

# RMIT

# End of Semester Test

## Faculty of Business School of Marketing

**PR220: Data Analysis 2 — 1st Semester 2000**

**This test counts for 15% of the total marks in PR220**

### INSTRUCTIONS

1. Clearly print your full name and student number in the space provided below.
2. **All questions must be answered.**  
The test consists of 10 multiple choice questions and **2 short answer questions**.  
For the short answer questions all workings should be shown.  
You are required to do either **Question 1A** or **Question 1B** but **not both**.
3. Calculators are allowed.
4. All questions must be completed in the booklet that forms part of this test paper.
4. Stationary supplied: — test questions and booklet  
normal distribution table  
t - distribution table  
F - distribution table  
Critical values of the rank correlation coefficient,  $r_s$

Student name.....

Student number.....

**Time Allowed 1.5 hours**

**This test paper must not be removed from the examination centre, and should be given to your lecturer or examination supervisor at the end of the test.**

**Total marks all questions for this test = 32**

## Multiple Choice Questions

(10 marks)

In each case, circle the **response** you believe most accurately answers the question.

**The following information relates to questions 1 and 2.**

The mean success rate of an auctioneer in Melbourne during any month is 24 sales with a standard deviation of 6. Assume that the auctioneer's performance may be described by a normal distribution.

1. The probability that he successfully auctions up to 30 properties in a given month is:
  - A. 0.5
  - B. 0.75
  - C. 0.8413
  - D. 0.9321
  
2. If the probability of successfully auctioning  $x$  properties or more is 0.758, what is the number of properties sold (to the nearest whole number):
  - A. 20
  - B. 26
  - C. 28
  - D. 34
  
3. The t-distribution is formed from the ratio of:
  - A. The normal distribution to the F-distribution.
  - B. The ratio of two independent chi-square distributions
  - C. The ratio of the standard normal distribution to the chi-squared distribution.
  - D. None of the above.
  
4. The presence of multicollinearity in a regression equation:
  - A. means that the equation may not be used for any purpose
  - B. may cause the coefficients and their t-ratios to be inflated.
  - C. may cause the coefficients to have the incorrect signs and render tests on the coefficients incorrect.
  - D. it is a data problem therefore, it does not effect the model.
  
5. In evaluating a regression equation which of the following should be adopted:
  - A. The adjusted  $R^2$  should be close to 1, the t-ratios should be positive and greater than 2, and the standard error of the estimate close to 0.
  - B. The F statistic and t-ratios should be significant at the chosen level of significance, the coefficients should possess the expected signs and the least squares assumptions should not be violated.
  - C. The signs and magnitudes of the regressions coefficients should be as expected, the F statistic should be significant and the adjusted  $R^2$  close to 1.
  - D. All of the above.
  
6. Two variables are said to be inversely related if:
  - A. The rate of change in one variable is less than the change in the other variable.
  - B. The rate of change in one variable is the same as the change in the other variable.
  - C. The rate of change in one variable decreases while the change in the other variable increases.
  - D. The rate of change in one variable decreases while the change in the other variable also decreases but at a different rate.

7. The Central Limit Theorem (CLT) is one of the most important theorems in statistics. Select the most appropriate response from among the following:
- A. The sampling distribution of the mean may be approximated by a normal distribution, for large samples, regardless of the distribution of the parent population.
  - B. If the population is large but not normally distributed, the distribution of sample means approaches a normal distribution provided that the sample is large.
  - C. "It has been established mathematically that the normal distribution may be used as a basis for approximating the sampling distribution of the mean of  $\bar{X}$ ." This fact alone makes the normal distribution so fundamentally important in statistics.
  - D. All of the above.
8. Which of the following statements best describe the relationship between the normal, Chi-square, t and F- distributions:
- A. The normal, Chi-square, t and F distributions are formed from standardised values of the sampling distribution of the random variable  $X$ , when the sample size is at least 30.
  - B. The Chi-square is a continuous distribution which is formed from the sum of squares of independent standard normal variables. The t-distribution is formed from the ratio of the standard normal to the Chi-square while the F distribution is formed from the ratio of two independent Chi-square distributions.
  - C. The t-distribution is identical to the normal when the degrees of freedom are at least 30. The Chi-square depends entirely on degrees of freedom while the F distribution is not affected by degrees of freedom.
  - D. All of the above.
9. Which of the following statements are correct regarding heteroscedasticity :
- A. The presence of heteroscedasticity gives rise to the standard errors of the coefficients being incorrect, which may lead to falsely accepting a hypothesis when in fact it should be rejected.
  - B. Heteroscedasticity is essentially theoretical in nature and does not lead to problems in regression models when real data is used for the analysis.
  - C. Heteroscedasticity causes the standard error of the estimate to take on different values for different observations in the model and hence has no impact on the model overall.
  - D. Heteroscedasticity reduces the  $R^2$  and F statistic below what they would otherwise be causing the model overall to be insignificant.
10. Which of the following statements about regression analysis is/are correct:
- A. Regression may be regarded as the decomposition of variance, represented by the component explained by the independent variables and the unexplained component containing the error.
  - B. The regression equation represents the conditional expectation of  $Y$  for given values of the  $X_i$ .
  - C. Ordinary Least Squares (OLS) regression is the most robust technique to estimate parameters in the presence of error, provided that the OLS assumptions are not violated.
  - D. All of the above.

