



UNIVERSITAS
GADJAH MADA

Statistika

Herni Utami

Sleman, 13 Agustus 2020

Departemen Matematika
Fakultas Matematika dan Ilmu Pengetahuan Alam,

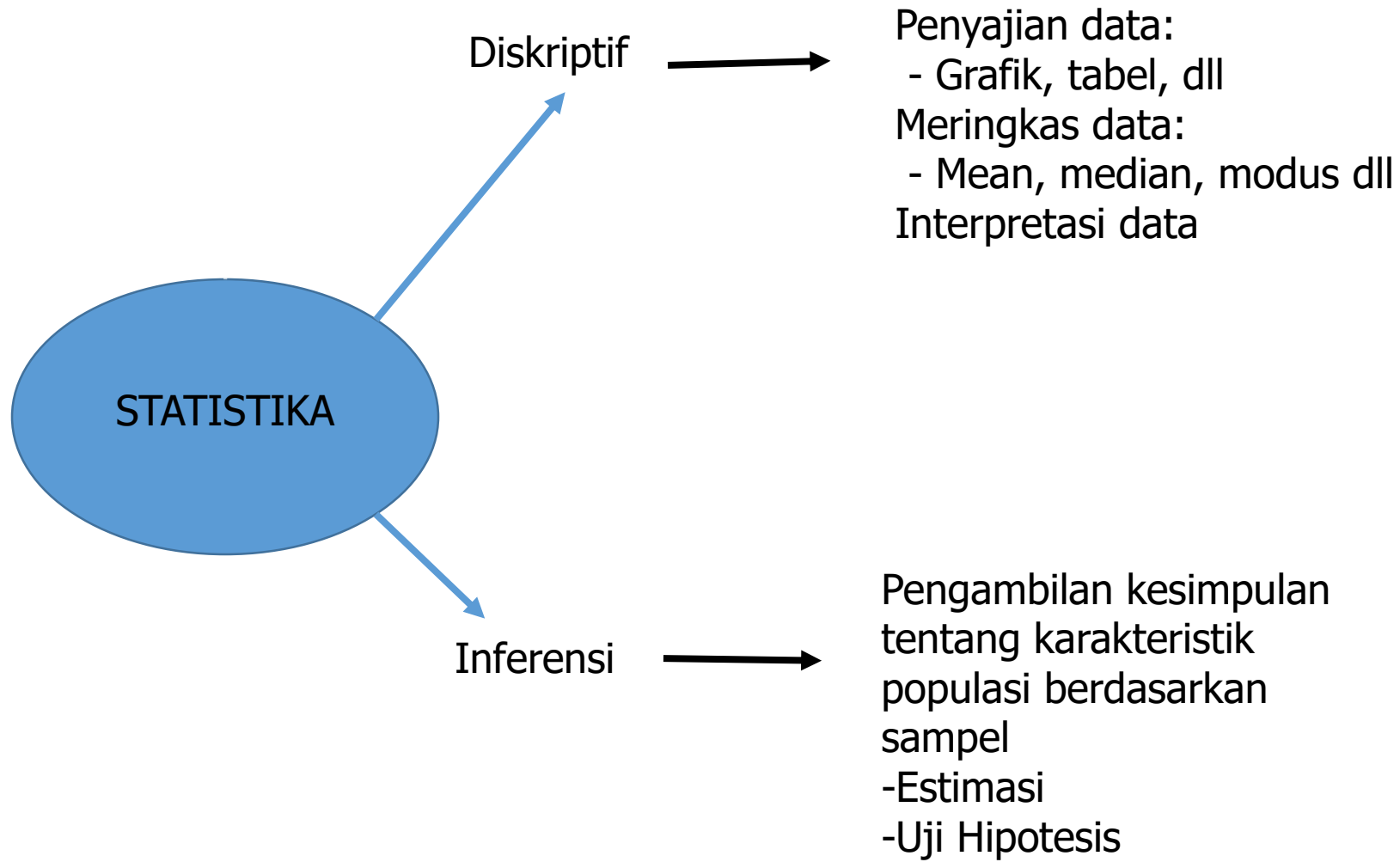


APAKAH STATISTIKA?



“Statistics is the study of the collection, analysis, interpretation, presentation and organization of data”.

(Dodge, Y. (2006) The Oxford Dictionary of Statistical Terms)





PERANAN STATISTIKA DALAM PENELITIAN



Tahapan Penelitian

- Pertanyaan Penelitian (substansi)
- Rancangan Penelitian
- Variabel dan Pengukuran
- Pengumpulan data



Eksplorasi dan Analisis Data



Kesimpulan



Tujuan Analisis

(1) Deskriptif

- Berapa nilai rata-rata ujian nasional SMA untuk setiap mata ujian di Kabupaten Sleman?
- Berapa persentase kelulusan SMA di Kabupaten Sleman?

(2) Konfirmasi

- Apakah nilai ujian nasional dipengaruhi status sekolah negeri atau swasta?
- Apakah kemampuan menulis esai dalam Bahasa Indonesia meningkat setelah metode *project based learning* diterapkan di kelas?

