KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DEPARTMENT OF MATHEMATICS

STAT 212: Business Statistics II

Term 232, Major Exam II

Sunday March 24, 2024, 09:30 PM

Time allowed **2 hours**

Name: _____ ID #: _____ Sec #: _____

Important instructions:

- Check that this exam has 20 questions.
- All types of mobile phones or smart watches are **not** allowed during the examination.
- Using HB 2.5 pencil is recommended for bubbling your answers.
- Use a good eraser. It is recommended **not** to use the erasers attached to the pencil.
- Write your name, ID number and Section number on the examination paper and in the upper left corner of the Test Answer Form (bubbling sheet).
- When bubbling your ID and Section numbers, be sure that the bubbles match with the numbers that you write.
- The Test Code Number is already bubbled in your Test Answer Form. Make sure that it is the same as that printed on your question paper.
- When bubbling, make sure that the bubbled space is fully covered.
- When erasing a bubble, make sure that you do not leave any trace of penciling.
- A formula sheet will be supplied during the exam, and you are prohibited from bringing any printed or written materials, including a formula sheet, with you.
- You are required to bring a standard scientific calculator with basic functions for the exam; however, calculators with more advanced features are **not** permissible.

[Blank page]

(Q12.7) A survey was conducted of 549 consumer magazines on the practices of their websites. Of these, 189 magazines reported that online-only content is copy-edited as rigorously as print content. Suppose that a sample of 489 newspapers revealed that 193 reported that online-only content is copy-edited as rigorously as print content.

Observed Frequencies		Type of Print Media		Total
		Magazines	Newspapers	
Is online-only content copy-edited as rigorously as print content?	Yes	189	193	
	No			
Total		549	489	

At 0.1 level of significance, it is to be tested if there is a difference between consumer magazines and newspapers in the proportion of online-only content that is copy-edited as rigorously as print content. Based on this information, **answer the next <u>three</u> questions**:

Q1: What is the alternative hypothesis, H_1?

(A) $\boldsymbol{\pi_1 \neq \pi_2}$
(B) $\pi_1 = \pi_2$
(C) $\pi_1 > \pi_2$
(D) $\pi_1 < \pi_2$
(E) $\pi_1 \geq \pi_2$

where π_1 denotes the proportion of consumer magazines reporting that online-only content is copy-edited as rigorously as print content and π_2 denotes the proportion of newspapers reporting that online-only content is copy-edited as rigorously as print content.

Q2: What is the computed value of test statistic?

(A) **2.8271**
(B) 0.8417
(C) 3.841
(D) 10.073
(E) -0.0093

Q3: Which one of the following is correct decision rule?

(A) **Reject H_0 if $\chi_0^2 > 2.706$**
(B) Reject H_0 if $|\chi_0^2| > 3.841$
(C) Reject H_0 if $\chi_0^2 < 2.706$
(D) Reject H_0 if $\chi_0^2 < 3.841$
(E) Reject H_0 if $|\chi_0^2| > 5.991$

(Q12.16) More shoppers do the majority of their grocery shopping on Saturday than any other day of the week. However, is there a difference in the various age groups in the proportion of people who do the majority of their grocery shopping on Saturday? A study showed the results for the different age groups.

f_o	Age			Total
Major Shopping Day	Under 35 (Group 1)	35-54 (Group 2)	Over 54 (Group 3)	
Saturday	119	62	36	217
Not Saturday	221	232	152	605
Total	340	294	188	822

$\frac{(f_o-f_e)^2}{f_e}$	Age			
Major Shopping Day	Under 35	35-54	Over 54	
Saturday	9.5277	3.1408	3.7433	
Not Saturday	3.4174	1.1265	1.3427	
Total				22.2984

Marascuilo's Procedure						
Groups	p_i	p_j	n_i	n_j	Critical Range	$\lvert p_i - p_j \rvert$
1 and 2	0.35	0.210884	340	294	0.1067	0.1391
1 and 3	0.35	0.191489	340	188	$\boldsymbol{CR_{1,3}}$	0.1585
2 and 3	0.210884	0.191489	294	188	0.1131	0.0194

Comparing the computed value of test statistic with the critical value, there is enough evidence to conclude that there is a significant relationship between age and major grocery shopping day.

At 0.01 level of significance, we want to determine which age groups are different. Based on this information, **answer the next <u>two</u> questions**:

Q4: What is the critical range ($\boldsymbol{CR_{1,3}}$) for testing the difference between Under 35 and Over 54 groups?

(A) 0.1172
(B) 0.0946
(C) 3.0348
(D) 0.0995
(E) 0.1131

Q5: Which one of the following conclusions is true?

