V. Shannon	SN
Sarbajeet	

G.J.

THE UNIVERSITY OF CALGARY

SCHULICH SCHOOL OF ENGINEERING

Final Examination

PROBABILITY AND STATISTICS FOR ENGINEERS
ENGG 319

December 17, 2016 Time: 3.0 h 12:00 – 15:00

INSTRUCTIONS:

- The exam is OPEN textbook & CLOSED notes.
- You are allowed a single page (double-sided) formula (cheat) sheet.
- Only standard Schulich calculators are allowed.
- Write your name and UCID at the top of every page.
- You must show all of your work to receive full marks.

There are **22** Questions

Show your *approach & computations* in the space provided for each question

Total possible marks for this exam is **60**
This exam counts for **45%** of the overall course grade

PART A (20 Questions) – 2 Marks Each (Circle the BEST answer):

1. Find k such that P (-2.771 < T < k) = 0.045 for a sample size of 15. Assume that T follows t-distribution.

   (a) 2.069
   (b) 3.104
   (c) 1.714
   (d) 2.807
   (e) 2.649
   (f) none of the above

2. In a study with regard to a chi-squared variable X^2, a certain number of observations were made. It was found that P (X^2 < 3.325) = 0.05. Determine the number of observations in the study.

   (a) 6
   (b) 7
   (c) 8
   (d) 9
   (e) 10
   (f) none of the above

3. The determination of the percent canopy closure (PCC) of a forest is essential for wildlife habitat assessment, watershed runoff estimation, erosion control and other forest management activities. One way in which scientists estimate PCC is through the use of satellite and aircraft sensor imaging. A study of the PCC in a certain forest was conducted by examining the readings of 32 imaging data collected by aircraft at various forest sites which resulted in a mean and a standard deviation of 121.74 and 27.52, respectively. Find the probability that the mean PCC will fall between 118 and 130. Consider four decimal places in your answer.

   (a) 0.2794
   (b) 0.1760
   (c) 0.7348
   (d) 0.9554
   (e) 0.4554
   (f) none of the above

4. A hybrid algorithm for solving polynomial 0-1 mathematical programming problems was presented in the IEEE (*Institute of Electrical and Electronics Engineers*) Transactions in June 1990. Using such algorithm, the solution time for a randomly selected problem has a normal probability distribution with a mean and standard deviation of 0.8 s and 1.5 s, respectively. For a random sample of 8 problems solved by the hybrid algorithm, find the probability that the variance will exceed 5.343 s^2.

   (a) ~ 0.1%
   (b) ~ 2.0%
   (c) 80% - 90.0%
   (d) 0.1% - 0.5%
   (e) 2.5% - 5%
   (f) none of the above

5. A machine produces metal pieces that are cylindrical in shape. A sample of 9 pieces is taken. The sample mean and standard deviation are found to be 1.006 and 0.025. Find 70% confidence interval on the mean (to three decimal places).

   (a) 1.006 ± 0.920
   (b) 1.006 ± 0.935
   (c) 1.006 ± 0.955
   (d) 1.006 ± 0.645
   (e) 1.006 ± 0.455
   (f) None of the above

6. In *Question 5* above, how large a sample is required to be 95% confident that the estimated mean is within 0.001 of the true value.

   (a) 25
   (b) 23
   (c) 21
   (d) 19
   (e) 17
   (f) none of the above

7. Geologists analyze fluid inclusions in rock to infer the compositions of fluids present when the rocks crystallized. A new technique was developed for that purpose and needs to be tested. An experiment was conducted to estimate the precision of the new technique where a chip of natural quartz with several artificially produced fluid inclusions was tested. The amount of liquid carbon dioxide present in the inclusion was recorded for the same inclusion on four different days which resulted in the following percentage data: 86.6, 84.6, 85.5 and 85.9, respectively. Construct the confidence interval at 1% level of significance (to three decimal places) for the standard deviation of the liquid carbon dioxide concentration measurements.

