Homework Assignments
Applied Regression Analysis

WEEK 2

The table gives the systolic blood pressure (SBP), body size (QUET), age (AGE), and smoking history (SMK = 0 if nonsmoker, SMK = 1 if a current or previous smoker) for a hypothetical sample of 32 white males over 40 years old from the town of Angina.

Person	SBP	QUET	AGE	SMK
1	135	2.876	45	0
2	122	3.251	41	0
3	130	3.100	49	0
4	148	3.768	52	0
5	146	2.979	54	1
6	129	2.790	47	1
7	162	3.668	60	1
8	160	3.612	48	1
9	144	2.368	44	1
10	180	4.637	64	1
11	166	3.877	59	1
12	138	4.032	51	1
13	152	4.116	64	0
14	138	3.673	56	0
15	140	3.562	54	1
16	134	2.998	50	1
17	145	3.360	49	1
18	142	3.024	46	1
19	135	3.171	57	0
20	142	3.401	56	0
21	150	3.628	56	1
22	144	3.751	58	0
23	137	3.296	53	0
24	132	3.210	50	0
25	149	3.301	54	1
26	132	3.017	48	1
27	120	2.789	43	0
28	126	2.956	43	1
29	161	3.800	63	0
30	170	4.132	63	1
31	152	3.962	62	0
32	164	4.010	65	0

Exercise One

Generate scatter diagrams for each of the following variable pairs:

1. SPB *(Y)* vs. QUET *(X)*
2. SBP *(Y)* vs. SMK *(X)*
3. QUET *(Y)* vs. AGE *(X)*
4. SBP *(Y)* vs. AGE *(X)*

Exercise Two

Sketch a line

Using scatter diagrams #1, #3, and #4 that you generated above, use paper and pencil to roughly sketch a line that fits the data reasonably well. Use the homework forum to share your sketches and comment on the relationships described.

Exercise Three

Comparing Blood Pressure with Smoking History

1. Determine the least-squares estimates of slope (β_1) and intercept (β_0) for the straight-line regression of SBP *(Y)* on SMK *(X)*.
2. Compare the value of $\hat{\beta}_0$ with the mean SBP for nonsmokers. Compare the value of $\hat{\beta}_0 + \hat{\beta}_1$ with the mean SBP for smokers. Explain the results of these comparisons.
3. Test the hypothesis that the true slope (β_1) is 0.
4. Is the test in part (e) equivalent to the usual two-sample *t* test for the equality of two population means assuming equal but unknown variances? Demonstrate your answer numerically.

Exercise Four

Comparing Blood Pressure with Body Size

1. Determine the least-squares estimates of slope and intercept for the straight-line regression of SBP *(Y)* on QUET *(X)*.
2. Sketch the estimated regression line on the scatter diagram involving SBP and QUET.
3. Test the hypothesis of zero slope.
4. Find a 95% confidence interval for $\mu_{y|\bar{x}}$.
5. Calculate 95% prediction bands.
6. Based on the above, would you conclude that blood pressure increases as body size increases?
7. Are any of the assumptions for straight-line regression clearly not satisfied in this example?

Week Two Homework Solutions

Help Center

If you need help and answers to the exercises of week two, please click below on the selected exercise:

- **Exercise One**
- **Exercise Two**
- **Exercise Three**
- **Exercise Four**

Click the button below to return to this week's homework exercise.

←Homework Page

Created Tue 10 Mar 2015 8:53 AM PDT

Last Modified Mon 30 Mar 2015 10:03 AM PDT

Homework Solutions
Applied Regression Analysis

WEEK 2

Exercise One

Generate scatter diagrams for each of the following variable pairs:

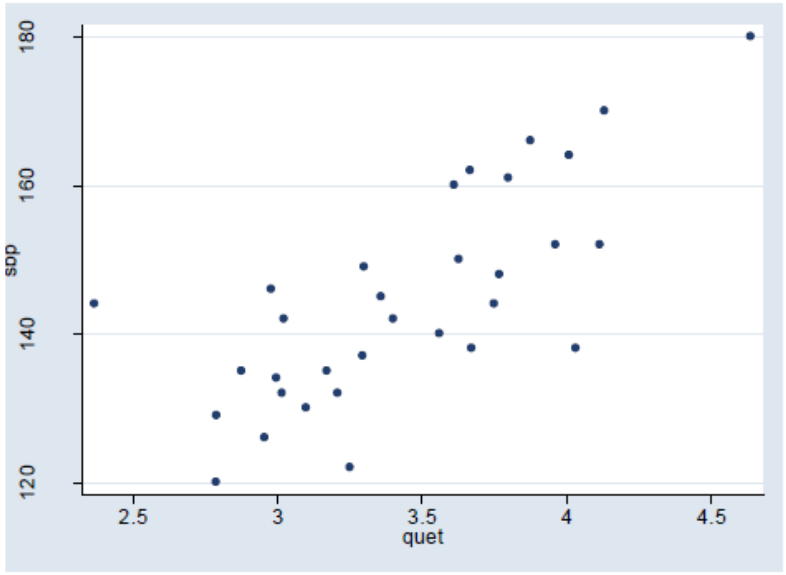
1. **SBP *(Y)* vs. QUET *(X)***

   In the command window, enter '.scatter sbp quet'

