EE325 : INTRODUCTORY ECONOMETRICS

SEMESTER 2/2011

SOLUTION FOR HOMEWORK ASSIGNMENT 1&2

1. $E(X) = \sum p_i X_i$

$E(S) = .05(1) + .09(2) + .03(3) + .42(4) + .41(5) = 4.05$

$E(M) = .04(1) + .10(2) + .12(3) + .46(4) + .28(5) = 3.84$

b) $\sigma^2 = \sum p_i (X_i - \mu)^2$

$\sigma_S^2 = .05(1-4.05)^2 + .09(2-4.05)^2 + .03(3-4.05)^2 + .42(4-4.05)^2 + .41(5-4.05)^2 = 1.2475$

$\sigma_M^2 = .04(1-3.84)^2 + .10(2-3.84)^2 + .12(3-3.84)^2 + .46(4-3.84)^2 + .28(5-3.84)^2 = 1.1344$

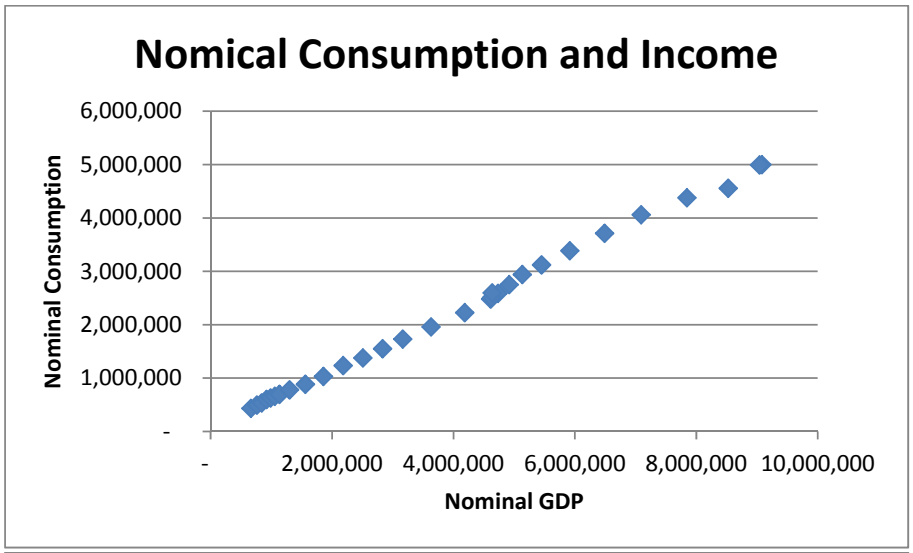
$\therefore \sigma_S = \sqrt{1.2475} = 1.1170$ and $\sigma_M = \sqrt{1.1344} = 1.0651$

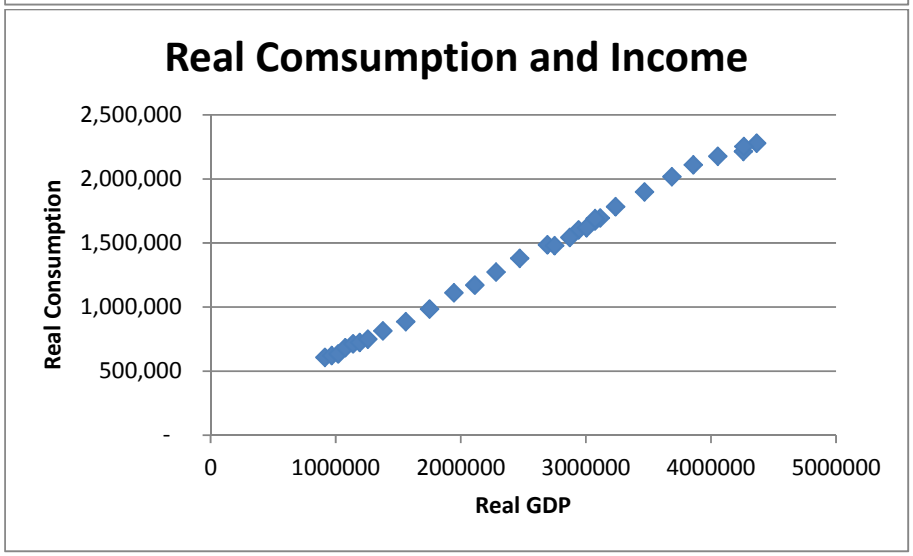
c) The customers satisfy with the senior executives more than middle managers because $E(S) > E(M)$. However, the s.d. of senior executives is more than middle managers. So, there is a higher degree of variation in senior executives satisfaction compared with middle managers.

