Statistics for Describing, Exploring, and Comparing Data

3rd Lecture

Applied Statistics - PKN STAN - Class 5.37 and 5.38

By: **Erika Siregar**

October 13 & 15, 2020

Review

1. Parameter vs Statistcs?
2. Stratified (homogen) vs Cluster (heterogen) Sampling?
3. Histogram vs Bar Chart? → histogram (distribusi data) , bar chart (visualiasi kategori)
4. What is the relationship between Frequency Table and Histogram?
5. What is the importance of drawing a **histogram**?
6. Apa itu trend? → waktu ke waktu
7. **Revision for lecture 02**

Review for Frequency Table Term

1. Lower class limit
2. Upper class limit
3. Class boundary
4. Class width (Interval)

Class interval x (weight in kg)	Tally	Frequency f
40 - 44	II	2
45 - 49	IIII	4
50 - 54	~~IIII~~	5
55 - 59	~~IIII~~ III	8
60 - 64	~~IIII~~	5
65 - 69	IIII	4
70 - 74	II	2
		30

What is the class width (Interval)?

Pendapatan (ribu rupiah)	Banyaknya
40-47	4
48-55	6
56-63	13
64-71	8
72-79	7
80-87	8
88-95	4
Jumlah	50

Marks	*No. of students*
30–39	*2*
40–49	*9*
50–59	*14*
60–69	*7*
70–79	*8*
	40

Covered Topics

01.Intro to Statistics, population, samples

02. Data Summarizing and Graphing

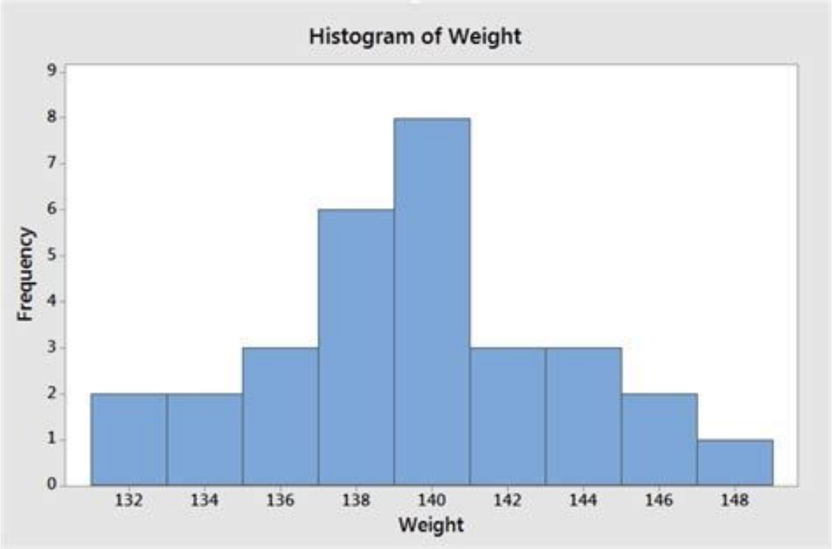
03. Exploring and Comparing Data (Center dan Spread)

Data Distribution → Center Vs Spread

Center → overview

Spread → variasi

Visualizing Distribution



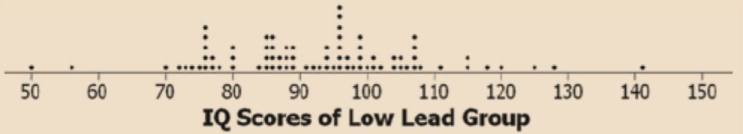


Figure 2-10 Dotplot: IQ Scores of Low Lead Group

Measures of Central Tendency (Mencari nilai tengah dari data)

What is The Measures of Central Tendency?

- Important characteristics that can explain the **general overview** of the data

- The center of the data
    - **Mean** (rata-rata)
        - arithmetic mean
        - Weighted arithmetic mean
    - **Median** (nilai tengah)
    - **Mode** (nilai yang paling sering muncul)

Arithmetic Mean

For survey

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

For population

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

Both are the same only different in symbols

$$\bar{x} = \frac{\Sigma x}{n} = \frac{22 + 22 + 26 + 24 + 23}{5} = \frac{117}{5}$$

= 23.4

Weighted Arithmetic Mean (Rata-rata tertimbang)

- Sometimes, different number could have different "**importance**"
- These **"impotance"** are called the **"weight"**.
- For these cases, it will be **fairer** to consider the weight when calculating the average.
- Example:
  - GPA vs the scale (SKS).
  - Expenditure vs the number of household members.
  - Price vs number of products

$$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i},$$

No	Courses	Score	SKS
1	Statistics	80	3
2	Accountancy	75	3
3	Audit	65	2
4	Indonesian	70	3
5	Management	60	3
6	Economy	70	1

$$\bar{x} = \frac{(80*3+75*3+65*2+70*3+60*3+70*1)}{(3+3+2+3+3+1)}$$

= 70.3

TRY IT WITH R

Another Example

No	Komponen	Score	Bobot
1	UAS	89	0.35
2	UTS	93	0.35
3	Presensi	100	0.03
4	Partisipasi	80	0.06
5	Quiz	96	0.105
6	Tugas	80	0.105

Let’s compute the weighted mean together

Mean of Grouped Data

- = mean from a frequency distribution
- $$\bar{x} = \frac{\Sigma(f \cdot x)}{\Sigma f}$$
-

Table 3-2 IQ Scores of Low Lead Group

IQ Score	Frequency f	Class Midpoint x	$f \cdot x$
50–69	2	59.5	119.0
70–89	33	79.5	2623.5
90–109	35	99.5	3482.5
110–129	7	119.5	836.5
130–149	1	139.5	139.5
Totals:	$\Sigma f = 78$		$\Sigma(f \cdot x) = 7201.0$

$$\bar{x} = \frac{\Sigma(f \cdot x)}{\Sigma f} = \frac{7201.0}{78} = 92.3$$

Median

- **Middle** value of a pre-sorted data
- often denoted by $\tilde{x}$ (pronounced ‘x-tilde’)
- Always **sort** the data before identifying median.
- The median = the **(n +1)/2** -th data
- If the number of data values is **odd**,
  - 1, 5, 7, 12, 20 → middle position = 3 → median = the 3rd data = 7
- If the number of data values is **even**, the median is found by computing **the mean of the two middle numbers**. → 1, 5, **7, 12**, 20, 24
  - Middle = → median = (7 + 12)/2 = 9.5

