STATISTIKA - 1

MFM DATA SCIENCE INDONESIA

"DEMOCRATIZING STATISTICS ACROSS INDONESIA"

SUBSCRIBE

STATISTICS-101 ID

OUTLINE

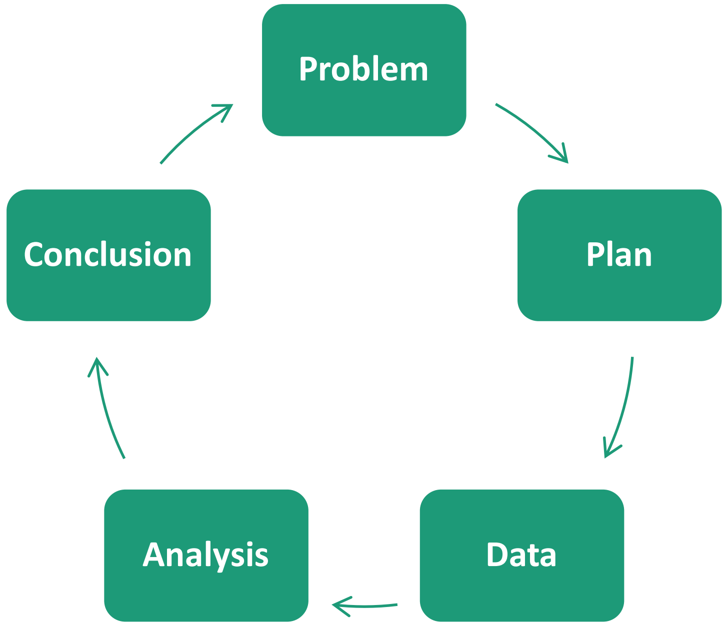
- Apa itu Statistika?
- Populasi vs Sampel
- Tipe Data dan Skala Pengukuran Data
- Statistika Deskriptif
- Peluang
- Mengukur Peluang
- Kejadian Independen dan Bersyarat
- Teorema Bayes
- Distribusi Peluang
- Distribusi Peluang Diksrit dan Kontinu
- Ekspektasi dan Varians
- Berbagai Macam Fungsi Distribusi Peluang

Apa itu statistika?

"Ilmu yang mempelajari data, mengukur, mengontrol, dan mengomunikasikan ketidakpastian"

American Statistical Association (ASA)

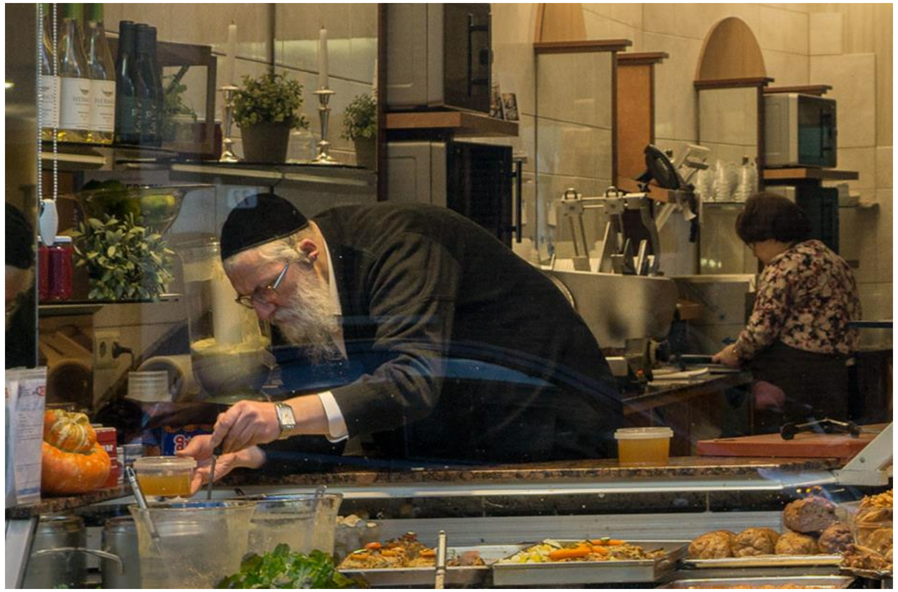
Apa itu statistika?

