Concordia University
John Molson
School of Business
Department of Supply Chain and Business Technology Management

COMM 215: Business Statistics
Fall 2013

Final Examination
19:00 – 22:00, December 6, 2013

Last Name: _____ **First Name**_____

Student No.:_____

IMPORTANT: ALSO use pencil to write your name and ID (AND mark the matching circles below the numbers/alphabets filled) in the appropriate spaces on the provided multiple choice answer sheet.

Circle your section

A	M-W	10:15 - 11:30	Prof. M. Amir
B	M-W	13:15 - 14:30	Prof. Samie Ly
C	W	08:45 - 11:30	Prof. Samie Ly
D	W	10:15 - 13:00	Prof. S. Hodai-Hemami
E	W	11:45 - 14:30	Prof. C. Bayne
AA	T	17:45 - 20:15	Prof. S. Hodai-Hemami
BB	W	17:45 - 20:15	Prof. Samie Ly

INSTRUCTIONS

a) **Attempt all questions**. Show your work for FULL credit.
b) This is a closed book, closed notes examination. You are allowed to use non-programmable calculators during the examination. Sharing of calculators is not allowed.
c) For the **multiple choice questions, use the separate answer sheet** provided.
d) For **PART II, write your answers in the space provided** below each question. Use **both sides of the paper if necessary. Do not include extra pages**
e) Tables and formulas are appended.
f) **Return the exam booklet intact (18 pages including the Table and formula sheet).**
g) **No questions about the examination are allowed.**

		Marked Obtained
Part I		**/60**
Part II		**/40**
Question 1	**/9**	
Question 2	**/10**	
Question 3	**/9**	
Question 4	**/12**	

PART I: Multiple choice questions (One mark for each of questions 1 to 10 and two marks for each of questions 11 to 35). 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. Which of the following statements is true?
   a. If the distribution of a data set is positively skewed, the numerical value of the median is larger than the value of the mean.
   b. The mean of the Binomial distribution is np(1 - p).
   c. If A and B are two independent events, and if P(A and B) = 0.3 and P(B) = 0.6, then P(A) = 0.5.
   d. For any distribution, less than 75% of the data will be contained within two standard deviations from the mean.

2. If two events are mutually exclusive, then their joint probability:
   a. will always be equal to zero.
   b. is equal to the product of the two marginal probabilities.
   c. must be larger than zero, but less than one.
   d. cannot be determined without knowing the probability of occurrence of each one.

3. In the construction of confidence intervals, if all other quantities are unchanged, a decrease in the sample size will lead to a:
   a. narrower interval.
   b. wider interval.
   c. less significant interval.
   d. more significant interval.

4. Which of the following values of the sample correlation coefficient r indicates the strongest linear relationship?
   a. -0.983
   b. -0.005
   c. +0.004
   d. +0.777

5. The standard deviation of the sample means
   a. is more than or equal to the standard deviation of the population
   b. increases as the sample size increases
   c. measures the variability of the mean from sample to sample
   d. none of the above.

6. In simple linear regression, the numerical value of the coefficient of determination
   a. is always larger than the coefficient of correlation
   b. is always smaller than the coefficient of correlation
   c. is negative if the coefficient of correlation is negative
   d. can be larger or smaller than the coefficient of correlation

7. Which of the following statements is FALSE:
   a. The set of possible events of an experiment is called a sample space
   b. An event that cannot be decomposed must have probability zero
   c. The sum of probabilities of all outcomes in a sample space is 1
   d. The complement of the sample space is the empty space

8. Which of the following chi-square test values are likely to lead to rejecting the null hypothesis in a goodness-of-fit test?
   a. 0
   b. 1.1
   c. 2.7
   d. 55

9. A chi-square test is always conducted as:
   a. a left tailed test
   b. a right tailed test
   c. a two-tailed test
   d. depends on the statement of the problem

10. The distribution of income is usually considered to be positively skewed. The measure of central tendency that best represents typical income is:

   a. the mode
   b. the mean
   c. the median
   d. the midhinge
   e. any of the above

11. In a regression model involving more than one independent variable, which of the following tests must be used in order to determine if the relationship between the dependent variable and the set of independent variables is significant?
   a. t test
   b. F test
   c. either a t test or a chi-square test can be used.
   d. Chi-square test

12. In a distribution, the proportion of the total area which must be to the left of the median is:
   a. between 0.25 and 0.75 only when the distribution is symmetric
   b. Answer depends on the value of the mean.
   c. less than 0.5 if the distribution is left skewed
   d. 0.5
   e. more than 0.5 if the distribution is right skew

13. A statistician has collected the following set of data:

   1 7 4 4 5 6 9 28

   Consider the following statements:
   (A) The data are right skewed.
   (B) The mean is 8, the mode is 4, and the median is 4.5.

