Testing of Hypothesis

Test of significance for a single mean(μ) **when n≥30** use z-test: Test statistics $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$

Test of significance for a single mean(μ) **when n<30** use t-test: Test statistics $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$

Test of significance for different of two means ($\mu_1 - \mu_2$), **when ($n_1 \geq 30$, $n_2 \geq 30$)** use z-test

Test Statistics $z = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$,

Test of significance for different of two means ($\mu_1 - \mu_2$) **when ($n_1 < 30$, $n_2 < 30$)** use t-test.

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

Test of significance for a single proportion(P):TS $z=\frac{p-P}{\sqrt{\frac{PQ}{n}}}$ when two n are given then $z=\frac{p1-P2}{\sqrt{PQ(\frac{1}{n1}+\frac{1}{n2})}}$

Multiple Correlation and Multiple Regression

Partial Correlation Coefficient

$r_{12.3}=\frac{r_{12}-r_{13}r_{23}}{\sqrt{1-r_{13}^2}\sqrt{1-r_{23}^2}}$

Partial Correlation Coefficient

$r_{13.2}=\frac{r_{13}-r_{12}r_{32}}{\sqrt{1-r_{12}^2}\sqrt{1-r_{32}^2}}$

Partial Correlation Coefficient

$r_{23.1}=\frac{r_{23}-r_{21}r_{31}}{\sqrt{1-r_{21}^2}\sqrt{1-r_{31}^2}}$

Multiple Correlation

$R_{1.23}=\sqrt{\frac{r_{12}^2+r_{13}^2-2r_{12}r_{13}r_{23}}{1-r_{23}^2}}$

$0 \leq R_{1.23} \leq +1$

Multiple Correlation

$R_{2.13}=\sqrt{\frac{r_{21}^2+r_{23}^2-2r_{21}r_{23}r_{13}}{1-r_{13}^2}}$

$0 \leq R_{2.13} \leq +1$

Multiple Correlation

$R_{12.3}=\sqrt{\frac{r_{31}^2+r_{32}^2-2r_{31}r_{32}r_{12}}{1-r_{12}^2}}$

$0 \leq R_{3.12} \leq +1$

$r_{12}=\frac{n\sum u_1u_2-\sum u_1\sum u_2}{\sqrt{n\sum u_{11}-(\sum u_1)^2\sum}\sqrt{n\sum u_{21}-(\sum u_1)^2}}$

$r_{13}=\frac{n\sum u_1u_3-\sum u_1\sum u_3}{\sqrt{n\sum u_{11}-(\sum u_1)^2\sum}\sqrt{n\sum u_{31}-(\sum u_3)^2}}$

$r_{23}=\frac{n\sum u_2u_3-\sum u_2\sum u_3}{\sqrt{n\sum u_{21}-(\sum u_2)^2\sum}\sqrt{n\sum u_{31}-(\sum u_3)^2}}$

Multiple Linear Regression

$Y= b_0 + b_1x_1 + b_2x_2 + e$

Estimation of coff. in multiple Linear Regression: $y= b_0 + b_1x_1 + b_2x_2 + e_i$

$\sum y= nb_0 + b_1\sum X_1 + b_2\sum X_2$, $\sum X_1 y = b_0\sum X_1 + b_1\sum X_1^2 + b_2\sum X_1X_3$

$\sum X_2 y = b_0\sum X_2 + b_1\sum X_1X_2 + b_2\sum X_2^2$ where $b_0=\frac{D_1}{D}$, $b_1=\frac{D_2}{D}$, $b_2=\frac{D_3}{D}$

ANOVA Table Of Regression Analysis

Source of Variation	df	SS	MSS	F. ratio
due to regression	K(no. inde va	SSR	MSR=SSR/K	
due to error	n-k-1	SSE	MSE= SSE/(n-k-1)	F= MSR/MSE
Total	n-1	TSS		

