Carleton UNIVERSITY

FINAL EXAMINATION
August 16, 2008

DURATION: **3 HOURS** Number of Students: **43**

Department Name: Mathematics & Statistics
Course Number: STAT2507 A
Course Instructor: Wayne Horn

Non-Programmable Calculators Allowed

Students MUST count the number of pages in this examination question paper before beginning to write, and report any discrepancy to a proctor. This question paper has six (6) two-sided pages.

Students will be provided with a formula sheet and a sheet of tables by the professor.

This examination question paper may not be taken from the examination room.

This examination question paper may not be released to the library.

In addition to this question paper, students require: absolutely nothing.

Name: _____

Student Number: _____

There are **12** multiple choice questions worth a total of **12** marks.
There are **11** written questions worth a total of **68** marks.
The exam is out of **80** marks.
You have **3 hours** to complete the exam. **GOOD LUCK!**

DO NOT OPEN THE EXAM UNTIL YOU HAVE BEEN TOLD TO DO SO

Part A – Multiple Choice Questions

INSTRUCTIONS: Correct answers are worth **1 mark**. Incorrect/blank answers are worth **0 marks**. Please read each question carefully and clearly circle only one answer.

1. When is the t-distribution used in statistical inference?

(A) When the population standard deviations are not known and the sample sizes are small.

(B) When the population standard deviations are known.

(C) Either the standard normal or t-distribution may be used at anytime.

(D) When the samples are not random or when the population distributions are strongly skewed.

2. Weekly sales of regular ground coffee at a supermarket have in the recent past varied according to a normal distribution with mean of 354 units per week. In an attempt to increase sales, the supermarket reduces the price of its coffee by 5%. Is there good evidence that average sales are now higher? If a test is carried out, which of the following hypotheses should be tested to answer this question?

(A) $H_0 : p \geq 0.05$ versus $H_A : p < 0.05$

(B) $H_0 : p \leq 0.05$ versus $H_A : p > 0.05$

(C) $H_0 : \mu \geq 354$ versus $H_A : \mu < 354$

(D) $H_0 : \mu \leq 354$ versus $H_A : \mu > 354$

3. A researcher wants to conduct a hypothesis test at the $\alpha = 0.01$ significance level. She designs her study so

that the probability of a Type II error is 0.05 (for some fixed value of the parameter of interest). If the alternative hypothesis is correct, the probability that the researcher will reject the null hypothesis is

(A) 0.05 (B) 0.01 (C) 0.99 (D) 0.95

4. Which of the following statements about the Central Limit Theorem (CLT) is correct?

(A) The CLT states that the sample mean $\overline{X}$ is always equal to the population mean μ.

(B) The CLT states that the sample mean $\overline{X}$ is equal to the population mean μ provided that the sample size is large enough ($n \geq 30$).

(C) The CLT states that the sampling distribution of the sample mean $\overline{X}$ is exactly normal for large sample sizes ($n \geq 30$).

(D) The CLT states that the sampling distribution of the population mean μ is approximately normal when $n \geq 30$.

(E) The CLT states that the sampling distribution of the sample mean $\overline{X}$ is approximately normal for large sample sizes ($n \geq 30$).

5. Suppose a random variable follows a binomial distribution with 10 trials and probability of success as 0.8. Which of the following statements is true?

(A) The shape of the distribution is skewed to the right and normal approximation to binomial is not appropriate.

(B) The shape of the distribution is symmetrical and normal approximation to binomial is not appropriate.

(C) The shape of the distribution is skewed to the left and normal approximation to binomial is appropriate.

(D) The shape of the distribution is skewed to the left and normal approximation to binomial is not appropriate.

6. Suppose you have obtained a confidence interval for μ, but wish to obtain a greater degree of precision. Which of the following would result in a narrower confidence interval?

(1) Increasing the sample size while keeping the confidence level fixed.

(2) Decreasing the sample size while keeping the confidence level fixed.

(3) Increasing the confidence level while keeping the sample size fixed.

(4) Decreasing the confidence level while keeping the sample size fixed.

(A) 1 or 4 (B) 1 or 3 (C) 2 or 4 (D) 2 or 3 (E) None of these.

7. In testing $H_0 : \mu = 50$ versus $H_A : \mu \neq 50$, a random sample of 9 observations from a normally distributed population yielded a test statistic of $t_0 = 2.03$. The p-value for this test is

(A) greater than 0.10.

(B) between 0.05 and 0.10.

(C) between 0.025 and 0.05.

(D) between 0.01 and 0.025.

(E) less than 0.01.

8. You have measured the systolic blood pressure of a random sample of 25 employees of a company. A 95% confidence interval for the mean systolic blood pressure for the employees is computed to be (122, 138). Which of the following statements gives a valid interpretation of this interval?