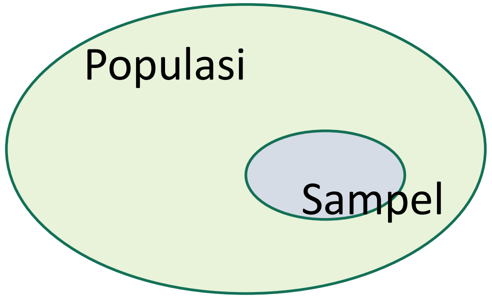


Statistical Cycle

- Problem : Mendefinisikan permasalahan
- Plan : Sistem pengukuran, desain sampling, manajemen data
- Data : Pengumpulan, manajemen, cleaning
- Analysis : Eksplorasi data, generalisasi hipotesis
- Conclusion : Interpretasi, kesimpulan, ide baru

Populasi vs Sampel





Karakteristik

Populasi
“Parameter”

Sampel
“Statistik”

Tipe Data dan Skala Pengukuran Data

Berdasarkan Sifat

- Diskrit

  Contoh : Jumlah Kendaraan, Jumlah Siswa

- Kontinu

  Contoh : Tingkat Kelembaban, Suhu

Berdasarkan Jenis

- Kualitatif
- Kuantitatif

Skala Pengukuran Data

- Nominal

  Contoh : Jenis Kelamin, Agama

- Ordinal

  Contoh : Tingkat Pendidikan, Tingkat Pendapatan

- Interval

  Contoh : Suhu

- Rasio

  Contoh : Tinggi Badan, Berat Badan, Panjang Jalan

Statistika Deskriptif

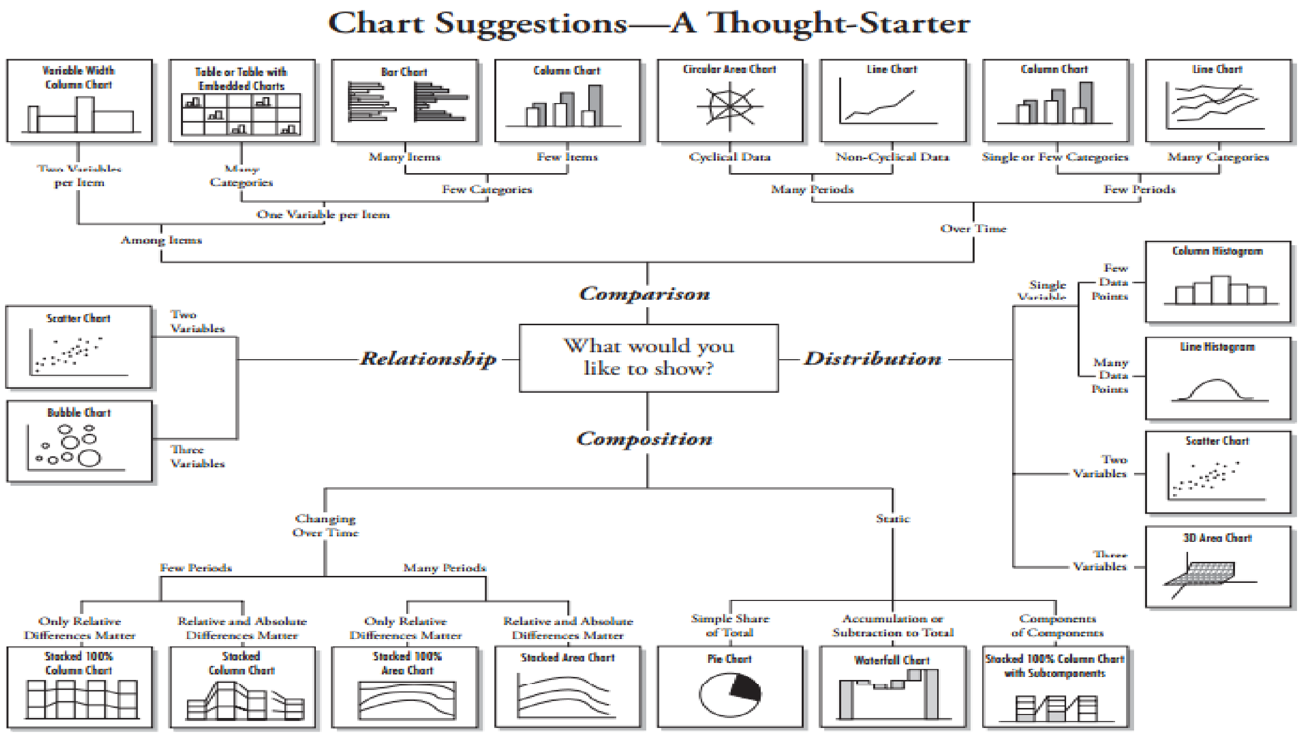
"Merangkum karakteristik sampel"