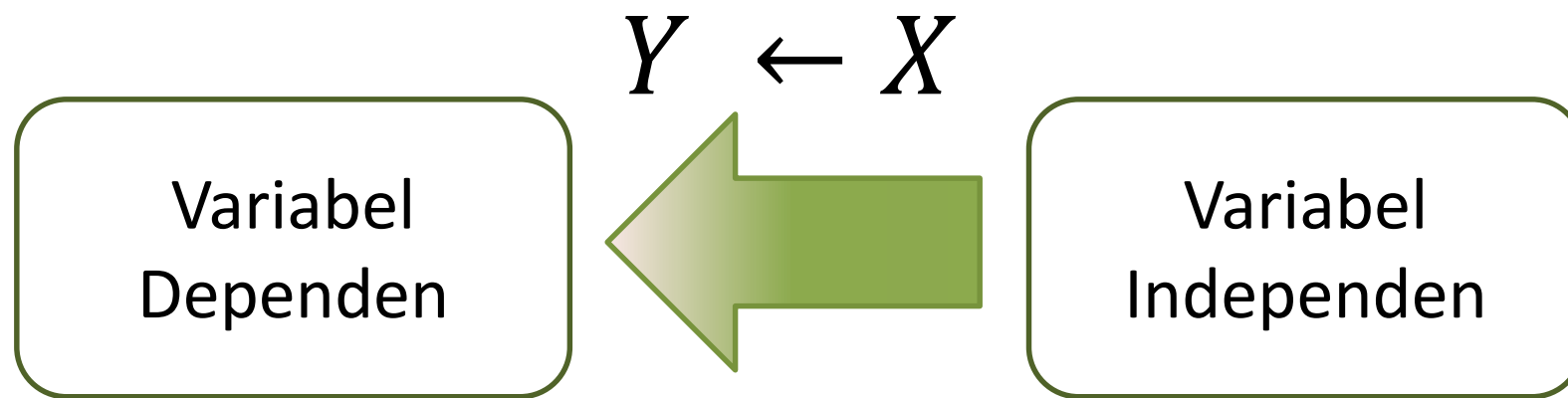
(3) Prediksi

- Nilai rata-rata kelas ujian nasional Bahasa Inggris akan naik minimal 10 *point*, jika pada suatu kelas dikenakan metode pembelajaran Bahasa Inggris *SMART*
- Waktu tempuh *sprint* 100 m akan lebih cepat paling tidak 2 detik, jika metode latihan *Super-Sprint* diterapkan pada siswa olah raga



Variabel dan Data

Variabel: karakteristik atau informasi yang diperoleh dari individu/unit sampel



Indikator, *outcome*, respon,
performance, status
output, hasil

prediktor, variabel penjelas,
perlakuan, *exposure*, faktor,
Determinan, *stimulus*,
task, input

Skala Pengukuran

Data: realisasi atau nilai variabel



Variabel dan Data

Data sampel Nilai UN 300 siswa

variabel

No	ID-Peserta	JK	B-Ind	Mtk	B-Ing	IPA	Neg/Sw
1	02-152-001	L	64,00	72,50	50,00	50,00	Negeri
2	03-141-014	P	78,00	57,50	92,00	72,50	Negeri
3	07-082-031	L	74,00	85,00	74,00	87,50	Swasta
4	02-110-025	P	68,00	72,50	54,00	55,00	Negeri
...
300	08-115-001	L	88,00	87,50	88,00	72,50	Negeri

data



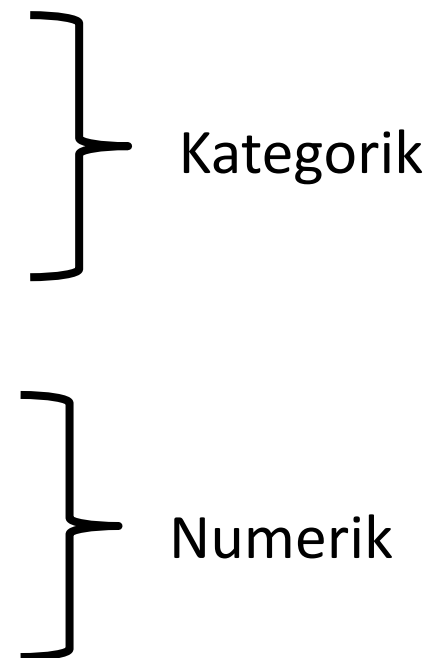
Tipe Data dan Skala Pengukuran

- Data kategorik
- Data numerik
 - Data diskret
 - Data kontinu



Skala Pengukuran

Skala	Contoh
Nominal	Jenis kelamin, status kelulusan, status kepemilikan sekolah
Ordinal	tingkat pendidikan, status akreditasi, kepangkatan
Interval	tahun kalender, tanggal lahir, temperature
Rasio	Biaya pendidikan, banyaknya siswa, luas tanah





Tahapan Analisis Data

- Identikasi variabel yang menjadi perhatian (jenis, skala pengukuran, asumsi hubungan antar variabel)
- Lakukan **deskripsi** dan eksplorasi data (ringkasan statistik, grafik, tabel) pada variabel-variabel tersebut
- Tentukan metode statistik yang sesuai (mempertimbangkan rancangan penelitian, cara pengumpulan data, dan asumsi data)
- Analisis data dan kesimpulan



Deskripsi Data

- Ringkasan Statistik
 - Ukuran Pemusatan: *mean, median, modus*
 - Ukuran Penyebaran: *range, sd.*
- Tabel
 - Tabel Distribusi frekuensi
 - Tabel ringkasan statistik
- Grafik
 - Distribusi (bentuk) data
 - Hubungan antar variabel dalam data



Ringkasan Statistik

Statistik	Penggunaan Tepat	Penggunaan Tidak tepat
Mean	Data interval atau rasio Data simetris terhadap nilai mean	Data kategorikal atau ordinal Ada data pencilan (<i>outliers</i>) atau data yang sangat menceng (<i>skewed</i>)
	Diperlukan untuk penghitungan lebih lanjut (variansi, estimasi regresi)	
Median	Data Interval, rasio atau ordinal, Ada data pencilan (<i>outliers</i>) atau data yang sangat menceng (<i>skewed</i>)	Diperlukan untuk penghitungan lebih lanjut (variansi, estimasi regresi)
Modus	Kategorik	Kontinu (bilangan real)



Grafik

Grafik	Penggunaan
Histogram/boxplot	Untuk mendeskripsikan distribusi data untuk skala pengukuran data interval atau rasio
<i>Bar plot</i> (diagram batang)	Untuk mendeskripsikan frekuensi data nominal atau ordinal
<i>Pie chart</i> (diagram lingkaran)	Untuk mendeskripsikan distribusi frekuensi data nominal atau data kategorik dengan beberapa kategori
<i>Scatter plot</i> (diagram pencar)	Untuk menunjukkan hubungan antar dua variabel dalam skala pengukuran interval, rasio atau ordinal. Dua atau lebih variabel diperlukan untuk setiap unit sampel
<i>Line graph</i> (grafik garis)	Untuk menunjukkan perubahan pola atau kecenderungan (<i>trend</i>) terhadap waktu untuk skala data interval, rasio atau ordinal. Bermanfaat untuk menampilkan <i>trend</i> sekaligus untuk beberapa variabel. `



Contoh Data

Data: Tes Pemantapan Persiapan Ujian SMP/MTs

Nilai TPPU mata ujian (Bahasa Indonesia, Matematika, Bahasa Inggris, IPA) untuk setiap peserta didik di wilayah DIY (Kota Yogyakarta, Sleman, Bantul, Kulon Progo, Gunung Kidul); Tahap 1 dan Tahap 2. Ada 50.323 siswa pada 215 sekolah Negeri; 216 Swasta dan 92 Mts.