   (a) 0.163, 29.151
   (b) 0.566, 4.732
   (c) 0.320, 22.387
   (d) 0.404, 5.399
   (e) 0.429, 4.263
   (f) none of the above

8. The following data set (x, y) = {(1, 1), (2, 1) ,(3, 2), (4, 2),(5, 4)} is fitted to a simple linear regression model. The fitted equation is:

$$\hat{y} = -\mathbf{0.1} + \mathbf{0.7}\, x.$$

   Calculate the sum of the residuals.
   (a) 0
   (b) 1.1
   (c) 4.9
   (d) 6
   (e) 2.8
   (f) none of the above

9. The methodology of conducting stress analysis of newly designed timber structures is well known, however, few data is available on the actual or allowable stress for repairing damaged structures. Therefore, design engineers often propose a repair scheme such as gluing without any knowledge of its structural effectiveness. In an attempt to partially fill this gap, tests were conducted on epoxy-bonded truss joints made of various types of wood species to determine actual glue-line shear stress recorded in psi (*pounds per square inch*). The results of two wood types are summarized in the following table:

Wood Type	Type I	Type II
Sample size	121	41
Mean Shear Stress (psi)	1312	1352
Standard Deviation (psi)	422	271

Obtain the 90% confidence interval for the ratio of the shear stress variances for the two typed of wood (to three decimal places)

(a) 1.263, 4.268
(b) 1.616, 3.831
(c) 0.986, 2.336
(d) 1.038, 2.460
(e) 1.535, 3.637
(f) none of the above

10. Two methods were used to produce solid catalyst pellets to enhance a gas phase heterogeneous chemical reaction. 50 out of a sample of 1000 pellets produced under method A, and 40 out of a sample of 1000 pellets produced under method B did not meet the specified pellet size requirements. Find 95% confidence interval for the difference in the true proportion of defective pellets produced under method A and method B (to three decimal places).

(a) 0.008, 0.028
(b) 1.616, 3.831
(c) 0.986, 2.336
(d) 1.038, 2.460
(e) 1.535, 3.637
(f) none of the above

11. A drilling process is called deep hole drilling when the ratio of the drilling hole depth to drilling hole diameter exceeds 10. The successful drilling in this process depends on the satisfactory discharge of the drill chip. An experiment was conducted to investigate the performance of deep hole drilling when chip congestion exists where the length of 50 drill chips resulted in a mean of 81.2 mm and a standard deviation of 50.2 mm. Is the true mean drill chip length different from 75 mm at a significance level of 0.01 (to three decimal places)?

(a) Yes; $Statistic_{obs} = 0.873 < Statistic_{cri} = 2.330$
(b) No; $Statistic_{obs} = 0.873 < Statistic_{cri} = 2.575$
(c) Yes; $Statistic_{obs} = 0.124 < Statistic_{cri} = 2.575$
(d) No; $Statistic_{obs} = 0.124 < Statistic_{cri} = 2.330$
(e) No; $Statistic_{obs} = 0.873 < Statistic_{cri} = 2.580$
(f) none of the above

12. The length of some machine parts is normally distributed with mean 200 millimeters and a standard deviation of 15 millimeters. The manufacturing process is periodically checked by taking a sample of 25 machine parts and computing the average. If the sample mean is greater than 191, but smaller than 209, the manufacturing process is thought to be operating satisfactorily; otherwise, we conclude that the population mean is not 200 millimeters. Find the probability of committing a type I error when population mean is 200 millimeters.

(a) 0.07
(b) 0.05
(c) 0.004
(d) 0.0026
(e) 0.0013
(f) none of the above

13. Find the power of the test (to three decimal places) in *Question 12* above if the true population mean is 215 millimeters.