When Y is dependent , X_1 and X_2 independent

$TSS = \sum(Y-\bar{Y})^2 = \sum Y^2 - n\bar{Y}^2$

$SSE = \sum(Y-\hat{Y})^2 = \sum Y^2 - b_0\sum Y - b_1\sum YX_1 - b_2\sum YX_2$

SSR = TSS – SSE

When X_1 is dependent, X_2 and X_3 independent

$TSS = \sum(X_1-\bar{X}_1)^2 = \sum X_1^2 - n\bar{X}_1^2$

$SSE = \sum(X_1-\hat{X}_1)^2 = \sum X_1^2 - a\sum X_1 - b_2\sum X_1X_2 - b_3\sum X_1X_3$

SSR = TSS - SSE

When Y is dependent , X_1 and X_2 independent

$TSS = \sum(X_1-\bar{X}_2)^2 = \sum X_2^2 - n\bar{X}_2^2$

$SSE = \sum(X_1-\hat{X}_2)^2 = \sum X_2^2 - a\sum X_2 - b_2\sum X_1X_2 - b_3\sum X_2X_3$

SSR = TSS – SSE

When X_1 is dependent, X_2 and X_3 independent

$TSS = \sum(X_1-\bar{X}_3)^2 = \sum X_3^2 - n\bar{X}_3^2$

$SSE = \sum(X_1-\hat{X}_3)^2 = \sum X_3^2 - a\sum X_3 - b_2\sum X_1X_3 - b_3\sum X_2X_3$

SSR = TSS - SSE

Standard Error of the Estimation

$S_e=\sqrt{MSE} = \sqrt{\frac{SSE}{n-k-1}}$; = no. of independent variable in RM

Coefficient of Determination

$R^2_{adjusted}(\bar{R})^2= 1-\frac{(n-1)}{(n-k-1)}[1-R^2]$; $R^2=\frac{SSR}{TSS}$

Test of Significance for Regression Coefficients at α% level of significance:

Equation: $y = b_0 + b_1x_1 + b_2x_2$; **Test Statistics:** $t=\frac{b_1}{Sb_1}$; **Critical Value:** $t_{tabulated} = t_{\alpha/2(n-k-1)}$

Test of Overall Significance of the Regression Coefficients(independent variables):

Test Statistics $F=\frac{MSR}{MSE}$, $F=\frac{MSR}{MSE} = \frac{(n-k-1)}{k} * \frac{R^2}{1-R^2}$

ANOVA Table for regression analysis

Source of Variation	df	SS	MSS	F. ratio	$F_{tabulated}$
due to regression	K(no. inde va	SSR	MSR=$\frac{SSR}{K}$		
due to error	n-k-1	SSE	MSE= $\frac{SSE}{n-k-1}$	F= $\frac{MSR}{MSE}$	$F_{\alpha(k,n-k-1)}$
Total	n-1	TSS			

Non Parametric Test

One Sample Test: for sample (n_1,n_2≤20); Test Statistics: no. of runs(r), Critical value: $\bar{r}$ r_{-}

For sample size (n_1 or n_2 > 20): in case of large sample size is approximately normally distributed with mean $\mu_r = \frac{2n_1n_2}{n_1+n_2} + 1$ And variance $\sigma_r^2 = \frac{2n_1n_2(2n_1n_2-n_1-n_2)}{(n_1+n_2)2\ (n_1+n_2-1)}$ **Test Statistics:** $z = \frac{r-\mu_r}{\sigma_r} \sim N(0,1)$; Md=$\frac{(n+1)}{2}$ $th\ item$

Binomial Test: Small Sample Size(n≤25)TS: x_0=min{n1,n2}, CV:p=prob(X≤x_0)=$\sum_{x=0}^{x_0} c(n,x)p^x(1-p)^{n-x} \sum_{x=0}^{x_0} C(n, x)\left(\frac{1}{2}\right)^n$

Large Sample Size(n>25); test statistics $Z = \frac{(x_0 \pm 0.5)-np}{\sqrt{npq}}$use +0.5 if x_0<np & use -0.5 if x_0 >np

Kolmogorov Smirnov Test: TS: D_0=Max|F_e -F_0|; Decision: Reject H_0 if $D_0 \geq D_n$, accept otherwise.

