Final Assignment

Due **9:30am** December 21, 2016

201-922-DW (Introduction to Statistical Methods)

Instructor: Émilie Richer (Department of Mathematics)

Worth 10 marks on final grade

This assignment is OPTIONAL

If you do not submit the assignment, your grade will be the one currently posted on *Omnivox*.

Please note that if you currently have a grade of **LESS THAN 44 out of 80** you are no longer in a position to pass this class, even with the successful completion of this assignment. Nevertheless you can still hand in the assignment to improve your grade.

Submission Guidelines:

The assignment can be submitted electronically as a PDF document to ericher@dawsoncollege.qc.ca OR can be dropped off at **office 8A.11**

Show all of your work, no marks will be awarded if complete solutions are not provided.

[Question 1]

A medical clinic in Chicago classifies the patients' files by gender and by type of diabetes (A or B). The number of patients in each classification is shown below:

	Type of Diabetes	
Gender	A	B
Male (M)	50	40
Female (F)	70	40

One file is selected at random.

i) Find the probability that the selected individual is female.

ii) Find the probability that the selected individual is Type B.

iii) Find the probability that the selected individual is Type A male.

[Question 2]

An experiment consists of drawing a card at random from a standard shuffled deck of 52 cards. Find the probability of the following events:

i. Drawing a diamond

ii. Drawing a queen

iii. Drawing a diamond or a queen

iv. Drawing at least two diamonds

[Question 3]

Consider the experiment of tossing a fair die two times, with the outcomes being observing either a number less than 4 or a number greater or equal to 4.

i. Construct a tree diagram for the experiment.

ii. Construct the sample space for the experiment.

iii. Find the probability of observing one number less than 4 and one number greater or equal to 4.

iv. Find the probability of observing two numbers less than 4.

v. Construct the probability distribution for the number of numbers less than 4 observed.

[Question 4]

Events A, B and C are events of a sample space S with A and C mutually exclusive, B and C mutually exclusive, $P(A) = 0.32$, $P(B) = 0.11$, $P(A \cap B) = 0.08$, and $P(C) = 0.42$. Find the following probabilities:

$P(A \cup B)$

$P(A \cup C)$

$P(B \cup C)$

$P(A')$

$P(C')$

$P(A \cap C)$

[Question 5]
Consider the following set of sample data given as a frequency table:

Value	Frequency
5	9
-3	6
4	10
3	12
2	23

For the data set, using your calculator, compute:

a) $\sum(x_i)$

b) $\sum(x_i)^2$

c) the sample mean $\bar{x}$

d) the sample standard deviation s_x

e) the total variation SS(X)

[Question 6]
SAT scores of students at an Ivy League college are distributed with a standard deviation of 250 points. Two statistics students, Raina and Luke, want to estimate the average SAT score of students at this college as part of a class project. They want their margin of error (E) to be no more than 25 points.

(a) Raina wants to use a 90% confidence interval. How large a sample should she collect?

(b) Luke wants to use a 99% confidence interval. Without calculating the actual sample size, determine whether his sample should be larger or smaller than Raina's, and explain your reasoning.

(c) Calculate the minimum required sample size for Luke

[Question 7]
New York is known as "the city that never sleeps". A random sample of 42 New Yorkers were asked how much sleep they get per night. Statistical summaries of these data are shown below.

$\bar{x} = 7.73$

$s = 0.77$

Do these data provide strong evidence that New Yorkers sleep less than 8 hours a night on average? Conduct a hypothesis test with $\alpha = 0.01$.

[Question 8]
Friends Brian and Melanie both completed the Montreal marathon, where Brian competed in the Men, Ages 30–34 Group while Melanie competed in the Women, Ages 25-29 Group. Brian completed the race in 3:05:28 (11,128 seconds), while Melanie completed the race in 3:31:53 (12,713 seconds). Here is some information on the performance of their groups:

- The finishing times of the Men, Ages 30-34 Group has a mean of 12,521 seconds with a standard deviation of 1028 seconds.
- The finishing times of the Women, Ages 25-29 Group has a mean of 13,210 seconds with a standard deviation of 1201 seconds.
- The distributions of finishing times for both groups are approximately Normal.

Remember: **a better performance corresponds to a faster finish.**

(a) What are the Z-scores for Brian's and Melanie's finishing times? What do these Z-scores tell you?

(b) Did Brian or Melanie rank better in their respective groups? Explain your reasoning.

(c) What percent of the marathoners did Brian finish faster than in his group?

(d) What percent of the marathoners did Mary finish faster than in her group?

(e) Compute the "cut-off time" for the **fastest 5%** of athletes in the men's group, i.e. those who took the shortest 5% of time to finish.

(f) The "cut-off time" for the **slowest 10%** of athletes in the women's group

[Question 9]
A 95% confidence interval for a population mean, μ, is given as (18.985, 21.015). This confidence interval is based on a simple random sample of 36 observations. Calculate the sample mean and standard deviation. Assume that all conditions necessary for inference are satisfied. Use the t-distribution in any calculations.