   This will produce the scatterplot of SBP (Y) and QUET (X). The resulting scatterplot displayed should resemble the screenshot depicted below

   (Note: Stata is case-sensitive so the variables should be lower-case if that is how they are coded in the dataset).

```
.scatter sbp quet
```

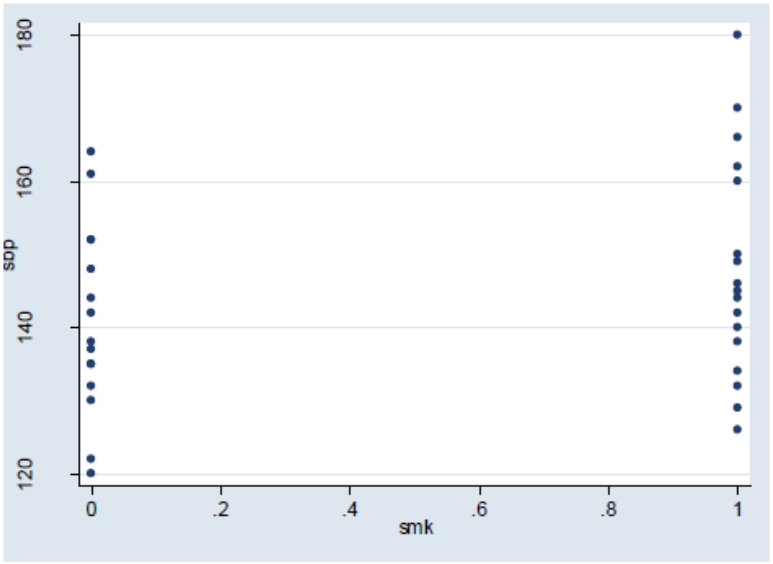


2. SBP *(Y)* vs. SMK *(X)*

In the command window, enter '.scatter sbp smk'.

This will produce the scatterplot of SBP (Y) and SMK (X). The resulting scatterplot displayed should resemble the screenshot depicted below.

```
. scatter sbp smk
```



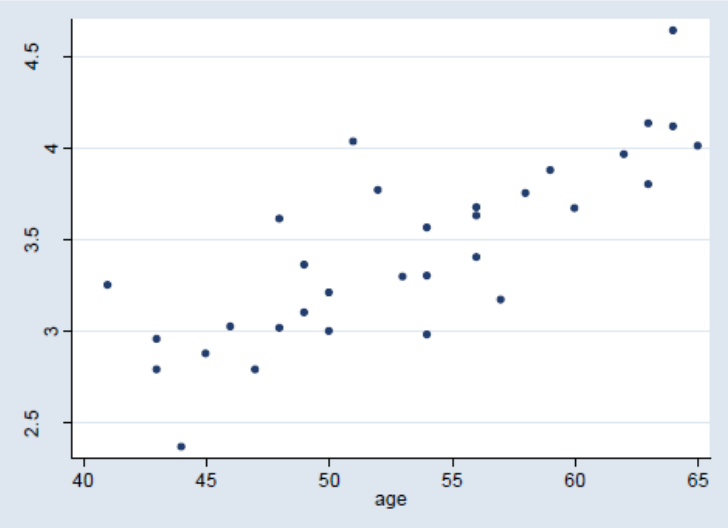
Note: Smoking is a binary variable, therefore we should not expect to see an even scatter of observations as with the other plots containing continuous variables.

3. QUET *(Y)* vs. AGE *(X)*

In the command window, enter '.scatter quet age'.

This will produce the scatterplot of QUET (Y) and AGE (X). The resulting scatterplot displayed should resemble the screenshot depicted below.

```
. scatter quet age
```

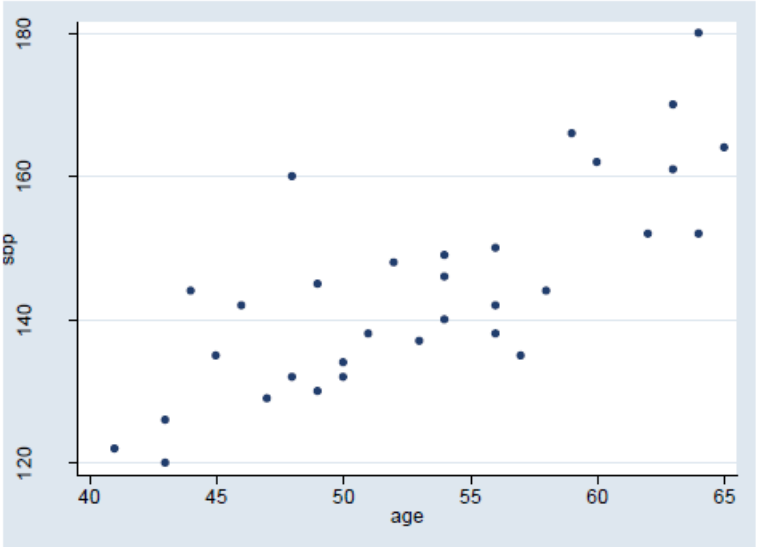


4. SBP *(Y)* vs. AGE *(X)*

In the command window, enter '.scatter sbp age'.

