Ec 5040: Exercise 0 Answers - Revising statistics and the 2 Variable Linear Model

1. Give definitions of the following properties of estimators
   Bias; Efficiency, Mean Square Error; Consistency

Given any estimation process we would like the estimates produced by this process to be as close to the true value as possible

Bias: $E(\hat{\theta}) \neq \theta$ Unbiased: $E(\hat{\theta}) = \theta$

If samples of size N are drawn repeatedly and the estimator $\hat{\theta}$ computed then for an unbiased estimator the average (mean) of these values will equal θ

We would also like these estimators to have the minimum variance property.
An estimator is said to be efficient if it has the smallest variance.

Sometimes estimators may be unbiased and not have minimum variance or vice versa

The mean squared error is a measure of both the bias and variance of an estimator

$$M.S.E. = E(\hat{\theta}-\theta)^2$$

Let $\mu = \hat{\theta}$

then can always re-write $M.S.E. = E(\hat{\theta}-\theta)^2 = E\left[(\hat{\theta}-\mu)+(\mu-\theta)\right]^2$

$$. = E\left[(\hat{\theta}-\mu)^2 + 2(\mu-\theta)(\hat{\theta}-\mu)+(\mu-\theta)^2\right]^2$$

which since $E(\hat{\theta}-\mu) = E(\hat{\theta}\_-E(\mu) = 0$ implies middle term vanishes so

$$M.S.E. = Var(\hat{\theta}) + E\left[(\mu-\theta)\right]^2$$

$$M.S.E. = Var(\hat{\theta}) + Bias^2$$

Consistency:

An estimator $\hat{\theta}$ is consistent estimator of the true value θ if its bais and variance both go to zero as the sample size, N, gets larger (tends to infinity)

$$\Pr\left[\left|\hat{\theta}-\theta\right| < \delta\right] \to 1 \qquad \text{as } N \to \infty$$

where δ is any positive number

Alternatively can write that an estimator is consistent if

$$\hat{\theta} \xrightarrow{p} \theta \qquad \text{or} \quad p\lim(\hat{\theta}) = \theta$$

Often estimators may be biased in small samples but have desirable properties in larger samples

Consistency is often easier to establish than efficiency or bias

2. Given the joint density function of 2 continuous random variables, x and y, is
   f(x,y) = 4-x-y 0<= x <=1 , 0<= y <=1
   Show that E(x) =1.42

We know $E(X) = \int_x xf(x)dx$

the expected value of is the sum of the set of values of X and the probability of that value occurring

First find the marginal distribution of x using $f(X) = \int_y f(x,y)dy$

$$f(x) = \int_0^1 f(x,y)dy = \int_0^1 (4-x-y)dy$$

$$= \left[(4y - xy - y^2)\right]_0^1$$

$$= (4 - x - 1/2)$$

f(x) = 7/2 – x

Hence $E(X) = \int_0^1 xf(x)dx = \int_0^1 x(7/2 - x)dx = \int_0^1 (7/2x - x^2)dx$

$$= \left[(\frac{7x^2}{2} - \frac{x^3}{3} \right]_0^1 = \frac{7}{4} - \frac{1}{3}$$

ie E(X) = 17/12 = 1.42

[

General rules on expectations

Mean and variance of a function of X

E(g(X)] = $\sum_x g(X)\Pr(X = x) = \int_x g(x) f(x)dx$

Var[g(X)] = $\sum_x (g(X) - E(g(x))^2 \Pr(X = x) = \int_x (g(X) - E(g(x))^2 f(x)dx$

Hence

$$E(a + bX) = \int_x (a + bX) f(X)dx = \sum_x (a + bX) f(X)$$

$$E(a + bX) = \sum_x af(x) + \sum_x bXf(x)$$

$$E(a + bX) = a + bE(X)$$

and

$$Var(a + bX) = \int_x \left[(a + bX) - E(a + bX)\right]^2 f(X)dx = \sum_x \left[(a + bX) - E(a + bX)\right]^2 f(X)$$

$$Var(a + bX) = \sum_x \left[(a + bX) - (a + bE(X))\right]^2 f(X)$$

$$Var(a + bX) = \sum_x \left[b(X - E(X))\right]^2 f(X)$$

$$Var(a + bX) = bVar(X)^2$$

so for any constant a then Var(a) = 0

Let Z = aX + bY

Var(Z) = Cov(Z,Z) = $\int_z \left[(Z - E(Z)\right]^2 f(Z)dz$

= Cov(Z, aX+bY)
=Cov(aZX) + Cov(bZY)
=Cov(aX+bY, aX) + Cov(bY, aX) + Cov(aX,bY) = Cov(bY, bY)
=a^2Var(X) + b^2Var(Y) + 2Cov(aX, bY)

Hence Cov(aX + bY)

=a^2Var(X) + b^2Var(Y) + 2abCov(X, Y)

