MAT 2375
Final Examination

April 2013 **Professor G. Ivanoff**
Time: 3 hours

Student Number:____________ **Seat number:**____________

Family Name: ____________ **First Name:** ____________

- **This is an open book examination. Faculty of Science approved calculators are the only electronic device permitted.**
- **The exam consists of two parts: A (multiple choice) and B (long answer). For part A, record your answer to each question in the table provided on the next page. For part B, write your answers in the space provided on the questionnaire.**
- **At the end of the examination, hand in the entire questionnaire.**

****

Professor's use only:

Part A		
Part B	1	
	2	
	3	
	4	
	5	
TOTAL		

PART A: Multiple choice. 5 marks per question.

Put the letter corresponding to the correct answer in the space provided in the table below.

Question	1	2	3	4	5	6	7	8	9	10
Answer										

1. Porter Airlines knows from experience that 10% of passengers booked on their flights will be "no-shows" (i.e. passengers who do not show up for their flights). However, Porter is concerned that the proportion of no-shows on the small flights between Sudbury and Thunder Bay is greater than 0.10. Before changing its departure schedule, Porter has decided to carry out a test of hypothesis of $H_0 : p = 0.10$ against $H_1 : p > 0.10$. The company will reject H_0 and modify its schedule if there are 5 or more no-shows in a sample of 25 randomly chosen passengers booked on flights between Sudbury and Thunder Bay. What is α, the probability of a type I error? What $\beta(.20)$, the probability of a type II error if the real proportion of no-shows is $p = 0.20$?

   (A) $\alpha = 0.0334, \beta = 0.3833$
   (B) $\alpha = 0.0980, \beta = 0.5793$
   (C) $\alpha = 0.9020, \beta = 0.4207$
   (D) $\alpha = 0.0334, \beta = 0.6167$
   (E) $\alpha = 0.0980, \beta = 0.4207$

2. Let Y_n be the largest order statistic from a random sample of size n from a distribution with cumulative distribution function

$$F(x) = \frac{x^2}{a^2},\ 0 \leq x \leq a,$$

   where $a > 0$. (In other words, $Y_n = \max(X_1, ..., X_n)$, where $X_1, ..., X_n$ are i.i.d. F.) Find a constant c such that cY_n is an *unbiased* estimator of a.

   (A) $\frac{2n+1}{2n}$ (B) $\frac{n+1}{n}$ (C) $\frac{2n}{2n+1}$ (D) $\frac{n}{n+1}$ (E) 1

3. Many people believe that students gain weight during their first year at university. Six students were chosen at random to take part in a study. The students were weighed during the first week of the semester then again 12 weeks later. The following data was obtained:

Subject	Initial Weight=X	Final Weight=Y
1	171	168
2	110	111
3	134	136
4	115	119
5	150	155
6	104	106

Is there enough evidence to conclude that students gain weight during their first year at university? Test the null hypothesis that the average weight gain is 0 against the appropriate alternative hypothesis at level $\alpha = 0.05$. Give bounds for the p-value of your test statistic. (Assume that the difference in weight is normally distributed.)

(A) yes; p-value < 0.01
(B) yes; $0.01 <$ p-value < 0.025
(C) yes; $0.025 <$ p-value < 0.05
(D) no; $0.05 <$ p-value < 0.10
(E) no; p-value > 0.10

4. We want to test whether a random number generator produces samples from the uniform distribution on $[0, 1]$. We simulated 1000 numbers and counted how many observations fall into non-overlapping subintervals of $[0, 1]$. We obtained the following frequencies:

Interval	$[0, 0.2]$	$(0.2, 0.4]$	$(0.4, 0.6]$	$(0.6, 0.8]$	$(0.8, 1]$
Frequency	200	190	210	170	230

Is it reasonable to assume that the data come from the standard $U(0, 1)$ distribution? Perform a goodness of fit test with level $\alpha = 0.05$, report the observed value of the test statistic and state your conclusion. Hint: The cumulative distribution function F_0 of the uniform distribution on $[0, 1]$ is given by: $F_0(x) = x$ for all x in $[0, 1]$.

(A) $q = 10.0$, yes (B) $q = 10.0$, no (C) $q = 2.45$, yes (D) $q = 2.45$, no
(E) $q = 7.56$, no

5. A study is carried out to determine if there is a relationship between the regular consumption of various beverages and the recurrence of urinary infections (UI) in women. 100 women with urinary infections were first treated with antibiotics and then assigned randomly to one of three groups The subjects in the first group drank cranberry juice daily for 6 months. The subjects of the second group drank a lactobacillus beverage daily for 6 months. The subjects of the third group drank vitamin enriched water daily for 6 months.The number of cases of recurrence of UI was observed for each group, with the following results:

	Recurrence of UI	No UI	Total
Cranberry juice	8	22	30
Lactobacillus beverage	17	15	32
Vitamin water	25	13	38
Total	50	50	100

Can we conclude that the beverage affects the rate of UI recurrence? Calculate the observed value of the χ^2 test statistic and report your conclusion at level $\alpha = 0.05$.

