%+13.59%) terletak diantara deviasi standar -2σ dan +2σ dan seterusnya.

Kurva normal juga memiliki fungsi kepadatan (*density function*) yang berfungsi menunjukkan berapa persen luas wilayah di bawah kurva normal, dari titik nol ke luar ke suatu titik tertentu lainnya. Sebagaimana yang dapat kita lihat, sekitar dua per tiga (2/3) distribusi berada pada wilayah dimulai dari 1σ di bawah titik nol hingga 1σ di atas titik nol dengan total wilayah yang tercakup seluas 68.26%.

Suatu wilayah yang mencakup 95.44% di bawah kurva normal terletak pada -2σ hingga +2σ. Hampir keseluruhan wilayah di bawah kurva normal, hanya kurang 0.28% saja untuk mencapai 100%, terletak pada -3σ hingga +3σ. Fungsi kepadatan ini dapat dilihat pada tabel kurva normal yang terdapat pada hampir semua buku tentang statistik (lihat appendik).

Grafik 4.2 : Kurva normal dan bagian-bagiannya



Pengetahun mengenai wilayah di bawah kurva dan distribusi normal memungkinkan peneliti untuk melakukan prediksi atau perkiraan yang berguna. Dengan menghitung *mean* dan deviasi standar kurva normal (distribusi normal), peneliti dapat menghitung nilai atau skor standar atau disebut juga dengan skor-*z* (*z-scores*) dari setiap distribusi data. Skor standar memungkinkan peneliti untuk membandingkan berbagai skor atau pengukuran yang diperoleh dengan menggunakan berbagai metode yang sama sekali berbeda. Hal ini dimungkinkan karena seluruh penghitungan skor standar berdasarkan standar pengukuran yang sama; semua penghitungan memiliki *mean* 0 dan standar deviasi 1.

Penghitungan skor standar atau skor-*z* selain sangat bermanfaat, juga mudah dihitung dan mudah diinterpretasikan dan menjadi salah satu instrumen statistik yang paling luas digunakan dalam berbagai kegiatan riset. Namun demikian, peneliti biasanya tidak selalu beruntung memperoleh data yang memiliki *mean* atau nilai rata-rata 0 dan deviasi standar 1. Namun kurva normal masih bisa digunakan untuk membuat keputusan dengan cara mengubah data ke dalam 'skor-*z*' (*z* scores) yang memungkinkan kita menggunakan kurva normal untuk membantu membuat keputusan, skor-*z* ini kadang-kadang dinamakan 'skor standar' (*standard scores*).

6.2 Rumus Skor-*z*

Skor-*z* memungkinkan kita menyajikan skor data sebagai unit dibawah kurva normal. Jadi, kita dapat menggunakan skor-*z* untuk membandingkan berbagai pola umum skor atau nilai. Rumus untuk mencari skor-*z* adalah sebagai berikut:

$$z = \frac{\mathrm{K} - \mu}{\sigma} \quad \text{atau} \quad z = \frac{\mathrm{K} - \bar{\mathrm{K}}}{\mathrm{S}}$$

Interpretasi mudah dilakukan karena setiap skor menunjukkan berapa banyak unit-unit deviasi standar berada di atas atau di bawah *mean* data. Penghitungan skor-*z* dan kemampuannya untuk membandingkan berbagai nilai sebagai hasil dari berbagai pengukuran atau metode yang berbeda dapat ditunjukkan melalui ilustrasi singkat berikut ini.

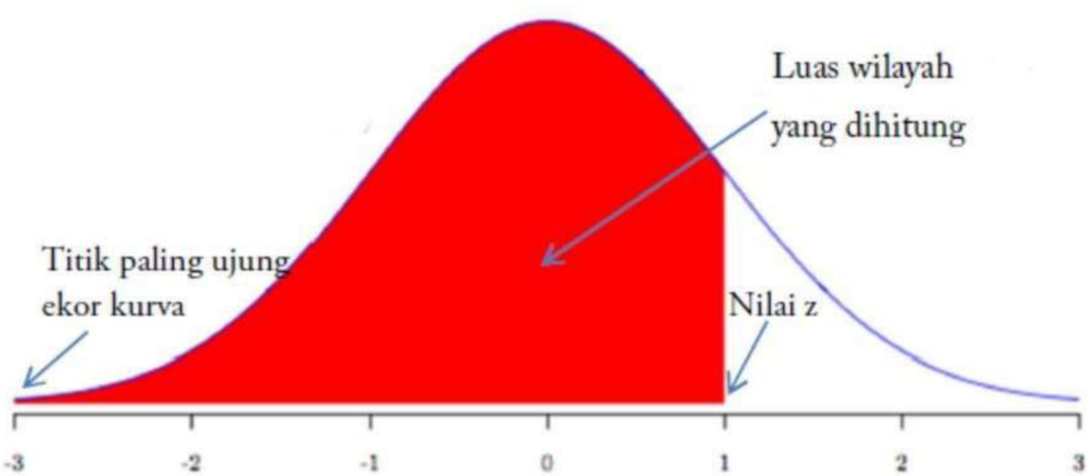
Misal, suatu penelitian menunjukkan bahwa rata-rata durasi orang menonton televisi adalah dua jam per hari dengan standar deviasi 0.5. Berapa banyak populasi yang menonton televisi antara dua hingga 2.5 jam per hari? Untuk menjawab pertanyaan ini maka kita perlu mengubah skor 'mentah' tersebut menjadi skor-*z*:

$$z = \frac{\mathrm{K} - \bar{\mathrm{K}}}{\mathrm{S}}$$

$$z = \frac{2 - 2}{0.5} = 0 \quad \text{dan} \quad \frac{2.5 - 2}{0.5} = 1.00$$

Dengan demikian, skor-*z* untuk nilai 2 jam menonton TV adalah 0 karena skor-*z* untuk nilai rata-rata adalah selalu 0, dan skor-*z* bagi 2.5 jam menonton TV adalah 1.00. Grafik 4.2 menunjukkan sekitar 34% wilayah di bawah kurva terletak di antara *mean* dan deviasi standar 1. Dengan demikian 34% populasi menonton televisi antara 2 dan 2.5 jam televisi setiap harinya. Selanjutnya, untuk mengetahui jumlah populasi yang menonton TV lebih dari 3 jam per hari. Sekali lagi, langkah pertama adalah menerjemahkan skor mentah menjadi skor-*z* yang menghasilkan nilai 2 (3-2/0.5). Dengan demikian, tiga jam menonton TV berhubungan dengan skor-*z* dengan nilai 2. Grafik 4.2 menunjukkan sekitar 98% wilayah di bawah kurva terletak di bawah skor 2 (50% wilayah berada di setengah bagian kurva ditambah dengan 48%, atau tepatnya 47.72%, wilayah yang dimulai dari *mean* hingga nilai 2). Dengan demikian hanya sekitar 2.28% populasi yang menonton lebih dari tiga jam per hari.

