

Statistik

Untuk Program Studi Pendidikan Geografi

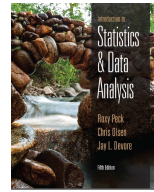
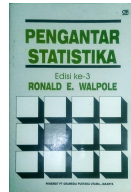
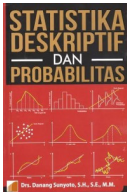
Nanda Arista Rizki, S.Si., M.Si.

Fakultas Keguruan dan Ilmu Pendidikan
Universitas Mulawarman
2020

Penilaian: tugas, kuis, UTS, UAS, afektif, dan wajib hadir (min. 80%).

Pengertian Statistik
Memahami Data
Ukuran Pemusatan Data (nilai sentral)
Ukuran Penyebaran Data (dispersi)
Populasi dan Sampel
Konsep Dasar Pengujian Hipotesis
Normalisasi Data

Buku referensi



Outline

- Pertemuan ke 1 : Pengantar statistik
- Pertemuan ke 2 : Memahami data
- Pertemuan ke 3 : Ukuran pemusatan
- Pertemuan ke 4 : Ukuran penyebaran
- Pertemuan ke 5 : **Kuis**, Materi (Populasi dan sampel)
- Pertemuan ke 6 : Pengujian hipotesis
- Pertemuan ke 7 : Ujian tengah semester (UTS)
- Pertemuan ke 8 : Normalisasi data

Pengertian statistik



Sumber: <https://werbas-ag.com>

Pengertian menurut KBBI

statistik /sta-tis-tik/ *n* **1** catatan angka-angka (bilangan); perangkaan; **2** data yang berupa angka yang dikumpulkan, ditabulasi, digolong-golongkan sehingga dapat memberi informasi yang berarti mengenai suatu masalah atau gejala

statistika /sta-tis-ti-ka/ *n* **1** ilmu tentang cara mengumpulkan, menabulasi, menggolong-golongkan, menganalisis, dan mencari kete-rangan yang berarti dari data yang berupa angka; **2** pengetahuan yang berhubungan dengan pengumpulan data, penyelidikan dan kesimpulannya berdasarkan bukti, berupa catatan bilangan (angka-angka)

Statistik

- merupakan kumpulan angka-angka (ringkasan) dari sampel berdasarkan fenomena kehidupan. (dapat juga berupa kategorik)
- digunakan untuk menduga parameter populasi.
- berperan dalam pembuatan kebijakan dan pengambilan keputusan.

Statistika

Statistika adalah **ilmu** yang mempelajari

- 1 cara mengumpulkan data
- 2 cara mengolah data
- 3 cara menyajikan data
- 4 cara menganalisis data
- 5 cara menginterpretasikan data

Fungsi statistik

Fungsi statistik ada dua, yaitu **fungsi deskriptif** dan **fungsi inferensial**.

Ciri khas statistik

Ada 3 ciri khas statistik:

- 1 bekerja dengan angka (bilangan). Data kualitatif sekalipun seperti kurang (1), cukup (2), dan lebih (3) dapat diubah menjadi bilangan.
- 2 objektif. Apa adanya, karena sebagai alat penilai kenyataan.
- 3 universal, dapat digunakan dalam hampir semua cabang kegiatan hidup manusia. Seperti biologi, kimia, fisika, geografi, sosial, dan lain-lain. Pada bidang kependudukan juga ada statistik kelahiran, statistik nikah, talak, cerai dan rujuk, statistik kematian, dan sebagainya.

Statistik deskriptif

Statistik deskriptif adalah metode statistik yang berkaitan dengan **pengumpulan dan penyajian data**.

Statistik deskriptif berfungsi untuk mendeskripsikan, menerangkan data dan peristiwa, yang dikumpulkan melalui proses penelitian dan penyelidikan (dimana belum sampai generalisasi atau mengambil kesimpulan tentang populasi yang diteliti).

Statistika deskriptif adalah informasi mengenai **gambaran umum** data dan sama sekali **tidak menarik kesimpulan** apapun.

Perhatikan ringkasan nilai geografi siswa berikut:

- $Mean = 68$
- $Median = 66$
- $Modus = 66$
- $Variansi = 61$
- $Deviasi\ standar = 8$
- $Nilai\ terendah = 60$
- $Nilai\ tertinggi = 92$

Tabel dan diagram batang

Nilai	Frekuensi
60	2
62	3
64	7
66	8
68	4
70	1
72	1
80	1
84	1
88	1
92	1

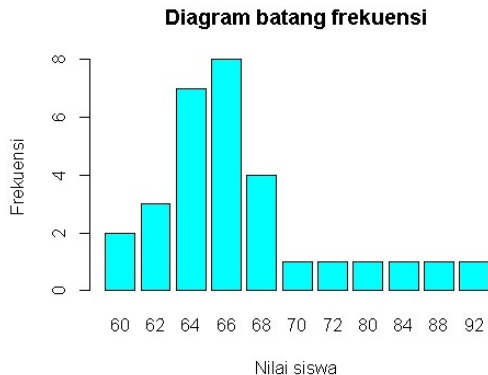
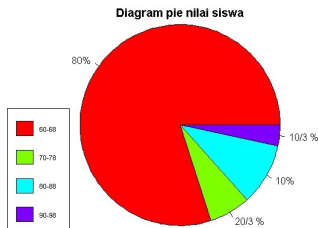


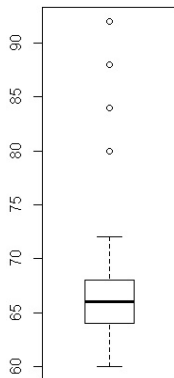
Diagram pie

Nilai	60	62	64	66	68	70	72	80	84	88	92
Frek.	2	3	7	8	4	1	1	1	1	1	1

Interval	60–68	70–78	80–88	90–98
Frek.	24	2	3	1



Box-plot

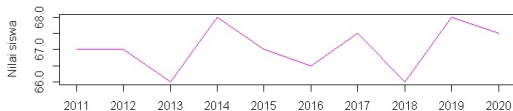


Perhatikan bahwa ada 4 titik di atas batas, artinya "nilai mereka ketinggian (*outlier*/pencilan)". Ada 4 nilai **ekstrim** yaitu 80, 84, 88, 92.

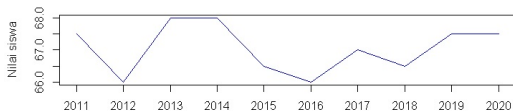
Grafik garis

	Tahun									
Kelas	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020
A	67	67	66	68	67	66.5	67.5	66	68	67.5
B	67.5	66	68	68	66.5	66	67	66.5	67.5	67.5

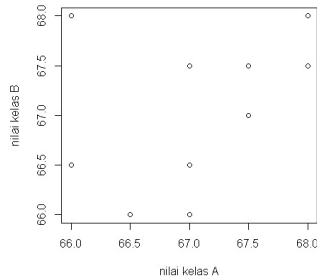
Nilai rata-rata ujian siswa kelas A per tahun



Nilai rata-rata ujian siswa kelas B per tahun



Scatterplot (diagram pencar)



Statistik inferensial (induktif)

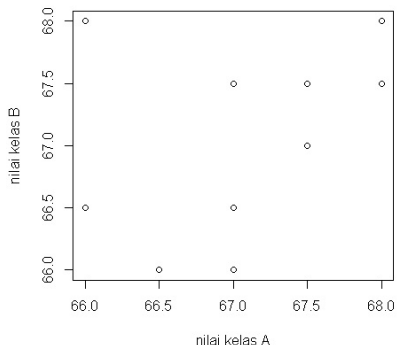
Statistik inferensial merupakan metode statistik yang berkaitan dengan analisis sampel **untuk menarik kesimpulan** tentang karakteristik populasi.

Statistik inferensial berfungsi **untuk memprediksi dan mengendalikan seluruh populasi** berdasarkan data, gejala, dan peristiwa yang ada pada proses penelitian. Fungsi ini dimulai dengan membuat suatu estimasi dan hipotesis.

Kesimpulan ilmiah adalah interpretasi terhadap sebuah populasi berdasarkan struktur teori.

Ujian	Nomor absen siswa									
	1	2	3	4	5	6	7	8	9	10
MTK	68	59	75	65	71	67	73	68	71	82
Geografi	71	66	76	70	68	77	86	69	82	81

- Apakah yakin ukuran pemusatan dan ukuran penyebaran antara hasil ujian Matematika dan hasil ujian Geografi adalah sama?
- Apakah yakin tidak ada perbedaan nilai antara hasil ujian Matematika dan hasil ujian Geografi?

Scatter plot nilai ujian siswa kelas A dan kelas B

Apakah yakin tidak ada hubungan antara nilai di kelas A dan nilai di kelas B?