3. If X and Y are continuous random variables with joint density f(XY) and a and b are constants, show that

E(aX + bY) = aE(X) + bE(Y)

$$E(aX+bY)=\int_x\int_y (aX+bY)f(X,Y)dxdy$$

$$=\int_x\int_y aXf(X,Y)dxdy+\int_x\int_y bYf(X,Y)dxdy \qquad (1)$$

since the order of integration does not matter

$$=a\int_x X\int_y f(X,Y)dxdy+b\int_x Y\int_y f(X,Y)dxdy \qquad (2)$$

[

Eg $\int_x\int_y(3X*XY^2)dxdy=\int_x\frac{X^2Y^3}{3}dx=\frac{X^3Y^3}{3}=3\left[\int_x X*\frac{XY^3}{3}dx\right]$

]

But $=\int_y f(X,Y)dy=f_x(x)$ (the marginal distribution of x, summing the joint pdf over all y values at a given x value)

and

$$\int_x f(xy)dx=f_y(y)$$

so (2) becomes $E(aX+bY)=a\int_x Xf_x(x)dx+b\int_y Yf_y(y)dy$

so

E(aX + bY) = aE(X) + bE(Y)

3. If X and Y are independently distributed random variables, show that
   a) E(X/Y) = E(X)
   b) Corr(X,Y) = 0

   If f(X,Y) = 2 and 0 < X < Y < 1
   find the marginal probability density functions and the conditional pdf of X given Y.

If X and Y are independent then Pr(X/Y) = Pr(X) and Pr(Y/X) = Pr(Y)
f(X/Y) = f(X) and f(Y/X) = f(Y)
(the conditional pdf equals the marginal pdf)
f(X/Y) = f(X,Y)/f(Y) and f(Y/X) = f(X,Y)/f(X)

f(X,Y) = f(X)*f(Y)

Hence

$$E(X/Y) = \int_x Xf(X/Y)dx = \int_x Xf(X)dx = E(X)$$

$$Cov(X/Y) = \int_y\int_x (X - E(X))(Y - E(Y))f(X,Y)dxdy$$
$$Cov(X/Y) = \int_y\int_x (X - E(X))(Y - E(Y))f(X)f(Y)dxdy$$
$$Cov(X/Y) = \int_x (X - E(X))f(X)dx\int_y (Y - E(Y))f(Y)dy$$

since E(X) = a constant value and E(Y) = a constant value

$$\int_x (X - E(X))f(X)dx = E(X) - E(X) = 0$$

$$\int_y (Y - E(Y))f(Y)dy = E(Y) - E(Y) = 0$$

Hence Cov(X,Y) = 0 and so

$$corr(X,Y) = \frac{Cov(X,Y)}{\sqrt{Var(X)*Var(Y)}} = 0$$

If f(X,Y) = 2 and 0 < X < Y < 1

Using $f(X) = \int_y f(x,y)dy = \int_x^1 2dy = [2y + c]_x^1$

So f(x) = 2(1-x) o<x<1

$$f(y) = \int_x f(x,y)dx = \int_0^y 2dx = [2x + c]_0^y$$

So f(y) = 2y o < y < 1

and the conditional pdf $f(x/y) = \frac{f(x,y)}{f(y)} = \frac{2}{2y} = \frac{1}{y}$ 0 < x < y

5. What is the connection between the chi squared, t and F distributions?

- they can all be derived from the normal distribution

The form of the normal distribution is given by

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{(x-\mu)^2}{2\sigma^2}\right]$$

and the random variable x is said to be normally distributed with mean μ and variance σ^2

$$X \sim N(\mu, \sigma^2)$$

Since the normal distribution is preserved under a linear transformation of a random variable

Then $(a + bX) \sim N(a+b\mu, b^2\sigma^2)$

If $a = -\mu/\sigma$ and $b = 1/\sigma$ then

The transformation $Z = \frac{x-\mu}{\sigma} \sim N(0,1)$

is said to give a standard normal variable

If $Z \sim N(0,1)$ then $X = Z^2$ is said to follow a Chi_squared distribution with 1 degree of freedom

$$X \sim \chi^2(1)$$

The sum of N independent standard normal variables has a χ^2 distribution with N degrees of freedom