Ukuran Pemusatan

- Mean
- Median
- Modus

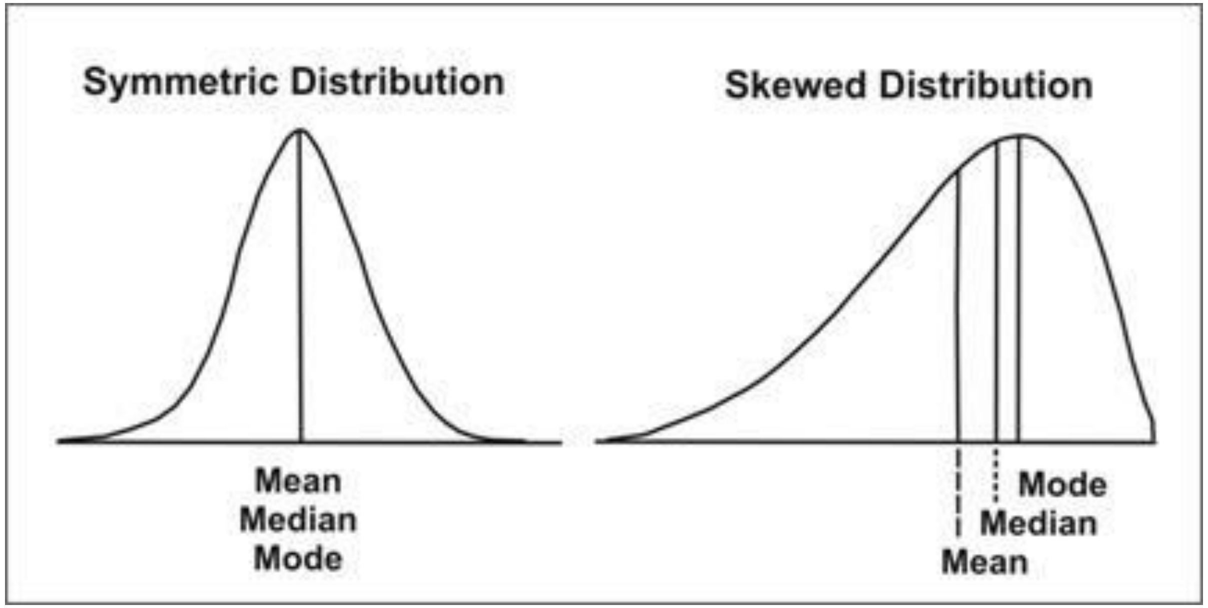
Ukuran Penyebaran

- Varians
- Standar Deviasi
- Range
- Interquartile Range



https://www.analyticsvidhya.com/learning-paths-data-science-business-analytics-business-intelligence-big-data/tableau-learning-path/

Statistika Deskriptif



https://www.cdc.gov/csels/dsepd/ss1978/Lesson2/Section8.html#ALT210

Misalkan vektor data x

12, 14, 14, 15, 17, 18, 18, 19, 22

Q1 Median Q3

$IQR = Q3 - Q1 = 4.5$

$$\text{Mean} = \bar{x} = \frac{1}{\text{n}}\sum_{i=1}^{n} x_i = 16.56$$

$$\text{Varians} = \sigma^2 = \frac{1}{\text{n}-1}\sum_{i=1}^{n}(x_i - \bar{x})^2 = 9.52$$

STATISTICS
101 ID

Apa itu Peluang?