   Which of the following statements is correct?
   a. only (A) is true.
   b. only (B) is true.
   c. both (A) and (B) are true.
   d. both (A) and (B) are false.

14. A 90% confidence interval for the mean percentage of airline reservations being canceled on the day of the flight is (2.9%, 6.6%). What is the point estimator of the mean percentage of reservations that are canceled on the day of the flight?

   a. 3.75%
   b. 1.85%
   c. 3.30%
   d. 4.75%
   e. None of the above

15. For the following multiple regression equation:

$$\hat{Y}_i = 50 - 2X_1 + 7X_2$$

Which of the following is an interpretation of the intercept?

a. The X-Intercept 50 is the estimate of the mean value of Y if at least one X_i is equal to 0
b. The Y-Intercept 50 is the estimate of the mean value of X_1 if X_2 is equal to 0
c. The Y-Intercept 50 is the estimate of the mean value of Y if X_1 and X_2 are both 0
d. The Y-Intercept 50 is the estimate of the mean value of Y if X_1 is equal to 0
e. None of the suggested answers are correct

Refer to the following in answering questions 16 and 18.

The following ANOVA table summary is for a multiple regression model with two independent variables:

Source	Df	Sum of Squares	Mean squares	F
Regression	2	30		
Error	10	120		
Total				

16. What is the sample size?
a. 12
b. 13
c. 9
d. 14
e. None of the suggested answers are correct

17. What is the overall F test statistic?
a. 1.75
b. 13.1
c. 1.25
d. 1.17
e. None of the suggested answers are correct

18. Which statement is correct?

a. 25% of variation in the Y variable is explained by the 2 independent variables
b. 20% of variation the Y variable is explained by the 10 independent variables
c. 20% of variation the Y variable is explained by the 2 dependent variables
d. 20% of variation the Y variable is explained by the 2 independent variables
e. None of the suggested answers are correct

19. To estimate the average time that teenagers spend on the internet per week, a random sample of 100 teenagers is selected, and a 95% confidence interval is computed. What happens to the width of the confidence interval if we decide to increase the confidence level?

a. Widens
b. Narrows
c. No change
d. Widens or Narrows
e. None of the suggested answers are correct

20. The quality control manager at a light bulb factory needs to determine whether the mean life of a large shipment of light bulbs is equal to 375 hours. The population standard deviation is 100 hours. A random sample of 64 light bulbs indicates a sample mean life of 350 hours. The p-value is given to be 0.0455.
Which of the following is correct?
    a. We should reject the null hypothesis and conclude that the mean life of a large shipment of light bulbs is equal to 375 hours.
    b. We should not reject the null hypothesis and conclude that the mean life of a large shipment of light bulbs is equal to 375 hours.
    c. We should reject the null hypothesis and conclude that the mean life of a large shipment of light bulbs differs from 375 hours.
    d. Since p-value is less than 0.05, we should reject the alternative hypothesis
    e. None of the suggested answers are correct

21. Which of the following is the most correct. A sampling error is
    a. the result of errors in measurement of variables in a statistical experiment.
    b. due to difference between the sample means and the population mean.
    c. the result or errors from incomplete sample returns for a census.
    d. none of the above.

22. Two variables X and Y are found to be correlated and the coefficient of correlation found to be 0.98. The researcher wishes to construct a simple linear regression equation and test the impact of x on Y. The coefficient of determination is:
    a. - 0.98
    b. 0. 96
    c. $\sqrt{0.98}$
    d. None of the above.

23. Statistics Canada would like to estimate average weekly wages within a margin of error equal to $20 for Canadian adults. Assume the population standard deviation for weekly wage is $160. The sample size needed to construct a confidence interval for the population mean at the 90 percent confidence level is approximately:
    a. 174
    b. 3464
    c. 246
    d. 256

24. The two events E1 and E2 are mutually exclusive. Which of the following is true:
    a. P(E1 and E2) is always positive.
    b. P(E1 and E2) = 1
    c. P(E1| E2) = P(E1)
    d. None of the above.

25. For Acme airline, the confidence interval for the mean percentage of reservations being canceled on the day of flight is estimated to be (2.9%, 6.6%). What is the margin of error in this estimation?
    a. 3.5 %
    b. 3.30%
    c. 1.85%
    d. 2.85%
    e. none of the above

Refer to the following in answering questions 26 and 27.
In a multiple regression model involving 44 observations, the estimated regression equation is:
$\hat{Y} = 29 + 18X_1 + 43X_2 + 87X_3$
For this model SSR = 600 and SSE = 400.