$$(Z^2_1 + Z^2_2 + \ldots Z^2_N+) \sim \chi^2(N)$$

If $Z \sim N(0,1)$ and $Y \sim \chi^2(N)$

Then the transformation $t = \frac{Z\sqrt{N}}{\sqrt{Y}} \sim t(N)$

is said to follow a (Student's) t distribution with N degrees of freedom

If $Y_1 \sim \chi^2(N_1)$ and $Y_2 \sim \chi^2(N_2)$

then the transformation $F = \frac{Y_1 / N_1}{Y_2 / N_2} \sim F[N_1, N_2]$

is said to be an F distribution with N_1 and N_2 degrees of freedom

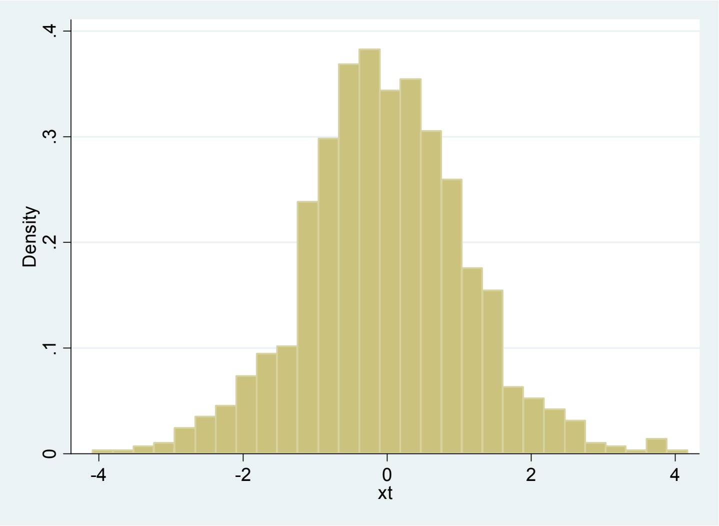
(note that the square of the t distribution $t^2 = \frac{Z^2/1}{Y/N} \sim F[1,N]$ follows an F distribution)

You can get a sense of what these distributions look like by using the "rnd" commands in stata

Eg to see what a t distribution with 10 degrees of freedom looks like type

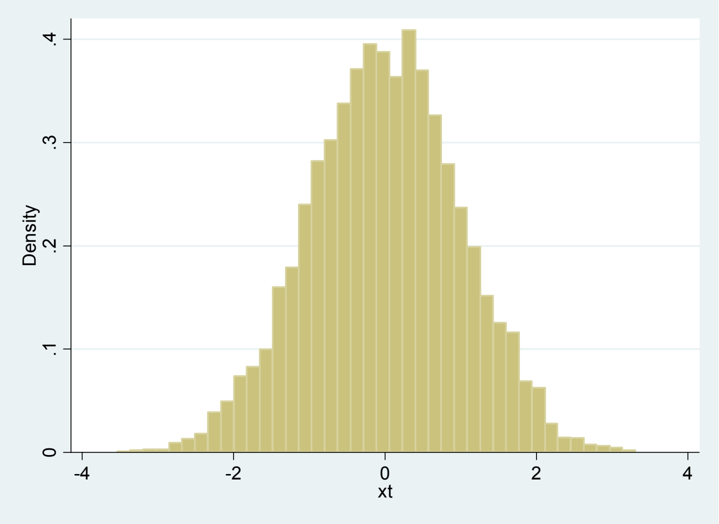
```
rndt 1000 10
hist xt
```

which will produce the following



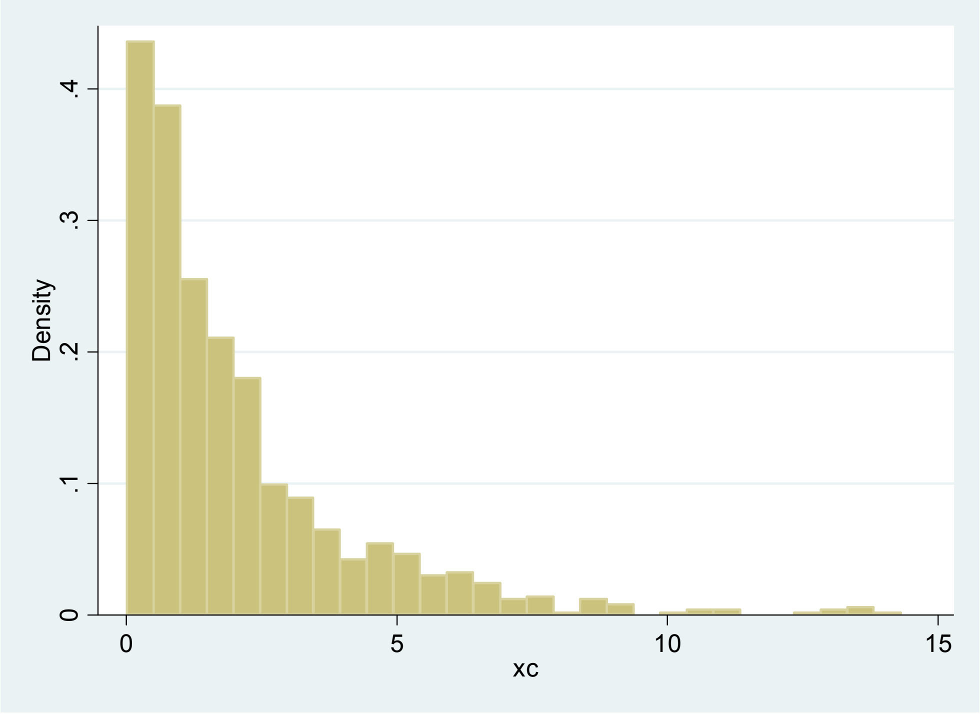
the t distribution approaches that of a normal distribution as the degrees of freedom get larger

```
rndt 10000 1000
hist xt
```



