National Exams December 2009

98-Ind-B1, Applied Probability & Statistics

3 hours duration

NOTES:

1. If doubt exists as to the interpretation of any question, the candidate is urged to submit with the answer paper, a clear statement of any assumptions made.

2. This is a closed book exam. Candidates are permitted to use one of the two permitted calculator types (Sharp or Casio models).

3. Candidates are permitted to have one 8.5x11.0 aid sheet. Candidates may write materials on both sides of the sheet.

4. This exam consists of five sections (A→E). Within each section, candidates will be given a choice of questions to answer. Please read the instructions for each section carefully. A breakdown of questions and marks is as follows:

Section A:	Do 2 of 3 Questions.	Total marks: 20
Section B:	Do 2 of 3 Questions.	Total marks: 20
Section C:	Do 1 of 2 Questions.	Total marks: 20
Section D:	Do 2 of 3 Questions.	Total marks: 40
Exam:	7 Questions.	Total marks: 100

4. The value of each question is listed in the exam. Remember to check the instructions for each section. DO NOT ATTEMPT TO DO ALL QUESTIONS.

5. Statistical tables are provided.

Section A: Complete two of the following three questions. This section is worth a total of 20 marks. Do not attempt all questions.

1. A function is given as: $f(t) = c(t^3 - t/2)$ for $0 < t < 4$.
   a. Find the value of c that makes this function a valid probability density function
   b. Find the mean of this function
   c. Find the variance of this function

10 Marks

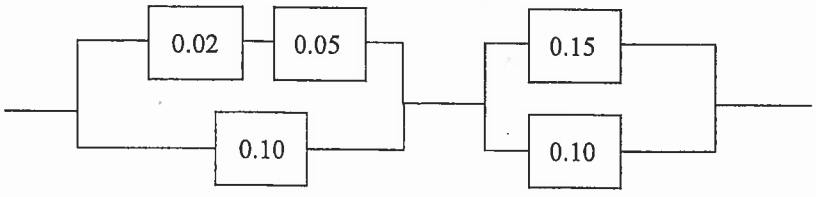
2. Two variables have a joint probability density function f(x,y) given by:

x	y	
	2	3
1	0.20	0.40
2	0.30	0.10

   a. Find the marginal probability mass function of X.
   b. Find the marginal probability mass function of Y.
   c. Are X and Y independent? Why or why not?
   d. Find the covariance of the random variables X and Y.
   e. Find the correlation of the random variables X and Y.

10 Marks

3. A circuit (shown below) functions if and only if there is a functional path from the left side of the circuit to the right side of the circuit. Assume that the devices in the circuit (shown as boxes) fail independently and that the probability of failure of each device is as shown below:



   a. Find the probability that this circuit functions.
   b. If the application for which this circuit is used requires a 99.99% reliability, how many redundant circuits of this type are required?

10 Marks

Section B: Complete two of the following three questions. This section is worth a total of 20 marks. Do not attempt all questions.

1. Bill Hazelwood is conducting an experiment on the number of customers to arrive at a bank between the 08:00 and 09:00. Bill has taken 100 observations over the course of the past four months and has the following data set:

Number of Arrivals	Number of Instances
0	32
1	41
2	18
3	7
4	2

   a. Use a χ^2 goodness of fit test (α=0.05) to determine if a binomial distribution with p = 0.01 and n = 100 would adequately model this dataset.
   b. Use a χ^2 goodness of fit test (α=0.05) to determine if a Poisson distribution with $\lambda = 1$ would adequately model this dataset.
   c. Based on your answers to (a) and (b), which of these two models would you select? Why?

10 Marks

2. A machine dispensing a liquid adhesive at a tire plant in Rockwood, Ontario has been completely overhauled recently to fix a problem in one of the heating elements. Measurements of the amount of adhesive dispensed (in ml) before and after the overhaul are listed below:

Before:	12.9	15.1	14.2	15.3	14.3	13.5	15.0	13.7
After:	17.0	19.7	11.1	18.8	10.7	19.3	12.2	17.6

   a. Can you conclude at the 5% level that the variance after the overhaul differs from the variance before the overhaul?
   b. Can you conclude at the 5% level that the process mean has changed? State any assumptions you have made in developing your solution.

10 Marks

3. A computer system administrator notices that computers running a particular operating system seem to freeze up more often as the installation of the operating system ages. She measures the time (in minutes) before freeze-up for a sample of machines at t = 1 and t = 7 months post installation. This data is given below:

t=1	t=7
207.4	84.3
233.1	53.2
215.9	127.3
235.1	201.3
225.6	174.2
244.4	246.2
245.3	149.4
	156.4
	103.3

a. Develop a 99% confidence interval for the difference in mean time to freeze-up between the 1st and 7th month.
b. State any assumptions you made to develop your solution to (a)
c. Does the data support the administrator's hypothesis? Why or why not?

10 Marks

Section C: Complete one of the following two questions. This section is worth a total of 20 marks. Do not attempt all questions.

1. Sarah Sleen, an industrial engineering student at the University of Wala Wala, is starting a training program to run a marathon. Accordingly, Sarah runs the following training distances (in kilometres) and records her time (in minutes).