Tabel skor-*z* pada lampiran buku ini memuat daftar luas wilayah yang berada di bawah kurva normal yang dimulai dari titik paling ujung ekor kurva normal yang berada di wilayah negatif dengan berbagai nilai *z* pada sumbu horizontal. Untuk menggunakan tabel ini, kita harus memperhatikan kolom dan baris pada tabel yang memuat sejumlah nilai atau skor yang sudah standar. Misal, umpamakan kita memiliki skor-*z* dengan nilai 1.79. Berapakah luas wilayah di bawah kurva normal mulai dari titik paling ujung ekor kurva normal yang berada di wilayah negatif hingga *z* = 1.79. Karena 1.79 merupakan penjumlahan 1.7 + 0.09 maka pada tabel skor-*z*, kita harus menemukan baris yang bertuliskan 1.7, dan kemudian temukan kolom berlabel angka 0.09. Pada persimpangan antara baris 1.7 dengan kolom 0.09 terdapat 0.9633 yang berarti luas wilayah yang tercakup adalah 96.33% (Lihat Tabel 4.1). Jika kita ingin mengetahui luas wilayah antara *mean* dengan nilai *z* = 1.79 maka kita tinggal mengurangkan 0.9633 dengan skor-z untuk *mean* yaitu 0. Pada tabel, wilayah yang tercakup mulai dari titik paling ujung ekor kurva normal yang berada di wilayah negatif hingga *z* = 0 adalah 0.5000. Luas wilayah antara *mean* dengan nilai *z* = 1.79 adalah 0.9633 – 0.5000 = 0.4633 atau sekitar 46%, maka luas wilayah di bawah kurva normal mulai dari titik paling ujung ekor kurva normal yang berada di wilayah negatif hingga *z* = 1.79 adalah 50% + 46% = 96% (lihat Grafik 4.3)



Grafik 4.3 : Menghitung luas wilayah tertentu di kurva normal

Contoh lain, suatu penelitian ingin mengetahui pengaruh tingkat kecerdasan (IQ) seseorang terhadap tanggapannya ketika menerima pesan. Skor uji tingkat kecerdasan yang populer yaitu Stanford-Binet IQ test tersebar secara normal dalam suatu populasi dengan nilai rata-rata (μ) 100 dan suatu deviasi standar populasi (σ) 15. Seandainya anda memiliki IQ 115. Seberapa jauh skor IQ anda melampaui IQ orang lain. Kita dapat memasukkan nilai tersebut ke dalam rumus:

$$z = \frac{\mathrm{K} - \mu}{\sigma} = \frac{115 - 100}{15} = \frac{15}{15} = 1.00$$

Jadi, seberapa banyak skor IQ anda mengungguli skor lainnya? Anda perlu melihat ke tabel skor-*z* dimana anda harus mencari kolom z dan lihat pada baris dengan nilai 1.0. Untuk menemukan nilai desimal kedua, lanjutkan penelusuran pada baris paling atas pada kolom ".00". Persimpangan kolom dan baris ini menunjukkan nilai 0.8413 yang berarti luas wilayah yang tercakup mulai dari ujung ekor negatif kurva normal hingga z = 1.00 adalah 84.13%. Hal ini berarti IQ anda lebih tinggi dari keseluruhan 84.13% populasi. Skor anda lebih besar dari 84% skor lainnya.

Seandainya anda tidak memiliki nilai rata-rata populasi (μ) atau deviasi standar populasi (σ)? Tidak masalah. Nilai rata-rata sampel ($\bar{X}$) merupakan estimator atau penaksir tidak bias dari nilai rata-rata populasi (μ). Deviasi standar sampel (s) adalah 'penaksir tidak bias' (dengan syarat rumus yang benar digunakan bagi deviasi standar sampel) dari deviasi standar populasi. Anda dapat mengganti nilai rata-rata sampel dan deviasi standar dengan nila rata-rata populasi dan deviasi standar untuk menghitung jawababnya. Kembali kita gunakan contoh sebelumnya mengenai penilaian siswa di kelas. Andaikan hasil tes rata-rata siswa di kelas adalah 50 sedangkan nilai tes anda adalah 61, dan deviasi standar 10. Berapa nilai yang anda dapatkan? Untuk menghitung kita masukkan nilai-nilai tersebut ke dalam rumus:

Tabel 4.1
Tabel skor-z untuk menghitung luas wilayah di bawah kurva normal

	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	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	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	0.7580	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	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	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	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	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	0.9265	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	0.9649	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
[illegible]	0.9918	0.9920	0.9922	0.9925	[illegible]	[illegible]	0.9931	0.9932	[illegible]	[illegible]

$$z = \frac{K - \bar{K}}{s} = \frac{61 - 50}{10} = \frac{11}{10} = 1.10$$

Tabel 4.2 menunjukkan perumpamaan suatu data hasil tes mahasiswa dari dua kelas berbeda dimana setiap kelas memiliki 20 mahasiswa. Hasil tes mahasiwa kelas pertama menunjukkan skor terendah 68 dan tertinggi 84. Dengan demikian jarak atau *range* kedua nilai tersebut adalah 16 (84-68), sedangkan kelas kedua menunjukkan skor terendah 38 dan tertinggi 73, dengan demikian jarak kedua nilai tersebut adalah 35 (73-38). Perbedaan skor diantara soal yang berbeda, kemampuan mahasiswa dalam menyelesaikan soal yang tidak sama, dan kemampuan pengajar menyampaikan materi kuliah.