(A) 95% of the sample of employees has a systolic blood pressure between 122 and 138.

(B) 95 % of the employees in the company have a systolic blood pressure between 122 and 138.

(C) If the sampling procedure were repeated 100 times, then approximately 95 of the sample means would be between 122 and 138.

(D) If the sampling procedure were repeated 100 times, then approximately 95 of the resulting 100 confidence intervals would contain the true mean systolic blood pressure for all employees of the company.

(E) We are 95% confident the sample mean is between 122 and 138.

9. Insurance companies are becoming concerned that increased use of cell phones is resulting in more car accidents. They want to estimate the proportion of drivers who talk on the cell phones while they are driving with 95% confidence. The acceptable margin of error is set at 3%. How many drivers must be sampled?

(A) 752 (B) 385 (C) 1068 (D) 1844 (E) 1067

10. You are told that the p-value for a hypothesis test was 0.03. This means:

(A) There is only a probability of 0.03 that the null hypothesis is true.

(B) If H_A is true, we have observed outcomes that occur 3% of the time.

(C) If H_0 is true, test statistic values that are at least as extreme as the value found from this sample occur 3% of the time.

(D) There is a probability of 0.03 that the alternative hypothesis is true.

11. Two populations follow normal distributions with unknown means μ_1 and μ_2, but known standard deviations σ_1 and σ_2, respectively. You take a random sample from each population and perform the following hypothesis test: $H_0 : \mu_1 - \mu_2 = 1$ versus $H_A : \mu_1 - \mu_2 \neq 1$. You obtain a p-value of 0.0418. Which of the following statements is/are definitely true?

(1) The corresponding two-sided 90% confidence interval for $\mu_1 - \mu_2$ will include the value 1.
(2) The corresponding two-sided 95% confidence interval for $\mu_1 - \mu_2$ will include the value 1.
(3) The corresponding two-sided 99% confidence interval for $\mu_1 - \mu_2$ will include the value 1.

(A) 1 and 2 (B) 2 and 3 (C) 1, 2, and 3 (D) 2 only (E) 3 only

12. A sample of 99 distances has a mean of 24 meters and a median of 24.5 meters. Unfortunately, it has just been discovered that an observation which was incorrectly recorded as "30" actually had a value of "35". If we correct this mistake in the data, which of the following things will happen?

(A) The mean remains the same, but the median is increased.

(B) The mean and median both remain the same.

(C) The median remains the same, but the mean is increased.

(D) We do not know how the mean and median are affected without access to the complete set of

measurements, but we know that the variance is increased.

(E) We can't say anything for certain without access to the complete set of measurements.

Part B – Written Questions

INSTRUCTIONS: Each question outlines its own marking scheme. Please read each question carefully. **For full marks, be sure to show all of your steps.**

1. Suppose the weights of packages of Oreo cookies have a normal distribution with a mean of 252 grams and a standard deviation of 9 grams. Let X represent the weight of a randomly selected package.

(A) (**2 marks**) What weight should be advertised on the packages so that only 5% of packages are below that weight?

(B) (**2 marks**) When selecting a random sample of 4 packages of Oreo cookies, what is the sampling distribution of the sample mean $\bar{X}$ and why?

(C) (**3 marks**) If you select 4 packages of Oreo cookies at random, what is the probability the average weight of these 4 packages will be between 250 grams and 255 grams?

2. It is known that a particular brand of computer has a 10% chance of being returned for repair/replacement during the warranty period, and that these computers malfunction independently of each other. 1500 of these computers have been sold to ABC Manufacturing.

(A) (**1 mark**) Let X represent the number of computers that are returned by ABC Manufacturing during the warranty period. What is the exact distribution of X?

(B) (**2 marks**) What is the approximate distribution of X? Justify your answer.

(C) (**2 marks**) What is the approximate probability that ABC Manufacturing will return at least 160 of these computers during the warranty period?

3. A quality control technician at Moo Moo Milk Inc. took a random sample of 21 "2 litre" milk containers from a one hour production run. The weights (in litres) of the 21 containers are given in the following (split) stem-and-leaf plot.

LEAF UNIT = 0.01

19 | 123
19 | 5566789
20 | 111244
20 | 5778
21 |
21 | 6

(A) (**3 marks**) What are the values of the median, lower quartile (Q_1), and upper quartile (Q_3) for this data set?

(B) (**2 marks**) Determine if there are any outliers in this data set using an appropriate method.

4. A student is preparing for an upcoming exam. The professor for the course has given the class 30 questions to study from and plans to select 10 of the questions at random for use on the actual exam. Suppose the student knows how to solve 25 of the 30 questions.

