# Econometrics Test Exercise 2

Jader Martins Camboim de Sá (jader.martins@ipea.gov.br)

January 6, 2021

### 1. Prove that $E(b_R) = \beta 1 + P\beta 2$.

First we have that $b_R = (X_1^T X_1)^{-1} X_1^T y$, and $y = X_1 b_1 + X_2 b_2 + \epsilon$,replacing one in another we have

$$b_R = (X_1^T X_1)^{-1} X_1^T (X_1 b_1 + X_2 b_2 + \epsilon), \tag{1}$$

distributing the multiplying factor we have

$$b_R = (X_1^T X_1)^{-1} X_1^T X_1 b_1 + (X_1^T X_1)^{-1} X_1^T X_2 b_2 + (X_1^T X_1)^{-1} X_1^T \epsilon. \tag{2}$$

The term $(X_1^T X_1)^{-1} X_1^T X_1 = I$, because $A^{-1}A = I$, so we simplify for

$$b_R = b_1 + (X_1^T X_1)^{-1} X_1^T X_2 b_2 + (X_1^T X_1)^{-1} X_1^T \epsilon. \tag{3}$$

The expected value of this equation is given by

$$E(b_R) = \beta_1 + (X_1^T X_1)^{-1} X_1^T X_2 \beta_2, \tag{4}$$

as the error $\epsilon$ is expected to have 0 mean. Replacing $(X_1^T X_1)^{-1} X_1^T X_2$ by $P$ we have

$$E(b_R) = \beta_1 + P\beta_2 \quad \square. \tag{5}$$

### 2. Prove that $var(b_R) = \sigma^2 (X_1^T X_1)^{-1}$.

We have that $var(A) = E[(A - E[A])^2] = E[A^2] - E[A]^2$. So we have

$$var(b_R) = E[(b_R - E[b_R])^2], \tag{6}$$

with the equation from previous question we have

$$var(b_R) = E[(b_1 + Pb_2 + (X_1^T X_1)^{-1} X_1^T \epsilon)^2] - (\beta_1 + P\beta_2)^2, \tag{7}$$

which we replace $a = b_1 + Pb_2$ and $c = (X_1^T X_1)^{-1} X_1^T \epsilon$, now we have

$$var(b_R) = E[a^2 + 2ac + c^2] - \beta_1^2 - P\beta_2^2 - 2\beta_1 P\beta_2. \tag{8}$$

With the linearity of expectation we get

$$var(b_R) = E[a^2] + E[2ac] + E[c^2] - \beta_1^2 - P\beta_2^2 - 2\beta_1 P\beta_2. \tag{9}$$

Replacing back

$$\begin{aligned} var(b_R) = E[(b_1 + Pb_2)^2] + E[2(b_1 + Pb_2)((X_1^T X_1)^{-1} X_1^T \epsilon)] + \\ E[((X_1^T X_1)^{-1} X_1^T \epsilon)^2] - \beta_1^2 - P\beta_2^2 - 2\beta_1 P\beta_2 \end{aligned} \tag{10}$$

The term $E[2(b_1 + Pb_2)((X_1^T X_1)^{-1} X_1^T \epsilon)] = 0$ as the $\epsilon$ has zero mean. We open the other terms

$$var(b_R) = E[b_1^2 + Pb_2^2 + 2b_1 Pb_2] + E[((X_1^T X_1)^{-1} X_1^T \epsilon)^2] - \beta_1^2 - P\beta_2^2 - 2\beta_1 P\beta_2. \tag{11}$$

The first term cancel with the last terms and then

$$var(b_R) = E[(X_1^T X_1)^{-1} X_1^T \epsilon\epsilon^T X_1 (X_1^T X_1)^{-1}], \tag{12}$$

with $(X_1^T X_1)^{-1}(X_1^T X_1) = I$ and $E[\epsilon^2] = \sigma^2$ we have

$$var(b_R) = \sigma^2 (X_1^T X_1)^{-1} \quad \square. \tag{13}$$

### 3. Prove that $b_R = b_1 + Pb_2$.

By (1.) we have

$$b_R = b_1 + (X_1^T X_1)^{-1} X_1^T X_2 b_2 + (X_1^T X_1)^{-1} X_1^T \epsilon. \tag{14}$$

and

$$b_R = b_1 + Pb_2 + (X_1^T X_1)^{-1} X_1^T \epsilon. \tag{15}$$

the product between the vector $\epsilon$ leds to a vetor of zeros, as it scale terms to cancel themselves (zero mean), so

$$bR = b1 + Pb2 \quad \square. \tag{16}$$

### 4. Argue that the columns of the (2 x 3) matrix P are obtained by regressing each of the variables 'Age', 'Educ', and 'Parttime' on a constant term and the variable 'Female'.

The Female and constant term, $X_1$, have dimensionality (500 x 2) those three variables, $X_2$, has dimensionality (500 x 3), for the regression equation we have

$$P = (X_1^T X_1)^{-1} X_1^T X_2 \tag{17}$$

As $X_1^T$ das simensionality (2 x 500), the dimensionality of $(X_1^T X_1)^{-1}$ is (2 x 2) now $(X_1^T X_1)^{-1} X_1^T$ has dimensionality (2 x 500), and as $X_2$ have dim equal (500 x 3) we get P with dim equal (2 x 3)

### 5. Determine the values of P from the results in Lecture 2.1.

Resultas are reported in Table 1.

| | Age | Educ | Parttime |
|---|---|---|---|
| Constant | 40.05 | 2.26 | 0.20 |
| Female | -0.11 | -0.49 | 0.25 |

Table 1: Caption

### 6. Check the numerical validity of the result in part (4.).

Regressing the data in R we have

```
## Call:
## lm(formula = LogWage ~ Female + Age + Educ + Parttime, data = df)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.73843 -0.15752 -0.00406  0.16491  0.77868
```

```
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.052685   0.055334  55.168   <2e-16 ***
## Female      -0.041101   0.024711  -1.663   0.0969 .
## Age          0.030606   0.001273  24.041   <2e-16 ***
## Educ         0.233178   0.010660  21.874   <2e-16 ***
## Parttime    -0.365449   0.031571 -11.576   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2452 on 495 degrees of freedom
## Multiple R-squared:  0.704,  Adjusted R-squared:  0.7016
## F-statistic: 294.3 on 4 and 495 DF,  p-value: < 2.2e-16
```

And the values

```
##                 Age       Educ  Parttime
## Constant 40.0506329  2.2594937 0.1962025
## Female   -0.1104155 -0.4931893 0.2494496
```

```
##      [,1]
## b1  4.73
## b2 -0.25
```

so it is valid.