Assignment - 02

Q1. Explain the following terms.

i) Null hypothesis - Null hypothesis is formulated for the purpose of its rejection under the assumption that it is true and denoted by H_0.

$$H_0 : \mu_1 = \mu_2$$

H_0 = the null hypothesis,
μ_1 = the mean of population 1
μ_2 = the mean of population 2

ii) Type I and type 2 error -
When we test a statistic confidence level, there are chances of taking wrong decisions due to small sample size or sampling fluctuations etc.

Type I Error is the incorrect rejection of a true null hypothesis, i.e. we reject H_0, when it is true, whereas Type II error is the incorrect acceptance of a false null hypothesis, i.e. we accept H_0 when its false.

iii) Confidence Limits.
Confidence limits show how accurate an estimation of the mean is or is likely to be. Confidence limits are the lowest and the highest numbers at the end of a confidence interval.

(iv) Alternative hypothesis -
Alternative hypothesis which is complementary to the null hypothesis called as "H_1" and is denoted by "H_1"

$$H_1 : \mu_1 \neq \mu_2$$

H_1 = Alternative hypothesis
μ_1 = the mean of population 1.
μ_2 = the mean of poplation 2.

v) Significance level -

The significance level of an event (such as statistical test) is the probability that the event could have occurred by chance. If the level is quite low, that is the probability of occurring by chance is quite small we say the event is significant.

The level of significant is the measurement of the statistical significance. It defines whether the null hypothesis is assumed to be accepted or rejected. It is expected to identify if the result is statistically significant for the null hypothesis to be false or rejected

$\alpha = 5\%$ $\quad a = 1\%$ $\quad a = 0.21\%$

Q2. A machinist is making engine parts with axle diameter of 0.7 inch. A random sample of 10 parts shows mean diameter 0.742 inch with a S.D of 0.04 inch. Based on this sample, would you say that the axle is inferior.

Given $n = 10 < 30$ The sample is small sample $\therefore$ we use t test

$\mu = 0.700$

$\bar{x} = 0.742$

$s = 0.040$

Null hypothesis (H_0)

$H_0 : \mu = 0.700$

Alternative hypothesis (H_1)

$H_1 : \mu \neq 0.7$

Level of significance :

$\alpha = 5\%$ $\quad d.f = n-1 = 10-1 = 9.$

Critical Region ~~10%~~ $= 2.262$

Test statistics, $t = \dfrac{\bar{x} - \mu}{\left(\frac{s}{\sqrt{n-1}}\right)} = \dfrac{0.742 - 0.7}{\left(\frac{0.040}{\sqrt{9}}\right)}$

$= 3.1500$

Conclusion.

If $-t_{\alpha/2} < t < t_{\alpha/2}$ then we accept H_0

$-2.262 < 3.15 \nless 2.262$

So we reject H_0.

$\therefore$ The product is not meeting specification.

Q3. A manufacture clamied that at least 95% Equipements which he supplied to a factory conformed to specifications. An examination of a sample of 200 pieces of equipement revealed that 18 of them were faulty. Test his claim at significance level of 1% & 5%.

<u>Given</u> $n = 200$

$H_0 : P = 95\% = 0.95$

$H_1 : P \neq 0.95$

The level of significance $\alpha = 0.05 \ (5\%)$

The test statistic $Z = \dfrac{p - P}{\sqrt{\dfrac{PQ}{n}}}$ where $P + Q = 1 \Rightarrow Q = 1 - P$

$Q = 1 - 0.95 = 0.05$

$\therefore$ Total defective less products (non defective)

$= 200 - 18 = 182.$

$$p = \frac{182}{200} = 0.91$$

$$Z = \frac{0.91 - 0.95}{\sqrt{\dfrac{0.95 \times 0.05}{200}}} = \frac{0.04}{\sqrt{0.0002375}} = -2.5955$$

At 5% level, the tabulated value of Z is 1.96

Since $|Z| = 2.5955 > 1.96$.

Hence, the null hypothesis is rejected at 5% level of significance.

4. Genetic theory states that the Children having one parent of blood type M and other blood type N will be always one of the 3 type M, MN, N and that the proportions of these type will on an avg be 1:2:1. A report states that out of 300 children have one M parent and other N parent, 30% were found to be of type M 45% of type MN and remaining of type N. Test the theory by χ^2 (chi square test)

Given M = 30% MN = 45% N = 25%

corresponding frequencies

$\frac{30}{100} \times 300$ $\frac{45}{100} \times 300$ $\frac{25}{100} \times 300$ total $\underline{300}$

$= 90$ $= 135$ $= 75$

Proportions of this types are 1:2:1

we have $\chi^2 = \sum_{1}^{n} \frac{(O_i - E_i)^2}{E_i} = \frac{(90-75)^2}{75} + \frac{(135-150)^2}{150}$

$+ \frac{(75-75)^2}{75} = \underline{\underline{4.5}}$

$\chi^2_{0.05} = 5.99$ at dof = 2   d.f = n-1 = 3-1 = 2

hence Calculated value < theoritical value

$4.5 < 5.99$

$\therefore$ The Hypothesis is accepted

---

05 State Central Limit theory. Use the theorm to evaluate $P[50 < \bar{x} < 56]$ where $\bar{x}$ represents the mean of random sample of size 100 from an infinite population with mean $\mu = 53$ and variance $\sigma^2 = 400$.