Pengertian data

← → ↺ 🏠 <https://kbbi.web.id/data> 📄 ⋮ 📌 ☆ 📖 📄 📄 ☰

data /da-ta/ *n* 1 keterangan yang benar dan nyata: *pengumpulan -- untuk memperoleh keterangan tentang kehidupan petani*; 2 keterangan atau bahan nyata yang dapat dijadikan dasar kajian (analisis atau kesimpulan);
-- **digital** data yang berhubungan dengan angka untuk sistem perhitungan tertentu;
-- **kualitatif** data tidak berbentuk angka yang diperoleh dari rekaman, pengamatan, wawancara, atau bahan tertulis;
-- **kuantitatif** data berbentuk angka yang diperoleh dari perhitungan data kualitatif;
-- **lapangan** 1 kumpulan data yang diperoleh dengan cara melakukan pengukuran langsung (tidak menggunakan satelit); 2 kumpulan data yang diperoleh dengan cara melakukan pengukuran langsung di lapangan;
-- **lisan** data yang diperoleh dari informan;
-- **pribadi** data yang berkenaan dengan ciri seseorang, misalnya nama, umur, jenis kelamin, pendidikan, pekerjaan, alamat, dan kedudukan dalam keluarga;
-- **primer** data yang diperoleh seorang peneliti langsung dari objeknya;
-- **relevan** data yang ada hubungan langsung dengan persoalan yang sedang diteliti;
-- **sekunder** data yang diperoleh seorang peneliti secara tidak langsung dari objeknya, tetapi melalui sumber lain, baik lisan maupun tulis;
-- **tertulis** data yang diperoleh dari sumber tertulis;
mendata /men-da-ta/ *v* melakukan pendataan;
mendatakan /men-da-ta-kan/ *v* 1 mengambil sebagai data; 2 memasukkan data;
pendataan /pen-da-ta-an/ *n* 1 proses, cara, perbuatan mendata; 2 pengumpulan data; pencarian data

variabel

karakteristik yang memiliki nilai yang dapat berubah dari satu pengamatan ke pengamatan lainnya.

data

kumpulan dari pengamatan pada satu atau lebih variabel.

Suatu kumpulan (atau himpunan) data yang terdiri dari berbagai pengamatan pada 1 karakteristik, disebut kumpulan data univariat (*univariate dataset*).

Jika dataset univariat tersebut berisi pengamatan-pengamatan kategorik, maka disebut *dataset* kategorik.

Namun jika dataset univariat tersebut berisi pengamatan-pengamatan numerik (kuantitatif), maka disebut *dataset* numerik.

Variabel numerik diskrit

Variabel numerik menghasilkan data diskrit jika nilai variabel yang mungkin diperoleh dengan cara **menghitung**.

Variabel numerik kontinu

Variabel numerik menghasilkan data kontinu jika nilai variabel yang mungkin diperoleh dengan cara **mengukur**.

Kualitas pengukuran

Ada dua kualitas yang dikehendaki dari suatu alat atau proses pengukuran, yaitu:

- reliabilitas
- validitas (keabsahan)



Reliable not Valid
Precise not Accurate



Not Reliable but Valid
Not Precise but Accurate



Not Reliable and not Valid
Not Precise and not Accurate



Reliable and Valid
Precise and Accurate

Reliabilitas

Reliabilitas adalah suatu ukuran konsistensi dari serangkaian pengukuran atau serangkaian alat ukur. Semakin konsisten hasil yang diperoleh dalam pengukuran yang berulang maka semakin tinggi reliabilitas dari alat, prosedur, atau proses pengukuran tersebut.

Uji reliabilitas adalah uji yang digunakan untuk menunjukkan sejauh mana pengukuran dari suatu tes tetap konsisten setelah dilakukan berulang-ulang terhadap subjek dan dalam kondisi yang sama.

Validitas

Validitas adalah suatu ukuran yang menunjukkan bahwa variabel yang diukur memang benar-benar variabel yang hendak diteliti (oleh Pene-liti). *"It measures what it purports to measure"*.

Uji validitas adalah uji yang digunakan untuk menunjukkan sejauh ma-na alat ukur yang digunakan dalam suatu mengukur apa yang diukur.

Syarat data yang baik

Data merupakan bahan yang sangat-sangat penting dalam pembuatan kebijakan dan pengambilan keputusan. Syaratnya,

- 1 Objektif
- 2 Representatif
- 3 Memiliki *sampling error* yang kecil
- 4 *Up-to-date*
- 5 Relevan

Pengumpulan data

Ada dua pendekatan dalam mengumpulkan data:

- Pengamatan lengkap, atau sensus. Pengumpulan data terhadap seluruh elemen/unsur/unit dalam populasi.
- Pengamatan sebagian, atau survei sampel. Pengumpulan data terhadap sebuah sampel yang merupakan bagian dari populasi.

Alasan menggunakan pengamatan sebagian (sampling):

- biaya (*cost*)
- tenaga (*human resources*)
- waktu (*time constraint*)
- bersifat merusak (*destructive case*), seperti percobaan daya tahan lampu
- ketidakmungkinan melakukan pengamatan keseluruhan (*impossibility*), seperti pengamatan batu di Kalimantan.

Ada dua jenis kesalahan dalam pengamatan sebagian:

- Kesalahan sampling (*sampling error*), yaitu terjadi perbedaan dari suatu nilai duga yang diperoleh berdasarkan sampel (**statistik/penduga**) dengan nilai yang berdasarkan populasi (**parameter**) akibat dilakukan pengamatan sebagian.
- Kesalahan bukan sampling (*non-sampling error*), yaitu terjadi perbedaan dari suatu nilai duga yang diperoleh berdasarkan sampel (**statistik/penduga**) dengan nilai yang berdasarkan populasi (**parameter**) **bukan dari** akibat dilakukan pengamatan sebagian.

Statistik vs parameter

- Populasi = tinggi badan mahasiswa Geografi angkatan 2019
 - Parameter = rata-rata seluruh mahasiswa
 - Sampel = tinggi badan 10 mahasiswa Geografi angkatan 2019
 - Statistik = rata-rata tinggi badan dari 10 mahasiswa
-
- Populasi = nilai ujian statistik untuk mahasiswa Geografi angkatan 2019
 - Parameter = variansi nilai ujian dari seluruh mahasiswa
 - Sampel = nilai ujian statistik untuk 10 mahasiswa Geografi angkatan 2019
 - Statistik = variansi nilai ujian dari 10 mahasiswa

***Sampling error* dapat dihindari dan diminimalisasi**, dengan cara menambah jumlah sampel atau mengubah menjadi sampel yang lebih representatif.

***Non-sampling error* sulit dihindari** dan dalam menghadapinya harus subyektif. Kesalahan sampling ini sebagian besar disebabkan oleh kesalahan manusia (*human error*), yang dapat terjadi pada setiap bagian penelitian (mulai dari penentuan masalah, desain penelitian, hingga penarikan kesimpulan).

Contoh *non-sampling error*:

- 1 **Ada unit pengamatan yang seharusnya dicacah**, namun karena batas daerah penelitian yang kurang jelas maka unit tersebut tidak teramati.
- 2 **Ada unit pengamatan yang seharusnya tidak masuk** dalam ruang lingkup penelitian, namun masuk dalam proses pengumpulan data.
- 3 **Ada unit pengamatan yang tercacah lebih dari satu kali**, yang diakibatkan oleh batas pembagian kerja yang tidak jelas antar pencacah.

Akurasi

berkaitan dengan sejauh mana sebuah penduga/estimator (berdasarkan sampel) berbeda dengan nilai parameter yang akan diduga (ringkasan populasi).

Presisi

berkaitan dengan seberapa dekatnya beberapa nilai duga (berdasarkan sampel) terhadap satu dengan yang lain. Presisi diukur dengan variansi dan kesalahan baku dari penduga.



Reliable not Valid
Precise not Accurate



Not Reliable but Valid
Not Precise but Accurate



Not Reliable and not Valid
Not Precise and not Accurate



Reliable and Valid
Precise and Accurate

$$Akurasi = Presisi + Bias,$$

dengan *Bias* adalah selisih rata-rata dari nilai duga dengan parameter yang akan diduga.

Pengertian nilai sentral

Nilai semua data biasanya tidak jauh dengan ukuran pemusatannya (nilai sentralnya). Ukuran ini dapat disebut sebagai ukuran lokasi, yaitu suatu nilai yang menggambarkan ciri atau nilai umum dari suatu kelompok pengamatan. Nilai ini mewakili kelompok pengamatannya. Mean (rata-rata), median, dan modus merupakan ukuran pemusatan data.

Tabel: Nilai Ulangan Harian Siswa

Kelas A			Kelas B			Kelas C		
94	94	90	16	94	22	92	30	84
86	84	36	42	26	82	74	78	14
64	56	54	50	16	32	44	12	40
76	90	94	80	60	84	10	46	24
42	28		62	78	20	48	24	12
72	46		12	46	52	94	88	76
84	20		56	44		90	60	28
64	26		18	64		80	18	58

Tabel: Nilai Ulangan Harian Siswa

Kelas A			Kelas B			Kelas C		
94	94	90	16	94	22	92	30	84
86	84	36	42	26	82	74	78	14
64	56	54	50	16	32	44	12	40
76	90	94	80	60	84	10	46	24
42	28		62	78	20	48	24	12
72	46		12	46	52	94	88	76
84	20		56	44		90	60	28
64	26		18	64		80	18	58

Apa yang terjadi jika nilai KKM=70? Jika diajar oleh guru yang sama (kenapa?), kelas mana yang paling butuh perhatian?