Ringkasan Statistik (numerik)

SMP (Salah satu sekolah swasta)

Stat.	B.Indo	Mtk	B.Ing	IPA	Total
Min	44,00	17,50	34,00	32,50	130,50
Median	76,00	58,75	76,00	57,50	262,25
Max	92,00	95,00	92,00	87,50	345,50
Mean	76,00	61,74	72,83	59,84	270,40
Sd	8,64	20,29	13,47	12,65	43,99
NA	-	-	-	-	-

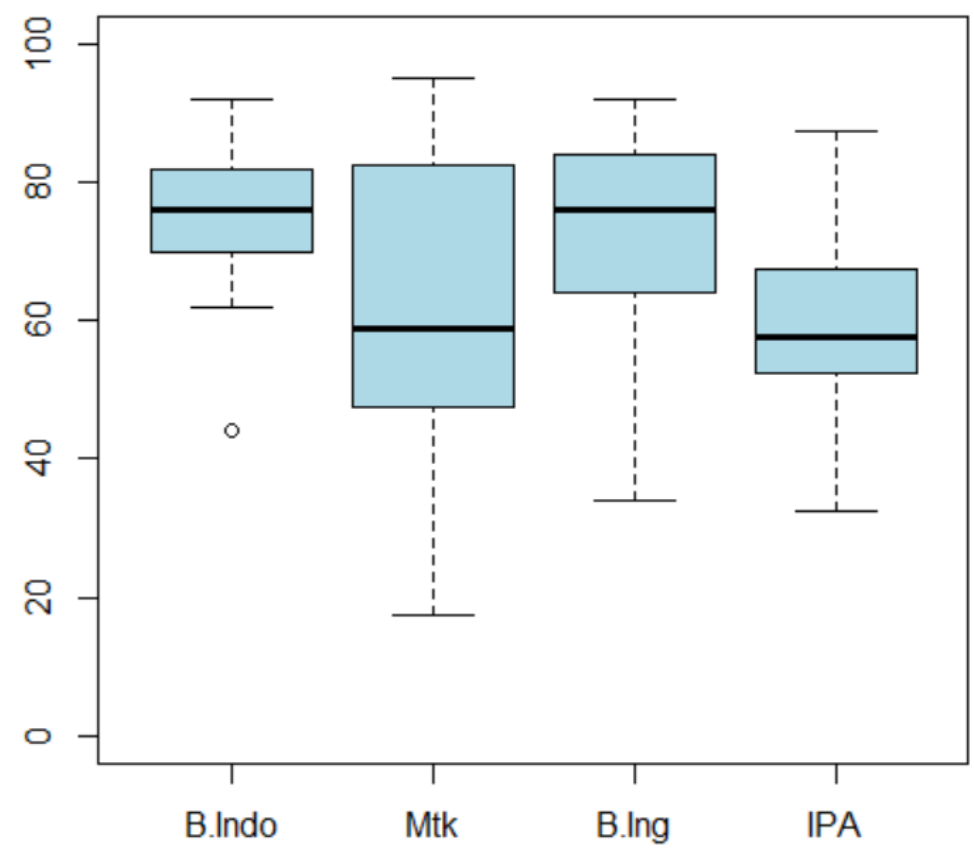
SMP/MTs (seluruh DIY)

Stat.	B.Indo	Mtk	B.Ing	IPA	Total
Min	0,00	2,50	8,00	2,50	12,00
Median	70,00	35,00	48,00	45,00	197,00
Max	100	100	100	100	387,00
Mean	68,13	41,74	50,60	47,56	207,16
Sd	12,69	19,57	17,66	16,65	59,27
NA	236	255	166	186	---

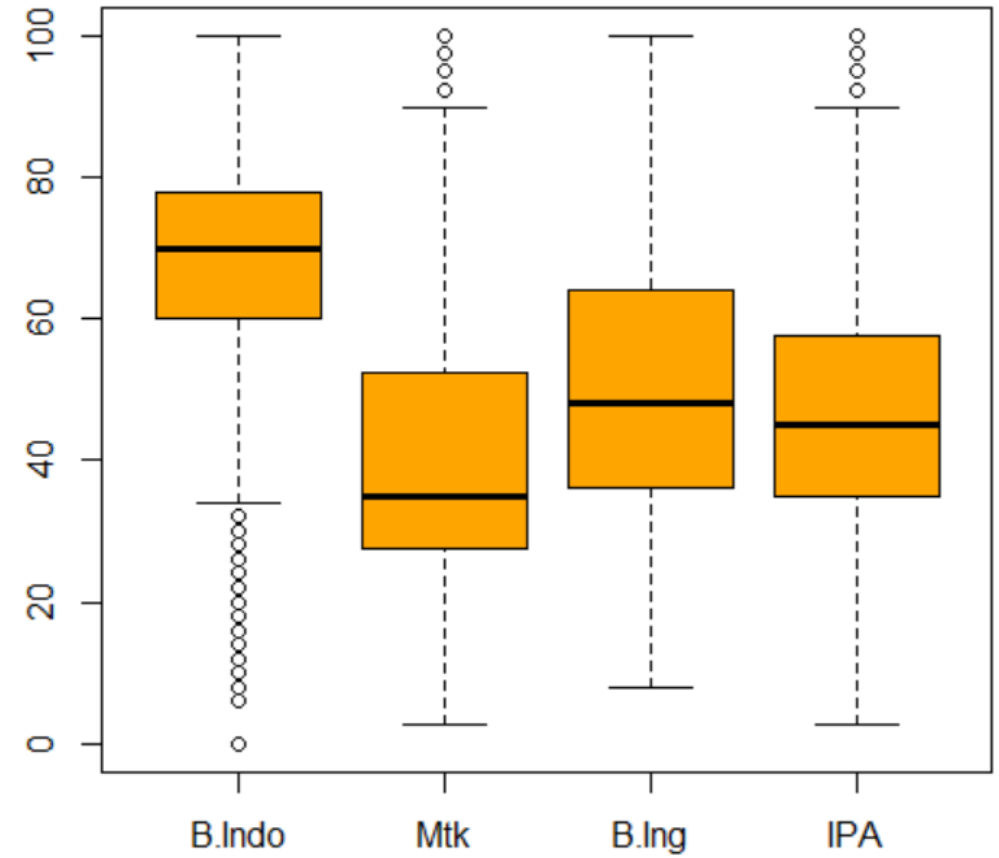


Grafik (Boxplot)

SMP (Salah satu sekolah swasta)