What is the median of

1. 23, 22, 27, 22, 27, 26 → sorted: 22 22 23 26 27 27 → median = 24
2. 22, 23, 26, 24, 22 → sorted: 22 22 23 24 26 → median = 23

Median of Grouped Data

$$Med = L_o + c\left(\frac{\frac{1}{2}n - F}{f}\right)$$

Med = median
Lo = tepi bawah kelas median
→ tepi bawah ≠ batas bawah
→ tepi bawah =
(batas bawah kelas sekarang + batas bawah kelas sebelumnya)/2
c = panjang kelas interval kelas median
n = banyaknya data pengamatan
F = jumlah frekuensi sebelum kelas median
f = frekuensi kelas median
½ n = Kelas median

IQ Score	Freq	CumFreq
50 - 69	2	2
70 - 89	33	35
90 - 109	35	70
110 - 129	7	77
130 - 149	1	78
Total	**78**	**78**

→ Kelas median

$$Med = 89{,}5 + 20\left(\frac{\frac{1}{2}\cdot 78 - 35}{35}\right)$$
$$= 89{,}5 + 20\left(\frac{39-35}{35}\right)$$
$$= 89{,}5 + 20\left(\frac{4}{35}\right)$$
$$= 91{,}29$$

Mean vs Median

1. Median is preferably used when we have extreme values.
   Score = 5, 6, 6, 6, 102
   Mean = 25
   Median = 6
2. Interpretation:
   a. Mean of income = Rp. 10,000 → each person on average earn 10,000
   b. Median of income = Rp. 10,000 → 50% of population has income < 10,000 while the other 50% has income > 10,000.

Mode (Modus)

- Value with highest frequency.
- A dataset could be
  - Unimodal → 1 mode
  - Bimodal → 2 modes
  - Multimodal → > 2 modes
  - No modal → 0 mode
- Examples:
  - 22 22 26 24 23 → Mode is 22
  - 22 22 22 23 23 23 24 24 26 27 → Bimodal: 22 & 23
  - 22 23 24 26 27 → No Mode

Mode of Grouped Data

$$\text{Mod} = L_0 + c\left(\frac{b_1}{b_1 + b_2}\right)$$

IQ Score	Freq	CumFreq
50 - 69	2	2
70 - 89	33	35
90 - 109	35	70
110 - 129	7	77
130 - 149	1	78
Total	**78**	**78**

Kelas modus

Mod = modus
Lo = tepi bawah kelas modus
c = panjang kelas interval kelas modus
n = banyaknya data pengamatan
b1 = selisih antara frekuensi kelas modus dengan frekuensi kelas sebelum kelas modus
b2 = selisih antara frekuensi kelas modus dengan frekuensi kelas setelah kelas modus
Kelas modus = kelas dengan frekuensi tertinggi

