

Assignment (Statistical Inference – Fall – 2023)

Due Date: 01.12.2023 (Till 10:00 PM)

Weight: 10%

- Q1. The restaurant management tells its customers that the average cost of dinner is Rs. 500 with a standard deviation of Rs. 75. A group of concerned customers think that the average cost is higher. To test the restaurant claims, 100 customers purchase dinner at the store and the mean price is Rs. 520. Perform a hypothesis test at 1% level of significance.
- Q2. A random sample of 12 shearing pins is taken in a study of the Rockwell hardness of the pin head. Measurements on the Rockwell hardness are made for each of the 12, yielding an average value of 48.50 with a sample standard deviation of 1.5. Assuming the measurements to be normally distributed, construct a 90% confidence interval for the mean Rockwell hardness .
- Q3. A large chain of supermarkets sells 5000 packets of cereal in each of its stores each month. It decides to test-market a different brand of cereal in 15 of its stores. After a month the 15 stores sold an average of 5200 packets each, with a standard deviation of 500 packets. Test the hypothesis that $\mu > 5,000$ at 0.05 level of significance.
- Q4. An electrical firm manufactures light bulbs that have a length of life that is approximately normally distributed, with a mean equal to 800 hours and a standard deviation of 40 hours. Find the probability that a random sample of 16 bulbs will have an average life of less than 775 hours.
- Q5. The production manager of Northern Windows, Inc., has asked you to evaluate a proposed new procedure for producing its Regal line of double-hung windows. The present process has a mean production of 80 units per hour with a population standard deviation of $\sigma = 8$. The manager does not want to change to a new procedure unless there is strong evidence that the mean production level is higher with the new process. Test the hypothesis that the mean production level is higher with the new process at 0.01 level of significance.
- Q6. If all possible samples of size 16 are drawn from a normal population with mean equal to 50 and standard deviation equal to 5, what is the probability that a sample mean \bar{X} will fall in the interval from $\mu_{\bar{X}} - 1.9\sigma_{\bar{X}}$ and $\mu_{\bar{X}} - 0.4\sigma_{\bar{X}}$?
- Q7. A machine produces metal pieces that are cylindrical in shape. A sample of pieces is taken, and the diameters are found to be 1.01, 0.97, 1.03, 1.04, 0.99, 0.98, 0.99, 1.01, and 1.03 centimeters. Find a 99% confidence interval for the mean diameter of pieces from this machine, assuming an approximately normal distribution.
- Q8. According to *Reader's Digest*, 42% of primary care doctors think their patients receive unnecessary medical care.
- a) Suppose a sample of 300 primary care doctors was taken. Show the sampling distribution of the proportion of the doctors who think their patients receive unnecessary medical care.
 - b) What is the probability that the sample proportion will be within 0.03 of the population proportion?
 - c) What is the probability that the sample proportion will be within 0.05 of the population proportion?

- d) What would be the effect of taking a larger sample on the probabilities in parts (b) and (c)? Why?
- Q9. To test a two-tailed test, a random sample of $n = 24$ is obtained from a population that is known to be normally distributed with $\sigma = 12$, and we got a sample mean of 47.1. Will the null hypothesis be rejected? Use critical value approach.
- Q10. A sample of 25 workers found the average overtime hours worked in the previous week was 7.8, with a standard deviation of 4.1 hours. Test the hypothesis that the average for all workers is 5 hours or less. Use a 0.05 level of significance.
- Q11. In a market study for BGI, a local department store, you select a sample of 60 actual and potential clients to interview. Among the questions you wish to answer is whether the clients and non- clients differ in their incomes. The table below gives summary statistics. Can you conclude that there is a difference in the mean incomes of clients and non- clients? Use $\alpha = 0.05$.

	Clients	Non-Clients
Mean income (in \$1000s)	58.7	50.4
Standard deviations (in \$1000s)	16.8	9.8
Number	29	31

- Q12. The average zinc concentration recovered from a sample of measurements taken in 36 different locations in a river is found to be 2.6 grams per milliliter. Find the 95% and 99% confidence intervals for the mean zinc concentration in the river. Assume that the population standard deviation is 0.3 gram per milliliter. What would be the effect on the margin of error when the confidence coefficient is increased from 95% to 99%?
- Q13. The Economic Policy Institute periodically issues reports on worker's wages. The institute reported that mean wages for male college graduates were \$37.39 per hour and for female college graduates were \$27.83 per hour in 2017. Assume the standard deviation for male graduates is \$4.60, and for female graduates it is \$4.10.
1. What is the probability that a sample of 50 male graduates will provide a sample mean within \$1.00 of the population mean, \$37.39?
 2. What is the probability that a sample of 50 female graduates will provide a sample mean within \$1.00 of the population mean, \$27.83?
 3. In which of the preceding two cases, part (a) or part (b), do we have a higher probability of obtaining a sample estimate within \$1.00 of the population mean? Why?
- Q14. A simple random sample of 40 items resulted in a sample mean of 25. The population standard deviation is 5.
1. What is the standard error of mean?
 2. At 95% confidence, what is the margin of error?

