YORK UNIVERSITY
School of Administrative Studies
AP/ADMS 2320.3.0 - Quantitative Methods 1
FALL 2009 - FINAL EXAMINATION – Part 1

December 12, 2009 TWO PARTS: 2PM-5PM

H. Bartel Section D ☐ Section E ☐ Section F ☐ Section G ☐ Section K ☐

P. Ng Section A ☐ Section H ☐ Section I ☐ Section J ☐

R. Huang Section C ☐ Section L ☐

M. Rochon Section B ☐

INSTRUCTIONS:

1. **You must write this exam in the section in which you are enrolled.**
2. **This is a closed book examination.**
3. Record Multiple Choice answers **in pencil** on the Scantron sheet provided.
4. Aids allowed: writing utensils, eraser, silent, **non-programmable** calculator, English paper-based dictionary and a two sided reference/formula sheet, not to exceed 8½" by 11", of your own construction. No restrictions on content. No flaps; all pasted or taped on pieces must be properly glued or taped down on all four sides.
5. Students found with any of the following will be deemed to have unauthorized materials as defined in the University Policy on Academic Honesty:
   - cellphones or other communication devices,
   - personal information managers (PIMs),
   - electronic dictionaries,
   - user-programmable calculators
   - pencil cases, unless a clear plastic bag
6. **YOU MUST HAND IN THIS EXAM BEFORE LEAVING THE EXAMINATION ROOM. FAILURE TO DO SO WILL RESULT IN AN AUTOMATIC F GRADE.**

_____
Student's Name (Print)

_____
Student Number

_____
Signature

THIS PART CONTAINS 10 PAGES. DO NOT REMOVE ANY PAGES

MULTIPLE CHOICE – ***Answer on Scantron Sheet***
(ONLY use PENCIL, fill in LAST NAME, FIRST NAME, STUDENT NUMBER, Answers)

1. When conducting a hypothesis test about a single mean, at a given level of significance, as the sample size n increases, the probability of a Type II error:
   a. Will decrease.
   b. Will increase.
   c. May increase or decrease
   d. Remains the same

2. Researchers determined that 60 Kleenex tissues is the average number of tissues used during a cold. Suppose a random sample of 100 Kleenex users yielded the following data on the number of tissues used during a cold: $\bar{x} = 52$ and $s = 22$. Suppose the alternative we wanted to test was H_1: $\mu < 60$. The correct rejection region for $\alpha = 0.05$ is:
   a. reject H_0 if $t > 1.660$.
   b. reject H_0 if $t < -1.660$.
   c. reject H_0 if $t > 1.984$ or $Z < -1.984$.
   d. reject H_0 if $t < -1.984$.

3. A random sample of 25 observations is selected from a normally distributed population. The sample variance is 10. In the 95% confidence interval for the population variance, the upper limit is:
   a. 17.110
   b. 6.097
   c. 17.331
   d. 19.353

4. In a hypothesis test for the population variance, the hypotheses are H_0: $\sigma^2 = 100$ vs. H_1: $\sigma^2 \neq 100$. If the sample size is 15 and the test is being carried out at the 10% level of significance, the rejection region is:
   a. $\chi^2 < 6.571$ or $\chi^2 > 23.685$.
   b. $\chi^2 < 7.790$ or $\chi^2 > 21.064$.
   c. $\chi^2 < 8.547$ or $\chi^2 > 22.307$.
   d. $\chi^2 < 7.261$ or $\chi^2 < 24.996$.

5. As a manufacturer of golf clubs, a major corporation wants to estimate the proportion of golfers who are right-handed. How many golfers must be surveyed if they want to be within 0.02, with a 95% confidence?
   a. 2100
   b. 2401
   c. 1801
   d. Not enough information

6. There is little difference between the values of $t_{\alpha/2}$ and $z_{\alpha/2}$ when:
   a. The sample size is small
   b. The sample size is large
   c. The sample mean is small
   d. The sample mean is large

7. For a given hypothesis test, if we do not reject H_0, and H_0 is true.
   a. No error has been committed.
   b. Type I error has been committed.
   c. Type II error has been committed.
   d. Type III error has been committed.

