**CONCORDIA UNIVERSITY BUSINESS STATISTICS
 COMM 215 ALL SECTIONS**

**FINAL EXAMINATION
FALL 2014**

|  |  |  |
| --- | --- | --- |
| **Part** | **Topic** | **Total Marks** |
| 1 | Multiple Choice | 58 |
| 2 | Long Answer | 42 |
|  | Total | 100 |

**PART I: Multiple choice questions (One mark for each of questions 1 to 14 and two marks for each of questions 15 to 36). Some of the numbers in the provided choices have been rounded.**

**Indicate your answers on the answer sheet provided.** Use **pencil** to make black marks that fill the circle completely. Erase cleanly any answer you wish to change. Make no stray marks on the answer sheet.

**1.** In sampling when the sample size is increased,

a) the potential for extreme sampling error is reduced.
b) the amount of sampling error remains the same.
c) the population standard deviation will increase.
d) There is no specific relationship between sample size and sampling error.

**2.** A population, with an unknown distribution, has a mean of 80 and a standard deviation of 7. For a sample size of 49, the probability that the sample mean will be larger than 82 is:

a) 0.5228
b) 0.9772
c) 0.4772
d) 0.0228

**3.** If an economist wishes to determine whether there is evidence that average family income in a community exceeds $25,000. The best null hypothesis is:

a) μ = 25,000
b) μ > 25,000
c) μ ≤ 25,000
d) μ ≥ 25,000

**4.** A hypothesis test is to be conducted using an α = .05. This means,

a) there is a 5 percent chance that the null hypothesis is true.
b) there is a 5 percent chance that the alternative hypothesis is true.
c) there is a maximum 5 percent chance that a true null hypothesis will be rejected.
d) there is a 5 percent chance of not rejecting a false null hypothesis.
e) None of the above.

**5.** For a given sample size *n*, if the level of significance α, is decreased, the power of the test

a) will increase.
b) will decrease.
c) will remain the same.
d) cannot be determined.

**6.** To test the fairness of a die a researcher tossed the die 100 times and observed the outcomes. The degrees of freedom for the test is:

a) 99
b) 6
c) 5
d) 1
e) None of the above

Every spring semester, the School of Business coordinates a luncheon with local business leaders for graduating students, their families, and friends. Corporate sponsorship pays for the lunches of all students, but students have to purchase tickets to cover the cost of lunches served to guests they ring with them. The following graph represents the attendance at the student luncheon, where X axis is the number of guests each graduating student invited to the luncheon and the Y axis is the number of graduating students in each category. Use the following diagram to answer questions **7** to **9**. Note to student: the X axis goes up by 1.

**7.** How many graduating seniors attended the luncheon?

a) 388
b) 4
c) 275
d) 152

**8.** What percentage of students invited at least 2 guests?

a) 30.91%
b) 7.64%
c) 61.45%
d) 38.55%
e) None of the above

**9.** How many guests attended the luncheon?

a) 275
b) 388
c) 663
d) None of the above

**10.** The power of a test is measured by its capability of,

a) rejecting a null hypothesis that is true.
b) not rejecting a null hypothesis that is true.
c) rejecting a null hypothesis that is false.
d) not rejecting a null hypothesis that is false.
e) None of the above.

**Refer to the following in answering questions 11 to 14.**

“Colour Association of Male and Female School Children”, where children were asked to indicate what emotion they associated with the colour red. The response and the sex of the child are noted and summarized below. The first number in each cell is the count (frequency), the second number is the row percent.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **Anger** | **Happy** | **Love** | **Pain** | **Total** |
| **Female** | 2726.47% (= 27/102) | 1918.63% | 3938.24% | 1716.67% | 102 |
| **Male** | 3430.36% (= 34/112) | 1210.71% | 3833.93% | 2825.00% | 112 |

**Total** 61 31 77 45 214

**11.** Under a suitable null hypothesis, the expected frequency for the cell corresponding to Anger and Males is:

a) 15.9
b) 55.7
c) 30.4
d) 31.9
e) 29.1

**12.** The null hypothesis will be rejected at α=0.05 if the test statistic exceeds:

a) 3.84
b) 5.99
c) 7.81
d) 9.49
e) 14.07

**13.** Which of the following is NOT CORRECT?

a) The children were cross-classified by gender and emotion associated with red. Each child was counted in one and only one cell.
b) The null hypothesis is that the type of emotion associated with red is independent of the sex of the child.
c) The null hypothesis is that the proportion of emotions associated with red is the same for both sexes.
d) All expected cell counts should be greater than five in order to use the chi-square test statistic.
e) If we reject the null hypothesis then we have proven that there is no relation between gender and emotion associate with the colour red.

**14.** Which of the following is NOT CORRECT?

a) A lower percentage of female students associate the emotion “anger” with the color red than do male students.
b) More students associate the color red with the emotion “love” than with the emotion “anger”.
c) Each student was classified by gender and by emotion association.
d) We will be unable to computer a correlation for this data because both variables are not interval or ratio scale.
e) We compute row or column percentages by dividing the cell count by the table total (214).