26. The coefficient of determination for the above model is:

a. 0.667
b. 0.600
c. 0.336
d. 0.400
e. none of the above

27. The F ratio is:

a. 20
b. 25
c. 30
d. 10
e. none of the above

28. A statistics professor has stated that over the years 90% of his students pass the class. To check this claim, a random sample of 150 students who have taken his class indicated that 129 passed the class. If the professor's claim is correct, what is the probability that 129 or fewer will pass the class this semester?

a. 0.9484
b. 0.0516
c. 0.5516
d. 0.4484
e. none of the above

29. An auditor is sampling 64 charge accounts from the many active accounts at a large department store. Suppose the average balance owed among all the store's accounts is \$75 with a standard deviation of \$100. What is the probability that the sample mean of the balances owed among the 64 accounts sampled is less than \$70?

a. 0.6554
b. very close to zero
c. 0.4801
d. 0.3446
e. none of the above

30. Suppose that the hourly dollar amount of food sold by a successful Fast Food restaurant follows a bell-shaped distribution with mean sale level of \$400 per hour and a standard deviation of \$60 per hour. Then, the percentage of the working hours for which the dollar amount of food sold falls between \$280 and \$520 is approximately:

a. This question cannot be answered from the information provided
b. 95%
c. 68%
d. Over 98%

31. In hypothesis testing a type I error is:

a. The margin of error in the confidence interval test of the null hypothesis.
b. The differences between the sample mean and the population mean caused by measurement errors.
c. The power of the test of hypothesis.
d. The rejection of the null hypothesis when it is in fact true.

32. The management of a company needs a sample size in order to determine the 98% confidence interval for the mean sales of its products. Unfortunately the files for product sales are not complete for the period under consideration. However, the sales manager believes that the minimum order was $1 546 and the largest was $2 746. Using your knowledge of the Empirical distribution the smallest approximate value of the standard deviation of sales is:

a. 300
b. 200
c. 150.4
d. None of the above.

33. Volkswagen would like to estimate the proportion of drivers of the new VW Beetle that are women. A random sample of 250 Beetle owners is taken and 140 are found to be women. The company decided that it would construct a 95 percent confidence interval for the true population proportion. The lower and upper bounds of the interval are:

a. (0.4985, 0.6215)
b. (0.4972 , 0.6228)
c. (0.50, 0.73)
d. None of the above

34. Indicate which of the following pairs of hypotheses is not a valid hypothesis:

a. Ho: $\mu > 20$ Ha: $\mu < 20$
b. Ho: $\mu = 30$ Ha: $\mu > 30$
c. Ho: $\mu \leq 55$ Ha: $\mu > 55$
d. Ho: $\mu = 30$ Ha: $\mu \neq 30$

35. Which of the following is a measure of the variability around the regression line?

a. Coefficient of determination
b. Regression sum of squares (SSR)
c. Standard error of the estimate
d. Slope of the regression line
e. None of the above

Part II: Use both sides of the paper if necessary.

Question 1. (9 marks)

In BOBLOS supermarket, during peak hours, the average waiting time for an available cashier (human or non-human) is 9.46 minutes with a standard deviation of 2.47 minutes. In order to reduce the waiting time to less than 6 minutes, BOBLOS manager introduces a new operational system. To assess effectiveness of the new system, a random sample of 75 customers' waiting is obtained which reveals a sample mean of 5.30 minutes with a standard deviation of 2.47 minutes.

a) Is there sufficient evidence to conclude that the manager was successful in reducing the waiting time under the new system? Test using a 5% level of significance. Use the critical value approach.

b) The manager would like to use automatic cashier (booths with automated service where customers can simply tap the item and pay without the interaction of a human). However, she believes that at least 50% of customers prefer human cashiers. In the sample of 75 customers, 28 said they prefer automatic cashiers. Is there sufficient evidence that the manager's belief is wrong? Use the p-value approach and 1% level significance.

c) Referring to your answer in part b) which of the two statistical errors might have been made in this case?

Question 2. (10 marks)

A random sample of 1008 students were asked, , to indicate their preferred drink during exam periods. Their preferences are summarized in the following table by their program status (part-time or Full-time).

	Drink Preferences			
Status	**Juices**	**Soft Drink**	**Tea/Coffee**	**Total**
Part-Time	130	79	190	
Full-Time	89	130	390	
Total				

a) At the 5% level of significance, test for a relationship between drink preferences and program status. (you may use the above table)

b) It has been claimed that more than 60% of students prefer tea/coffee during the exam period. Using a 5% significance level, based on this sample, what can you conclude? (state the null and the alternative hypotheses)

Question 3. (9 marks)

The following is a sample of scores (X) for 36 students in a marketing course.

54, 77, 84, 62, 77, 86, 69, 78, 86, 69, 81, 87, 72, 82, 89, 72, 82, 90, 76, 82, 90, 74, 82, 92, 75, 82, 92, 76, 82, 92, 76, 82, 94, 76, 84, 96.

