classmate
Date
Page

SM : U3 – Classification of Data.

Mean $= \dfrac{\text{all sum}}{\text{no. of terms.}}$

Median :=

1. Raw Data : Ascending Order

   if $N =$ even $\Rightarrow \left(\dfrac{N}{2}\right)^{th}$ term

   $N =$ odd $\Rightarrow \left(\dfrac{N+1}{2}\right)^{th}$ term.

2. Ungrouped Data : calculate CF

   $\dfrac{N}{2}$ (ie. ΣF) $=$ midde value

   for CF just $>$ middle value

3. Grouped Data :

$$L + \frac{H}{f}\left[\frac{N}{2} - C\right]$$

L : lower bound

N : total frequency

H : difference (UL – LL)

C : CF – 1

f : frequency of a median class.

Date
Page

Mode:

Raw Data : ~~m~~ value of the most appearing ~~v~~ term.

Ungrouped Data : ~~max. freq.~~
value of max. frequency

Grouped Data : $L + \frac{H(f_1 - f_0)}{2f_1 - f_0 - f_2}$

L : lower bound of highest freq.
H : width (UL − LL)
f_0 : freq. of interval before highest freq.
f_1 : highest freq.
f_2 : freq. of interval after ~~befo~~ highest freq.

Quartile : $Q_i = L + \frac{H}{f}\left(\frac{iQ}{4} - c\right)$

Decile : $D_i = L + \frac{H}{f}\left(\frac{iD}{10} - c\right)$

Percentile : $P_i = L + \frac{H}{f}\left(\frac{iP}{100} - c\right)$

Geometric Mean :

Raw Data : $\sqrt[n]{x_1 . x_2 . x_3 \dots x_n}$.

Ungrouped Data : $\sqrt[n]{x_1^{f1} . x_2^{f_2} . x_3^{f_3} \dots x_n^{f_n}}$

$$n = \Sigma f.$$

Grouped Data : anti-log $\left[\frac{\Sigma f \cdot \log x}{N} \right]$

Wrong Entry : $G_{new} = \left[\frac{(G)^N \cdot new}{old} \right]^{1/N}$

Harmonic Data :

Raw : $\dfrac{N}{\Sigma \frac{1}{x}}$

Ungrouped : $\dfrac{N}{\Sigma \frac{f}{x}}$

$(GM)^2 = (AM) \times (HM)$

classmate
Date
Page

SM: U3 - Presentation of Data.

Histogram:

Step 1: make class-interval continuous.
Step 2: plot.

Frequency:

Step 1: make class-interval continuous.
Step 2: find midpoint of class-intervals.
Step 3: plot and connect dots.
Step 4: connect the start ~~to~~ ~~a~~ and end to
~~Ogive~~ the ~~upper~~ lower limit and upper limit of class boundary to x-axis.

Ogive:

Step 1: make class-interval continuous.
Step 2: find cumulative frequency.
Step 3: ~~po~~ plot point at upper class boundary.
Step 4: extend graph to the first lower class boundary.

Stem & Leaf:

Step 1: ~~sigh~~ write the tens or greater place in stem while units in leaf.
Step 2: arrange in ascending order.
Step 3: key ⇒ eg. 2|3 means 23.

classmate

Date
Page

Skewness:

- S_k = mean − median = mean − mode.
  $= (Q_3 - Q_2) - (Q_2 - Q_1)$

Note: central moments (μ_r) (about mean)

$$\therefore \mu_r = \frac{1}{N} \Sigma f(x - \bar{x})^r, \ r = 0, 1, 2 \ldots$$

$$\mu_0 = 1 \ ; \ \mu_1 = 0 \ ; \ \mu_2 = \text{var.}$$

★ mode = 3median − 2mean.

- $\beta_1 = \dfrac{\mu_3^2}{\mu_2^3}$

- $\gamma_1 = \pm \sqrt{\beta_1}$

Karl Pearson's coeff of Skewness:

- $S_k = \dfrac{\text{mean} - \text{mode}}{\sigma} = \dfrac{3(\text{mean} - \text{median})}{\sigma}$

  $= \dfrac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$

Kurtosis:

I. Karl Pearson's co-eff / measure:

- $\beta_2 = \frac{\mu_4}{\mu_2^2}$

- $\gamma_2 = \beta_2 - 3$.

