Ec. 2203 – Quantitative Methods II

Problem Set 0. Review of Variance, Covariance and Correlation

Learning Outcomes
This problem set familiarises you with the concepts of variance, covariance and correlation. In some ways, these exercises are untypical of the rest of the course, but these statistics and rules are used extensively throughout the rest of the course so it is essential that you learn them now.

1. Show that the sample covariance, (a measure of the association), between the variables X and Y, $\frac{1}{N}\sum_i (X_i - \bar{X})(Y_i - \bar{Y})$

   can also be written as $\frac{1}{N}\sum_i (X_i Y_i) - \bar{X}\bar{Y}$

Hence, given the following information, find Cov(X,Y) ie the association between years of education and job tenure. Are they positively or negatively related?

Individual	W = Age	X = Years of Education	Y = Job Tenure
1	18	11	1
2	29	14	6
3	33	12	8
4	35	16	10
5	45	12	5

2. Show that the sample variance of X, Var(X) = $\frac{1}{N}\sum_i (X_i - \bar{X})^2$

   Can also be written as $\frac{1}{N}\sum_i (X_i)^2 - \bar{X}^2$

   Hence calculate Var(X) ie a measure of the extent of the dispersion in the years of education in the sample, using the data in question 1

What happens to your estimate of the variance if you first multiply the data in the X column by 10?
What happens to your estimate of the covariance?

Now see if you can replicate your answers using the commands in Stata

Turn over

(Hints: type in the numbers into Stata's internal spreadsheet then:

a) if you type

summarise *variable name*, detail

then Stata will give you an estimate of the sample variance. However Stata uses the formula $\frac{1}{N-1}\sum_i (X_i - \bar{X})^2$ not $\frac{1}{N}\sum_i (X_i - \bar{X})^2$ to calculate the sample variance, so you will have to adjust this estimate in some way. How?)

b) if you type

corr *variable1name variable2name, cov*

Stata will give you an estimate of the covariance. Again Stata uses the formula

$$\frac{1}{N-1}\sum_i (X_i - \bar{X})(Y_i - \bar{Y})$$

so you will have to adjust this estimate

3. If Y= A+B, show that

Cov(X,Y) = Cov(X,A) + Cov(X,B)

Hence show that

i) Var(Y)=Var(A) +Var(B) + 2Cov(A,B)
ii) Cov(X,Y) = aCov(X,Z) if y = aZ, where a is a constant
iii) Cov(X,Y) = 0 if y = a

4. Find the sample correlation coefficient between X and Y using the original data in the Table, ie a scale invariant measure of the association between years of education and job tenure.

5. Suppose yearly after-tax Income, Y, is related to yearly income before tax, X, by the equation

Y = 4000 + 0.7X

and that X is a random variable, (so that wages vary over the years) with an expected (ie mean) value , μ_X , and variance σ^2_X

i) Interpret the equation
ii) Find the expected value of after tax income
iii) The variance and standard deviation of after-tax income