1. A sample of 900 members has a mean of 3.4 cm and S.D. 2.61 cm. Is sample from a large population of mean 3.25 cm and S.D. 2.61 cm? Is the population normal and if its mean is unknown, find the 95% fiducial limits of true mean.

Solution:-

Given,

Mean of the population, $\mu = 3.25$, Standard deviation of the population, $\sigma = 2.61$,
Sample size, n = 900, Sample mean, $\bar{x} = 3.4$

Null hypothesis: $H_0: \mu = 3.25$ **(No difference between $\bar{x}$ and μ)**
Alternative hypothesis: $H_1: \mu \neq 3.25$(there is difference) --- **Two Tailed Test**

Level of significance: $\alpha = 0.05$ (5%) Z_{tab}=1.96

Test Statistic: $Z_{cal} = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} = \frac{3.4-3.25}{2.61/\sqrt{900}} = 1.724$

Conclusion: Here, $|Z_{cal}| < |Z_{tab}|$ (1.724 <1.96)
$\therefore H_0$ is accepted
$\therefore$ There is no difference between $\bar{x}$ and μ.

And 95% confidence limits are $\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}}$

$$= 3.4 \pm 1.96\frac{2.61}{\sqrt{900}} = 3.4 \pm 0.1705$$
$$= 3.2295 \text{ and } 3.57$$

2. A soft-drink machine is regulated so that it discharges an average of 200 millilitres per cups. If the amount of drink is normally distributed with a standard deviation equal to 15 millilitres,

a) What fraction of cups will have more than 224 millilitres?
b) What is the probably that a cup contains between 191 and 209 millilitres?
c) How many cups will probably overflow if 230 millilitres cups are used for the next 1000 drinks?

Solution:-

Let X represent the amount of drink distributed.

$$\mu = 200; \qquad \sigma = 15$$

(a) The fraction of the cups will contain more than 224 millilitres.

$$P(X > 224) = 1 - P(X < 224)$$
$$= 1 - P\left(Z < \frac{224-200}{15}\right)$$
$$= 1 - P(Z < 1.6)$$
$$= 1 - 0.9452$$
$$= 0.0548$$

(b) The probability that a cup contains between 191 and 209 millilitres.

$$P(191 < X < 209) = P(X < 209) - P(X < 191)$$
$$= P\left(Z < \frac{209-200}{15}\right) - P\left(Z < \frac{191-200}{15}\right)$$
$$= P(Z < 0.6) - P(Z < -0.6)$$
$$= 0.7257 - 0.2743$$
$$= 0.4514$$

(c) How many cups will probably overflow if 230 millilitres cups are used for the next 1000 drinks?

$$P(X > 230) = 1 - P(X < 230)$$
$$= 1 - P\left(Z < \frac{230-200}{15}\right)$$
$$= 1 - P(Z < 2)$$
$$= 1 - 0.9772$$
$$= 0.0228$$

Hence, using the binomial property, we get:

$$E(X) = n.p = 1000 * 0.0228 = 22.8 \approx 23$$

3. The bank has a head office in Delhi and a branch in Mumbai. There are long customer queues at one office, while customer

queues are short at the other. The Operations Manager of the bank wonders if the customers at one branch are more variable than the number of customers at another. He carries out a research study of customers. The variance of Delhi head office customers is 31, and that for the Mumbai branch is 20. The sample size for the Delhi head office is 11, and that for the Mumbai branch is 21. Carry out a two-tailed F-test with a level of significance of 10%.

Solution:-

- **Step 1:** Null Hypothesis H_0: $\sigma_1^2 = \sigma_2^2$

Alternate Hypothesis H_a: $\sigma_1^2 \neq \sigma_2^2$

- **Step 2:** F statistic = F Value = σ_1^2 / σ_2^2 = 31/20 = **1.55**
- **Step 3:** $df_1 = n_1 - 1$ = 11-1 = 10

$df_2 = n_2 - 1$ = 21-1 = 20

- **Step 4:** Since it is a two-tailed test, alpha level = 0.10/2 = 0.05. The F value from the F Table with degrees of freedom as 10 and 20 is 2.348.
- **Step 5:** Since the F statistic (1.55) is lesser than the table value obtained (2.348), we cannot reject the null hypothesis.

4

The average weight of a dumbbell in a gym is 90lbs. However, a physical trainer believes that the average weight might be higher. A random sample of 5 dumbbells with an average weight of 110lbs and a standard deviation of 18lbs. Using hypothesis testing check if the physical trainer's claim can be supported for a 95% confidence level.

Solution: As the sample size is lesser than 30, the t-test is used.

$H_0: \mu = 90$, $H_1: \mu > 90$

$\overline{x} = 110$, $\mu = 90$, $n = 5$, $s = 18$.

$\alpha = 0.05$

Using the t-distribution table, the critical value is 2.132

$$t = \frac{\overline{x} - \mu}{\frac{s}{\sqrt{n}}}$$

$t = 2.484$

As 2.484 > 2.132, the null hypothesis is rejected.