Two Independent Sample Test: 1. Median Test; TS: $\frac{c(n_1,a)c((n_2,k-a)}{c(n_1+n_2,k)} a = 0,1,2 \ldots .\min(n_1,k) = \frac{n_1+n_2}{2} = \frac{n}{2}$

Large sample size(n_1>10, n_2 >10)

	No. of obs≤Md	No. of obs≤Md	Total
Sample x	a	C	a+c
Sample y	b	D	b+d
Total	a+b	c+d	N=a+b+c+d

Test Statistics:

$X^2 = \frac{N(ab-bc)^2}{(a+c)(b+d)(a+b)(c+d)} \sim x^2(1)$

if any cell frequency is less than 5 then

$X^2 = \frac{N(|ad-bc|-\frac{N}{2})^2}{(a+c)(b+d)(a+b)(c+d)} \sim X^2(1)$

Two Sample Kolmogorov Smirnov Test:Small Sample test(n_1=n_2<40,&n_2≤20for$n_1 \neq n_2$):TS:D_0=maximum{|F_x-F_y|}

Large Sample Test(n_1=n_2>40,&n_2>20for $n_1 \neq n_2$): Test Statistics; D_0=maximum{|F(x)-F(y)|} for two tail test

$X^2 = 4D_0^2\frac{n_1n_2}{n_1+n_2}$; Critical Value: $D_\alpha=1.36\sqrt{\frac{n_1+n_2}{n_1n_2}}$for two tail with α = 5%

Mann Whitey U Test: small sample size(n_1 ≤10,n_2≤10) TS: U_0=min{U_1, U_2}; CV: p = Prob(U≤U_0)

U_1= $n_1n_2 + \frac{n_1(n_1+1)}{2} - R_1$ and **U_2=** $n_1n_2 + \frac{n_2(n_2+1)}{2} - R_2$ such that $n_1n_2 = U_1U_2$

<u>Large sample size(n_1 >10,n_2>10)</u> variance $\sigma_u^2 = \frac{n_1n_2(n_1+n_2+1)}{12} = \frac{n_1n_2}{n(n-1)}\left\{\frac{n^3-n}{12} - \frac{\sum t_i^3-ti}{12}\right\}$,TS:**Z**=$\frac{U_0-\mu_\alpha}{\sigma_n} = \frac{U_0-\frac{n_1n_2}{2}}{\sqrt{\frac{n_1n_2(n_1+n_2+1}{12}}}$

Chi Square Test for Goodness of Fit: TS: $X^2=\sum_{i=0}^{k}\frac{(O_i-E_i)^2}{E_i} \sim X^2(k-1)$

Chi Square Test for Independence of Attributes: $X^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} \sim X^2_{(r-1)(c-1)}$ **E_{ij}=($O_{i.}$ * $O_{.j}$)/N**

Paired Sample Test:

1. **Wilcoxon Matched Pair Signed Rank Test:**

Small Sample size(n≤25):TS= min{S(+), S(-)}, Decision: Reject H_o at level of significance if T <= $T_{\alpha,n}$ accept otherwise.

Large Sample size n>25: $\mu_T = \frac{n(n+1)}{4}$ and $\sigma_T^2 = \frac{n(n+1)(2n+1)}{24}$ **Test Statistic : Z =** $\frac{T-\mu_T}{\sigma_T} = \frac{T-\frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}} \sim N(0,1)$

Cochran Q test:TS:Q=$\frac{(k-1)[K\sum R_i^2-(\sum R_i)^2]}{K\sum C_j-\sum C_j^2} \sim X^2$, **CV:** $X^2_{\alpha(k-1)}$;**Decision:**reject H_0atα% level of sign, if Q>$X^2_{\alpha(k-1)}$

Kruskal Wallis H Test: TS:

H=$\frac{12}{n(n+1)}\sum\frac{R_i^2}{n_i} - 3(n+1) \sim X^2(k-1)$,

if tied occurs the corrected test Statistics is

H=$\frac{\frac{12}{n(n+1)}\sum\frac{R_i^2}{n_i}-3(n+1)}{1-\sum\frac{t_i^3-t_i}{n^3-n}}$, t_i=no.of times i^{th} rank is repeated

Friedman F test:

H=$\frac{12}{nk(k+1)}\sum_{i=1}^{k} R_i^2 - 3n(k+1)$

if tied occurs then corrected test statistics is

H=$\frac{\frac{12}{nk(k+1)}\sum_{i=1}^{k} R_i^2-3n(k+1)}{1-\sum\frac{t_i^3-t_i}{n(k^3-k)}}$,$t_i$= number of times ith rank is repeated.

