ECO383
Economics of Education
*Sample Test**

December 201N

Please fill in your full name and student number in the spaces below.

NAME:

STUDENT NUMBER:

You have 1 hour and 45 minutes to complete this **closed-book** test. When the instructor asks everyone to stop, you must stop writing. This should be entirely your own work – **no conferring please**.

The test is worth 106 points. Please attempt all the questions, and be sure to read each question carefully. The number of points each question is worth is indicated in brackets.

You should answer the test directly on this booklet, in the spaces provided. Please write legibly, and answer in proper sentences where sentences are required. If you need additional space, indicate clearly which question you are answering, and write on the reverse side of the page. [Approximate guidelines as to the length of the 'perfect' answer are given in square brackets, where necessary.]

*Instructor: Robert McMillan, email: mcmillan@chass.utoronto.ca

Please write clearly in the spaces below (or overleaf if necessary) and answer in proper sentences, rather than in some shorthand code that only you may understand.

Question 1: Summations (worth 10 points)

a) Suppose you have two random variables, X and Y, with N=2, and so two observations on each, written as $\{x_1, x_2, y_1, y_2\}$. Can

$$\sum_{i=1}^{N}(x_i + y_i) = \sum_{i=1}^{N} x_i y_i$$

in this case? If so, under what conditions? (6 points)

b) Generalizing from part a), suppose you have two random variables, X and Y, with N now being any positive number, the observations on each being written as $\{x_i, y_i\}_{i=1}^{N}$. Can

$$\sum_{i=1}^{N}(x_i + y_i) = \sum_{i=1}^{N} x_i y_i$$

in this case? If so, under what general conditions? (4 points)

Question 2: Statistical Properties of Least Squares Estimator (worth 21 points)

Many researchers have been interested in assessing whether schools that have smaller classes produce better academic outcomes.

To that end, you are given data on a sample of N students drawn from a variety of schools. For each student i, you observe the student's class size (C_i) and a test score (T_i). Imagine you plotted the sample data points on a graph, with class size, C, on the horizontal axis and test scores, T, on the vertical axis.

Suppose the underlying (or 'true') model is given by

$$T_i = \alpha + \beta C_i + u_i, \tag{1}$$

where u_i is a random error term, and α and β are unknown parameters to be estimated.

a) Show that the ordinary least squares slope coefficient, b^{ols}, is a random variable. [Hint: start from the formula for b^{ols}, then make a relevant substitution, then simplify the resulting expression.] (4 points)

b) Derive the expected value of the least squares intercept, a^{ols}. (3 points)

c) **State** the formula for the variance of b^{ols}. (2 points)

d) **Derive** the variance of b^{ols} carefully. In doing so, please make clear what assumption or assumptions you require in your derivation *and at what precise points*, and show all the relevant steps. [Hint: start from the expression you derived as your answer to part a).] (12 points)

Question 3: Continuous Random Variables (worth 17 points)

We write the probability density function of a continuous random variable X as $f(x)$, for $a < x < b$.

a) Write down a general expression for the expected value of a continuous random variable, $E(X)$. (3 points)

b) Write down a general expression for the variance of a continuous random variable, $Var(X)$. (3 points)

For the remainder of this question, consider the specific case of a uniformly distributed random variable, $X \sim U(-2, 3)$, where $-2 \leq x \leq 3$. [Hint: recall that the pdf of such a variable has a rectangular shape.]

c) What is the value of $f(x)$, for $-2 \leq x \leq 3$? (3 points)

d) In this instance, what is the expected value of the uniformly distributed random variable X? Please show your calculations. (3 points)

e) What is the variance of the same uniformly distributed random variable X? Please show your calculations. (5 points)

Question 4: Continuous Random Variables – the Normal Distribution (worth 10 points)

Consider a random error term X that follows a normal distribution, with mean μ and variance σ^2.

a) What is the precise expression for the probability density function (in this case, $f(x)$) of this normal random variable? (4 points)