$$\text{Mod} = 89{,}5 + 20\left(\frac{2}{2+28}\right)$$

$$= 89{,}5 + 20\left(\frac{2}{30}\right)$$

$$= 90{,}83$$

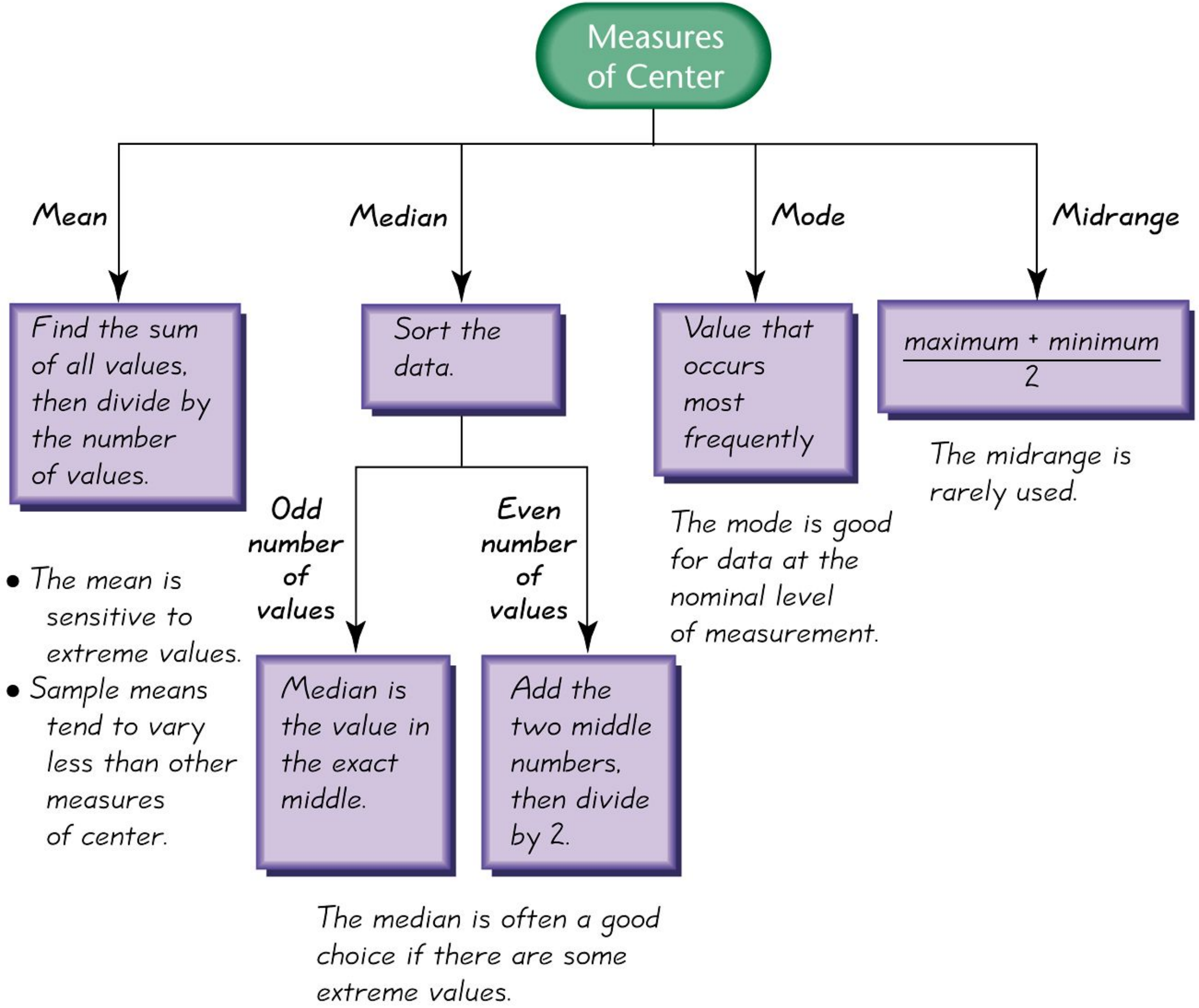
Midrange

$$\text{Midrange} = \frac{\text{maximum value} + \text{minimum value}}{2}$$

example: 22 22 23 24 26

$$\frac{26 + 22}{2}$$

Midrange is 24

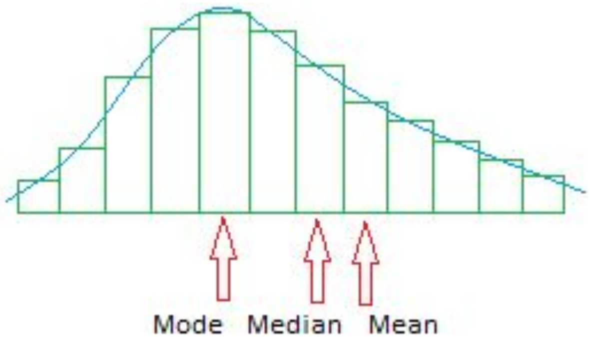

Measures of Center
Mean
Median
Mode
Midrange
Find the sum of all values, then divide by the number of values.
Sort the data.
Value that occurs most frequently
$\frac{\text{maximum} + \text{minimum}}{2}$
The midrange is rarely used.
Odd number of values
Even number of values
The mode is good for data at the nominal level of measurement.
• The mean is sensitive to extreme values.
• Sample means tend to vary less than other measures of center.
Median is the value in the exact middle.
Add the two middle numbers, then divide by 2.
The median is often a good choice if there are some extreme values.

Central Measures Summary (Pros & Cons)

No	Indicators	pros	cons
1	Mean ($\bar{x}$)	Reliable - Takes every data value into account - means of samples drawn from the same population **don't vary as much as other** measures of center	Not resistant - **sensitive** to every data value - one extreme value can affect it dramatically
2	Median ($\tilde{x}$)	Resistant - not affected by an extreme value	- Must sort the data in ascending order.
3	Mode	can be used with nominal data	
4	Midrange	Easy to compute	Sensitive to extremes (coz we only consider max & min)

Why Bother Learning Central Measure?

- Mean
- Median
- modus



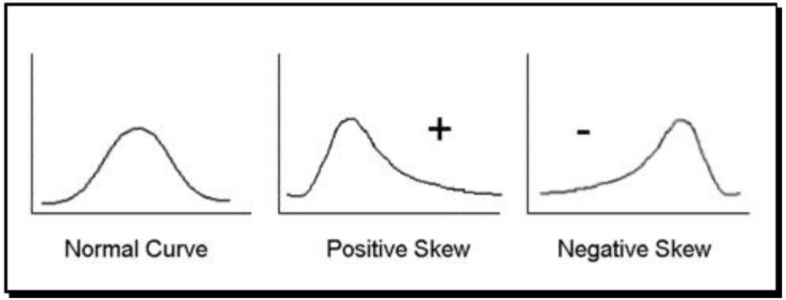
- **To describe data and get the general overview of the data**
- **Understand how the data distributes → symmetrical or skewed?**
- **Compare between populations**

What is Skewness?

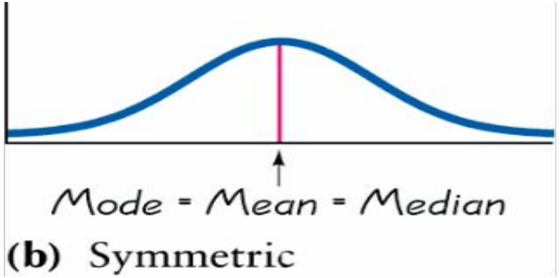
- KEMELENCENGAN
  - a real-life data will not be always perfect
  - it will be skewed towards some side

Skewness is a measure of the asymmetry of a distribution.

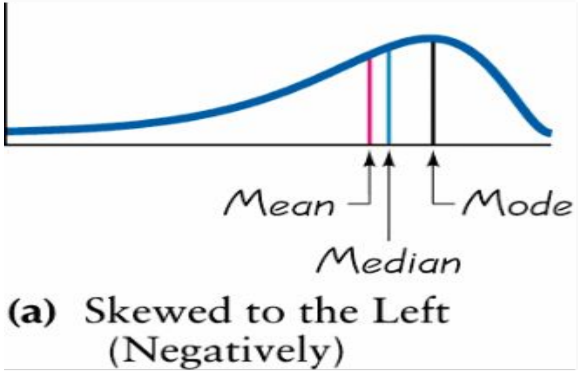
- It basically explains us **to what extent there is tilt in data towards one side**, and how it differs from a perfect Normal distribution.