Given that:
$\sum X = 2900$
and
$\sum (X - \bar{X})^2 = 2852.889$

a) Find the 90 percent confidence interval for the population mean μ.

b) Suppose that the average score for all students taking the course is 75 and the standard deviation is 16 what is the probability of getting a sample with a mean as large or greater than 80.56.

c) What is the probability of getting two independent samples each with sample means greater than or equal to 80.56.

Question 4. (12 marks)

An automobile dealer wants to see if there is a relationship between monthly sales and the interest rate. A random sample of 4 months was taken. The results of the sample are presented in the table below:

	Y= Monthly Sales	X= Interest Rate %	Y^2	X^2	XY
	22	9.2	484	84.64	202.4
	20	7.6	400	57.76	152
	10	10.4	100	108.16	104
	45	5.3	2025	28.09	238.5
Total	97	32.5	3009	278.65	696.9

a. Find the estimated least squares regression equation and interpret the meaning of the values obtained for the coefficients b_0 and b_1 in the problem.

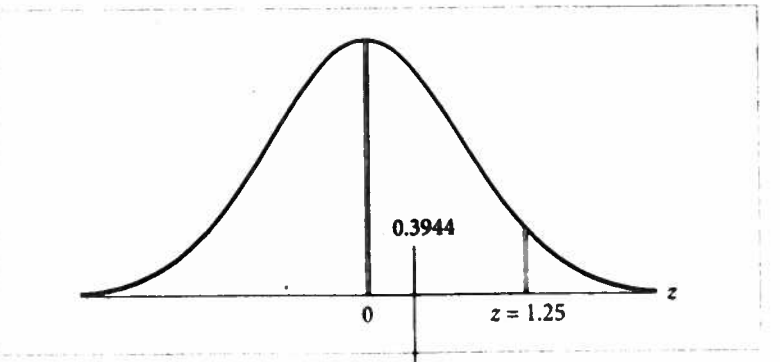
b. Obtain a measure of how well the estimated regression line fits the data.

c. At the 1% level of significance, test whether there is a significant relationship between the interest rate and monthly sales. State the null and alternative hypotheses.

d. Construct a 99% confidence interval for the average monthly sales for all months with a 10% interest rate.

APPENDIX D

Standard Normal Distribution Table

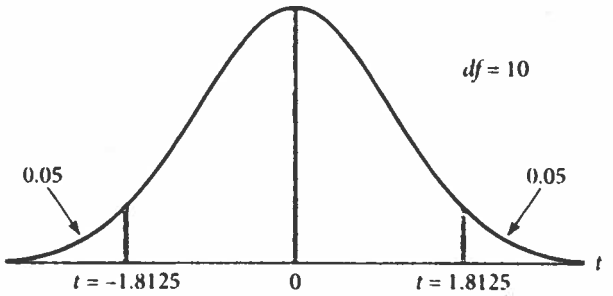


z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0753
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2257	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2517	0.2549
0.7	0.2580	0.2611	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2995	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990

13

APPENDIX F

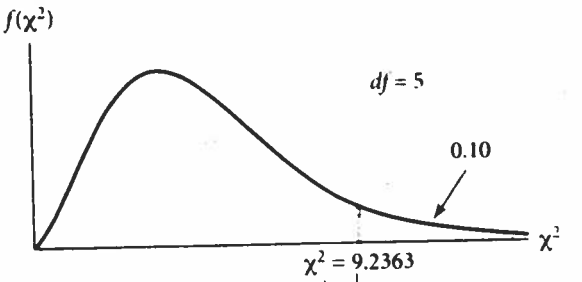
Values of t for Selected Probabilities



PROBABILITIES (OR AREAS UNDER t-DISTRIBUTION CURVE)