**15.** When someone is on trial for suspicion of committing a crime, the hypotheses are:
 *H0*: innocent
 *HA* : guilty
 Which of the following is correct?

a) Type I error is acquitting a guilty person.
b) Type I error is convicting an innocent person.
c) Type II error is acquitting an innocent person.
d) Type II error is convicting an innocent person.

**16.** A multiple regression is estimated using x1, x2, x3. A second multiple regression was estimated using x1, x2, x3, x4. Which regression will have higher F-statistic?

a) Not enough information to answer this question.
b) The first regression which uses 3 independent variables.
c) The second regression which uses 4 independent variables.
d) Both will have the same F-statistic.
e) None of the above.

**17.** A population has a mean of 75 and a standard deviation of 8. A random sample of 800 is selected. The expected value of x̄ is:

a) 8
b) 75
c) 800
d) None of these alternatives is correct.

**18.** The sampling error is the,

a) same as the standard error of the mean.
b) difference between the value of the sample mean and the value of the population mean.
c) error caused by selecting a bad sample.
d) standard deviation multiplied by the sample size.

**19.** A theorem that allows us to use the normal probability distribution to approximate the sampling distribution of sample means and sample proportions whenever the sample size is large is known as the,

a) approximation theorem.
b) normal probability theorem.
c) central limit theorem.
d) central normality theorem.

**20.** In a regression and correlation analysis if r2 = 1, the

a) SSE must also be equal to one.
b) SSE must be equal to zero.
c) SSE can be any positive value.
d) Not enough information to answer this question.

**21.** Regression analysis was applied between demand for a product (Y) and the price of the product (X), and the following estimated regression equation was obtained.

 Ŷ = 120 – 10X

Based on the above estimated regression equation, if price is increased by 2 units, then demand is expected to,

a) increase by 120 units.
b) increase by 100 units.
c) increase by 20 units.
d) decrease by 20 units.
e) None of the above.

**Refer to the following in answering questions 22 to 24.**
Regression analysis was applied between sales data (in $1,000s) and advertising data (in $100s) and the following information were obtained.

ŷ = 1.2 + 1.8x
n = 17
SSR = 225
SSE = 75
Sb1 = 0.2683

**22.** The t statistic for testing the significance of the slope is close to:

a) 4.473
b) 3
c) 6.708
d) 0.555
e) None of the foregoing.

**23.** The F statistic computed from the above data is:

a) 3
b) 45
c) 48
d) 50
e) None of the above

**24.** The critical value of the F statistic at 0.05 level of significance is:

a) 4.600
b) 4.451
c) 3.682
d) 4.543
e) None of the above.

**25.** The standard error of the mean for a sample of 100 is 30. In order to cut the standard error of the mean to 15, we would,

a) decrease the sample to 25.
b) decrease the sample size to 50.
c) increase the sample size to 200.
d) increase the sample size to 400.

**26.** Which of the following is NOT a reason for the need for sampling?

a) It is sometimes destructive to observe the entire population.
b) It is always more informative by investigating a sample than the entire population.
c) It is usually too time consuming to look at the whole population.
d) It is usually too costly to study the whole population.

**27.** Researchers believe that the number of tissues used during a cold by tissue users is less than 60. Give the null and alternate hypotheses for testing this belief.

a) H0: x̄ = 52 and HA: x̄ ≠ 52.
b) H0: μ ≤ 60 and HA: μ > 60.
c) H0: x̄ ≥ 60 and HA: x̄ < 60.
d) H0: μ ≥ 60 and HA: μ < 60.

**28.** The least squares method minimizes which of the following?

a) SSE
b) SSR
c) SST
d) all of the above

**Refer to the following in answering questions 29 and 30.**

In a popular restaurant located in a large metropolitan city, a 90% confidence interval for the proportion of reservations being cancelled on a busy Saturday night is: (0.29. 0.66).

**29.** What is the margin of error in this estimation?

a) 0.285
b) 10%
c) 0.185
d) 0.475
e) None of the above.

**30.** What is the standard error of the proportion of reservations that are cancelled?

a) 0.1985
b) 0.1125
c) 0.3301
d) 0.0943

**31.** For a fixed sample size, when we interpret a 95% confidence interval for μ, what does the phrase 95% confident mean?

a) 95% of similarly constructed intervals would contain the value of the sampled mean.
b) In repeated sampling, 95% of similarly constructed intervals contain the value of the population mean.
c) 95% of the observation in the population fall within the bounds of the calculated interval.
d) The probability that the sample mean falls in the calculated interval is 0.95.