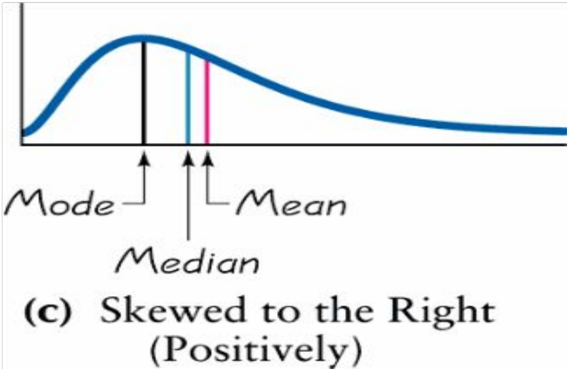
Skewness



Skewness dilihat dari posisi
mean - median -mode



have a **longer left tail**, mean and median are to the left of the mode
→ **mean, median, mode**



have a **longer right tail**, mean and median are to the right of the mode
→ **mode, median, mean**

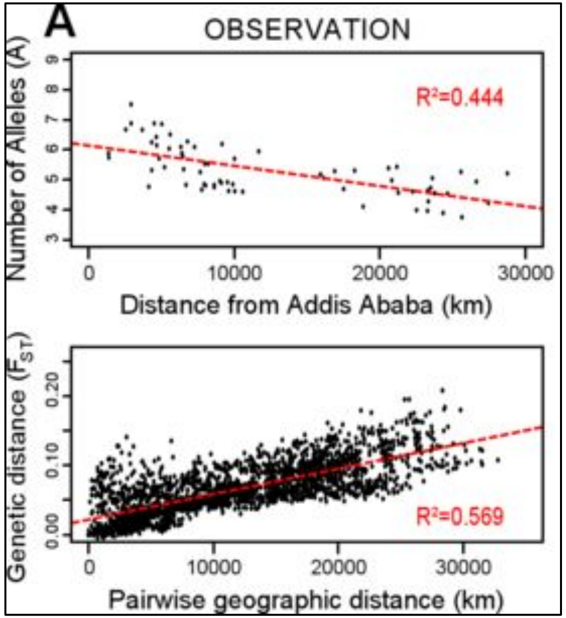
Why Bother with Skewness?

- Skewness can tell us about:
  - the presence of **outliers**
  - **“Penumpukan”** pada value tertentu → data is not equally distributed.
- Skewness can **affect statistics model** → regression, etc.
  - Many model building techniques have the **assumption** that predictor values are **distributed normally** and have a symmetrical shape. Hence, it is sometimes paramount to deal with skewness.
- Type of Skewness
  - Mean = median = modus → normal = simetris
  - Mean < median < modus
  - Mean > median > modus

Variation

Let's Talk about Variation

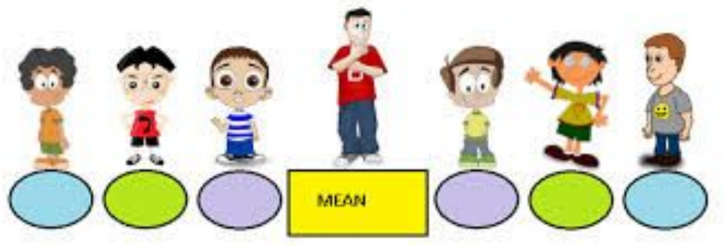
1. It's about "how homogen/heterogen" the data is.
2. The measures:
   a. **Range** → max - min
      i. **very sensitive to extreme values**; therefore not as useful as other measures of variation
      ii. 50, 40, 30, 60, 70, 1 → range??
   b. **Standard Deviation** → a measure of variation of all values from the mean
      i. σ (population)
      ii. s (sample)
   c. **Variance** = the square of standard deviation
      i. σ^2
      ii. s^2



Range

- Range = max - min
- The higher range = the more heterogenous
- very sensitive to **extreme values**; therefore not as useful as other measures of variation
- Example:
    - A: 5, 5, 5, 5, 5, 5, 100, 5, 5, 5, 5 → range = 95
    - B: 5, 8, 9, 12, 13, 7, 15, 20, 25, 31, 18 → range =26
    - Which one has a higher range?
    - According to the range, which one is more heterogeneous?
    - Does it make sense?
- 50, 40, 30, 60, 70, 1 → range??

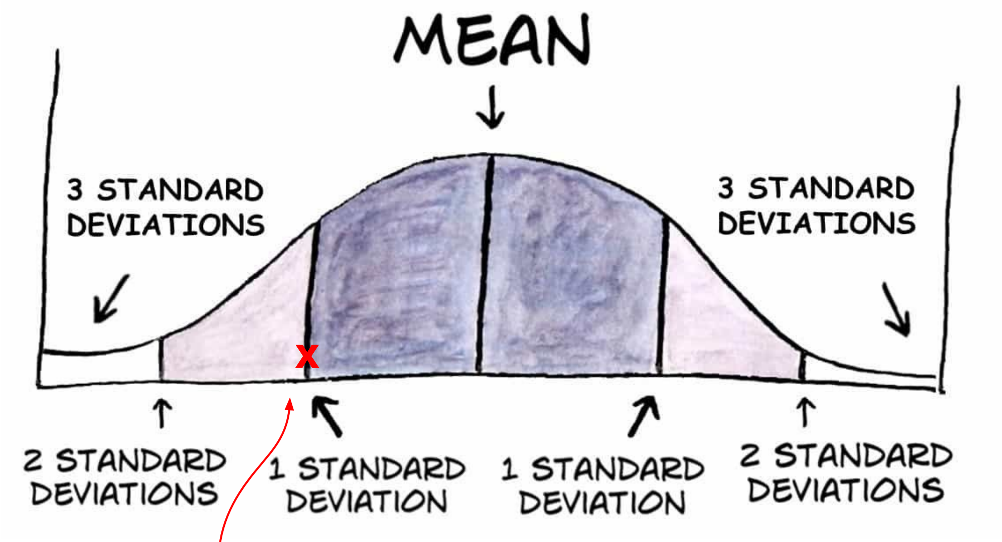
Standard Deviation (Concept)



seberapa jauh sih data point ini dari mean-nya??