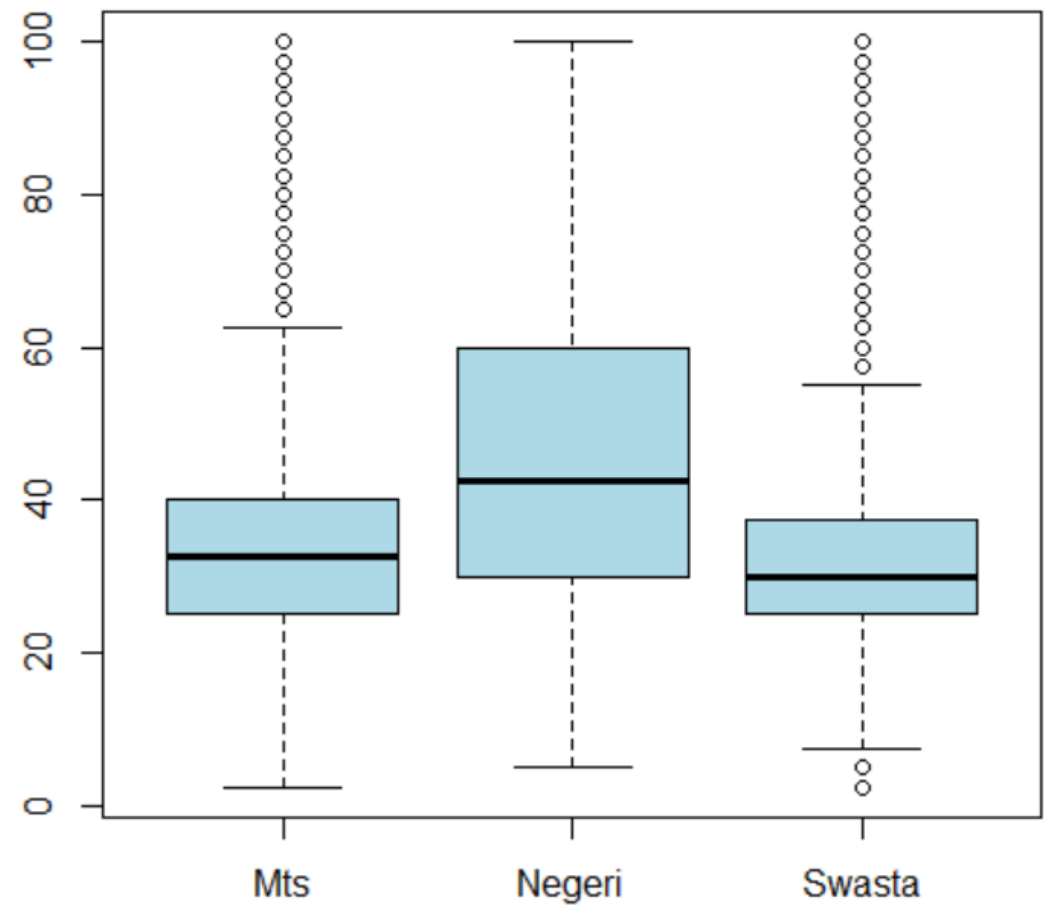
SMP/MTs (seluruh DIY)





Grafik (Boxplot)

Matematika (Tahap 1)



Matematika (Tahap 2)

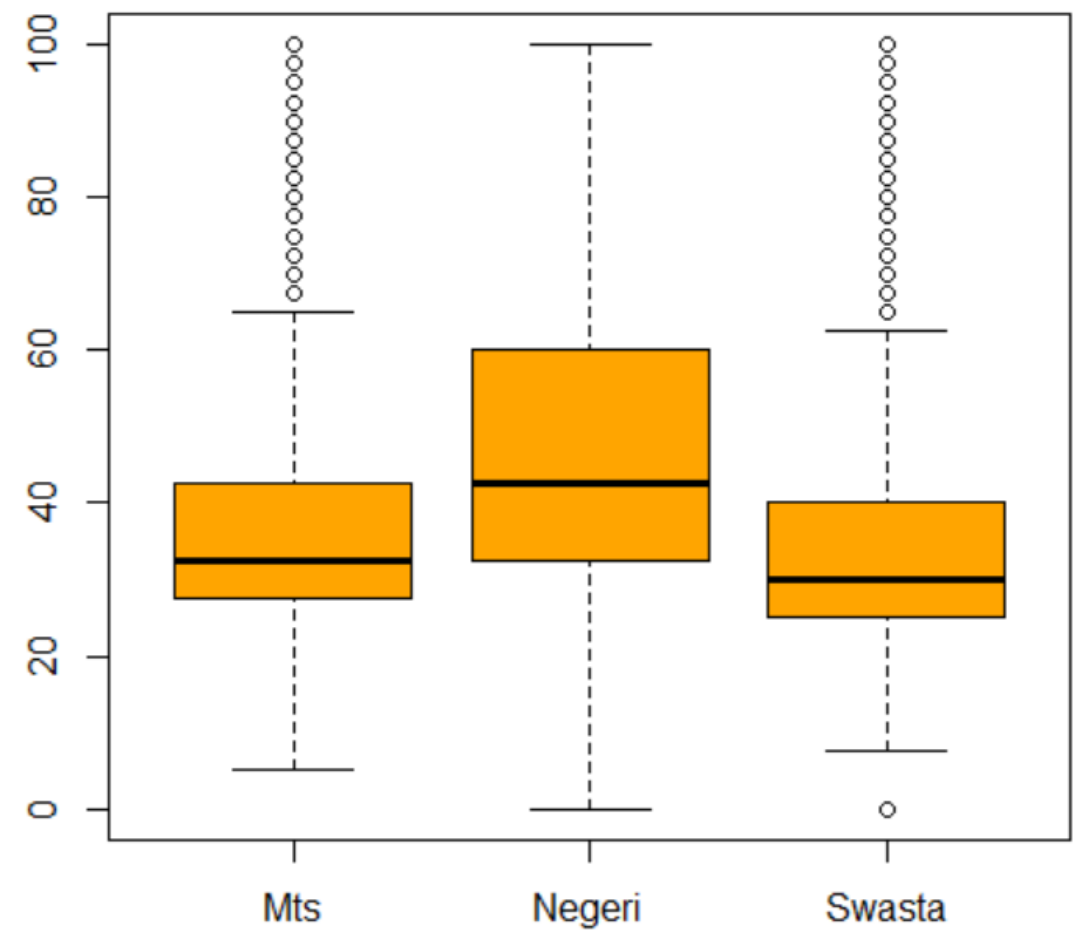




Diagram Batang (data kategorik)