Similarly to sketch a Chi_Squared distribution with 2 degrees of freedom

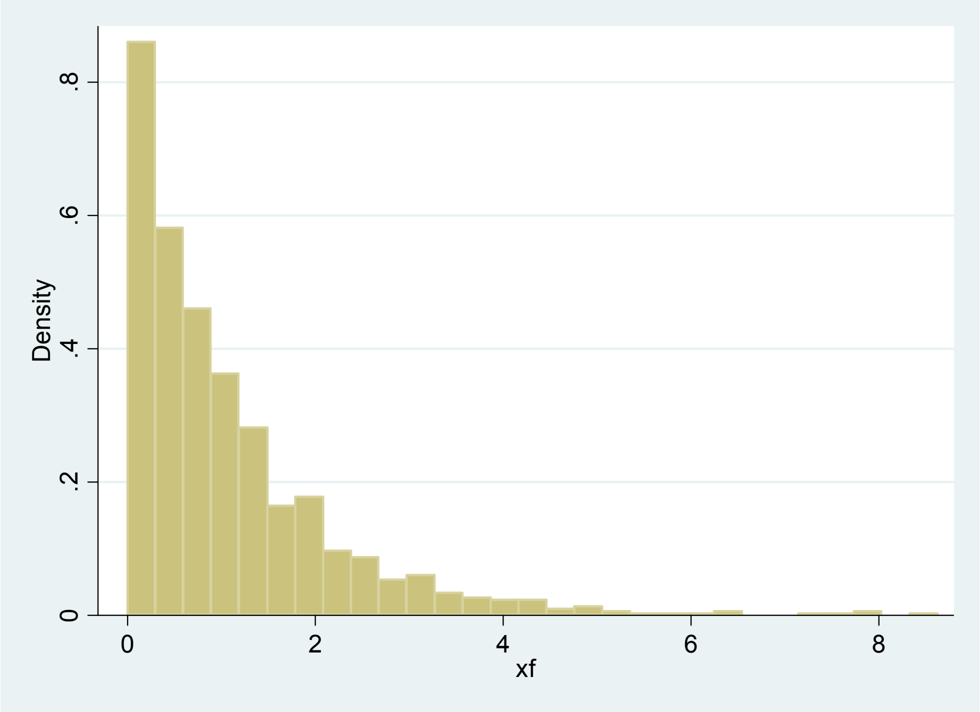
```
rndchi 1000 2
hist xc
```



Can see that the Chi-Squared distribution lower bound is zero (as it should be since all values are effectively the result of squared transformations)

Like wise the F distribution

```
rndf 1000 2 25
hist xf
```



6. What do you understand by the following:

a) The null hypothesis

is the hypothesis of interest to be tested (denoted by H0)
It is presumed to be true until statistical evidence indicates otherwise

b) Type I and Type II error

Type I Error: Rejecting the null when it is true
Type II Error: Failing to reject the null when it is false

c) The significance level of a test

the probability α of making Type I error
(also known as the *size* of the test)

the size of the test is under the control of the researcher through the specification of the area of the acceptance region

Pr(Lower bound <= θ <= Upper bound) = 1 – α

Type I error can be reduced by making the acceptance region very large
But
reducing Type I error means increasing Type II error

d) The power of a test

is the probability that a test will correctly lead to a rejection of a false null
= 1 – pr(Type II Error)

7. If A and B are square, non-singular matrices of the same order, show

i) $(AB)^{-1} = B^{-1}A^{-1}$
ii) $(A')^{-1} = (A^{-1})'$

i) Given that

$$AB(AB)^{-1} = I \tag{1}$$

Pre-multiply (1) by A^{-1}

$$A^{-1}AB(AB)^{-1} = A^{-1}I$$
$$B(AB)^{-1} = A^{-1} \tag{2}$$

Pre-multiply (2) by B^{-1}

$$B^{-1}B(AB)^{-1} = B^{-1}A^{-1}$$
$$(AB)^{-1} = B^{-1}A^{-1}$$

ie to invert a product reverse the order of multiplication and multiply the inverse of the second matrix by the inverse of the first matrix in the product

ii) We know

$$A'(A')^{-1} = I \qquad (1)$$

Transposing (1) using the rule that (AB)' = B'A'
(for proof of this rule see Johnston & DiNardo p 461)