Distance (Km)	Time (min)
4	12.0
10	45.0
7	28.0
23	90.0
8	16.0
11	42.0
27	140.0
9	35.0
21	90.0
35	165.0

a. Assuming a linear relationship between distance and run time, fit a model of the form $y = ax+b$ to this data.
b. Are the parameters of the fit model significant at the $\alpha = 0.05$ level?
c. Complete an ANOVA table for your model and determine if a linear model is indeed a good fit for this data.
d. Determine a 95% confidence interval for the time for Sarah to complete a marathon of 42.2 km.
e. Determine a 95% prediction interval for the time for Sarah to complete a marathon of 42.2 km.
f. Explain the difference between your answers for (d) and (e)

20 Marks

2. The following data was used in an experiment studying the relationship between the number of grams of fertilizer per metre^2 applied to a field and the yield of pumpkins in kgs per metre^2:

x Fertilizer	y Yield
5	16
10	21
15	27
20	25
25	21

a. Set up the equations needed to calculate a least-squares quadratic model of the form: $y = b_0 + b_1x + b_2x^2$.

Assume:
$\Sigma x = 75$
$\Sigma y = 110$
$\Sigma x^2 = 1375$
$\Sigma xy = 1720$
$\Sigma x^2y = 3170$
$S_{yy} = 72$

b. Assuming that the $(X'X)^{-1}$ matrix is of the following form:

4.6000	-0.6600	0.0200
-0.6600	0.1069	-0.0034
*	-0.0034	0.0001

Compute the coefficients for the least-squares quadratic model (where * indicates a value you must find)

c. Using your equation from (b), compute the residuals.
d. Find SST, SSE, and SSR.
e. Create the ANOVA table for your model to test the hypothesis H_0: $B_1 = B_2 = 0$, using $\alpha = 0.01$.
f. Find the likely range for a single value (i.e. the next value) of $x = 22$ ($\alpha = 0.01$).
g. Find the likely range for a large group of values for $x = 50$ ($\alpha = 0.01$).

20 Marks

Section D: Complete two of the following three questions. This section is worth a total of 40 marks. Do not attempt all questions.

1. The following dataset is part of a 2^3 designed experiment to determine the effects of cutting speed (A), tool geometry (B), and cutting angle (C) on the life of a machine tool. Two levels of each factor were chosen and two replications of each design point were run.

Run	A	B	C	Rep 1	Rep 2
1	-1	-1	-1	22	31
2	1	-1	-1	32	43
3	-1	1	-1	35	34
4	1	1	-1	35	47
5	-1	-1	1	44	45
6	1	-1	1	40	37
7	-1	1	1	60	50
8	1	1	1	39	41

a. Complete the rest of the design matrix
b. Compute the contrast and mean effects for the A factor, the AB factor, and the ABC factor.
c. Calculate the sum of squares for the A factor, the AB factor, and the ABC factor.
d. Assuming the Contrasts and Sums of Squares listed below, calculate the ANOVA table for this model. Assuming a value of $\alpha = 0.05$, determine which effects, if any are significant. Please note that you will have to fill in the missing values in the Contrast and Sum of Squares tables before you can complete the ANOVA.

Contrasts

A	B	C	AB	AC	BC	ABC
*	47	77	**	-77	1	**

Sum of Squares

SS(A)	SS(B)	SS(C)	SS(AB)	SS(AC)	SS(BC)	SS(ABC)	SST
*	138.063	370.56	**	370.56	0.063	**	1163.44

20 Marks

2. Tom Thorton Inc. is a Hamilton based donut company. The firm is considering expanding its range of offerings and has conducted a consumer survey. The firm is testing three donut bases: cake, phyllo pastry, and yeast based along with four icing flavours: chocolate, vanilla, lemon and blueberry.

   Test versions of each of the potential offerings have been created and a sample of consumers has been engaged to rate the results on a scale of 0-100. The survey results appear below.

		Icing			
		Chocolate	Vanilla	Lemon	Blueberry
Base	Cake	84	60	88	90
		82	72	84	95
	Phyllo	44	88	22	10
		43	84	28	25
	Yeast	78	80	80	92
		79	92	83	94

   a. Perform an analysis of variance with $\alpha = 0.05$ to test for the significance of the main effects.
   b. Calculate R^2.
   c. Is this a good model? Why or why not?
   d. Is there an interaction effect between base and icing?

   To simplify your calculations, you may assume that SST = 15664.625.

20 Marks

3. Burr Industries of Pembroke, ON manufactures a variety of specialty metal based products for the Canadian Armed Forces, including magnesium snowshoes, aluminum sledges, and specialty scaling ladders. The firm is considering a change in their manufacturing process that would make their scaling ladders cheaper to manufacture, but the effect on the quality of the product (as measured by the maximum load in kgs) is unknown.

   Accordingly, the Burr Industries had ten prototypes built and tested. Results were then compared to a sample from the last batch of regular ladders produced. The results are given below:

Regular	Prototype
709	787
696	771
693	776
656	758
704	780
655	764
721	766
681	790
704	761
713	780

   a. Determine at the $\alpha = 0.05$ level whether the variance in the maximum load is different between the regular and prototype ladders
   b. Determine at the $\alpha = 0.05$ level whether the mean value of the maximum load between the regular and prototype ladders is significant. State any assumptions you make.

