Date # PROBABILITY & STATISTICS:

Grouped Data:

Class Interval $(K) = 1 + 3.31 * \log(n) = \sqrt{n} = \log(n)/\log 2$

Range $=$ Max $-$ Min

Height $=$ Range $/K$

Measure of Centeral Tendency:

* Ungrouped Data:

even $\Rightarrow n$ odd $\Rightarrow n$

Mean $= \bar{X} = \frac{\Sigma x}{n}$ Median $: \tilde{X} = \left(\frac{n}{2}\right)^{th} + \left(\frac{n+2}{2}\right)^{th} / 2$ $\quad \frac{(n+1)}{2}^{th}$

Variance $= \frac{\Sigma x^2}{n} - \left(\frac{\Sigma x}{n}\right)^2 = \frac{\Sigma x^2}{n} - \bar{X}^2$ Standard dev $= \sqrt{Var}$

$G.M = \text{Antilog}\left(\frac{\Sigma \log x}{n}\right)$ $H.M = \frac{n}{\Sigma(1/x)}$

$-\infty \quad \infty$

HM GM AM

* Grouped Data:

$\bar{X} = \frac{\Sigma fx}{\Sigma f} = \frac{\Sigma fx}{n}$ Median $= LL + \frac{h}{f}\left(\frac{\Sigma f}{2} - CF_p\right)$

Mode $= LL + h\left[\frac{fm - f_1}{2fm - f_1 - f_2}\right]$ $GM = \left(\frac{\Sigma \log fx}{\Sigma f}\right)$ Antilog

$H.M = \frac{\Sigma f}{\Sigma(f/x)}$ $Var = \frac{\Sigma fx^2}{n} - \bar{X}^2$ $S.D = \sqrt{Var}$

Measure of Position:

* Quartile : $\frac{i(n+1)}{4}^{th}$ $= LL + \frac{h}{f}\left(\frac{i\Sigma f}{4} - CF_p\right)$.

for group $= \frac{i\Sigma f}{4}^{th}$

Date

* Deciles $= \frac{i(n+1)}{10}^{th} = \frac{i\Sigma f}{10}^{th} \rightarrow LL + \frac{h}{f}\left(\frac{i\Sigma f}{10} - Cf_p\right)$.

* Percentile $= \frac{i(n+1)}{100}^{th} = \frac{i\Sigma f}{100}^{th} \rightarrow LL + \frac{h}{f}\left(\frac{i\Sigma f}{100} - Cf_p\right)$

$Q_1 = P_{25}$ | $Q_2 = P_{50}$ | $Q_3 = P_{75}$ | $\tilde{X} = Q_2 = D_5 = P_{50}$

$D_1 = P_{10}$ | $D_2 = P_{20}$ | $D_3 = P_{30}$ | $IQR = Q_3 - Q_1$

Measure of Disperssion:

*** Absolute:**

Range = Max − Min

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

Var = (Same as above)

SD = $\sqrt{var}$

Mean Dev $= \frac{\Sigma(X - \bar{x})}{n} = \frac{\Sigma(fx - \bar{x})}{n}$

Median Dev $= \frac{\Sigma(X - \tilde{X})}{n} = \frac{\Sigma(fx - \tilde{X})}{n}$

*** Relative:**

Co. of Range $= \frac{Max - Min}{Max + Min}$

Co. of Quartile Dev $= \frac{Q_3 - Q_1}{Q_3 + Q_1}$

Co. of ~~Var~~ $= \frac{SD}{\bar{x}} \times 100$

Co. of SD $= \frac{SD}{\bar{x}}$

Co. of Mean Dev $= \frac{MD}{\bar{x}}$

Co. of Median Dev $= \frac{MD}{\tilde{x}}$

Moments About Mean:

*** Ungrouped Data:**

$\mu_1 = \frac{\Sigma(X - \bar{X})}{n} = 0$; $\mu_3 = \frac{\Sigma(X - \bar{X})^3}{n}$

$\mu_2 = \frac{\Sigma(X - \bar{X})^2}{n}$; $\mu_4 = \frac{\Sigma(X - \bar{X})^4}{n}$

*** Grouped Data.**

$\mu_1 = \frac{\Sigma(fx - \bar{x})}{n} = 0$; $\mu_3 = \frac{\Sigma(fx - \bar{x})^3}{n}$