(A) There is a significant difference between the under 35 and 35-54 groups, and between under 35 and over 54 groups.
(B) There is a significant difference between the under 35 and 35-54 groups, and between 35-54 and over 54 groups.
(C) There is a significant difference between 35-54 and over 54 groups, only.
(D) There is a significant difference between under 35 and over 54, and between 35-54 and over 54 groups.
(E) There is a significant difference between all three pairs of groups.

Q6: (Q12.22) The owner of a restaurant serving Continental-style entrées has the business objective of learning more about the patterns of patron demand during the Friday-to-Sunday weekend time period. Data were collected from 615 customers on the type of entrée ordered and the type of dessert ordered and organized into the following table:

	Type of Entrée				Total
Type of Dessert	Beef	Poultry	Fish	Pasta	
Ice cream	29	9	17	11	66
Cake	65	20	29	11	125
Fruit	9	10	9	3	31
None	155	65	140	33	393
Total	258	104	195	58	615

At the 0.01 level of significance, we want to test if there is evidence of a relationship between type of dessert and type of entrée. The computed value of test statistic came out to be 19.4497. Identify the option that accurately presents the critical value along with the corresponding decision.

(A) Critical value is 21.666 and we fail to reject H_0.
(B) Critical value is 14.684 and we fail to reject H_0.
(C) Critical value is 21.666 and we reject H_0.
(D) Critical value is 23.589 and we to reject H_0.
(E) Critical value is 14.684 and we to reject H_0.

(Q12.32) The personnel director of a large department store wants to reduce absenteeism among sales associates. She decides to institute an incentive plan that provides financial rewards for sales associates who are absent fewer than five days in a given calendar year. A sample of 113 sales associates selected at the end of the second year reveals the following:

	Year 2		Total
Year 1	< 5 Days Absent	≥ 5 Days Absent	
< 5 Days Absent	35	13	48
≥ 5 Days Absent	21	44	65
Total	56	57	113

We want to test if the proportion of employees absent on fewer than five days was lower in year 1 than in year 2. Based on this information, **answer the next <u>two</u> questions**:

Q7: What is the computed value of test statistic?

(A) -1.372
(B) 18.214
(C) -0.235
(D) 10.976
(E) 5.831

Q8: Compute the *p-value*.

(A) 0.08534
(B) 0.00050
(C) 0.40517
(D) More than 0.99999
(E) Less than 0.00001

Q9: (Q12.73) Under what conditions should you use the χ^2 test of independence?

(A) All expected frequencies are at least one
(B) The number of categories is more than two
(C) The variable of interest follows a Chi-square distribution
(D) Overall sample size is more than 30
(E) All observed frequencies as less than the expected frequencies

Q10: (Q13.64) What is the interpretation of the coefficient of determination in the simple linear regression?

(A) It represents the proportion of variation in Y that is explained by the independent variable X in the regression model.
(B) It represents the estimated expected change in Y per unit change in X.
(C) It represents the estimated mean value of Y when X equals 0.
(D) It represents the strength and direction of linear relationship between Y and X.
(E) It represents the average value of Y when the independent variable $X = x_0$.

(Q13.87) Can demographic information be helpful in predicting sales at sporting goods stores? The file contains the monthly sales totals from a random sample of 24 stores in a large chain of nationwide sporting goods stores. All stores in the franchise, and thus within the sample, are approximately the same size and carry the same merchandise. The data set contains the following variables:

Sales → Latest one-month sales total (thousand dollars)
HS →Percentage of customer base with a high school diploma

Summary of data is given as:
$\sum Y = 38079.96, \quad \sum X = 1856.26, \quad \sum Y^2 = 81951358.74, \quad \sum X^2 = 145143.23,$
$\sum XY = 3057339.42$

Based on this information, **answer the next <u>four</u> questions**:

Q11: Assuming a linear relationship, use the least-squares method to compute the regression coefficient b_1.

(A) 71.2798
(B) -3926.4117
(C) 112076.6471
(D) 0.014
(E) 0.005

Q12: We want to test if there is evidence of a linear relationship between the independent variable and the dependent variable. What is the computed value of T-test statistic?

