# BS2280 – Econometrics I
# Homework 6: Multiple Regression Model II

## 1

It seems that the first study in hedonic pricing was Waugh's investigation of the prices of vegetables in the Boston wholesale market (Waugh, 1929). Waugh was an economist working for the Bureau of Agricultural Economics and he was surprised to find that one box of cucumbers might sell for \$7 while another for only \$1. Being told that thinner cucumbers had better texture and taste than fat ones, he fitted the following regression (standard errors in parentheses, data from 1925):

$$\begin{array}{llll} \hat{P}_i = & 508.0 & +32.3 L_i & -8.80 D_i \qquad R^2 = 0.35 \\ & (272.0) & (20.1) & (4.45) \qquad F(2, 47) = 12.43 \end{array}$$

where $P_i$ is the price, in cents, of a box of cucumbers and $L_i$ and $D_i$ are the length in inches and the diameter/length ratio, as a percentage, of the cucumbers in the box. The boxes in the market were carefully sorted so that their contents were uniform in terms of these characteristics. Give an interpretation of the regression results.

An extra inch of length increases the value of the box by 32.3 cents. A one percent increase in the diameter/length ratio reduces the value of the box by 8.8 cents. Note that the number of cucumbers in a box will be different for different dimensions. The volume and approximate total weight (approximately a bushel) are fixed.

## 2

We estimate an educational attainment function by regressing $S$ (educational attainment) on a general ability score ($ASVABC$), and the educational attainment of the mother ($SM$) and father ($SF$).

a. Undertake an F-test (using the ANOVA table below) to check the overall significance of the estimated model. The critical F value at the 5% significance level is $F_{crit} = 2.62$.

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

Residuals:
    Min      1Q  Median      3Q     Max 
-6.2094 -1.6985  0.0289  1.5765  6.5142 

Coefficients:
            Estimate Std. Error t value Pr(>|t|)    
(Intercept) 10.64842    0.60201  17.688  < 2e-16 ***
ASVABC       1.26116    0.11458  11.006  < 2e-16 ***
SM           0.18212    0.04834   3.768 0.000185 ***
SF           0.09049    0.04164   2.173 0.030254 *  
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.267 on 496 degrees of freedom
Multiple R-squared:  0.3335,	Adjusted R-squared:  0.3295 
F-statistic: 82.74 on 3 and 496 DF,  p-value: < 2.2e-16
```

```
Analysis of Variance Table

Response: S
           Df  Sum Sq Mean Sq  F value    Pr(>F)    
ASVABC      1 1089.37 1089.37 212.0396 < 2.2e-16 ***
SM          1  161.56  161.56  31.4475 3.408e-08 ***
SF          1   24.26   24.26   4.7219   0.03025 *  
Residuals 496 2548.24    5.14                       
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Write hypothesis test for $R^2$: $H_0 : R^2 = 0; H_1 : R^2 \neq 0$
Significance level: 5%
Calculate F-statistic:
You can use $R^2$ to calculate the F-statistic (or find it in the output table).
$R^2 = 0.3335$

$$F(3, 496) = \frac{R^2/(k-1)}{(1-R^2)/(n-k)} = \frac{0.3335/(4-1)}{(1-0.3335)/496} \approx 82.74$$

F-statistic = 82.74.
Compare test statistic with the critical F-value: $82.74 > 2.62$
The F statistic is greater than the critical F value, therefore we can reject the null hypothesis. The estimated model is statistically significant.

b. In another attempt to improve our estimated model, we add two speed test scores, $ASVABNO$ and $ASVABCS$, to the regression model:

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

Residuals:
    Min      1Q  Median      3Q     Max
-6.3751 -1.6673 -0.0382  1.4315  6.6593

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 10.63489    0.59645  17.830  < 2e-16 ***
ASVABC       0.97934    0.13973   7.009 7.93e-12 ***
SM           0.18155    0.04798   3.784 0.000173 ***
SF           0.09134    0.04127   2.213 0.027323 *
ASVABNO      0.29969    0.13752   2.179 0.029784 *
ASVABCS      0.20889    0.12787   1.634 0.102982
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 2.244 on 494 degrees of freedom
Multiple R-squared:  0.3492,    Adjusted R-squared:  0.3427
F-statistic: 53.02 on 5 and 494 DF,  p-value: < 2.2e-16
```

*Analysis of Variance Table*

```
Response: S
           Df  Sum Sq Mean Sq  F value    Pr(>F)
ASVABC      1 1089.37 1089.37 216.2885 < 2.2e-16 ***
SM          1  161.56  161.56  32.0777 2.517e-08 ***
SF          1   24.26   24.26   4.8165  0.028653 *
ASVABNO     1   46.69   46.69   9.2703  0.002453 **
ASVABCS     1   13.44   13.44   2.6686  0.102982
Residuals 494 2488.11    5.04
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Using the $ANOVA$ table below, perform an F test of the joint explanatory power of $ASVABNO$ and $ASVABCS$. The critical F statistics at the 1% significance level is $F_{crit} = 3.01$.

Write hypothesis test: $H_0 : \beta_4 = \beta_5 = 0; H_1 : \beta_4 \neq \beta_5 \neq 0$
Significance level: 1%
Calculate F-statistic:
You need to use the information in two Analysis of Variance Tables
From the first Analysis of Variance Table, $RSS_1 = 2548.24$
From the second Analysis of Variance Table, $RSS_2 = 2488.11$

$$F(cost\ in\ d.f., d.f.\ remaining) = \frac{reduction\ in\ RSS/cost\ in\ d.f.}{RSS\ remaining/degrees\ of\ freedom\ remaining}$$

$$F(2, 494) = \frac{(RSS_1 - RSS_2)/2}{RSS_2/494} = \frac{(2548.24 - 2488.11)/2}{2488.11/494} \approx 5.97$$

Compare test statistic with the critical F-value: $5.97 > 3.01$
The F statistic is greater than the critical F value, therefore we can reject the null hypothesis. We conclude that $ASVABNO$ and $ASVBCS$ do have significant joint explanatory power.