2. a) NCONS, RCONS, NGDP, and RGDP. All of them are continuous.

b) Nominal consumption and real consumption should be dependent variables. Nominal GDP and real GDP should be independent variables. The reason is that the consumption should be proportional to income. (Some may argue that the consumption also causes income by citing the identity $Y \equiv C + I + G + X - M$. There is no right or wrong here since it is actually two ways causation. However, it is conventional to assume that consumption depends on income.)

c)





They have positive relationship which is consistent with the expectation. When people have more income, we would expect them to consume more.

d) Nominal Consumption and GDP

$E(NGDP) = 3,922,722$

$E(NCONS) = 2,196,165$

$Var(NGDP) = 7.046x10^{12}$

$Var(NCONS) = 2.109x10^{12}$

$Covar = 3.722x10^{12}$

$Corr = 0.9989$

Real Consumption and GDP

$E(NGDP) = 2,525,924$

$\text{E(NCONS)} = 1,395,373$

$Var(NGDP) = 1.240x10^{12}$

$Var(NCONS) = 3.115x10^{11}$

$\text{Covar} = 6.002x10^{11}$

$Corr = 0.999$

3. a) We usually observe that a country with the high crime rate also has high number of police officers. Equivalently, we usually observe that a country with high number of police officers also has the high crime rate.

   b) It does not mean that if we increase or reduce one variable, the other variable will be affected. For example, if we reduce the number of police, it does not mean that the crime rate will be reduced. It is also possible that there is a 3rd variable that affect both variables.

4. a) The discrete probability density function is the probability function that is associated with a discrete random variable; on the other hand, the continuous probability density function is associated with continuous variable. Thus, a discrete probability density function is a discrete function, and a continuous probability density function is a continuous function.

   b) Outcomes are 2, 3, 4, … , 12. The PDF is

Outcome	Probability
2	1/36
3	2/36
4	3/36
5	4/36
6	5/36
7	6/36

Outcome	Probability
8	5/36
9	4/36
10	3/36
11	2/36
12	1/36

   c) It is a discrete PDF because it assumes that the outcomes are only discrete values.

5. a) $P(X = 0) = 0.1 + 0.2 + 0.18 = 0.48$

   $P(X = 1) = 0.04 + 0.09 + 0.08 = 0.21$

   $P(X = 2) = 0.06 + 0.11 + 0.14 = 0.31$

   b) X and Y are not independent. It can be proved by showing that $P(X = 0 \text{ and } Y = 0) \neq P(X = 0) \cdot P(Y = 0) \rightarrow 0.1 \neq (0.2)(0.48)$.

   c) $E(X + Y) = E(X) + E(Y) = [.48(0) + .21(1) + .31(2)] + [0.2(0) + 0.4(1) + 0.4(2)] = 2.03$

   (For the summation of expected values, you do not have to take into account the covariance, but you have to take into account when you find var(X+Y).)