Conf. Level	0.1	0.3	0.5	0.7	0.8	0.9	0.95	0.98	0.99
One Tail	0.45	0.35	0.25	0.15	0.1	0.05	0.025	0.01	0.005
Two Tails	0.9	0.7	0.5	0.3	0.2	0.1	0.05	0.02	0.01
df	Values of t								
1	0.1584	0.5095	1.0000	1.9626	3.0777	6.3137	12.7062	31.8210	63.6559
2	0.1421	0.4447	0.8165	1.3862	1.8856	2.9200	4.3027	6.9645	9.9250
3	0.1366	0.4242	0.7649	1.2498	1.6377	2.3534	3.1824	4.5407	5.8408
4	0.1338	0.4142	0.7407	1.1896	1.5332	2.1318	2.7765	3.7469	4.6041
5	0.1322	0.4082	0.7267	1.1558	1.4759	2.0150	2.5706	3.3649	4.0321
6	0.1311	0.4043	0.7176	1.1342	1.4398	1.9432	2.4469	3.1427	3.7074
7	0.1303	0.4015	0.7111	1.1192	1.4149	1.8946	2.3646	2.9979	3.4995
8	0.1297	0.3995	0.7064	1.1081	1.3968	1.8595	2.3060	2.8965	3.3554
9	0.1293	0.3979	0.7027	1.0997	1.3830	1.8331	2.2622	2.8214	3.2498
10	0.1289	0.3966	0.6998	1.0931	1.3722	1.8125 ←	2.2281	2.7638	3.1693
11	0.1286	0.3956	0.6974	1.0877	1.3634	1.7959	2.2010	2.7181	3.1058
12	0.1283	0.3947	0.6955	1.0832	1.3562	1.7823	2.1788	2.6810	3.0545
13	0.1281	0.3940	0.6938	1.0795	1.3502	1.7709	2.1604	2.6503	3.0123
14	0.1280	0.3933	0.6924	1.0763	1.3450	1.7613	2.1448	2.6245	2.9768
15	0.1278	0.3928	0.6912	1.0735	1.3406	1.7531	2.1315	2.6025	2.9467
16	0.1277	0.3923	0.6901	1.0711	1.3368	1.7459	2.1199	2.5835	2.9208
17	0.1276	0.3919	0.6892	1.0690	1.3334	1.7396	2.1098	2.5669	2.8982
18	0.1274	0.3915	0.6884	1.0672	1.3304	1.7341	2.1009	2.5524	2.8784
19	0.1274	0.3912	0.6876	1.0655	1.3277	1.7291	2.0930	2.5395	2.8609
20	0.1273	0.3909	0.6870	1.0640	1.3253	1.7247	2.0860	2.5280	2.8453
21	0.1272	0.3906	0.6864	1.0627	1.3232	1.7207	2.0796	2.5176	2.8314
22	0.1271	0.3904	0.6858	1.0614	1.3212	1.7171	2.0739	2.5083	2.8188
23	0.1271	0.3902	0.6853	1.0603	1.3195	1.7139	2.0687	2.4999	2.8073
24	0.1270	0.3900	0.6848	1.0593	1.3178	1.7109	2.0639	2.4922	2.7970
25	0.1269	0.3898	0.6844	1.0584	1.3163	1.7081	2.0595	2.4851	2.7874
26	0.1269	0.3896	0.6840	1.0575	1.3150	1.7056	2.0555	2.4786	2.7787
27	0.1268	0.3894	0.6837	1.0567	1.3137	1.7033	2.0518	2.4727	2.7707
28	0.1268	0.3893	0.6834	1.0560	1.3125	1.7011	2.0484	2.4671	2.7633
29	0.1268	0.3892	0.6830	1.0553	1.3114	1.6991	2.0452	2.4620	2.7564
30	0.1267	0.3890	0.6828	1.0547	1.3104	1.6973	2.0423	2.4573	2.7500
40	0.1265	0.3881	0.6807	1.0500	1.3031	1.6839	2.0211	2.4233	2.7045
50	0.1263	0.3875	0.6794	1.0473	1.2987	1.6759	2.0086	2.4033	2.6778
60	0.1262	0.3872	0.6786	1.0455	1.2958	1.6706	2.0003	2.3901	2.6603
70	0.1261	0.3869	0.6780	1.0442	1.2938	1.6669	1.9944	2.3808	2.6479
80	0.1261	0.3867	0.6776	1.0432	1.2922	1.6641	1.9901	2.3739	2.6387
90	0.1260	0.3866	0.6772	1.0424	1.2910	1.6620	1.9867	2.3685	2.6316
100	0.1260	0.3864	0.6770	1.0418	1.2901	1.6602	1.9840	2.3642	2.6259
250	0.1258	0.3858	0.6755	1.0386	1.2849	1.6510	1.9695	2.3414	2.5956
500	0.1257	0.3855	0.6750	1.0375	1.2832	1.6479	1.9647	2.3338	2.5857
∞	0.1257	0.3853	0.6745	1.0364	1.2816	1.6449	1.9600	2.3263	2.5758

Values of χ^2 for Selected Probabilities

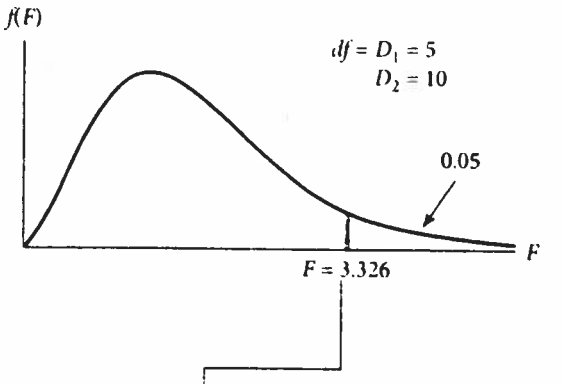


PROBABILITIES (OR AREAS UNDER CHI-SQUARE DISTRIBUTION CURVE ABOVE GIVEN CHI-SQUARE VALUES)