Nilai rata-rata atau *mean* mahasiswa kelas pertama adalah 74.6 dengan deviasi standar 4.9 sedangkan mean mahasiswa kelas kedua adalah 43.9 dengan deviasi standar 7.5. Tabel 4.2 menunjukkan kinerja mahasiswa kelas pertama dimana Ali berada dan mendapatkan skor 73. Perhatikan nilai yang diperoleh Ali merupakan skor rata-rata di kelas tersebut. Tabel juga menunjukkan kinerja kelas kedua dimana Budi berada dan, sebagaimana Ali, juga mendapatkan skor 73. Namun skor yang diperoleh Budi berada di atas rata-rata dan bahkan tertinggi untuk kelas terse

Tabel 4.2 menunjukkan perumpamaan suatu data hasil tes mahasiswa dari dua kelas berbeda dimana setiap kelas memiliki 20 mahasiswa. Hasil tes mahasiwa kelas pertama menunjukkan skor terendah 68 dan tertinggi 84. Dengan demikian jarak atau *range* kedua nilai tersebut adalah 16 (84-68), sedangkan kelas kedua menunjukkan skor terendah 38 dan tertinggi 73, dengan demikian jarak kedua nilai tersebut adalah 35 (73-38). Perbedaan skor diantara soal yang berbeda, kemampuan mahasiwa dalam menyelesaikan soal yang tidak sama, dan kemampuan pengajar menyampaikan materi kuliah.

Nilai rata-rata atau *mean* mahasiswa kelas pertama adalah 74.6 dengan deviasi standar 4.9 sedangkan mean mahasiswa kelas kedua adalah 43.9 dengan deviasi standar 7.5. Tabel 4.2 menunjukkan kinerja mahasiswa kelas pertama dimana Ali berada dan mendapatkan skor 73. Perhatikan nilai yang diperoleh Ali merupakan skor rata-rata di kelas tersebut. Tabel juga menunjukkan kinerja kelas kedua dimana Budi berada dan, sebagaimana Ali, juga mendapatkan skor 73. Namun skor yang diperoleh Budi berada di atas rata-rata dan bahkan tertinggi untuk kelas tersebut.

Ketika setiap skor pada distribusi diubah menjadi skor-*z* maka skor tersebut menunjukkan suatu karakteristik tertentu. Setiap skor yang berada di bawah *mean* menjadi skor-*z* negatif, dan setiap skor yang terletak di atas *mean* menjadi skor-*z* positif. Nilai *mean* skor-*z* dari suatu distribusi adalah selalu 0. Hal ini juga berlaku terhadap mahasiswa yang memiliki skor-*z* yang sama dengan *mean*. Sebagaimana telah dijelaskan sebelumnya, varian dan deviasi standar dari suatu distribusi skor-*z* keduanya adalah selalu 1.0. Dalam hal ini, skor-*z* selalu dinyatakan dalam berbagai unit standar deviasi. Jadi, suatu skor-*z* 3.00 berarti skor tersebut memiliki 3 unit deviasi standar di atas *mean*.

	Data Skor-z						
	Kelas Pertama			Kelas Kedua			
	Skor	*Penghitungan*	*Skor z*	*Skor*	*Penghitungan*	*Skor z*	
Nilai B	84	(84 - 74.6)/4.9	1.9	73	(73 – 43.9)/7.5	3.9	**Nilai A**
	81	(81 - 74.6)/4.9	1.3	50	(50 – 43.9)/7.5	0.8	Nilai C
	81	(81 - 74.6)/4.9	1.3	50	(50 – 43.9)/7.5	0.8	
Nilai C	79	(79 – 74.6)/4.9	0.9	47	(47 – 43.9)/7.5	0.4	
	79	(79 – 74.6)/4.9	0.9	46	(46 – 43.9)/7.5	0.3	
	79	(79 – 74.6)/4.9	0.9	45	(45 – 43.9)/7.5	0.2	
	78	(78 – 74.6)/4.9	0.7	43	(43 – 43.9)/7.5	-0.1	
	77	(77 – 74.6)/4.9	0.5	43	(43 – 43.9)/7.5	-0.1	
	77	(77 – 74.6)/4.9	0.5	42	(42 – 43.9)/7.5	-0.2	
	75	(75 – 74.6)/4.9	0.1	41	(41 – 43.9)/7.5	-0.4	
	73	(73 – 74.6)/4.9	-0.3	41	(41 – 43.9)/7.5	-0.4	
	71	(71 – 74.6)/4.9	-0.7	41	(41 – 43.9)/7.5	-0.4	
	71	(71 – 74.6)/4.9	-0.7	40	(40 – 43.9)/7.5	-0.5	
	71	(71 – 74.6)/4.9	-0.7	40	(40 – 43.9)/7.5	-0.5	
	70	(70 – 74.6)/4.9	-0.9	40	(40 – 43.9)/7.5	-0.5	
	70	(70 – 74.6)/4.9	-0.9	40	(40 – 43.9)/7.5	-0.5	
	70	(70 – 74.6)/4.9	-0.9	40	(40 – 43.9)/7.5	-0.5	
Nilai D	69	(69 – 74.6)/4.9	-1.1	39	(39 – 43.9)/7.5	-0.6	
	68	(68 – 74.6)/4.9	-1.3	38	(38 – 43.9)/7.5	-0.8	
	68	(68 – 74.6)/4.9	-1.3	38	(38 – 43.9)/7.5	-0.8	
			Mean 74.6 S 4.9	Mean 43.9 S 7.5			

Soal Latihan

1. Lihat tabel skor-z dan tentukan proporsi (luas wilayah) dari suatu distribusi normal yang berhubungan dengan setiap skor-z berikut ini, sertakan pula grafik kurva normal yang menunjukkan proporsi wilayah yang hendak dihitung luasnya.

