Year: 2066

Group A

Attempt any two$(\mathbf{10 \times 2 = 20})$

Q.N.1) Describe Simple random sampling with and without replacement for drawing a random sample of size n from a population of size N. In both cases show that sample mean is unbiased estimate of population mean. Derive the variance of the sample mean in both cases. Show that the ratio of the variance of sample mean in sampling without replacement to that sampling with replacement is $\frac{N-n}{N-1}$. Comment on this result.

ANSWER:

If 1^{st} selected sample is return to population before the selection of 2nd sampling unit, then the sampling method is called simple random sampling with replacement (SRSWR).
If N be population size and n be sample size then there are N^n ways of selecting samples so each unit have equal probability of being selected as sample and is equal to $1/N^n$.

If 2^{nd} sampling unit is selected without replacing the 1^{st} unit selected as sample to the population, then the method of sampling is called simple random sampling without replacement (SRSWOR).
Here if N be population size and n be sample size then there are NC_n ways selecting n samples out of N population. Hence each population unit has equal chance of being selected as sample and is equals to $1/(NC_n)$

Sample mean is unbiased estimator of population mean

Proof:

In simple random sampling, the samples are drawn directly from population units. There may be two cases- SRSWOR and SRSWR. In SRSWOR, the sample unit occurred once is not replaced back in population before making the next draw, and in SRSWR, the sample unit once drawn is again placed on the population unit and then another draw is made. The SRS is useful for homogeneous collection of data.

i) Let $x_1, x_2, \dots, x_n$ be the sample observation drawn from the population units $X_1, X_2, \dots, X_N$ of size N.

Then

Sample Mean $(\overline{x}) = \left(\sum_{i=1}^{n} \frac{x_i}{n}\right)$ ……… (i)

Population Mean $\left(\overline{X}\right) = \frac{(X_1+X_2+\cdots+X_N)}{N} = \sum_{i=1}^{N} \frac{X_i}{N}$

Now taking expectations on both sides of equation (i)

$$E(\overline{x}) = E\left(\frac{1}{n}\sum_{i=1}^{n} x_i\right) \rightarrow E(\overline{x}) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) \quad \dots\dots\dots (ii)$$

Here, x_i is the ith term of sample unit which is drawn from the population unit X_i $(i = 1, 2, \dots, N)$ with probability of occurance $\frac{1}{N}$.

So, $E(x_i) = X_1\frac{1}{N} + X_2\frac{1}{N} + \cdots + X_N\frac{1}{N} = \sum_{i=1}^{N}\frac{X_i}{N} = \overline{X}\ (or\ \mu)$

Now from equation (ii);

$$E(\overline{x}) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \frac{1}{n}.n\mu = \mu$$

Thus sample mean is an unbiased estimator of population mean.

We know that,

$$V(\overline{x}) = E[\overline{x} - E(\overline{x})]^2 = E(\overline{x} - \mu)^2 = E\left(\frac{x_1 + x_2 + \cdots + x_n}{n} - \mu\right)^2$$

$$= \frac{1}{n^2}\ E[(x_1 - \mu) + (x_2 - \mu) + \cdots + (x_n - \mu)]^2 = \frac{1}{n^2}\ E\left[\sum_{i=1}^{n}(x_i - \mu)\right]^2$$

$$V(\overline{x}) = \frac{1}{n^2}\ E\left[\sum_{i=1}^{n}(x_i-\mu)^2 + 2\sum_{i<}^{n-1}\sum_{j}^{n}(x_i-\mu)(x_j-\mu)\right]\ \dots\dots\dots(i)$$

For SRSWR;

$$V(\overline{x}) = \frac{1}{n^2}\ E\left[\sum_{i=1}^{n}(x_i-\mu)^2\right] + 0$$

Here $(x_i-\mu)^2$ is i[th] sample unit expected to occur from i[th] population unit $(x_i-\mu)^2\ \forall\ i = 1, 2, \dots, N$ with the probability of occurance $\frac{1}{N}$, then

$$E(x_i-\mu)^2 = \frac{1}{N}\sum(x_i-\mu)^2 = \frac{1}{N}\sum\left(x_i-\overline{X}\right) = \sigma^2$$

$$V(\overline{x})_{SRSWR} = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \frac{1}{n^2}n\sigma^2 = \frac{\sigma^2}{n}$$

$$V(\overline{x})_{SRSWR} = \frac{N-1}{N}.\frac{s^2}{n}\ \ \dots\dots\dots(ii)$$

For SRSWOR;

