uOttawa
Faculté des sciences sociales | Faculty of Social Sciences
Département de science économique
Department of Economics

ECO3150
Introduction to Probability and Statistics

David Gray — Fall 2014

There are two parts to this examination – one with written questions and one with problems to solve. There are a total of 100 points. Explain your work and show your calculations. You can have a total of 4 hours to complete the exam. Please budget your time accordingly, and do not allocate too much time to any one question. I will be evaluating them on the basis of style as well as content. Please explain your work carefully, using sentences to justify your steps. Skimpy responses and sloppy work will cause points to be deducted. If you do not show how a certain result was obtained, you will not receive much credit. You may not consult with any written documents. I have attached the tables for the standard normal distribution. No conversation with other students is permitted when the examination is in progress. Glancing at other students' papers is forbidden. Calculators are permitted, but they should not be passed between students. If you are in need of some formula, you may approach me. I may or may not satisfy your curiosity. Good luck.

A. Problems (48 points)

1. (9 points) Suppose that a person's IQ (intelligence quotient) is a random variable that follows a normal distribution with mean μ = 100 and variance σ^2 = 225.

a) Find the probability that the realisation for this variable is greater than 125, which indicates that he/she is pretty darn smart.
b) Suppose that the probability that the IQ of a person is greater than a certain threshold is 0.75, which would place him/her among the three uppers quartile of the distribution. What is this value?
c) Suppose that you have a random sample comprised of 9 persons. Find the probability that the realisation for that variable (i.e. the sample mean) is greater than 125. Compare that result with your response for part a). It should be lower. Why?

2. (4 points) Consider random variables X and Y. Demonstrate the validity of the following identity, and explain your work. Note that in posing this question to you, I am giving you the formula for the covariance.

$$COX(X_i, Y_i) = E[(X_i - \mu_x)(Y_i - \mu_y)] = E(X_i Y_i) - \mu_x \mu_y$$

3. (6 points) According to a recent survey that deals with the death penalty, 64 % of the respondents indicated that they were in favor. 48 % of all of the respondents were women, and conditioned on being a woman, 46 % were in favor of the death penalty. Hint: Note that these are proportions.

a) What is the probability that a person selected randomly is <u>both</u> a woman <u>and</u> a supporter of the death penalty?
b) If you were to select in random fashion a sample of 250 women from the population, what would be the probability that the proportion that supports the death penalty surpasses 40 %?

4. (10 points) Given that Z is a standardized normal random variable, find the value of z (the threshold) in the following cases. The figures below are probabilities.

a) the probability for the area to the right of z is 0.025
b) the probability for the area to the right of z is 0.10
c) the probability for the area to the left of z is 0.025
d) the probability for the area to the left of z is 0.10
e) In the table for the standard normal distribution, one does not see probabilities for Z that are greater than 4 in absolute value. How can one obtain values for the probabilities corresponding to Z values that are greater than 4.5 or less than negative 4.5?

5. (6 points) Consider an establishment that sells poutines for 1.45 $ apiece. (This is a midnight madness sale.) The total sales per day have a mean value of 530 units and a standard deviation of 69 units (number of servings of poutine).

a) What is the average value for total revenues per day?
b) What is the standard error for total revenues per day?
c) The production costs in $ per day are calculated by the following function:

C = 100 + 0.95 X, where X is the number of servings of poutine that are sold. What are the values of the mean and the standard error of the variable C per day?

6. (13 points) Consider the following joint probability distribution for the random variables X and Y:

		X = 1	X = 2
Y	0	0.7	0.0
	1	0.0	0.3

a) Are these random variables discrete or continuous? Why?
b) Compute the marginal probability functions for these two random variables.
c) Compute the population means of X and Y.
d) Compute the covariance between X and Y. Interpret it in words.
e) Compute the standard deviations of X and Y.
f) Compute the correlation coefficient between X and Y, and interpret it in words. The formula appears below.
g) Determine whether these two variables are statistically independent by using two methods.

$$\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}$$

B. Written Questions (52)

7. (8 points) Consider a random variable Z that follows the standard normal distribution. The probability density function for Z is expressed as:

$$\frac{1}{\sqrt{2\pi}} \; exp^{-0.5\, Z^2}$$

a) What is the interpretation of the following expression? Do NOT attempt to evaluate this integral, as that is impossible. $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{1.24} \exp^{-\frac{1}{2}z^2} \, dz$

b) Evaluate the following expression, but do NOT try to integrate the function that appears below.

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} exp^{-0.5\,Z^2} dZ$$

c) Evaluate the following expression, but do NOT try to integrate the function that appears below.

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} Z \times exp^{-0.5\,Z^2} dZ$$

d) Evaluate the following expression, but do NOT try to integrate the function that appears below.

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} Z^2 \times exp^{-0.5\,Z^2} dZ$$

8. (6 points) Consider the following random variable with the following distribution.

$$f\left(x\right) = 5 \; if \; 1 \le X \le 1.2 \; and \; 0 \; otherwise$$

a) Is this a cumulative density function or a probability density function? Why?
b) Which type of probability distribution is this? How do you know?
c) What is the interpretation of the figure 5?
d) Why is the domain of possible X values restricted?

9. (4 points) Briefly discuss the binomial distribution. Under what circumstances do we use it? This question appeared on the second midterm examination. I hope that you do better this time.

10. (10 points) Consider a random variable $\overline{X}$, which is the mean of a sample of N observations. This question appeared on the second midterm examination.

a) Explain why $\overline{X}$ is a random variable.
b) Define the term random sample.
c) Show that the expected value of $\overline{X}$ is μ_x, the population mean for X. Explain and verify your steps.
d) Show that the variance of $\overline{X}$ is σ^2/n, for which sigma is the standard deviation of the population for X. Explain and verify your steps.

e) Explain intuitively (and not mathematically) why it is logical that the variance of $\overline{X}$ is smaller than the variance of X.

11. (10 points) This is a synthesis-type question. In a paragraph, write an outline of the essence of the discipline of probability and statistics for your friend who is to take the course in the future. The object of this question is to LINK the following elements, for which I provide you with the sequence: population, uncertainty, random variable, probabilities, parameters, sample, estimator.

12. (4 points) Consider the estimator $\hat{\theta}$ for an unknown parameter θ. There are several desirable properties of an estimator, including efficiency and unbiasedness.

a) Define in words what is meant by the term unbiasedness.
b) Define in words what is meant by the term efficiency.

For both of these questions, mathematical symbols will not receive credit.

13. (4 points) Identify and give the significance of the central limit theorem. In other words, explain why it is so useful.

14. (6 points) Consider the following elements X_1, X_2, X_3, and X_4 that are drawn from a random sample of a population of mean μ_x and standard error σ. Consider the following two point estimators for μ_x :

$$\hat{\theta}_1 = 0.1(X_1) + 0.4(X_2) + 0.4(X_3) + 0.1(X_4)$$
$$\hat{\theta}_2 = 0.2(X_1) + 0.3(X_2) + 0.3(X_3) + 0.2(X_4)$$

a) Which one has a larger variance, and why?
b) Which one is unbiased, and why ?
c) Neither one corresponds to $\overline{X}$ - why not?