**Short Answer Questions**

**Show all workings where necessary**

**Do either Question 1A or 1B but not both**

**Question 1A.**

**(7 + 3 = 10 marks)**

A claim being made by shopping centre tenants is that within certain limits sales revenue per square metre of floor space is inversely related to the size of the floor space for each shop. If this is the case then rentals on the larger shops should be reduced.

- (i) calculate a rank-correlation coefficient
- (ii) test the coefficient's significance.

Data required have been collected randomly from various shops in a large centre.

Shop ID	Area (sq metres)	Revenue per Sq. Metre (\$)
S27	2156	542
N66	687	636
N34	728	700
E74	539	750
S59	2762	500
S56	3023	484
E01	585	744
N28	930	301
E42	611	670

$$r_s = 1 - \frac{6 d^2}{n(n^2 - 1)}$$

**Do either Question 1A or 1B but not both**

**Question 1B.**

**(6 + 4 = 10 marks)**

A residential property manager classifies tenants according to those who are on 6-months lease and those who have leases for periods longer than 6 months. The manager's data base revealed that 340 tenants were on 6-month leases and had a probability of 25% of being late with their rent.

- (i) What is the proportion of tenants on 6-month leases who pay on time, if the total number of tenants with leases longer than 6 months is 560 ?
- (ii) If the number of late payers is 408, what is the joint probability of a tenant who has a lease greater than 6-months being a late payer ?

**Question 2.**

**(6 + 4 + 2 = 12 marks)**

RENT BREAKDOWN: We are attempting to set rentals in a project consisting of small warehouse and unserviced office units. Rentals for a very similar development, and the respective office and warehouse areas, measured in square feet, were collected to develop a model to estimate rentals.

$Y$  = annual rent,  $X_1$  = office area,  $X_2$  = warehouse area

- (i) From the two models provided below, **Model 1** and **Model 2**, select the model (**one only**), providing an explanation for your choice, you believe should be used to set rentals.
- (ii) Provide an explanation for the inclusion of the following variables, stating what you believe should be the expected signs and magnitudes of their respective coefficients:

in Model 1:  $X_1^2$ ,  $X_2^2$       in Model 2:  $X_1X_2$

The following two models were estimated, using Excel.

**Model 1**

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>		
Intercept	-1720.4716	958.3141	-1.80	Observations	20
$X_1$	12.4450	1.8169	6.85	$S_e$	660.5
$X_2$	2.5204	0.5816	4.33	R Square	0.829
$X_1^2$	-0.0072	0.0009	-7.55	Adj. R Square	0.784
$X_2^2$	-0.0003	0.0001	-2.68	F	18.216

**Model 2**

	<i>Coefficients</i>	<i>Standard Error</i>	<i>t Stat</i>		
Intercept	1190.7955	1017.2916	1.17	Observations	20
$X_1$	5.6992	1.8715	3.05	$S_e$	1022.
$X_2$	2.0269	0.4844	4.18	R Square	0.564
$X_1X_2$	-0.0034	0.0009	-3.76	Adj. R Square	0.482
				F	6.886

- (iii) For **Model 2** test the hypothesis that the interaction term ( $X_1X_2$ ) is not significant in explaining warehouse rents, use a 1% level of significance.