From equation (i);

$$V(\overline{x})_{SRSWOR} = \frac{1}{n^2}\left[\sum_{i=1}^{n}E(x_i-\mu)^2 + 2\sum_{i<}^{n-1}\sum_{j}^{n}(x_i-\mu)(x_j-\mu)\right]$$

$(x_i-\mu)^2$ is the ith sample unit expected to occur from population unit $(x_i - \mu 2\ \forall\ i=1, 2, \dots, N$ with probability $1N$ to occur.

Then

$$E(x_i-\mu)^2 = \frac{1}{N}\sum_{i=1}^{n}(x_i-\mu) = \sigma^2$$

And $(x_i - \mu)(x_j - \mu)$ is the sample unit expected to occur from the population unit $(x_i - \mu)(x_j - \mu)$ $\forall\, i = 1, 2, \dots, N-1 \,;\, j = 1, 2, \dots, N$ with probability of occurance $\frac{1}{N}.\frac{1}{N-1}$

Then,

$$E\left[(x_i - \mu)(x_j - \mu)\right] = \frac{1}{N}.\frac{1}{N-1}\sum_{i<}^{N-1}\sum_{j}^{N}(x_i - \mu)(x_j - \mu)$$
$$= \frac{1}{N}.\frac{1}{N-1}\left(-\frac{N\sigma^2}{2}\right)$$

Therefore,

$$V(\overline{x})_{SRSWOR} = \frac{1}{n^2}\left[\sum_{i=1}^{n}\sigma^2 + 2\sum_{i<}^{n-1}\sum_{j}^{n}\frac{1}{N}.\frac{1}{N-1}\left(-\frac{N\sigma^2}{2}\right)\right] = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$
$$= \left(1 - \frac{n}{N}\right).\frac{s^2}{n}$$

Therfore

$$\frac{V(\bar{x})_{SRSWOR}}{V(\bar{x})_{SRSWR}} = \frac{\left(\left(1 - \frac{n}{N}\right)\frac{s^2}{n}\right)}{\frac{N-1}{N}\left(\frac{s^2}{n}\right)}$$
$$= \frac{N-n}{N} \times \frac{N}{N-1}$$
$$= \frac{N-n}{N-1}$$

Which is the required answer

Q.N.2)Describe the function and the procedure of Kruskal-Wallis one-way ANOVA test by ranks. The original data of three independent samples were collectively converted to ranks as shown in the adjacent table. Set up

appropriate null and alternative hypothesis and carry out the Kruskal-Wallis test at 5% level.

Sample 1	Sample 2	Sample 3
1	2	5
4	3	9
6	7	12
10	8	18
11	13	20
14	15	22
16	17	23
19	21	24

This test is alternative test of One-way ANOVA. Here we test the significance difference between three or more than three treatments classified as one way criteria.

The process of Kruskal-Wallis H test is as follow:

a. Set the Null and Alternative hypothesis.
b. Test Statistics: Under H_0, test statistics is

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

When ranks are repeated,

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

$T=t^3-t$; t = no. of times that rank is repeated.

c. Critical value:
   i. Small sample case: i.e. when k=3 & $n_1 \leq 5$
      Critical value is obtained from Kruskal Wallis table as: $P_0 = P(H \geq H_{cal})$
      Decision: If $P_0 < \alpha$, then H_0 is rejected.
   ii. Large sample case: i.e. when k>3 & $n_1 > 5$
      Then, $H \sim \chi^2_{k-1}$ obtained from chi-square table.

Decision: If $H \geq \chi^2_{k-1}$, then H_0 is rejected.

