Page No.
Date: / /

PR-3 - Derivable Judgement.

Q.1 What is inferential statistics?

→ It can use sample data to draw conclusion about a larger population, usually with probability statement.

Ex :- from 200 patients. to estimate the average diabetes patients in the whole city.

Q:2 What is Hypothesis Testing and its component.

→ Hypothesis testing is a formal process to decide if the data give strong enough evidence to ~~decide~~ reject a null hypothesis Ho in favour of an alternative H1.

Ex :- "BMI vs exercise"
Suppose you want to check if mean bmi differ between people who exercise Daily and Never. ~~in~~

$H_0 : \mu$ Daily $= \mu$ Never - (no difference)
$H_1 : \mu$ Daily $\neq \mu$ Never (some difference)

Q.3. Explain confidence interval and critical value

→ A confidence interval is a range of plausible values for a population parameter (e.g., mean or proportion) based on sample data plus a margin of error.

→ critical value is the cutoff of the sampling distribution that separate the "do not reject Ho" and ~~Ho~~ " reject Ho" region at a chosen α.

Ex :- 95% is a confidence interval &
z - critical value = 1.96.

Q.4 Define P - value

→ The P- value is the probability of observing the sample result assuming the null hypothesis is true. A smaller p-value provides stronger evidence against Ho.

Ex :- if $P = 0.03 < 0.05$, reject the null hypothesis.

Q.5. Differentiate type 1 and type II error

→ Type I = Rejecting a true Ho (false positive),
probability = α.
Type II = Not rejecting a false Ho (false negative),
probability = β.

Ex :- Type II = Smoking doesn't increase the risk of diabetes when it actually does.

Type I = Saying the treatment work on 3rd stage cancer when it actually doesn't.

Q.6. Brief description of z-test, t-test, chi-square, and ANOVA Test.

→ z test

~~who~~ used when sample size is large and population variance is known.

Ex - ~~A~~ ~~company~~ DO younger adult (18-35) years and older adult (60+ years) have differ X

Ex:- Is the average fasting glucose level in our sample different from the guideline value of 100 mg/dL?

hypothesis

- H_0 - The population mean fasting glucose is 100 mg/dL, $\mu = 100$.
- H_1 - The population mean fasting glucose is not 100 mg/dl, $\mu \neq 100$. (two-tailed)

we assume the population standard deviation of glucose is known and equal to 15 mg/dL.

sample information (from 200 patient)

- Sample size : $n = 200$
- Sample mean : $\bar{x} = 103$ mg/dL
- population standard deviation (assumed known) : $\sigma = 15$ mg/dL

Test statistics (z-test)

$$SE = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{200}} \approx 1.06$$

$$z = \frac{\bar{x} - \mu_0}{SE} = \frac{103 - 100}{1.06} \approx 2.83$$

Page No.
Date: / /

Decision
since $t = 3.80 > 1.98$, The p-value is less than 0.001, which below 0.05.

Conclusion
we reject the null hypothesis

Chi-square Test :
- Test ~~class~~ association between categorical variables.

Ex - Teachers wants to know if boys and girls like sports equally.

	Like Sports	Don't like	Total
Boys	12	8	20
Girls	6	14	20
Total	18	22	40

- H_0 - Gender and liking sports are independent
- H_1 - Gender & liking sports aren't independent.

Expected counts.

$$E = \frac{\text{row total} \times \text{column total}}{\text{grand total}}$$

- Boys & like $= \frac{20 \times 18}{40} = 9$
- Boys & Not like $= \frac{20 \times 22}{40} = 11$
- Girls & like $= \frac{20 \times 18}{40} = 9$

Page No.______ Date: __/__/__ Shape

- Girls & Not like : E $= \frac{20 \times 22}{40} = 11$

$$\chi^2 = \sum \frac{(O-E)^2}{E}$$

1. Boys & like : $\frac{(12-9)^2}{9} = \frac{3^2}{9} = \frac{9}{9} = 1$

2. Boys & Not like : $\frac{(8-11)^2}{11} = \frac{(-3)^2}{11} = \frac{9}{11} = 0.82$

3. Girls & like : $\frac{(-3)^2}{11} = \frac{(-3)^2}{9} = \frac{9}{9} = 1$

4. Girls & Not like : $\frac{(14-11)^2}{11} \quad \frac{3^2}{11} = \frac{9}{11} = 0.82$

Adding:

$$\chi^2 = 1 + 0.82 + 1 + 0.82 = 3.64$$

Degree of freedom:

$$df = (r-1)(c-1) = (2-1)(2-1) = 1$$

At $\alpha = 0.05$, critical value for $df = 1$ is about 3.84.

