



MySkill | *#RintisKarirImpian*

Portfolio - Intensive Bootcamp

Let's Learn Basic Statistics

Owner: Cornelia Riasdita Valentina

Course Summary



Mean	The sum of all data divided by number of data
Median	The most middle value of arranged data in ascending order. If there are even number of data, the median is the mean of the most middle two values.
Mode	A value which has the highest number of data.
Variance	A mean of squared difference between each value and the mean. Used to measure how far each value of data from the mean (a measure of data dispersion).
Standard Deviation	Square root of variance. Used to determine limit value of outliers.
Percentile	A measure to compare a value with other values from the same data.
Outlier	Values which are very different from the rest of the data, usually values above upper quartile and below lower quartile.
Linear Regression	A model to predict values based on known data, using a straight line. A gradient and Y-interception must be estimated to find the position of the line.



Key Differences Between Standard Deviation and Variance

	Standard Deviation	Variance
What Is it?	The square root of the variance	The average of the squared differences from the mean
What Does it Indicate?	The spread between numbers in a data set	The average degree to which each point differs from the mean
How Is it Expressed?	The same as the units in the data set	In squared units or as a percentage
What Does it Mean?	A low standard deviation (spread) means low volatility while a high standard deviation (spread) means higher volatility	The degree to which returns vary or change over time

Source:
<https://www.investopedia.com/ask/answers/021215/what-difference-between-standard-deviation-and-variance.asp>

Mean

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

Variance

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

Standard Deviation

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

Percentile

$$a = \left(\frac{q}{100} \right) (n + 1)$$

Nomenclature

x_i : value of i^{th} data

\bar{x} : mean atau rata-rata

s^2 : variance

s : standard deviation

a : index (ranking from sorted data)

q : q^{th} percentile (between 0 to 100)

n : number of data

15 MINS: EXERCISE TIME!



Mini Task for Introduction to Statistics:

<https://docs.google.com/spreadsheets/d/1vPinnFq1zZhMcEDHE5BOGkotYb39oAZs/edit?usp=sharing&oid=104232444782941200439&rtpof=true&sd=true>



DATA 1

Date	Voucher Redeem
2023-01-01	98
2023-01-02	112
2023-01-03	1
2023-01-04	41
2023-01-05	106
2023-01-06	111
2023-01-07	86
2023-01-08	142
2023-01-09	143
2023-01-10	88
2023-01-11	32
2023-01-12	66
2023-01-13	92
2023-01-14	94
2023-01-15	118
2023-01-16	53
2023-01-17	119
2023-01-18	12
2023-01-19	37
2023-01-20	90
2023-01-21	128
2023-01-22	79
2023-01-23	86
2023-01-24	45
2023-01-25	33
2023-01-26	54
2023-01-27	115
2023-01-28	129
2023-01-29	56
2023-01-30	83

Case:

You are a part of CRM Team and asked to evaluate the redeemed voucher of the day.

1. What are the **mean**, **median**, **mode**, **standard deviation**, and **outlier** threshold?
2. Is there any outlier from the last month performance?
3. If yes? How many?

Answer			
Mean	81.6	Mean	=AVERAGE(B2:B31)
Median	87	Median	=MEDIAN(B2:B31)
Mode	86	Mode	=MODE(B2:B31)
Var	1460.309195	Var	=VAR(B2:B31)
Std Dev S	38.21399214	Std Dev S	=STDEV(B2:B31)
Percentile 25	53.25	Percentile 25	=PERCENTILE(B2:B31,0.25)
Percentile 75	111.75	Percentile 75	=PERCENTILE(B2:B31,0.75)
IQR	58.5	IQR	=E21-E20
OUTLIER-LOWER LIMIT	-34.5	OUTLIER-LOWER LIMIT	=E20-(1.5*E22)
OUTLIER-UPPER LIMIT	199.5	OUTLIER-UPPER LIMIT	=E21+(1.5*E22)
COUNTIF	0	COUNTIF	=COUNTIF(B2:B31,">199.5")
		No outlier	

*IQR = Interquartal Range

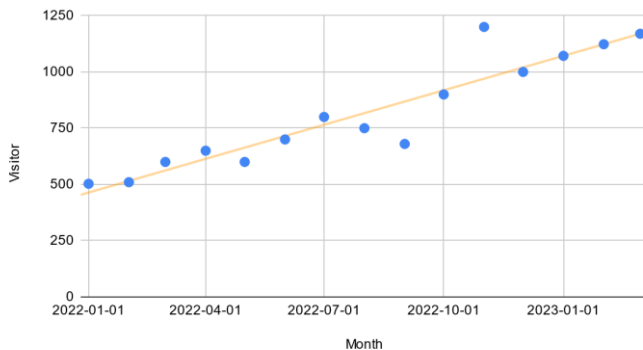
DATA 2

Case:

Kamu adalah seorang manager museum, kamu meminta untuk memperkirakan dan memvisualisasikan perkiraan pengunjung yang potensial pada Q1 2022.

Month	Visitor
2022-01-01	503
2022-02-01	510
2022-03-01	600
2022-04-01	650
2022-05-01	600
2022-06-01	700
2022-07-01	800
2022-08-01	750
2022-09-01	680
2022-10-01	900
2022-11-01	1200
2022-12-01	1000

Visitor vs Month



2023-01-01 =FORECAST.LINEAR(A14, \$B\$2:\$B13, \$A\$2:\$A13)

2023-02-01 =FORECAST.LINEAR(A15, \$B\$2:\$B14, \$A\$2:\$A14)

2023-03-01 =FORECAST.LINEAR(A16, \$B\$2:\$B15, \$A\$2:\$A15)

Hasil Forecast:

Month	Visitor
2022-01-01	503
2022-02-01	510
2022-03-01	600
2022-04-01	650
2022-05-01	600
2022-06-01	700
2022-07-01	800
2022-08-01	750
2022-09-01	680
2022-10-01	900
2022-11-01	1200
2022-12-01	1000
2023-01-01	1071
2023-02-01	1123
2023-03-01	1170

Follow me!

LinkedIn : Cornelia Riasdita

Bootcamp Data Analysis
by @myskill.id