   Each day, the machine is expected to be broken by 2.03 times.

d) $P(X>1|Y<1)=\dfrac{P(X=2,Y=0)}{P(Y=0)}=\dfrac{0.06}{0.2}=0.3$

6. a) PRF, in this case, is a linear function. It tells us how does Y indeed response to X. For the SRF, it is the estimated version of PRF. We will never know the actual parameters in the PRF, so we use the estimators estimated in SRF as a proxy for parameters in PRF.

   b) No because β_1 and β_2 are the population concept. As a result, we cannot find them from the sample.

   c) If we have a sample, we can use the sample to find $\hat{\beta}_1$ and $\hat{\beta}_2$ because they are the sample concept.

   d) No. It is important to remember that the PRF is something fixed, but unknown, in the population. Because the SRF is obtained from a given sample of data, a new sample will generate different slope and intercept.

   e) Unbias ($E(\hat{\theta})=\theta$) and efficient (lowest variance)

   f) u_i is usually called an error term or disturbance term. $\hat{u}_i$ is called a residual. They are different because the error term is a population concept, but the residual is for sample concept.

7. a) Intrinsically linear model: $\ln(Y_i)=\beta_1+\beta_2 X_i+u_i$

   b) Intrinsically linear model:

$$1-Y_i=\frac{e^{\beta_1+\beta_2 X_i+u_i}}{1+e^{\beta_1+\beta_2 X_i+u_i}}$$

$$\frac{Y_i}{1-Y_i}=\frac{\dfrac{1}{1+e^{\beta_1+\beta_2 X_i+u_i}}}{\dfrac{e^{\beta_1+\beta_2 X_i+u_i}}{1+e^{\beta_1+\beta_2 X_i+u_i}}}=\frac{1}{e^{\beta_1+\beta_2 X_i+u_i}}=e^{-(\beta_1+\beta_2 X_i+u_i)}$$

$$\ln\left(\frac{Y_i}{1-Y_i}\right)=-(\beta_1+\beta_2 X_i+u_i)$$

$$\ln\left(\frac{Y_i}{1-Y_i}\right)=-\beta_1-\beta_2 X_i-u_i$$

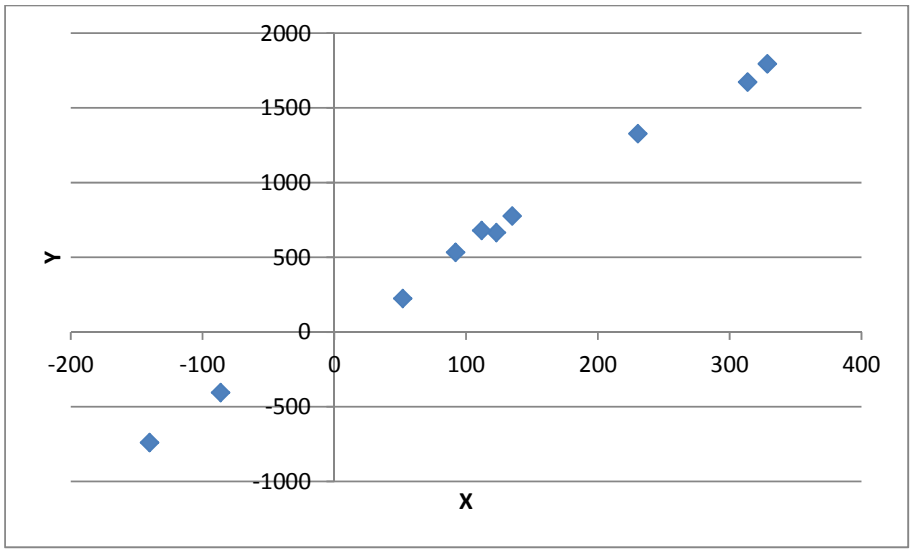
$$\ln\left(\frac{Y_i}{1-Y_i}\right)=\alpha_1+\alpha_2 X_i+e_i;\ where\ \alpha_1=-\beta_1,\alpha_2=-\beta_2,e_i=-u_i$$