$\mu_2 = \frac{\Sigma(fx - \bar{x})^2}{n}$; $\mu_4 = \frac{\Sigma(fx - \bar{x})^4}{n}$

Scanned with CamScanner

Skewness $= \beta_1 = \frac{\mu_3^2}{\mu_2^3}$ Kurtosis $= \beta_2 = \frac{\mu_4}{\mu_2^2}$

Counting Technique:

Permutation (arrangment):

$${}^nP_r = \frac{n!}{(n-r)!}$$

Combination (Selection):

$${}^nC_r = \frac{n!}{r!(n-r)!}$$

* For all different objects:

repetation: n^r

no repetation $= {}^nP_r = \frac{n!}{(n-r)!}$

For same objects:

$\cancel{P(n,r)} = \frac{n!}{n_1!\, n_2! \ldots n_n!}$

Circular Permutation: $(n-1)!$

* if rep. count them 1.

Axioms of Probability:

i) $0 \leq P(\text{Event}) \leq 1$

ii) $P(\text{Sample}) = 1$.

iii) $P(A') = 1 - P(A)$.

Additive Rule:

* $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
* $P(A \cup B \cup C) = P(A) + P(B) + P(C) + P(A \cap B \cap C) - \big(P(A \cap B) + P(B \cap C) + P(A \cap C)\big)$.

Conditional Probability:

$$P(D|I) = \frac{P(I \cap D)}{P(I)}$$

$P(D \cap I) = P(D|I) \cdot P(I)$ } * Dependent condition.

$$P(D|I') = \frac{P(D) - P(D \cap I)}{P(I')} = \frac{P(D) - P(D \cap I)}{1 - P(I)}.$$

$P(D'|I) = 1 - P(D|I)$

$P(D|I) = P(D)$

$P(D \cap I) = P(D) \cdot P(I)$ } independent condition.

$$P(B_r|A) = \frac{\overset{P(A \cap B_r)}{\overline{P(B_r) \cdot P(A|B_r)}}}{\underbrace{\sum_{i=1}^{k} P(B_i)\, P(A|B_i)}}$$ $P(A \cap B_i)$

Bayes' Rule: $P(B_r|A) = \dfrac{P(B_r) \cdot P(A|B_r)}{\sum_{i=1}^{k} P(B_i)\, P(A|B_i)}$

Continuous Probability Distribution:

① $P(a < x < b) = \int_a^b f(x)\,dx$ $P(X=x)$ ÷ does not exist.

② $f(x) \geq 0$.

③ $\int_{-\infty}^{+\infty} f(x)\,dx = 1$

CDF for discrete PD:

CDF = $F(x) = P(X \leq x)$

CDF = $F(x) = \int_{-\infty}^{x} f(x)\,dx$

Joint Probability Distribuition / Probability Mass Function.

* Discrete :

(1) $f(x,y) \geq 0$

(2) $\sum_x \sum_y f(x,y) = 1$

(3) $P(X=x, Y=y) = f(x,y)$.

* Continuous:

(1) $f(x,y) \geq 0$

(2) $\int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} f(x,y)\,dx\,dy = 1$

(3) $P(a<x<b, c<y<d) = \int_a^b \int_c^d f(x,y)\,dy\,dx$

Marginal Probability Distribuition:

* Discrete

$g(x) = \sum_y f(x,y)$

$h(y) = \sum_x f(x,y)$

* Continuous:

$g(x) = \int_{-\infty}^{\infty} f(x,y)\,dy$

$h(y) = \int_{-\infty}^{+\infty} f(x,y)\,dx$

Scanned with CamScanner

Continuous Probability Distribuition:

$$f(x|y) = \frac{f(x,y)}{h(y)} \qquad f(y|x) = \frac{f(x,y)}{g(x)}$$

Mean & Variance of Random Variable:

*Discrete

Mean $= E(x) = \sum_x x f(x)$.

Var $= E(x^2) - (E(x))^2$

$\therefore E(x^2) = \sum_x x^2 f(x)$.