Design and Experiment

Completely randomized design:

S.V	d.f	S.S	M.S	F_{cal}	F_{tab}
Due to Treatment	t-1	SST	MST = $\frac{SST}{t-1}$	$F_T = \frac{MST}{MSE}$	$F_{\alpha\{(t-1),\, t(r-1)\}}$
Due to error	t(r-1)	SSE	MSE = $\frac{SSE}{t(r-1)}$		
Total	r t-1	TSS			

Calculation of completely randomized design: **SSE = TSS - SSE**

$$\text{TSS} = \sum_{i=1}^{t}\sum_{j=1}^{r} y_{ij}^2 - \frac{(T)^2}{n}\ ,\ \text{SST} = \frac{\sum_{i=1}^{t} Ti^2}{r} - CF\ ,\text{where CF} = \frac{(T_i)^2}{n}$$

Randomized block design:

S.V	d.f	S.S	M.S	F_{cal}	F_{tab}
Due to Treatment	t-1	SST	MST = $\frac{SST}{t-1}$	$F_T = \frac{MST}{MSE}$	$F_{\alpha\{(t-1),\, (t-1)(r-1)\}}$
Due to block	r-1	SSB	MSB = $\frac{SSB}{r-1}$	$F_B = \frac{MSB}{MSE}$	$F_{\alpha\{(r-1),\, (t-1)(r-1)\}}$
Due to error	(t-1) *(r-1)	SSE	MSE = $\frac{SSE}{(t-1)(r-1)}$		
Total	r t-1	TSS			

Calculation of randomized block design: **SSE = TSS – SST - SSB**

$$\text{TSS} = \sum_{i=1}^{t}\sum_{j=1}^{r} y_{ij}^2 - \frac{(T)^2}{n}\ ,\ \text{SST} = \frac{\sum_{i=1}^{t} Ti^2}{r} - CF,\text{where CF} = \frac{(T_i)^2}{n}\ ,\ \text{SSB} = \frac{\sum_{j=1}^{t} Tj^2}{r} - CF$$

Efficiency of RBD relative to CRD	Efficiency of LSD relative to CRD	Efficiency of LSD relative to RBD
$\frac{\delta_e'^2}{\delta_e^2} = \frac{r(t-1)*\text{MSE} + (r-1)*\text{MSB}}{(rt-1)*MSE}$	$\frac{\delta_e'^2}{\delta_e^2} = \frac{(m-1)*\text{MSE} + \text{MSR} + \text{MSC}}{(m+1)MSE}$	$\frac{\delta_e'^2}{\delta_e^2} = \frac{(m-1)*\text{MSE} + \text{MSR}}{m*MSE}$
$\frac{\delta_e'^2}{\delta_e^2} < 1$ => RBD is less efficient than CRD	$\frac{\delta_e'^2}{\delta_e^2} < 1$ => LSD is less efficient than CRD	$\frac{\delta_e'^2}{\delta_e^2} < 1$ => LSD is less efficient than RBD
$\frac{\delta_e'^2}{\delta_e^2} > 1$ => RBD is more efficient than CRD	$\frac{\delta_e'^2}{\delta_e^2} > 1$ => LSD is more efficient than CRD	$\frac{\delta_e'^2}{\delta_e^2} > 1$ => LSD is more efficient than RBD
$\frac{\delta_e'^2}{\delta_e^2} = 1$ => RBD and CRD are equally effective	$\frac{\delta_e'^2}{\delta_e^2} = 1$ => LSD and CRD are equally effective	$\frac{\delta_e'^2}{\delta_e^2} = 1$ => LSD and RBD are equally effective

Latin Square design:

Calculation of Latin square design: **SSE = TSS – SSR – SSC - SST**

$$\text{TSS} = \sum_{(i,j,k)} y_{ijk}2\ ,\ \text{SSR} = \frac{\sum_i T_{i..}^2}{m} - CF\ ,\text{where CF} = \frac{(T_i)^2}{n}\ ,\ \text{SSC} = \frac{\sum_j T_{.j.}^2}{m} - CF, \text{SST} = \frac{\sum_k T_{..k}^2}{m} - CF,$$

Reject H_{0R} at α% level of significance if $F_R > F_{\alpha\{(m-1),\, (m-1)(m-2)\}}$, accept otherwise.