"Ukuran yang mengukur ketidakpastian"

"Mengungkapkan kemungkinan kejadian yang tidak dapat diprediksi dengan pasti. Bahkan kejadian yang tidak biasa kadang terjadi"

"Cara manusia untuk mengetahui matematika Tuhan"

STATISTICS
101 ID

Apa itu Peluang?

Ruang Sampel

"Kumpulan semua hasil percobaan yang mungkin berbeda"

Kejadian

"Sebagian hasil percobaan yang terjadi dari ruang sampel"

2 kali toss

Ruang Sampel

$$\{HH, HT, TH, TT\}$$

Kejadian

$$\{TH, TT\}$$

STATISTICS
101 ID

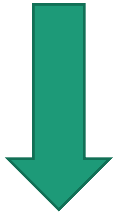
Apa itu Peluang?

Peluang adalah **fungsi real** bernilai- P yang menetapkan untuk setiap kejadian A dalam ruang sampel S, nilai $P\,(A)$, yang disebut peluang kejadian A, sedemikian sehingga memenuhi sifat-sifat berikut :

- $0 \leq P(A) \leq 1$
- $P(S) = 1$
- jika $A_1, A_2, A_3, \ldots$ adalah kejadian dan $A_i \cap A_j = \emptyset, i = j$, maka $P(A_1 \cup A_2 \cup \cdots \cup A_k) = P(A_1) + P(A_2) + \cdots + P(A_k)$

Mengukur Peluang

George Mendel



Merah	**Pink**	**Pink**	**Putih**

$P(Merah) = 0.25$ $P(Pink) = 0.50$ $P(Putih) = 0.25$

Kasus Uniform

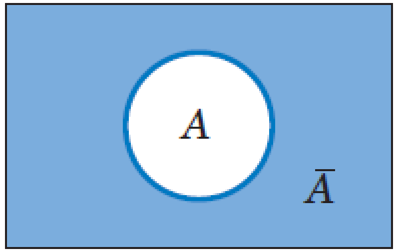
$$P(A) = \frac{m}{k}$$

Dimana :
m = banyak anggota himpunan kejadian A
k = banyak anggota dalam himpunan ruang sampel S

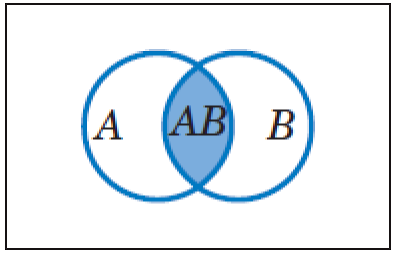
Mengukur Peluang

Sifat-sifat lain

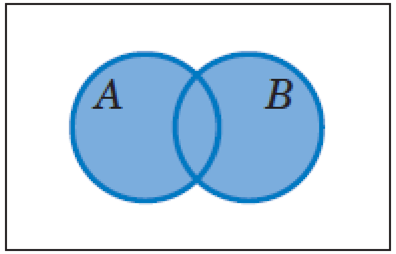
Law of complement : $P(\bar{A}) = 1 - P\,(A)$
Addition law : $P\,(A \cup B\,) = P(A) + P\,(B) - P\,(AB)$
Multiplication law : $P\,(AB) = (B)P\,(A|B)$



Complement $\bar{A}$



Intersection AB



Union AUB

Peluang Bersyarat dan Kejadian Independen

"Jika orang tua memiliki Riwayat diabetes maka Sang anak memiliki peluang terkena diabetes xx%"