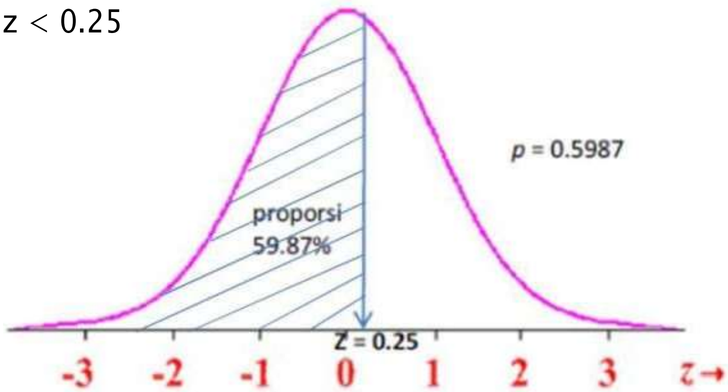
   a) $z < 0.25$ b) $z > 0.80$ c) $z < -1.50$ d) $z > -0.75$

2. Lihat tabel skor-z dan tentukan lokasi skor-z yang membagi distribusi normal di bawah ini (gambarkan pula grafik kurva normalnya):
   a) skor-z yang memisahkan 20% skor tertinggi dari skor yang lebih rendah.
   b) skor-z yang memisahkan 60% skor tertinggi dari skor lainnya yang lebih rendah.
   c) skor-z yang memisahkan 70% skor wilayah tengah dari skor lainnya
3. Suatu perusahaan pengelola jalan bebas hambatan melakukan penelitian untuk mengetahui kecepatan kendaraan yang melewati salah satu ruas jalan tol. Data yang diperoleh menunjukkan suatu distribusi normal dengan kecepatan rata-rata kendaraan $\mu = 58$ Km/ jam dan deviasi standar $\sigma = 10$. Berdasarkan data tersebut, berapa proporsi kendaraan yang melaju dengan kecepatan antara 55 dan 65 Km/jam?
4. Suatu distribusi normal memiliki nilai rata-rata $\mu = 60$ dan suatu deviasi standar $\sigma = 12$, tentukan probabilitas setiap nilai X berikut ini
   a) $p(X > 66)$
   b) $p(X < 75)$
   c) $p(X < 57)$
   d) $p(48 < X < 72)$

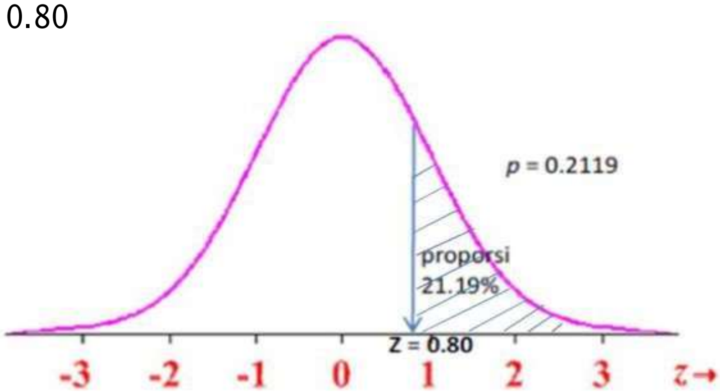
5. Nilai tes kemampuan bahasa Inggris TOEFL terhadap calon karyawan menunjukkan nilai rata-rata $\mu = 500$ dan deviasi standar $\sigma = 100$
   a. Jika perusahaan hanya akan menerima sebanyak 60% peserta dengan nilai tertinggi. Berapa nilai minimal agar bisa diterima?
   b. Berapa nilai minimal yang harus dimiliki agar seorang peserta dapat masuk dalam kategori 10% dengan nilai tertinggi?
   c. Berapa skor terendah yang harus dimiliki agar masuk dalam kategori kelompok nilai menengah yang jumlahnya 50%? Berapa nilai tertinggi untuk kategori ini?

Jawaban

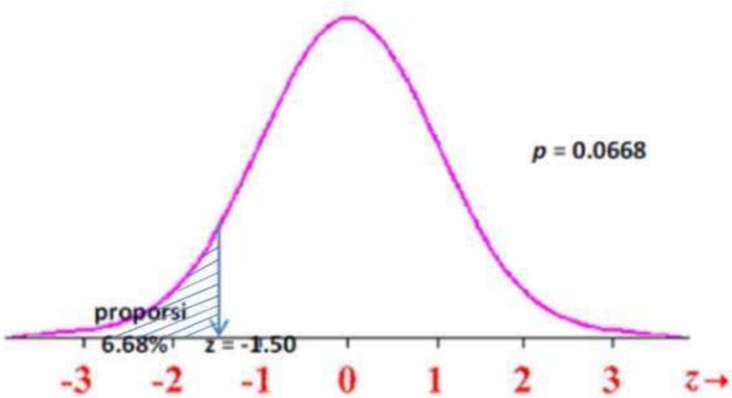
1. a) $z < 0.25$



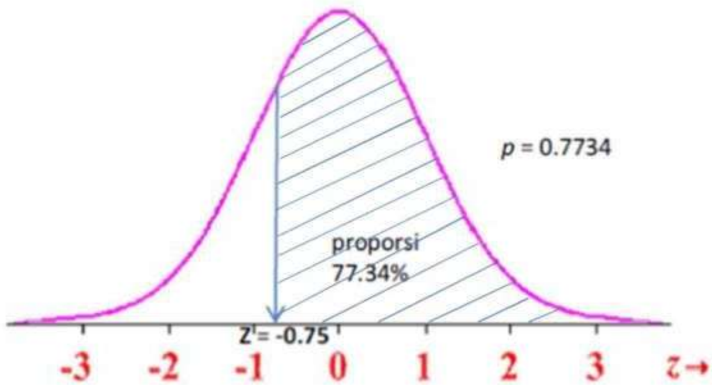
b) $z > 0.80$



c) $z < -1.50$



d) $z > -0.75$



2 a) Untuk menentukan lokasi skor-z yang memisahkan 20% skor tertinggi dari skor lainnya yang lebih rendah, kita harus melihat tabel skor-z pada proporsi yang mencakup 80% atau 0.8000 wilayah kurva. Pada tabel, nilai yang paling mendekati adalah 0.7995 atau 79.95% yang merupakan titik pertemuan skor-z untuk kolom 0.8 dan baris atas 0.04 sehingga lokasi skor z adalah 0.84 (0.8 + 0.04).

e) Untuk menentukan lokasi skor-z yang memisahkan 60% skor tertinggi dari skor lainnya, kita harus melihat tabel skor-z pada proporsi yang mencakup 40% atau 0.4000 wilayah kurva. Pada tabel, nilai yang paling mendekati adalah 0.4013 atau 40.13% yang merupakan titik pertemuan skor-z untuk kolom -0.2 dan baris atas 0.05 sehingga lokasi skor z adalah -0.25 (-0.2 + (-0.05)).