Hence

$$[(A')^{-1}]'A = I$$

(since transposing the identity matrix leaves the matrix unchaged, I'=I)

$$[(A')^{-1}]' = A^{-1}$$

transposing this

$$(A')^{-1} = (A^{-1})'$$

so the transpose of an inverse matrix equals the inverse of the transpose matrix

8. For a square matrix A, show
   - i) tr(A') = tr(A)
   - ii) tr(A+B) = tr(A) + tr(B)
   - iii) tr(AB) = tr(BA)

i) The trace is the sum of the elements on the main diagonal of a square matrix

Since when transposing a matrix the elements on the main diagonal are unchanged then the elements

$$a_{ij} = a_{ji} \qquad \text{for all } i= j$$

in the two matrices

It follows that

$$\text{tr}(A') = \sum_{i=1}^{N} a_{ii} = \text{tr}(A)$$

ii) Given the matrix sum A + B then the ijth element of this matrix is $a_{ij} + b_{ij}$

where a_{ij} is the ijth element of the matrix A
b_{ij} is the ijth element of the matrix B

$$\text{then tr}(A+B) = \sum_{i=1}^{N} a_{ii} + b_{ii} = \sum_{i=1}^{N} a_{ii} + \sum_{i=1}^{N} b_{ii} \quad = \text{tr}(A) = \text{tr}(B)$$

iii) Given AB

then the ijth element of this matrix is the product of the ith row of A and the Jth column of B

$$= \sum_{k=1}^{N} a_{ik} b_{jk}$$

Follows that tr(AB) = $= \sum_{i=1}^{N} \sum_{k=1}^{N} a_{ik} b_{jk} = \sum_{i=1}^{N} \sum_{k=1}^{N} b_{ki} a_{ik} = tr(BA)$

9. Find the inverse of the following matrix

$$\begin{bmatrix} 3 & -1 & 1 \\ 4 & 2 & 8 \\ 1 & 2 & 3 \end{bmatrix}$$

First find the matrix of co-factors where the co-factor of the ijth element is given by

$$C_{ij} = (-1)^{i+j} M_{ij}$$

Where M_{ij} is the determinant of the submatrix formed when row I and column j are deleted from the matrix A

Follows that the matrix of cofactors is given by

$$\begin{bmatrix} 6-16 & -(12-8) & 8-2 \\ -(-3-2) & 9-1 & -(6--1) \\ -8-2 & -(24-4) & 6--4 \end{bmatrix} = \begin{bmatrix} -10 & -4 & 6 \\ 5 & 8 & -7 \\ -10 & -20 & 10 \end{bmatrix}$$

transposing this matrix of co-factors gives the ***adjoint*** of A

$$adj.A = \begin{bmatrix} -10 & 5 & -10 \\ 5-4 & 8 & -20 \\ 6 & -7 & 10 \end{bmatrix}$$

dividing by the determinant of a gives the inverse

$$A^{-1} = \frac{adjA}{|A|}$$

where $|A|$ = 3(6-16)-*-1)(12-8)+1(8-2)
= -30 + 4 + 6

$= -20$

(the determinant = elements of row I multiplied by their cofactors)

Hence

$$.A^{-1} = \begin{bmatrix} 1/2 & -1/4 & 1/2 \\ 1/5 & -2/5 & 1 \\ -3/10 & 7/20 & -1/2 \end{bmatrix}$$

10. Find the rank and determinants of the following matrices

$$A = \begin{bmatrix} 1 & 3 \\ 2 & 8 \end{bmatrix} \qquad B = \begin{bmatrix} 1 & 3 \\ 2 & 6 \end{bmatrix}$$

the rank is the maximum number of linearly independent columns (or rows) in the matrix <=Min(no. of rows, no. of columns)

a matrix is said to have full rank if the rank equals (rather than is less than) Min(no. of rows, no. of columns)

Given the matrix A it is not possible to obtain any column (or row) by a scalar multiplication of the other column (row) so

Rank(A) = 2 (full rank)

However in matrix B can see that

Col. 2 = 3* col. 1
Col. 1 = 1/3*col. 2

So the columns are linearly dependent and rank(B) = 1 (rank of order 1)

Note then that the determinants

$$|A| = (1*8) - (3*2) = 2$$

$$|B| = (1*6) - (3*2) = 0$$

so the determinant of a matrix is non-zero only if it has full rank

Hence the rank can also be given by the maximum order of a non-vanishing determinant that can be constructed from the rows and columns of that matrix

