BS2280 – Econometrics I
Homework 5: Multiple Regression Model I

1

When we interpret the coefficients of a multiple regression model, we always add "holding everything else constant". Using an example, explain what it means.

2

Data on 935 individuals was collected to identify what factors can explain the variation in wage data. Firstly, a simple regression is run regressing wages (monthly earnings in USD) on years of education. Secondly, a multiple regression is run regressing wages on years of education and years of work experience.

i Interpret the intercepts and coefficients of both regressions and make reference to their statistical significance.

ii Explain why the omission of the work experience variable in the simple regression model led to an underestimation of the impact of education on wages.

iii Using the multiple regression model, predict the wage of someone who has 12 years of education and 1 year of work experience.

Simple regression model:

```
Call:
lm(formula = wages2$wage ~ wages2$educ)

Residuals:
    Min      1Q  Median      3Q     Max
-877.38 -268.63  -38.38  207.05 2148.26

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  146.952     77.715   1.891   0.0589 .
wages2$educ   60.214      5.695  10.573   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 382.3 on 933 degrees of freedom
Multiple R-squared:  0.107,     Adjusted R-squared:  0.106
F-statistic: 111.8 on 1 and 933 DF,  p-value: < 2.2e-16
```

Multiple regression model:

```
Call:
lm(formula = wages2$wage ~ wages2$educ + wages2$exper)

Residuals:
    Min      1Q  Median      3Q     Max
-924.38 -252.74  -40.88  198.16 2165.70

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  -272.528    107.263  -2.541   0.0112 *
wages2$educ    76.216      6.297  12.104  < 2e-16 ***
wages2$exper   17.638      3.162   5.578 3.18e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 376.3 on 932 degrees of freedom
Multiple R-squared:  0.1359,    Adjusted R-squared:  0.134
F-statistic: 73.26 on 2 and 932 DF,  p-value: < 2.2e-16
```

3

The output below is the result of fitting an educational attainment function, regressing S on $ASVABC$, a measure of cognitive ability, SM, and SF, years of schooling (highest grade completed) of the respondent's mother and father, respectively.

a. Give an interpretation of the regression coefficients.

b. Undertake hypothesis tests to show whether the coefficients are statistically significant. The critical $t - value = 1.965$ (5% significance level).

c. Is the R^2 is statistically significant? The critical F value at the 5% significance level is 2.62. Interpret your result.

d. Calculate the 95% confidence interval for each coefficient.

```
Call:
lm(formula = EAWE21$S ~ EAWE21$ASVABC + EAWE21$SM + EAWE21$SF)

Residuals:
    Min      1Q  Median      3Q     Max
-5.9387 -1.6521  0.0186  1.5161  7.1553

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  10.59674    0.61428
EAWE21$ASVABC  1.24253    0.12359
EAWE21$SM      0.09135    0.04593
EAWE21$SF      0.20289    0.04251
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.254 on 496 degrees of freedom
Multiple R-squared:  0.329,     Adjusted R-squared:  0.3249
F-statistic: 81.06 on 3 and 496 DF,  p-value: < 2.2e-16
```

4

Explain the differences between R^2 and adjusted R^2 and calculate adjusted R^2 using the information from question 3. The formula of adjusted R^2 is

$$\bar{R}^2 = R^2 - \frac{k-1}{n-k}(1-R^2)$$