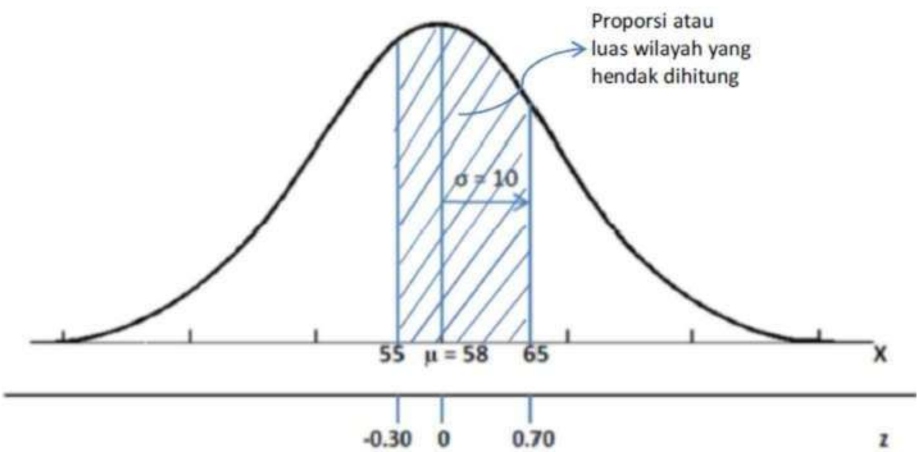
f) Untuk menentukan lokasi skor-z yang memisahkan 70% wilayah tengah dengan wilayah lainnya, kita harus menghitung sisa wilayah pada kurva yaitu 100% - 70% = 30% selanjutnya membagi 30% menjadi dua bagian yang masing-masing 15% untuk di kedua ujung kurva, masing-masing untuk wilayah negatif dan positif. Pada tabel, nilai yang paling mendekati 15% adalah 0.1492 atau 14.92% yang merupakan titik pertemuan skor-z untuk kolom -0.1 dan baris atas 0.04 sehingga lokasi skor z adalah -1.04. Karena kurva norma bersifat simetris maka skor z untuk wilayah positif adalah +1.04.

3. Dengan menggunakan notasi probabilitas, kita dapat menyatakan pertanyaan tersebut dengan persamaan $p(55 < X < 65) = ?$

Distribusi kecepatan kendaraan dapat ditunjukkan pada grafik kurva normal XXX dimana bagian kurva yang diarsir (garis-garis) adalah nilai kecepatan yang ingin diketahui. Langkah pertama adalah menentukan skor-z yang berhubungan dengan nilai X pada kedua nilai interval sbb:

Untuk $X_1 = 55$: $z = \frac{K-\mu}{\sigma} = \frac{55-58}{10} = \frac{-3}{10} = -0.30$

Untuk $X_2 = 65$: $z = \frac{K-\mu}{\sigma} = \frac{65-58}{10} = \frac{7}{10} = 0.70$



Untuk mengetahui proporsi atau luas wilayah yang diarsir, kita terlebih dahulu harus menghitung luas wilayah dari ujung kurva sebelah kiri hingga ke nilai z = 0.70. Pada tabel skor-z kita menemukan nilai 0.7580 atau 75.80%. Selanjutnya, kita hitung pula luas wilayah dari ujung kurva sebelah kiri hingga ke nilai z = -0.30. Pada tabel skor-z kita menemukan nilai 0.3821 atau 38.21%. Dengan demikian luas wilayah yang diarsir $p(55 < X < 65) = p(-0.30 < z < +0.70) =$ 75.80% - 38.21% = 37.59%.

1. *a)* $p(X > 66)$: $z = \frac{K-\mu}{\sigma} = \frac{66-60}{12} = \frac{6}{12} = 0.5$ maka $p(X > 66) = p(z > 0.5)$ pada tabel, proporsi z = 0.5 adalah 69,15%. Proporsi untuk z > 0.5 adalah 100% - 69.15% = 30.85%

   a) $p(X < 75)$: $z = \frac{K-\mu}{\sigma} = \frac{75-60}{12} = \frac{15}{12} = 1.25$ maka $p(X < 75) = p(z < 1.25)$ pada tabel, proporsi z = 1.25 adalah 89.44%. Proporsi untuk z < 1.25 adalah 89.44%.

b) $p(X < 57)$: $z = \frac{K-\mu}{\sigma} = \frac{57-60}{12} = \frac{-3}{12} = -0.25$ maka $p(X < 57) = p(X < -0.25) = 40.13\%$

c) $p(48 < X < 72)$ = $z_1 = \frac{K-\mu}{\sigma} = \frac{48-60}{12} = \frac{-12}{12} = -1$, karena kurva memiliki bentuk simetris maka $z_2 = +1$. Luas wilayah (proporsi) z2 adalah 84,13%, proporsi z2 adalah 15.87% sehingga $p(48 < X < 72) = p(-1 < X < +1)$ = 84.13% - 15.87% = 68.26%

2. a) Nilai rata-rata TOEFL adalah μ = 500, σ = 100. Jika perusahaan hanya akan menerima sebanyak 60% peserta dengan nilai tertinggi maka batas nilai terendah ada pada 40%, dan pada tabel skor-z, 40% menunjuk pada z = -0.25. Untuk mencari nilai terendah adalah X = z.σ + μ = -0.25. 100 + 500 = -25 + 500 = 475.

b) Jika perusahaan hanya akan menerima sebanyak 10% peserta dengan nilai tertinggi maka batas nilai terendah ada pada 90%, dan pada tabel skor-z, 90% menunjuk pada z = 1.28. Untuk mencari nilai terendah adalah X = z.σ + μ = 1.28. 100 + 500 = 128 + 500 = 628.

c) Nilai tengah 50% terdiri atas 25% di bawah nilai rata-rata dan 25% di atas nilai rata-rata, maka batas nilai terendah ada pada 25% dan pada tabel skor-z, 25% menunjuk pada z = -0.67 maka nilai terendah adalah X = z.σ + μ = -0.67. 100 + 500 = 433. Karena kurva normal bersifat simetris maka batas nilai tertingi adalah +0.67, maka batas nilai tertinggi adalah X = z.σ + μ = +0.67. 100 + 500 = 567.