   To manufacture the new style of ladder in larger volumes, Burr will have to purchase a new computer controlled machining centre. Again, because quality is an issue, Burr has asked three machine vendors (A, B, & C) to produce a sample of 15 items, which were subjected to destructive testing to identify the maximum load of the scaling ladders. The results of this test are shown below:

	Vendor		
Sample	*A*	*B*	*C*
1	716	716	714
2	713	764	714
3	684	742	730
4	725	742	694
5	715	750	691
6	746	730	718
7	719	763	701
8	707	730	726
9	746	751	706
10	725	719	694

c. Conduct an analysis of variance (α =0.05) to determine if the machine vendor has a significant influence on the maximum load of the scaling ladder

20 Marks

APPENDIX 1 of 1

Runs Test

P(V <= v* when H_0 is true) In the Runs Test

	v*								
(n_1, n_2)	2	3	4	5	6	7	8	9	10
(2,3)	0.200	0.500	0.900	1.000					
(2,4)	0.133	0.400	0.800	1.000					
(2,5)	0.905	0.333	0.714	1.000					
(2,6)	0.071	0.286	0.643	1.000					
(2,7)	0.056	0.250	0.583	1.000					
(2,8)	0.044	0.222	0.533	1.000					
(2,9)	0.036	0.200	0.491	1.000					
(2,10)	0.030	0.182	0.455	1.000					
(3,3)	0.100	0.300	0.700	0.900	1.000	1.000			
(3,4)	0.057	0.200	0.543	0.800	0.971	1.000			
(3,5)	0.036	0.143	0.429	0.714	0.929	1.000			
(3,6)	0.024	0.107	0.345	0.643	0.881	1.000			
(3,7)	0.017	0.083	0.283	0.583	0.833	1.000			
(3,8)	0.012	0.067	0.236	0.533	0.788	1.000			
(3,9)	0.009	0.055	0.200	0.491	0.745	1.000			
(3,10)	0.007	0.045	0.171	0.455	0.706	1.000			
(4,4)	0.029	0.114	0.371	0.629	0.886	0.971	1.000	1.000	
(4,5)	0.016	0.071	0.262	0.500	0.786	0.929	0.992	1.000	
(4,6)	0.010	0.048	0.190	0.405	0.690	0.881	0.976	1.000	
(4,7)	0.006	0.033	0.142	0.333	0.606	0.833	0.954	1.000	
(4,8)	0.004	0.024	0.109	0.279	0.533	0.788	0.929	1.000	
(4,9)	0.003	0.018	0.085	0.236	0.471	0.745	0.902	1.000	
(4,10)	0.002	0.014	0.068	0.203	0.419	0.706	0.874	1.000	
(5,5)	0.008	0.040	0.167	0.357	0.643	0.833	0.960	0.992	1.000
(5,6)	0.004	0.024	0.110	0.262	0.522	0.738	0.911	0.976	0.998
(5,7)	0.003	0.015	0.076	0.197	0.424	0.652	0.854	0.955	0.992
(5,8)	0.002	0.010	0.054	0.152	0.347	0.576	0.793	0.929	0.984
(5,9)	0.001	0.007	0.039	0.119	0.287	0.510	0.734	0.902	0.972
(5,10)	0.001	0.005	0.029	0.095	0.239	0.455	0.678	0.874	0.958
(6,6)	0.002	0.013	0.067	0.175	0.392	0.608	0.825	0.933	0.987
(6,7)	0.001	0.008	0.043	0.121	0.296	0.500	0.733	0.879	0.966
(6,8)	0.001	0.005	0.028	0.086	0.226	0.413	0.646	0.821	0.937
(6,9)	0.000	0.003	0.019	0.063	0.175	0.343	0.566	0.762	0.902
(6,10)	0.000	0.002	0.013	0.047	0.137	0.288	0.497	0.706	0.864
(7,7)	0.001	0.004	0.025	0.078	0.209	0.383	0.617	0.791	0.922
(7,8)	0.000	0.002	0.015	0.051	0.149	0.296	0.514	0.704	0.867
(7,9)	0.000	0.001	0.010	0.035	0.108	0.231	0.427	0.622	0.806
(7,10)	0.000	0.001	0.006	0.024	0.080	0.182	0.355	0.549	0.743
(8,8)	0.000	0.001	0.009	0.032	0.100	0.214	0.405	0.595	0.786
(8,9)	0.000	0.001	0.005	0.020	0.069	0.157	0.319	0.500	0.702
(8,10)	0.000	0.000	0.003	0.013	0.048	0.117	0.251	0.419	0.621
(9,9)	0.000	0.000	0.003	0.012	0.044	0.109	0.238	0.399	0.601
(9,10)	0.000	0.000	0.002	0.008	0.029	0.077	0.179	0.319	0.510
(10,10)	0.000	0.000	0.001	0.004	0.019	0.051	0.128	0.242	0.414