- Q15. A 95% confidence interval for a population mean was reported to be 152 to 160. If $\sigma = 15$, what sample size was used in this study?
- Q16. A manufacturer claims that the average tensile strength of thread A exceeds the average tensile strength of thread B by at least 12 kilograms. To test this claim, 50 pieces of each type of thread were tested under similar conditions. Type A thread had an average tensile strength of 86.7 kilograms with a standard deviation of 6.28 kilograms, while type B thread had an average tensile strength of 77.8 kilograms with a standard deviation of 5.61 kilograms. Test the manufacturer's claim using a 0.05 level of significance.
- Q17. A group of students' marks on two tests, before and after instruction, were as follows:

Student	1	2	3	4	5	6	7	8	9	10	11	12
Before	14	16	11	8	20	19	6	11	13	16	9	13
After	15	18	15	11	19	18	9	12	16	16	12	13

Test the hypothesis that the instruction has no effect on students' marks at 0.05 level of significance.

- Q18. The following measurements were recorded for the drying time, in hours, of a certain brand of latex paint:

3.4, 2.5, 4.8, 2.9, 3.6, 2.8, 3.3, 5.6, 3.7, 2.8, 4.4, 4.0, 5.2, 3.0, 4.8.

Assuming that the measurements represent a random sample from a normal population, find a 95% confidence interval for the drying time for the next trial of the paint.

- Q19. Suppose a new production method will be implemented if a hypothesis test supports the conclusion that the new method reduces the mean operating cost per hour.
1. State the appropriate null and alternative hypotheses if the mean cost for the current production method is \$220 per hour.
 2. What is the Type I error in this situation? What are the consequences of making this error?
 3. What is the Type II error in this situation? What are the consequences of making this error?
- Q20. The heights of 1000 students are approximately normally distributed with a mean of 174.5 centimeters and a standard deviation of 6.9 centimeters. Suppose random samples of size 25 are drawn from this population. Determine:
1. The mean and standard deviation of the sampling distribution of mean.
 2. The number of samples means falling between 172.5 and 175.8 centimeters inclusive.
 3. The number of samples means falling below 172.0 centimeters.
- Q21. The amount of time that a drive-through bank teller spends on a customer is a random variable with a mean $\mu = 3.2$ minutes and a standard deviation $\sigma = 1.6$ minutes. If a random sample of 64 customers is observed, find the probability that their mean time at the teller's window is:
1. At most 2.7 minutes.

2. More than 3.5 minutes.
 3. At least 3.2 minutes but less than 3.4 minutes.
- Q22. The label on a 3-quart container of orange juice states that the orange juice contains an average of 1 gram of fat or less. Answer the following questions for a hypothesis test that could be used to test the claim on the label.
1. State the appropriate null and alternative hypotheses.
 2. What is the Type I error in this situation? What are the consequences of making this error?
 3. What is the Type II error in this situation? What are the consequences of making this error?
- Q23. *Amstat News* (December 2004) lists median salaries for associate professors of statistics at research institutions and at liberal arts and other institutions in the United States. Assume that a sample of 200 associate professors from research institutions has an average salary of \$70,750 per year with a standard deviation of \$6000. Assume also that a sample of 200 associate professors from other types of institutions has an average salary of \$65,200 with a standard deviation of \$5000. Test the hypothesis that the mean salary for associate professors in research institutions is \$2000 higher than for those in other institutions. Use a 0.01 level of significance.
- Q24. On 12 randomly selected cars, one of each type of tyre is installed on the rear wheels and, as before, the cars are driven until the tyres wear out. The number of kilometers until wear-out occurred is shown in the following table. Can we conclude from these data that the new-design tyre is superior at 0.05 level of significance?

Car	New Design	Existing Design	Car	New Design	Existing Design
1	65	56	7	108	106
2	72	58	8	98	94
3	110	97	9	91	86
4	70	64	10	92	98
5	90	87	11	94	106
6	95	83	12	70	66

- Q25. Suppose we want to test the usefulness of IQ and gender on the test score. The regression outputs are as follows:

	Coef	Std err	t Stat	P-value
Intercept	38.6	9.535	4.048	0.005
IQ	0.4	0.094	4.278	0.004
Gender	7	2.686	2.606	0.035

SUMMARY OUTPUT	
<i>Regression Statistics</i>	
Multiple R	0.900113
R Square	0.810203
Adjusted R Square	0.755975
Standard Error	4.181592
Observations	10

F-test, p-value = 0.00000001

*Please note that female is the reference category.