	0.995	0.99	0.975	0.95	0.90	0.10	0.05	0.025	0.01	0.005
df	*Values of Chi-Squared*									
1	0.0000	0.0002	0.0010	0.0039	0.0158	2.7055	3.8415	5.0239	6.6349	7.8794
2	0.0100	0.0201	0.0506	0.1026	0.2107	4.6052	5.9915	7.3778	9.2104	10.5965
3	0.0717	0.1148	0.2158	0.3518	0.5844	6.2514	7.8147	9.3484	11.3449	12.8381
4	0.2070	0.2971	0.4844	0.7107	1.0636	7.7794	9.4877	11.1433	13.2767	14.8602
5	0.4118	0.5543	0.8312	1.1455	1.6103	9.2363	11.0705	12.8325	15.0863	16.7496
6	0.6757	0.8721	1.2373	1.6354	2.2041	10.6446	12.5916	14.4494	16.8119	18.5475
7	0.9893	1.2390	1.6899	2.1673	2.8331	12.0170	14.0671	16.0128	18.4753	20.2777
8	1.3444	1.6465	2.1797	2.7326	3.4895	13.3616	15.5073	17.5345	20.0902	21.9549
9	1.7349	2.0879	2.7004	3.3251	4.1682	14.6837	16.9190	19.0228	21.6660	23.5893
10	2.1558	2.5582	3.2470	3.9403	4.8652	15.9872	18.3070	20.4832	23.2093	25.1881
11	2.6032	3.0535	3.8157	4.5748	5.5778	17.2750	19.6752	21.9200	24.7250	26.7569
12	3.0738	3.5706	4.4038	5.2260	6.3038	18.5493	21.0261	23.3367	26.2170	28.2997
13	3.5650	4.1069	5.0087	5.8919	7.0415	19.8119	22.3620	24.7356	27.6882	29.8193
14	4.0747	4.6604	5.6287	6.5706	7.7895	21.0641	23.6848	26.1189	29.1412	31.3194
15	4.6009	5.2294	6.2621	7.2609	8.5468	22.3071	24.9958	27.4884	30.5780	32.8015
16	5.1422	5.8122	6.9077	7.9616	9.3122	23.5418	26.2962	28.8453	31.9999	34.2671
17	5.6973	6.4077	7.5642	8.6718	10.0852	24.7690	27.5871	30.1910	33.4087	35.7184
18	6.2648	7.0149	8.2307	9.3904	10.8649	25.9894	28.8693	31.5264	34.8052	37.1564
19	6.8439	7.6327	8.9065	10.1170	11.6509	27.2036	30.1435	32.8523	36.1908	38.5821
20	7.4338	8.2604	9.5908	10.8508	12.4426	28.4120	31.4104	34.1696	37.5663	39.9969
21	8.0336	8.8972	10.2829	11.5913	13.2396	29.6151	32.6706	35.4789	38.9322	41.4009
22	8.6427	9.5425	10.9823	12.3380	14.0415	30.8133	33.9245	36.7807	40.2894	42.7957
23	9.2604	10.1957	11.6885	13.0905	14.8480	32.0069	35.1725	38.0756	41.6383	44.1813
24	9.8862	10.8563	12.4011	13.8484	15.6587	33.1962	36.4150	39.3641	42.9798	45.5584
25	10.5196	11.5240	13.1197	14.6114	16.4734	34.3816	37.6525	40.6465	44.3140	46.9280
26	11.1602	12.1982	13.8439	15.3792	17.2919	35.5632	38.8851	41.9231	45.6416	48.2898
27	11.8077	12.8785	14.5734	16.1514	18.1139	36.7412	40.1133	43.1945	46.9628	49.6450
28	12.4613	13.5647	15.3079	16.9279	18.9392	37.9159	41.3372	44.4608	48.2782	50.9936
29	13.1211	14.2564	16.0471	17.7084	19.7677	39.0875	42.5569	45.7223	49.5878	52.3355
30	13.7867	14.9535	16.7908	18.4927	20.5992	40.2560	43.7730	46.9792	50.8922	53.6719

APPENDIX H

F-Distribution Table: Upper 5% Probability (or 5% Area) under *F*-Distribution Curve