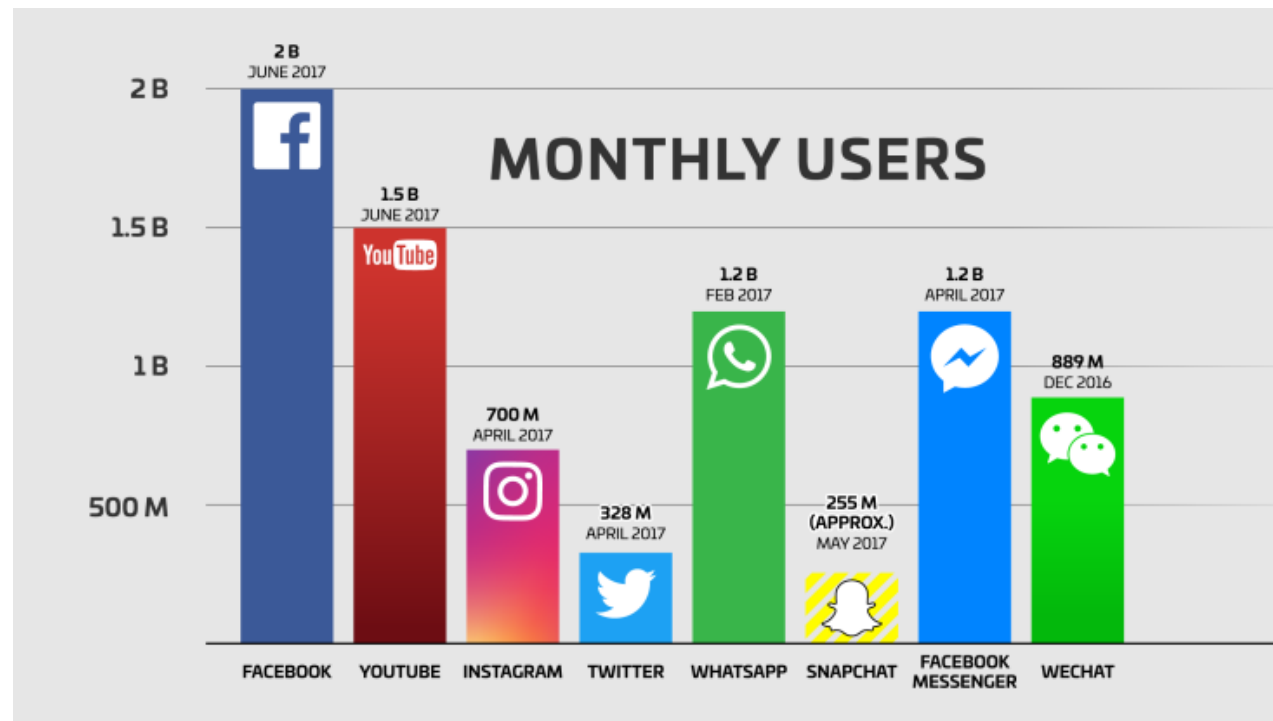


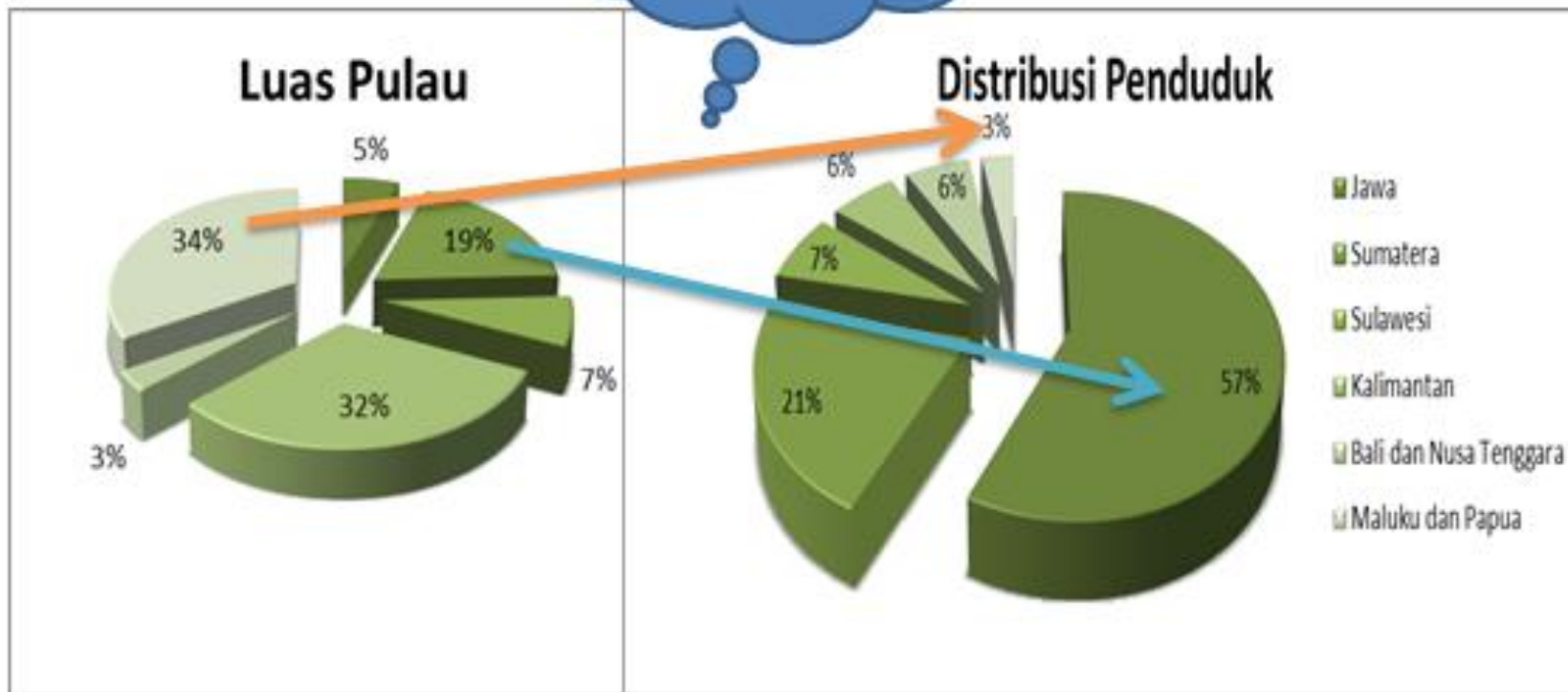


DIAGRAM LINGKARAN





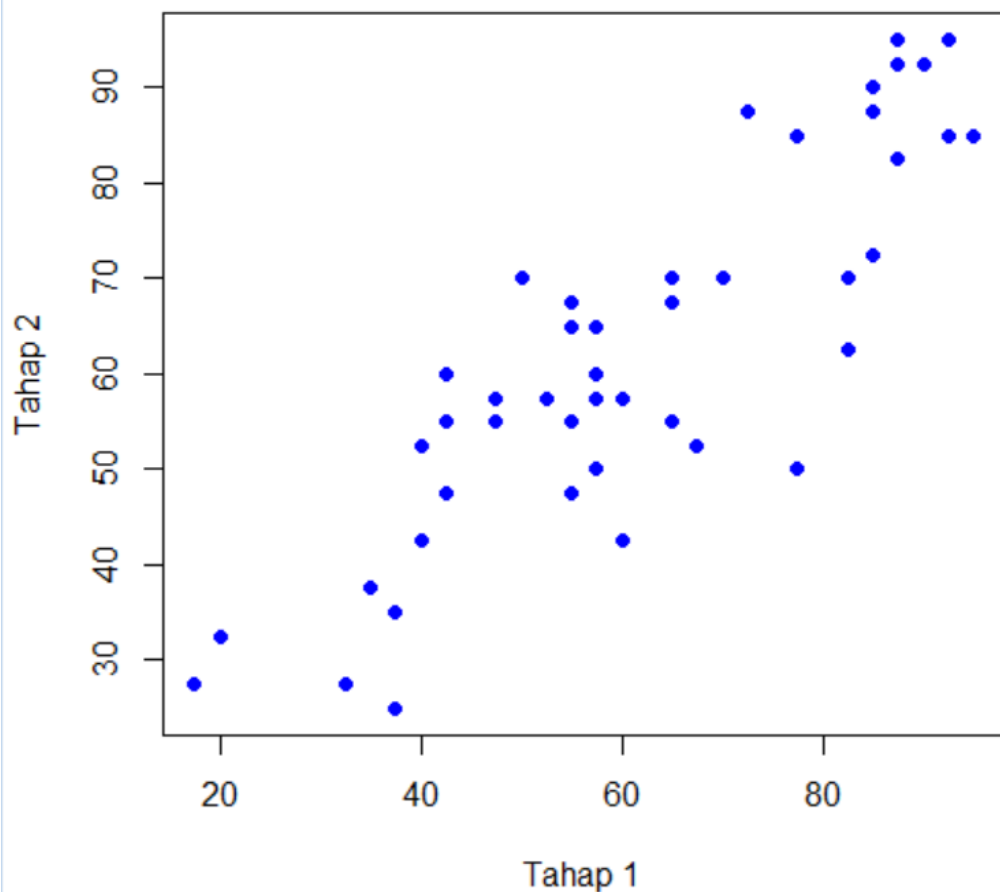