8. A decision in a hypothesis test can be made by using a
   a. p-value
   b. Rejection region
   c. Either a or b
   d. None of the above

9. A ____ is the likelihood of a sample result assuming that the null hypothesis is true.
   a. Type II error
   b. Type I error
   c. Rejection point
   d. p-value

10. When carrying out a large sample test about a population proportion p where we are testing H_0: p = .4 versus H_1: p < .4 and z is the calculated test statistic, we reject H_0 at level of significance α when:
   a. $z < -z_{\alpha/2}$
   b. $z < -z_{\alpha}$
   c. p-value < α
   d. Both b and c

11. In testing for the equality of means from two independent populations, if the null hypothesis is rejected, the test could result in:
   a. A Type I error.
   b. Either a Type I error or a Type II error.
   c. Neither a Type I error or a Type II error.
   d. A Type II error

12. If we are testing the hypothesis about the mean of a population of paired differences with samples of n1 = 10, n2 = 10, the degrees of freedom for the t statistic is ____.
   a. 19
   b. 18
   c. 9
   d. 8

13. In the ANOVA, different levels of a factor are called
   a. Treatments
   b. Variables
   c. Responses
   d. Observations

14. When using one -way analysis of variance, the calculated F statistic will decrease when:
   a. The variability among the groups decreases relative to the variability within the groups
   b. The total variability increases
   c. The total variability decreases
   d. The variability among the groups increases relative to the variability within the groups

15. For the interval estimation of μ when σ is known and the sample is large, the proper distribution to use is
   a. the normal distribution
   b. the t distribution with n degrees of freedom
   c. the t distribution with n + 1 degrees of freedom
   d. the t distribution with n + 2 degrees of freedom

16. A random sample of 1000 people was taken. Four hundred fifty of the people in the sample favored Candidate A. The 95% confidence interval for the true proportion of people who favors Candidate A is
    a. 0.419 to 0.481
    b. 0.40 to 0.50
    c. 0.45 to 0.55
    d. 1.645 to 1.96

17. When using simple linear regression, we would like to use confidence intervals for the _____ and prediction intervals for the _____ at a given value of x.
    a. individual y-value, mean y-value
    b. Mean y-value, individual y-value
    c. Slope, mean slope
    d. y-intercept, mean y-intercept

18. Which of the following is not an example illustrating the use of variance?
    a. As a measure of risk.
    b. As a judge of consistency.
    c. To search for and reduce variability in a process.
    d. All of these choices are true.

19. In simple regression analysis, if the correlation coefficient is a positive value, then
    a. The Y intercept must also be a positive value.
    b. The coefficient of determination can be either positive or negative, depending on the value of the slope.
    c. The least squares regression equation could either have a positive or a negative slope.
    d. The slope of the regression line must also be positive.

20. A 95% confidence interval for a population mean is determined to be 100 to 120. If the confidence coefficient is reduced to 0.90, the interval for μ
    a. becomes narrower
    b. becomes wider
    c. does not change
    d. becomes 0.1

21. After computing a confidence interval, the user believes the results are meaningless because the width of the interval is too large. Which one of the following is the best recommendation?
    a. Increase the level of confidence for the interval.
    b. Decrease the sample size.
    c. Increase the sample size.
    d. Reduce the population variance.

22. After calculating the sample size needed to estimate a population proportion to within 0.04, your statistics professor told you the maximum allowable error must be reduced to just .01. If the original calculation led to a sample size of 800, the sample size will now have to be:
    a. 800
    b. 3200
    c. 12,800
    d. 6400

23. If a hypothesis is rejected at 95% confidence, it
    a. will always be accepted at 90% confidence
    b. will always be rejected at 90% confidence
    c. will sometimes be rejected at 90% confidence

d. None of these alternatives is correct.

24. In a two-tailed hypothesis test the test statistic is determined to be Z = -2.5. The p-value for this test
   a. is 0.0062
   b. is 0.0124
   c. is 0.4938
   d. cannot be determined, since the level of confidence is not given.

25. If the lower and upper confidence limits of the population proportion p, using a sample of size 1500, are 0.184 and 0.238, respectively, then the lower and upper confidence limits of the total number of successes in the population, given that the population size is 750,000, are respectively:
   a. 276 and 357
   b. 137,724 and 178,143
   c. 138,000 and 178,500
   d. 138,276 and 179,857

26. In a regression and correlation analysis if $r^2 = 1$, then
   a. SSE must also be equal to one
   b. SSE must be equal to zero
   c. SSE can be any positive value
   d. SSE must be negative

27. If we want to conduct a test to determine whether a population mean is greater than another population mean, we
   a. can use the analysis of variance
   b. can use the independent samples *t*-test for difference between two means
   c. can use the chi-squared test
   d. either a or b.