*Continuous:

Mean $= E(x) = \int_{-\infty}^{+\infty} x f(x)\, dx$

Var $= E(x^2) - (E(x))^2$

$E(x^2) = \int_{-\infty}^{+\infty} x^2 f(x)\, dx$

Mean & Variance of Joint Variable:

*Discrete:

Mean $= E(x,y) = \sum_x \sum_y f(x,y)\,(x \cdot y)$

Var $= E(x,y) - \mu_x \mu_y$.

$\mu_x = \sum_x x\, g(x)$

$\mu_y = \sum_y y\, h(y)$

Mean $= E(x,y) = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} (xy)\, f(x,y)\, dy\, dx$

Var $= E(x,y) - \mu_x \mu_y$

$\mu_x = \int_{-\infty}^{+\infty} x\, g(x)\, dx \; ; \; \mu_y = \int_{-\infty}^{+\infty} y\, h(y)\, dy$

Discrete Probability Distribution:

*Uniform distribuition:

$P(x, k) = 1/k$

Scanned with CamScanner

Date

* Binomial Distribution:

$$b(x,n,p) = \binom{n}{x} p^x q^{n-x} \qquad p+q=1.$$

$$\mu = np \qquad \sigma^2 = npq.$$

* Multinomial Distribuition:

$$P(x_1, x_2 \dots x_n ; p_1, p_2, \dots p_n ; n) = \binom{n}{x_1, x_2 \dots x_n} p_1^{x_1} p_2^{x_2} \dots p_n^{x_n}$$

$\Rightarrow \Sigma x = n$

$\Rightarrow \Sigma p = 1$

* Hyper geometric Distribuition: (replacement is not possible).

$$P(X=x) = h(x, N, n, K) = \frac{\binom{K}{x}\binom{N-K}{n-x}}{\binom{N}{n}}$$

N = population

n = Sample

K = # of sucess

x = random vaiable

$$\mu = n\frac{K}{N} \qquad 6^2 = \frac{N-n}{N-1} \cdot n \cdot \frac{K}{N}\left(1 - \frac{K}{N}\right) = \frac{N-n}{N-1} npq$$

$p = K/N \qquad P \Rightarrow$ sucess probability.

if $\frac{n}{N} \leq 0.005$ apply binomial. $b(x, n, K/N)$.

* Poisson Distribution:

$$P(x, \lambda t) = \frac{e^{-\lambda t}(\lambda t)^x}{x!} = \frac{e^{-\lambda}(\lambda)^x}{x!} \qquad \mu = \lambda t = \lambda = 6^2$$

if $n \to \infty$ & $p \to 0$ $\qquad \lambda = np = \mu$

$$b(x, n, p) \xrightarrow{\lambda = np} p(x, \lambda).$$

Scanned with CamScanner

Continuous Uniform Distribuition

$$f(x, A, B) = \begin{cases} 1/B-A & A \leq x \leq B \\ 0 & \text{else where} \end{cases} \qquad \mu = \frac{A+B}{2} \qquad \sigma^2 = \frac{(B-A)^2}{12}$$

Normal Distribuition:

$$n(x, \mu, \sigma) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \qquad z = \frac{x-\mu}{\sigma}$$

Standard Normal Distribuition:

$\mu = 0$ && $\sigma^2 = 1$

Normal Approximation to Binomial:

$n \to \infty$ $np > 5$ $p, q \neq 0$.
$nq > 5$

use law of Continuity = $P(x - 0.5 < x < x + 0.5)$.

$$P(X \leq x) = \sum_{k=0}^{x} b(k, n, p) = P\left(Z \leq \frac{x + 0.5 - np}{\sqrt{npq}}\right)$$

* $P(X \leq x) = P\left(Z \leq \frac{x + 0.5 - \mu}{\sigma}\right)$.

* $P(X \geq x) = P\left(z \geq \frac{x - 0.5 - \mu}{\sigma}\right)$

* $P(X < x) = P\left(z \leq \frac{x - 1 + 0.5 - \mu}{\sigma}\right)$ $P(X > x) = P\left(z \geq \frac{x + 1 - 0.5 - \mu}{\sigma}\right)$

Scanned with CamScanner

CHI - SQUARED Distribuition :

$\chi_K^2 = \sum_{i=1} \frac{(O_i - E_i)^2}{E_i}$ K = # of free variable = $n-1$.

n = # of outcomes

Square of normal dist. E_i = expected value. α = level of significance

$K \to \infty$ (almost bell shaped, Normal Dist.). O_i = Observed vale

$\mu = K$ $\sigma^2 = 2K$ $E_i = (\Sigma f) n_i$

* Indepency check of Independent variable on multiple dependent variable.