The Central limit Theory states that the samp[le] mean $\bar{x}$ follows approximately the normal distrib[ution] with mean μ and standard deviation $\frac{\sigma}{\sqrt{n}}$ (is also called standard error), i.e., $\bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ where μ, σ are mean and standard deviation of the population from where the sample.

Given,

sample size $n = 100$

$\mu = 53$

$\sigma^2 = 400 \Rightarrow \sigma = \sqrt{400} = 20$

$\therefore \bar{x} \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$

$\Rightarrow \bar{x} \sim N\left(53, \frac{20}{\sqrt{100}}\right)$

$\Rightarrow \bar{x} \sim N(53, 2)$

$\therefore$ we know that $Z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$\Rightarrow Z = \frac{\bar{x} - 53}{2}$

$\therefore$ At $\bar{x} = 50 \Rightarrow Z = \frac{50-53}{2} = \frac{-3}{2} = -1.5 = Z_1$

At $\bar{x} = 56 \Rightarrow Z = \frac{56-53}{2} = \frac{3}{2} = 1.5 = Z_2$

$\therefore P(50 < \bar{x} < 56) = P(-1.5 < Z < 1.5)$

$= 2P(0 < Z < 1.5)$

$= 2A(1.5)$

$= 2 \times 0.4332$

$\therefore P(50 < \bar{x} < 56) = 0.8664$

A random Sample of Size 25 from a normal distr-ution $N(\mu, \sigma^2 = 4)$ yield; Sample mean $\bar{x} = 78.3$. obtain a 99 % Confidence interval for μ.

Given the sample size $n = 25$

$\bar{x} = 78.3$

$\sigma = 2$

we know, confidence level of 99 %. the corresponding z value is 2.58. This is determined from the normal distribution table.

Confidence interval $C.I = \mu = \text{Mean} \pm z = \bar{x} \pm z \frac{\sigma}{\sqrt{n}}$

$C.I = \mu = 78.3 \pm \left(2.58 \times \frac{2}{\sqrt{25}}\right)$

$\mu = 78.3 \pm 1032$

$C.I \Rightarrow (78.3 - 1.032,\ 78.3 + 1.032)$

$= (77.268,\ 79.332)$

Q7. Let the Observed value of the mean $\bar{x}$ of a random sample of size 20 from a normal distribution with mean μ and variance $\sigma^2 = 80$ be 81.2. Find a 90 % and 95 % Confidence intervals for μ.

Given, $n = 20$

$\bar{x} = 81.2$

$\sigma^2 = 80 = \sqrt{80} = 8.9442.$

we know, confidence level of 95 %, 90 %. the corresponding z values are 1.96, 1.642. This is determined from the normal distribution table

for 95 %,

$CI = \mu = 81.2 \pm \left(1.96 \times \frac{8.9442}{\sqrt{20}}\right)$

$\mu = 81.2 \pm 3.92$

$CI = (81.2 - 3.928,\ 81.2 + 3.92)$

$(77.28,\ 85.12)$

for 90%

$\therefore$ C.I : $\mu = 81.2 \pm \left(1.645 \times \frac{8.9992}{\sqrt{20}}\right)$

$= \mu = 81.2 \pm 3.29.$

$= CI = (81.2 - 3.29,\ 81.2 + 3.29) =$

$= (77.91,\ 84.49)$

Q8. In a recent study reported on the Flurry Blog, the mean age of tablet users is 34 years. Suppose the standard deviation is 15 years. Take a sample of size n = 100. Using central limit theory, find the probability that the sample mean age is more than 30 years.

Sol Given, $\mu = 34$

$\sigma = 15.$

$n = 100$

$$Z = \frac{x - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$Z = \frac{\bar{x} - 34}{\frac{15}{\sqrt{100}}} = \frac{\bar{x} - 34}{1.5}$$

$P(\bar{X} > 30)$

At $\bar{x} = 30$

$$Z = \frac{30 - 34}{1.5}$$

$Z = -2.66.$

$P(\bar{x} > 30) = P(Z > -2.66)$

$= 0.5 + \phi(2.66)$

$= 0.5 + 0.4962$

$P(Z > -a)$

$= 0.5 + \phi(a)$

$= 0.5 + \phi(2.66)$

$P(\bar{x} > 30) = 0.9962$

Three different kinds of food are tested on three group of rats for 5 weeks. The objective is to check the difference in mean weight (in grams) of the rats per week. Apply one-way ANOVA using a 0.05 significance level to following data.