how concentrated the data is around the mean

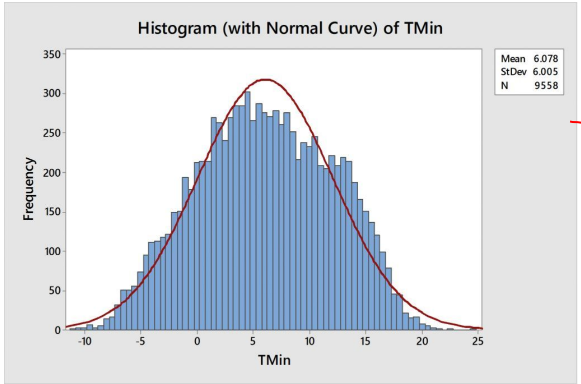
Mencari rata-rata dari total jarak tiap value ke mean-nya

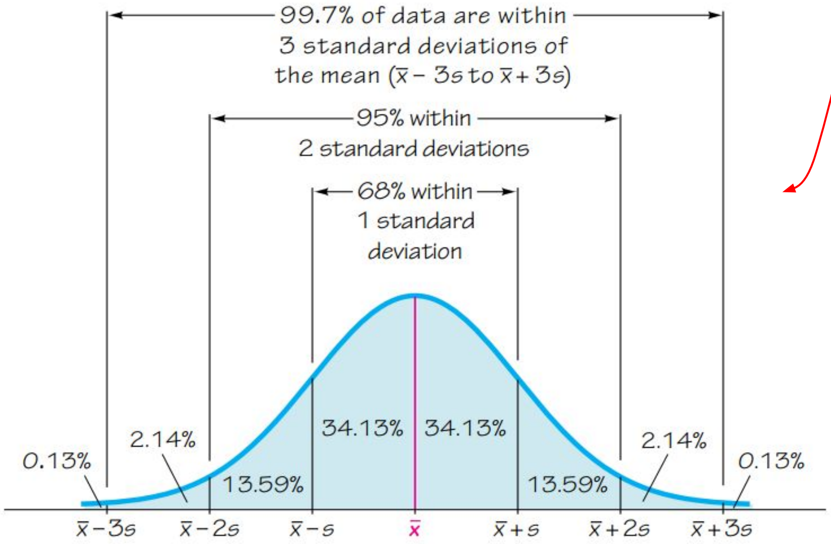


x berbeda sebanyak 1sd dari mean-nya, yang mana masih OK

More about Standard Deviation

- for many data sets, the vast majority (± 95%) of sample values **lie within 2 sd** of the mean.
- Values **close together have a small standard deviation**, but values with much more variation have a larger standard deviation
- a value is considered unusual if it **differs from the mean** by > 2 sd





Standard Deviation (Formula)

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

$$s = \sqrt{\frac{n\Sigma(x_i^2) - (\Sigma x_i)^2}{n(n-1)}}$$

Formula for sample

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

Formula for population

Why divide by (n-1) instead of n?
→ Bessel's Correction
→ Simulation proves that **(n-1)** as denominator provides **a more unbiased estimator** compare to n.
→ $\bar{x}$ < **σ** → n as denominator→ makes the s becomes much more smaller than **σ** → underestimate.
→ because of degree of freedom (d.o.f)

Why Divide by $n - 1$? After finding all of the individual values of $(x - \bar{x})^2$, we combine them by finding their sum. We then divide by $n - 1$ because there are only $n - 1$ independent values. With a given mean, only $n - 1$ values can be freely assigned any number before the last value is determined. Exercise 45 illustrates that division by $n - 1$ yields a better result than division by n. That exercise shows how division by $n - 1$ causes the sample variance s^2 to target the value of the population variance σ^2, whereas division by n causes the sample variance s^2 to underestimate the value of the population variance σ^2.

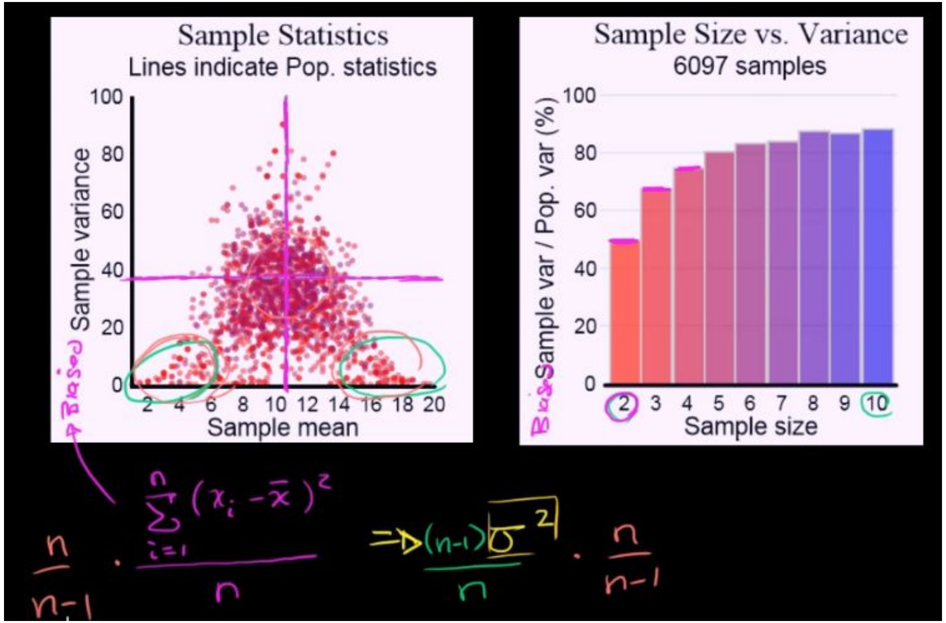
Buku Triola, p. 103

there are
only n - 1 independent values

Degree of freedom = banyaknya independent values yang bisa di-assign.

