# Homework Solutions
# Applied Regression Analysis

## WEEK 5

### Exercise One

Earlier in the course we studied the multiple regression relationship of SBP *(Y)* to AGE *($X_1$)*, SMK *($X_2$)*, and QUET *($X_3$)* using the data in Homework 1 of Week 2. *Please refer to the dataset from Week 2 homework if you need to.*

Three regression models were considered:

| Model | Independent Variables Used |
|---|---|
| 1 | AGE *($X_1$)* |
| 2 | AGE *($X_1$)*, SMK *($X_2$)*, |
| 3 | AGE *($X_1$)*, SMK *($X_2$)*, QUET *($X_3$)* |

**First**, use your computer to generate each of the above models.

In order to fit the three models, you have to consider each model separately.

### Model 1

Type 'regress sbp age' in the command window. From the output, you can obtain the coefficient for the slope ($\beta_1$) as well as the intercept ($\beta_0$) in the bottom right corner of the output in the "Coef." column.

**Model 1**

```
. regress sbp age

  Source |       SS       df       MS              Number of obs =      32
---------+------------------------------           F(  1,    30) =   45.18
   Model |  3861.63038     1  3861.63038           Prob > F      =  0.0000
Residual |  2564.33838    30  85.4779458           R-squared     =  0.6009
---------+------------------------------           Adj R-squared =  0.5876
   Total |  6425.96875    31  207.289315           Root MSE      =  9.2454

------------------------------------------------------------------------------
     sbp |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |     1.6045   .2387159      6.721   0.000       1.116977    2.092023
   _cons |   59.09162   12.81626      4.611   0.000       32.91733    85.26592
------------------------------------------------------------------------------
```

## Model 2

Type ‘regress sbp age smk’ in the command window. From the output, you can obtain the coefficient for $\beta_1$ and $\beta_2$ as well as the intercept ($\beta_0$) in the bottom right corner of the output in the “Coef.” column.

**Model 2**

```
. regress sbp age smk

  Source |       SS       df       MS              Number of obs =      32
---------+------------------------------           F(  2,    29) =   39.16
   Model |  4689.68423     2  2344.84211           Prob > F      =  0.0000
Residual |  1736.28452    29    59.87188           R-squared     =  0.7298
---------+------------------------------           Adj R-squared =  0.7112
   Total |  6425.96875    31  207.289315           Root MSE      =  7.7377

------------------------------------------------------------------------------
     sbp |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |    1.70916   .2017587      8.471   0.000       1.296517    2.121803
     smk |   10.29439   2.768107      3.719   0.001       4.632978    15.95581
   _cons |    48.0496   11.12956      4.317   0.000        25.2871    70.81211
------------------------------------------------------------------------------
```

## Model 3

Type ‘regress sbp age smk quet’ in the command window. From the output, you can obtain the coefficient for $\beta_1$, $\beta_2$ and $\beta_3$ as well as the intercept ($\beta_0$) in the bottom right corner of the output in the “Coef.” column.

**Model 3**

```
. regress sbp age smk quet

  Source |       SS       df       MS              Number of obs =      32
---------+------------------------------           F(  3,    28) =   29.71
   Model |  4889.82567     3  1629.94189           Prob > F      =  0.0000
Residual |  1536.14308    28  54.8622529           R-squared     =  0.7609
---------+------------------------------           Adj R-squared =  0.7353
   Total |  6425.96875    31  207.289315           Root MSE      =  7.4069

------------------------------------------------------------------------------
     sbp |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     age |   1.212715   .3238192      3.745   0.001        .549401    1.876028
     smk |   9.945568   2.656057      3.744   0.001       4.504882    15.38625
    quet |   8.592448   4.498681      1.910   0.066      -.6226827    17.80758
   _cons |   45.10319   10.76488      4.190   0.000       23.05235    67.15404
------------------------------------------------------------------------------
```

**Then**, complete the following:

A. Use model 3 for the following:

(1) What is the predicted SBP for a 50-year old smoker with a quetelet (QUET) index of 3.5?

The regression equation for the third model is

$$\hat{y} = \widehat{\beta_0} + \widehat{\beta_1}\, age + \widehat{\beta_2}\, smk + \widehat{\beta_3}\, quet$$

substituting the different value of the predictor variables, we get the predicted value of SBP

for AGE=50, SMK=1 and QUET=3.5

y=45.103+1.2127(50)+9.945568(1)+8.592448(3.5)

= 145.76

(2) What is the predicted SBP for a 50-year-old non-smoker with a quetelet index of 3.5?

Using the same regression equation again

$$\hat{y} = \widehat{\beta_0} + \widehat{\beta_1}\, age + \widehat{\beta_2}\, smk + \widehat{\beta_3}\, quet$$

and substituting the value of the predictors, we get the new predicted value of the SBP

for AGE=50, SMK=0 and QUET=3.5

y=45.103+1.2127(50)+8.592448(3.5)

= 135.81

(3) For 50-year-old smokers, give an estimate of the change in SBP corresponding to an increase in quetelet index from 3.0 to 3.5.

Making use of the regression equation $\hat{y} = \widehat{\beta_0} + \widehat{\beta_1}\, age + \widehat{\beta_2}\, smk + \widehat{\beta_3}\, quet$ again and just changing the value of QUET from 3.5 to 3.0 from the previous question, we will get the predicted value of the SBP. We will then make use of this and take the difference between the previous value and the present value of the predicted SBP to obtain the change due to increase in QUET.

for AGE=50, SMK=1 and QUET=3.0

y=45.103+1.2127(50)+9.945568(1)+8.592448(3.0)

= 141.46

Thus the change is SBP = 145.76 – 141.46 = 4.30

B. Using the ANOVA tables, compute and compare the $R^2$-values for models 1, 2, and 3.

Now we will make use of the three outputs of the regression models that we had obtained earlier. Make sure that you match the correct output to the models. The value of $R^2$ for each of the model can be obtained from the fourth line of the right hand side (RHS) of the outputs.

| Model | Independent Variables Used | $R^2$ |
|---|---|---|
| 1 | AGE $(X_1)$ | 0.6009 |
| 2 | AGE $(X_1)$, SMK $(X_2)$, | 0.7298 |
| 3 | AGE $(X_1)$, SMK $(X_2)$, QUET $(X_3)$ | 0.7609 |

The $R^2$ value increases with each variable added to the model.

C. Conduct (separately) the overall *F* tests for significant regression under models 1,2, and 3. Be sure to state your null hypothesis for each model in terms of regression coefficients.

In order to conduct the overall F test, we essentially check to see the null hypothesis that the slope coefficients simultaneously equal to zero. The value of the F statistic for each model can be obtained from the second line of the RHS of the output. The corresponding p-value can be obtained from the third line of the RHS of the output.

| Model | Independent Variables Used | $H_0$ | F | p |
|---|---|---|---|---|
| 1 | AGE $(X_1)$ | $\beta_{age}=0$ | 45.18 | $<0.001$ |
| 2 | AGE $(X_1)$, SMK $(X_2)$, | $\beta_{age}=\beta_{smk}=0$ | 39.16 | $<0.001$ |
| 3 | AGE $(X_1)$, SMK $(X_2)$, QUET $(X_3)$ | $\beta_{age}=\beta_{smk}=\beta_{quet}=0$ | 29.71 | $<0.001$ |