DENOMINATOR $df = D_2$ | NUMERATOR $df = D_1$

	1	2	3	4	5	6	7	8	9	10
1	161.446	199.499	215.707	224.583	230.160	233.988	236.767	238.884	240.543	241.882
2	18.513	19.000	19.164	19.247	19.296	19.329	19.353	19.371	19.385	19.396
3	10.128	9.552	9.277	9.117	9.013	8.941	8.887	8.845	8.812	8.785
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	5.964
5	6.608	5.786	5.409	5.192	5.050	4.950	4.876	4.818	4.772	4.735
6	5.987	5.143	4.757	4.534	4.387	4.284	4.207	4.147	4.099	4.060
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.637
8	5.318	4.459	4.066	3.838	3.688	3.581	3.500	3.438	3.388	3.347
9	5.117	4.256	3.863	3.633	3.482	3.374	3.293	3.230	3.179	3.137
10	4.965	4.103	3.708	3.478	3.326	3.217	3.135	3.072	3.020	2.978
11	4.844	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.854
12	4.747	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	2.753
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.671
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	2.602
15	4.543	3.682	3.287	3.056	2.901	2.790	2.707	2.641	2.588	2.544
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	2.494
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.450
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.412
19	4.381	3.522	3.127	2.895	2.740	2.628	2.544	2.477	2.423	2.378
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.348
24	4.260	3.403	3.009	2.776	2.621	2.508	2.423	2.355	2.300	2.255
30	4.171	3.316	2.922	2.690	2.534	2.421	2.334	2.266	2.211	2.165
40	4.085	3.232	2.839	2.606	2.449	2.336	2.249	2.180	2.124	2.077
50	4.034	3.183	2.790	2.557	2.400	2.286	2.199	2.130	2.073	2.026
100	3.936	3.087	2.696	2.463	2.305	2.191	2.103	2.032	1.975	1.927
200	3.888	3.041	2.650	2.417	2.259	2.144	2.056	1.985	1.927	1.878
300	3.873	3.026	2.635	2.402	2.244	2.129	2.040	1.969	1.911	1.862

DENOMINATOR $df = D_2$ | NUMERATOR $df = D_1$

	11	12	13	14	15	16	17	18	19	20
1	242.981	243.905	244.690	245.363	245.949	246.466	246.917	247.324	247.688	248.016
2	19.405	19.412	19.419	19.424	19.429	19.433	19.437	19.440	19.443	19.446
3	8.763	8.745	8.729	8.715	8.703	8.692	8.683	8.675	8.667	8.660
4	5.936	5.912	5.891	5.873	5.858	5.844	5.832	5.821	5.811	5.803
5	4.704	4.678	4.655	4.636	4.619	4.604	4.590	4.579	4.568	4.558
6	4.027	4.000	3.976	3.956	3.938	3.922	3.908	3.896	3.884	3.874
7	3.603	3.575	3.550	3.529	3.511	3.494	3.480	3.467	3.455	3.445
8	3.313	3.284	3.259	3.237	3.218	3.202	3.187	3.173	3.161	3.150
9	3.102	3.073	3.048	3.025	3.006	2.989	2.974	2.960	2.948	2.936
10	2.943	2.913	2.887	2.865	2.845	2.828	2.812	2.798	2.785	2.774
11	2.818	2.788	2.761	2.739	2.719	2.701	2.685	2.671	2.658	2.646
12	2.717	2.687	2.660	2.637	2.617	2.599	2.583	2.568	2.555	2.544
13	2.635	2.604	2.577	2.554	2.533	2.515	2.499	2.484	2.471	2.459
14	2.565	2.534	2.507	2.484	2.463	2.445	2.428	2.413	2.400	2.388
15	2.507	2.475	2.448	2.424	2.403	2.385	2.368	2.353	2.340	2.328
16	2.456	2.425	2.397	2.373	2.352	2.333	2.317	2.302	2.288	2.276

(continued)

COMM 215 Business Statistics
List of formula provided in the Final Examination

Chapter 3 Describing Data

Sample mean: $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$

Sample variance:

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{1}{n-1}\left[\sum x^2 - \frac{(\sum x)^2}{n}\right]$$

Sample standard deviation: $s = \sqrt{s^2}$

Z score: $z = \frac{x - \text{mean}}{\text{standard deviation}}$

Coefficient of variation: $\frac{\text{standard deviation}}{\text{mean}} \times 100\ \%$