	Kelas A	Kelas B	Kelas C
Banyak siswa	20	22	24
Banyak siswa lulus	10	5	9
Median	68	48	47
Modus	94	16	12 dan 24
Mean	65	48	51

Fungsi nilai sentral

Fungsi nilai sentral adalah:

- Untuk mengetahui keadaan suatu kelompok atau kelas pengamatan, sehingga diperoleh gambaran ringkas mengenai kelompok tersebut.
- Untuk membandingkan nilai suatu variabel pada beberapa (dua atau lebih) kelompok.

Rata-rata (*Mean*)

Rata-rata hitung (*arithmetic mean*), atau sering disebut rata-rata hitung atau rata-rata saja, diperoleh dengan menjumlahkan nilai seluruh pengamatan dan membagi jumlah tersebut dengan banyaknya pengamatan.

Rata-rata (*Mean*)

Rata-rata hitung (*arithmetic mean*), atau sering disebut rata-rata hitung atau rata-rata saja, diperoleh dengan menjumlahkan nilai seluruh pengamatan dan membagi jumlah tersebut dengan banyaknya pengamatan.

Misal diberikan hasil belajar dari 10 siswa berikut

No. Urut	1	2	3	4	5	6	7	8	9	10
Hasil belajar siswa	84	79	3	65	65	13	48	69	86	68

Bagaimana kondisi sampel tersebut jika nilai KKM=70?

Diberikan data sampel hasil ujian Matematika dan Geografi dari 10 siswa kelas X sebagai berikut

Ujian	Nomor absen siswa									
	1	2	3	4	5	6	7	8	9	10
MTK	68	59	75	65	71	67	73	68	71	82
Geografi	71	66	76	70	68	77	86	69	82	81

Nilai rata-rata ujian Matematika adalah sebagai berikut

$$\begin{aligned}\bar{x}_{mtk} &= \frac{1}{10}(68 + 59 + 75 + 65 + 71 + 67 + 73 + 68 + 71 + 82) \\ &= \frac{699}{10} = 69.9\end{aligned}$$

Sedangkan nilai rata-rata ujian Geografi adalah sebagai berikut

$$\begin{aligned}\bar{x}_{geo} &= \frac{1}{10}(71 + 66 + 76 + 70 + 68 + 77 + 86 + 69 + 71 + 81) \\ &= \frac{735}{10} = 73.5\end{aligned}$$

Rata-rata hitung tertimbang

Rata-rata hitung tertimbang adalah rata-rata hitung dari sekumpulan nilai data namun diperoleh dengan memberikan timbangan (bobot) terhadap setiap nilai tersebut.

Sebagai contoh, Annisa memperoleh nilai untuk mata kuliah statistik sebagai berikut

Afektif	UTS	UAS
75	60	80

Berapakah nilai akhir Annisa jika dihitung berdasarkan acuan pemberian nilai mutu berikut?

Skema	Kuis/UTS	UAS	Afektif
III	45%	45%	10%
IV	40%	50%	10%

Misal berikut rincian mata kuliah dengan SKS dan angka mutu Annisa *)

Smt	Mata Kuliah	SKS	Angka Mutu	Nilai Bobot	Nilai Huruf
I	A	3	85	4,00	A
I	B	2	77	3,50	B
II	C	3	68	2,50	C
II	D	4	72	3,00	B
II	E	3	50	1,50	D

(*) fiktif belaka

Hitunglah Indeks Prestasi (IP) dan Indeks Prestasi Kumulatif (IPK) Annisa!

Jawab:

- IP semester I Annisa adalah

$$\frac{3 \cdot 4,00 + 2 \cdot 3,50}{3 + 2} = \frac{19}{5} = 3,80$$

- IP semester II Annisa adalah

$$\frac{3 \cdot 2,50 + 4 \cdot 3,00 + 3 \cdot 1,50}{3 + 4 + 3} = \frac{24}{10} = 2,40$$

- IPK semester II Annisa adalah

$$\frac{3 \cdot 4,00 + 2 \cdot 3,50 + 3 \cdot 2,50 + 4 \cdot 3,00 + 3 \cdot 1,50}{3 + 2 + 3 + 4 + 3} = \frac{43}{15} = 2,87$$

Rata-rata hitung tertimbang digunakan ketika perlakuan untuk nilai-nilai pengamatan adalah tidak sama. Dalam hal ini, diberi suatu bobot untuk setiap nilai dalam suatu pengamatan sehingga mempengaruhi nilai rata-ratanya.

Rata-rata harmoni

Rata-rata harmoni (*harmonic mean*) adalah teknik perhitungan rata-rata yang baik digunakan jika ada beberapa data pencilan. Tidak seperti rata-rata hitung, rata-rata harmoni sangat tidak berpengaruh oleh pencilan yang bernilai tinggi, sehingga memberikan gambaran umum yang lebih baik.

Rumusnya,

$$RH(\text{Rata-rata Harmoni}) = \frac{n}{\sum_{i=1}^n \frac{1}{x_i}}$$

Diberikan data sebagai berikut

1, 2, 3, 4, 5

Hitunglah rata-rata harmoninya!

Jawab:

$$RH = \frac{5}{\frac{1}{1} + \frac{1}{2} + \frac{1}{3} + \frac{1}{4} + \frac{1}{5}} = \frac{5}{137/60} \approx 2.19$$

Median

Misal diberikan sampel nilai ujian suatu mata pelajaran berikut.

49 99 45 47
44 43 100

Jika nilai standar kelulusannya adalah 60, apakah kelas ini dapat dikatakan lulus?

Median

Misal diberikan sampel nilai ujian suatu mata pelajaran berikut.

49 99 45 47
44 43 100

Jika nilai standar kelulusannya adalah 60, apakah kelas ini dapat dikatakan lulus? Berapa nilai tepat untuk mewakili kelas ini?

Median

Misal diberikan sampel nilai ujian suatu mata pelajaran berikut.

49 99 45 47
 44 43 100

Jika nilai standar kelulusannya adalah 60, apakah kelas ini dapat dikatakan lulus? Berapa nilai tepat untuk mewakili kelas ini?

mean	RH	median
61	54	47

Ujian	Nomor absen siswa									
	1	2	3	4	5	6	7	8	9	10
MTK	68	59	75	65	71	67	73	68	71	82
Geografi	71	66	76	70	68	77	86	69	82	81

Berikut nilai ujian Matematika yang telah diurutkan

Nilai	59	65	67	68	71	73	75	82
Frek.	1	1	1	2	2	1	1	1

Karena banyaknya siswa adalah 10, maka nilai median untuk ujian Matematika adalah rata-rata dari siswa ke 5 dan siswa ke 6 (yang telah diurutkan). Dengan demikian, diperoleh

$$Median_{mtk} = \frac{1}{2}(68 + 71) = 69.5$$

Berikut nilai ujian Geografi yang telah diurutkan

Nilai	66	68	69	70	71	76	77	81	86
Frek.	1	1	1	1	2	1	1	1	1

Karena banyaknya siswa adalah 10, maka nilai median untuk ujian Geografi juga adalah rata-rata dari siswa ke 5 dan siswa ke 6 (yang telah diurutkan). Dengan demikian, diperoleh

$$Median_{geo} = \frac{1}{2}(71 + 71) = 71$$

Rumus dalam penentuan median adalah sebagai berikut

$$\frac{k - 1}{n - 1} = 0.5$$
$$k = 1 + 0.5(n - 1).$$

Untuk $n = 10$, maka $k = 1 + 0.5(9) = 5.5$ yang berarti data ke 5.5. Dengan menggunakan interpolasi, maka nilai median adalah rata-rata dari data ke 5 dan data ke 6.

Perhatikan 2 kelas berikut

No. Urut Siswa	Kelas A	Kelas B
1	45	51
2	66	71
3	54	40
4	54	73
5	100	74
6	90	71
7	69	72
8	66	73
9	48	71
10	98	34

Jika KKM=70, kelas manakah yang lebih baik?

Kelas	mean	median
A	69	66
B	63	71

Hitunglah banyaknya siswa yang lulus (dengan nilai KKM=70)!

Modus (*Mode*)

Perhatikan kembali hasil ujian 10 siswa kelas X berikut

Ujian	Nilai									
MTK	68	59	75	65	71	67	73	68	71	82
Geografi	71	66	76	70	68	77	86	69	82	81

Modus (*Mode*)

Perhatikan kembali hasil ujian 10 siswa kelas X berikut

Ujian	Nilai									
MTK	68	59	75	65	71	67	73	68	71	82
Geografi	71	66	76	70	68	77	86	69	82	81

Pertama urutkan data (juga untuk memudahkan dalam penentuan median), lalu hitung frekuensinya.

