

Student's name _____

Total value of all 7 problems is 100 points.

1. (10 pts) Four data points are given:

$$(0, 0), (4, -1), (2, -1), (-2, 6),$$

Estimate the parameters α_1 and β of the regression line $y = \alpha_1 + \beta x$. Draw the scatter plot, marking the data points and the regression line.

Answer: $y = 2.1 - 1.1x$.

2. (15 pts) Given a random sample

12.2, 2.5, -5.0, 2.7, 3.3, -7.5, 5.6, 2.6, 3.1, 2.4, 13.2

(a) Find point estimates for the median m and both quartiles q_1 and q_3 .

Answer: $\hat{m} = 2.7$, $\hat{q}_1 = 2.4$, $\hat{q}_3 = 5.6$.

(b) Find an approximate 93% confidence interval for the median. What is the exact confidence level of your interval?

Answer: CI is $[2.4, 5.6]$ at 93.46%.

(c) Determine the probability

$$P(2.5 < \pi_{0.55} < 13.2)$$

Answer: $P(2.5 < \pi_{0.55} < 13.2) = P(4 \leq b(11, 0.55) \leq 10) = 0.9376$.

3. (15 pts) In the following table, 500 individuals are classified by gender and by whether they answer Yes, No, or Not Sure in a certain poll. Test the null hypothesis that the overall distributions of the answers are the same for both sexes. Let $\alpha = 0.01$.

Gender	Yes	No	Not Sure
Male	128	78	14
Female	122	147	11

[Bonus] What can you say about the p-value of the test?

Answer: $Q = 14.68$. The critical region is $Q > \chi_{0.01}^2 = 9.210$, thus we reject H_0 . The p-value is less than 1%.

4. (15 pts) Two samples from random variables X and Y are recorded:

X : 18, -4, 34, -7, 2, 40, -16, 9, -3

Y : 36, -7, 64, 58, 18, 26

Use the Wilcoxon test to test the hypothesis $H_0 : m_X = m_Y$ against $H_1 : m_X < m_Y$. Determine the p-value. What does the p-value mean? How would you conclude the test depending on the significance level α ?

Answer: $W = 62$, $\mu = 48$, $\sigma^2 = 72$, $Z = 1.65$, p-value= 0.0495.

5. (15 pts) Let X_1, X_2, X_3 be three independent random variables that have some distributions with mean values μ_1, μ_2, μ_3 respectively. Test the hypothesis

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

at $\alpha = 10\%$. The observed data are given in the table below:

$X_1 :$	8	5	6	5		
$X_2 :$	2	4	5	4	5	4
$X_3 :$	11	8	11			

Construct an ANOVA table (without p-value) and state your conclusion.

Answer: $SS(E) = 18$, $SS(T) = 72$, $F = 20$. Critical region is $F > F_{0.1}(2, 10)$, which is not in the table VII, but its value is clearly less than 19. So we reject H_0 .

6. (15 pts) In a regression problem, $n = 20$ data points are observed and the following accumulated values are found:

$$\sum x_i = 80, \sum y_i = 10, \sum x_i^2 = 480, \sum x_i y_i = 48, \sum y_i^2 = 12.8$$

(a) Find $\hat{\alpha}$, $\hat{\beta}$, and $\hat{\sigma}^2$.

Answer: $\hat{\alpha} = 0.5$, $\hat{\beta} = 0.05$, $\hat{\sigma}^2 = 0.37$.

(b) Find 90% confidence intervals for α and β .

Answer: $[0.25, 0.75]$ and $[-0.038, 0.138]$.

(Bonus) Find a 90% confidence interval for σ^2 .

Answer: $[0.256, 0.788]$.

7. (15 pts) A computer program supposedly generates a standard normal random variable $N(0, 1)$. The following numbers were produced by this program:

0.2, 1.2, -1.4, -0.6, 2.4, -1.0, -0.6, 0.8

Use the Kolmogorov-Smirnov test to test the hypothesis that the program works right. Let $\alpha = 5\%$. Sketch an empirical distribution function. Indicate how you would construct a 95% confidence band around the empirical distribution function.

[Bonus] Give a formula for computing the p-value of the test and find the p-value approximately.

Answer: $D = 0.2257$. The critical region is $D > 0.46$. We accept H_0 .