Say whether the following matrices are positive or negative definite

i) B= $\begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix}$ x=$\begin{bmatrix} 1 \\ -2 \end{bmatrix}$ ii) C = $\begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$ x=$\begin{bmatrix} 1 \\ 1 \end{bmatrix}$

iii) C’C

(Hint: convert the matrix into a quadratic form x’Ax using the column vector x)

i) B is square and symmetric (B=B’)

we know that any square symmetric matrix is +ve (-ve) definite if the product

x’Bx > (<) 0

for *any* x

Since

$$\begin{bmatrix} 1 & -2 \end{bmatrix} \begin{bmatrix} 1 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = \begin{bmatrix} -3 & -6 \end{bmatrix} \begin{bmatrix} 1 \\ -2 \end{bmatrix} = 9$$

then B is +ve definite

(note that the result of a quadratic form transformation $\underset{(1xK)}{x'}\underset{(KxK)}{B}\underset{(Kx1)}{x}$ is always a 1x1 scalar)

ii) $\begin{bmatrix} -2 & 1 \\ 1 & -2 \end{bmatrix}$

$$x'Cx = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} -2 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} -1 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = -1$$

Hence C is –ve definite

iii) C’C

$$= \begin{bmatrix} -2 & 1 \\ 1 & -1 \end{bmatrix} \begin{bmatrix} -2 & 1 \\ 1 & -1 \end{bmatrix} = \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix}$$

$$x'C'Cx = \begin{bmatrix} 1 & 1 \end{bmatrix} \begin{bmatrix} 5 & -3 \\ -3 & 2 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 2 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \end{bmatrix} = 1$$

so C’C is +ve definite

this is an example of a general rule that says the product A’A of ***any*** square symmetric matrix is always +ve semi-definite

$$x'(A'A)x = (Ax)'(Ax)$$

since

$$\underset{(KxK)}{A}\,\underset{(Kx1)}{x}$$ is a K_x1 vector and (Ax)’ is 1_xK

then their product is a scalar whose elements equal the sum of squares $\sum_i (Ax)i^2$

and since squares are always>=0 it follows that

$$x'A'Ax>=0$$

11. Given a kx1 column vector, b, differentiate the function a’b with respect to b, where a is a kx1 vector of constants.

Differentiate b’Ab with respect to b, where A is a symmetric k_xk matrix of constants

Matrix differentiation implies differentiating partially with respect to each element in the vector b

$$\frac{\delta(a'b)}{\delta b} = \begin{bmatrix} \frac{\delta(a'b)}{\delta b_1} \\ \vdots \\ \\ \vdots \\ \frac{\delta(a'b)}{\delta b_k} \end{bmatrix} = \begin{bmatrix} a_1 \\ \vdots \\ \vdots \\ . \\ a_k \end{bmatrix} = a = \frac{\delta(b'a)}{\delta b}$$

Given a quadratic f(b)=b’Ab

$$= \begin{bmatrix} b_1 & b_2 & . & . & b_k \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & & a_{1k} \\ \\ \\ a_{k1} & & & a_{kk} \end{bmatrix} \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ \vdots \\ b_k \end{bmatrix}$$

$$= \begin{bmatrix} b_1 & b_2 & . & . & b_k \end{bmatrix} \begin{bmatrix} a_{11}b_1 & +a_{12}b_2 & + & .. & +a_{1k}b_k \\ a_{21}b_1 & +a_{22}b_2 & + & .. & +a_{2k}b_k \\ & & & & \\ & & & & \\ a_{k1}b_1 + & + & . & . & +a_{kk}b_k \end{bmatrix}$$

$= b'(Ab)$

$= a_{11}b_1^2 + a_{22}b_2^2 + \ldots + a_{kk}b_k^2 + 2a_{12}b_1b_2 + 2a_{13}b_1b_3 + \ldots$

(if A is symmetric and so $a_{ij} = a_{ji}$)

It follows that the vector of partial derivatives

$$\frac{\delta[b'Ab]}{\delta b} = \begin{bmatrix} 2a_{11}b_1 & +2a_{12}b_2 & + & .. & +2a_{1k}b_k \\ 2a_{21}b_1 & +2a_{22}b_2 & + & .. & +2a_{2k}b_k \\ & & & & \\ & & & & \\ 2a_{k1}b_1 + & + & . & . & +2a_{kk}b_k \end{bmatrix} = 2Ab$$

13. Consider the following equations

a) $y_t = b_1 + b_2x_t + e_t$ b) $y_t = \hat{b}_1 + \hat{b}_2X_t + \hat{e}_t$ c) $\hat{y}_t = \hat{b}_1 + \hat{b}_2X_t$
d) $y_t = \hat{b}_1 + \hat{b}_2X_t + e_t$ e) $y_t = b_1 + b_2X_t$ f) $y_t = b_1 + b_2X_t + \hat{e}_t$
g) $\Sigma e_t = 0$ h) $\Sigma_t X_t e_t = 0$

(OLS estimated values are denoted by a ^ above the parameter/variable)

Say which equations are true and which are false.