28. Two samples of sizes 32 and 38 are independently drawn from two normal populations, where the unknown population variances are assumed to be equal. The number of degrees of freedom of the equal-variances t test statistic is:
   a. 70
   b. 65
   c. 64
   d. None of the above

29. In a goodness-of-fit test, suppose that a sample showed that the observed frequency f_i and expected frequency e_i were equal for each cell i. Then, the null hypothesis is
   a. rejected at α=0.05 but is not rejected at α=0.025
   b. not rejected at α= 0.05 but is rejected at α= 0.025
   c. rejected at any level α
   d. not rejected at any α level

30. The analysis of variance is a procedure that allows statisticians to compare two or more population
   a. Means
   b. Proportions
   c. variances
   d. standard deviations

31. If an estimated regression line has a y-intercept of 10 and a slope of 4, then when x = 2 the actual value of y is:
    a. 18
    b. 15
    c. 14
    d. unknown.

32. In order to estimate with 95% confidence the expected value of y for a given value of x in a simple linear regression problem, a random sample of 10 observations is taken. Which of the following t-table values listed below would be used?
    a. 2.228
    b. 2.306
    c. 1.860
    d. 1.812

33. The least squares method for determining the best fit minimizes:
    a. total variation in the dependent variable
    b. sum of squares for error
    c. sum of squares for regression
    d. All of these choices are true.

34. In a single-factor analysis of variance, MST is the mean square for treatments and MSE is the mean square for error. The null hypothesis of equal population means is rejected if:
    a. MST is much smaller than MSE.
    b. MST is much larger than MSE.
    c. MST is equal to MSE.
    d. None of these choices.

35. Which of the following is not a required condition for one-way ANOVA?
    a. The sample sizes must be equal.
    b. The populations must all be normally distributed.
    c. The population variances must be equal.
    d. The samples for each treatment must be selected randomly and independently.

36. The confidence interval estimate of the expected value of y for a given value x, compared to the prediction interval of y for the same given value of x and confidence level, will be:
    a. wider.
    b. narrower.
    c. the same.
    d. impossible to know.

37. The number of degrees of freedom for a contingency table with 4 rows and 8 columns is
    a. 32
    b. 12
    c. 21
    d. 10

38. A regression analysis between weight (y in pounds) and height (x in inches) resulted in the following least squares line: $\hat{y} = 120 + 5x$. This implies that if the height is increased by 1 inch, the weight, on average, is expected to:
    a. increase by 1 pound.
    b. decrease by 1 pound.
    c. increase by 5 pounds.
    d. increase by 24 pounds.

39. In performing a regression analysis which of the following must be true about the distribution of the error variable?
    a. The distribution is normal with mean zero.
    b. The errors associated with one y value are independent of errors associated with another y value.
    c. The standard deviation is constant for each value of x.
    d. All of these choices are true.

40. Assuming that all necessary conditions are met, what needs to be changed in the formula $\hat{p} \pm t_{\alpha} \sqrt{\hat{p}(1-\hat{p})/n}$ so that we can use it to construct a (1-α) confidence interval estimate for the population proportion p?
    a. $\hat{p}$ should be replaced by p.
    b. t_{α} should be replaced by z_{α}.
    c. t_{α} should be replaced by $t_{\alpha/2}$.
    d. t_{α} should be replaced by $z_{\alpha/2}$.

41. How do you calculate the expected frequency for one cell in a goodness-of-fit test?
    a. The expected frequency is equal to the proportion specified in H_0 for that cell.
    b. Use the total number of observations divided by the number of categories.
    c. Multiply the specified proportion for that cell (found in H_0) by the total sample size.
    d. None of these choices.

42. The number of degrees of freedom for the denominator in one-way ANOVA test involving 4 population means with 15 observations sampled from each population is:
    a. 60
    b. 19
    c. 56
    d. 45

43. In simple linear regression, which of the following statements indicates there is no linear relationship between the variables x and y?
    a. Coefficient of determination is −1.0.
    b. Coefficient of correlation is 0.0.
    c. Sum of squares for error is 0.0.
    d. None of these choices.

