JOHN ABBOTT
CEGEP/COLLEGE

Department of Mathematics
Final Examination

May 21, 2013

14h–17h

Statistical Methods
201-DDD-05

Instructors: Olivier Dubois, Moushira Guindi

Student name: _____

Student number: _____

Instructor: _____

Instructions

1. Do not open this booklet before the examination begins.

2. Check that this booklet contains 13 pages, excluding this cover page.

3. Write all of your solutions in this booklet and show all supporting work.

4. If the space provided is not sufficient, continue the solution on the opposite page.

5. Keep at least 4 decimals in all calculations.

1. (3 points) Suppose you have n temperature readings, in degrees Celsius ($^\circ$C). The sample mean is $37\,^\circ$C and the sample standard deviation is $2\,^\circ$C. What would be the mean and standard deviation if the readings are first converted to degrees Fahrenheit ($^\circ$F)? Carefully prove your answer. (Recall the conversion formula $^\circ$F $= \frac{9}{5}\,^\circ$C $+ 32$.)

2. (8 points) The degree of muscle deterioration for astronauts is classified as either negligible, moderate or severe. It is known that astronauts suffer from severe deterioration with probability 0.25, and moderate deterioration with probability 0.6. Two astronauts are given medical exams after returning from the space station. What is the probability that:

(a) both astronauts have severe muscle deterioration?

(b) both astronauts have the same level of muscle deterioration?

(c) at least one astronaut has severe muscle deterioration?

(d) both astronauts have negligible muscle deterioration given it is not severe for either?

3. (5 points) Three technicians regularly make repairs when breakdowns occur on an automated production line. Janet, who services 20% of the breakdowns, makes an incomplete repair 1 time in 20; Tom, who services 60% of the breakdowns, makes an incomplete repair 1 time in 10 and Georgia, who services 20% of the breakdowns, makes an incomplete repair 1 time in 15. A problem now occurs with the production line, and it is diagnosed as being due to an incomplete repair; what is the probability that this incomplete repair was made by Janet?

4. (4 points) Five engineering firms $(F_1, F_2, F_3, F_4, F_5)$ are bidding for three separate city contracts (C_1, C_2, C_3).

(a) Suppose the city randomly assigns contracts to the firms, allowing a firm to get multiple contracts. Find the probability that firm F_1 obtains exactly one contract.

(b) Suppose the city randomly assigns contracts to the firms, but this time without allowing a firm to get multiple contracts. Find the probability that the firm F_1 gets the first contract C_1.

5. The wind velocity (in km/h) is measured at a certain fixed location at different times. At a given time, the wind velocity (X) can be modeled as a continuous random variable with the following **cumulative** distribution function

$$F(x) = \begin{cases} 0 & \text{for } x < 0, \\ \dfrac{(x-100)^3}{100^3} + 1 & \text{for } 0 \le x \le 100, \\ 1 & \text{for } x > 100. \end{cases}$$

(a) (2 points) Find the probability that the wind velocity is between 10km/h and 80km/h.

(b) (3 points) Find the median wind velocity.

(c) (2 points) Recover the density function $f(x)$ for this random variable.

6. You have a box containing 5 lug nuts and 10 bolts. You grab a random handful of 4 pieces without looking. If you get no bolts, then you look and take one bolt from the box. Let X be the number of bolts you obtain this way. (Note that it cannot be 0!).

  (a) (4 points) Find the probability distribution $p(x)$ of this random variable. Is it symmetric?

  (b) (2 points) Plot the cumulative distribution $F(x)$ for this random variable.

  (c) (3 points) Calculate $E(X)$ and $V(X)$.

7. Suppose that across all students in CEGEP, the R-score is normally distributed with mean 25 and standard deviation 3.

(a) (3 points) Suppose that McGill only considers applicants with R-scores of at least 32 for its medical school, and with R-scores of at least 27 for its mathematics program. Find what proportion of students are eligible for each one of those programs.

(b) (3 points) Suppose a specific program wants to accept only students among the top 3% in terms of R-scores. What is the minimum R-score they should require?

(c) (3 points) If a random sample of 35 students was taken, what is the probability that the sample mean R-score is between 24 and 26?

8. A dentist makes appointments with his patients; however, each patient has a probability of 0.06 of cancelling the appointment.

   (a) (2 points) Suppose that 12 appointments are scheduled for tomorrow. Find the probability that at least two of those appointments will be cancelled.