(A) $q = 10.448$, no (B) $q = 5.991$, yes (C) $q = 3.842$, no
(D) $q = 3.842$, yes (E) $q = 10.448$, yes

6. A professor taught the same simulation course twice last year. A group of 10 students from the first class (class X) and 13 students from the second (class Y) are comparing their final grades. The first group had an average of $\bar{x} = 76.3$ and $s_X = 14.89$ while the other had an average of $\overline{y} = 69.2$ and $s_Y = 8.27$. Assuming that the grades are normally distributed, the students want to test the null hypothesis H_0 that the variances of the grades were equal for the two groups against the alternative H_1 that they were different. Which of the following statements is correct?

(A) Do not reject H_0 at level 10%.
(B) Reject H_0 at level 10% but not at level 5%.
(C) Reject H_0 at level 5% but not at level 2%.
(D) Reject H_0 at level 2% but not at level 1%.
(E) Reject H_0 at level 1%.

7. Among patients with lung cancer, at most 10% will survive for three years after diagnosis. Roche Holding, one of the world's largest manufacturers of cancer drugs, claims that its Tarceva medicine can extend the lifetime of patients with non-small-cell lung cancer. Assume that in a study of 150 patients with non-small-cell lung cancer treated with Tarceva, 22 survived for 3 years. Compute an 85% confidence interval for the proportion p of lung cancer patients treated with Tarceva, who will survive for 3 years after diagnosis. Based on this interval, can we conclude that the proportion p is greater than .10?

(A) [.090, .203]; we cannot conclude that p is greater than .10
(B) [.099, .194]; we cannot conclude that p is greater than .10
(C) [.105, .188]; we conclude that p is greater than .10
(D) [.113, .175]; we conclude that p is greater than .10
(E) [.087, .231]; we cannot conclude that p is greater than .10

8. The following observations are from a random sample from a continuous distribution F.

$$\begin{array}{cccccc} 77.6 & 78.1 & 75.3 & 76.2 & 76.5 & 79.2 \\ 74.8 & 75.6 & 76.4 & 77.8 & 78.2 & 78.6 \end{array}$$

Based on order statistics, what is the confidence coefficient of the interval $(75.3, 78.1)$ as a confidence interval for $\pi_{.4}$, the 40^{th} percentile of F?

(A) .9013 (B) .9238 (C) .9776 (D) .9651 (E) .9500

9. A paper producer sells paper that should have a mean weight of 75.0 gm/m^2 and a variance of 4. During an inspection, a sample of 30 is taken and the sample mean and sample variance are found to be 74.43 gm/m^2 and 4.389, respectively. The producer is concerned that the variance is too large and constructs a 90% lower confidence bound for σ^2. What is the bound, and would this lead the producer to conclude that the variance is greater than 4?

(A) 6.438, yes (B) 3.256, no (C) 2.991, no (D) 7.187, yes (E) 3.270, no

10. In a study that examines the damage caused by insects on cereals, researchers recorded the number of beetle larvae found on the oat stem for two groups of plants: the first group of plants was not treated with any pesticide, the second group was treated with an insecticide called malathion. For both groups, it can be assumed that the data is normally distributed but it is *not* assumed that the variances are equal. The data is summarized below:

	n	$\overline{x}$	s
no pesticide	53	3.47	1.21
malathion	64	2.36	0.52

Is there enough evidence to conclude that the insecticide has reduced the mean number of beetle larvae per plant? Test the null hypothesis that the means are equal at level $\alpha = 0.01$ against the appropriate alternative. Give bounds on the p-value of the test statistic and state your conclusion.

(A) p-value< 0.005; the insecticide has reduced the number of larvae
(B) 0.005$<$ p-value< 0.01; the insecticide has reduced the number of larvae
(C) $0.01 < p$-value < 0.025; not enough evidence that the insecticide has reduced the number of larvae
(D) $0.025 < p$-value < 0.05; not enough evidence that the insecticide has reduced the number of larvae
(E) $0.05 < p$-value < 0.10; not enough evidence that the insecticide has reduced the number of larvae

Part B: Long answer 10 marks per question

Write your answers directly on the questionnaire. Clearly define your notation and show all of your calculations.

1. (a) Let $X_1, ..., X_n$ be a random sample from a continuous distribution with density function

$$f(x) = \frac{1}{2\beta},\ \alpha - \beta \leq x \leq \alpha + \beta, \alpha \in \mathbf{R}, \beta \in \mathbf{R}_+.$$

(This is a two-parameter uniform distribution.) Find the method of moments estimators of α and β.