Nilai ujian Matematika	59	65	67	68	71	73	75	82
Frek.	1	1	1	2	2	1	1	1

Nilai ujian Geografi	66	68	69	70	71	76	77	81	86
Frek.	1	1	1	1	2	1	1	1	1

Karena nilai 68 dan 71 pada ujian Matematika masing-masing memiliki frekuensi tertinggi yang sama banyaknya, yaitu 2, maka nilai modus untuk nilai ujian Matematika adalah 68 dan 71. Hal ini merupakan bimodus atau *bimodal*, karena memiliki dua nilai modus.

Sementara pada nilai ujian Geografi, frekuensi terbanyak berasal dari nilai 71. Oleh karena itu, nilai modus untuk ujian Geografi adalah 71. Dalam hal ini, disebut *unimodal*.

Modus memberi gambaran mengenai konsisten nilai dari data yang ada. Semakin besar frekuensi modus maka semakin konsisten data yang ada.

Perhatikan 2 kelas berikut

No. Urut Siswa	Kelas A	Kelas B
1	70	72
2	70	72
3	70	72
4	69	72
5	68	72
6	68	69
7	65	66
8	76	72
9	69	59
10	75	54

Jika nilai KKM=70, kelas manakah yang lebih baik?

Kelas	mean	modus
A	70	70
B	68	72

Hitunglah banyaknya siswa yang lulus (dengan nilai KKM=70)!

Selain untuk data kuantitatif, modus sering digunakan untuk data kualitatif (nominal atau ordinal).

Berikut warna favorit siswa di suatu sekolah

Hitam, Biru, Ungu, Biru, Merah, Biru,
Merah, Biru, Hijau, Putih, Kuning, Merah,
Biru, Biru, Biru, Hijau, Kuning,

Tentukan modusnya!

Contoh Soal

1. Berikut adalah nilai ulangan harian dari 5 kelas.

No. Urut	A	B	C	D	E
1	98	59	6	84	1
2	45	27	74	84	56
3	98	94	63	79	3
4	74	93	32	39	81
5	59	76	67	63	98
6	84	65	7	68	72
7	8	93	85	69	92
8	98	93	60	84	80
9	98	23	32	83	92
10	22	65	95	3	33

Contoh Soal

No. Urut	A	B	C	D	E
11	93	49	58	47	30
12	78	2	24	28	50
13	10	64	25	20	1
14	24	75	23	84	45
15	98	11	73	84	31
16	98	7	21	21	69
17	39	23	58	76	33
18	52	25	67	84	76
19	98	55	47	84	12
20	98	54	52	84	24

Contoh Soal

2. Berikut adalah nilai hasil ujian remedial dari suatu kelas.

52	74	4	28	48	56	16
74	52	28	16	74	74	

Contoh Soal

3. Berikut adalah nilai ujian akhir dari suatu kelas.

90	90	90	90	90	56	90
90	84	60	12	90	36	80
64	90	90	90			

Contoh Soal

4. Berikut adalah nilai ujian nasional dari suatu sampel.

60	98	60	98	60	60	60
60	62	60	68	60	69	60
89	60	98	97	60	60	83
60	98	74				

Contoh Soal

5. Bukalah akun Anda di <https://sia.unmul.ac.id>, lalu hitunglah nilai IPK secara manual!

Pengertian dispersi

Ukuran penyebaran data disebut juga ukuran dispersi, yang diperlukan untuk mengukur seberapa jauh data dari ukuran pemusatannya. Oleh karena itu, kisaran data 'bisa ditebak' dari ukuran pemusatan dan ukuran penyebarannya.

Pengertian dispersi

Ukuran penyebaran data disebut juga ukuran dispersi, yang diperlukan untuk mengukur seberapa jauh data dari ukuran pemusatannya. Oleh karena itu, kisaran data 'bisa ditebak' dari ukuran pemusatan dan ukuran penyebarannya.

Ukuran ini berperan dalam menjelaskan homogenitas suatu sampel. Semakin kecil ukuran penyebarannya, maka sampel tersebut homogen. Sebaliknya, semakin besar ukuran penyebarannya, maka sampel tersebut dikatakan heterogen.

Perhatikan data hasil pengukuran tinggi badan sejumlah mahasiswa/i di suatu kelas berikut

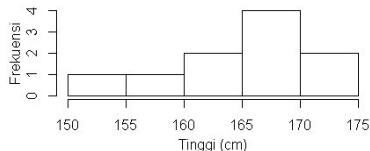
Mahasiswa		Mahasiswi		
169	175	148	148	159
159	173	138	154	159
170	168	142	151	153
165	153	148	159	147
161	170	147	148	152

Statistik	Mahasiswa	Mahasiswi
min	153	138
median	168.5	148
maks	175	159
<i>range</i>	22	21
<i>n</i>	10	15
$k = 1 + 3.3(\log n)$	4	5
<i>length</i>	6	4

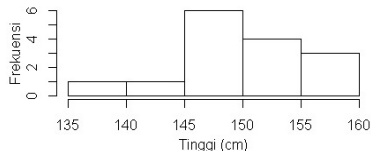
Mahasiswa		Mahasiswi	
Interval	Frekuensi	Interval	Frekuensi
153-158	1	138-141	1
159-164	2	142-145	1
165-170	5	146-149	6
171-176	2	150-153	3
		154-157	1
		158-161	3

Mahasiswa		Mahasiswi	
Interval	Frekuensi	Interval	Frekuensi
153-158	1	138-141	1
159-164	2	142-145	1
165-170	5	146-149	6
171-176	2	150-153	3
		154-157	1
		158-161	3

Histogram Tinggi Badan Sejumlah Mahasiswa



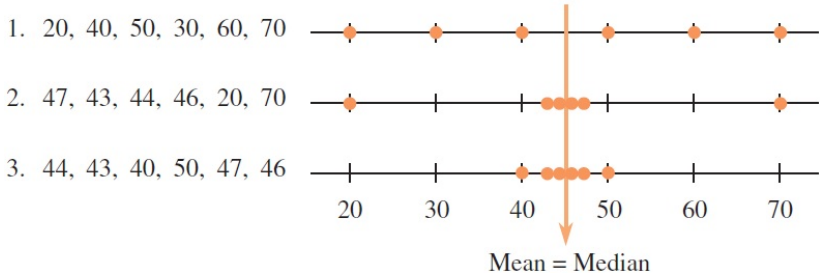
Histogram Tinggi Badan Sejumlah Mahasiswi



Rentang (*Range*)

Rentang data diperoleh dari selisih dari nilai maksimum dan nilai minimum sampel.

Sample



Perhatikan kembali data sampel hasil ujian Matematika dan Geografi dari 10 siswa kelas X berikut

Ujian	Nomor absen siswa									
	1	2	3	4	5	6	7	8	9	10
MTK	68	59	75	65	71	67	73	68	71	82
Geografi	71	66	76	70	68	77	86	69	82	81

Perhatikan kembali data sampel hasil ujian Matematika dan Geografi dari 10 siswa kelas X berikut

Ujian	Nomor absen siswa									
	1	2	3	4	5	6	7	8	9	10
MTK	68	59	75	65	71	67	73	68	71	82
Geografi	71	66	76	70	68	77	86	69	82	81

Rentang untuk nilai ujian Matematika adalah $82 - 59 = 23$, sedangkan untuk nilai ujian Geografi adalah $82 - 66 = 16$.

Artinya, nilai ujian Matematika **lebih menyebar** daripada nilai ujian Geografi berdasarkan rentang datanya.

Kelemahan dari jangkauan (*range*) adalah hanya bergantung pada dua nilai saja (nilai minimum dan maksimum), sehingga tidak cocok ketika ada nilai ekstrim (pencilan).

Variansi dan deviasi standar

Rumus untuk menghitung variansi (*variance*) dari suatu sampel:

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

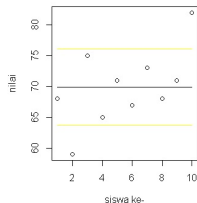
Rumus untuk menghitung deviasi standar (*standard deviation*):

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

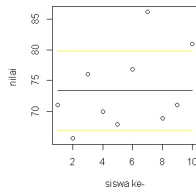
Perhatikan kembali data sampel hasil ujian Matematika dan Geografi dari 10 siswa kelas X berikut

Ujian	Nomor absen siswa									
	1	2	3	4	5	6	7	8	9	10
MTK	68	59	75	65	71	67	73	68	71	82
Geografi	71	66	76	70	68	77	86	69	82	81

Nilai ujian Matematika kelas X



Nilai ujian Geografi kelas X



Perhatikan bahwa garis hitam tengah merupakan *mean* (rata-rata). Garis kuning menandakan penyebaran data (deviasi standar) terhadap rata-ratanya.

