BS2280 – Econometrics I
Homework 8: Dummy Variables - Solution

1

Does the sex of an individual affect educational attainment? We regress *S (educational attainment in years)* on *ASVABC (Ability score)*, *SM (educational attainment of mother in years)*, *SF (educational attainment of father in years)*, and *MALE*, a dummy variable that is 1 for male respondents and 0 for female ones. Interpret the coefficients and perform t-tests. The critical t value at the 5% significance level is 1.96. Is there any evidence that the educational attainment of males is different from that of females?

```
Call:
lm(formula = S ~ ASVABC + SM + SF + MALE, data = EAWE22)

Residuals:
    Min      1Q  Median      3Q     Max
-6.6240 -1.5514  0.0377  1.4935  6.3454

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.90114    0.59470  18.331  < 2e-16 ***
ASVABC       1.20327    0.11344  10.607  < 2e-16 ***
SM           0.17453    0.04755   3.671 0.000268 ***
SF           0.11214    0.04125   2.719 0.006782 **
MALE        -0.86372    0.20170  -4.282 2.22e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.228 on 495 degrees of freedom
Multiple R-squared:  0.3573,    Adjusted R-squared:  0.3521
F-statistic: 68.81 on 4 and 495 DF,  p-value: < 2.2e-16
```

ASVABC:
Interpretation: A one score higher ability score increases educational attainment on average by 1.2 more years, everything else constant.

$H_0 : \beta_2 = 0; H_1 : \beta_2 \neq 0$
$t - value = \frac{\hat{\beta}_2 - \beta_2^0}{s.e.(\hat{\beta}_2)} = \frac{1.20327 - 0}{0.11344} \approx 10.607 > t_{crit,5\%} = 1.96$
therefore we reject the null hypothesis at the 5% significance level. The coefficient of ability score is statistically significant.

SM:
Interpretation: One more year of educational attainment of the mother leads on average to 0.17 years of educational attainment, ceteris paribus.

$H_0 : \beta_3 = 0; H_1 : \beta_3 \neq 0$
$t - value = \frac{\hat{\beta}_3 - \beta_3^0}{s.e.(\hat{\beta}_3)} = \frac{0.17396 - 0}{0.04802} \approx 3.623 > t_{crit,5\%} = 1.96$
therefore we reject the null hypothesis at the 5% significance level. The coefficient of educational attainment of the mother is statistically significant.

SF:
Interpretation: One more year of educational attainment of the father leads on average to 0.11 years of educational attainment, ceteris paribus.

$H_0 : \beta_4 = 0; H_1 : \beta_4 \neq 0$
$t - value = \frac{\hat{\beta}_4 - \beta_4^0}{s.e.(\hat{\beta}_4)} = \frac{0.11385 - 0}{0.04196} \approx 2.7143 > t_{crit,5\%} = 1.96$
therefore we reject the null hypothesis at the 5% significance level. The coefficient of educational attainment of the father is statistically significant.

$MALE$:
Interpretation: Men have on average 0.84 years less education than women, ceteris paribus

$H_0 : \delta = 0; H_1 : \delta \neq 0$
$t - value = \frac{\hat{\delta} - \delta^0}{s.e.(\hat{\delta})} = \frac{-0.86372 - 0}{0.20170} \approx |-4.282| > t_{crit,5\%} = 1.96$
therefore we reject the null hypothesis at the 5% significance level. The coefficient of the male dummy is statistically significant.

$Intercept$:
Interpretation: A person with an ability score of 0, who has parents with no educational attainment and who is female will have on average 10.9 years of education.

$H_0 : \beta_1 = 0; H_1 : \beta_1 \neq 0$
$t - value = \frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)} = \frac{10.90114 - 0}{0.59470} \approx 18.331 > t_{crit,5\%} = 1.96$
therefore we reject the null hypothesis at the 5% significance level. The intercept is statistically significant.