Here

H_0: $There\ is\ no\ significant\ difference\ between\ three\ sample$
H_1: $There\ is\ significant\ difference\ between\ three\ sample$

Test statistics: Under H_0

Here k=3, $n_1 = 8, n_2 = 8\ and\ n_3 = 8, n = n_1 + n_2 + n_3 = 24$

Kruskal wallis H is

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

$$= \frac{12}{24 \times 25}\left[\frac{(81)^2}{8} + \frac{(86)^2}{8} + \frac{(133)^2}{8} - 3(24+1)\right]$$
$$= 77.115$$

Q.N.3)In order to establish the functional relationship between annual salaries(Y), years of educated past high school,(X_1), and years of experience with the firm(X_2), data on these three variables were collected from a random sample of 10 persons working in a large firm. Analysis of data produces the following results.

a) **The total sum of squares, $\sum(Y_i - \bar{Y})^2$ is 397.6 and sum of squares due to error $\sum(Y_i - \hat{Y})^2$, is 23.5. Compute the value of R^2 and interpret the result. Also compute the value of F statistic for testing the significance of the model. Interpret the model and carry out the test or significance of the two slope regression coefficients at 5% level by stating the null and alternative hypothesis explicitly.**
$$\hat{Y} = -8.883 + 1.85X_1 + 2.92X_2$$
$$(4.94) \qquad (0.59) \qquad (0.61)$$

Group B

Answer any eight question:$(8 \times 5 = 40)$

Q.N.4) Show that sample variance in simple random sampling method is an unbiased estimator of population Variance.

Solution:

T0 proove; $E(S^2) = \sigma^2$

Where, $s^2 = \frac{1}{n-1} \sum(x_i - \bar{x})^2$ & $\sigma^2 = \frac{1}{N}\sum(X_i - \mu)^2, S^2 = \frac{1}{N-1}\sum(X_i - \mu)^2$

Proof ,

Here, $s^2 = \frac{1}{n-1} \sum(x_i - \bar{x})^2$

By taking expectation on both sides of above relation, we get

$$E(s^2) = \frac{1}{n-1}\, E\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]$$

$$= \frac{1}{n-1}\, E\left[\sum_{i=1}^{n}(x_i - \bar{X} + \bar{X} - \bar{x})^2\right]$$

$$= \frac{1}{n-1}\, E\left[\sum_{i=1}^{n}\{(x_i - \bar{X})^2 + 2(x_i - \bar{X}).(\bar{X} - \bar{x}) + \bar{X} - \bar{x})^2\}\right]$$

$$= \frac{1}{n-1}\, E\left[\sum_{i=1}^{n}(x_i - \bar{X})^2 + n(\bar{x}_i - \bar{X})^2 - 2(\bar{x} - \bar{X})\sum_{i=1}^{n}(x_i - \bar{X})\right]$$

$$= \frac{1}{n-1}\, E\left[\sum_{i=1}^{n}(x_i - \bar{X})^2 + n(\bar{x}_i - \bar{X})^2 - 2(\bar{x} - \bar{X})n(\bar{x} - \bar{X})\right]$$

$$= \frac{1}{n-1}\, E\left[\sum_{i=1}^{n}(x_i - \bar{X})^2 - n(\bar{x} - \bar{X})^2\right]$$

$$= \frac{1}{n-1} E\left[\sum_{i=1}^{n} \sigma^2 - nE(\bar{x} - E(\bar{x}))^2\right]$$

$$= \frac{1}{n-1}[n\sigma^2 - nv(\bar{x})_{wr}]$$

$$= \frac{1}{n-1}[n\sigma^2 - \frac{n\sigma^2}{n}$$

$$= \frac{\sigma^2(n-1)}{n-1}$$

$$= \sigma^2$$

$\boxed{\therefore E(s^2) = \sigma^2}$ .Hence , s^2 is an unbiased estimator of σ^2 in case of SRS.

Q.N.5) Write the sample multiple linear regression model of Y on X_1, X_2 *and* X_3 based on a sample of size n. What are the assumption to be made on this model for estimation and test of significance?

Q.N.6) Test whether the color of son's eyes is associated with that of the fathers at 5% level of significance using the data available in the following table.