Kondisi yang miris



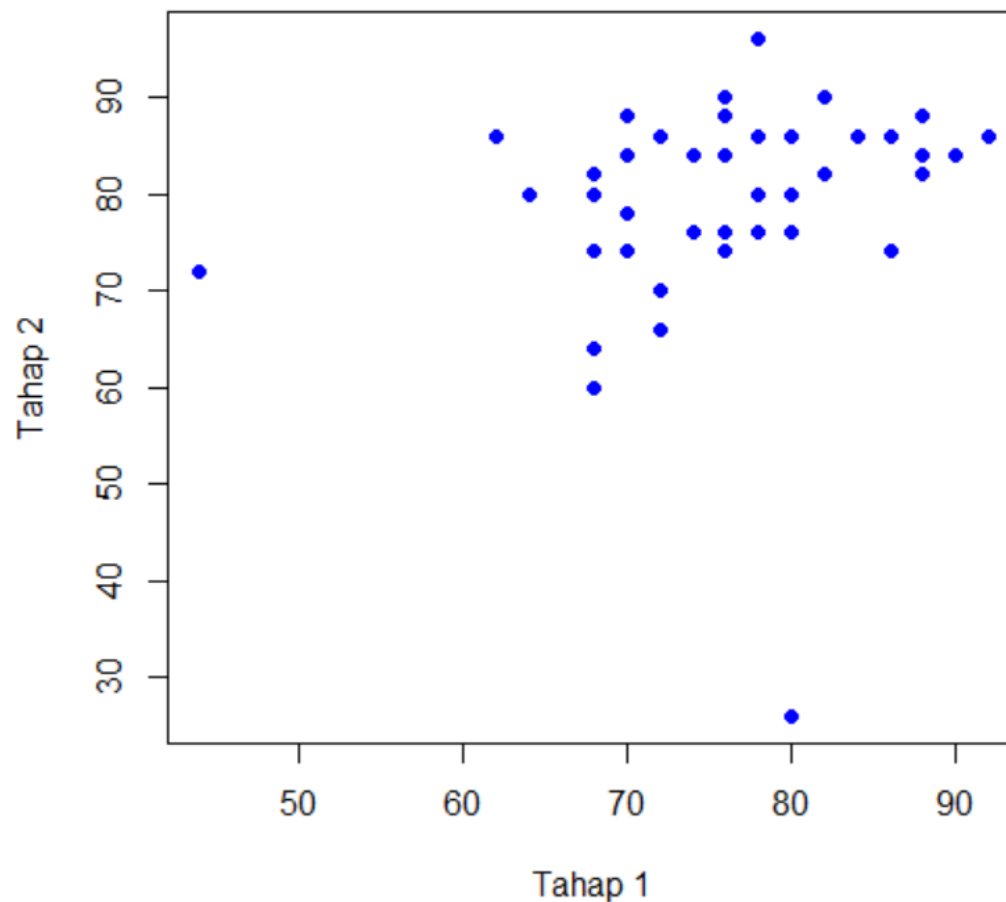


Grafik (Scatterplot)

Matematika (Satu sekolah – Swasta)