44. Consider the following ANOVA table:

Source of Variation	*SS*	*df*	*MS*	*F*
Treatments	128	4	32	2.963
Error	270	25	10.8	
Total	398	29		

The total number of observations is:
a. 25
b. 29
c. 30
d. 32

45. Consider a multinomial experiment with 200 trials, where the outcome of each trial is classified into one of 5 categories. The number of degrees of freedom associated with the chi-squared goodness-of-fit test equals:
a. 195
b. 40
c. 5
d. 4

46. In which case can an F-test be used to compare two population means?
a. For one tail tests only.
b. For two tail tests only.
c. For either one or two tail tests.
d. None of these choices.

47. Testing whether the slope of the population regression line could be zero is equivalent to testing whether the:
a. sample coefficient of correlation could be zero
b. standard error of estimate could be zero
c. population coefficient of correlation could be zero
d. sum of squares for error could be zero

48. If the standard error of estimate $s_\varepsilon = 20$ and n = 10, then the sum of squares for error, SSE, is:
a. 400
b. 3,200
c. 4,000
d. 40,000

49. To address whether two variables are related in a contingency table, the null hypothesis, H0, says that
a. The two variables are independent.
b. The two variables are dependent.
c. The two variables are equal.
d. None of these choices.

50. In a simple linear regression problem, the following statistics are calculated from a sample of 10 observations: $\sum (x - \bar{x})(y - \bar{y}) = 2250$, $s_x = 10$, $\sum x = 50$, $\sum y = 75$. The least squares estimates of the slope and y-intercept are, respectively,
a. 1.5 and 0.5
b. 2.5 and 1.5
c. 1.5 and 2.5
d. 2.5 and –5.0

51. In a simple linear regression problem, the following sum of squares are produced: $\sum (y_i - \bar{y})^2 = 200$, $\sum (y_i - \hat{y}_i)^2 = 50$, and $\sum (\hat{y}_i - \bar{y})^2 = 150$. The percentage of the variation in y that is explained by the variation in x is:
   a. 25%
   b. 75%
   c. 33%
   d. 50%

52. Given that the sum of squares for error is 60 and the sum of squares for regression is 140, then the coefficient of determination is:
   a. 0.429
   b. 0.300
   c. 0.700
   d. None of these choices.

53. Given the least squares regression line $\hat{y} = 2.48 - 1.63x$, and a coefficient of determination of 0.81, the coefficient of correlation is:
   a. –0.66
   b. 0.81
   c. –0.90
   d. 0.90

54. Which of the following statements is correct regarding the percentile points of the F-distribution?
   a. $F_{0.05,10,20} = 1/F_{0.95,10,20}$
   b. $F_{0.05,10,20} = 1/F_{0.05,20,10}$
   c. $F_{0.95,10,20} = 1/F_{0.95,20,10}$
   d. $F_{0.95,10,20} = 1/F_{0.05,20,10}$

55. The width of the confidence interval estimate for the predicted value of y depends on
   a. the standard error of the estimate
   b. the value of x for which the prediction is being made
   c. the sample size
   d. All of these choices are true.

56. Given that $s_x^2 = 500, s_y^2 = 750, s_{xy} = 100$, and n = 6, the standard error of estimate is:
   a. 3,749.00
   b. 937.25
   c. 30.61
   d. None of these choices.

57. If the coefficient of determination is 0.975, then which of the following is true regarding the slope of the regression line?
   a. All we can tell is that it must be positive.
   b. It must be 0.975.
   c. It must be 0.987.
   d. Cannot tell the sign or the value.

58. If all the points in a scatter diagram lie on the least squares regression line, then the coefficient of correlation must be:
    a. 1.0
    b. −1.0
    c. either 1.0 or −1.0
    d. 0.0

59. Which of the following tests does not use the chi-squared distribution?
    a. Test of a contingency table.
    b. Goodness-of-fit test.
    c. One tailed test for two proportions.
    d. None of these choices.

60. In the one-way ANOVA where there are k treatments and n observations, the degrees of freedom for the F-statistic are equal to, respectively:
    a. n and k.
    b. k and n.
    c. $n - k$ and $k - 1$.
    d. $k - 1$ and $n - k$.

