

Statistical Data Processing – TEST – Part 1 – 30 points

Name:

1) Descriptive statistics: Count the measures of central tendency and measures of variability.

25	25	22	28	31	23
26	26	23	29	29	25

Population mean:

Mode:

Median:

Range:

Population variance:

Population standard deviation:

2) A company wants to know if customer preferences for five types of products—Product A, Product B, Product C, Product D, Product E—match the market's expected preference distribution.

They believe that 30% of customers prefer Product A, 20% prefer Product B, 20% prefer Product C, 20% prefer Product D, and 10% prefer Product E.

They conduct a survey of 500 customers to test this hypothesis.

Test at the alpha significance level 0.1

Null hypothesis (H0): Customer preferences match the expected distribution.

Alternative hypothesis (H1): Customer preferences do not match the expected distribution

	O _i (Observed)	E _i (Expected)	Test criterion
Product A	110		
Product B	100		
Product C	90		
Product D	110		
Product E	90		

Test criterion =

Critical Value =

Do you reject or fail to reject the null hypothesis? Why?

3) A researcher wants to determine if there is an association between gender and preference for two types of books.

The researcher surveys 150 people, and the data collected is summarized in the table below

	Book 1	Book 2	Total
Male	40	35	
Female	30	45	
Total			

Test at the alpha significance level 0.01

Null hypothesis (H_0): Gender and book preference are independent (no association).

Alternative hypothesis (H_1): Gender and book preference are not independent (there is an association).

Expected frequencies		
	Book 1	Book 2
Male		
Female		

Test criterion		
	Book 1	Book 2
Male		
Female		

Test criterion =

Critical Value =

Do you reject or fail to reject the null hypothesis? Why?

4) Analysis of variance:

1st car	2nd car	3rd car
4.851	5.899	3.952
4.745	4.256	4.052
4.781	4.052	3.825
4.782	4.125	3.932

We tested the mileage of three cars, and you are given three samples of observations.

- At the significance level of $\alpha = 1\%$, test the null hypothesis that all three cars have the same results.

- Formulate the null hypothesis $\mu_1 = \mu_2 = \mu_3$ that all three cars have the same mileages.

- Formulate the alternative hypothesis that the mileage of at least one mileages is different.

- Calculate:

- the group sums and the group sample means \bar{y}_i

- the grand sum and the grand sample mean \bar{y}

- Calculate the quantity SS_B = the sum of squares “between” = RegSS.

- Calculate the quantity SS_W = the sum of squares “within” = RSS.

- Calculate the Coefficient of Determination ($R^2 = \text{RegSS} / \text{TSS}$, where $\text{TSS} = \text{RegSS} + \text{RSS}$).

Is the fit “good”/“poor”?

- Determine the quantity DF_B = the degrees of freedom “between”.

- Determine the quantity DF_W = the degrees of freedom “within”.

- Calculate the quantity MS_B = the mean squares “between”,

i.e. the the sample variance between the groups = SS_B / DF_B .

- Calculate the quantity MS_W = the mean squares “within”,

i.e. the the sample variance within the groups = SS_W / DF_W .

- Calculate the statistic $F = MS_B / MS_W = (SS_B / SS_W) / (DF_B / DF_W)$.

- Calculate the critical value c .

- Calculate the p -value of the test.

- Do you reject or fail to reject the null hypothesis?