

ECO383
Economics of Education
*In-Class Test – Sample Solutions**

October 2011

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 5 points)

Let the sample mean of a variable X be defined by

$$\bar{X} \equiv \frac{1}{N} \sum_{i=1}^N X_i,$$

given a sample of N realizations of the variable.

a) Can you re-write the following:

$$\sum_{i=1}^N \alpha X_i,$$

where α is a fixed parameter, as some function of the sample mean of X ? Please set out your reasoning clearly. (2 points)

First, we can factor out the α from the summation, so that

$$\sum_{i=1}^N \alpha X_i = \alpha \sum_{i=1}^N X_i.$$

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

[You should make sure you can see why.] Further, given that

$$\bar{X} \equiv \frac{1}{N} \sum_{i=1}^N X_i,$$

we can re-arrange this as

$$\sum_{i=1}^N X_i = N\bar{X}.$$

Substituting, we get the answer:

$$\sum_{i=1}^N \alpha X_i = \alpha N\bar{X}.$$

b) Can you re-write the following:

$$\sum_{i=1}^N \alpha(X_i + \gamma Y_i),$$

where α and γ are fixed parameters, as some function of the sample means of X and of Y ? (Obviously, we have that

$$\bar{Y} \equiv \frac{1}{N} \sum_{i=1}^N Y_i.$$

Please set out your reasoning carefully. (3 points)

First, we can expand the expression, according to which

$$\sum_{i=1}^N \alpha(X_i + \gamma Y_i) = \sum_{i=1}^N (\alpha X_i + \alpha\gamma Y_i).$$

Next, we can break the summation into two (as it were), giving

$$\sum_{i=1}^N (\alpha X_i + \alpha\gamma Y_i) = \sum_{i=1}^N \alpha X_i + \sum_{i=1}^N \alpha\gamma Y_i$$

– I hope everyone is happy with that.

Then we can apply the same type of reasoning as in the solution to part a), factoring out the constants, and replacing the summation signs. That is,

$$\sum_{i=1}^N \alpha X_i + \sum_{i=1}^N \alpha\gamma Y_i = \alpha \sum_{i=1}^N X_i + \alpha\gamma \sum_{i=1}^N Y_i = \alpha N\bar{X} + \alpha\gamma N\bar{Y}.$$

So, in short,

$$\sum_{i=1}^N \alpha(X_i + \gamma Y_i) = \alpha N\bar{X} + \alpha\gamma N\bar{Y}.$$

(Actually, we can further factor the right hand side of the last line as $\alpha N(\bar{X} + \gamma\bar{Y})$.)

Question 2: ‘Noise’ (worth 14 points)

Suppose you have test score data on a sample of high school students. Let a student’s performance on a test be given by the following very simple (purely random!) model:

$$S_i = \alpha u_i, \tag{1}$$

where S_i is a student i ’s score, u_i is a random error term that captures factors assumed to influence test scores in an unsystematic way, and α is a positive constant.

Suppose that the random error term follows a discrete distribution, taking on three possible values (irrespective of the student), given by

$u_i = 1$ with probability = 0.25; $u_i = 2$ with probability = 0.5; and $u_i = 3$ with probability = 0.25.

a) What is the mean (or ‘expected value’) of the random error term u_i , given the above discrete distribution? Please show your calculations. [Hint: start by writing down the expression for the expected value of a discrete random variable.] (2 points)

Generally speaking, the expected value of a discrete random variable X that takes on J possible values (x_1, x_2, \dots, x_J) is given by

$$E(X) = \sum_{j=1}^J x_j P(X = x_j) = \sum_{j=1}^J x_j p_j.$$

In this context, let us sharpen the notation to fit the case in hand. Thus we can write

$u_{i1} = 1$ with probability $(p_1) = 0.25$; $u_{i2} = 2$ with probability $(p_2) = 0.5$; and $u_{i3} = 3$ with probability $(p_3) = 0.25$.

Then we have

$$E(u_i) = \sum_{j=1}^3 u_{ij} p_j = 1 \times 0.25 + 2 \times 0.5 + 3 \times 0.25 = 2.$$

(Note: this is unsurprising. The distribution is symmetric around 2, the mid-point.)

b) What is the mean (or ‘expected value’) of student i ’s score? Again, please show your calculations. (3 points)

$$E(S_i) = \sum_{j=1}^3 \alpha u_{ij} p_j = \alpha(1 \times 0.25 + 2 \times 0.5 + 3 \times 0.25) = 2\alpha.$$

[Aside: here, the key is that one can factor the α out. This follows simply from the linear nature of the expectations operator.]

c) Write down an expression for the variance of student i ’s score. [Hint: again, start from the relevant formula, this time for the *variance* of a discrete random variable.] (3 points)

Let us start with the general expression, tailored to this context:

$$\text{Var}(S_i) = E((S_i - E(S_i))^2) = \sum_{j=1}^J (\alpha u_{ij} - 2\alpha)^2 p_j$$

Thus

$$\text{Var}(S_i) = (\alpha \times 1 - 2\alpha)^2 \times 0.25 + (\alpha \times 2 - 2\alpha)^2 \times 0.5 + (\alpha \times 3 - 2\alpha)^2 \times 0.25.$$

But we can re-write the latter as

$$\alpha^2(1 - 2)^2 \times 0.25 + \alpha^2(2 - 2)^2 \times 0.5 + \alpha^2(3 - 2)^2 \times 0.25 = \alpha^2(1 \times 0.25 + 1 \times 0.25) = \alpha^2/2.$$

(Note: $\text{Var}(S_i) = E((S_i - E(S_i))^2) = E((\alpha u_i - E(\alpha u_i))^2)$, the expected value of the squared deviations from the mean. In turn, we can factor out the constant, which is now squared, to give $\alpha^2 E((u_i - E(u_i))^2)$, but this is simply $\alpha^2 \text{Var}(u_i)$.)

d) Given your answer to part c), what would happen to the variance as α became more positive? (2 points)

It is straightforward to see from the previous expression that the variance would increase (in the square of α).

e) If the score were instead given by

$$S_i = 10 + \alpha u_i, \tag{2}$$

what would the expected value of the score be? Please explain, with reference to the relevant formula. [Hint: what is the expected value of a constant?] (4 points)

Here, we just need to take the expected value of S_i :

$$E(S_i) = E(10 + \alpha u_i) = E(10) + E(\alpha u_i) = 10 + \alpha E(u_i) = 10 + 2\alpha.$$

Thus, the expected score just shifts up by a constant of 10.

Question 3: the Mechanics of Curve Fitting (worth 16 points)

Suppose we are interested (as many researchers have been) in assessing whether schools that have smaller classes produce better academic outcomes.

To that end, you are given 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

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

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

Suppose that researchers pick the slope parameters to minimize the least squares objective, given by

$$S = \sum_{i=1}^N (e_i(a, b))^2,$$

where e_i is the residual associated with observation i , given the line parameterized by (a, b) , and there are N observations in total.

a) What is the least squares formula for the intercept, a^{ols} ? (2 points)

[See class notes.]

b) What is the least squares formula for the slope coefficient, b^{ols} ? Write down the relevant expression accurately. (3 points)

[See notes.]

c) **Derive** the least squares intercept *and* slope estimators using calculus, showing the relevant steps. (11 points)

[See notes. You will need this for the test, for sure.]