Version A

Math 203: Principles of Statistics I

February 28, 2014

Time: 50 Minutes

Please print your name and student number legibly.

Last Name: _____

First Name: _____

Student Number: _____

Instructions:

1. Answer all questions
2. Show your reasoning briefly
3. Hand calculators are permitted
4. Crib sheets are NOT permitted.
5. Leave your answers in unsimplified form.
6. The marks allocated are shown in [**brackets**].
7. Do not forget to write your name and student number above.

This exam comprises 6 pages including the cover page.

Question 1: **[10]** Aviation inspectors are given information on the ages of the pilots of planes involved in 60 crashes. It is known that in 20% of these crashes the pilot was younger than 50 years old. The inspectors pick 10 of these crashes at random, without replacement. Assuming that the same pilot was not involved in more than one crash:

i) What is the probability that exactly 2 of the 10 selected crashes involved a pilot who was at least 50 years old?

ii) What is the probability that at most 2 of the selected crashes involved a pilot who was at least 50 years old?

Question 2: **[10]** Canada possibly has the highest prevalence of multiple sclerosis (MS) in the world. Two factors that are thought to be associated with the risk of developing MS are low levels of vitamin D and having had mononucleosis (mono). Suppose that 10% of Canadians have low vitamin D levels and 5% have had mono.

i) What proportion of Canadians will have at least one of the risk factors above? What assumption have you made in order to compute this proportion?

ii) Under scenario i), what proportion of Canadians will have had neither mono nor low levels of Vitamin D?

Question 3: [**10**] The following data were collected by food inspectors on the number of parts per million of a certain contaminant in a brand of bottled drinking water:
10.2, 11.3, 8.4, 9.6, 30.4, 11.2, 8.5, 7.5, 7.2, 6.7, 1.3, 11.9, 10.4

i) Considering two well known measures of centrality, in a **single** sentence give a reason why you think one measure is to be preferred over the other for these data.

ii) Suppose that these data had been converted into units of parts per billion (1000 million). If you had calculated the sample standard deviation using the original data set and the data set with the converted units, would you have noticed a difference in your sample standard deviation? Give a reason for your answer. There is no need to actually carry out any calculations.

Question 4: **[10]** It is known that 70% of cervical cancers are caused by a particular class of human papilloma viruses. It is also known that if a woman does not have cervical cancer there is a probability of 0.90 that she has not been infected by a virus in this class. Assume that 1% of women have cervical cancer.

If a woman is found to have been infected by a virus in this class, what is the probability that she has cervical cancer?

Question 5: **[10]** Refer to Question 4. If, in this very large population of women, a sample of 10 is selected without replacement, what is the probability that at least two will have cervical cancer?