KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
MATHEMATICS DEPARTMENT

CODE 000

STAT 212 BUSINESS STATISTICS II
Semester 223, Final Exam
August 17, 2023

CODE 000

Time allowed 150 minutes

Name: _____ ID #:_____

Section #: _____ Serial #:_____

Important Instructions:

Check that this exam has 24 questions

- All types of mobile phones or smart watches are NOT allowed during the examination.
- Use HB 2.5 pencils only.
- Use a good eraser. DO NOT 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 answer sheet.
- When bubbling your ID number and Section number, be sure that the bubbles match with the numbers that you write.
- The Test Code Number is already bubbled in your answer sheet. 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
- Formula sheet will be provided to you in exam. You are not allowed to bring, with you, formula sheet or any other printed/written paper.

The marketing manager of a large supermarket chain faced the business problem of determining the effect on the sales of pet food of shelf space (X_1) and whether the location of the product (X_2) was placed at the front $(X_2 = 1)$ or back $(X_2 = 0)$ of the aisle. Data are collected from a random sample of 12 equal-sized stores. The results of the regression model and the ANOVA are shown in the following tables:

Analysis of Variance

Source	DF	SS	MS	F-Value	P-Value
Regression	2		12967.5		0.000
Error	9	4090			
Total		30025			

Coefficients

Term	Coef	SE Coef	T-Value	P-Value
Constant		15.7	8.29	0.000
x1		1.10	6.72	
x2	45.0	13.1		

Using the data above, answer the following **6** questions:

1. What is the interpretation of the coefficient 45?

   A. When the product is placed at the front of the aisle, the average sales is 45$ more than the sales when the product is placed at the back of the aisle, given that the shelf space is kept constant.
   B. When the location of the product increases by one feet, the average sales hours increases by 45$, given that the shelf space is kept constant.
   C. When the product is placed at the back of the aisle, the average sales is 45$ more than the sales when the product is placed at the front of the aisle, given that the shelf space is kept constant.
   D. When the product is placed at the front of the aisle, the average sales is 45$ more than the sales when the product is placed at the back of the aisle.
   E. When the location of the product decreases by one feet, the average sales hours increases by 45$, given that the shelf space is kept constant.

2. The predicted sales for the product when the shelf space is 8.5 feet, and it is placed at the front is:

   A. 237.99
   B. 192.99
   C. 182.99
   D. 229.99
   E. 259.99

3. Is it reasonable to drop the dummy variable from model?

A. No, because X_2 is significantly related to Y since the $p-value < 0.01$.
B. Yes, because X_2 is significantly related to Y since the $p-value < 0.01$.
C. No, because X_2 is not significantly related to Y since the $p-value < 0.01$.
D. Yes, because X_2 is not significantly correlated to X_1 since the $p-value < 0.05$.
E. No, because X_2 is significantly correlated to X_1 since the $p-value < 0.05$.

4. The percentage of variation in the sales that is explained by the variation of the shelf space and the location, adjusted to the sample size and the number of variables is:

A. 83.35%
B. 86.38%
C. 97.45%
D. 90.13%
E. 80.45%

5. At 0.025 level of significance, to test the significance of the shelf space to the regression model. The p-value is:

A. $p-value < 0.01$
B. $0.01 < p-value < 0.025$
C. $0.025 < p-value < 0.05$
D. $0.05 < p-value < 0.1$
E. $0.1 < p-value < 0.15$

6. Given that $SSR_{(x1)}$ =20526 and $SSR_{(x2)}$ = 5400, the proportion of variation in the weekly sales is explained by the location of the product while controlling for (holding constant) the shelf space.