Apakah penyebaran data nilai ujian Matematika dan Geografi adalah sama?

Ujian Matematika

Diketahui $\bar{x}_{mtk} = 69.9$, maka

$$s_{mtk}^2 = \frac{(68 - 69.9)^2 + (59 - 69.9)^2 + \dots + (82 - 69.9)^2}{10 - 1} = 38.1$$

Sehingga $s_{mtk} = \sqrt{s_{mtk}^2} = \sqrt{38.1} = 6.17$

Ujian Geografi

Diketahui $\bar{x}_{geo} = 74.6$, maka

$$s_{geo}^2 = \frac{(71 - 74.6)^2 + (66 - 74.6)^2 + \dots + (81 - 74.6)^2}{10 - 1} = 46.27$$

Sehingga $s_{geo} = \sqrt{s_{geo}^2} = \sqrt{40.28} = 6.80$

Koefisien variasi

Koefisien variasi (*Coefficient of variation*) adalah salah satu ukuran variasi yang dianjurkan untuk membandingkan tingkat variasi dari dua kelompok data atau lebih. Rumusnya,

$$KV = \frac{s}{\bar{x}} \times 100\%.$$

Sehingga nilai KV untuk nilai ujian Matematika adalah $\frac{6.17}{69.9} \approx 8.83\%$, sedangkan nilai KV untuk nilai ujian Geografi adalah $\frac{6.80}{74.6} \approx 9.12\%$. Artinya nilai ujian Geografi lebih heterogen dibanding dengan nilai ujian Matematika.

Nilai deviasi standar sangat tergantung pada nilai rata-ratanya, namun nilai KV (yang telah dibagi dengan mean) dianggap telah 'menghilangkan' efek rata-rata tersebut. Sehingga nilai KV bisa lebih baik untuk digunakan dalam membandingkan penyebaran dari beberapa kelompok data.

Sebagai contoh, ada sekelompok mahasiswa yang mengambil 2 ujian yang berbeda, yaitu ujian skripsi dan *Test of English as a Foreign* (TOEFL). Nilai rata-rata ujian skripsinya adalah 80 dengan deviasi standar 6, sedangkan nilai rata-rata TOEFL yang diperoleh 550 dengan deviasi standar 7. Hal ini menyebabkan nilai KV untuk ujian skripsi dan TOEFL berturut-turut adalah 7.5% dan 1.27%.

Sebagai contoh, ada sekelompok mahasiswa yang mengambil 2 ujian yang berbeda, yaitu ujian skripsi dan *Test of English as a Foreign* (TOEFL). Nilai rata-rata ujian skripsinya adalah 80 dengan deviasi standar 6, sedangkan nilai rata-rata TOEFL yang diperoleh 550 dengan deviasi standar 7. Hal ini menyebabkan nilai KV untuk ujian skripsi dan TOEFL berturut-turut adalah 7.5% dan 1.27%.

Jika hanya dilihat dari deviasinya saja, maka nilai TOEFL lebih heterogen. Namun jika diperhatikan secara relatif terhadap nilai rata-ratanya, maka nilai TOEFL lebih homogen.

Aturan Chebyshev

Chebyshev's Rule

Consider any number k , where $k \geq 1$. Then the percentage of observations that are within k standard deviations of the mean is at least $100\left(1 - \frac{1}{k^2}\right)\%$.

Substituting selected values of k gives the following results.

Number of Standard Deviations, k	$1 - \frac{1}{k^2}$	Percentage Within k Standard Deviations of the Mean
2	$1 - \frac{1}{4} = .75$	at least 75%
3	$1 - \frac{1}{9} = .89$	at least 89%
4	$1 - \frac{1}{16} = .94$	at least 94%
4.472	$1 - \frac{1}{20} = .95$	at least 95%
5	$1 - \frac{1}{25} = .96$	at least 96%

Aturan empiris

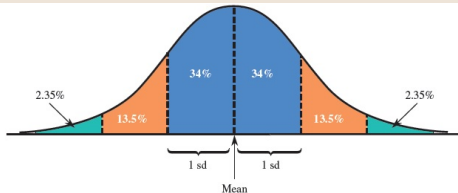
The Empirical Rule

If the histogram of values in a data set can be reasonably well approximated by a normal curve, then

Approximately 68% of the observations are within 1 standard deviation of the mean.

Approximately 95% of the observations are within 2 standard deviations of the mean.

Approximately 99.7% of the observations are within 3 standard deviations of the mean.



	Mahasiswa	Mahasiswa
\bar{x}	166.3	150.2
s	6.82	6.06
$\bar{x} - 4s$	139.04	125.95
$\bar{x} - 3s$	145.85	132.02
$\bar{x} - 2s$	152.67	138.08
$\bar{x} - s$	159.48	144.14
\bar{x}	166.30	150.20
$\bar{x} + s$	173.12	156.26
$\bar{x} + 2s$	179.93	162.32
$\bar{x} + 3s$	186.75	168.38
$\bar{x} + 4s$	193.56	174.45

Aturan Chebyshev

- Untuk data tinggi badan mahasiswa, sedikitnya 75% data berada dalam rentang 152.67 sampai 179.93, sedikitnya 89% data berada dalam rentang 145.85 sampai 186.75, dan sedikitnya 94% data berada dalam rentang 139.04 sampai 193.56
- untuk data tinggi badan mahasiswi, sedikitnya 75% data berada dalam rentang 138.08 sampai 162.32, sedikitnya 89% data berada dalam rentang 132.02 sampai 168.38, dan sedikitnya 94% data berada dalam rentang 139.04 sampai 174.45

Aturan empiris

- Untuk data tinggi badan mahasiswa, diestimasi sekitar 68% data berada dalam rentang 159.48 sampai 173.12, sekitar 95% data berada dalam rentang 152.67 sampai 179.93, dan sekitar 99.7% data berada dalam rentang 145.85 sampai 186.75
- untuk data tinggi badan mahasiswi, diestimasi sekitar 68% data berada dalam rentang 144.14 sampai 156.26, sekitar 95% data berada dalam rentang 138.08 sampai 162.32, dan sekitar 99.7% data berada dalam rentang 132.02 sampai 168.38

Contoh Soal

1. Berikut adalah nilai ulangan harian dari 5 kelas.

No. Urut	A	B	C	D	E
1	98	59	6	84	1
2	45	27	74	84	56
3	98	94	63	79	3
4	74	93	32	39	81
5	59	76	67	63	98
6	84	65	7	68	72
7	8	93	85	69	92
8	98	93	60	84	80
9	98	23	32	83	92
10	22	65	95	3	33

Contoh Soal

No. Urut	A	B	C	D	E
11	93	49	58	47	30
12	78	2	24	28	50
13	10	64	25	20	1
14	24	75	23	84	45
15	98	11	73	84	31
16	98	7	21	21	69
17	39	23	58	76	33
18	52	25	67	84	76
19	98	55	47	84	12
20	98	54	52	84	24

Contoh Soal

2. Berikut adalah nilai hasil ujian remedial dari suatu kelas.

52	74	4	28	48	56	16
74	52	28	16	74	74	

Contoh Soal

3. Berikut adalah nilai ujian akhir dari suatu kelas.

90	90	90	90	90	56	90
90	84	60	12	90	36	80
64	90	90	90			

Pengertian populasi dan sampel

Populasi

Populasi adalah **seluruh** elemen atau unit pengamatan yang akan diteliti dan dipelajari. Misalnya, semua nilai Geografi siswa kelas X.

Sampel

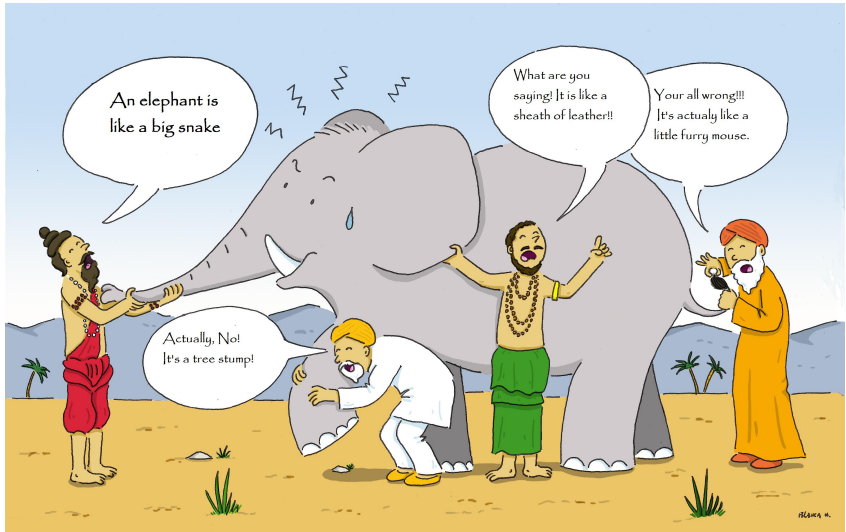
Sampel adalah **sebagian** elemen unit pengamatan dari populasi. Misalnya, nilai Geografi dari 10 siswa kelas X pada tahun ajaran 2019/2020 (yang dipilih secara acak).