(a) 0.801
(b) 0.023
(c) 0.112
(d) 0.211
(e) 0.024
(f) none of the above

14. The testing department of a tire and rubber company schedules tires for durability tests. Currently, the tires are scheduled twice a week using the SPT (Shortest Processing Time) approach where the tire with the shortest processing time is scheduled first. The company researchers have developed a new proposed scheduling rule (PSR) which they believe will reduce the average flow time (completion time of a test) and lead to a reduction in the average tartness of a scheduled test. To compare the two scheduled rules, 64 tires were randomly selected and divided into two sets of equal size where one set was scheduled using SPT and the other using PSR. The summary of the flow times and tardiness in hours of the tire tests is given in the table below:

	Flow Time		Tardiness	
	Mean	Variance	Mean	Variance
SPT	158.28	8532.80	5.26	452.09
PSR	117.07	5208.53	4.52	319.41

Do you think at the 5% significance level that the average flow time is less under the PSR than under the SPT approach (to three decimal places)?

(a) Yes; $\text{Statistic}_{obs} = 1.989 > \text{Statistic}_{cri} = 1.960$
(b) No; $\text{Statistic}_{obs} = -1.989 < \text{Statistic}_{cri} = 1.645$
(c) No; $\text{Statistic}_{obs} = 0.151 < \text{Statistic}_{cri} = 1.645$
(d) Yes; $\text{Statistic}_{obs} = 1.989 > \text{Statistic}_{cri} = 1.645$
(e) No; $\text{Statistic}_{obs} = 0.151 < \text{Statistic}_{cri} = 1.960$
(f) none of the above

15. A problem that occurs with curtain types of mining is that some byproducts tend to be mildly radioactive and these products sometimes get into freshwater supply. The EPA (Environmental Protection Agency) has issued regulations concerning a limit on the amount of radioactivity in supplies of drinking water, particularly the maximum level for naturally occurring radiation is 5 ppl (picocuries per liter of water). A random sample of 28 water specimens from a city's water supply produced a mean of 4.61 ppl and a variance of 0.76 ppl^2. Is there sufficient evidence from the data to indicate that the mean level of radiation is safe at 99.5% confidence (to three decimal places)?

(a) Yes; $Statistic_{obs} = -2.367 < Statistic_{cri} = 3.057$
(b) No; $Statistic_{obs} = -2.715 < Statistic_{cri} = 2.771$
(c) Yes; $Statistic_{obs} = -0.448 > Statistic_{cri} = -2.575$
(d) No; $Statistic_{obs} = -2.367 < Statistic_{cri} = -2.771$
(e) No; $Statistic_{obs} = -2.715 > Statistic_{cri} = -3.057$
(f) none of the above

16. An experiment was conducted to study the effect of reinforced flanges on the torsional capacity of reinforced concrete T-beams. Several different types of T-beams were tested until cracking where each type had a different flange width. The cracking torsion moment at the top of the flange of two types of T-beams is shown in the table below:

T-Beam Type	Cracking Torsion Moment							
28 inches	6.0	7.2	10.2	13.2	11.4	13.6	9.2	11.2
40 inches	6.8	9.2	8.8	13.2	11.2	14.9	10.2	11.8

Is there a sufficient evidence to indicate a difference in the variation in the cracking torsion moments of the two types of T-beams at a 90% confidence? Consider three decimal places.

(a) No; $Statistic_{obs} = 1.085 < Statistic_{cri} = 3.440$
(b) Yes; $Statistic_{obs} = 1.085 < Statistic_{cri} = 3.790$
(c) Yes; $Statistic_{obs} = 0.921 < Statistic_{cri} = 3.790$
(d) No; $Statistic_{obs} = 1.041 < Statistic_{cri} = 3.790$
(e) No; $Statistic_{obs} = 1.085 < Statistic_{cri} = 3.790$
(f) none of the above

17. Two alloys (A1 and A2) are used in manufacturing steel bars. Assume that a steel producer wants to compare the two alloys on the basis of average load capacity. Steel bars containing alloys A1 and A2 were randomly selected and tested for load capacity and the following results were obtained:

Alloy	A1	A2
Sample size	11	17
Mean	43.7	48.5
Standard Deviation	4.94	4.46

Find the confidence interval for the difference between the true average load capacities for the two alloys at the 1% significance level.