Chapter 4 Probability

The rule of complement: $P(\bar{E}) = 1 - P(E)$

The addition rule for two events:

$$P(E_1 or E_2) = P(E_1) + P(E_2) - P(E_1 and E_2)$$

Conditional probability: $P(E_1 \mid E_2) = \frac{P(E_1 and E_2)}{P(E_2)}$

The general multiplication rule:

$$P(E_1 and E_2) = P(E_2)P(E_1 \mid E_2)$$

Chapter 5 Discrete Probability Distributions

Mean (expected value) of a discrete random variable

$$E(x) = \sum xP(x)$$

Variance and standard deviation of a discrete random variable

$$\sigma_x^2 = \sum [x - E(x)]^2 P(x) \qquad \sigma_x = \sqrt{\sigma_x^2}$$

Binomial probability formula

$$p(x) = \frac{n!}{x!(n-x)!} p^x q^{n-x}$$

Mean, variance, and standard deviation of a binomial random variable

$$\mu_x = np,\ \sigma_x^2 = npq, \text{ and } \sigma_x = \sqrt{npq}$$

Chapter 6 Continuous Probability Distribution

Standard normal random variable: $z = \frac{x - \mu}{\sigma}$

Chapter 7 Sampling distribution

$$\mu_{\bar{x}} = \mu$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

Sample proportion: $\bar{p} = \frac{x}{n}$

$$\mu_{\bar{p}} = p$$

Standard error of $\bar{p}$:

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Chapter 8 Estimating Single Population Parameters

A *z*-based confidence interval for a population mean μ with σ known:

Margin of error: $e = z\frac{\sigma}{\sqrt{n}}$

Confidence interval = $\bar{x} \pm z\frac{\sigma}{\sqrt{n}}$

A *t*-based confidence interval for a population mean μ with σ unknown: $\bar{x} \pm t\frac{s}{\sqrt{n}}$

Confidence interval for the proportion: $\bar{p} \pm z\sqrt{\frac{\bar{p}(1-\bar{p})}{n}}$

Chapter 9 Hypothesis Testing

$$z - Test\ for\ \ mean\ \ z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$$

$$t - Test\ for\ \ mean\ \ t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$$

$$z - Test\ \ for\ \ proportion\ z = \frac{\bar{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$$

Chapter 13 Goodness-of-Fit Tests

Chi-square goodness-of-fit test statistic

$$\chi^2 = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$$

Chi-square contingency test statistic

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(o_{ij} - e_{ij})^2}{e_{ij}} \text{ with } df = (r-1)(c-1)$$

Chapter 14 Simple Linear Regression and Correlation Analysis

Sample correlation coefficient

$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\left[\sum (x-\bar{x})^2\right]\left[\sum (y-\bar{y})^2\right]}}$$
$$= \frac{n\sum xy - \sum x \sum y}{\sqrt{\left[n(\sum x^2) - (\sum x)^2\right]\left[n(\sum y^2) - (\sum y)^2\right]}}$$

Simple linear regression model: $y = \beta_0 + \beta_1 x + \varepsilon$

Least squares point estimates

$$b_1 = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} = \frac{\sum xy - \frac{\sum x \sum y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Sum of squared residuals (Sum of squares error)

$$SSE = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 = \sum y^2 - b_0 \sum y - b_1 \sum xy$$

Total sum of squares: $SST = \sum_{i=1}^{n} (y_i - \bar{y})^2$

Sum of squares regression: $SSR = \sum_{i=1}^{n} (\hat{y}_i - \bar{y})^2$

Simple regression estimator for the standard error of the estimate: $s_\varepsilon = \sqrt{\frac{SSE}{n-2}}$

Coefficient of determination: $R^2 = r^2 = \frac{SSR}{SST}$

F test for the simple linear regression model:

$$F = \frac{\frac{SSR}{1}}{\frac{SSE}{(n-2)}} \qquad df = (D_1 = 1, D_2 = n-2)$$

Simple regression estimator for the standard error of the slope: $s_{b_1} = \frac{s_\varepsilon}{\sqrt{\sum (x-\bar{x})^2}} = \frac{s_\varepsilon}{\sqrt{\sum x^2 - \frac{(\sum x)^2}{n}}}$

Test of hypothesis for slope: $t = \frac{b_1 - \beta_1}{s_{b_1}} \qquad df = n-2$

Confidence interval for $E(y) \mid x_p$:

$$\hat{y} \pm t s_\varepsilon \sqrt{\frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x-\bar{x})^2}}$$

Prediction interval for $y \mid x_p$: $\hat{y} \pm t s_\varepsilon \sqrt{1 + \frac{1}{n} + \frac{(x_p - \bar{x})^2}{\sum (x-\bar{x})^2}}$

Chapter 15 Multiple regression

The multiple regression model:

$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + \varepsilon$

Standard error: $s_\varepsilon = \sqrt{\frac{SSE}{n-k-1}} = \sqrt{MSE}$

Multiple coefficient of determination: $R^2 = \frac{SSR}{SST}$

An *F* test for the linear regression model:

$$F = \frac{\frac{SSR}{k}}{\frac{SSE}{n-k-1}} \qquad df = (D_1 = k, D_2 = n-k-1)$$

Testing the significance of each regression coefficient:

$$t = \frac{b_1 - 0}{s_{b_1}} \qquad df = n-k-1$$