Econometrics Test Exercise 2

a) Prove: $E(b_R) = \beta + P\beta_2$

Expressing b_R in term of

$$b_R = (X_1'X_1)^{-1}X_1'y$$

$$= (X_1'X_1)^{-1}X_1'(X_1\beta_1 + X_2\beta_2 + e)$$

by definition $y = X_1\beta_1 + X_2\beta_2 + \varepsilon$

$$= (X_1'X_1)^{-1}X_1'X_1\beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 + (X_1'X_1)^{-1}X_1'\varepsilon$$

$$= I\beta_1 + (X_1'X_1)^{-1}X_1'X_2\beta_2 + (X_1'X_1)^{-1}X_1'\varepsilon$$

$$b_R = \beta_1 + P\beta_2 + (X_1'X_1)^{-1}X_1'\varepsilon$$

as $(X_1'X_1)^{-1}(X_1'X_1) = I$

$(X_1'X_1)^{-1}(X_1'X_2) = P$

consequently $E(b_R)$ is

$$E(b_\gamma) = E\left(\beta + P\beta_2 + (x_1'x_1)^{-1}x_1'\varepsilon\right)$$

$$E(b_\gamma) = \beta_1 + P\beta_2 + (x_1'x_1)^{-1}x_1' E(\varepsilon)$$

β, x fixed (A_2, A_6)

$$E(b_\gamma) = \beta + P\beta_2$$

as $E(\varepsilon) = 0$ (A_3)

~~to Prove~~

b) Prove $Var(b_R) = \sigma^2 (X_1'X_1)^{-1}$

first define $b_R - E(b_R)$ expression in term of X_1, using part (a) findings

$$b_R - E(b_R) = [\beta_1 + P\beta_2 + (X_1'X_1)^{-1} X_1'\varepsilon] - (\beta_1 + P\beta_2) \quad \text{by (a)}$$

$$b_R - E(b_R) = (X_1'X_1)^{-1} X_1'\varepsilon$$

then we determine $Var(b_R)$

$$Var(b_R) = E([b_R - E(b_R)][b_R - E(b_R)]')$$

$$= E([(X_1'X_1)^{-1} X_1'\varepsilon][(X_1'X_1)^{-1} X_1'\varepsilon]')$$

$= E\left([(X_1'X_1)^{-1} X_1'\varepsilon][\varepsilon' X_1 (X_1'X_1)^{-1}] \right)$ as

as $(X_1'X_1)^{-1}$ Symmetric Matrix

$= E\left((X_1'X_1)^{-1} X_1' \varepsilon\varepsilon' X_1 (X_1'X_1)^{-1} \right)$.

$= (X_1'X_1)^{-1} X_1' E(\varepsilon\varepsilon') X_1 (X_1'X_1)^{-1}$ $\times X_1$ fixed (A2)

$= (X_1'X_1)^{-1} X_1' \sigma^2 I X_1 (X_1'X_1)^{-1}$ $\sigma^2 = E(\varepsilon^2)$. (A4, A5)

$= \sigma^2 (X_1'X_1)^{-1} X_1'X_1 (X_1'X_1)^{-1}$ σ^2 Scalar

$= \sigma^2 (X_1'X_1)^{-1}$ as $(X_1'X_1)^{-1} X_1'X_1 = I$

c) Prove $b_R = b_1 + P b_2$

Determine b_R

$$b_R = (x_1' x_1)^{-1} x_1' y$$

$$= (x_1' x_1)^{-1} x_1' (x_1 b_1 + x_2 b_2 + e)$$

by definition $y = x_1 b_1 + x_2 b_2 + e$

$$= (x_1' x_1) x_1' x b_1 + (x_1 x_1)^{-1} x_1' x_2 b_2 + (x_1' x_1)^{-1} x_1' e$$

$$= b_1 + P b_2 + (x_1' x_1)^{-1} x_1' e.$$

as $(x_1' x_1)^{-1} x_1' x_1 = I$

and

$$(x_1' x_1)^{-1} x_1' x_2 = P$$

$x_1' e = 0$ due to orthogonality

$$= b_1 + P b_2$$

a) Argue that Column .... Variable 'Female'.

The fact about X_1 and X_2 matrices are:

X_1 is a $(n \times 2)$ matrix are:

→ X_1 is a $(n \times 2)$ matrix, with 2 Columns of Constant term and Female Variable

→ X_1' is a $(2 \times n)$ ~~matr~~ transpose

→ X_2 is a $(n \times 3)$ matrix with 3 Column for 'Age' 'Edu' 'Parttime' Variables

Consequently $(X_1'X_1)^{-1}$ is a (2×2) matrix

So, $P \doteq (X_1'X_1)^{-1} X_1'X_2$ is a (2×3) matrix. with Column as such.