**Refer to the following exhibit in answering questions 32 to 35.**

n = 51 x̄ = 54.8 s = 28 H0: μ ≤ 50
 HA: μ > 50

**32.** The test statistic is:

a) 0.1714
b) 0.3849
c) -1.2
d) 1.2

**33.** The *p*-value of the above test is between,

a) 0.01 and 0.025
b) 0.025 and 0.05
c) 0.05 and 0.1
d) 0.1 and 0.15
e) None of the above.

**34.** If the level of significance of a hypothesis test is changed from .01 to .05, the probability of a Type II error,

a) will also change from .01 to .05
b) will not change.
c) will decrease.
d) will increase.

**35.** Which of these are most accurate? In multiple regression model building with Y as the dependent variable and X1, X2, X3 as dependent variables.

a) Multiple R is equivalent to the square of the correlation coefficient.
b) R square measures the extent to which the independent variables collectively explain variation in Y.
c) A high R square is a great indicator that each of the independent variable make a significant contribution to explaining in Y.
d) All of the above.

**36.** In a multiple regression analysis if the confidence intervals for the two population coefficients *β*1 *and* *β*2 contain 0 that means,

a) That the two coefficients are significantly different from zero.
b) The two variables associated with the coefficients are excellent at explain the dependent variable.
c) That the two coefficients are not significant.
d) All of the above.

**Question 1 (11 marks total)**

**A.** (6 marks)
 A new computer chip has a mean clock speed of 210 Mhz and a standard deviation of 21 Mhz. Management wants to make sure that the product line is in control and does not produce chips whose average speed is significantly less than 210 Mhz. For this purpose a sample of 9 chips is taken and their average speed is calculated to be 200 Mhz.

a) Is the process in control? Use α = 0.05 and state your hypothesis.

b) State any assumption(s) you may have to make for answering (a).

c) Calculate the p-value for the above test.

**B.** (5 marks)
 A poll was conducted by the marketing department of a video game company to determine the popularity of a new game that was targeted to be launched in three months. Telephone interviews with 1,500 young adults were conducted and 735 of them said that they would purchase the new game.

a) Construct a 99% confidence interval for the percentage of the targeted adults who said that they would purchase the new game and report the margin of error associated with this estimate.

b) What minimum sample size is needed to obtain a 95% confidence interval estimate of the above parameter with a margin of error of 3%?

**Questions 2 (8 marks)**

It is believed that a generation gap exist in music. To test this a study is conducted on three different age groups. The study shows that 47% of the 15 – 28 year olds, 45% of the 29 – 48 year olds and 30% of the 40 – 65 year olds listen to a fusion of jazz and rock. The study was based on 500 respondents in each age group, a total of 1500 individuals.

a) Is there sufficient evidence that a significant difference exist among the age groups with respect to the proportions who listen to the fusion of rock and jazz? Use α = 0.05.

b) Would you say that age is independent of music preference?

c) What is the p-value of the above test? Interpret it.

**Question 3 (12 marks)**

An experiment was conducted to study the relationship between the sales performance in the first year and the average sales performance of the subsequent four years of salesperson in a company. The data are given below where:
x = First year sales in millions of dollars,
y = Average annual sales in the four subsequent years in millions of dollars.

(x) 104.7, 81.5, 60.5, 71.2, 62.6, 65.5, 102.8, 86.4, 101.7, 93.0, 89.1, 86.4

(y) 130.0, 95.0, 71.9, 86.2, 77.5, 80.0, 155.0, 139.0, 98.5, 105.0, 116.7, 109.0

You are given the following summary:

Σx = 1005.4 Σy = 1263.8

Σx2 = 86971.5 Σy2 = 140504.4 Σxy = 109474.01

a) Write the sample linear regression equation and interpret the slope of the regression equation in the context of this problem.

b) Compute and interpret the coefficient of determination.

c) Test for the significance of the linear relationship between the two variables using a 5% level of significance.

d) Set up 90% prediction interval for the average annual sales in the four subsequent years for a salesman who’s fiscal year sales is 80 million dollars.

**Question 4 (11 marks)**

The annual percent of cars stolen (Y), and the annual percent of people unemployed (X1) are collected for 21 years, X2 is a time trend which takes the value 1 for the first year, 2 for the second year and 21 for the last year. Here are the partial Excel regression results where both X1 and X2 are included as independent variables.

Multiple R 0.5802
R Square
Adjusted R Square
Standard Error 2.9327
Observations 21

**ANOVA**
df SS MS F

Regression 78.56

Residual

Total

**Estimates
 Coefficients Standard Error t Stat P-value**

**Intercept** 6.7131 2.1117 **Unemployment Rate** 1.6188 0.5366 **Time trend** -0.4115 0.05

a) Write out the regression equation and interpret the coefficients for both unemployment and the time trend.

b) Using the above results estimate the theft rate in the third year given the unemployment rate was 5%.

c) Compute the R2 and interpret it.

d) Determine if the overall model is significant. Use significance level of 0.05.

e) Does the unemployment rate play a significant role in this model?