Son's eye color			Row Total
Father's eye color	**No light**	**Light**	
No light	**230**	**148**	**378**
Light	**151**	**471**	**622**
Column Total	**381**	**619**	**1000**

H_o: ***color of son'sey is associatd with that of fathers***

H_1: **color of son's eye is not associated with that of fathers**

Now

$$correct\ \gamma^2 = \frac{n\left[|ad - bc| - \frac{n}{2}\right]^2}{(a+b)(a+c)(b+c)(b+d)}$$

$$= \frac{1000\left[230 \times 471 + 148 \times 151 - \frac{1000}{2}\right]^2}{(230+148)(230+151)(148+151)(148+471)}$$

= please calculate…

Q.N.7) Define the first order autocorrelation? Estimate the first order autocorrelation from the data available in the table below.

The first order coefficient of autocorrelation is given by

$$\rho = \frac{cov(e_t.e_{t-1})}{var(e_t)} = \sum_{i=2}^{n} e_t.e_{t-1}/(\sum_{i=1}^{n} e_{t^2})$$

Time(t)	1	2	3	4	5	6	7	8	9	10
Residuals($\hat{u}_i$)	-5	-4	-3	-2	-1	0	1	2	3	4

S.N.	e_t	e_{t-1}	e_{t-2}	$e_t.e_{t-1}$	$e_t.e_{t-2}$	e_t^2	$e_t - e_{t-1}$	$(e_t - e_{t-1})^2$
1	-5					25		
2	-4	-5		20		16	-1	1
3	-3	-4	-5	12	15	9	-1	1
4	-2	-3	-4	6	8	4	-1	1
5	-1	-2	-3	2	3	1	-1	1
6	0	-1	-2	0	0	0	-1	1
7	1	0	-1	1	-1	1	-1	1
8	2	1	0	2	0	4	-1	1
9	3	2	1	6	3	9	-1	1
10	4	3	2	12	8	16	-1	1
Total				$\sum e_t.e_{t-1} = 61$	$\sum e_t.e_{t(-2)} = 36$	$\sum e_t^2 = 85$		$\sum (e_t - e_{t-1})^2 = 9$

$$\rho = \frac{\sum_{t=2}^{10} e_t . e_{t-1}}{\sum_{t=1}^{10} e_t^2} = \frac{61}{85} \cong 0.72$$

Q.N.8) In a stratified sampling using simple random sampling without replacement method in each stratum, show that $Var\ (\bar{y}_{st}) = \sum \left(\frac{N_h}{N}\right)^2 \Big(1 -$ $nhNhSh2nh$ Simplify this expression when the total sample of size n is allocated according to proportional allocation across strata.

Stratified sampling is the process of dividing the population into number of sub-population called strata and from each strata a sample is drawn independently.

For stratified sampling;

$$\overline{y_{st}} = \sum_{h=1}^{L} \omega_h \overline{y_h}$$

Where, $\overline{y_{st}}$ = unbiased estimator of population mean

$\omega_h = \frac{N_h}{N}$

$\overline{y_h}$ = mean of the sample of hth strata

For the expression of $Var(\overline{y_{st}})$ adopting SRSWOR, we know that;

$$\overline{y_{st}} = \sum_{h=1}^{L} \omega_h \overline{y_h}$$

Taking variance on both sides,

$$V(\overline{y_{st}}) = V\left[\sum_{h=1}^{L} \omega_h \overline{y_h}\right] = \sum_{h=1}^{L} \omega_h^2\, V(\overline{y_h}) = \sum_{h=1}^{L} \frac{N_h^2}{N^2} . \frac{N_h - n_h}{N_h} . \frac{S_h^2}{n_h} \quad \ldots\ldots\ldots (i)$$

$$V(\overline{y_{st}}) = \frac{1}{N^2}\sum_{h=1}^{L}\frac{N_h(N_h - n_h)S_h^2}{n_h}$$