(a) -2.482, 7.118
(b) -9.261, -0.339
(c) -9.801, 0.201
(d) -9.794, 0.195
(e) -9.434, -0.166
(f) none of the above

18. A multiple linear regression model is sought to relate the dependent variable y with three independent variables x_1, x_2 and x_3. In a study, 19 observations were recorded and the following fitted regression model is obtained:

$$\hat{y} = b_0 + b_1x_1 + b_2x_2 + b_3x_3.$$

The numerical values of the constants b_i ($i = 0, 1, 2, 3$) and their estimated variance are given below:

i	b_i	*variance of* b_i
0	-2.555125535	12.57603253
1	7.3964548	3.098749472
2	5.937955521	1.382073416
3	0.113620433	0.183883848

$SSR = 8745.317676$
$SSE = 907.3139034$

Calculate the adjusted coefficient of determination (to three decimal places).

(a) 0.932
(b) 0.806
(c) 0.906
(d) 0.887
(e) 0.831
(f) none of the above

19. Using the information given in *Question 18* above, determine the value of the appropriate statistic (to two decimal places) to test the hypothesis that the coefficient of x_3 is zero.

(a) 2.05
(b) 0.22
(c) 4.31
(d) 0.62
(e) 0.40
(f) none of the above

20. Using the information given in *Question 18* above, determine the value (to two decimal places) of the appropriate statistic to test the hypothesis that the overall regression model is not significant.

(a) 0.05
(b) 0.04
(c) 20.62
(d) 0.02
(e) 0.01
(f) none of the above

PART B (2 Questions) – Q21 (10 Marks) & Q22 (10 Marks):

21. The relationship between temperature (x) and the amount of converted sugar (y) in a certain process is determined using the following data point:

y	x
8	1
10	1.3
10.4	1.7
10.5	1.9
11	2

$\sum x_i = 7.9,\ \sum y_i = 49.9,\ \sum x_i y_i = 80.63,\ \sum x_i^2 = 13.19,\ \sum y_i^2 = 503.41.$

(a) What is the equation of the fitted least square regression line? **(2 marks)**

(b) What is the residual at x=1.3? **(1 mark)**

(c) Estimate the value of y when x = 1.2. **(1 mark)**

(d) If the actual (or true) value of β_1' (the coefficient of x of the true regression line) were 2.5, what is the probability that the calculated slope value, b_1, in part (a) is between 1.0 and 1.5? **(1 mark)**

(e) Construct a 90% confidence interval around β_1, the true coefficient of x. **(2 marks)**

(f) Using the estimated value of the intercept of the true regression line, test the hypothesis that the intercept of the true regression line is zero at the 0.05 level of significance. Show critical region and P-value. **(3 marks)**

22. The lifetime of 10 randomly selected televisions from company B were measured and the results are shown below:

Average life of televisions for company B (years)	5.1	3.8	5.5	4.6	4.8	4.0	4.3	3.4	4.9	5.3

The lifetime of 10 randomly selected televisions from another company (company A) were also measured, but most the data were lost afterwards due to a mishap, except few results from some statistical analyses. The p-value for the hypothesis test whether the population mean of the televisions of company A is 3.5 years is 0.005. The F-statistic (Company A/Company B) is 1.89, assuming the population variances from both Company A and B are equal.

(a) Determine the sample mean and standard deviation for company B. **(2 marks)**

(b) Determine the sample mean ($\bar{X}_A$) for Company A. **(2 marks)**

(c) Determine the confidence interval for $\mu_A - \mu_B$ at a 5% level of significance. **(2 marks)**

(d) Based on the given information, determine using hypothesis testing at the 1% significance level whether $\mu_A - \mu_B \leq 0$. **(4 marks)**