

2nd EXAM IN STATISTICS Answers Key

Exercise 1 (5pts): Analyse each scenario carefully and identify the error (s) (**circle the letter corresponding to your answer**)

Scenario 1 : Two groups of Algerian students are taught grammar using different methods: Group A receives explicit instruction, and Group B receives implicit instruction. The researcher uses an independent t-test to assess the efficacy of the two instructions by comparing the student's opinion collected in semi-structured interviews.

- a. Inappropriate test selection.
- b. **The collected data is unsuitable for t-test.**
- c. The participants do not belong to the same group.

Scenario 2 : To assess the impact of three different vocabulary learning apps, an instructor measured EFL students' post-intervention scores and ran a series of independent t-tests to compare each pair of apps.

- a. **Multiple comparisons inflate the risk of Type I error.**
- b. Independent t-tests are inappropriate for comparing more than two groups.
- c. Vocabulary learning cannot be accurately assessed through scores.

Scenario 3 : A study compares speaking confidence between Algerian EFL students in private and public schools. A standardized test was used to measure students' performance among 80 students (40 per group). Mann-Whitney U test was used to analyze differences between the two groups. The investigation aims to highlight how school environments impact students' confidence in speaking English in Algeria.

- a. **The collected data is unsuitable for the test.**
- b. Sample size too large for the test
- c. Inappropriate test selection.

Scenario 4 : A researcher investigates gender differences in perceptions of politeness strategies in Algerian discourse. 50 male and 50 female participants rated the politeness of conversational excerpts on a 5-point Likert scale (1 = not polite at all, 5 = very polite). Excerpts were selected from authentic interactions and validated for representativeness. Ratings were compared using ANOVA to identify significant differences between genders. The study aims to explore how gender influences sociolinguistic evaluations in the Algerian context.

- a. **Inappropriate test selection.**
- b. The collected data is unsuitable for the test.
- c. The research objective doesn't match with the procedure.

Scenario 5 : A study aims to explore whether Algerian students' self-reported anxiety levels differ before and after participating in a speaking-focused language camp. The same group of 30 students completed a Likert-scale anxiety questionnaire before and after the intervention. The researcher applied a Mann-Whitney U test to compare the two sets of scores.

- a. **The test does not account for the repeated measures design.**
- b. The Likert-scale format violates the assumptions of non-parametric tests.
- c. The sample size is too small for meaningful analysis.

Exercise 2 (4,5 pts): Indicate what test is most suitable for the following research questions.

RQ1: Do Algerian EFL students show significant improvement in their vocabulary test scores after participating in a four-week intensive English workshop? **1**

- a- Dependent t-test
- b- Chi-square for goodness of fit
- c- Wilcoxon signed-rank test
- d- Chi-square homogeneity test

RQ2: Are there significant differences in text classification accuracy when processing Algerian Arabic data using three different machine learning models? **1**

- a- ANOVA
- b- Chi-square test of independence
- c- Kruskal- Wallis test
- d- Pearson's coefficient of correlation

RQ3: Are there significant differences in perceived academic stress levels between Algerian university students living in dormitories and those commuting from home? **1**

- a- Regression
- b- Mann-Whitney U test
- c- Brown and Smythe's test
- d- t-test

Exercise 3 (4 pts): Use the following fictional SPSS output to answer the questions:

Scenario: A two-way ANOVA examines the effect of “teaching method” (traditional vs. project-based) and “gender” (male vs. female) on speaking test scores.

Key Output:

- Main effect of method:	$F(1, 76) = 5.01, p = .028$
- Main effect of gender:	$F(1, 76) = 1.45, p = .232$
- Interaction (method*gender):	$F(1, 76) = 7.62, p = .007$

Questions:

1. What does the interaction result suggest?

- A) Gender has a significant effect regardless of method
B) The effect of method is different for males and females
 C) No significant relationship exists
 D) There is no interaction

2. Which effect is NOT statistically significant?

- A) Teaching method
B) Gender
 C) Method*Gender
 D) All are significant

3. How should the main effect of teaching method be interpreted?

- A) There is no significant difference
 B) Project-based learning is equally effective for both genders
C) There is a significant difference in means due to method
 D) Gender moderates the method effect

4. How would you report the significant interaction effect in APA style?

- A) $F(1, 76) = 7.62, p = .007$**
 B) $t(76) = 7.62, p = .007$
 C) $p < .05$
 D) ANOVA interaction $p = .007$

Exercise 1 (6,5 pts): Determine whether the following statements are true (T) or false (F) and correct the false statements.

N°	Statements	T/F
1	While commonly cited as a robust alternative to the t-test, the Mann-Whitney U test’s utility for comparing central tendencies extends only to two independent samples with similarly shaped distributions, making its use for comparing medians across three or more groups fundamentally flawed.	F
	The Mann-Whitney U test is specifically limited to comparing two independent groups. For more than two groups, the Kruskal-Wallis test is appropriate. Additionally, valid comparison of medians using Mann-Whitney U requires similarly shaped distributions.	1
2	Parametric tests typically exhibit greater power when their underlying assumptions, such as normality, are met. Under these conditions, parametric tests are usually <u>less</u> efficient at detecting genuine effects compared to nonparametric tests.	F
	Parametric tests typically exhibit greater power when their underlying assumptions, such as normality, are met. Under these conditions, parametric tests are usually <u>more</u> efficient at detecting genuine effects compared to nonparametric tests.	1
3	Mann-Whitney U test is ideal for comparing two <u>dependent</u> groups when data is <u>categorical</u> , continuous, skewed, or violates normality and equal variance assumptions.	F
	Mann-Whitney U test is ideal for comparing two <u>independent</u> groups when data is <u>ordinal</u>, continuous, skewed, or violates normality and equal variance assumptions.	1
4	ANOVA assumes normality, homogeneity of variance, and independence. Violations—especially with unequal sample sizes—can lead to incorrect inferences, even if the F-ratio appears significant.	T
		1
5	When the distributions of the two groups differ in shape, the Mann-Whitney U test no longer tests for a shift in medians but instead for differences in the distribution as a whole.	T
		1