

# FINAL EXAMINATION

**Course: Statistics and Probability**

**Based on Week 9 – Week 15 Materials**

**Lecture: Alfrian C Talakua**

## **PART A — MULTIPLE CHOICE QUESTIONS (40 QUESTIONS)**

**1. The main purpose of a confidence interval is to ...**

- A. Eliminate data errors
  - B. Determine population probability
  - C. Provide a range estimate for a population parameter
  - D. Reduce sample size
- 

**2. As sample size (n) increases, the standard error will ...**

- A. Increase
  - B. Stay the same
  - C. Decrease
  - D. Cannot be determined
- 

**3. The formula for standard error is:**

$$SE = \frac{\sigma}{\sqrt{n}}$$

- A.  $\sigma/n$
  - B.  $\sigma\sqrt{n}$
  - C.  $\sigma/\sqrt{n}$
  - D.  $\sqrt{\sigma}/n$
- 

**4. If p-value < 0.05, the correct decision is ...**

- A. Accept  $H_0$
  - B. Reject  $H_0$
  - C. Reject the sample
  - D. Data is invalid
- 

**5. A Type I Error occurs when ...**

- A. Failing to reject a false  $H_0$
  - B. Rejecting a true  $H_0$
  - C. Mean is too small
  - D. Sample size is too large
- 

**6. The null hypothesis usually states ...**

- A. There is an effect
  - B. There is a relationship
  - C. There is no effect or difference
  - D. All are correct
- 

**7. A Z-test is used when ...**

- A. Sample is small and  $\sigma$  unknown
  - B. Sample is large and  $\sigma$  known
  - C. Data is ordinal
  - D. Data is categorical
-

**8. A T-test is used when ...**

- A. Population variance is known
  - B. Data is nominal
  - C. Population variance is unknown
  - D. All are incorrect
- 

**9. The Pearson correlation coefficient ranges from ...**

- A. 0 to 100
  - B. -1 to +1
  - C. -10 to 10
  - D. 1 to 100
- 

**10. A value of  $r = 0.95$  indicates ...**

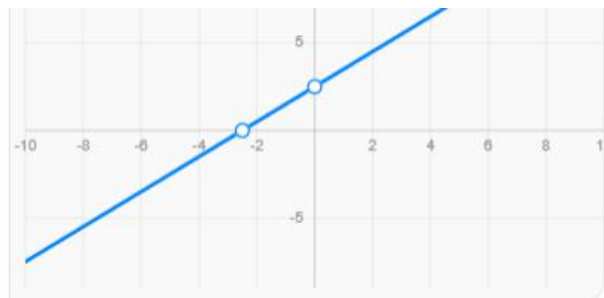
- A. Strong negative correlation
  - B. No relationship
  - C. Very strong positive correlation
  - D. Weak correlation
- 

**11. The simple linear regression equation is:**

$$Y = a + bX$$

$a$     2.5   

$b$     1.0   



- A.  $Y = aX$
- B.  $Y = a + bX$

C.  $X = a + bY$

D.  $Y = ab$

---

**12. The slope (b) in regression represents ...**

A. Initial value

B. Change in Y for each unit increase in X

C. Sample size

D. Data variance

---

**13.  $R^2$  is used to ...**

A. Measure sample size

B. Measure model prediction strength

C. Measure sampling error

D. Measure p-value

---

**14. ANOVA is used to ...**

A. Compare two means

B. Compare three or more means

C. Measure correlation

D. Calculate median

---

**15. The main statistic used in ANOVA is ...**

A. Z

B. t

C. F

D. r

---

**16. If the F-value is large, it means ...**

- A. Groups are likely significantly different
  - B. No difference exists
  - C. Variance is small
  - D. Data is invalid
- 

**17. Within-group variance is also called ...**

- A. Signal
  - B. Noise
  - C. Prediction
  - D. Correlation
- 

**18. Between-group variance represents ...**

- A. Variation between groups
  - B. Variation within groups
  - C. Measurement error
  - D. Population standard deviation
- 

**19. Chi-Square is used for ...**

- A. Continuous numerical data
  - B. Categorical data
  - C. Ratio data
  - D. Interval data
- 

**20. The Chi-Square test of independence is used to ...**

- A. Measure means
- B. Analyze relationships between categorical variables
- C. Measure regression
- D. Measure median

---

**21. If the Chi-Square p-value < 0.05, then ...**

- A. Variables are independent
  - B. No relationship exists
  - C. There is a significant relationship
  - D. Means are equal
- 

**22. The Mann-Whitney U test is used for ...**

- A. Two independent non-parametric groups
  - B. Linear regression
  - C. Correlation analysis
  - D. Confidence interval analysis
- 

**23. Kruskal-Wallis is a non-parametric alternative to ...**

- A. Regression
  - B. Correlation
  - C. ANOVA
  - D. Z-test
- 

**24. Ordinal data is best analyzed using ...**

- A. ANOVA
  - B. Regression
  - C. Non-parametric tests
  - D. Pearson correlation
- 

**25. Correlation does not imply ...**

- A. Prediction
- B. Causation

- C. Regression
  - D. Estimation
- 

**26. If p-value = 0.48, the correct decision is ...**

- A. Reject  $H_0$
  - B.  $H_0$  is significant
  - C. Fail to reject  $H_0$
  - D. Data is invalid
- 

**27. A 99% confidence level produces an interval that is ...**

- A. Narrower
  - B. Wider
  - C. The same
  - D. Impossible to determine
- 

**28. As standard deviation increases, the confidence interval becomes ...**

- A. Narrower
  - B. Wider
  - C. Constant
  - D. Negative
- 

**29. In regression analysis, X is called the ...**

- A. Dependent variable
  - B. Response variable
  - C. Independent variable
  - D. Random variable
-