Reject H_{0C} at α% level of significance if $F_C > F_{\alpha\{(m-1),\, (m-1)(m-2)\}}$, accept otherwise.

Reject H_{0T} at α% level of significance if $F_T > F_{\alpha\{(m-1),\, (m-1)(m-2)\}}$, accept otherwise.

S.V	d.f	S.S	M.S	F_{cal}	F_{tab}
Due to row	m-1	SSR	MSR = $\frac{SST}{m-1}$	$F_R = \frac{MSR}{MSE}$	$F_{\alpha\{(m-1), (m-1)(m-2)\}}$
Due to column	m-1	SSC	MSC = $\frac{SSC}{m-1}$	$F_C = \frac{MSC}{MSE}$	$F_{\alpha\{(m-1), (m-1)(m-2)\}}$
Due to treatment	m-1	SST	MST = $\frac{SST}{m-1}$	$F_T = \frac{MST}{MSE}$	$F_{\alpha\{(m-1), (m-1)(m-2)\}}$
Due to error	(m-1) *(m-2)	SSE	MSE = $\frac{SSE}{(m-1)(m-2)}$		
Total	$m^2 - 1$	TSS			

Stochastic Process

N step transition probability:

$$P_{ij}(n) = \sum_{k=1}^{m} \boldsymbol{P_{ik}(n-1)\, P_{kj}(1)}\ ,\ P_{ij}(2) = \sum_{k=1}^{m} \boldsymbol{P_{ik}\, P_{kj}}\ \&\ P_{ij}(3) = \sum_{k=1}^{m} \sum_{l=1}^{m} \boldsymbol{P_{jk}\, P_{kl}\, P_{lj}}$$

N step transition probability matrix:

$P^{(2)} = P * P$ & $P^{(3)} = P^{(2)} * P$ Markov Chain Steady State distribution : $\pi_x = \lim_{h\to 0} P_h(x)$.

When steady state distribution exists $\pi P = \pi$.

<u>Binomial Process:</u>

$\lambda = arrival\ rate\ (p/\Delta), \Delta =\ frame\ size, P = probability\ of\ success\ during\ one\ frame,$

$X\left(\frac{t}{\Delta}\right) =\ number\ of\ arrivals\ by\ time\ t, T =\ inter\ arrival\ time,$ n = t/Δ

$T = Y\Delta$, $E(T) = E(Y\Delta) \Rightarrow \Delta E(Y) \Rightarrow \frac{\Delta}{p} = 1/\lambda$, $V(T) = V(Y\Delta) \Rightarrow \Delta^2 V(Y) = (1-p)(\Delta/p)^2 \Rightarrow (1-p)/\lambda^2$

Sampling Distribution and Estimation

Statistic	Standard error
Mean (when σ known and population size infinite)	$S.E.(\bar{X}) = \frac{\sigma}{\sqrt{n}}$
Mean (when σ known and population size finite i.e. N)	$S.E.(\bar{X}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$
Mean (when σ unknown and population size infinite)	$S.E.(\bar{X}) = \frac{s}{\sqrt{n}}$
Mean (when σ unknown and population size finite i.e. N)	$S.E.(\bar{X}) = \frac{s}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$
Difference of means(when σ's are known)	$S.E.(\bar{X}_1 - \bar{X}_2) = \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$
Difference of means (when σ's are unknown)	$S.E.(\bar{X}_1 - \bar{X}_2) = \sqrt{\left(s^2\left\{\frac{1}{n_1} + \frac{1}{n_2}\right\}\right)}$
Proportion (when population size is infinite)	$S.E.(p) = \sqrt{\frac{PQ}{n}}$
Proportion (when population size is finite i.e. N)	$S.E.(p) = \sqrt{\frac{PQ}{n}}\sqrt{\frac{N-n}{N-1}}$
Difference of proportions	$S.E.(p_1 - p_2) = \sqrt{\left(PQ\left\{\frac{1}{n_1} + \frac{1}{n_2}\right\}\right)}$