

1. (7 marks) The boxplots below were constructed from data collected on the calories contained in beef, meat (mixture) and poultry hot dogs of the same weight.
 - (a) What are the approximate median calorie counts for each of the hotdog types?
 - (b) Can you be certain that by eating a poultry hot dog your calorie consumption will be lower than if you were to eat either of the remaining types. Explain your answer briefly.

2. (10 marks) You have torn a tendon and are facing surgery to repair it. The orthopedic surgeon explains the risks to you. Infection occurs in 3% of such operations, the repair fails in 14%, and both infection and failure occur together in 1%. What percent of these operations succeed and are free from infection?

3. (10 marks) A study of Georgia residents suggests that those who worked in shipyards during World War II were subjected to a higher risk of lung cancer (*Wall Street Journal*, September 21, 1978). It was found that approximately 22% of those persons who had lung cancer worked in a shipyard at some prior time. In contrast, only 14% of those who had no lung cancer had worked in a shipyard. Suppose that the proportion of all Georgians living during World War II who have contracted or will contract lung cancer is .04 percent. Find the percentage of Georgians living during the same period who will contract (or have contracted) lung cancer, given that they worked in a shipyard at some time.

4. (5 marks) A heavy-equipment salesman can contact either one or two customers per day with probability $1/3$ and $2/3$, respectively. Each contact will, independently, result in either no sale or a \$50,000 sale with probability $9/10$ and $1/10$, respectively. What is the expected value of his daily sales?
- (Hint: Write down the three possible values for the total daily sale together with their probabilities, by using the law of total probability.)
5. (10 marks) A patient is said to be hypokalemic (low potassium in the blood) if the measured level of potassium is 3.5 or less. (The units for this measure are meq/l, or milliequivalents per liter.) An individual's potassium level is not a constant, however, but varies from day to day. In addition, the measurement procedure itself has some variation. Suppose that the overall variation follows a normal distribution with a mean potassium level of 3.8 with a standard deviation of 0.2.
- (a) What proportion of the population are hypokalemic?
- (b) If 10 people are selected at random, what is the probability that at least 3 will be hypokalemic? (Leave your answer in unsimplified form.)
6. (10 marks) Suppose that the times between “exceptionally cold” winters are roughly independent and have a mean of 15.2 yrs with a standard deviation of 4.6 yrs. What is the (approximate) probability that the time elapsed from now until 40 such winters have occurred will be greater than 630 years?
7. (15 marks) Poisoning by the pesticide DDT causes tremors and convulsions. In a study of DDT poisoning, researchers fed several rats a measured amount of DDT. They then measured electrical characteristics of the rats' nervous systems that might explain how DDT poisoning causes tremors. One important variable was the “absolutely refractory period”, the time required for a nerve to recover after a stimulus. This period varies normally. Measurements on four rats gave the data below (in milliseconds). (Data from D.L. Shankland, “Involvement of spinal cord and peripheral nerves in DDT-poisoning syndrome in albino rats”, *Toxicology and Applied Pharmacology*, 6 (1964), pp. 97-213.)

1.6 1.7 1.8 1.9

- (a) Give a 90% confidence interval for the mean “absolutely refractory period” for all rats of this strain when subjected to the same treatment.
- (b) Interpret this interval.
- (c) It is desired to test the hypothesis that the mean refractory period for the population of rats from which these came is 1.3ms against the hypothesis that DDT increases the refractory period. Carry out such a test at the $\alpha = .05$ level. (That is, it is suspected that the population of rats was contaminated with DDT.)

8. (5 marks) An economist wishes to find a 95% confidence interval of maximum width 6 percentage points, for the proportion of unemployed Canadians. How many Canadians should be included in her sample to achieve her specifications?
9. (5 marks) Jan Lindhe, D.M.D., conducted a study on the effect of an oral antiplaque rinse on plaque buildup on teeth (“Clinical Assessment of Anti-plaque Agents”, *The Compendium of Continuing Education in Dentistry*, suppl. no. 5[1984]). Fourteen subjects, whose teeth were thoroughly cleaned and polished, were randomly assigned to two groups of seven subjects each. Both groups were assigned to use oral rinses (no brushing) for a two-week period. Group 1 used a rinse that contained an anti-plaque agent. Group 2, the control group, received a similar rinse except that, unknown to the subjects, the rinse contained no antiplaque agent. A plaque index y , a measure of plaque buildup, was recorded at 4, 7 and 14 days. The mean and standard deviation for the 14-day plaque measurements are shown below for the two groups.

	Control group	Antiplaque group
Sample Size	7	7
Mean	1.26	.78
Standard Deviation	.32	.32

- (a) State the null and alternative hypotheses that should be used to test the effectiveness of the antiplaque oral rinse.
- (b) Do the data provide sufficient evidence to indicate that the oral antiplaque rinse is effective? Test using $\alpha = .05$.
10. (15 marks) A pollution control inspector suspected that a river community town was releasing amounts of semitreated sewage into a river. To check his theory, he drew five randomly selected specimens of river water at five different distances above the town and another five below, at distances corresponding to those above the town. The dissolved oxygen readings, in parts per million, are as follows:

Distance	1	2	3	4	5
Above town	4.8	5.2	5.0	4.9	5.1
Below town	5.0	4.7	4.9	4.8	4.9

- (a) Do the data provide sufficient evidence to indicate that mean oxygen content between locations above and below the town is less than the mean oxygen content above? Test using $\alpha = .05$.
- (b) Suppose that you prefer estimation as a method of inference. Estimate the difference in mean dissolved oxygen content between locations above and below the town. Use a 95% confidence interval.

11. (8 marks) Indicate which of the statements below is true and which is false. For those that are false, write the correct statement.
- (a) Concerning the 95% confidence interval $(1.23, 1.87)$, for μ , the most that can be said is that roughly 95% of the time in repetitions of the experiment, this interval will capture μ .
 - (b) The random confidence interval for \bar{X} , $\left(\bar{X} - z_{\alpha/2} \frac{s}{\sqrt{n}}, \bar{X} + z_{\alpha/2} \frac{s}{\sqrt{n}}\right)$ will, with probability .95 contain \bar{X} .
 - (c) If the sample size is large enough, the Central Limit Theorem says the observations can be regarded as being approximately Normally distributed.
 - (d) In carrying out an hypothesis test, your test statistic falls in the rejection region and the level of significance is $\alpha = .05$. The best you can say is that the rejection rule you have used, will, on repeated use, lead to a Type I error about 5% of the time.

McGILL UNIVERSITY
FACULTY OF SCIENCE

FINAL EXAMINATION

MATHEMATICS 189-203B

PRINCIPLES OF STATISTICS I

Examiner: Professor D. Wolfson
Associate Examiner: Professor K. Worsley

Date: Friday, April 23, 1999
Time: 2:00 P.M. - 5:00 P.M.

INSTRUCTIONS

Answer all questions.
Tables have been provided.

This exam comprises the cover, 4 pages of questions and 2 pages of tables.