(b) Let $X_1, ..., X_n$ be a random sample from a normal distribution with *known* mean $\mu = 0$ and unknown variance $\sigma^2 = \theta$. Find the maximum likelihood estimator of θ.

2. Crystalline forms of certain chemical compounds are used in various electronic devices. It is often more desirable to have large crystals rather than small ones. In a laboratory study, 14 crystals of the same initial size were allowed to grow for certain periods of time. The following data gives the weight y of the crystal (in grams) and the period x of time (in hours) that was used to grow each crystal.

Time (x)	Weight (y)	Time (x)	Weight (y)
2	0.08	16	8.4
4	1.12	18	8.81
6	4.43	20	10.81
8	4.98	22	11.16
10	4.92	24	10.12
12	7.18	26	13.12
14	5.57	28	15.04

We have the following summary data:

$$\bar{x} = 15, \quad \bar{y} = 7.55, \quad \sum_{i=1}^{14}(x_i - \bar{x})^2 = 910, \quad \sum_{i=1}^{14}(x_i - \bar{x})(y_i - \bar{y}) = 458.12$$

$$\text{sum of squares of residuals} = \sum_{i=1}^{14}(y_i - \hat{y}_i)^2 = 13.53$$

(a) Calculate the least squares regression line for these data:

$$\hat{y} = \hat{\alpha} + \hat{\beta}(x - \bar{x}).$$

(b) Test $H_0 : \beta = 0$ against $H_1 : \beta > 0$ at level of significance $\alpha = 0.05$.

3. A manufacturer of paper used for making grocery bags is interested in improving the tensile strength of the bags. A team of engineers decides to investigate four levels of hardwood concentration in the paper: 5%, 10%, 15% and 20%. A sample of size 6 is collected for each of the four concentrations and the tensile strength of the paper is measured. The data is given in the following table:

	5%	10%	15%	20%	Totals
	7	12	14	19	
	8	17	18	25	
	15	13	19	22	
	11	18	17	23	
	9	19	16	18	
	10	15	18	20	
n_i	6	6	6	6	$n = 24$
$\sum_{j=1}^{6} x_{ij}$	60	94	102	127	$\sum_{i=1}^{4} \sum_{j=1}^{6} x_{ij} = 383$
$\sum_{j=1}^{6} x_{ij}^2$	640	1512	1750	2723	$\sum_{i=1}^{4} \sum_{j=1}^{6} x_{ij}^2 = 6625$

Are the four concentrations different from the point of view of the resulting tensile strength? Test the hypothesis: $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$, where μ_i is the mean tensile strength for the i^{th} concentration, $i = 1, 2, 3, 4$. Give bounds on the p-value of the test statistic, and state your conclusion at level $\alpha = 0.05$.

4. Cyclazocine is a drug used to treat heroine addicts. The effectiveness of the drug is evaluated using a test that measures the addict's psychological dependence on heroin. The result of the test is a Q-score; *high* Q-scores represent *less* dependence on heroin. The median Q-score for addicts not given cyclazocine is known to be 28. Fourteen (14) heroin addicts had the following Q-scores after cyclazocine therapy:

55	55	53	51	45	43	43
39	36	27	26	25	22	21

We will use a Wilcoxon statistic to test the null hypothesis that cyclazocine has no effect on the median Q-score m against the alternative that it lessens the addict's dependence on heroin.

(a) State an appropriate null hypothesis and one-sided alternative.

(b) Find the value of the test statistic.

(c) Find the approximate p-value of your test statistic, using a normal approximation and the appropriate continuity correction.

(d) State your conclusion about the effectiveness of cyclazocine at level $\alpha = 0.05$.

5. The oil company Shell is looking for additives that will increase gas mileage. As a pilot study, they send thirty-six (36) identical cars fueled with a new additive on a road trip from New York to Los Angeles. It is known from experience that without the additive, the same cars average 25.0 miles per gallon (mpg) with a standard deviation of 2.4 mpg. Let $X_i, i = 1, ..., 36$ denote the mpg of the cars when using the additive. It is assumed that the distribution of the X_i's will be normal with mean μ and the same standard deviation $\sigma = 2.4$ mpg. Shell will test the null hypothesis $H_0 : \mu = 25.0$ against the alternative $H_1 : \mu > 25.0$.

   (a) Find an appropriate test statistic and critical region for a test of level $\alpha = .10$.

(b) Find an expression for the power function $K(\mu)$ of your test. Find the values $K(25.25)$, $K(25.5)$, $K(25.75)$ and $K(26)$.

(c) Sketch a graph of the power function.