Now consider a random variable $Z = (X - \mu)/\sigma$. This is a standardized version of X.

b) *Derive* the variance of Z. Please show your calculations. (6 points)

Question 5: 'Noise' Application (worth 12 points)

Suppose you have test score data on high school students over time. Let student performance on a test be given by the following very simple model:

$$S_{it} = A_i + u_{it}, \tag{2}$$

where S_{it} is a student i's score at time t, A_i is student i's ability, assumed to be unchanging *and* unobserved, and u_{it} is a random error term that captures factors assumed to influence test scores in an unsystematic way. Assume that this error term has mean zero – that is $E(u_{it}) = 0$, for all (i, t) – and a constant variance, σ^2. And also assume that the error terms across tests are independent, and so have a zero covariance.

a) Suppose N tests are set, to be taken separately at times $t = \{1, ..., N\}$. For student i , what is the expected value of the score she obtains on a given test at time t? In other words, what is $E(S_{it})$? Please set out your reasoning. (4 points)

b) What is the expected value of the average score for student i, where the average score is given by

$$AS_i \equiv \frac{1}{N}\sum_{t=1}^{N} S_{it}?$$

Here, assume $N = 3$. (3 points)

c) In the case where $N = 3$, what is $Var(AS_i)$, where AS_i is as defined in part b)? Please show your calculations. (5 points)

Question 6: Hypothesis Testing (worth 11 points)

As in Question 1, imagine you are given data on a sample of N students drawn from a variety of schools. For each student i, you observe the student's class size (C_i) and a test score (T_i). Consider the underlying (or 'true') model describing the relation between class size and student performance, given by

$$T_i = \alpha + \beta C_i + u_i, \tag{3}$$

where u_i is a random error term, and α and β are unknown parameters to be estimated. Suppose the error term is distributed as a normal random variable, with mean zero and variance equal to σ^2, written $u_i\ N(0, \sigma^2)$. Assume that $\sigma^2)$ is known.

a) What is the implied distribution of b^{ols} in this case? Please be precise. (3 points)

b) Define the standardized version of b^{ols} as z, formed by deducting off the relevant mean and dividing through by the standard deviation. What is the implied distribution of the resulting standardized variable, z? (2 points)

c) What is the probability that z is greater than 1.645? [Hint: think what sort of null hypothesis is implied here.] [One sentence.] (2 points)

d) Now suppose that σ^2 is unknown. What implications would that have for your answer to part a)? Please explain. (4 points)

Question 7: Lazear etc. (worth 7 points)

a) Please describe Lazear's rationalization for Hanushek's broad finding in his well-known (1986) paper. [Hint: omitted variables bias.] [Three sentences.] (5 points)

b) How might randomization help when studying the effects of class size on student performance? Please be specific, both about how you would implement the randomization and why it would help. (5 points)

Question 8: Discrete Random Variables (worth 15 points)

Consider a discrete random variable, X, that takes on J discrete values $\{x_1, x_2, ..., x_J\}$, with the probability that X takes on the specific value x_j being written $P(X = x_j) \equiv p_j$.

a) Write down a general expression for the expected value of the random variable, $E(X)$, in the general case described above. (2 points)

b) Write down a general expression for the *variance* of the random variable, $Var(X)$, in the general case described above. (3 points)

For the remainder of this question, consider a discrete random variable that takes on just two possible values (known as a Bernoulli random variable). Let its probability density function be given by $X = x_1 = -a$ with probability b and $X = x_2 = a$ with probability $(1 - b)$.

c) Write down an expression for the mean (expected value) of this Bernoulli random variable if $b = 1/3$. (3 points)

d) Write down an expression for the variance of this Bernoulli random variable if $b = 1/3$. (3 points)

e) Building on your answer to part d), if a became a *larger* positive number, what would happen to the variance of the random variable? Please explain clearly, with reference to a relevant formula or expression. (4 points)