15, 6, 15, 12, 7 → mean = 11

14, 10, 20, 3, **x** → mean = 11



- Simulation output: (n-1) provides a more unbiased estimator
- Rata2 sampel < rata2 populasi
- Degree of freedom → always lose 1 d.o.f when estimating population from sample

https://www.khanacademy.org/math/ap-statistics/summarizing-quantitative-data-ap/more-standard-deviation/v/simulation-showing-bias-in-sample-variance

Coefficient of Variation

- Comparing Data from 2 Different Groups
- Used when:
  - Example:
    - Sample 1: berat (kg)
    - Sample 2: umur
    - Which sample is more varied?
  - We want to compare 2 populations/samples whose means are significantly varied

Sample

$$CV = \frac{S}{\overline{X}} \cdot 100\%$$

Population

$$CV = \frac{\sigma}{\mu} \cdot 100\%$$

Example:
Sekelompok mahasiswa mengikuti 2 ujian.
- Ujian 1 (statistik): mean = 6, sd = 0.6 (nilai max. 10)
- Ujian 2 (ekonomi): mean = 70, sd = 0.7 (nilai max. 100)

Measures of Relative Standing and Boxplots

Measures of Relative Standing and Boxplots

- numbers showing the location of data values relative to the other values within a data set. → *bagaimana posisi suatu data jika dibandingkan dengan data lainnya*.
- What is used for:
  - **compare values** from different data sets
  - compare values within the same data set.
- How to:
  - Z score
  - Measures of location: Percentile, quartile, decile
  - boxplot

Z-Score

- Basically, **menstandarkan/menormalkan** data dengan membuang efek/pengaruh dari mean dan sd-nya
- Berguna ketika kita ingin membandingkan data dari 2 kelompok yang memiliki **rata-rata dan standar deviasi** yang berbeda **signifikan.**
- Result: new data with **mean = 0** dan **sd = 1**
- Round z scores to **two decimal places** (such as **2.31**)

$$z = \frac{x - \bar{x}}{s}$$ sample

$$z = \frac{x - \mu}{\sigma}$$ population

normal

Example of Using z-score for Comparing Values

14. Oscars As of this writing, Sandra Bullock was the last woman to win an Oscar for Best Actress and Jeff Bridges was the last man to win for Best Actor. At the time of the awards ceremony, Sandra Bullock was 45 years of age and Jeff Bridges was 60 years of age. Based on Data Set 11 in Appendix B, the Best Actresses have a mean age of 35.9 years and a standard deviation of 11.1 years. The Best Actors have a mean age of 44.1 years and a standard deviation of 9.0 years. (All ages are determined at the time of the awards ceremony.) Relative to their genders, who was younger when winning the Oscar: Sandra Bullock or Jeff Bridges? Explain.

SB vs. JB

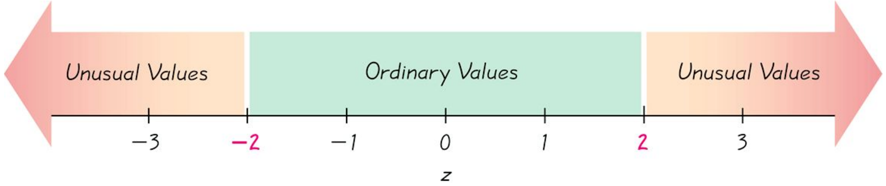
$$z = \frac{45 - 35{,}9}{11{,}1} \qquad \frac{60 - 44.1}{9}$$

$$= 0{,}82 \qquad 1{,}77$$

Andi - akuntansi

budi - pajak

Using z-score to identify unusual values



Ordinary values: –2 ≤ z score ≤ 2
Unusual Values: z score < –2 or z score > 2

Example 2 **Is a Pulse Rate of 48 Unusual?**

As the author was creating this example, he measured his pulse rate to be 48 beats per minute. (The author has too much time on his hands.) Is that pulse rate unusual? (Based on the pulse rates of males from Data Set 1 in Appendix B, assume that a large sample of adult males has pulse rates with a mean of 67.3 beats per minute and a standard deviation of 10.3 beats per minute.)

Solution

The author's pulse rate of 48 beats per minute is converted to a z score as shown below:

Pulse rate of 48:

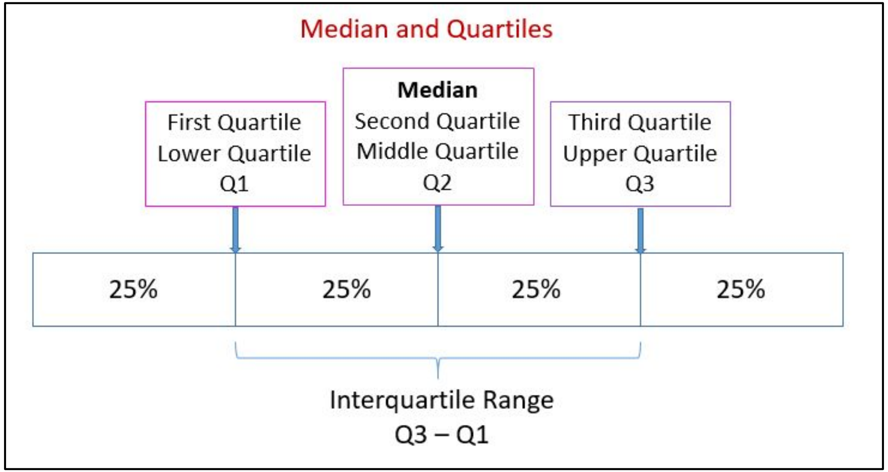
$$z = \frac{x - \bar{x}}{s} = \frac{48 \text{ beats per minute} - 67.3 \text{ beats per minute}}{10.3 \text{ beats per minute}} = -1.87$$