- (1) What are the dependent and independent variables?
 - (2) Write down the regression equation by defining the dummy variable.
 - (3) Interpret the partial regression coefficients.
 - (4) Check the significance of the overall regression model.
 - (5) Check the significance of the individual β coefficients.
 - (6) Interpret the R – squared value.
- Q26. Suppose we have three variables for 12 different basketball players as follows:
- Points scored, hours spent practicing, and training program (program 1, program 2, program 3) used. The regression outputs are as follows:

Residuals:

Min	1Q	Median	3Q	Max
-1.5192	-1.0064	-0.3590	0.8269	2.4551

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.3013	0.9462	6.660	0.000159 ***
hours	0.9744	0.3176	3.068	0.015401 *
program2	2.2949	1.1369	2.019	0.078234 .
program3	6.8462	1.5499	4.417	0.002235 **

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.403 on 8 degrees of freedom

Multiple R-squared: 0.9392, Adjusted R-squared: 0.9164

F-statistic: 41.21 on 3 and 8 DF, p-value: 3.276e-05

- (1) What are the dependent and independent variables?
- (2) Which program is the reference category?
- (3) Write down the regression equation.
- (4) Interpret the partial regression coefficients.
- (5) Check the significance of the overall regression model.
- (6) Check the significance of the individual β coefficients.
- (7) Interpret the multiple R – squared value.

Q27. The outputs are related to the impact of the amount of money spent (in US\$) on three advertising media (youtube, facebook and newspapers) on sales units.

youtube facebook newspaper sales

276.1 45.4 83.0 26.5

53.4 47.2 54.1 12.5

20.6 55.1 83.2 11.2

181.8 49.6 70.2 22.2

Residuals

Min 1Q Median 3Q Max

-10.59 -1.07 0.29 1.43 3.40

Coefficients

Estimate Std. Error t value Pr(>|t|)

(Intercept) 3.52667 0.37429 9.42 2e-16 ***

youtube 0.04576 0.00139 32.81 2e-16 ***

facebook 0.18853 0.00861 21.89 2e-16 ***

newspaper -0.00104 0.00587 -0.18 0.86

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.02 on 196 degrees of freedom

Multiple R-squared: 0.897, Adjusted R-squared: 0.896

F-statistic: 570 on 3 and 196 DF, p-value: 2e-16

- (1) What are the dependent and independent variables?
- (2) Write down the regression equation.
- (3) Interpret the intercept.
- (4) Interpret the partial regression coefficients.
- (5) Check the significance of the overall regression model.
- (6) Check the significance of the individual β coefficients.
- (7) Interpret the multiple R – squared value.

Q28. A clinical study was conducted to see whether an antibody called denosumab is effective in treatment of osteoporosis of postmenopausal women, as reported in the article “Denosumab in Postmenopausal Women with Low Bone Mineral Density (*New England Journal of Medicine*, Vol. 354, No. 8, pp. 821–831) by M. McClung et al. Postmenopausal women with osteoporosis were randomly assigned into groups that received either a placebo, or a six-month regimen of Denosumab at doses of 14 mg, 60 mg, 100 mg, or 210 mg. The following table provides summary statistics for the body-mass indexes (BMI) of the women in each treatment group.

(F – Table Value = 2.41)

Treatment	n_j	\bar{x}_j	s_j
Placebo	46	25.9	4.3
14 mg	54	25.8	5.3
60 mg	47	27.5	5.8
100 mg	42	26.0	4.6
210 mg	47	25.9	4.3

At the 5% significance level, do the data provide sufficient evidence to conclude that a difference exists in mean BMI for women in the five different treatment groups?

Q29. Gender discrimination in hiring. Suppose you are investigating allegations of gender discrimination in the hiring practices of a particular firm. An equal-rights group claims that females are less likely to be hired than males with the same background, experience, and other qualifications. Data collected on 28 former applicants will be used to fit the model.

$$E(y) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3, \text{ where}$$

$$y = \begin{cases} 1 & \text{if hired} \\ 0 & \text{if not} \end{cases}$$

$$x_1 = \text{Years of higher education (4, 6, or 8)}$$

$$x_2 = \text{Years of experience}$$

$$x_3 = \begin{cases} 1 & \text{if male applicant} \\ 0 & \text{if female applicant} \end{cases}$$

- 1) Estimate the probability that female candidate who has 5 years of higher education and 6 years of experience is hired.
 - 2) Estimate the probability that male candidate who has 5 years of higher education and 6 years of experience is hired.
 - 3) Find and interpret the marginal effect of candidate's gender on hiring probability for a candidate who has 5 years of higher education and 6 years of experience.
 - 4) Find and interpret the marginal effect of another year of experience on hiring probability for a candidate who has 5 years of higher education and 6 years of experience.
 - 5) Predict whether a female candidate who has who has 5 years of higher education and 6 years of experience is hired or not hired.
 - 6) Find and interpret the odds ratio of experience and gender variables.
- Q30. One of the issues that came up in a recent national election (and is likely to arise in many future elections) is how to improve Australian economic growth. Specifically, should governments cut public spending and/or introduce tax reforms, job creation, and increase education funding? Politicians need to know which sectors of the electorate support these options to improve economic growth. Suppose that a random sample of 1000 voters were asked which option they support and their political affiliations: Labor, Liberal-National Coalition or Others (which included a variety of political persuasions). The responses were summarized in a table called a contingency table and shown below. Do these results allow us to conclude that political affiliation affects support for the economic options at 0.05 level of significance?

Economic options	Political affiliation		
	Labor	Coalition	Others
Cut public spending	101	282	61
Introduce tax reforms	38	67	25
Job creation	131	88	31
Increase education funding	61	90	25