Peluang Bersyarat A dengan syarat B:

$$P\ (A|B) = \frac{P(AB)}{P(B)}$$

Kedua Kejadian A dan B dikatakan Independen jika

$P\ (A|B) = P(A)$ atau $P\ (B|A) = P(B)$

Sehingga

$$P(AB) = P(A)P(B)$$

STATISTICS 101 ID

Peluang Bersyarat dan Kejadian Independen

- Peluang bahwa penerbangan yang dijadwalkan secara teratur berangkat tepat waktu adalah $P(D) = 0.83$; peluang bahwa penerbangan tersebut datang tepat waktu adalah $P(A) = 0.82$; dan peluang berangkat dan tiba tepat waktu adalah $P(D \cap A) = 0.78$. Berapakah peluang bahwa sebuah pesawat tiba tepat waktu, mengingat pesawat itu berangkat tepat waktu?

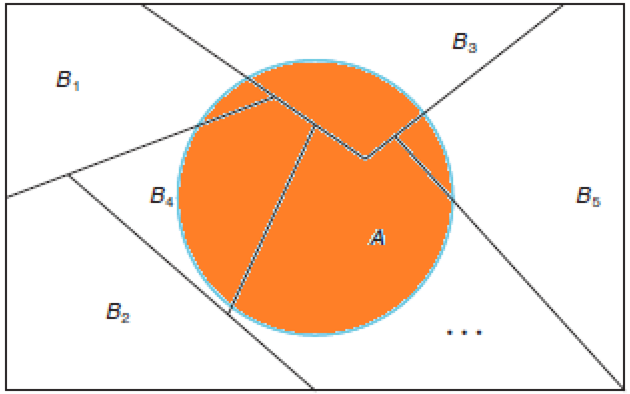
$$P(A|D) = \frac{P(D \cap A)}{P(D)} = \frac{0.78}{0.83} = 0.94$$

- Misalkan sistem mekanik terdiri dari dua komponen yang berfungsi secara independen. Berdasarkan pengujian, komponen 1 memiliki reliabilitas 0.98 dan komponen 2 memiliki reliabilitas 0.95. Jika sistem hanya dapat berfungsi jika kedua komponen berfungsi, berapakah keandalan sistem?

$$P(S) = P(A_1)P(A_2) = 0.98 \times 0.95 = .931$$

Teorema Bayes

Thomas Bayes



Rule of Total Probability

Jika kejadian $B_1, B_2, \ldots, B_k$ merupakan partisi ruang sampel S sedemikian rupa sehingga $P(B_i) \neq 0$ untuk $i = 1, 2, \ldots, k$, lalu untuk kejadian apa pun A dari S,

$$P(A) = \sum_i^k P(B_i \cap A) = \sum_i^k P(B_i)P(A|B_i)$$

Bayes Rule

$$P(B_r|A) = \frac{P(B_r \cap A)}{\sum_i^k P(B_i \cap A)} = \frac{P(B_r)P(A|B_r)}{\sum_i^k P(B_i)P(A|B_i)}$$

