DMPA

Date: / /

(6)
$$SD = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n}}$$

$$Cov(x,y) = \frac{\Sigma(x-\bar{x})(y-\bar{y})}{n}$$

$$r = \frac{Cov(x,y)}{\delta x \, \delta y}$$ (Co-relation coefficient)

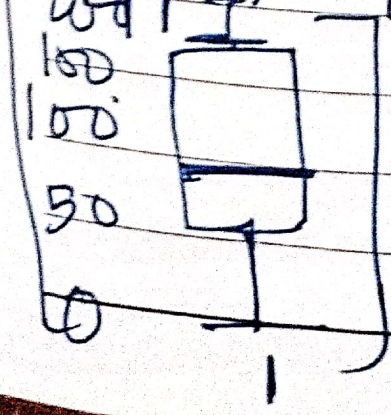
$$Z\text{-}Score = \frac{x-\bar{x}}{\delta}$$

$$Median = l + \left(\frac{N/2 - cf}{f}\right) \times h.$$

Outlies detection (IQR)

lower limit = Q1 − 1.5 × IQR

Upper limit = Q3 + 1.5 × IQR



Boxplot.

Mo Tu We Th Fr Sa Su Date: / /

Min - max Normalized Stock Price

X = Stock Price

$$X = \frac{X - min}{max - min}$$

Mid - range Stock price

$$= \frac{min + max}{2}$$

~~Z score~~ Z-Score (25)

$$= \frac{x - \bar{x}}{\sigma}$$

Decimal Scaling

$$= X' = \frac{X}{10^j}$$

Mean > median (Right Skewed +)

median > mean (-)

Mo Tu We Th Fr Sa Su

Date: / /

$$CI = \bar{x} \pm t_{\alpha/2, df} \times \left(\frac{s}{\sqrt{n}}\right)$$

Sample Proportion $= \hat{p} = \frac{\text{Success}}{\text{size}}$.

$\hat{q} = 1 - \hat{p}$

$$CI = \hat{p} \pm z \times \sqrt{\frac{\hat{p}\hat{q}}{n}}$$

Population Proportion

Test Statistic (z-score)

$$z = \frac{\hat{p} - \pi_0}{\sqrt{\frac{\pi_0(1-\pi_0)}{n}}}$$

— P-Value?

Test Statistic $= T = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$

Mo Tu We Th Fr Sa Su

Date: / /

Null hypothesis (H_0):

$$H_0 : \mu \leq 50$$
$$H_1 : \mu > 50$$

The Test Statistic 'T' measures how many Standard Errors the Sample mean is away from the hypothesized population mean ($50)

$$T = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1^2} + \frac{s_2^2}{n_2}}}$$

Chi-Square

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

P-Value $\leq \alpha$ reject H_0

p-value $> \alpha$ fail to reject H_0

E_{ij} = (Row Total)(Column Total) / Grand Total

Simple linear Regression

- one independent Variable (X)
- and one dependent Variable (Y)

$$Y = \beta_0 + \beta_1 X + \varepsilon$$

Multiple linear Regression

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n + \varepsilon$$

2 or more independent Variables and one dependent Variable.

Least Squares Estimation Method

$$Y = a + bX$$

$$b = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2}$$

$$a = \frac{\sum Y - b\sum X}{n}$$

SST = SSR + SSE

Mo Tu We Th Fr Sa Su

Date: / /

$SSE = \sum (y - \hat{y})^2$

$SSR = \sum (\hat{y} - \bar{y})^2$

$SST = \sum (y - \bar{y})^2$

$\hat{y}$ = Put the x value and find this

Coefficient of Determination

$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST}$

Mean Square Error

$= MSE = \frac{SSE}{n-2}$

Standard Error $= \sqrt{MSE} = \sqrt{\frac{SSE}{n-2}}$

Pearson Correlation formula

$$r = \frac{n\sum xy - (\sum x)(\sum y)}{\sqrt{[n\sum x^2 - (\sum x)^2][n\sum y^2 - (\sum y)^2]}}$$

Co-relation

Coefficient = $P = \sqrt{R^2}$ Date: / /

↳ Coefficient of determination

Min - Max.

Normalization

$$X' = \frac{x - X_{min}}{X_{max} - X_{min}}$$

Euclidean Distance

$$d = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$$

Manhattan Distance:

$$d(x,y) = \sum |x_i - y_i|$$

minkowski distance

$$= d(x,y) = \left\{ \sum |x_i - y_i|^k \right\}^{1/k}$$

$(k \in N)$

Discrete Metric

$d(x,y) = 1$

if $x \neq y$ and $d(x,y) = 0$

Weighted Average

$$= \frac{\sum_{i=1}^{k} w_i y_i}{\sum_{i=1}^{k} w_i}$$

Robust Scaling

$$= \frac{X - Q_2}{IQR}$$

Types of Errors in Estimation

1. Residual error
2. Measurement error
3. Model error
4. Random error