a) True. Actual y value equals model plus true residual
b) True. Actual y value equals predicted value from model plus predicted residual
c) True. Predicted y value equals predicted value from model
d) False True y value does not equal predicted value from model plus true residual
e) False. True y value does not equal actual value from model (always add a residual)
f) False. True y value does not equal actual value from model plus estimated residual
g) False. Sum of actual residuals need not be zero (this is only an assumption)
h) False. Correlation between observed X and actual residuals need not be zero (this is only an assumption)

14. Say whether the following models are linear in parameters.

a) $y = b_0 + b_1x_1 + e$
b) $y = b_0 + b_1x_1x_2 + e$

c) $y = b_0 + b_1x_1^2 + e$
d) $y = \exp[b_0 + b_1x_1] + e$

a) Yes
b) Yes model is non-linear in variables (the interaction of X_1 & X_2) but linear in parameters
c) Yes model is non-linear in X but linear in parameters
d) No. Parameters are non-linear

Often non-linear models can be made linear in parameters by algebraic transformation (eg by taking logs)

15. Given the model Y = a + bX + e and the following information, obtain the OLS estimates of a and b.

X	1	2	3	4	5
Y	11	15	17	17	20

Ols in 2 variable model gives

$$\hat{a} = \bar{Y} + \hat{b}\,\bar{X} \qquad \hat{b} = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{N}(X_i - \bar{X})^2}$$

From the data $\bar{X} = 3 \quad \bar{Y} = 16 \quad \sum_{i=1}^{N} x_i^2 = 10 \quad \sum_{i=1}^{N} x_i\,y_i = 10$

Since $\hat{b} = \dfrac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{N}(X_i - \bar{X})^2} = \dfrac{\sum_i x_iy_i}{\sum_i x_i^2} = \dfrac{\sum_i X_iY_i - N\bar{X}\bar{Y}}{\sum_i X_i^2 - N\bar{X}^2} = \dfrac{20}{10} = 2$

So $\hat{a} = \bar{Y} + \hat{b}\,\bar{X} = 16 - (2 * 3) = 10$

16. Suppose the data for a regression are given in mean deviation form. What is the OLS estimate of the slope b in y = a + bx, (lower case letters denote mean deviation)? How does the estimate of the slope change if only the independent variable is in mean deviation form?

Given

$$Y_i = X_i\beta + u_i$$

$$Y_i = X_i\beta + e_i$$

Follows that summing and averaging within each group of T observations for individual i

$$\bar{Y}_i = \bar{X}_i \beta + \bar{e}_i$$

so

$$Y_i - \bar{Y}_i = \left(X_i - \bar{X}_i \right) \beta + \left(e_i - \bar{e}_i \right)$$

Let $y_i = Y_i - \bar{Y}_i \qquad x_i = \left(X_i - \bar{X}_i \right)$

So $y_i = \hat{\beta} x_i + \hat{e}_i$ (since $\bar{\hat{e}} = 0$)

Then OLS estimates given by $\min \sum_{i=1}^{N} \hat{e}_i^2 \Rightarrow \frac{\delta \sum_i \hat{e}_i^2}{\delta \beta} = 0 = \frac{\delta \sum (y_i - \hat{\beta} x_i)^2}{\delta \beta}$

$$0 = -2 \sum x_i (y_i - \hat{\beta} x_i)$$

Hence

$$\hat{b} = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$$

so OLS estimate of the slope is identical whether the model is estimated in mean deviation form or in levels

If only the X variable is in mean deviation form then the model becomes

$$Y_i = \left(X_i - \bar{X}_i \right) \beta + e_i$$

and the 1st order condition for minimisation becomes

$$\min \sum_{i=1}^{N} \hat{e}_i^2 \Rightarrow \frac{\delta \sum_i \hat{e}_i^2}{\delta \beta} = 0 = \frac{\delta \sum (Y_i - \hat{\beta} x_i)^2}{\delta \beta}$$

$$0 = -2 \sum x_i (Y_i - \hat{\beta} x_i)$$

Hence

$$\hat{b} = \frac{\sum_i x_i Y_i}{\sum_i x_i^2}$$

Since $\sum_i x_i Y_i = \sum_i (X_i - \bar{X})Y_i = \sum_i (X_i Y_i) - N\bar{X}\bar{Y} \equiv \sum_i (X_i - \bar{X})(Y_i - \bar{Y})$

Then $\hat{b} = \frac{\sum_i x_i Y_i}{\sum_i x_i^2} \equiv \frac{\sum_i x_i y_i}{\sum_i x_i^2}$

So OLS estimate of the slope is the same if only the X variable is measured in mean deviation form

[N.B. if only the Y variable is measured in mean deviation form then

$$\hat{b} = \frac{\sum_i X_i y_i}{\sum_i x_i^2} \neq \frac{\sum_i x_i y_i}{\sum_i x_i^2}$$

]

17. Given the model Y = a + bX + e and the following information from a sample of 22 firms, obtain the OLS estimates of a and b and test the hypotheses that i) b = 0 ii) b==1