$E_i = \frac{(\text{row total})(\text{col total})}{(\text{grand total})}$ $K = (r-1)(c-1)$

r = # of rows c = # of columns.

* Fitting a distribuition :

$\boxed{\begin{array}{l} Max = \mu + 3\sigma \\ Min = \mu - 3\sigma \end{array}}$

* Binomial Distribuition:

$p = \frac{\Sigma f x}{n \Sigma f} = \frac{\bar{x}}{n}$ $P = {}^nC_x \, p^x q^{n-x}$

* Poisson Distribuition:

$\lambda = \frac{\Sigma f x}{\Sigma f} = \bar{x}$ $P = \frac{e^{-\lambda} \lambda^x}{x!}$

* Normal Distribuition:

$\mu = \frac{\Sigma f x}{\Sigma f} = \bar{x}$ $z = \frac{\text{Upper CB} - \mu}{\sigma}$

$\sigma = \frac{\Sigma f x^2}{\Sigma f} - \mu^2$ $P(x) = P(x_n - x_{n-1})$

Scanned with CamScanner

Centeral Limit Theorem:

$$t, z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

$n \geq 30 \Rightarrow$ CLT (Z-Test)
$n < 30 \Rightarrow$ T-Test

$$S = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$$

deg. of freedom: $V = n-1$] T-Test

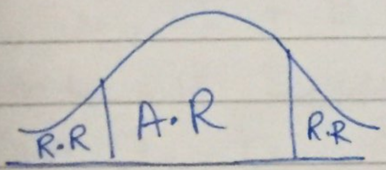
Hypothesis Testing:

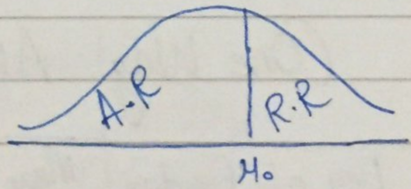
H_0 = Null Hyp (given) H_a = Alternate Hyp. μ = parameters of hyp testing.

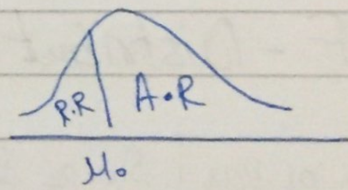
$H_0 : \mu = X \Rightarrow H_A : \mu \neq X$
$H_0 : \mu \geq X \Rightarrow H_A : \mu \leq X$
$H_0 : \mu \leq x \Rightarrow H_A : \mu \geq X$

$=, \neq$ condition $\Rightarrow$ two tail test.

$\leq \ \geq$ condition $\Rightarrow$ one tail test.
$\mu_0 < \mu_A$ (RR $\Rightarrow$ RHS)
$\mu_0 > \mu_A$ (RR $\Rightarrow$ LHS)







* Confidence Interval.

Lvl of Confiden = 1 − Lvl of Sign.

$$\bar{X} - z_{\alpha/2} \cdot \frac{S}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2} \cdot \frac{S}{\sqrt{n}}$$

$\frac{\alpha}{2} \Rightarrow$ divide value of z by 2

$$\bar{X} - t_{\alpha/2} \cdot \frac{S}{\sqrt{n}} < \mu < \bar{X} + t_{\alpha/2} \frac{S}{\sqrt{n}}$$

$\frac{\alpha}{2} \Rightarrow$ refers to two tail test.

Scanned with CamScanner

Date

Hypothesis Testing for 2 mean.

$$z_{cal} = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$$df = n_1 + n_2 - 2$$

$$t_{cal} = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{sp\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$sp = \sqrt{\frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}}$$

* Confidence Interval:

$$(\bar{X}_1 - \bar{X}_2) - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$$

$$(\bar{X}_1 - \bar{X}_2) - t_{\frac{\alpha}{2}} \cdot sp \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{X}_1 - \bar{X}_2) + t_{\frac{\alpha}{2}} \cdot sp \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

F-Distribution (One Way ANOVA).