Sampel diambil untuk dipelajari informasinya, untuk:

- memperkirakan karakteristik dari populasi
- menguji hipotesis tentang karakteristik populasi tersebut.

Syarat-syarat data sampel yang baik bisa dilihat kembali pada materi

syarat data yang baik



Misal ingin melihat perubahan iklim di Kalimantan Timur. Data sampel apa yang diperlukan?

- Perubahan nilai curah hujan musiman
- Perubahan nilai frekuensi hujan lebat musiman
- Perubahan jumlah hari tanpa hujan musiman
- Perubahan jumlah indeks *consecutive dry days* musiman
- Perubahan jumlah indeks *consecutive wet days* musiman
- Perubahan nilai fraksi hujan lebat musiman
- Perubahan nilai rata-rata suhu

Statistik vs parameter

- Populasi = tinggi badan mahasiswa Geografi angkatan 2019
 - Parameter = rata-rata seluruh mahasiswa
 - Sampel = tinggi badan 10 mahasiswa Geografi angkatan 2019
 - Statistik = rata-rata tinggi badan dari 10 mahasiswa
-
- Populasi = nilai ujian statistik untuk mahasiswa Geografi angkatan 2019
 - Parameter = variansi nilai ujian dari seluruh mahasiswa
 - Sampel = nilai ujian statistik untuk 10 mahasiswa Geografi angkatan 2019
 - Statistik = variansi nilai ujian dari 10 mahasiswa

Teknik sampling

Teknik sampling adalah cara menentukan banyaknya sampel dan pemilihan calon anggota sampel. Sampel yang diharapkan adalah representatif. Teknik pemilihan sampel ada dua, yaitu *probability sampling* dan *non probability sampling*.

Probability sampling

- setiap elemen populasi memiliki kesempatan yang sama untuk menjadi sampel
- hasilnya lebih objektif
- contohnya: *simple random sampling, systematic random sampling, stratified random sampling, cluster random sampling, multi stage sampling*

Non probability sampling

- setiap elemen populasi tidak memiliki kesempatan yang sama untuk menjadi sampel
- hasilnya lebih subjektif
- dilakukan karena kemampuan dan batasan peneliti dalam menarik sampel
- contohnya: *purposive sampling, accidental sampling, quota sampling, snowball sampling*

Menentukan ukuran sampel

Berapa banyak sampel yang akan diambil agar memperoleh hasil yang bermakna?

Rumus Slovin

$$n = \frac{N}{1 + N\alpha^2},$$

biasanya $\alpha = 5\%$.

Rumus Issac dan Michael

$$n = \frac{\lambda N p q}{\alpha^2 (N - 1) + \lambda^2 p q},$$

dengan λ adalah nilai *Chi-Square* dengan derajat bebas $db = 1$ ($\alpha = 1\%$, atau 5% , atau 10%). Biasanya $\alpha = 5\%$, dan $p = q = 0.5$.

Tabel: $\lambda \sim \chi^2(db = 1, \alpha)$

α	λ
1%	6.635
5%	3.841
10%	2.706

Menurut Gay dan Diehl (1992)

- penelitian deskriptif, minimal 10% dari populasi
- penelitian korelasi (hubungan), minimal 30 pengamatan
- penelitian komparatif, minimal 30 pengamatan
- penelitian eksperimental berkelompok, minimal 15 pengamatan per kelompok

Menurut Roscoe (1982)

- ukuran (banyaknya) sampel adalah antara 50 sampai 500
- jika sampel dibagi per kategori, maka ukuran sampel setiap kategori tersebut minimal 30
- untuk analisis multivariat, banyaknya sampel adalah 10 kali dari banyaknya variabel yang diteliti
- untuk eksperimen perbandingan sederhana (menggunakan eksperimen kontrol dan eksperimen perlakuan), maka ukuran sampel masing-masing eksperimen adalah antara 10 sampai 20

Contoh Soal

1. Jelaskan pengertian dari teknik sampling berikut dan berikan contohnya:
 - a) *simple random sampling*
 - b) *systematic random sampling*
 - c) *stratified random sampling*
 - d) *cluster random sampling*

Contoh Soal

2. Jelaskan pengertian dari teknik sampling berikut dan berikan contohnya:
- a) *purposive sampling*
 - b) *accidental sampling*
 - c) *quota sampling*
 - d) *snowball sampling*

Contoh Soal

3. Dengan menggunakan **rumus Slovin**, tentukan banyaknya sampel yang akan diambil jika ingin meneliti karakteristik siswa kelas X dari suatu sekolah yang terdiri dari 320 siswa!

4. Berdasarkan soal No. 3, tentukan ukuran sampelnya jika siswa kelas X tersebut terdiri dari 10 kelas dengan setiap kelasnya ada 32 siswa!

Contoh Soal

5. Dengan menggunakan **rumus Issac dan Michael**, tentukan banyaknya sampel yang akan diambil jika ingin meneliti karakteristik siswa kelas X dari suatu sekolah yang terdiri dari 320 siswa!

6. Berdasarkan soal No. 5, tentukan ukuran sampelnya jika siswa kelas X tersebut terdiri dari 10 kelas dengan setiap kelasnya ada 32 siswa!

Pengertian hipotesis

Hipotesis adalah jawaban sementara terhadap rumusan masalah penelitian, karena masih berdasarkan teori yang relevan, belum didasarkan pada fakta-fakta empiris.

Hipotesis penelitian vs hipotesis statistik

Hipotesis yang dirumuskan untuk populasi (keadaan data yang diambil), maka disebut hipotesis penelitian.

Namun jika yang diteliti merupakan sampel (dari suatu populasi), untuk mengetahui keadaan (generalisasi) populasi (yang dipilih dengan teknik sampling), maka disebut hipotesis statistik.

Dengan kata lain, hipotesis statistik digunakan untuk menguji apakah hipotesis penelitian (yang hanya diuji dengan data sampel tersebut) berlaku untuk populasi?

Contoh hipotesis penelitian

- 1 Ada kecenderungan mahasiswa baru untuk memilih ukuran baju L (hipotesis deskriptif)
- 2 Terdapat perbedaan nilai ujian sebelum (kontrol) dan sesudah (perlakuan) dilakukannya suatu metode pengajaran terbaru (hipotesis komparatif)
- 3 terdapat hubungan antara nilai ujian nasional dengan nilai IPK mahasiswa semester 2 (hipotesis asosiasif)

Contoh hipotesis statistik

- 1 Ada kecenderungan mahasiswa baru **yang signifikan** untuk memilih ukuran baju L (hipotesis deskriptif)
- 2 Terdapat perbedaan **yang signifikan** antara nilai ujian sebelum (kontrol) dan sesudah (perlakuan) dilakukannya suatu metode pengajaran terbaru (hipotesis komparatif)
- 3 terdapat hubungan **yang signifikan** antara nilai ujian nasional dengan nilai IPK mahasiswa semester 2 (hipotesis asosiasif)

Contoh hipotesis statistik

- 1 Ada kecenderungan mahasiswa baru **yang signifikan** untuk memilih ukuran baju L (hipotesis deskriptif)
- 2 Terdapat perbedaan **yang signifikan** antara nilai ujian sebelum (kontrol) dan sesudah (perlakuan) dilakukannya suatu metode pengajaran terbaru (hipotesis komparatif)
- 3 terdapat hubungan **yang signifikan** antara nilai ujian nasional dengan nilai IPK mahasiswa semester 2 (hipotesis asosiasif)

Apa itu ”signifikan”?



Kesepakatan ya 😊

Untuk selanjutnya, yang dimaksud hipotesis adalah hipotesis statistik.

Pengantar pengujian hipotesis

Hipotesis merupakan pernyataan sementara (mengenai populasi) yang mungkin benar atau tidak, sehingga perlu diuji kebenarannya. Hipotesis ada dua, yaitu hipotesis awal (H_0) dan hipotesis alternatif (H_1).

- Hipotesis nol (H_0): pernyataan yang mengandung tanda kesamaan, yaitu $=, \geq, \leq$
- Hipotesis tandingan (H_1): pernyataan yang mengandung tanda $\neq, >, <$

Hipotesis nol (H_0)

- menyatakan **tidak ada perbedaan** sesuatu kejadian antara dua kelompok.
Contoh: Tidak ada perbedaan keseriusan belajar antara siswa IPA dan IPS
- menyatakan **tidak ada hubungan** antara variabel satu dengan variabel yang lain.
Contoh: Tidak ada hubungan pelajaran dengan keseriusan belajar

Hipotesis alternatif (H_1) atau hipotesis tandingan

- menyatakan **ada perbedaan** sesuatu kejadian antara dua kelompok.
Contoh: Ada perbedaan keseriusan belajar antara siswa IPA dan IPS
- menyatakan **ada hubungan** antara variabel satu dengan variabel yang lain.
Contoh: Ada hubungan pelajaran dengan keseriusan belajar

Dalam pengujian hipotesis, selalu dihadapkan suatu kesalahan pengambilan keputusan.