Column 1 $(X_1'X_1)^{-1} X_1' (Age)$ → OLS formula for ~~Age~~ Regressing Age on X_1

Column 2 $(X_1'X_1)^{-1} X_1' (Edu)$ → OLS for Regressing Edu on X_1

Column 3) $(X_1 X_1)^{-1} X_1 (Parttime)$ → OLS for Regressing Parttime on X_1

Week2

September 9, 2018

```
In [4]: # getting the necessary libraries to read the data
        import pandas as pd
        import numpy as np

        # import the libraries for any potential mathematical operation
        import math

        # getting libraries for the scatter diagram
        import matplotlib.pyplot as plt
        import seaborn as sns

        # necessary function to display the chart here once it is generated

        %matplotlib inline
```

```
In [5]: testExer2=pd.read_excel("C:\Users\\vishe\Downloads\Dataset2.xls",sheet_name='Dataset L
        testExer2.head()
```

Out[5]:

	Observation	Wage	LogWage	Female	Age	Educ	Parttime
0	1	66	4.189655	0	49	1	1
1	2	34	3.526361	1	42	1	1
2	3	70	4.248495	1	42	1	1
3	4	47	3.850148	0	38	1	0
4	5	107	4.672829	1	54	1	1

```
In [7]: #creating a constant term variable of value=1
        testExer2['Constant']=1
        testExer2.head()
```

Out[7]:

	Observation	Wage	LogWage	Female	Age	Educ	Parttime	Constant
0	1	66	4.189655	0	49	1	1	1
1	2	34	3.526361	1	42	1	1	1
2	3	70	4.248495	1	42	1	1	1
3	4	47	3.850148	0	38	1	0	1
4	5	107	4.672829	1	54	1	1	1

Building X1 (nx2) matrix

```
In [ ]: X1=testExer2.as_matrix(columns={'Constant','Female'})
```

and its X1' (2×n) transpose...

In [13]:
```
X1transpose= X1.transpose()
```

Building X2 (n×3) matrix...

In [ ]:
```
X2=testExer2.as_matrix(columns={'Age','Educ','Parttime'})
```

Producing X1'X1 (2×2) matrix...

In [21]:
```
X3=np.matmul(X1transpose,X1)
```

... and its (X1'X1)^-1 (2×2) inverse...

In [30]:
```
X4=np.linalg.inv(X3)
X4
```

Out[30]:
```
array([[ 0.00316456, -0.00316456],
       [-0.00316456,  0.00859934]])
```

... then, the (X1'X1)^-1 *X'1 (2×n) matrix...

In [25]:
```
X5=np.matmul(X4,X1transpose)
```

Finally, producing the P=(X1'X1)^−1*X'1X2 (2×3) matrix

In [31]:
```
P=np.matmul(X5,X2)
P
```

Out[31]:
```
array([[40.05063291,  2.25949367,  0.19620253],
       [-0.11041552, -0.49318932,  0.24944964]])
```

first column represent age

second column represent educ

third column age represent parttime

So, P matrix can be rounded as:

$$P = \begin{bmatrix} 40.05 & 2.26 & 0.2 \\ -0.11 & -0.49 & 0.25 \end{bmatrix}$$

(f) Part (c) numerical validity checking.

(f) Check the numerical validity of the result in part (c). Note: This equation will not hold exactly because the coefficients have been rounded to two or three decimals; preciser results would have been obtained for higher precision coefficients.

In Lecture 2.5, the OLS regression equation was computed as:

$$Log(Wage)_i = 3.053 - 0.041 Female_i + 0.031 Age_i + 0.233 Educ_i - 0.365 Parttime_i + e_i$$

which determines $b_1 = \begin{bmatrix} 3.053 \\ -0.041 \end{bmatrix}$ and $b_2 = \begin{bmatrix} 0.031 \\ 0.233 \\ -0.365 \end{bmatrix}$.

In [35]:
```
b1= np.array([[3.053],[-0.041]])
b1
```

Out[35]:
```
array([[ 3.053],
       [-0.041]])
```

In [36]:
```
b2=np.array([[0.031],[0.233],[-0.365]])
b2
```

Out[36]:
```
array([[ 0.031],
       [ 0.233],
       [-0.365]])
```

In part *(e)*, P matrix was computed as $P = \begin{bmatrix} 40.05 & 2.26 & 0.2 \\ -0.11 & -0.49 & 0.25 \end{bmatrix}$.

In part *(c)* it was shown that: $b_R = b_1 + Pb_2$, which can be solved as:

$$\begin{aligned} b_R &= b_1 + Pb_2 \\ &= \begin{bmatrix} 3.053 \\ -0.041 \end{bmatrix} + \begin{bmatrix} 40.05 & 2.26 & 0.2 \\ -0.11 & -0.49 & 0.25 \end{bmatrix} \begin{bmatrix} 0.031 \\ 0.233 \\ -0.365 \end{bmatrix} \\ &= \begin{bmatrix} 3.053 \\ -0.041 \end{bmatrix} + \begin{bmatrix} 1.695 \\ -0.209 \end{bmatrix} \\ &= \begin{bmatrix} 4.748 \\ -0.25 \end{bmatrix} \end{aligned}$$

This aligns with the b_R defined in Lecture 2.1 by the the 'Log(Wage)' OLS regression on the constant and 'Female' variable:

$$Log(Wage)_i = 4.73 - 0.25Female_i + e_i$$

which determined b_R as: $b_R = \begin{bmatrix} 4.73 \\ -0.25 \end{bmatrix}$.

Minor b_R values difference is explained by reduced precission due to the OLS coefficients' computation rounding values.