(A) 3.6025
(B) 71.2798
(C) 0.6091
(D) -3926.4117
(E) 2.074

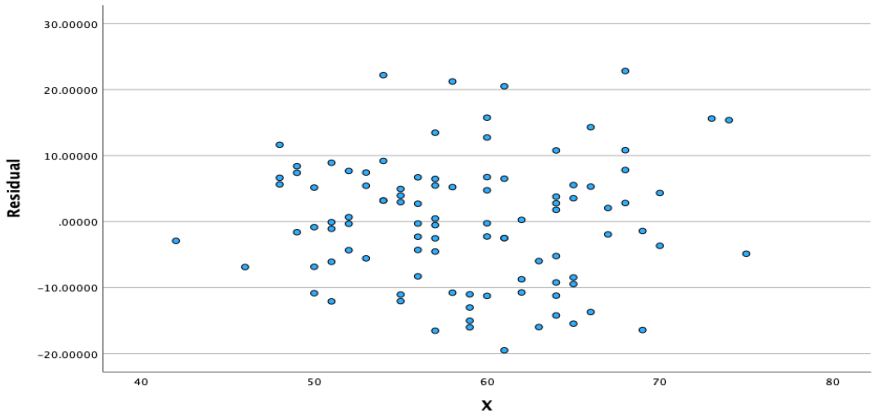
Q13: We want to test if there is evidence of a linear relationship between the independent variable and the dependent variable. At 0.01 level of significance, the null hypothesis is rejected if

(A) $|t_{STAT}| > 2.819$
(B) $|t_{STAT}| > 2.508$
(C) $|t_{STAT}| > 2.074$
(D) $|t_{STAT}| > 1.717$
(E) $|t_{STAT}| > 2.518$

Q14: Construct a 90% confidence interval estimate of the population slope. The lower limit is equal to

(A) 37.3069
(B) -600.9154
(C) 30.2432
(D) 45.1422
(E) 71.2798

Q15: (Q13.25) Zagat's publishes restaurant ratings for various locations in the United States. The data contains the Zagat rating for food, décor, service, and the cost per person for a sample of 100 restaurants located in New York City and in a suburb of New York City. We developed a regression model to predict the price per person, based on a variable that represents the sum of the ratings for food, décor, and service. For testing the assumptions of our regression model, a plot of residuals with X_i is constructed. Which one of the following statements best explain the graph?



(A) The assumption of constant variance is violated.
(B) The assumption of normality is violated.
(C) The assumption of linearity is violated.
(D) The assumption of independence is violated.
(E) There is no linear relationship between X and Y.

A critically important aspect of customer service in a supermarket is the waiting time at the checkout (defined as the time the customer enters the line until he or she is served). Data were collected during time periods in which a constant number of checkout counters were open. The total number of customers in the store (X) and the waiting times in minutes (Y) were recorded. Following summary statistics are available:

$n = 18, \quad \sum Y = 39.2, \quad \sum X = 379, \quad \sum Y^2 = 108.16, \quad \sum X^2 = 9367, \quad \sum XY = 983.7$

$SSX = 1386.944, \quad SSY = 22.7911, \quad SSXY = 158.3222, \quad SSE = 4.7107, \quad MSE = 0.2944$

$b_1 = 0.1142, \quad b_0 = -0.2268$

Based on this information, **answer the next <u>three</u> questions**:

Q16: (13.7c) Interpret the meaning of slope in this example.

(A) The estimated mean waiting time will increase by 0.1142 minutes for each additional customer.
(B) The estimated mean waiting time will decrease by 0.1142 minutes for each additional customer.
(C) The estimated mean waiting time is 0.1142 minutes for a store with no customer.
(D) 11.42% variation in waiting time is explained by the number of customers.
(E) The relationship between waiting time and number of customers is very weak.

Q17: (13.19a) Determine the coefficient of determination.

(A) 0.7933
(B) 0.8907
(C) 0.2944
(D) 0.5426
(E) 0.1142

Q18: (13.59a) Construct a 95% confidence interval estimate of the mean waiting time for all customers when there are 25 customers in the store. The critical value is equal to 2.12. What is the upper limit of constructed interval estimate?

(A) 2.9254
(B) 3.8163
(C) 2.6282
(D) 3.0317
(E) 2.3303

Q19: (Q13.69) Which one of the following is **not** an assumption of regression analysis?

(A) The variables X and Y are independent of each other.
(B) The relationship between variables is linear.
(C) The errors are independent of one another.
(D) The errors are normally distributed at each value of X.
(E) The variance of the errors be constant for all values of X.

Q20: (Q13.2) Fitting a straight line to a set of data yields the following prediction line:

$\hat{y}_i = 10.5 + 0.96X_i$

If the values of X range from 7 to 35, the fitted model should **not** be used to predict the mean value of Y when

(A) $X = 5$
(B) $X = 10$
(C) $X = 15$
(D) $X = 20$
(E) $X = 30$