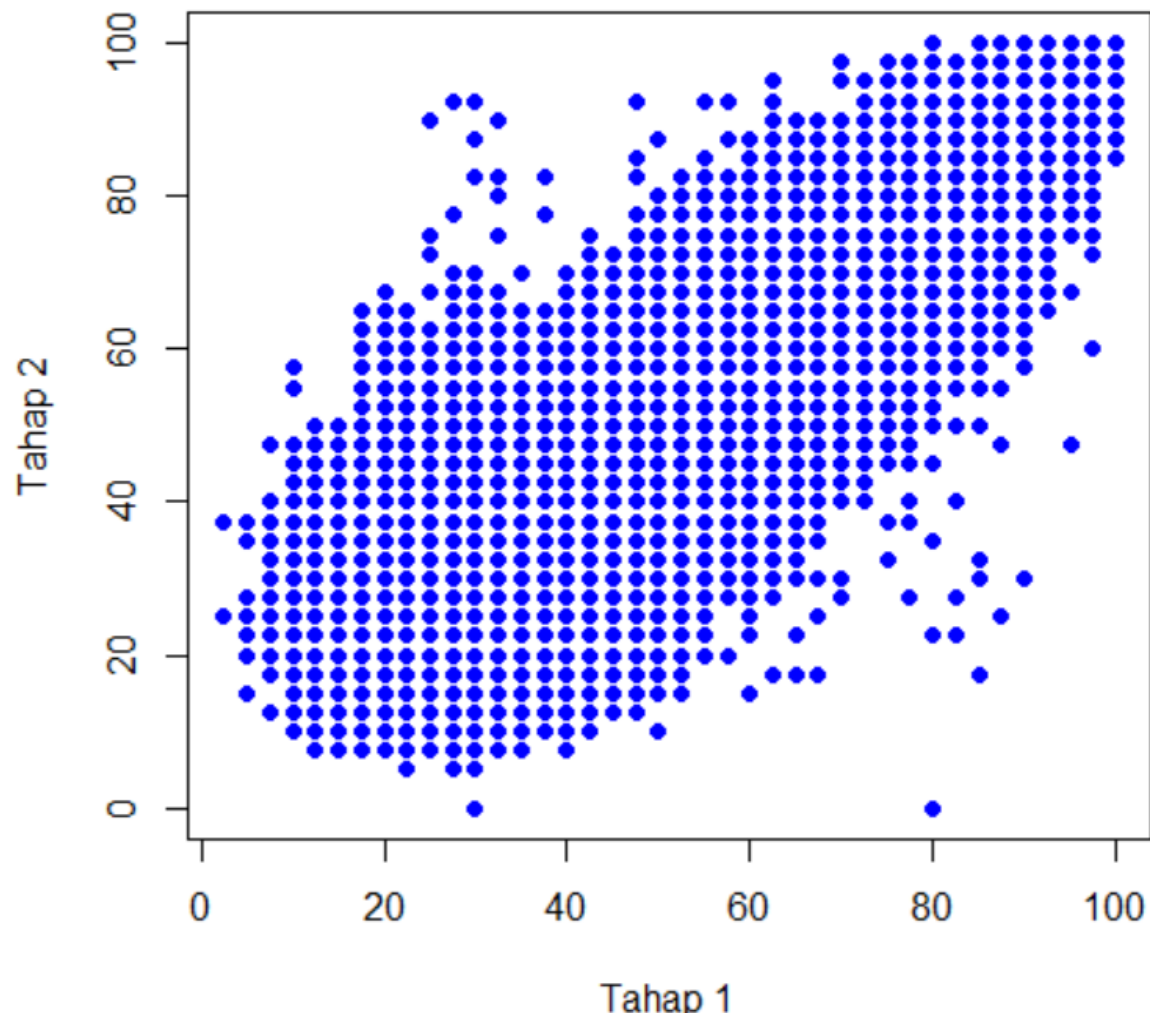
Bahasa Indonesia





Grafik (Scatterplot)

Matematika (DIY)





Ukuran pusat



Apakah Mean

- **Mean (Rata-rata)** digunakan sebagai salah satu sifat perbandingan statistik.
- Ini didefinisikan sebagai rata-rata semua klarifikasi.

Aplikasi dari mean



- Ini membantu para guru untuk melihat nilai rata-rata siswa.
- Ini digunakan pada pabrik-pabrik, agar manager perusahaan tahu kualitas barang yang diproduksinya berkualitas atau tidak.
- Ini juga digunakan untuk membedakan gaji para pekerja laki-laki dan wanita.
- Untuk menghitung kecepatan rata-rata apa pun.
- Ini juga digunakan oleh pemerintah untuk mengetahui rata-rata pemasukan atau pengeluaran negara pertahun.
- Dengan menggunakan ini pemerintah dapat menyeimbangkan rata-rata pengeluaran dengan pendapatan.



Apa itu median dan apa aplikasinya dalam kehidupan sehari-hari

Median didefinisikan sebagai nilai tengah dari suatu data/observasi.

Aplikasinya:

- Ini digunakan untuk mengukur distribusi pendapatan
- Digunakan untuk menemukan tinggi pemain misalnya pemain sepakbola.
- Untuk menemukan usia menengah dari siswa kelas.
- Juga digunakan untuk mencari garis kemiskinan.



Modus merupakan outcome yang paling sering muncul dalam suatu data.

Aplikasinya:

- Ini digunakan untuk masuknya transportasi umum.
- Banyaknya kemenangan suatu tim dalam pertandingan dengan tim lawan mana pun.
- Banyaknya kunjungan seorang pasien ke rumah sakit.



**Kapan sebaiknya
menggunakan mean,
median dan modus?**



Ukuran Sebaran Data

- Range

- Variansi

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

$$s^2 = \frac{1}{n-1} \left[\sum_{i=1}^n x_i^2 - \frac{1}{n} \left(\sum_{i=1}^n x_i \right)^2 \right]$$

- Standar Deviasi

$$s = \sqrt{s^2}$$

- Rata-rata deviasi

$$\frac{\sum_{i=1}^n |x_i - \bar{x}|}{n}$$



Konfirmasi: Metode dan Model Statistik

V-Dependen Y	V-Independen X	Metode
Kontinu	Biner	Uji-Z, Uji-t
	Kategorik, lebih dari 2 variabel	ANOVA
	Kategorik dan Kontinu	ANAKOVA Regresi Berganda (<i>multiple</i>)
Biner	Biner	Tabel 2×2 Regresi Logistik
	Kategorik	Tabel kontingensi Regresi Logistik
	Kategorik dan Kontinu	Regresi Logistik
Nominal, lebih dari 2 kategori	Kategorik dan Kontinu	Regresi Logistik Multinomial