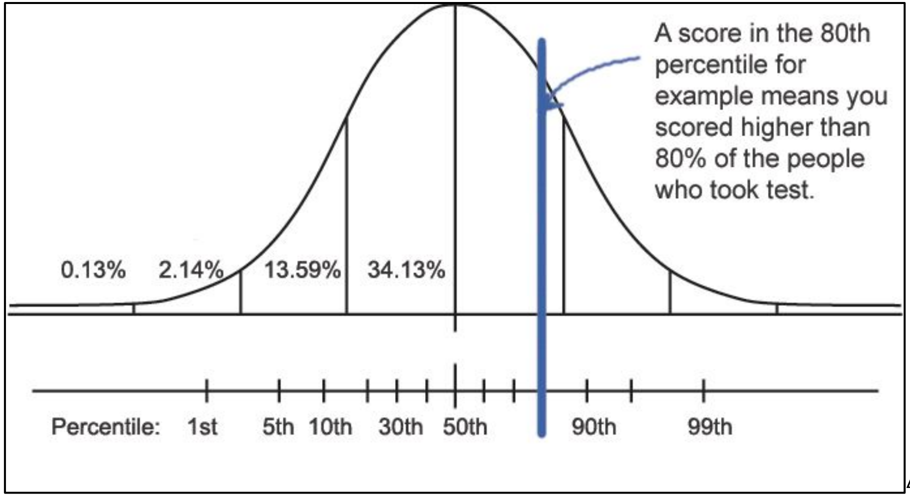
Interpretation

The result shows that the author's pulse rate of 48 beats per minute is converted to the z score of -1.87. Refer to Figure 3-4 to see that $z = -1.87$ is between -2 and $+2$, so it is not unusual.

Measures of Location

- **The 4 sisters:**
  - Median → bagi jadi 2
  - Quartile (Q) → bagi jadi 4 → Q1, Q2, Q3
  - Decile (D) → bagi jadi 10 → D1, D2, ..., D9
  - Percentile (P) → bagi jadi 100 → P1, P2, ... , P99
- In quartile-wise → Median = Q2
- In percentile-wise → median = ...?
- In decile-wise, median = ...?





Computing Percentile

$$\text{Percentile of value } x = \frac{\text{number of values less than } x}{\text{total number of values}} \cdot 100$$

Nilai 23 adalah percentile ke berapa?

Example 3 **Finding a Percentile**

Table 3-5 lists the same counts of chocolate chips in 40 Chips Ahoy regular cookies listed in Table 3-1, but in Table 3-5 those counts are arranged in increasing order. Find the percentile for a cookie with 23 chocolate chips.

Table 3-5 *Sorted* Counts of Chocolate Chips in Chips Ahoy (Regular) Cookies

19	19	20	20	20	20	22	22	22	22
23	23	23	23	23	23	23	24	24	24
24	24	25	25	25	25	25	25	25	26
26	26	26	26	26	27	27	28	28	30

Solution

From the sorted list of chocolate chip counts in Table 3-5, we see that there are 10 cookies with fewer than 23 chocolate chips, so

$$\text{Percentile of } 23 = \frac{10}{40} \cdot 100 = 25$$

Interpretation

A cookie with 23 chocolate chips is in the 25th percentile. This can be interpreted loosely as: A cookie with 23 chocolate chips separates the lowest 25% of cookies from the highest 75%.

Converting from the kth Percentile to the Corresponding Data Value

$$L = \frac{k}{100} \cdot n$$

- n = total number of values in the data set
- k = percentile being used
- L = locator that gives the position of a value
- k = kth percentile

P18, adalah data yang mana?

From Figure 3-5, we see that the sample data are already sorted, so we can proceed to find the value of the locator L. In this computation we use $k = 18$ because we are trying to find the value of the 18th percentile. We use $n = 40$ because there are 40 data values.

$$L = \frac{k}{100} \cdot n = \frac{18}{100} \cdot 40 = 7.2$$

Since $L = 7.2$ is not a whole number, we proceed to the next lower box where we change L by rounding it up from 7.2 to 8. (In this book we typically round off the usual way, but this is one of two cases where we round *up* instead of rounding *off*.) From the last box we see that the value of P_{18} is the 8th value, counting from the lowest. In Table 3-5, the 8th value is 22. That is, $P_{18} = 22$ chocolate chips. Roughly speaking, about 18% of the cookies have fewer than 22 chocolate chips and 82% of the cookies have more than 22 chocolate chips.

Percentile of Grouped Data

$$P_i = L_0 + c\left(\frac{\frac{i}{100}(n) - F}{f}\right)$$

Pi = persentil ke-i (i = 1, 2, …, 99)
L0 = tepi bawah kelas persentil
c = panjang kelas interval kelas persentil
n = banyaknya data pengamatan
F = jumlah frekuensi sebelum kelas persentil
f = frekuensi kelas persentil

Kelas	Frekuensi (f_i)	Frekuensi komulatif (X_i)
93 – 97	2	2
98 – 102	10	12
103 – 107	12	24
108 – 112	10	34
113 – 117	7	41
118 – 122	4	45
123 – 127	3	48
128 – 132	1	49
133 – 137	0	49
138 – 142	1	50

Diketahui besarnya tekanan darah dari 50 mahasiswa suatu universitas yang disajikan dalam bentuk tabel sebagai berikut. Tentukan besarnya P50 dari data di atas.