Untuk $r = 1,2,..,k$

STATISTICS 101 ID

Teorema Bayes

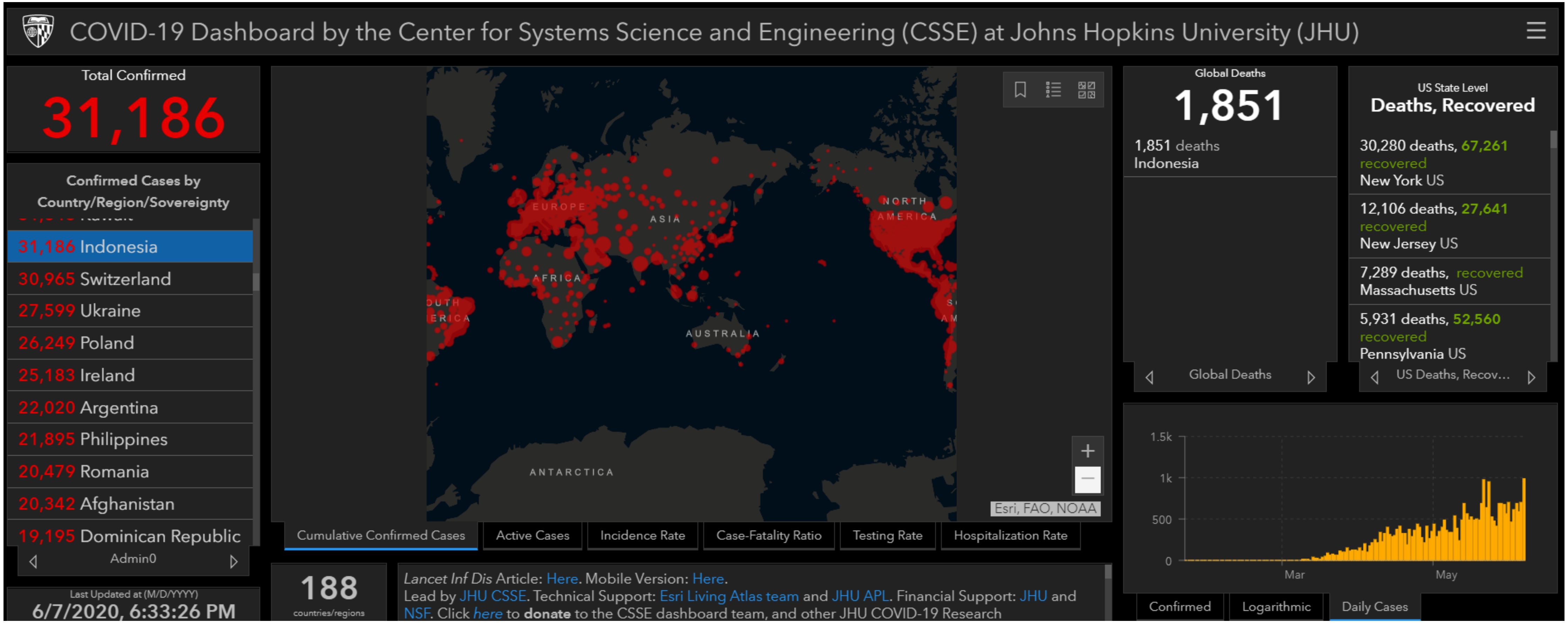
- Keseluruhan produk barang yang siap jual, diproduksi dari 3 mesin. M_1, M_2, M_3 masing-masing memroduksi 30%, 45%, dan 25%. Namun berdasarkan pengalaman masa lalu bahwa 2%, 3%, dan 2% dari produk yang dibuat oleh masing-masing mesin, rusak (cacat). Jika saat ini diambil produk secara acak. Berapa peluang produk tersebut rusak?

$$P(Rusak) = P(M_1)P(Rusak|M_1) + P(M_2)P(Rusak|M_2) + P(M_3)P(Rusak|M_3)$$

$$P(Rusak) = 0.3 \times 0.02 + 0.45 \times 0.03 + 0.25 \times 0.02 = 0.0245$$

- Jika suatu produk dipilih secara acak dan ternyata rusak, berapa peluang bahwa itu dibuat oleh mesin M_3?

$$P(M_3|Rusak) = \frac{P(M_3)P(Rusak|M_3)}{P(M_1)P(Rusak|M_1) + P(M_2)P(Rusak|M_2) + P(M_3)P(Rusak|M_3)} = \frac{0.005}{0.0245} = 0.204$$



STATISTICS
101 ID

Apa itu Distribusi Peluang?

"Persebaran** dari suatu **ketidakpastian"

Apa itu Distribusi Peluang?