Summary: The slope coefficients are all significant at the 1 percent level or higher. That for $MALE$ indicates that males tend to have nearly a year less schooling than females, controlling for $ASVABC$ score and parental education. The reason is that although males are now under-represented at the postgraduate level, a relatively recent phenomenon, as well as over-represented among high school drop-outs, a long-standing one.

2

Does ethnicity affect educational attainment? We add the following ethnic dummy variables to the regression model above:

$ETHHISP$	1 if hispanic, 0 otherwise
$ETHBLACK$	1 if black, 0 otherwise
$ETHWHITE$	1 if not hispanic or black, 0 otherwise

We regress S on $ASVABC$, $MALE$, SM, SF, $ETHBLACK$, and $ETHHISP$. In this specification $ETHWHITE$ has been chosen as the reference category, and so it is omitted. Interpret the regression results and perform t tests on the coefficients. The critical t value at the 5% significance level is 1.96.

```
Call:
lm(formula = S ~ ASVABC + SM + SF + MALE + ETHBLACK + ETHHISP,
    data = EAWE22)

Residuals:
    Min      1Q  Median      3Q     Max
-6.5677 -1.5150  0.0058  1.4156  6.4117

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.78365    0.63926  16.869  < 2e-16 ***
ASVABC       1.26416    0.12042  10.498  < 2e-16 ***
SM           0.17396    0.04802   3.623 0.000321 ***
SF           0.11385    0.04196   2.714 0.006889 **
MALE        -0.83509    0.20254  -4.123 4.39e-05 ***
ETHBLACK     0.51097    0.34022   1.502 0.133760
ETHHISP      0.18325    0.33330   0.550 0.582705
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.227 on 493 degrees of freedom
Multiple R-squared:  0.3604,    Adjusted R-squared:  0.3526
F-statistic: 46.29 on 6 and 493 DF,  p-value: < 2.2e-16
```

$ASVABC$:
Interpretation: A one score higher ability score increases educational attainment on average by 1.2 more years, everything else constant.

$H_0 : \beta_2 = 0; H_1 : \beta_2 \neq 0$
$t - value = \frac{\hat{\beta}_2 - \beta_2^0}{s.e.(\hat{\beta}_2)} = \frac{1.26416-0}{0.12042} \approx 10.498 > t_{crit,5\%} = 1.96$
therefore we reject the null hypothesis at the 5% significance level. The coefficient of ability score is statistically significant.
SM:
Interpretation: One more year of educational attainment of the mother leads on average to 0.17 years of educational attainment, ceteris paribus.

$H_0 : \beta_3 = 0; H_1 : \beta_3 \neq 0$
$t - value = \frac{\hat{\beta}_3 - \beta_3^0}{s.e.(\hat{\beta}_3)} = \frac{0.17396 - 0}{0.04802} \approx 3.623 > t_{crit,5\%} = 1.96$
therefore we reject the null hypothesis at the 5% significance level. The coefficient of educational attainment of the mother is statistically significant.

SF:
Interpretation: One more year of educational attainment of the father leads on average to 0.11 years of educational attainment, ceteris paribus.

$H_0 : \beta_4 = 0; H_1 : \beta_4 \neq 0$
$t - value = \frac{\hat{\beta}_4 - \beta_4^0}{s.e.(\hat{\beta}_4)} = \frac{0.11385 - 0}{0.04196} \approx 2.7143 > t_{crit,5\%} = 1.96$
therefore we reject the null hypothesis at the 5% significance level. The coefficient of educational attainment of the father is statistically significant.

$MALE$:
Interpretation: Men have on average 0.84 years less education than women, ceteris paribus

$H_0 : \delta = 0; H_1 : \delta \neq 0$
$t - value = \frac{\hat{\delta} - \delta^0}{s.e.(\hat{\delta})} = \frac{-0.83509 - 0}{0.20254} \approx |-4.123| > t_{crit,5\%} = 1.96$
therefore we reject the null hypothesis at the 5% significance level. The coefficient of the male dummy is statistically significant.