Again from equation (i);

$$V(\overline{y_{st}}) = \sum_{h=1}^{L}\frac{N_h^2}{N^2}\left(1 - \frac{n_h}{N_h}\right)\frac{S_h^2}{n_h}$$

$$V(\overline{y_{st}}) = \sum_{h=1}^{L}\omega_h^2(1 - f)\frac{S_h^2}{n_h}$$

Again from equation (i);

$$V(\overline{y_{st}}) = \sum_{h=1}^{L}\frac{N_h^2}{N^2}\left(\frac{N_h - n_h}{N_h.n_h}\right)S_h^2 = \sum_{h=1}^{L}\omega_h^2\left(\frac{1}{n_h} - \frac{1}{N_h}\right)S_h^2$$

$$\sum_{h=1}^{L}\frac{\omega_h^2 S_h^2}{n_h} - \sum_{h=1}^{L}\frac{\omega_h^2 S_h^2}{N_h} = \sum_{h=1}^{L}\frac{\omega_h^2 S_h^2}{n_h} - \sum_{h=1}^{L}\frac{N_h^2}{N^2}.\frac{S_h^2}{N_h} = \sum_{h=1}^{L}\frac{\omega_h^2 S_h^2}{n_h} - \sum_{h=1}^{L}\frac{\omega_h S_h^2}{N}$$

$$V(\overline{y_{st}}) = \sum_{h=1}^{L}\frac{\omega_h^2 S_h^2}{n_h} - \frac{1}{N}\sum_{h=1}^{L}\omega_h S_h^2 \quad \dots\dots\dots(ii)$$

For proportional allocation method;

$$n_h = \frac{n}{N}N_h$$

Now, from equation (ii);

$$V(\overline{y_{st}}) = \sum_{h=1}^{L}\frac{\omega_h^2 S_h^2}{n_h} - \frac{1}{N}\sum_{h=1}^{L}\omega_h S_h^2$$

$$V(\overline{y_{st}})_{prop} = \sum_{h=1}^{L}\frac{\omega_h^2 S_h^2}{\frac{n}{N}N_h} - \frac{1}{N}\sum_{h=1}^{L}\omega_h S_h^2$$

$$V(\overline{y_{st}})_{prop} = \frac{n}{N}\sum_{h=1}^{L}\frac{N_h^2}{N^2}.\frac{S_h^2}{N_h} - \frac{1}{N}\sum_{h=1}^{L}\omega_h S_h^2$$

$$V(\overline{y_{st}})_{prop} = \frac{1}{n}\sum_{h=1}^{L}\frac{N_h^2}{N}.S_h^2 - \frac{1}{N}\sum_{h=1}^{L}\omega_h S_h^2$$

$$V(\overline{y_{st}})_{prop} = \frac{1}{n}\sum_{h=1}^{L}\omega_h S_h^2 - \frac{1}{N}\sum_{h=1}^{L}\omega_h S_h^2$$

$$V(\overline{y_{st}})_{prop} = \left(\frac{1}{n} - \frac{1}{N}\right)\sum_{h=1}^{L}\omega_h S_h^2$$

Further continuing;

$$V(\overline{y_{st}})_{prop} = \left(1 - \frac{n}{N}\right).\frac{1}{n}\sum_{h=1}^{L}\omega_h S_h^2 = (1-f).\frac{1}{n}\sum_{h=1}^{L}\omega_h S_h^2$$

Q.N.9) To evaluate a speed reading course, a group of 10 subjects was asked to read two comparable articles – one before the course and one after the course. Their scores on reading test are follows.

Before course(X)	**57**	**80**	**64**	**7**	**90**	**59**	**76**	**98**	**70**	**83**
After Course (Y)	**63**	**90**	**62**	**7**	**95**	**58**	**80**	**99**	**75**	**94**

Test whether the course is beneficial using Wilcoxon Signed Rank test at 5% level of significance, given that $P(T^+ > 44|H_o) = 0.042$ where T^+ is the sum of the positive ranks of the difference $d_1 = (y_i - x_i)$.

Solution

Null Hypothesis(H_0): X and Y are equally effective
Alternative Hypothesis(H_1): X and Y are not equally effective.

Test statistic:- Under H_0, test statistic is obtained as follows:

Before Course(X_i)	After Course(Y_i)	$d_i = yi - x_i$	Ranking of d_i	(+) sign	(-) sign
57	63	6	3	3	
80	90	10	2	2	
64	62	-2	7		7
7	7	0	10	10	
90	95	5	4.5	4.5	
59	58	-1	8.5		8.5
76	80	+4	6	6	
98	99	+1	8.5	8.5	
70	75	+5	4.5	4.5	
83	94	+11	1	1	
				$\sum = 39.5$	$\sum 15.5$

Test statistic
$T_{cal} = min\{S(+), S(-)\}$
$= min\{39.5,24.5\}$
=24.5

Critical Value: At 5% level of significance P($T^+ < 44$) therefore
H_1 *is accepted*

Q.N.11) Describe the function and procedure of the median test.