6.3 Penerapan Skor-*z*

Pengukuran dengan menggunakan skor-*z* sering digunakan pada semua jenis riset. Skor-*z* memungkinkan peneliti untuk membandingkan secara langsung kinerja berbagai subjek yang berbeda yang diuji dengan menggunakan metode pengukuran yang berbeda namun dengan asumsi distribusinya sama. Misal, hasil panen sektor perkebunan pisang pada tahun tertentu tercatat sebanyak 24 kwintal per hektar. Hasil panen ini lebih tinggi dari rata-rata hasil panen tahunan sebanyak 22 kwintal per hektar dengan deviasi standar 10 (lihat Tabel 4.3). Pada tahun yang sama, hasil panen sektor perkebunan jeruk menghasilkan 18 kwintal per hektar dibandingkan dengan rata-rata hasil panen tahunan jeruk sebanyak 16 kwintal per hektar dengan deviasi standar 8. Kinerja sektor perkebunan manakah yang lebih baik? Rumus nilai standar menunjukkan skor-*z*

0.20 untuk pisang [(24-22)/10] dan 0.25 untuk jeruk [(18-16)/8]. Dengan demikian, sektor perkebunan jeruk memiliki kinerja lebih baik.

Tabel 4.3

	Pisang	Jeruk
Rata-rata produksi	22	16
Deviasi standar	10	8
Produksi tahun ini	24	18
Skor-z	0.20	0.25

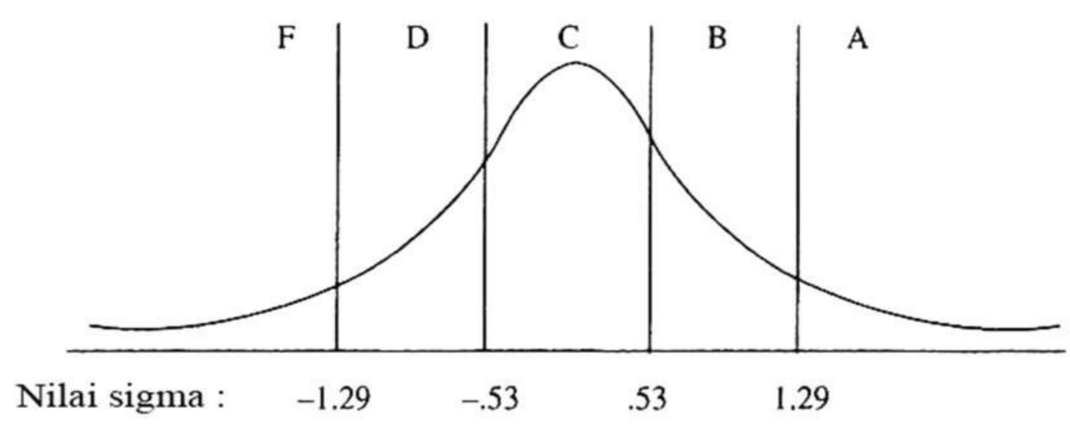
Kita gunakan contoh lain, berapa luas wilayah antara titik tengah kurva dengan skor-*z* yang bernilai -1.32? Menurut tabel skor-z, 0.0934 atau 9.34% wilayah kurva tercakup mulai dari ujung ekor negatif hingga z = -1.32. Cara menemukannya, lihat kolom sebelah kiri, cari angka -1.3, dan kemudian cari kolom yang bertuliskan 0.02. Perhatikan wilayah tersebut selalu positif walaupun skor-z dinyatakan dalam nilai negatif. Titik tengah kurva merupakan mean dengan nilai z = 0.5000 atau 50%. Dengan demikian, luas wilayah antara titik tengah kurva dengan skor-z yang bernilai -1.32 adalah 0.5000 - 0.0934 = 40.66%.

Mari kita gunakan contoh sebelumnya mengenai nilai kuliah metode riset dari tabel 5.8. Sebagaimana telah dijelaskan sebelumnya, pada kelas yang memiliki *mean* atau nilai rata-rata 74.6 dengan nilai deviasi standar 4.9, skor terdistribusi secara normal. Pengajar memutuskan memberikan nilai “C” kepada mahasiswa dengan nilai yang masuk ke wilayah tengah (skor z = +1 hingga -1). Skor berapa saja yang akan menerima nilai “C” ini? Pertama, kita harus menemukan batas atas nilai C dengan menggunakan *mean* atau nilai rata-rata kelas yaitu 74.6 dan menambahkannya dengan +1.00 x 4.9 yang menghasilkan 4.9 dan kemudian ditambahkan dengan 74.6 sehingga menghasilkan 79.5. Untuk mengetahui jarak batas bawah, kita harus mengurangi *mean* (74.6) dengan -1.00 x 4.9 = -4.9, sehingga 74.6 – 4.9 = 69.7. Setelah melakukan pembulatan kita menemukan bahwa semua mahasiswa yang memperoleh skor ujian antara 70 dan 80 akan menerima nilai akhir C.

Tabel kurva normal memiliki banyak manfaat antara lain, misalnya, membantu para guru atau pengajar memberi nilai terhadap kinerja siswa di kelas dengan asumsi bahwa nilai mahasiswa terdistribusi secara normal (biasanya memang demikian). Hal ini dilakukan dengan cara membagi-bagi kelompok nilai siswa dari nilai tertinggi hingga terendah berdasarkan skor ujian yang mereka peroleh. Dalam hal ini sebanyak 10% dari keseluruhan siswa dengan nilai tertinggi mendapat nilai A, dan 10% siswa dengan nilai terendah mendapat nilai F. Selanjutnya 20% mahasiswa dengan nilai setelah kedua nilai ekstrim tersebut menerima B dan D, dan sisanya 40% menerima nilai C. Grafik4.4 menunjukkan bagaimana skor yang terdistribusi secara normal diatur ketika pengajar memberi nilai berdasarakan kurva normal. Nilai C yang terletak di tengah-tengah distribusi yang mencakup sekitar 40% wilayah kurva berada dalam jarak $+0.53\sigma$ hingga ke -0.53σ.

Pada tabel kurva normal, simbol *z* pada tabel menunjukkan jumlah sigma atau deviasi standar yang diinginkan. Dengan mengggunakan contoh sebelumnya mengenai nilai siswa di kelas, kita dapat melihat kolom dengan tanda *z* hingga menemukan nilai 0.5 yang menjadi nilai desimal pertama. Selanjutnya baca baris paling atas untuk menemukan nilai .03 yang berarti kita memperoleh nilai desimal kedua sehingga kita mendapatkan nilai 0.53 (0.5 + 0.03).