Kelas	Frekuensi (f_i)	Frekuensi komulatif (X_i)
93 – 97	2	2
98 – 102	10	12
103 – 107	12	24
108 – 112	10	34
113 – 117	7	41
118 – 122	4	45
123 – 127	3	48
128 – 132	1	49
133 – 137	0	49
138 – 142	1	50

$$\boxed{P_i = L_0 + c\left(\frac{\frac{i}{100}(n) - F}{f}\right)}$$

$$Letak\ P_{50} = \frac{50}{100}(n) = \frac{50}{100}(50) = 25$$

$$Kelas\ P_{50} = 108 - 112$$

$$P_i = L_0 + c\left(\frac{\frac{i}{100}(n) - F}{f}\right)$$

$$P_{50} = 107{,}5 + 5\left(\frac{\frac{50}{100}(50) - 24}{10}\right)$$

$$P_{50} = 107{,}5 + 5\left(\frac{1}{10}\right)$$

$$P_{50} = 108$$

Quartile of Grouped Data

$$Q_i = L_0 + c\left(\frac{\frac{i}{4}(n) - F}{f}\right) \longrightarrow \text{Letak } Q_i = \frac{i}{4}(n)$$

Qi	=	kuartil ke-i (i = 1, 2, 3)
L0	=	tepi bawah kelas kuartil
c	=	panjang kelas interval kelas kuartil
n	=	banyaknya data pengamatan
F	=	jumlah frekuensi sebelum kelas kuartil
f	=	frekuensi kelas kuartil

Kelas	Frekuensi (f_i)	Frekuensi komulatif (X_i)
93 – 97	2	2
98 – 102	10	12
103 – 107	12	24
108 – 112	10	34
113 – 117	7	41
118 – 122	4	45
123 – 127	3	48
128 – 132	1	49
133 – 137	0	49
138 – 142	1	50

Diketahui besarnya tekanan darah dari 50 mahasiswa suatu universitas yang disajikan dalam bentuk tabel sebagai berikut. Tentukan besarnya Q1, dari data di atas.

$$Q_i = L_0 + c\left(\frac{\frac{i}{4}(n) - F}{f}\right)$$

$$Q_1 = 102{,}5 + 5\left(\frac{\frac{1}{4}(50) - 12}{12}\right)$$

$$Q_1 = 102{,}5 + 5\left(\frac{0{,}5}{12}\right)$$

$$Q_1 = 102{,}71$$

Decile of Grouped Data

$$D_i = L_0 + c\left(\frac{\frac{i}{10}(n) - F}{f}\right)$$

$$\text{Letak } D_i = \frac{i}{10}(n)$$

Di = desil ke-i (i = 1, 2, ..., 9)
L0 = tepi bawah kelas desil
c = panjang kelas interval kelas desil
n = banyaknya data pengamatan
F = jumlah frekuensi sebelum kelas desil
f = frekuensi kelas desil

Kelas	Frekuensi (f_i)	Frekuensi komulatif (X_i)
93 – 97	2	2
98 – 102	10	12
103 – 107	12	24
108 – 112	10	34
113 – 117	7	41
118 – 122	4	45
123 – 127	3	48
128 – 132	1	49
133 – 137	0	49
138 – 142	1	50

Diketahui besarnya tekanan darah dari 50 mahasiswa suatu universitas yang disajikan dalam bentuk tabel sebagai berikut. Tentukan besarnya D2

$$D_i = L_0 + c\left(\frac{\frac{i}{10}(n) - F}{f}\right)$$

$$D_2 = 97,5 + 5\left(\frac{\frac{2}{10}(50) - 2}{10}\right)$$

$$D_2 = 97,5 + 5\left(\frac{8}{10}\right)$$

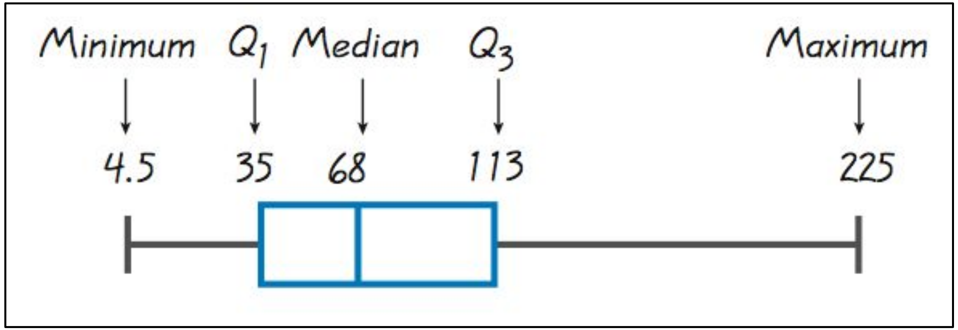
$$D_2 = 101,5$$

Some Other Statistics

- **Interquartile Range (or IQR):** $Q_3 - Q_1$
- **Semi-interquartile Range:** $\dfrac{Q_3 - Q_1}{2}$
- **Midquartile:** $\dfrac{Q_3 + Q_1}{2}$
- **10 - 90 Percentile Range:** $P_{90} - P_{10}$

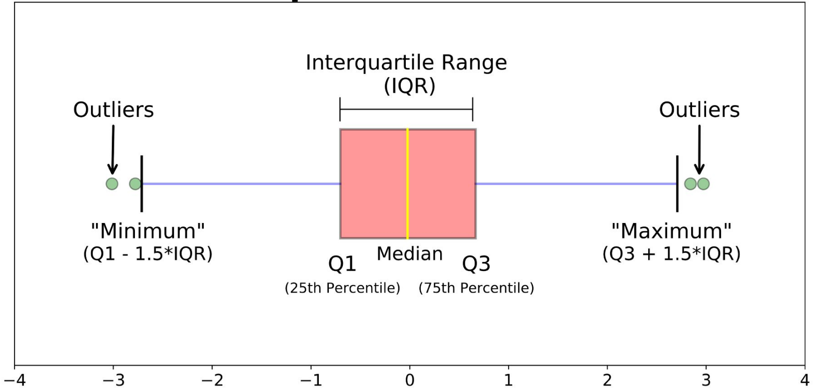
Boxplot

In R:
summary()



5-Number Summary:
1. min
2. Q1
3. Q2 (Median)
4. Q3
5. Max

Modified boxplot



Regular boxplot + **outliers**

Try to create it with R

Thank You