A. 56.9%
B. 86.4%
C. 68.4%
D. 17.9%
E. 83.4%

You constructed simple linear regression models to investigate the relationship between monthly sales for a chain of sporting goods stores and five independent variables.
Based on the following MINITAB output

Best Subsets Regression: y versus x1, x2, x3, x4, x5

Response is y

Vars	R-Sq	R-Sq (adj)	R-Sq (pred)	Mallows Cp	S	x1	x2	x3	x4	x5
1	24.1	21.9	17.2	-1.8	802004				X	
1	14.7	12.3	4.0	2.1	849860		X			
2	24.3	20.0	8.7	0.0	811809	X			X	
2	24.2	19.9	14.0	0.1	812572				X	X
3	24.4	17.8	5.4	2.0	823193	X			X	X
3	24.4	17.7	0.0	2.0	823626	X	X		X	
4	24.4	15.3	0.0	4.0	835508	X		X	X	X
4	24.4	15.3	0.0	4.0	835563	X	X		X	X
5	24.4	12.6	0.0	6.0	848433	X	X	X	X	X

Correlation: y, x1, x2, x3, x4, x5

	y	x1	x2	x3	x4
x1	0.041				
	0.805				
x2	0.112	-0.203			
	0.502	0.222			
x3	0.384	0.387	0.032		
	0.017	0.016	0.848		
x4	0.490	0.193	0.217	0.804	
	0.002	0.247	0.192	0.000	
x5	0.322	0.265	0.381	0.682	0.712
	0.049	0.108	0.018	0.000	0.000

Cell Contents: Pearson correlation
P-Value

solve the next 4 questions

7. If only one variable were be brought into the model, which variable should be if the goal is to explain the highest possible percentage of variation in the dependent variable?

   A. X_4 has the highest correlation with monthly sales.
   B. X_1 has the highest correlation with monthly sales.
   C. X_2 has the highest correlation with monthly sales.
   D. X_5 has the highest correlation with monthly sales.
   E. X_3 has the highest correlation with monthly sales.

8. Using R^2_{adj}, the model should be taken into consideration contains

   A. 1 variable .
   B. 4 variables.
   C. 2 variables.
   D. 3 variables.
   E. 5 variables.

9. If you are going to fit a regression model using the backward elimination selection method, what is the first predictor to be removed

   A. X_1
   B. X_2
   C. X_4
   D. X_5
   E. X_3

10. What can you say about the multicollinearity between the independent variables

A. The correlation between some of the independent variables is high and significant. there is highly potential multicollinearity problem
B. The correlation between the independent variables is low and not significant. there is no potential multicollinearity problem
C. The correlation between the independent variables is high and not significant. there is no potential multicollinearity problem
D. The correlation between the independent variables is low and not significant. there is potential multicollinearity problem
E. We need more information to tell.

11. Which of the following would be an appropriate alternative hypothesis?

A. The population proportion is less than 60 percent.
B. The sample proportion is less than 60 percent.
C. The population proportion is no less than 60 percent.
D. The sample proportion is no less than 60 percent.
E. The population proportion is no more than 60percent.

12. When testing $H_0: \mu_1 - \mu_2 \leq 0 \; vs \; H_A: \mu_1 - \mu_2 > 0$, the observed value of z – score was 2.52. at 10% level of significance, the p – value for the test would be

A. 0.0059
B. 0.9941
C. 0.0118
D. 0.0500
E. 0.9882

13. If the coefficient of determination between two independent variables is 75%, then the VIF equal to

A. 4
B. 5
C. 3
D. 2
E. 1

Betts Electronics purchases three replacement parts for robotic machines used in its manufacturing process. Information on the price of the replacement parts and the quantity purchased is given below:

	PRICE		QUANTITY	
PART	2000	2016	2000	2016
A	$ 2.00	$ 5.50	120	220
B	$ 2.50	$ 6.00	90	110
C	$ 4.00	$ 10.00	80	95

Using the above information to answer the following **4** questions:

14 What is the simple price index for part B in 2016, using 2000 as a base year:

A. 240%
B. 86.54%
C. 120%
D. 41.66%
E. 48.96%

15. The aggregate price index for 2016, using 2000 as a base year is equal to:

A. 252.94%
B. 39.53%
C. 86.45%
D. 210.96%
E. 198.78%

16. The Paasche price index for 2016 using 2000 as the base period is equal to:

A. 257.53%
B. 38.83%
C. 39.25%
D. 254.78%
E. 123.53%

17. The Laspeyres price index for 2016 using 2000 as the base period is equal to:

A. 254.78%
B. 257.53%
C. 38.83%
D. 39.25%
E. 270.45%

18. An entrepreneur is considering the purchase of a coin-operated laundry. The current owner claims that over the past 5 years, the average daily revenue was $675 with a standard deviation of $75. A sample of 30 days reveals a daily average revenue of $625. If you were to test the null hypothesis that the daily average revenue was $675 and decide not to reject the null hypothesis, what can you conclude?

A. There is evidence to conclude that the daily average revenue was $675.
B. There is not evidence to conclude that the daily average revenue was $675.
C. There is not evidence to conclude that the daily average revenue was not $675.
D. There is evidence to conclude that the daily average revenue was not $675.
E. Inconclusive, we cannot tell.

19. If, as a result of a hypothesis test, we reject the null hypothesis when it is false, then we have committed?

A. No error
B. Type I error
C. Type II error
D. An acceptance error
E. Type I error or Type II error

20. If a hypothesis test for a single population variance is to be conducted using a significance level of 0.10, a sample size of 16, and the test is a one-tailed upper-tail test, the critical value is

A. 22.307
B. 23.542
C. 1.28
D. 1.3406
E. 1.7531

An economist develop a multiplicative time-series model to forecast the petrol prices in the USA in future quarters, using quarterly data of petrol prices (in dollars per gallon) from January 2006 to April 2010. The following is the resulting regression equation:

$$\log \hat{y} = 2.623 + 0.0081\, X_i - 0.226\, Q1 + 0.233\, Q2 + 0.436\, Q3$$

Where:

- $\hat{y}$: *is the estimated price of petrol in quarter*
- X_i: *is the code quarterly value with* $X = 0$ *in the first quarter of* 2006
- $Q1$: *dummy variable equal to* $\mathbf{1}$ *in the first quarter of a year and* $\mathbf{0}$ *other wise*
- $Q2$: *dummy variable equal to* $\mathbf{1}$ *in the second quarter of a year and* $\mathbf{0}$ *other wise*
- $Q3$: *dummy variable equal to* $\mathbf{1}$ *in the third quarter of a year and* $\mathbf{0}$ *other wise*

Using the data above answer the following **4** questions:

21. The estimated quarterly compound growth rate in the price is:

A. 1.88%
B. 4.19%
C. 0.09%
D. 1.28%
E. 2.44%

22. The best interpretation of the coefficient of Q1 (-0.226) in the regression equation is:

A. The prices in the 1st quarter of a year is approximately 41% lower than it would be during the 4th quarter.
B. The prices in the 1st quarter of a year is approximately 59% less than it would be during the 4th quarter.
C. The prices in the 1st quarter of a year is approximately 41% higher than the average over all 4 quarters.
D. The prices in the 1st quarter of a year is approximately 59% less than the average over all 4 quarters.
E. The prices in the 1st quarter of a year is approximately 59% higher than it would be during the 4th quarter.

23. To obtain a forecast for the **third** quarter of 2007 using the model, which of the following sets of values should be used in the regression equation?

A. $X = 6, Q1 = 0, Q2 = 0, Q3 = 1$
B. $X = 7, Q1 = 0, Q2 = 0, Q3 = 1$
C. $X = 3, Q1 = 0, Q2 = 0, Q3 = 1$
D. $X = 10, Q1 = 1, Q2 = 0, Q3 = 1$
E. $X = 12, Q1 = 1, Q2 = 0, Q3 = 1$

24. Using the regression equation, the forecast for the prices in the third quarter of 2010 is:

A. 1602.51$
B. 1164.18$
C. 707.98$
D. 1005.38$
E. 1250.09$