Grafik 4.4



Pada pertemuan antara kolom dan baris tersebut kita mendapatkan nilai 0.7019. Hal ini berarti 70.19% total wilayah dibawah kurva normal terletak mulai dari ujung ekor negatif kurva hingga ke $+0.53\sigma$. Untuk mengetahui luas wilayah antara 0.53 sigma di bawah titik nol hingga 0.53 sigma di atas titik nol adalah dengan cara menentukan terlebih dahulu luas wilayah mulai dari ujung ekor hingga ke -0.53σ. Kita dapat melihat kolom dengan tanda z hingga menemukan nilai -0.5 yang menjadi nilai desimal pertama. Selanjutnya baca baris paling atas untuk menemukan nilai 0.03 yang berarti kita memperoleh nilai desimal kedua sehingga kita mendapatkan nilai 0.29806. Dengan demikian luas wilayah antara 0.53 sigma di bawah titik nol hingga 0.53 sigma di atas titik nol adalah 0.7019 – 0.29806 = 0.4038. Dengan demikian luas wilayah dari -0.53σ ke $+0.53\sigma$ adalah sekitar 40%. Untuk memperoleh tambahan 20% nilai sekitar C yaitu B dan D, kita harus bergerak keluar menuju ke nilai $\pm$ 1.29σ. Sisa wilayah 10% di kiri dan kanan kurva distribusi merupakan wilayah nilai A dan F.

Penghitungan dengan menggunakan kurva normal sangat penting bagi peneliti karena banyak variabel yang ditemukan dalam penelitian terdistribusi secara normal, atau cukup normal sehingga penyimpangan yang tidak signifikan (kecil) dapat diabaikan. Lebih jauh, penghitungan kurva normal merupakan contoh distribusi sampel probabilitas yang sangat penting dalam statistik inferensial.

6.4 Ringkasan

1. Setiap nilai *X* dapat diubah menjadi skor-z yang berfungsi menunjukkan lokasi *X* secara tepat dalam suatu distribusi. Tanda pada skor-z menunjukkan apakah lokasi *X* terletak di atas nilai rata-rata (positif) atau di bawah nilai rata-rata (negatif). Nilai skor-z menentukan jumlah deviasi standar antara *X* dan μ.
2. Rumus skor-*z* digunakan untuk mengubah nilai *X* menjadi skor-*z*. Rumus skor-*z* untuk populasi adalah: $z = \frac{K - \mu}{\sigma}$ dan untuk sampel adalah $z = \frac{K - \bar{K}}{s}$
3. Untuk mengubah skor-*z* kembali menjadi nilai *X* adalah dengan cara mengubah rumus skor-z di atas terlebih dahulu. Dalam hal ini, rumus mencari nilai X untuk populasi adalah: $X = \mu + z\sigma$ sedangkan untuk sampel adalah: $X = \mu + zs$
4. Jika seluruh distribusi nilai *X* diubah menjadi skor-z maka hasilnya adalah suatu distribusi skor-*z*. Distribusi skor-*z* akan memiliki bentuk yang sama dengan distribusi nilai X awal yang akan memiliki nilai rata-rata 0 dan deviasi standar 1.
5. Ketika membandingkan sejumlah nilai atau skor dari dua distribusi yang berbeda, maka kedua distribusi nilai tersebut hendaknya diubah menjadi skor-z terlebih dahulu. Hal ini dimaksudkan

agar kedua distribusi tersebut dapat dibandingkan satu sama lain karena keduanya telah memiliki parameter yang sama (μ= 0, σ = 1).

6. Jika suatu distribusi nilai telah diubah menjadi skor-*z*, selanjutnya skor-*z* tersebut dapat diubah menjadi suatu distribusi nilai baru dengan nilai rata-rata dan deviasi standar yang telah ditentukan terlebih dahulu.
7. Dalam statistik inferensial, skor-*z* memberikan metode yang objektif untuk menentukan seberapa baik suatu nilai tertentu mewakili populasi. Suatu skor-*z* yang mendekati nol menunjukkan bahwa nilai tersebut mendekati nilai rata-rata populasi sehingga nilai bersifat representatif. Suatu skor-*z* yang melampaui +2.00 (atau -2.00) merupakan nilai ekstrim dan secara jelas berbeda dengan nilai-nilai lainnya pada distribusi.

6.5 Soal Latihan

1. Tentukan nilai skor-*z* yang berhubungan dengan setiap lokasi berikut ini dalam suatu distribusi
   a. Dibawah nilai rata-rata dengan deviasi standar 2
   b. Diatas nilai rata-rata dengan deviasi standar $\frac{1}{2}$
   c. Dibawah nilai rata-rata dengan deviasi standar 1.50
2. Jelaskan lokasi pada distribusi untuk masing-masing skor-*z* berikut. (Misal, z = +1.00 terletak di atas nilai rata-rata dengan deviasi standar 1)
   a) z = - 1.50 b) z = 0.25 c) z = -2.50 d) z = 0.50
3. Untuk suatu populasi dengan μ = 30 dan σ = 8, tentukan skor-z untuk masing-masing skor berikut:
   a) X = 32 b) X = 26 c) X = 42
4. Untuk populasi dengan μ = 50 dan σ = 12, tentukan nilai X yang berhubungan setiap skor-*z* berikut:
   a) z = -0.25 b) z = 2.00 c) z = -1.50
5. Pada suatu distribusi dengan μ = 50, suatu skor X = 42 yang berhubungan dengan z = -2.00. Berapakah deviasi standar untuk distribusi ini?
6. Pada suatu distribusi dengan σ = 12, suatu skor X = 56 yang berhubungan dengan z = -0.25. Berapakah nilai rata-rata distribusi?

7. Suatu distribusi normal dengan $\mu = 40$ dan $\sigma = 8$ diubah menjadi skor-z. Jelaskan bentuk, nilai rata-rata, dan deviasi standar distribusi skor-z tersebut?
8. Pada distribusi skor-z, apa keuntungan memiliki suatu nilai rata-rata $\mu = 0$?
9. Suatu distribusi nilai hasil ujian bahasa Inggris memiliki $\mu = 70$ dan $\sigma = 4$. Distribusi nilai hasil ujian sejarah memiliki $\mu = 60$ dan $\sigma = 20$. Pada hasil ujian bahasa Inggris ataukah sejarah, suatu nilai $X = 78$ memiliki peringkat lebih tinggi? Jelaskan jawaban anda.
10. Suatu distribusi nilai hasil ujian bahasa Inggris memiliki $\mu = 50$ dan $\sigma = 12$. Distribusi nilai hasil ujian sejarah memiliki $\mu = 58$ dan $\sigma = 4$. Pada hasil ujian bahasa Inggris ataukah sejarah, suatu nilai $X = 62$ menjadi nilai tertinggi? Jelaskan jawaban anda.