- Our $\chi^2 = 3.64$
- Critical value $= 3.84$

$3.64 < 3.84 \rightarrow$ don't reject H_0.

ANOVA test

compares means of three or more group

Ex - teacher wants to know if the avg. test score is the same for three different teaching methods.

Sample :

Method	Scores
A	70, 75, 80
B	65, 70, 72
C	80, 85, 90

Total Students $N = 9$, groups $k = 3$.

- H_0 = All three population means are equal
  $\mu A = \mu B = \mu c$.
- H_1 - At least one mean is different.
  Significance level $\alpha = 0.05$.

Group means :

- Method A - $(70 + 75 + 80) / 3 = 75$
- Method B - $(65 + 70 + 72) / 3 = 69$
- Method C - $(80 + 85 + 90) / 3 = 85$

Overall grand mean :

$$\frac{70 + 75 + 80 + 65 + 90 + 72 + 80 + 85 + 90}{9} = \frac{687}{9} =$$

76.33 (approx

Between-group variation (SSB)

Formula : $SSB = \Sigma n_i (\bar{X}_i - \bar{X}_{overall})^2$

Each group has $n_i = 3$.

- Group A $= 3(75 - 76.33)^2 = 3(-1.33)^2 \approx 3 \times 1.77 = 5.31$
- Group B $= 3(69 - 76.33)^2 = 3(-7.33)^2 \approx 3 \times 53.73 = 161.19$
- Group C $= 3(85 - 76.33)^2 = 3(8.67)^2 \approx 3 \times 75.17 = 225.51$

$$SSB \approx 5.31 + 161.19 + 225.51 = 392.01$$

Degree of freedom between:

$$df_B = k - 1 = 3 - 1 = 2$$

$$MSB = \frac{SSB}{df_B} \approx \frac{392.01}{2} = 192.01$$

Within group variation

- Group A (mean 75)
  - $(70 - 75)^2 = 25$
  - $(75 - 75)^2 = 0$
  - $(80 - 75)^2 = 25$

  sum $= 50$

- Group B (mean 69)
  - $(65 - 69)^2 = 16$
  - $(70 - 69) = 1$
  - $(72 - 69) = 9$

  sum $= 26$

- Group C (mean 85)
  - $(80 - 85)^2 = 25$
  - $(85 - 85)^2 = 0$
  - $(90 - 85)^2 = 25$

  sum $= 50$

$$SSW = 50 + 26 + 50 = 126$$

Degree of freedom within:

$$df = N - k = 9 - 3 = 6$$

$$MSW = \frac{SSW}{df_w} = \frac{126}{6} = 21$$

Page No.

Date: / /

F - statistics

$$F = \frac{MSB}{MSW} = \frac{196.01}{21} \approx 9.33$$

for $\alpha = 0.05$, with $df_B = 2$ and $df_w = 6$, the critical F is about 5.14 (from F-table).

- Calculated $F = 9.33$
- Critical $F_{0.05, 2, 6} \approx 5.14$

since $9.33 > 5.14$, the test is significant.

Decision · Conclusion

Reject H_0.

component; uses the F statistic.

Q.7. What is Covariance ?

→ Covariance measures how two variable vary together, a positive value means they tend to move in the same direction, and a negative value means they move in opposite direction. Ex:- Height & weight usually have positive covariance

Q.8. What is Correlation

→ Correlation rescales covariance to a standardized coefficient between -1 and +1, showing both the strength and direction of the linear relationship, independent of units.

Ex:- Correlation = +0.85 mean between study hours and exam score indicates a strong positive relationship.