Decision-making Based on Historical Data

Student Name

2024-01-28

## Student Name

## Department

## Course Name

## Lecturer Name

## January 28, 2024

# Variance and Skewness

## Variance

Variance is a statistical measure of how far apart the numbers are in a set of data. It calculates each number’s deviation from the mean or average, and consequently from all other numbers in the data set denoted by symbol σ2.

### Example how to calculate variance

σ2=∑(Xi−Xˉ)2/ N #### where:

• Xi is each data point  
• Xˉ is the mean of all values in the data set  
• And N is the number of data points.

#### Example Calculation:

-Let’s set of data points be : 2, 4, 4, 4, 5, 5, 7, 9. -Mean(Xˉ)=(2+4+4+4+5+5+7+9)/8=5 -Variance=((2−5)2+(4−5)2+(4−5)2+(4−5)2+(5−5)2+(5−5)2+(7−5)2+(9−5)2)/8 -Variance=(9+1+1+1+4+16)/8=32/8=4

## Skewness

Skewness is a measurement of the asymmetry of the probability distribution of a real-valued random variable in a set of data. It is demonstrated is shown on a bell curve when the distribution of data points to the left and right of the median is not symmetrical.

#### Example how to calculate

Skewness=∑(Xi−Xˉ)3/N/(∑(Xi−Xˉ)2/N/(√ (∑(Xi−Xˉ)2/N))3

#### Example

lea data points be: 2, 3, 3, 4, 4, 5. Skewness=(2−3)3+(3−3)3+(3−3)3+(4−3)3+(4−3)3+(5−3)3/ (6×((2−3)2+(3−3)2+(3−3)2+(4−3)2+(4−3)2+(5−3)2 )/ 6)3/2 Skewness=−1+0+0+1+1+8/6(1+0+0+1+1+4)/6) Skewness=6×(67)3/29 Skewness≈6×2.389≈1.19

# Explaining Columns in dataG2.csv

data<-read.csv("/home/addis/Desktop/Projects/R/Data Analysis/dataG2.csv")  
head(data)

## X.0.01954083 X0.51413006 X0.50578123 X0.25652292 X0.045819789  
## 1 -0.5258383 0.64961147 0.5588736 0.2607941 0.63470894  
## 2 1.1221135 0.24065644 0.3108606 0.2065487 0.09730016  
## 3 0.8902922 0.20710699 0.3079907 0.2375051 0.80563653  
## 4 1.3825344 0.14337941 0.2525307 0.2449315 0.35563171  
## 5 1.9131062 0.07756285 0.1991516 0.2746975 0.08502010  
## 6 0.7300894 0.36294243 0.4147241 0.2134256 0.81564647

## 1. Skewness Column:

This represents numeric values indicating the skewness of the distribution of stock prices for each corresponding row. Skewness measures the asymmetry of the distribution. Positive values indicate a right-skewed distribution, while negative values indicate a left-skewed distribution.

## 2. Median Column:

Median column represents numeric values representing the median of the distribution of stock prices for each corresponding row. This is the middle value when the data is sorted in ascending order and it provides a measure of central tendency.

## 3. Mean Column:

This represent the mean (average) of the distribution of stock prices for each corresponding row. It is calculated by summing up all values and dividing by the number of values.

## 4. Standard Deviation Column:

The standard deviation shows the numeric values representing the standard deviation of the distribution of stock prices for each corresponding row. It measures the amount of variation in a set of values.

## 5. Price Column:

The price shows the numeric values representing the last observed stock price for each corresponding row. The last price is the most recent value in the dataset.

# Conclusions Based on 1. and 2.

## • Skewness in dataG2.csv:

Skewness describes the skewness of the stock prices. Positive skewness indicate a right-skewed distribution, suggesting the possibility of higher values. While, negative skewness indicates a left-skewed distribution, suggesting the possibility of lower values.

## • Comparison of Median and Mean:

Median represents the median of the stock prices while Mean indicates the mean of the stock prices. When median > mean, the distribution is left-skewed. while if mean > median, the distribution is right-skewed.

## • Standard Deviation:

Standard Deviation measures the dispersion of stock prices.Where higher standard deviation suggests greater variability in stock prices while, lower standard deviation suggests less variability.

## • Price:

The last observed stock price which provides a reference point for analyzing trends.

# Code Execution

imported\_data <- read.csv("/home/addis/Desktop/Projects/R/Data Analysis/dataG2.csv")  
S <- imported\_data[,5] - imported\_data[,3]  
I\_1<- which.min(S) # Buy when the difference is minimum  
I\_2 <- which.max(S) # Sell when the difference is maximum  
I\_1

## [1] 2049

I\_2

## [1] 1816

## The meaning of variables’ I\_1’ and ‘I\_2’ after code execution

### • I\_1 :

‘I\_1’ is the index (row number) where the difference is minimized. It corresponds to the row where the last price is relatively close to the mean price.

### • I\_2 :

‘I\_2’ is the index (row number) where the difference is maximized. This corresponds to the row where the last price deviates the most from the mean price.

# Decision: Buy and Sell Strategy:

## Buy Decision

Buy at the row with the index ‘I\_1’ (2049) where the last price is relatively close to the mean. This decision is based on the assumption that historical patterns suggest a potential for the last price to increase.

## Sell Decision (I\_2):

Sell at the row with the index ‘I\_2’ (1816) where the last price deviates the most from the mean. This decision is based on the assumption that historical patterns suggest a potential for the last price to decrease.

# Using Skewness to Decide about Buying or Selling:

## I\_1=2049:

The row with index 2049 has a positive skewness. This suggests a right-skewed distribution, indicating a potential for higher future values. Buying decision might be supported at this point by positive skewness.

## I\_2=1816:

• The row with index 1816 has a positive skewness. This also suggests a right-skewed distribution. Referencing the figure in file 1\_2.jpeg, index 1816 indicates a potential for the last price to decrease. Selling decision is supported by positive skewness.

# Decision Considerations:

To base our decision we consider the skewness of values where, positive skewness suggests a right-skewed distribution, indicating a potential for higher future values. And the differences between the last prices and mean which provide specific numerical information about how much each last price deviates from the mean.

## Decision:

### 1. I\_1 (I=2049):

Positive skewness suggests a potential for higher future values. The last price being near 0, while the median and mean prices are near 0.6, indicates a potential for the last price to increase. Buying decision would be supported by positive skewness and the relative position of the last price to the median and mean. Based on this information, I would buy rows with positive skewness.

### 2. I\_2 (I=1816):

Positive skewness suggests a potential for higher future values. From the image I\_2.jpeg, the last price being near 1, while the median and mean prices are near 0.1, indicates a potential for the last price to decrease.

## Consideration:

Selling decision is supported by positive skewness and the relative position of the last price to the median and mean.