   (b) (3 points) Suppose that 2000 appointments will be scheduled next year. Find the probability that at least 130 appointments will be cancelled. (Hint: use a normal approximation.)

9. Dr. Wilson writes prescriptions following a Poisson distribution with an average of 3 per day. Dr. House gives either exactly 3 prescriptions per day when he's in a good mood (which happens with probability 0.01), otherwise he gives no prescriptions.

   (a) (3 points) Find the probability that Dr. Wilson will give 13 prescriptions next week (consisting of 5 business days).

(b) (3 points) Find the probability that, tomorrow, Dr. House and Dr. Wilson will write a total of 6 prescriptions. Suppose that these doctors work independently.

10. A random sample of 100 two-month-old babies is obtained, and the mean head circumference is found to be 40.6 cm. Assume the distribution of head circumference is normal and the population standard deviation is 1.6 cm.

    (a) (4 points) Test using the classical approach the claim that the mean head circumference of all two-month-old babies is greater than 40 cm. (Use a 5% significance level.)

(b) (3 points) When doing a test as in part (a) but without knowing the sample mean yet, if the true mean head circumference is actually 40.2 cm, find and **interpret** the probability of making a type II error.

(c) (3 points) Find a 95% confidence interval for the true mean head circumference of two-month old babies.

(d) (3 points) What sample size is required if the width of the interval is not to exceed 0.5 and the confidence level should not be less than 90% ?

11. (3 points) A sample from a normal distribution was taken and the lower bound of the 95% confidence interval for the population standard deviation was found to be 35.4612. However, the raw data was lost. We recall that the sample size was 28. What was the sample standard deviation?

12. (2 points) A sample of 24 observations from a normal population was obtained to test whether or not the population variance equals 6. What is the p-value associated with the test statistic $\chi^2 = 41.637$?

13. A random sample of 16 females and 13 males that applied to get into pre-med at McGill was taken and their R-score was recorded. The following is a summary of the findings.

	Sample size	Sample mean	Sample standard deviation
Females	16	36.11	2.36
Males	13	35.13	1.47

Assume the distribution of the R-scores is normal.

(a) (4 points) Using the p-value approach, test whether or not the population variances are equal. Use a 10% significance level. **More space is available on the next page.**

(b) (4 points) Test the hypothesis that the average R-score of the girls applying to pre-med is higher than that of the boys. (Use a 1% significance level and the classical approach.)

14. Suppose a group of 900 smokers (who all wanted to give up smoking) were randomly assigned to either receive a placebo or an anti-depressant drug for 6 weeks. Of the 148 patients who received the antidepressant drug, 23 were no longer smoking a year later. Of the 752 patients who received the placebo, 97 were not smoking one year later.

  (a) (3 points) Find a 90% confidence interval for the true difference in the rates of success achieved by using a placebo and that achieved in using the antidepressant.

  (b) (2 points) From the interval found in part (a), is it plausible that there is no difference in the rates of success achieved by using a placebo and that achieved in using the antidepressant? Justify your answer.

15. (5 points) Students in a statistics class at a large university were classified by birth order (among their siblings) and by discipline they study. Is there a relationship between birth order and the discipline they study? Perform the appropriate test using the classical approach with a 5% level of significance, and interpret your conclusion.

		Birth order		
		First	Second or later	Total
Discipline	Arts/Science	34	23	57
	Agriculture	52	41	93
	Other	28	46	74
	Total	114	110	224

16. It is difficult to accurately determine a persons body fat percentage without immersing him or her in water. Researchers hoping to find ways to make a good estimate immersed 9 male subjects, then measured their waist (in inches) and recorded their weights (in pounds).

Waist (X)	32	36	38	33	39	40	41	35	38
Weight (Z)	175	181	200	159	196	192	205	193	187
Body Fat Percentage (Y)	6	21	15	3	22	31	32	21	25

$$\sum x_i = 332 \quad \sum x_i^2 = 12324 \quad \sum x_i y_i = 6712$$
$$\sum z_i = 1688 \quad \sum z_i^2 = 318190 \quad \sum z_i y_i = 33880$$
$$\sum y_i = 176 \quad \sum y_i^2 = 4246$$

(a) (5 points) Which of these, waist size (X) or weight (Z), give a better indicator of body fat percentage (Z)? In other words, if you were attempting to predict the body fat percentage of a person, would it be more accurate to use waist size or weight as a basis for your prediction? Justify using the notion of correlation.

(b) (3 points) Find the least-squares regression line using the better indicator as the independent variable.