Keputusan	Keadaan sebenarnya	
	H_0 benar	H_0 salah
H_0 ditolak	Kesalahan tipe I	Keputusan yang benar
H_0 tidak ditolak	Keputusan yang benar	Kesalahan tipe II

Ada dua jenis kesalahan pengambilan keputusan dalam uji statistik:

1. Kesalahan tipe I (α):

- **Kesalahan menolak H_0** padahal sesungguhnya justru H_0 yang benar.
- peluang kesalahan tipe ini adalah α (misal 5%). Disebut juga tingkat signifikansi.
- peluang untuk tidak membuat kesalahan tipe ini adalah $1 - \alpha$. Disebut juga tingkat kepercayaan.

2. Kesalahan tipe II (β):

- **Kesalahan tidak menolak H_0** padahal sesungguhnya H_0 salah.
- peluang kesalahan tipe ini adalah β . Disebut juga tingkat kekuatan uji (*power of the test*).

No. Absen	Kelas A	Kelas B
1	70	72
2	70	72
3	70	72
4	69	72
5	68	72
6	68	69
7	65	66
8	76	72
9	69	59
10	75	54

- Apakah yakin ukuran pemusatan dan ukuran penyebaran antara hasil ujian di kelas A dan hasil ujian di kelas B adalah sama?
- Apakah yakin tidak ada perbedaan nilai antara hasil ujian di kelas A dan hasil ujian di kelas B?

Hipotesis

H_0 : Tidak ada perbedaan nilai antara hasil ujian di kelas A dan hasil ujian di kelas B.

H_1 : Ada perbedaan nilai antara hasil ujian di kelas A dan hasil ujian di kelas B.

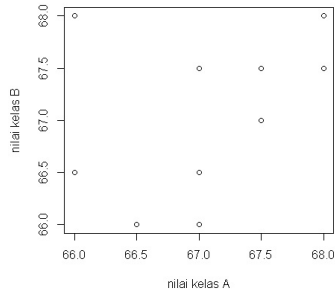
Hipotesis (menggunakan simbol)*

$$H_0: \mu_A = \mu_B$$

$$H_1: \mu_A \neq \mu_B$$

μ = rata-rata populasi

Scatter plot nilai ujian siswa kelas A dan kelas B



- Apakah yakin tidak ada hubungan antara nilai di kelas A dan nilai di kelas B?

Hipotesis

H_0 : Tidak ada hubungan antara nilai di kelas A dan nilai di kelas B.

H_1 : Ada hubungan antara nilai di kelas A dan nilai di kelas B.

Hipotesis (menggunakan simbol)*

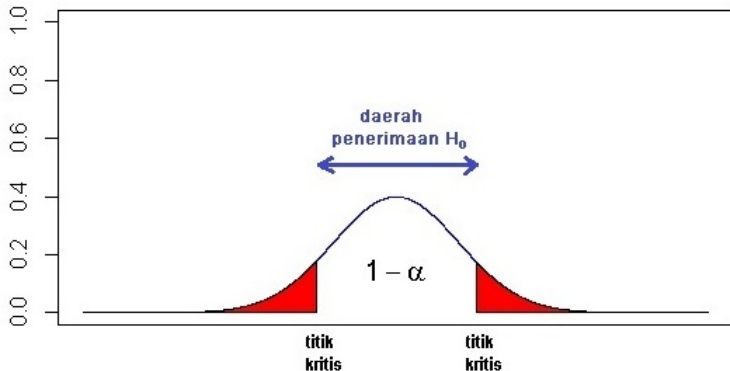
$H_0: \rho = 0$

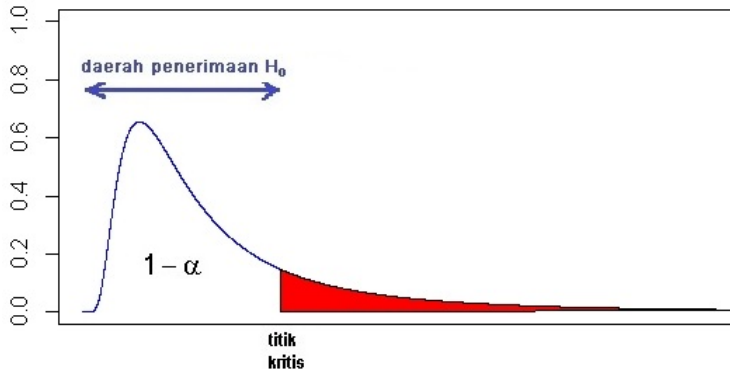
$H_1: \rho \neq 0$

ρ = korelasi, asosiasi atau hubungan

Statistik uji dan titik kritis

- Statistik uji digunakan untuk menguji hipotesis statistik yang telah dirumuskan. Notasinya berpadanan dengan jenis distribusi yang digunakan.
- Titik kritis membatasi daerah penolakan dan penerimaan H_0 .
- Titik kritis diperoleh dari tabel statistik yang bersangkutan.
- H_0 ditolak jika nilai statistik uji jatuh di daerah kritis.
- Daerah kritis = daerah penolakan H_0 .
- Selain titik kritis, juga dapat menggunakan peluang.





Penerimaan suatu hipotesis terjadi karena **tidak cukup bukti** untuk **menolak** hipotesis tersebut dan **bukan karena hipotesis itu benar.**

Penolakan suatu hipotesis terjadi karena **tidak cukup bukti** untuk **menerima** hipotesis tersebut dan **bukan karena hipotesis itu salah.**

Langkah pengujian hipotesis

Langkah pengujian hipotesis:

1. Rumuskan hipotesis
2. Pilih taraf signifikansi α (biasanya $\alpha = 5\%$ atau 0.05)
3. Pilih uji statistik yang sesuai dan cari daerah kritisnya
4. Hitung nilai statistik (atau peluangnya)
5. Buat kesimpulan berdasarkan kriteria penolakan

Contoh Soal

1. Tentukan kriteria hipotesis yang baik!

Jawab:

Contoh Soal

2. Tentukan kegunaan dari hipotesis!

Jawab:

Contoh Soal

3. Jelaskan perbedaan antara hipotesis deskriptif, komparatif, dan asosiatif!

Jawab:

Contoh Soal

4. Ambil data skripsi pendidikan Geografi, lalu buatlah hipotesis statistiknya yang terkait dengan perbedaan!

Jawab:

Contoh Soal

5. Ambil data skripsi pendidikan Geografi, lalu buatlah hipotesis statistiknya yang terkait dengan hubungan!

Jawab:

Normalisasi Data

Semua gejala di alam, diyakini mengikuti distribusi Normal. Kurva yang dihasilkan oleh distribusi Normal menyerupai lonceng. Distribusi ini disebut juga distribusi Gauss, karena ditemukan oleh Karl Friedrich Gauss (1777-1855).

Perhatikan data hasil pengukuran tinggi badan sejumlah mahasiswa/i di suatu kelas berikut

Mahasiswa		Mahasiswi		
169	175	148	148	159
159	173	138	154	159
170	168	142	151	153
165	153	148	159	147
161	170	147	148	152

Statistik	Mahasiswa	Mahasiswi
min	153	138
median	168.5	148
maks	175	159
<i>range</i>	22	21
<i>n</i>	10	15
$k = 1 + 3.3(\log n)$	4	5
<i>length</i>	6	4

Mahasiswa		Mahasiswi	
Interval	Frekuensi	Interval	Frekuensi
153-158	1	138-141	1
159-164	2	142-145	1
165-170	5	146-149	6
171-176	2	150-153	3
		154-157	1
		158-161	3

Histogram

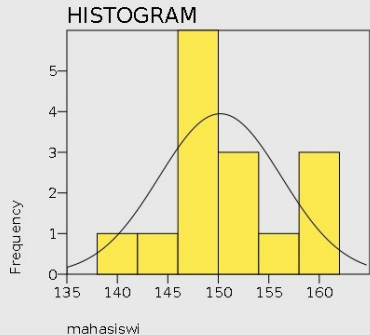
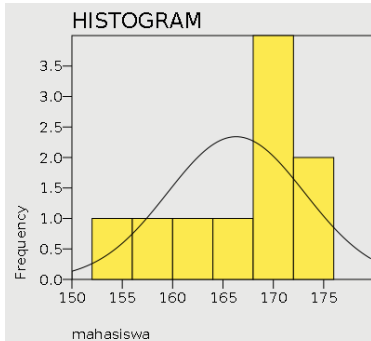


Diagram batang dan daun (*Stem and Leaf Plots*)

Mahasiswa	Mahasiswa	Mahasiswi	Mahasiswi
15 3, 9	15 3	13 8	13 8
16 1, 5, 8, 9	15 9	14 2, 7, 7, 8, 8, 8, 8	14 2
17 0, 0, 3, 5	16 1	15 1, 2, 3, 4, 9, 9, 9	14 7, 7, 8, 8, 8, 8
	16 5, 8, 9		15 1, 2, 3, 4
	17 0, 0, 3		15 9, 9, 9
	17 5		

Menguji apakah kedua distribusi sama?