   c) Linear model

   d) Nonlinear model

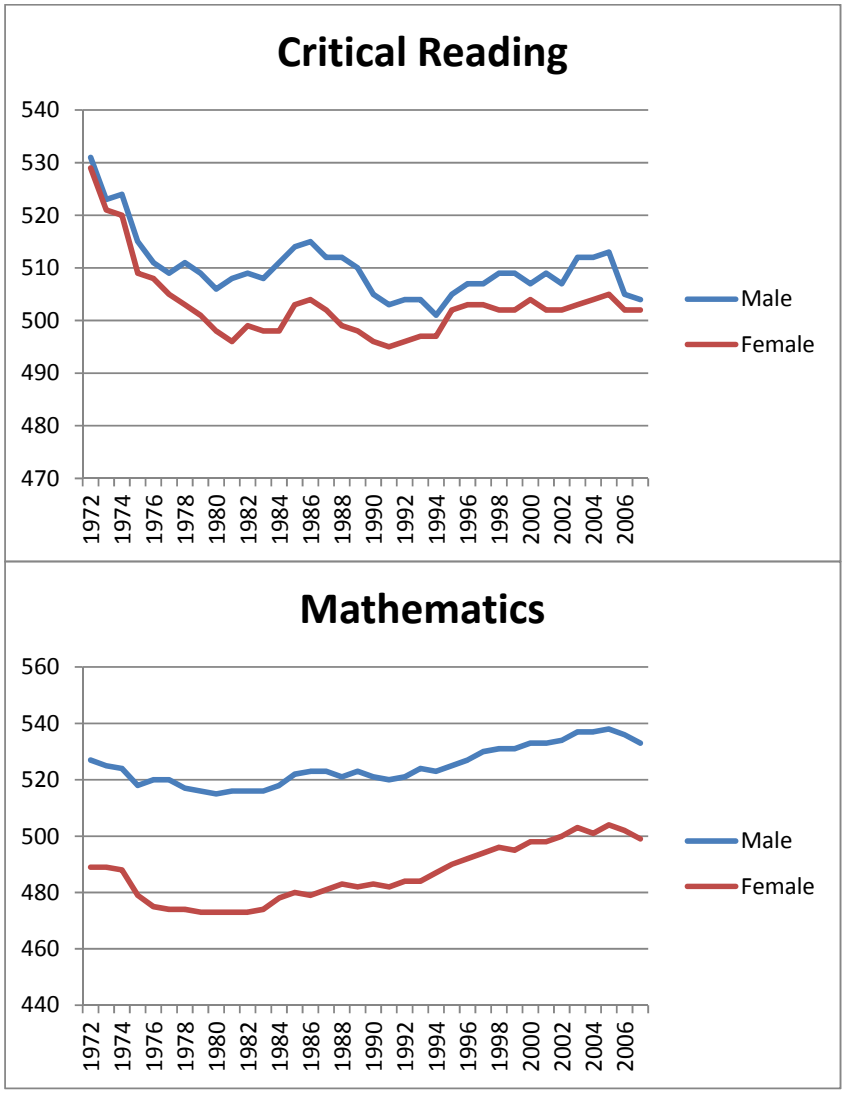
e) Nonlinear model

8. a)