Source of Var	Sum of Sq	Deg of freedom	Mean Sq	Computed f.
Treatment	SST	K−1	$S_1^2 = SST/K-1$	$f = \frac{S_1^2}{S_2^2}$
Error	SSE	N−K	$S_2^2 = SSE/N-K$	
Total	TSS	N−1		

SST (Sum of Square of Treatment) $= \sum_{i=1}^{K} \frac{Y_{i\cdot}^2}{n} - \frac{Y_{\cdot\cdot}^2}{N}$

Scanned with CamScanner

Total Sum of Squares. $TSS = \sum_{i=1}^{k} \sum_{j=1}^{n} Y_{ij}^2 - \frac{Y_{..}^2}{N}$

$SSE = TSS - SST.$

K = # of parameters. n = # of elements in each sample
N = Total Population. $Y_{..}$ = Sum of all values.

Co Relation: (Linear Co-Relation & Regression).

*Co. of Co-relation: $r = \frac{n \sum xy - \sum x \sum y}{\sqrt{[n \sum x^2 - (\sum x)^2][n \sum y^2 - (\sum y)^2]}}$

*Test the hypothesis using t-table. $K = n-2$

$\alpha = 0.05$

$t = r\sqrt{\frac{n-2}{1-r^2}}$

$H_0 : \rho = 0$

$H_a : \rho_A \neq 0.$

$r \Rightarrow$ sample

ρ = Population.

Simple Regression Model:

$Y_i = B_0 + B_1 x_i + \varepsilon_i$.

* Least Square parameters:

$a_0 = b_0 = \frac{\sum y \sum x^2 - \sum x \sum xy}{n \sum x^2 - (\sum x)^2}$

$b : b_1 = \frac{n(\sum xy) - \sum x \sum y}{n \sum x^2 - (\sum x)^2}$

Standard Error: $S_{xy} = \sqrt{\frac{\sum (Y - Y_i)^2}{n-2}}$

$SSE = \sum_{i=1}^{n} (Y - Y_i)^2$

n = # of Samples.

Multiple Regression:

$Y_i = a + b_1 x_1 + b_2 x_2$

Error Est $= S_{Y12} = \sqrt{\dfrac{\Sigma (Y - Y_i)^2}{n - (K+1)}}$

$a = \bar{y} + b_1 \bar{x}_1 + b_2 \bar{x}_2$

$b_1 = \dfrac{\Sigma x_1 y \, \Sigma x_2^2 - \Sigma x_2 y \, \Sigma x_1 x_2}{\Sigma x_1^2 \, \Sigma x_2^2 - (\Sigma x_1 x_2)^2}$

$b_2 = \dfrac{\Sigma x_2 y \, \Sigma x_1^2 - \Sigma x_1 y \, \Sigma x_1 x_2}{\Sigma x_1^2 \, \Sigma x_2^2 - (\Sigma x_1 x_2)^2}$

n = # of sample
K = # of independent var.

$\bar{y}$ = mean of Y
$\bar{x}_1$ = mean of x_1
$\bar{x}_2$ = mean of x_2

$R_{Y12} = \sqrt{\dfrac{a \Sigma y + b_1 \Sigma x_1 y + b_2 \Sigma x_2 y - (\Sigma y)^2/n}{\Sigma y^2 - (\Sigma y)^2/n}}$

$R = \sqrt{\dfrac{r_{yx_1}^2 + r_{yx_2}^2 - 2 r_{yx_1} r_{yx_2} r_{x_1 x_2}}{1 - r_{x_1 x_2}^2}}$

$r_{xy} = \dfrac{n \Sigma xy - \Sigma x \Sigma y}{\sqrt{[n \Sigma x^2 - (\Sigma x)^2][n \Sigma y^2 - (\Sigma y)^2]}}$

Anova Model for Simple Regression:

Types of Var	Sum of Sq	Deg of freedom	Mean Sq	Computed f.
Regression	$\Sigma (y_i - \bar{y})^2$	$K-1$	$\Sigma (y_i - \bar{y})^2 / K-1 = S_1^2$	$f = \dfrac{S_1^2}{S_2^2}$
Residual	$\Sigma (y - y_i)^2$	$n-2$	$\Sigma (y - y_i)^2 / n-2 = S_2^2$	
Total	$\Sigma (y - \bar{y})^2$	$n-1$	$\Sigma (y - \bar{y})^2 / n-1$	

Scanned with CamScanner