Mahasiswa		Mahasiswi
	13	8
	14	2
	14	7 7 8 8 8 8
3	15	1 2 3 4
9	15	9 9 9
1	16	
9 8 5	16	
3 0 0	17	
5	17	

Pengertian uji normalitas

Misalkan X adalah data sampel tunggal (univariat) kontinu. Untuk membuktikan secara ilmiah, apakah distribusi X berbentuk "lonceng", maka perlu dilakukan pengujian hipotesis (yaitu uji normalitas), antara lain:

- **uji Kolmogorov-Smirnov**
- uji Anderson-Darling
- uji Jarque-Bera
- uji Shapiro-Wilk
- uji Lilliefors
- uji Pearson's chi-squared

Fungsi kepadatan peluang (*probability density function*) dari variabel acak X berdistribusi Normal adalah

$$f(x : \mu, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

Jika $\mu = 0$ dan $\sigma = 1$, maka disebut berdistribusi Normal baku, ditulis $X \sim N(0, 1)$, dengan fungsi kepadatan peluangnya

$$f(x : \mu = 0, \sigma = 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Fungsi uji normalitas

Suatu data sampel perlu diuji apakah mengikuti distribusi Normal atau tidak. Jika data sampel tersebut berdistribusi Normal, maka

- berlaku aturan empiris
- dapat menggunakan tabel t atau z dalam **pengujian hipotesis**
- $\text{mean} \approx \text{median} \approx \text{modus}$ (ukuran pemusatan)
- $\mu = \bar{x}$ dan $\sigma = ns$ (dengan n adalah ukuran sampel)

Prosedur pengujian normalitas data

Uji Kolmogorov-Smirnov (KS) adalah uji normalitas yang paling umum. Untuk mempermudah dalam menggunakan uji ini, data harus diurutkan terlebih dahulu.

Dengan $\alpha = 5\%$, selidikilah apakah data tinggi badan Mahasiswa berikut diambil dari populasi yang berdistribusi Normal?

Mahasiswa	
169	175
159	173
170	168
165	153
161	170

Penyelesaian

Hipotesis

H_0 : Sampel tinggi badan Mahasiswa berdistribusi Normal.

H_1 : Sampel tinggi badan Mahasiswa tidak berdistribusi Normal.

Taraf signifikansi

Gunakan $\alpha = 5\%$.

Uji statistik

Penelitian ini menggunakan statistik uji Kolmogorov-Smirnov.

Kriteria penolakan

H_0 ditolak jika nilai $D = \max(|F_t - F_s|)$ (hitung) lebih dari D_{tabel} .

x	urutan kumulatif	z	F_t	F_s	$ F_t - F_s $
153	1	-1.95	0.0256	0.1	0.0744
159	2	-1.07	0.1423	0.2	0.0577
161	3	-0.78	0.2177	0.3	0.0823
165	4	-0.19	0.4247	0.4	0.0247
168	5	0.25	0.5987	0.5	0.0987
169	6	0.40	0.6554	0.6	0.0554
170	8	0.54	0.7054	0.8	0.0946
170	8	0.54	0.7054	0.8	0.0946
173	9	0.98	0.8365	0.9	0.0635
175	10	1.28	0.8997	1	0.1003

Dalam hal ini, $z = (x - \bar{x})/s$, F_t diperoleh dari tabel z , F_s = urutan kumulatif dibagi n . Sehingga diperoleh $D = \max(|F_t - F_s|) = 0.1003$.

Keputusan

Karena nilai $D_{hitung} = 0.1003 < 0.4090 = D_{tabel}$, maka H_0 tidak ditolak.

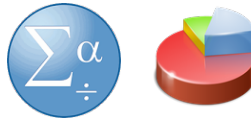
Kesimpulan

Data sampel tinggi badan Mahasiswa berdistribusi Normal.

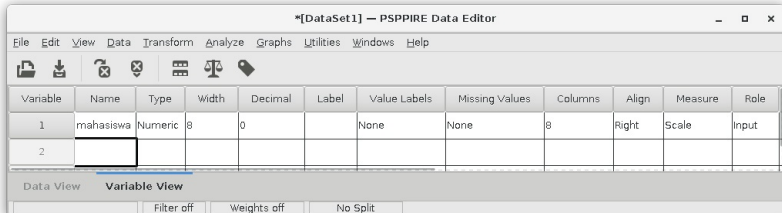
Pengertian Statistik
Memahami Data
Ukuran Pemusatan Data (nilai sentral)
Ukuran Penyebaran Data (dispersi)
Populasi dan Sampel
Konsep Dasar Pengujian Hipotesis
Normalisasi Data

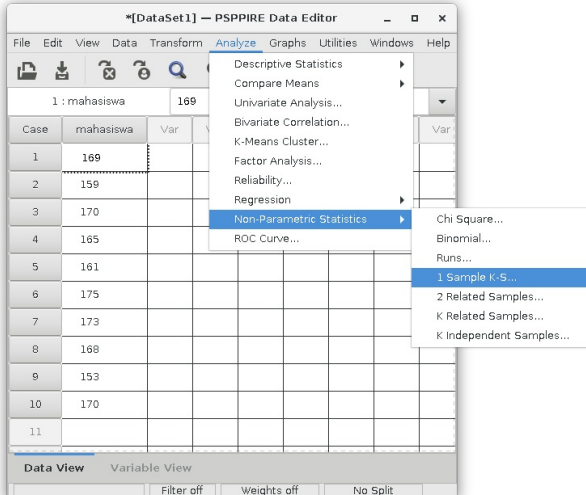
Pengertian uji normalitas
Fungsi uji normalitas
Prosedur pengujian normalitas data
Pengujian normalitas data menggunakan PSPP
Transformasi ke bentuk Normal Baku

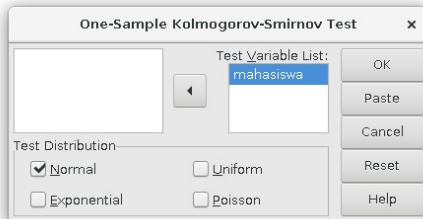
Pengujian normalitas data menggunakan PSPP



Gambar: SPSS dan PSPP







		<i>mahasiswa</i>
<i>N</i>		10
<i>Normal Parameters</i>		
	<i>Mean</i>	166.30
	<i>Std. Deviation</i>	6.82
<i>Most Extreme Differences</i>		
	<i>Absolute</i>	.20
	<i>Positive</i>	.10
	<i>Negative</i>	-.20
<i>Kolmogorov-Smirnov Z</i>		.63
<i>Asymp. Sig. (2-tailed)</i>		.826

Transformasi ke bentuk Normal Baku

Bagaimana jika data sampel tidak Normal?

- mengabaikannya
- reduksi data, dengan cara menghilangkan *outlier*
- *resampling*, dengan cara simulasi
- mengambil data ulang, untuk memperoleh sampel yang lebih representatif
- **transformasi ke bentuk Normal**

Misalkan X adalah data sampel asli (sebelum dilakukan transformasi), maka transformasinya adalah

$$z_i = \frac{x_i - \bar{x}}{s},$$

dengan \bar{x} dan s masing-masing adalah rata-rata dan deviasi standar.

Diketahui mean sampel tinggi badan Mahasiswa adalah 166.3 dengan simpangan baku (deviasi standar) 6.816.

Mahasiswa	z
169	0.396
159	-1.071
170	0.543
165	-0.191
161	-0.778
175	1.276
173	0.983
168	0.249
153	-1.951
170	0.543 $\approx \frac{170-166.3}{6.816}$

Contoh Soal

1. Berikut adalah nilai hasil ujian remedial dari suatu kelas.

52	74	4	28	48	56	16
74	52	28	16	74	74	

Dengan $\alpha = 5\%$, selidikilah apakah data tersebut diambil dari populasi yang berdistribusi Normal?

Contoh Soal

2. Berikut adalah nilai ujian akhir dari suatu kelas.

90	90	90	90	90	56	90
90	84	60	12	90	36	80
64	90	90	90			

Dengan $\alpha = 5\%$, selidikilah apakah data tersebut diambil dari populasi yang berdistribusi Normal?

Contoh Soal

3. Kampus mencatat partisipasi peserta senam beserta berat badan dengan sampel yang diambil sebanyak 27 orang diambil secara acak. Adapun data berat badan yang diperoleh adalah sebagai berikut:

78, 78, 95, 90, 78, 80, 82, 77, 72, 84, 68, 67, 87, 78, 77, 88, 97, 89
97, 98, 70, 72, 70, 69, 67, 90, 97

Dengan $\alpha = 5\%$, selidikilah apakah data tersebut diambil dari populasi yang berdistribusi Normal?