YORK UNIVERSITY
School of Administrative Studies
AP/ADMS 2320.3.0 - Quantitative Methods 1

FALL 2009 - FINAL EXAMINATION – Part 2

DECEMBER 12, 2009 TWO PARTS: 2PM-5PM

H. Bartel Section D ☐ Section E ☐ Section F ☐ Section G ☐ Section K ☐

P. Ng Section A ☐ Section H ☐ Section I ☐ Section J ☐

R. Huang Section C ☐ Section L ☐

M. Rochon Section B ☐

INSTRUCTIONS:

1. **You must write this exam in the section in which you are enrolled.**
2. Calculations are to be to the **4th decimal place if necessary.**
3. **This is a closed book examination.**
4. Aids allowed: writing utensils, eraser, silent, **non-programmable** calculator, English paper-based dictionary and a two sided reference/formula sheet, not to exceed 8½" by 11", of your own construction. No restrictions on content. No flaps; all pasted or taped on pieces must be properly glued or taped down on all four sides.
5. Students found with any of the following will be deemed to have unauthorized materials as defined in the University Policy on Academic Honesty:
   - cellphones or other communication devices,
   - personal information managers (PIMs),
   - electronic dictionaries,
   - user-programmable calculators
   - pencil cases, unless a clear plastic bag
6. **YOU MUST HAND IN THIS EXAM BEFORE LEAVING THE EXAMINATION ROOM. FAILURE TO DO SO WILL RESULT IN AN AUTOMATIC F GRADE.**

Student's Name (Print)

Student Number

Signature

MARKS:	
Question 1: (10)	
Question 2: (10)	
Question 3: (10)	
Question 4: (10)	
Multiple Choice: (60)	
TOTAL: (100)	

THIS PART CONTAINS 8 PAGES. DO NOT REMOVE ANY PAGES

QUESTION 1 (10 marks)

One is given a number x, which is randomly drawn from a normal distribution with mean μ and standard deviation 1. She needs to test $H_0 : \mu = 0$ against $H_1 : \mu \neq 0$. She decides to reject the null hypothesis in favour of the research hypothesis if $|x| \geq 2$.

	Question	Final Answer
a.	Compute the probability of committing a type I error.	
b.	Compute the probability of committing a type II error if the true value of μ is 1.	

Please show all of your calculations in the spaces provided below and place your answers in the appropriate boxes above.
(NOTE: if there is no answer inserted into the Box, the mark = 0)

QUESTION 2 (10 marks)

A group of babies all of whom weighed approximately the same at birth are randomly divided into two groups. The babies in sample 1 were fed formula 1; those in sample 2 were fed formula 2. The weight gains attained from birth to age six months were recorded for each baby. The results were given in the following table. We desire to know whether either formula 1 or formula 2 is more effective than the other in producing weight gains. Test for this at the 0.05 level of significance. Assume that the weight gains after feeding six months are normally distributed for each formula.

Formula 1	Formula 2
5	9
7	10
8	8
9	6
6	8
7	7
10	9
8	
6	

Part of the results is shown in the following table.

Formula 1	Formula 2
$\bar{x}_1 = 7.33$	$\bar{x}_2 = 8.14$
$s_1^2 = 2.50$	$s_2^2 = 1.81$

QUESTION 3 (10 marks)

Information regarding the ACT scores of samples of students in three different majors is given below.

	Major		
	Management	Finance	Accounting
1	28	22	29
2	26	23	27
3	25	24	26
4	27	22	28
5	21	24	25
6	19	26	26
7	27	27	28
8	17	29	20
9	17	28	20
10	23		24
11			28
12			28
13			29
$\bar{x}_j$	23	25	26
$\sum_{i=1}^{n_j} x_{ij}^2$	5452	5679	8900

At 95% confidence test to determine whether there is a significant difference in the means of the three populations.

Question 4 (10 marks)

A professor of economics suspects that education leads to higher income. She conducted a study to investigate the relationship between income (in $1000s) and education (in years). A random sample of eight individuals is taken and the results are shown below.

Observation #	1	2	3	4	5	6	7	8
Income	58	40	55	35	43	41	52	49
Education	16	11	15	8	12	10	13	14

Some intermediate calculation is given as follows:

$(58+40+55+35+43+41+52+49)=373$

$(58^2+40^2+55^2+35^2+43^2+41^2+52^2+49^2)=17849$

$(16+11+15+8+12+10+13+14)=99$

$(16^2+11^2+15^2+8^2+12^2+10^2+13^2+14^2)=1275$

$(16*58+11*40+15*55+8*35+12*43+10*41+13*52+14*49)=4761$

Use these intermediate results to solve the following questions.

a. Determine the least squares regression line and interpret the value of the slope of the regression line.

b. Can we infer at the 5% significance level that the professor's assertion is correct?

c. Determine the coefficient of determination and discuss what its value tells you.