Answer: The average weight of the dumbbells may be greater than 90lbs

5

In the population, the average IQ is 100 with a standard deviation of 15. A team of scientists want to test a new medication to see if it has either a positive or negative effect on intelligence, or not effect at all. A sample of 30 participants who have taken the medication has a mean of 140. Did the medication affect intelligence?

Step 1: Set up the null and alternate hypothesis

H_0: medication affects intelligence
H_a: medication does not affect intellignece

(not that the alternate hypothesis is always the opposite of the null hypothesis)

Step 2: Determine the type of test to use

Since the sample size is 30, we use the z-test. See why we use the z-test when sample size is 30 and above

$$z = \frac{\bar{x}_n - \mu_0}{\sigma}\sqrt{n}$$

Using the data given in the equation we would have the following:

$\mu_0 = 100$
$\sigma = 15$
$n = 30$
$\bar{x}_n = 140$

Plugging the values into the formular we have:

$$z = \frac{140 - 100}{15}\sqrt{30} = 14.606$$

Step 5: Draw a conclusion

In this case the tested statistic value of z calculated is more than the critical value obtained from statistical tables.

14.606 > 1.96

Therefore we reject the null hypothesis.
This means, from the question, that the medication administered does not affect intelligence.

6

A) Find the critical value for a left tailed z test where $\alpha = 0.012$.

Solution: First subtract α from 0.5. Thus, $0.5 - 0.012 = 0.488$.

Using the z distribution table, $z = 2.26$.

However, as this is a left-tailed z test thus, $z = -2.26$

Answer: Critical value = -2.26

B) Find the critical value for a two-tailed f test conducted on the following samples at a $\alpha = 0.025$, Variance = 110, Sample size = 41, Variance = 70, Sample size = 21

Solution: $n_1 = 41$, $n_2 = 21$,

$n_1 - 1 = 40$, $n_2 - 1 = 20$,

Sample 1 df = 40, Sample 2 df = 20

Using the F distribution table for $\alpha = 0.025$, the value at the intersection of the 40^{th} column and 20^{th} row is

$F(40, 20) = 2.287$

Answer: Critical Value = 2.287

7. Based on field experiments, a new variety of green gram is expected to given a yield of 12.0 quintals per hectare. The variety was tested on 10 randomly selected farmer's fields. The yield (quintals/hectare) were recorded as 14.3,12.6,13.7,10.9,13.7,12.0,11.4,12.0,12.6,13.1. Do the results conform to the expectation?

Solution

Null hypothesis H_0: $\mu=12.0$

(i.e) the average yield of the new variety of green gram is 12.0 quintals/hectare.

Alternative Hypothesis: H_1:$\mu \neq 12.0$

(i.e) the average yield is not 12.0 quintals/hectare, it may be less or more than 12 quintals / hectare

Level of significance: 5 %

Test statistic:

$$t = \left| \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \right|$$

From the given data

$$\sum x = 126.3 \qquad \sum x^2 = 1605.77$$

$$\bar{x} = \frac{\sum x}{n} = \frac{126.3}{10} = 12.63$$

$$s = \sqrt{\frac{\sum x^2 - \frac{\left(\sum x\right)^2}{n}}{n-1}} = \sqrt{\frac{1605.77 - 1595.169}{9}} = \sqrt{\frac{10.601}{9}}$$

$$= 1.0853$$

$$\frac{s}{\sqrt{n}} = \frac{1.0853}{\sqrt{10}} = 0.3432$$

$$\text{Now} \quad t = \left| \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} \right|$$

$$t = \frac{12.63 - 12}{0.3432} = 1.836$$

Table value for t corresponding to 5% level of significance and 9 d.f. is 2.262 (two tailed test)

Inference

$t < t_{tab}$

We accept the null hypothesis H_0

We conclude that the new variety of green gram will give an average yield of 12 quintals/hectare.

8 Calculate a t-test for the following data of the number of times people prefer coffee or tea in five time intervals.

Coffee	Tea
4	3
5	8
7	6
6	4
9	7

Solution: let x_1 be the sample of data that prefers coffee and x_2 be the sample of data that prefers tea.

let us find the mean, variance and the SD

x_1	$(x_1 - \overline{x_1})$	$(x_1 - \overline{x_1})^2$	x_2	$(x_2 - \overline{x_2})$	$(x_2 - \overline{x_2})^2$
4	-2.2	4.84	3	-2.6	6.76
5	-1.2	1.44	8	2.4	5.76
7	0.8	0.64	6	0.4	0.16
6	-0.2	0.04	4	-1.6	2.56
9	2.8	7.84	7	1.4	1.96
6.2		**14.8**	**5.6**		**17.20**

$\overline{x_1} = 31/5 = 6.2$

$\overline{x_2} = 28/5 = 5.6$

$\Sigma(x_1-\overline{x_1})^2 = 14.8$

$\Sigma(x_2-\overline{x_2})^2 = 17.2$

$S_1 = 14.8/4 = 3.7$

$S_2 = 17.2/4 = 4.3$

According to the t-test formula,

$$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)}}$$

Applying the known values in the t-test formula, we get

$$t = \frac{6.2 - 5.6}{\sqrt{\left(\frac{3.7}{5} + \frac{4.3}{5}\right)}}$$

$$= \frac{0.6}{\sqrt{1.6}} = 0.6/1.26 = 0.47$$

9

Suppose we want to know if gender has anything to do with political party preference. So, we poll 440 voters in a simple random sample to find out which political party they prefer. The results of the survey are provided in the table below.