$ETHBLACK$:
Interpretation: A black person has on average 0.51 years more education than a white person, ceteris paribus

$H_0 : \delta_b = 0; H_1 : \delta_b \neq 0$
$t - value = \frac{\hat{\delta_b} - \delta_b^0}{s.e.(\hat{\delta_b})} = \frac{0.51097 - 0}{0.34022} \approx 1.502 < t_{crit,5\%} = 1.96$
therefore we cannot reject the null hypothesis at the 5% significance level. The coefficient of the black ethnicity dummy is not statistically significant. This means that we cannot say that on average a black person has more education than a white person.

$ETHHISP$:
Interpretation: A hispanic person has on average 0.18 years more education than a white person, ceteris paribus

$H_0 : \delta_h = 0; H_1 : \delta_h \neq 0$
$t - value = \frac{\hat{\delta_h} - \delta_h^0}{s.e.(\hat{\delta_h})} = \frac{0.18325 - 0}{0.33330} \approx 0.550 < t_{crit,5\%} = 1.96$
therefore we cannot reject the null hypothesis at the 5% significance level. The coefficient of the hispanic ethnicity dummy is not statistically significant. This means that we cannot say that on average a hispanic person has more education than a white person.

$Intercept$:

Interpretation: A white person with an ability score of 0, who has parents with no educational attainment and who is female will have on average 10.8 years of education.

$H_0 : \beta_1 = 0; H_1 : \beta_1 \neq 0$
$t - value = \frac{\hat{\beta}_1 - \beta_1^0}{s.e.(\hat{\beta}_1)} = \frac{10.78365-0}{0.63926} \approx 16.869 > t_{crit,5\%} = 1.96$
therefore we reject the null hypothesis at the 5% significance level. The intercept is statistically significant.

Summary: Both blacks and Hispanics have more schooling than whites (short for non-black, non-hispanic), controlling for $ASVABC$ score, sex, and parental education, but in neither case is the difference significant, at least in this sample.

3

Using the ANOVA tables below, Evaluate whether the ethnicity dummies as a group have significant explanatory power for educational attainment by comparing the residual sums of squares in the regressions in Question 1 and 2. The critical F value at the 5% significance level is 3.01. Analysis of Variance Table

Model 1:

```
Response: S
           Df  Sum Sq Mean Sq  F value    Pr(>F)
ASVABC      1 1089.37 1089.37 219.4517 < 2.2e-16 ***
SM          1  161.56  161.56  32.5468 2.005e-08 ***
SF          1   24.26   24.26   4.8869   0.02752 *
MALE        1   91.03   91.03  18.3382 2.222e-05 ***
Residuals 495 2457.21    4.96
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Model 2:

```
Response: S
           Df  Sum Sq Mean Sq  F value    Pr(>F)
ASVABC      1 1089.37 1089.37 219.6059 < 2.2e-16 ***
SM          1  161.56  161.56  32.5697 1.987e-08 ***
SF          1   24.26   24.26   4.8904   0.02746 *
MALE        1   91.03   91.03  18.3511 2.209e-05 ***
ETHBLACK    1   10.15   10.15   2.0454   0.15330
ETHHISP     1    1.50    1.50   0.3023   0.58271
Residuals 493 2445.56    4.96
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

You have to undertake an F-test:
Step 1. State the null and alternative hypotheses
Step 2. Select the significance level. Significance level $\alpha = 5\%$

Null Hypothesis	$H_0 : \delta_b = \delta_h = 0$
Alternative Hypothesis	$H_1 : \delta_b \neq 0 \, or \, \delta_h \neq 0 \, or \, both \, \delta_b \, and \, \delta_h \neq 0$

Step 3. Select and calculate the test statistics

$$F(cost\,in\,dof, dof\,remaining) = \frac{reduction\,in\,RSS/cost\ in\,dof}{RSS\,remaining/dof\,remaining} = \frac{(RSS_1 - RSS_2)/cost\ in\,dof}{RSS_2/dof\,remaining} \quad (1)$$