$\bar{y}=20$ $\bar{x} = 10$ $\Sigma(y_i - \bar{y})^2 = 100$

$\Sigma(x_i - \bar{x})^2 = 60$ $\Sigma(y_i - \bar{y})(x_i - \bar{x}) = 30$

Calculate the R^2 and the standard error of the estimated equation

OLS implies

$$\hat{b} = \frac{\sum_{i=1}^{N}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{N}(X_i - \bar{X})^2} = \frac{30}{60} = 0.5$$

$$\hat{a} = \bar{Y} + \hat{b}\bar{X} = 20 - (0.5 * 10) = 15$$

So estimated regression line is

$$\hat{Y} = 15 + 0.5X$$

For t test need the estimate of the standard error of b since

$$t = \frac{\hat{b} - b^{null}}{s.e.(\hat{b})} \qquad \text{and} \qquad s.e.(\hat{b}) = \sqrt{\frac{s^2}{\sum_{i-1}^{N}(X_i - \bar{X})^2}}$$

where

$$s^2 = \sum_{i=1}^{N} \hat{e}_i^2 \Big/ N-2$$

Since $\hat{e}_i = Y_i - \hat{Y}_i$ and $\bar{\hat{e}} = \bar{Y} - \bar{X}\hat{b}$

Then $Y_i - \bar{Y} = \hat{b}(X_i - \bar{X}) + \hat{e}_i$

and $\sum_i \hat{e}_i = \sum_i \left(Y_i - \bar{Y} \right)^2 - \hat{b}^2 \sum_i (X_i - \bar{X})^2 = 100-(0.5)^2*60 = 85$

Hence $Var(\hat{b}) = 85/(22-2) = 0.0708$

and

$$s.e.(\hat{b}) = \sqrt{0.0708} = 0.2661$$

18. Suppose a regression of consumption on income using yearly observations in 1970 prices produces the following result

$$C = 100 + 0.6*\text{Income}, \qquad R^2=0.4$$

If the data were rescaled to 1997 prices using a rebasing index of 0.74 what would happen to the estimates of the slope, intercept and the coefficient of determination?

The rebased variables for any observation I are given by deflating the values in 1970 prices by 0.74

$$C_i^* = \frac{C_{i1970}}{0.74} \qquad I_i^* = \frac{I_{i1970}}{0.74}$$

It follows that the mean values

$$\bar{C}^{*} = \frac{\bar{C}1970}{0.74} \qquad \bar{I}^{*} = \frac{\bar{I}1970}{0.74}$$

Regressing C* on I gives the OLS estimate of the slope as

$$\hat{b} = \frac{\sum_{i=1}^{N}(I^{*}_{i} - \bar{I}^{*})(C^{*}_{i} - \bar{C}^{*})}{\sum_{i=1}^{N}(I^{*}_{i} - \bar{I}^{*})^{2}} = \frac{\sum_{i=1}^{N} {}^{1}\!/_{0.74}(I^{*}_{i} - \bar{I}^{*}) * {}^{1}\!/_{0.74}(C^{*}_{i} - \bar{C}^{*})}{\sum_{i=1}^{N}\left[{}^{1}\!/_{0.74}(I^{*}_{i} - \bar{I}^{*})\right]^{2}}$$

$$\Rightarrow \hat{b} = \frac{\sum_{i=1}^{N}(I_{i} - \bar{I})(C_{i} - \bar{C})}{\sum_{i=1}^{N}(I_{i} - \bar{I})^{2}}$$

So the OLS slope estimate is unchanged by the rebasing of ALL the variables in the model
But
the intercept in the rebased model does have a new value, since

$$\hat{a}^{*} = \bar{C}^{*} + \hat{b}\,\bar{I}^{*} = {}^{1}\!/_{.74}\,\bar{C}^{*} + \hat{b}\left({}^{1}\!/_{.74}\,\bar{I}^{*}\right) = \frac{\hat{a}}{0.74}$$

Note that the R^2 value is also unchanged since

$$R^{2} = \hat{\beta}^{2}\,\frac{\sum_{i=1}^{N}(I_{i} - \bar{I})^{2}}{\sum_{i=1}^{N}(C_{i} - \bar{C})^{2}} \quad \text{and} \quad R^{*2} = \hat{\beta}^{*2}\,\frac{\sum_{i=1}^{N}(I^{*}_{i} - \bar{I}^{*})^{2}}{\sum_{i=1}^{N}(C^{*}_{i} - \bar{C}^{*})^{2}}$$

So

$$R^{*2} = \hat{\beta}^{*2}\,\frac{\sum_{i=1}^{N}\left({}^{1}\!/_{0.74}(I^{*}_{i} - \bar{I}^{*})\right)^{2}}{\sum_{i=1}^{N}\left({}^{1}\!/_{0.74}(C^{*}_{i} - \bar{C}^{*})\right)^{2}} = R^{2}$$