Ruang Sampel

"Kumpulan semua hasil percobaan yang mungkin berbeda"

Kejadian

"Sebagian hasil percobaan yang terjadi dari ruang sampel"

Variabel Acak

"Nilai dari hasil percobaan"

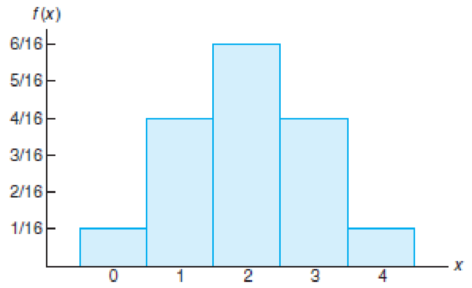
Distribusi Peluang Diksrit dan Kontinu

Distribusi peluang untuk **variabel random diskrit** X dinyatakan sebagai fungsi

$$f(x_i) = P(X = x_i)$$

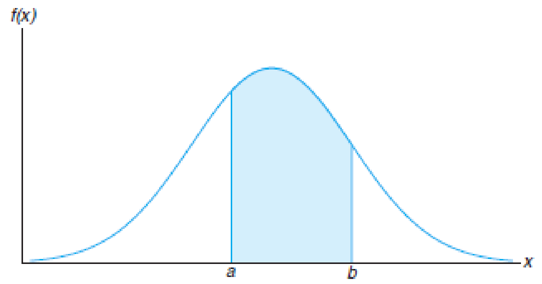
Yang memberikan peluang untuk setiap nilai dan memenuhi:

- $0 \leq f(x_i) \leq 1$ untuk setiap x_i dari vaiabel acak X
- $\sum_{i=1}^{k} f(x_i) = 1$



Distribusi peluang untuk **variabel random kontinu** X dinyatakan dalam fungsi $f(x)$. Fungsi distribusi tersebut memiliki syarat:

- $\int_{-\infty}^{\infty} f(x)\, dx = 1$
- $\int_{a}^{b} f(x)\, dx = P[a \leq X \leq b]$
- $f(x) \geq 0$ for all x.



Ekspektasi dan Varians

Diberikan variabel random X dengan distribusi peluang $f(x)$ maka **Ekspektasi** dari X adalah:

$$E(X) = \mu = \sum x\, f(x)$$

Sedangkan untuk kontinu

$$E(X) = \mu = \int x\, f(x)\, dx$$

Diberikan variabel random X dengan distribusi peluang $f(x)$ dan mean μ maka **Varians** dari X adalah:

$$Var(X) = \sigma^2 = \sum (x-\mu)^2 f(x)$$

Sedangkan untuk kontinu

$$Var(X) = \sigma^2 = \int (x-\mu)^2 f(x)\, dx$$

Beberapa Sifat Sifat

- $E(X^2) = \sum x^2\, f(x)$, diskrit.
- $E(X^2) = \mu = \int x^2 f(x)\, dx$, kontinu.
- $Var(X) = E(X^2) - \left(E(X)\right)^2$

Ekspektasi dan Varians

- Variabel acak X merupakan jumlah mobil yang digunakan untuk keperluan bisnis resmi pada hari kerja tertentu. Bagaimana Ekspektasi dan Variansnya apabila distribusi peluang penggunaan mobil untuk perusahaan A sebagai berikut?

x	1	2	3
$f(x)$	0.3	0.4	0.3

$$\mu_A = 1 \times 0.3 + 2 \times 0.4 + 3 \times 0.3 = 2 \qquad \sigma_A^2 = \sum (x-2)^2 f(x) = 0.6$$

- Permintaan mingguan untuk produk air minum dalam ribuan liter, dari Toko A adalah variabel acak kontinu X. Berapakah ekspektasi dan varians jika yang memiliki fungsi distribusi peluang seperti berikut?

$$f(x) = \begin{cases} 2(x-1), & 1 < x < 2, \\ 0, & \textit{lainnya} \end{cases}$$

$$\mu = E(X) = 2\int_1^2 x(x-1)\ dx = \frac{5}{3} \qquad E(X^2) = 2\int_1^2 x^2(x-1)\ dx = \frac{17}{6}. \qquad \sigma^2 = \frac{17}{6} - \left(\frac{5}{3}\right)^2 = \frac{1}{18}.$$