$RSS_1 = 2457.21$; $RSS_2 = 2445.56$; cost in dof $= 2$ dof remaining $n - k = 500 - 7 = 493$

$$F(2, 493) = \frac{(RSS_1 - RSS_2)/cost\ in\,dof}{RSS_2/dof\,remaining} = \frac{(2457.21 - 2445.56)/2}{2445.56/493} = 1.17 \quad (2)$$

Step 4. Set the decision rule.
$cost\,in\,dof = number\,of\,new\,variables\,added = 2$
$k = 7, n = 500, \; dof\,remaining = n - k = 500 - 7 = 493$

$F_{crit,5\%}(cost\,in\,dof, dof\,remaining) = F_{crit,5\%}(2, 493) = 3.01$

Step 5. Make statistical decisions.
$F = 1.17 < F_{crit,5\%}(2, 493) = 3.01$
We cannot reject the null $H_0 : \delta_b = \delta_h = 0$.
We conclude that Adding the ethnicity dummies does not improve the overall fit of the model significantly.

4

Is the effect of the $ASVABC$ score on educational attainment different for males and females? We define a slope dummy variable $MALEASVC$ as the product of $MALE$ and $ASVABC$:

$$MALEASVC = MALE \times ASVABC$$

Regress S on $ASVABC$, SM, SF, $ETHBLACK$, $ETHHISP$, $MALE$, and $MALEASVC$, interpret the equation.

```
Call:
lm(formula = S ~ SM + SF + ETHBLACK + ETHHISP + ASVABC + MALE +
    MALEASVC, data = EAWE22)

Residuals:
    Min      1Q  Median      3Q     Max
-6.6488 -1.5246 -0.0176  1.3740  6.4195

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.77864    0.63893  16.870  < 2e-16 ***
SM           0.16983    0.04811   3.530 0.000454 ***
SF           0.11518    0.04195   2.746 0.006258 **
ETHBLACK     0.52734    0.34029   1.550 0.121870
ETHHISP      0.19088    0.33318   0.573 0.566973
ASVABC       1.42912    0.17996   7.941 1.37e-14 ***
MALE        -0.78987    0.20573  -3.839 0.000140 ***
MALEASVC    -0.26292    0.21326  -1.233 0.218213
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.226 on 492 degrees of freedom
Multiple R-squared:  0.3623,    Adjusted R-squared:  0.3533
F-statistic: 39.94 on 7 and 492 DF,  p-value: < 2.2e-16
```

Combined regression model:
$\widehat{S}_i = 10.78 + 0.17SM_i + 0.12SF_i + 0.53ETHBLACK_i + 0.19ETHHISP_i + 1.43ASVABC_i - 0.78MALE_i - 0.26MALE_i \times ASVABC_i$

Female ($MALE = 0$) regression model:
$\widehat{S_{Fi}} = 10.78 + 0.17SM_i + 0.12SF_i + 0.53ETHBLACK_i + 0.19ETHHISP_i + 1.43ASVABC_i$

Male ($MALE = 1$) regression model:
$\widehat{S_{Mi}} = 10.78 + 0.17SM_i + 0.12SF_i + 0.53ETHBLACK_i + 0.19ETHHISP_i + 1.43ASVABC_i - 0.79 - 0.26ASVABC_i$

$\widehat{S_{Mi}} = 9.99 + 0.17SM_i + 0.12SF_i + 0.53ETHBLACK_i + 0.19ETHHISP_i + 1.17ASVABC_i$

We find a lower intercept for the male sample, meaning that if all covariates for our male and female sample were zero, then men would have 0.79 years less education than women. This difference is highly statistically significant.

While we find the slope-coefficient of the ability score to be lower for men than for women, this difference is not statistically significant at the 5% or even 10% significance level.

Advice: Attempt to draw the regression line for the male and female sample.