–	Republican	Democrat	Independent
Male	100	70	30
Female	140	60	20

Solution:

First we define the cases for different hypothesis such as,

H_0 = There is no link between gender and political party preference.

H_1 = There is a link between gender and political party preference.

Now we calculate the expected frequency.

$$Expected\ Value = \frac{(Row\ Total) \times (Column\ Total)}{Total\ Number\ of\ Observations}$$

For one case such as,

Expected Value of Male Republican = $\frac{240 \times 200}{440} = 109$

Similarly, we calculate the expected value for each of the cells such as,

–	Republican	Democrat	Independent	Total
Male	109	59	22.72	200
Female	120	65	25	220
Total	240	130	50	440

Expected Value

Now we use the formula latex] \frac{\left(O_i-E_i\right)^2}{E_i} [/latex] for each cell.

Where, O = Observed Value, and

E = Expected Value.

–	Republican	Democrat	Independent	Total
Male	0.7431197	2.050847	2.332676056	200
Female	3.3333333	0.384615	1	220
Total	240	130	50	440

Finally we calculate the test statistics X^2 which is the sum of the cell values from the above table.

$$X^2 = 0.743 + 2.05 + 2.33 + 3.33 + 0.384 + 1 = 9.837$$

Next we have to calculate the degrees of freedom such as,

$$(r-1) \times (c-1) = (3-1) \times (2-1) = 2$$

Where, r = number of column items, and

c = number of row items.

for an alpha level of 0.05 and 2 degrees of freedom, the critical statistic shown is 5.991, which is less than our obtained statistics of 9.83. Therefore we can reject the null hypothesis because the critical statistics is higher than the obtained statistic.

10. The average score of a class is 90. However, a teacher believes that the average score might be lower. The scores of 6 students were randomly measured. The mean was 82 with a standard deviation of 18. With a 0.05 significance level use hypothesis testing to check if this claim is true.

Solution: The t test will be used.

H_0: $\mu = 90$, H_1: $\mu < 90$

$\overline{x} = 110$, $\mu = 90$, $n = 6$, $s = 18$

The critical value from the t table is -2.015

$$t = \frac{\overline{x} - \mu}{\frac{s}{\sqrt{n}}}$$

$$t = \frac{82 - 90}{\frac{18}{\sqrt{6}}}$$

$t = -1.088$

As $-1.088 > -2.015$, we fail to reject the null hypothesis.

Answer: There is not enough evidence to support the claim.

Q11

The severity of a disease and blood group were studied in a research project. The findings are given in the following table, known as the m * n contingency table. Can the severity of the condition and blood group are associated?

Show the severity of a disease classified by blood group in 1500 patients using X^2.

Condition	Blood Groups			
	O	A	B	AB
Severe	51	40	10	9
Moderate	105	103	25	17
Mild	384	527	125	104

Solutions:

Solution:

First we define the cases for different hypothesis such as,

H_0 = There is no link between gender and political party preference.

H_1 = There is a link between gender and political party preference.

Now we calculate the expected frequency.

$$Expected\ Value = \frac{(Row\ Total) \times (Column\ Total)}{Total\ Number\ of\ Observations}$$

For one case such as,

Expected Value of Male Republican = $\frac{240 \times 200}{440} = 109$

Similarly, we calculate the expected value for each of the cells such as,

Finally we calculate the test statistics X^2 which is the sum of the cell values from the above table.

$$X^2 = 0.743 + 2.05 + 2.33 + 3.33 + 0.384 + 1 = 9.837$$

Next we have to calculate the degrees of freedom such as,

$$(r - 1) \times (c - 1) = (3 - 1) \times (2 - 1) = 2$$

Where, r = number of column items, and

c = number of row items.

$$\chi^2_{cal} < \chi^2_{tab}$$

Ans- X^2 =12.2347, X^2 table value=12.59, df=6 We accept the null hypothesis.

Q12

A gym trainer claimed that all the new boys in the gym are above average weight. A random sample of thirty boys weight have a mean score of 112.5 kg the population mean weight is 100 kg and the standard deviation is 15. Is there sufficient evidence to support the claim of gym trainer?

Solutions:

Step-1: State Null and Alternate Hypothesis

Null Hypothesis:

H_o: $\mu = 100$

Alternate Hypothesis:

H_a: $\mu > 100$

Step-2: Set the significance level (alpha-value)

Let alpha-value is 0.05, so corresponding z-score is 1.645

Step-3: Find the z-value

$$\boxed{z = \frac{\overline{X}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{112.5-100}{\frac{15}{\sqrt{30}}} = 4.56}$$

Step-4: Comparing with the significance level:

From step-3, we have

4.56 > 1.645

So, we have to reject the null hypothesis.

i.e. average weight of new boys are greater than 100 kg