**30. In regression analysis, Y is called the ...**

- A. Predictor
  - B. Independent variable
  - C. Dependent variable
  - D. Input variable
- 

**31. ANOVA is appropriate when the number of groups is ...**

- A. 1
  - B. 2
  - C. 3 or more
  - D. Unlimited
- 

**32. The alternative hypothesis in ANOVA states that ...**

- A. All means are equal
  - B. At least one mean is different
  - C. No variance exists
  - D. All groups are identical
- 

**33. Chi-Square is classified as a ...**

- A. Parametric test
  - B. Non-parametric test
  - C. Regression test
  - D. Correlation test
- 

**34. Expected frequencies in Chi-Square should generally be ...**

- A.  $< 1$
- B.  $< 3$
- C.  $\geq 5$
- D.  $= 0$

---

**35. If  $r = -0.80$ , the relationship is ...**

- A. Strong positive
  - B. Strong negative
  - C. Weak
  - D. No relationship
- 

**36. The p-value is used to ...**

- A. Measure sample size
  - B. Make hypothesis testing decisions
  - C. Calculate mean
  - D. Calculate median
- 

**37. In hypothesis testing,  $\alpha = 0.05$  means ...**

- A. 5% risk of Type I Error
  - B. 5% risk of Type II Error
  - C. 5% data error
  - D. Mean changes by 5%
- 

**38. Regression analysis is mainly used for ...**

- A. Testing categorical data
  - B. Prediction
  - C. Counting frequencies
  - D. Measuring median
-

**39. If the data is not normally distributed, it is more appropriate to use ...**

- A. Parametric tests
  - B. Non-parametric tests
  - C. Linear regression
  - D. Confidence intervals
- 

**40. The main purpose of statistics in Information Systems is to ...**

- A. Increase data volume
  - B. Create charts only
  - C. Support data-driven decision making
  - D. Eliminate all errors
- 

**PART B — CASE STUDY / ANALYTICAL QUESTIONS (10 QUESTIONS)**

**41.** An e-commerce company previously had an average transaction value of \$85. After redesigning the website, a sample of 100 customers showed an average transaction value of \$89 with  $\sigma = 15$ .

Conduct a Z-test at  $\alpha = 0.05$ .

---

**42.** A new server configuration has an average response time of 185 ms compared to the old standard of 200 ms with p-value = 0.012.

What is the statistical decision?

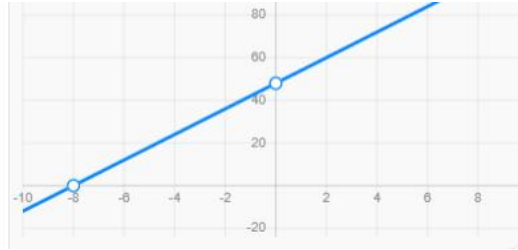
---

**43.** A study found a correlation coefficient of  $r = 0.92$  between study hours and exam scores. Interpret the result.

---

**44.** Given the regression equation:

$$Y = 48 + 6X$$



Predict the exam score if a student studies for 7 hours.

---

**45.** Three applications have the following mean satisfaction scores:

- App A = 7.8
- App B = 4.2
- App C = 6.0

ANOVA results show  $p = 0.001$ . What is the conclusion?

---

**46.** A Chi-Square test produced:

- $\chi^2 = 4.80$
- $p = 0.028$

Interpret the result.

---

**47.** In a cybersecurity study:

- Mobile breaches = 45
- Desktop breaches = 30

Why is Chi-Square appropriate for this analysis?

---

**48.** A company compares 3 marketing strategies using ANOVA. Why should they not use multiple t-tests instead?

---

**49.** A user satisfaction dataset uses a Likert scale (1–5) and is not normally distributed. Which statistical test is most appropriate?

---

**50.** Explain the main difference between correlation and regression.

## ANSWER KEY ( 1-50 )

### PART A

1. C

2. C

3. C

4. B

5. B

6. C

7. B

8. C

9. B

10. C

11. B

12. B

13. B

14. B

15. C

16. A

17. B

18. A

19. B

20. B

21. C

22. A

23. C

24. C

25. B

26. C

27. B

28. B

29. C

30. C

31. C

32. B

33. B

34. C

35. B

36. B

37. A

38. B

39. B

40. C

## PART B

41. **Answer:**

- $H_0: \mu = 85$
- $H_1: \mu \neq 85$

$$Z = \frac{89 - 85}{15/\sqrt{100}} = 2.67$$

Since p-value < 0.05 → Reject  $H_0$ .

Conclusion: The website redesign significantly increased transaction value.

42. **Answer:**

Because p-value 0.012 < 0.05, reject  $H_0$ .

Conclusion: The new server significantly reduces response time.

43. **Answer:**

$r = 0.92$  indicates a very strong positive correlation.

As study hours increase, exam scores also increase.

44. **Answer:**

$$Y = 48 + 6(7) = 90$$

Predicted exam score = 90.

45. **Answer:**

Since  $p < 0.05$ , reject  $H_0$ .

At least one application has a significantly different mean satisfaction score.

46. **Answer:**

Since  $p < 0.05$ , there is a significant relationship between the two categorical variables.

47. **Answer:**

Because the data consists of frequencies and categorical variables (mobile vs desktop), making Chi-Square the appropriate method.

**48. Answer:**

Multiple t-tests increase the risk of Type I Error.

ANOVA compares all groups simultaneously while controlling the overall error rate.

**49. Answer:**

Non-parametric tests such as Kruskal-Wallis or Mann-Whitney are more appropriate because the data is ordinal and non-normal.

**50. Answer:**

- Correlation measures the strength and direction of the relationship between variables.
- Regression is used for prediction and building equations that describe relationships between variables.