Median test is non parametric test used to test the difference in medians of two independent distribution. In other words this test is commended to testing whether two independent groups has been drawn from similar population or from different with same median or not. Also it is used to test whether two treatment applied is an experiment are equally effective or not.

Process for testing median test is.

i) Null Hypothesis H_0: $M_{d1}=M_{d2}$

Alternative Hypothesis H_1: $M_{d1} \neq M_{d2}$ (for two tailed test)

$M_{d1}>M_{d2}$(Right tailed test)

$M_{d1}<M_{d2}$(left tailed test)

Test statistic:

The test statistic under H_0,

i) combined the observation such that $n_2+n_2=n$

ii)compute the median of pooled data.

iii) find the no. of observation in the 1st sample which are less or equal to the median & denoted by a, which is the test statistic.

Critical value:

P-value $=P_0 = P(A \leq a) = \frac{\binom{n_1}{a}\binom{n_2}{k-a}}{\binom{n_1+n_2}{k}}$

Where $k = \frac{n_1+n_2}{2} = \frac{n}{2}$

Decision: if $P_0 > \alpha, then\ we\ accept\ H_0\ for\ one\ tail.$

$2P_0 > \alpha, then\ we\ accept\ H_0\ for\ two\ tail.$

Q.N.12) Define the problem of multicollinearity in a multiple regression model. How do you detect it and correct it?

If the model has several variables, it is likely that some of the explanatory variables will be approximately related. This property is known as multicollinearity. Multicollinearity refers to the situation where there is either an exact or approximately exact linear relationship among the regression.

The situation where the problem of multicollinearity arises are:

1. If the coefficient of multiple determination(R^2) is very high but the test of individual regression coefficient found to be insignificant then there could be the problem of multicollinearity.
2. The high value of standard error associated with individual regression coefficient also indicate the presence of multicollinearity.
3. If inclusion of additional variables make substantial change in the value of individual regression coefficients then the problem of multicollinearity may be present.

However, all above methods just give the tentative idea about the existence of multicollinear problem but not sufficient indicators.

Q.N.13) Write short notes on any two of the followings

Systematic Sampling

Systematic Sampling is the slight variance of the simple random sampling in which only the first sample unit is selected at random and the remaining units are automatically selected in a definite sequence at equal spacing from one another. This technique of drawing samples is usually recommended if the complete and up-to-date list of the sampling units i.e., the frame is available and the units are arranged in some systematic order such as alphabetical, chronological, geographical order, etc. this requires the sampling units in the population to be ordered in such a way that each item in the population is uniquely identified by its order, for examples the names of persons in a telephone directory, list of voters etc.

Let us suppose that N sampling units in the population are arranged in some systematic order and serially numbered from 1 to N and we want to draw a sample of size n form it such that

$$\square = \square\square \quad => \quad \square = \frac{\square}{\square}$$

Where k is usually called sample interval.

Order Statistics

In statistics, the kth order statistic of a statistical sample is equal to its kth-smallest value. Together with rank statistics, order statistics are among the most fundamental tools in non-parametric statistics and inference.

Important special cases of the order statistics are the minimum and maximum value of a sample, and (with some qualifications discussed below) the sample median and other sample quantiles.

When using probability theory to analyze order statistics of random samples from a continuous distribution, the cumulative distribution function is used to reduce the analysis to the case of order statistics of the uniform distribution.

Heteroscedasticity

The assumption of homoscedasticity demands for equal variance of each distribution term. If this assumption is not fulfilled, then we say that there is presence of heteroscedasticity. The problem of heteroscedasticity is likely to be more common in cross-sectional data than in time series data.

Consequences of heteroscedasticity

If heteroscedasticity among the stochastic disturbance term in a regression model is ignored and LS procedures is used to estimate the parameters then the following conditions occur.

(1) The estimates and forecast based on them will be consistent and unbiased.
(2) Forecast will be inefficient.

(3) The estimated variance and covariance will be biased and hence test of hypothesis are invalid.