V-Dependen Y	V-Independen X	Metode
Ordinal	Kategorik dan Kontinu	Regresi Logistik Ordinal
Cacah	Kategorik	Model Log-linear Regresi Poisson
	Kategorik dan Kontinu	Regresi Poisson
Durasi (antar-kejadian)	Kategorik	<i>Survival Analysis</i> Uji Log-Rank
	Kategorik dan Kontinu	<i>Survival Analysis</i> Regresi Cox <i>Parametric Survival Model</i>
Respon berkorelasi	Kategorik dan Kontinu	<i>Generalized Estimating Equation</i> ModelMultilevel



V-Dependen Y	V-Independen X	Metode
Rank/ordinal	Biner	Uji Mann-Whitney Uji McNemar (sampel berpasangan) Uji Wilcoxon (sampel berpasangan)
	Kategorik, 2 atau lebih	Uji Kruskal-Wallis Uji Friedman (sampel dependen)
	Kategorik dan Kontinu	<i>Rank ordered Regression</i>
Respon univariat dengan asumsi distribusi tertentu	Kategorik dan Kontinu	<i>Generalized Linear Models (GLM)</i>
Lebih dari satu variabel	Kategorik dan Kontinu	Analisis Multivariat



Contoh Analisis Data (Regresi)

Model: $Mtk_{thp2} = \beta_0 + \beta_1 Mtk_{thp1} + \beta_2 \text{status} + \beta_3 \text{daerah}$

Mtk_{thp2} : Nilai Matematika Tahap 2 (0 – 100, kontinu/skala rasio)

Mtk_{thp1} : Nilai Matematika Tahap 1 (0 – 100, kontinu/skala rasio)

status : Negeri, Swasta, MTs (kategorik)

daerah : Kota Yogyakarta, Sleman, Bantul, Kulon Progo, Gunung Kidul (kategorik)

Model: $Mtk_{thp2} = \beta_0 + \beta_1 Mtk_{thp1} + \beta_2 \text{Negeri} + \beta_3 \text{Swasta}$
 $+ \beta_4 \text{Kota} + \beta_5 \text{GKidul} + \beta_6 \text{KProgo} + \beta_7 \text{Sleman}$

Negeri, Swasta, Kota, Gkidul, Kprogo, Sleman adalah *dummy variable* bernilai 0 dan 1

Contoh: $\text{Negeri} = \begin{cases} 1 & \text{status sekolah Negeri} \\ 0 & \text{status sekolah bukan Negeri} \end{cases}$



Contoh Analisis Data (Regresi)

Hasil output analisis data:

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	8.532147	0.171869	49.643	< 2e-16	***
Mtk.x	0.791440	0.002585	306.142	< 2e-16	***
statusNegeri	1.723631	0.137425	12.542	< 2e-16	***
statusSwasta	-1.163406	0.155753	-7.470	8.19e-14	***
daerahG.KIDUL	-1.170846	0.142907	-8.193	2.61e-16	***
daerahK.PROGO	-0.945970	0.163212	-5.796	6.84e-09	***
daerahKOTA YK	1.704350	0.153269	11.120	< 2e-16	***
daerahSLEMAN	0.570592	0.130563	4.370	1.24e-05	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 9.892 on 45870 degrees of freedom

(4445 observations deleted due to missingness)

Multiple R-squared: 0.7308, Adjusted R-squared: 0.7308

F-statistic: 1.779e+04 on 7 and 45870 DF, p-value: < 2.2e-16

Analysis of Variance Table

Response: Mtk.y

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Mtk.x	1	12114013	12114013	123794.471	< 2.2e-16	***
status	2	34649	17325	177.043	< 2.2e-16	***
daerah	4	37662	9416	96.219	< 2.2e-16	***
Residuals	45870	4488648	98			

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1



Contoh Analisis Data (Regresi)

Variabel	Estimasi β	SE($\hat{\beta}$)	<i>p-value</i>
Intersep	8,532	0,172	< 0,001
MTK Tahap 1	0,791	0,003	< 0,001
Status			< 0,001
MTs	—		
Negeri	1,724	0,137	
Swasta	−1,163	0,156	
Daerah			< 0,001
Bantul	—		
Gunung Kidul	−1,171	0,143	
Kulon Progo	−0,946	0,163	
Kota Yogyakarta	1,704	0,153	
Sleman	0,571	0,131	



PELUANG



Definisi:

PELUANG: Harga angka yang menunjukkan seberapa besar kemungkinan suatu peristiwa akan terjadi.





Deskripsi Data

- Ruang Sampel (S)
- Titik Sampel
- Kejadian/Peristiwa \rightarrow subset dari ruang sampel



Peluang Suatu Peristiwa

Definisi klasik, dengan menganggap tiap-tiap elemen ruang sampel S mempunyai peluang yang sama untuk terjadi.

Peluang terjadinya peristiwa A ,

$$P(A) = \frac{n(A)}{n(S)}$$

dengan $n(A)$ = banyaknya anggota dalam peristiwa A , dan $n(S)$ = banyaknya anggota ruang sampel



Contoh 1.

Jika 5 dari 10 truk pengiriman dari suatu perusahaan tidak memenuhi standar emisi dan 3 diantaranya akan dipilih secara acak untuk inspeksi. Berapa probabilitas bahwa semua truk yang terpilih tidak memenuhi standar emisi?

Contoh 2.

Berikut ini Gaji dari karyawan perbulan suatu perusahaan (juta rupiah) :

Pria	2	2,5	2	1,7	3	2,7	1,5	2	3,2	2,1	3	2,8	2,5
wanita	3	2,1	1,3	2	1,6	2,3	3	1,3	1	1,7	2,5	1,5	2

Dengan analisis data eksploratif, bandingkan gaji karyawan Pria dan wanita di suatu perusahaan!



Contoh 3.

Suatu kelas terdiri dari 8 laki-laki dan 12 perempuan. Dipilih 4 siswa untuk menjadi pengurus kelas. Berapakah peluang bahwa 4 orang terpilih terdiri dari 2 laki-laki dan 2 perempuan.

Contoh 4.

Suatu kelas terdiri dari 8 laki-laki dan 12 perempuan. Dipilih 4 siswa untuk menjadi ketua, wakil ketua, sekretaris dan bendahara. Berapakah peluang bahwa 4 orang terpilih terdiri dari 2 laki-laki dan 2 perempuan.



UNIVERSITAS
GADJAH MADA

Terima kasih

LOCALLY ROOTED, GLOBALLY RESPECTED

UGM.AC.ID