II. Kelly's measure:

$$\beta_2 = \frac{P_{75} - P_{25}}{P_{90} - P_{10}}.$$

Measures of Dispersion:

1. Range : $X_{max} - X_{min}$.

   coeff of range : $\frac{X_{max} - X_{min}}{X_{max} + X_{min}}$

2. Quartile Deviation : $QD = \frac{Q_3 - Q_1}{2}$

   co-eff of QD $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$.

inter-quartile range : $Q_3 - Q_1$

* semi-inter-quartile range : quartile deviation.

classmate

Date

Page

3. Mean Deviation : $\frac{1}{n} \Sigma |x - A| \Rightarrow$ raw

$\frac{1}{N} \Sigma f|x - A| \Rightarrow$ grouped / ungrouped.

coeff MD = $\frac{\text{MD}}{\text{avg. about which it is calculated.}}$

coeff of MD about mean = $\frac{\text{MD}}{\text{mean}}$

coeff of MD about median = $\frac{\text{MD}}{\text{median}}$

4. Standard Deviation : $\sigma = \sqrt{\frac{1}{n} \Sigma (x - \bar{x})^2}$

σ^2 = variance.

$$\mu_1 = \mu_1' - \mu_1'$$

$$\mu_2 = \mu_2' - (\mu_1')^2$$

$$\mu_3 = \mu_3' - 3\mu_1'\mu_2' + 2\mu_1'^3$$

$$\mu_4 = \mu_4' - 4\mu_1'\mu_3' + 6\mu_2'\mu_1'^2 - 3\mu_1'^4$$

classmate
Date
Page

SM: U4 - Simple & Multiple Linear Regression Model.

1. coeff of correlation : r.

$$r = r(x,y) = \frac{n\sum xy - \sum x \cdot \sum y}{\sqrt{(n\sum x^2 - (\sum x)^2) \cdot (n\sum y^2 - (\sum y)^2)}}.$$

2. regression coeff : b_{xy} and b_{yx}

b_{xy}: x on y; b_{yx}: y on x.

~~3. equation of lines of regression : y and x.~~

$\cancel{x \text{ on } y : x - \bar{x} = b_{xy}(y - \bar{y})}$

$$b_{xy} = \frac{n\sum xy - \sum x \cdot \sum y}{\sum y^2 - (\sum y)^2} \; ; \; b_{yx} = \frac{n\sum xy - \sum x \sum y}{\sum x^2 - (\sum x)^2}.$$

$$= \frac{cov(x,y)}{\sigma_y^2} \qquad = \frac{cov(x,y)}{\sigma_x^2}$$

$$= \frac{r\sigma_x}{\sigma_y} \qquad = r\frac{\sigma_y}{\sigma_x}.$$

* $b_{xy} \cdot b_{yx} = r$.

3. equation of lines of regression : y and x.

x on y : $x - \bar{x} = b_{xy}(y - \bar{y})$

y on x : $y - \bar{y} = b_{yx}(x - \bar{x})$

Multiple Linear Regression Model:

$$\hat{y} = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_k x_k.$$

residual $= y - \hat{y}$.

$$SS_{xy} = \Sigma xy - \frac{\Sigma x \cdot \Sigma y}{n}$$

$$SS_{xx} = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$$

$$\therefore \quad \beta_1 = \frac{SS_{xy}}{SS_{xx}}$$

$$\beta_0 = \frac{\Sigma y}{n} - \beta_1 \frac{\Sigma x}{n}$$

- Sum of sq. of error : $\Sigma (y - \hat{y})^2$.

- Standard error : $S_e = \sqrt{\dfrac{SSE}{n-k-1}}$

here, n : no. of observations.
k : no. of independent variables.
$n-k-1$: degree of freedom

$$a = \frac{n \Sigma xy - \Sigma x \Sigma y}{n \Sigma x^2 - (\Sigma x)^2} \; ; \; b = \frac{1}{n} (\Sigma y - a \Sigma x)$$

—x—

classmate

Date
Page

SM : U5 - Multivariate Regression.

1. general equation : $Y = X\beta + \epsilon$.

extended equation : $y = \beta_0 + \beta_1 x_1 + \epsilon_i ; i = 0, 1, \dots n$

matrix : $\begin{bmatrix} y_1 \\ y_2 \\ y_n \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & x_{1m} \\ 1 & x_{21} & x_{2m} \\ 1 & x_{n1} & x_{nm} \end{bmatrix} \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_m \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_n \end{bmatrix}$

$(n \times m)$

2. $\beta = (X'X)^{-1} . (X'Y) = b = \begin{bmatrix} b_0 \\ b_1 \end{bmatrix}$

$\epsilon = Y - \hat{Y} \quad (\hat{Y} = X\hat{\beta})$

$RSS = \epsilon . \epsilon^T$

classmate

Date

Page

SM : U6 - Linear Discriminant Analysis.