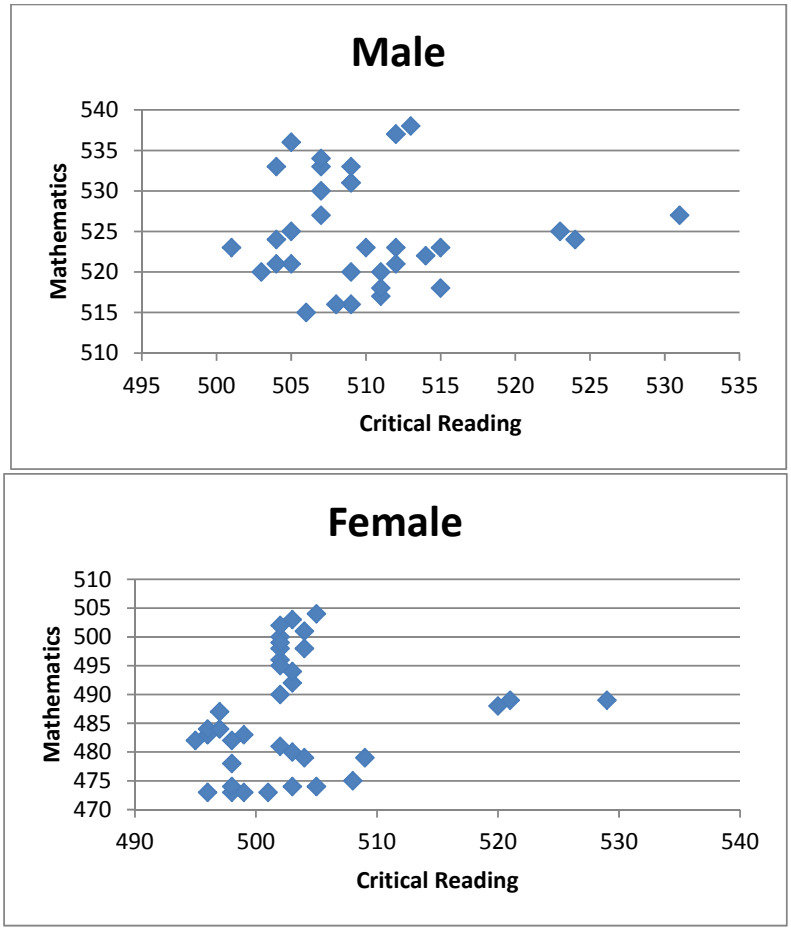
b) $E(Y) = \frac{\sum Y_i}{N} = 653.26$, $E(Y \mid X = 100) = 28.9015 + 5.3836(100) = 567.26$

9. a)


Critical Reading
540
530
520
510
500
490
480
470
1972
1974
1976
1978
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
Male
Female
Mathematics
560
540
520
500
480
460
440
1972
1974
1976
1978
1980
1982
1984
1986
1988
1990
1992
1994
1996
1998
2000
2002
2004
2006
Male
Female

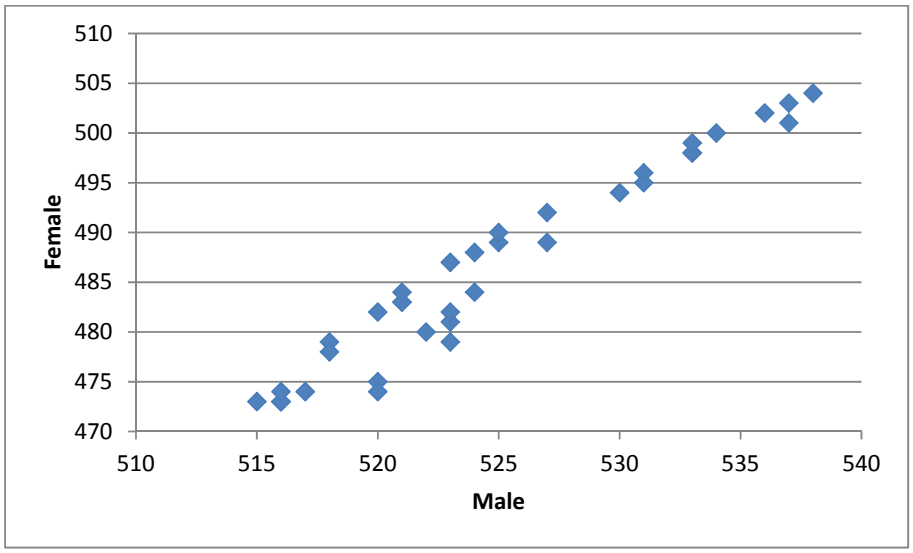
b) The performance of male and female tend to move together (This is a good example of what I said in 3.b that there may be the 3rd factor that affect 2variables. In this case, male causes female or female courses male are illogical. However, it may be more logical to assume that the 3rd factor, e.g. the difficulty of the exam, is responsible for the movement of these variables.).

c)



The relationships for both groups are not clear. However, it seems like there is only slightly positive relationship.

d) There is a highly positive linear relationship.


510
505
500
495
490
485
480
475
470
Female
510
515
520
525
530
535
540
Male