Jawaban

1. a) z = -2.00 b). z = +0.50 c) z = -1.50
2. a. Dibawah nilai rata-rata dengan deviasi standar $1\frac{1}{2}$

   b. Diatas nilai rata-rata dengan deviasi standar $\frac{1}{4}$

   c. Dibawah nilai rata-rata dengan deviasi standar $2\frac{1}{2}$
3. a) $z = \frac{K-\mu}{\sigma} = \frac{32-30}{8} = \frac{2}{8} = +0.25$ b) z = -0.50 c) z = +1.50
4. a) X = $\mu + z\sigma$ =50 + (-0.25) .12 = 50 -3 = 47

   b) X = 74 c) X = 56
5. $\sigma = \frac{K-\mu}{z} = \frac{42-50}{-2.00} = \frac{-8}{2} = 4$
6. $\mu = X - z\sigma = 56 - (-0.25 \cdot 12) = 56 - (-3) = 56 + 3 = 59$
7. Distribusi skor-z merupakan distribusi normal dengan nilai rata-rata 0 dan deviasi standar 1.

8 Distribusi skor-z yang memiliki nilai rata-rata 0 maka seluruh nilai positif berada di atas nilai rata-rata dan seluruh nilai negatif berada di bawah nilai rata-rata.

9. Skor-z untuk nilai ujian bahasa Inggris $z = \frac{K-\mu}{\sigma} = \frac{78-70}{4} = \frac{8}{4} = 2$

   Skor-z untuk nilai ujian sejarah = $z = \frac{K-\mu}{\sigma} = \frac{78-60}{20} = \frac{18}{20} = 0.90$

   Skor 78 untuk ujian bahasa Inggris memiliki peringkat lebih tinggi (skor-z = 2), skor 78 untuk nilai ujian sejarah merupakan nilai rata-rata (skor-z dibawah 1)

10. Nilai ujian $X = 62$ berhubungan dengan skor-z = + 1.00 untuk kedua distribusi. Nilai tersebut memiliki peringkat yang sama pada kedua distribusi.

REFERENSI

Bailey, K. D. (1978). Methods of social research. (3rd ed.). New York: The Free Press.

Grinnell (1988). Social work research and evaluation. Itasca, Illinois: F. E. Peacock Publishers.

Leedy, P. D. (1993). Practical research: planning and design. New Jersey: Prentice-Hall.

Kerlinger, F. (1973). The structure of scientific revolution. Chicago: University of Chicago Press.

Kumar, R. (2005). Research methodology. SAGE Publications.

Saunders, M., Lewis, P. & Thornhill, A. (2012) "Research Methods for Business Students" 6th edition, *Pearson Education Limited.*

Robson, C. (2011) *Real World Research: A Resource for Users of Social Research Methods in Applied Settings* (3rd edn). Chichester: John Wiley.

Dudovskiy, John (2018). The Ultimate Guide to Writing a Dissertation in Business Studies: A Step-by-Step Assistance. e-book. Https://research-methodology.net/about-us/ebook/

Thompson, Karl (January 13, 2016). Experiments in Sociology: An Introduction. Https://revisesociology.com/2016/01/13/experiments-in-sociology/

McCarney R, Warner J, Iliffe S, van Haselen R, Griffin M, Fisher P; Warner; Iliffe; Van Haselen; Griffin; Fisher (2007). *"The Hawthorne Effect: a randomised, controlled trial"*. BMC Med Res Methodol. **7**: 30. *doi:10.1186/1471-2288-7-30. PMC 1936999. PMID 17608932.*

Keizer, Kees (2010). The spreading of disorder. University of Groningen/UMCG research database. Https://www.rug.nl/research/portal/files/10436651/11complete.pdf

Khan, Omar (2019). Racism still festers in Britain's workplaces. It's time to get tough. Https://www.theguardian.com/commentisfree/2019/jan/22/racism-britain-workplaces-diversity-training-employers-sanctions.

Hutchinson, John (November 13, 2014). **Would you** intervene if you saw domestic abuse happening in public? Shocking secret experiment shows just how few of us would. Https://www.dailymail.co.uk/news/article-2832893/Would-intervene-saw-domestic-abuse-street-Shocking-social-experiment-shows-just-2-speak-up.html

Holsti, O. R. (1969). Content analysis for the social sciences and humanities, Reading, MA: Addison-Wesley.

Lombard, M., Snyder-Duch, J., & Bracken, C.C. (2004). Practical resources for assessing and reporting intercoder reliability in content analysis research projects. Retrieved on September 4, 2011, from http://astro.temple.edu/~lombard/reliability/.

Lombard, M., Snyder-Duch, J., & Bracken, C. C. (2004). A call for standardization in content analysis reliability. Human Communication Research, 30, 434-437.

Nueuendorf, K. A. (2002). The content analysis guidebook. Thousand Oaks, CA: Sage.

Wang, Weize (2011). A Content Analysis of Reliability in Advertising Content Analysis Studies. Electronic Theses and Dissertations. Paper 1375. http://dc.etsu.edu/etd/1375

Lani, James (2019). Quantitative Research Approach. Statistics Solution. Https://www.statisticssolutions.com/quantitative-research-approach/

Bhat, Adi (2019). Quantitative research: Definition, methods, types and examples. QuestionPro. Https://www.questionpro.com/blog/quantitative-research/

Socialcops (June 11, 2018). A complete guide to quantitative research methods. Https://blog.socialcops.com/academy/resources/quantitative-research-methods/

Jones & Bartlett (2019). *Quantitative reserach design: Experimental, quasi-experimental and descriptive*. Http://samples.jbpub.com/9781284101539/9781284101539_CH06_Drummond.pdf

Shuttleworth, Martin (2019). Quasi-Experimental Design. Https://explorable.com/quasi-experimental-design