Step 1: $\mu_1 = \frac{1}{N_1} \cdot \sum_{i=1}^{N_1} x_i$

$\mu_2 = \frac{1}{N_2} \sum_{i=1}^{N_2} x_i$

Step 2: find : $[x_1 - \mu_1]$, $[x_2 - \mu_2]$.

$[x_1 - \mu_1]^T$, $[x_2 - \mu_2]^T$.

Step 3: $S_1 = \frac{1}{N_1 - 1} (x_1 - \mu_1) \cdot (x_1 - \mu_1)^T$

$S_2 = \frac{1}{N_2 - 1} (x_2 - \mu_2) \cdot (x_2 - \mu_2)^T$.

Step 4: $S_W = S_1 + S_2$

Step 5: $w^* = S_W^{-1} (\mu_1 - \mu_2)$

Step 1: $\mu_1 = \frac{1}{N_1} \cdot \sum_{i=1}^{N_1} x_i$

$\mu_2 = \frac{1}{N_2} \cdot \sum_{i=1}^{N_2} x_i$

Step 2: $\mu = \frac{\mu_1 + \mu_2}{2}$.

Step 3: $x_i = x_i - \mu$.

step 4: $C_1 = \frac{1}{N_1} \cdot (x_i)^T (x_i)$

$C_2 = \frac{1}{N_2} \cdot (x_2^i)^T (x_2^i)$

step 5: $C = \frac{N_1}{N_1+N_2} C_1 + \frac{N_2}{N_1+N_2} C_2.$

and,

$C^{-1}.$

step 6: $f_i = \mu_i C^{-1} \cdot x_k^T - \frac{1}{2} \mu_i C^{-1} \mu_i + \ln(p_i)$

for,

$p_1 = \frac{n_1}{n_1+n_2}$

$p_2 = \frac{n_2}{n_1+n_2}.$

Note: x_k : given

classmate
Date
Page

SM : U7 - Principal Component Analysis.

Step 1: $\mu_1 = \frac{1}{N_1} \cdot \sum_{i=1}^{N_1} x_i$

$\mu_2 = \frac{1}{N_2} \cdot \sum_{i=1}^{N_2} x_i$

Step 2: $A = \begin{bmatrix} x_i - \mu_1 \\ x_j - \mu_2 \end{bmatrix}$

Step 3: method 1: $S = \frac{A \cdot A^T}{n}$

method 2: $\cancel{S = [cov(x_1, x_1)\ cov(x_1, x_2)\ cov}$

$$S = \begin{bmatrix} cov(x_1, x_1) & cov(x_1, x_2) \\ cov(x_2, x_1) & cov(x_2, x_2) \end{bmatrix}$$

where,

$$cov(x_i, x_j) = \frac{1}{N-1} \sum_{K=1}^{N} (x_{ik} - \bar{x}_i)(x_{jk} - \bar{x}_j)$$

Step 4: form characteristic equation.

ie. $|S - \lambda I| = 0$

and equation for λ and thus values of λ (ie. eigen value)

Step 5: for each λ,

find $X = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix}$;

hint : apply cramer's rule.

classmate
Date
Page

step 6: for highest λ,

$$\|X\| = \sqrt{x_1^2 + x_2^2}.$$

step 7: ~~for / each vector~~,

$$\text{normalization} \quad PCA = \begin{bmatrix} \dfrac{x_1}{\|X\|} \\[2ex] \dfrac{x_2}{\|X\|} \end{bmatrix} = e_1.$$

step 8: for each PCA, ; vector

$$\text{component} = e_1^T \cdot \begin{bmatrix} A_i \\ A_j \end{bmatrix}$$

step 9:

$$\left.\begin{aligned} z_1 &= e_1 \cdot X \\ z_2 &= e_2 \cdot X. \end{aligned}\right\} \text{principal component axis}$$

step 10: explained variance,

$$\lambda_1 = \frac{\lambda_1}{\lambda_1 + \lambda_2} \times 100\,\%.$$

$$\lambda_2 = \frac{\lambda_1 + \lambda_2}{\lambda_1 + \lambda_2} \times 100\,\%.$$

—X—