(A) (**2 marks**) What is the probability that the student will get perfect on the test?

(B) (**2 marks**) What is the probability that the student will get at least 8 questions correct on the test?

5. An agricultural researcher plants 25 randomly selected plots with a new variety of corn. From these 25 plots, the average yield was 150 bushels per acre. The population standard deviation is known to be 20 bushels per acre.

(A) (**4 marks**) Give a 99% confidence interval for the population mean yield μ bushels per acre. Interpret your answer in terms of the question. What do you need to assume in order for your confidence interval to be valid?

(B) (**2 marks**) Suppose the researcher wanted to estimate the true average yield, to within 5 bushels per acre, with 99% confidence. How many plots of land should he use in order to achieve the desired margin of error?

6. Suppose you take a simple random sample of n observations from a normal population with known standard deviation σ. Based on this SRS you compute a 95% confidence interval for the population mean μ.

(A) (**1 mark**) What is the length of this interval, in terms of n and σ?

(B) (**3 marks**) Describe what would happen to the length of the confidence interval, relative to the length of the interval in part (A), if the sample size is increased to $4n$ and the confidence level is increased to 97%, everything else remaining constant.

7. A fashion industry analyst suspects that the models featuring Liz Claiborne clothing earn less, on average, than models featuring clothes designed by Calvin Klein. For a given period of time, a random sample of 8 Liz Claiborne models revealed average earnings of \$3,888 with a standard deviation of \$1002. For the same period, an independent random sample of 10 Calvin Klein models had mean earnings of \$4,238 and a standard deviation of \$876. The distributions of salaries for both groups are approximately normally distributed.

(A) (**3 marks**) Conduct an appropriate hypothesis test at $\alpha = 0.05$ to determine whether or not you should assume $\sigma_1^2 = \sigma_2^2$.

(B) (**5 marks**) Conduct an appropriate hypothesis test to test analyst's suspicion at the 5% level of significance.

8. An experiment was conducted to test the effect of a new drug on a viral infection. The infection was induced in 100 mice, and the mice were randomly split into two groups of 50. The first group, the control group, received no treatment for the infection. The second group received the drug. After a 30-day period, the proportions of survivors, p_1 and p_2, in the two groups were found to be 0.36 and 0.60, respectively.

(A) (**3 marks**) Give a 95% confidence interval for the difference between the cure rates. Interpret the result in terms of the question.

(B) (**4 marks**) Is there sufficient evidence to indicate that the drug is effective in treating the viral infection? Conduct a hypothesis test at the 5% level of significance.

9. (**5 marks**) In a test of the effectiveness of a device intended to increase gas mileage, four cars were operated with and without the device. The mileages (in mpg), as well as some sample statistics, are given in the table below.

CAR	WITHOUT DEVICE	WITH DEVICE	DIFFERENCE
1	18.5	22.5	–4
2	30.7	31.7	–1
3	27.8	28.8	–1
4	31.5	33.5	–2
$\bar{x}$	27.125	29.125	–2
s	5.97	4.82	1.41

The researcher wishes to show that the device actually increases the average mileage. Conduct an appropriate hypothesis test at $\alpha = 0.10$. (HINT: Are the measurements dependent or independent?)

10. Building contracts were issued last year to three contractors starting up a new subdivision of 20 houses. Tara Construction (T) and Westview Construction (W) built 5 houses each. The most active contractor, Hearthstone Construction (H), built ten houses. Tara's houses have a 20% probability of developing leaky basements (L); homes built by Westview and Hearthstone have the same problem 10% and 5% of the time, respectively.

(A) (**3 marks**) What is the probability that a randomly selected home will develop a leaky basement?

(B) (**3 marks**) The Better Business Bureau received a complaint from one of the new homeowners that his basement is leaking. What is the probability that the house was built by Hearthstone Construction?

11. A computer supply house receives a large shipment of blank CDs each week. Past experience has shown that the number of flaws per CD can be described by the following probability distribution, where the random variable X represents the number of flaws per CD.

x	$P(X=x)$
0	0.65
1	0.20
2	0.10
3	0.05

(A) (**3 marks**) Calculate the mean and standard deviation of the number of flaws per CD.

(B) (**4 marks**) Suppose you select two CDs at random. What is the sampling distribution of the average number of flaws on the two CDs? That is, what is the sampling distribution of the sample mean $\overline{X}$?

(C) (**2 marks**) Suppose you select 100 CDs at random. What is the (approximate) sampling distribution of the sample mean $\overline{X}$? Why?

(D) (**2 marks**) Suppose you select 100 CDs at random. Approximate the probability that the sample mean $\overline{X}$ is less than 0.4.