Berbagai Macam Fungsi Distribusi Peluang

Diskrit
- Binomial
- Poisson
- Binomial Negatif
- Geometrik
- Hypergeometrik
- dll

Kontinu
- Normal
- Eksponensial
- Chi-Square
- Student-t
- Log Normal
- Weibull
- dll

Untuk Distribusi Peluang Diskrit

X	*X Counts*	*p(x)*	*Values of X*	*E(x)*	*V(x)*
Discrete uniform	Outcomes that are equally likely (finite)	$\frac{1}{b-a+1}$	$a \le x \le b$	$\frac{b+a}{2}$	$\frac{(b-a+2)(b-a)}{12}$
Binomial	Number of sucesses in n fixed trials	$\binom{n}{x} p^x (1-p)^{n-x}$	$x = 0, 1, \ldots, n$	np	$np(1-p)$
Poisson	Number of arrivals in a fixed time period	$\frac{e^{-\lambda}\lambda^x}{x!}$	$x = 0, 1, 2, \ldots$	λ	λ
Geometric	Number of trials up through 1st success	$(1-p)^{x-1} p$	$x = 1, 2, 3, \ldots$	$\frac{1}{p}$	$\frac{1-p}{p^2}$
Negative Binomial	Number of trials up through kth success	$\binom{x-1}{k-1}(1-p)^{x-k} p^k$	$x = k, k+1, \ldots$	$\frac{k}{p}$	$\frac{k(1-p)}{p^2}$
Hyper - geometric	Number of marked individuals in sample taken without replacement	$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$	$\max(0, M+n-N) \le x \le \min(M, n)$	$n * \frac{M}{N}$	$\frac{nM(N-M)(N-n)}{N^2(N-1)}$

Sumber :https://www.dummies.com/education/math/probability/probability-for-dummies-cheat-sheet/

Untuk Distribusi Peluang Kontinu

X	X Measures	f(x)	Values of X	E(x)	V(x)
Continuous uniform	Outcomes with equal density (continuous)	$\frac{1}{b-a}$	$a \le x \le b$	$\frac{b+a}{2}$	$\frac{(b-a)^2}{12}$
Exponential	Time between events; time until an event	$\lambda e^{-\lambda x}$	$x \ge 0$	$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$
Normal	Values with a bell-shaped distribution (continuous)	$\frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$	$-\infty < x < \infty$	μ	σ
Standard normal (Z)	Standard scores	$\frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$	$Z = \frac{x-\mu}{\sigma}$	0	1
Binomial approximation	Number of successes in large number of trials	Approx. normal if $np \ge 5$ and $n(1-p) \ge 5$ by CLT	$Z = \frac{x-np}{\sqrt{np(1-p)}}$	np	$np(1-p)$
Poisson approximation	Number of occurrences in a fixed time period (large average)	Approx. normal if $\lambda > 30$	$z = \frac{x-\lambda}{\sqrt{\lambda}}$	λ	λ
$\bar{x}$	Average of $x_1, x_2, \ldots, x_n$	Exactly normal if x is normal. Approx. normal if $n \ge 30$ by CLT	$Z = \frac{\bar{x}-\mu_x}{\sigma_x/\sqrt{n}}$	μ_x	$\frac{\sigma_x^2}{n}$
$\hat{p}$	Proportion or percentage of successes in binomial with $np \ge 5$, $n(1-p) \ge 5$	Approx. normal if $np \ge 5$ and $n(1-p) \ge 5$ by CLT	$Z = \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}}$	p	$\frac{p(1-p)}{n}$

Sumber :https://www.dummies.com/education/math/probability/probability-for-dummies-cheat-sheet/

Thank You and Let’s Practice

STATISTICS 101 ID

“DEMOCRATIZING STATISTICS ACROSS INDONESIA”

Medium https://medium.com/@alfanstatistika

https://www.linkedin.com/in/mohammad-alfan-alfian-riyadi

SUBSCRIBE **STATISTICS-101 ID**