Food 1	8	12	19	8	6	11
Food 2	4	5	4	6	9	7
Food 3	11	8	7	13	7	9

							Total	Square
F1	8	12	19	8	6	11	T_1 64	T_1^2 4096
F2	4	5	4	6	9	7	T_2 35	T_2^2 1225
F3	11	8	7	13	7	9	T_3 55	T_3^2 3035
							154	

							Sum of sq
F_1	64	144	361	64	36	121	790
F_2	16	25	16	36 ~~144~~	81	49	223
F_3	121	64	49	169	49	81	533
							1546

Grand total $- \Sigma_i \Sigma_j x_{ij}^2$

Set the null hypothesis. $H_0 : \mu_1 = \mu_2 = \mu_3$

Correction Factor $\underline{\underline{CF}} = \frac{T^2}{N} = \frac{(154)^2}{18} = \frac{23716}{18} = 1317.55$

Total sum of squares $TSS = \Sigma_i \Sigma_j x_{ij}^2 - CF$

$= TSS = 1546 - 1317.55$

$TSS = 228.45$

Sum of the square of between the treatments.

$SST = \Sigma_i \frac{T_i^2}{n_i} - CF$

$SST = \frac{4096}{6} + \frac{1225}{6} + \frac{3035}{6} - 1317.55$

$= 1391 - 1317.55 = \underline{73.4500}$

Sum of square due to Error SEE = TSS − SST

$\therefore SSE = 228.95 - 73.95 = 155$

$\therefore SSE = 155$

Sources variation	df	SS	MSS	F ratio
Between treatments	3-1 = 2	SST = 73.95	$MST = \frac{73.95}{2}$ = 36.725	$F = \frac{36.725}{10.33}$ = 3.55.
Error	~~18-3~~ = 18-3 = 15	SSE = 155	$MSE = \frac{155}{15} = 10.33$	
Total	18-1 = 17	-	-	

Q10. Set up an analysis of variance table for the following per acer production data for 3 varieties of wheat, each grow on 4 plots and state if the variety difference are significant at 5% sig level

Plot of Land

	A	B	C
1	6	5	5
2	7	5	4
3	3	3	3
4	8	7	4

Sol to carry out the analysis of variance, we for

Plot of Land	A	B	C	T	T^2
1	6	5	5	16	256
2	7	5	4	16	256
3	3	3	3	9	81
4	8	7	4	19	361
P	P 24 - P^2 576	P 20 - P^2 400	16 - 256	60	-

The Square are as follows
Varience

A	B	C	Grand Total
36	25	25	86
49	25	16	90
9	9	9	27
64	49	16.	129.
		=	332

$\Sigma_i \Sigma_j x_{ij}^2 = 332$

Set the null hypothesis $H_0 : \mu_1 = \mu_2 = \mu_3$, $N = 12$

Correction Factor $CF = \frac{T^2}{N} = \frac{(60)^2}{12} = \frac{3600}{12} = 300.$

Total sum of square $TSS = \Sigma_i \Sigma_j x_{ij}^2 - CF$

$332 - 300$

$TSS. = 32$

Sum of the row square $SSR = \Sigma_i \frac{T_i^2}{n_i} - CF$

$SSR = \frac{256}{3} + \frac{256}{3} + \frac{81}{3} + \frac{361}{3} - 300.$

$SSR = 318 - 300.$

$SSR = 18.$

Sum of the coloum square $SSC = \Sigma_j \frac{T_j^2}{n_j} - CF$

$SSC = \frac{576}{4} + \frac{400}{4} + \frac{256}{4} + \frac{60}{4} - 300.$

$SSC = 308 - 300.$

$SSC = 8.$

$\therefore SSE = TSS - SSR - SSC$

$SSE = 32 - 18 - 8 = 16.$

Source Variation	df	SS	MSS	F Ratio
Rows	$4-1=3$	$SSR=18$	$MSR=\frac{13}{3}=6$	$F_r=\frac{6}{1}=6.$
~~to~~ Coloums	$3-1=2$	$SSC=8$	$MSC=\frac{8}{2}=4$	$F_c=\frac{4}{1}=4$
Error	$3\times2=6$	$SSE=6$	$MSE=\frac{6}{6}=1$	
Total	$12-1=1$			

$F_r = 6 > F(3,6) = 4.76$

$F_c = 4 < F(6,2) = 19.33.$