This will produce the scatterplot of SBP (Y) and AGE (X). The resulting scatterplot displayed should resemble the screenshot depicted below.

```
. scatter sbp age
```



Homework Solutions
Applied Regression Analysis

WEEK 2

Exercise Two

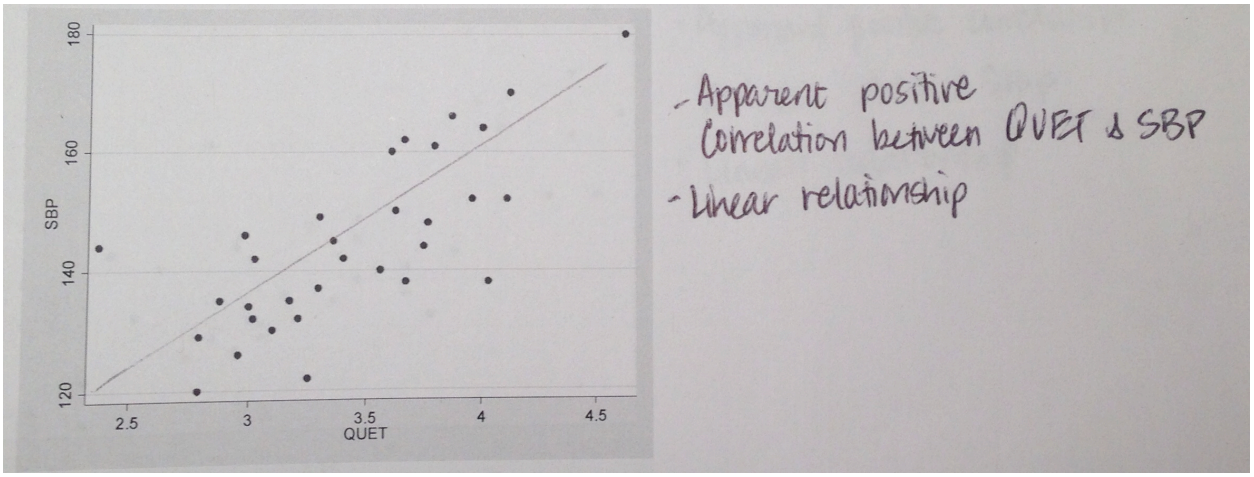
Sketch a line

Using scatter diagrams #1, #3, and #4 that you generated in exercise one, use paper and pencil to roughly sketch a line that fits the data reasonably well.

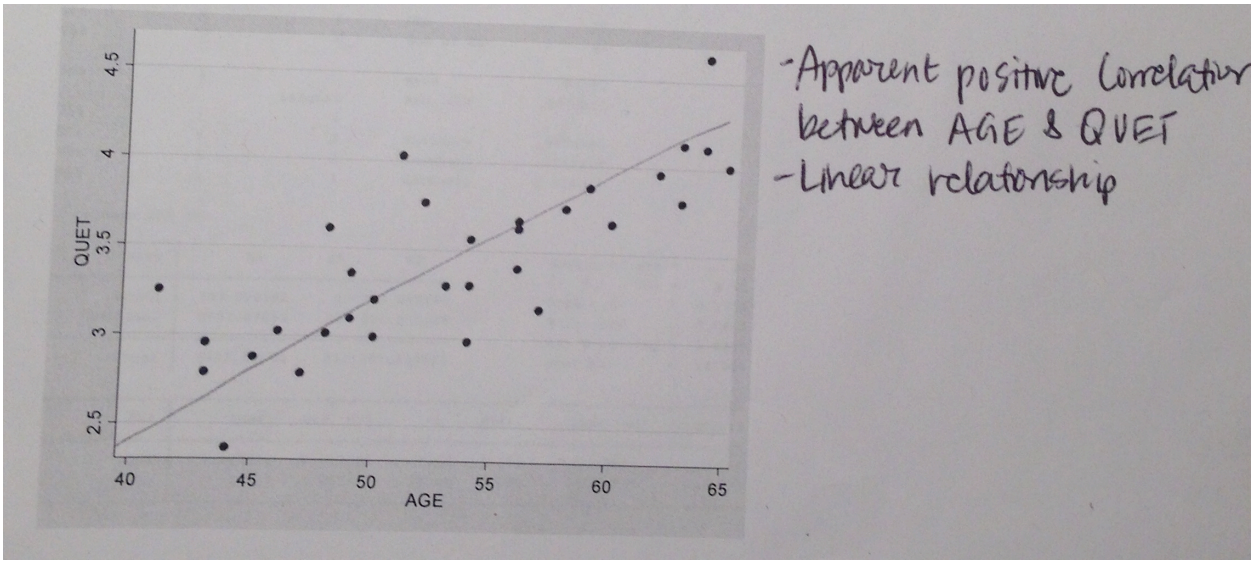
Use the **homework forum** to share your sketches and comment on the relationships described.

Below are examples of how this could look when you hand-draw a line.

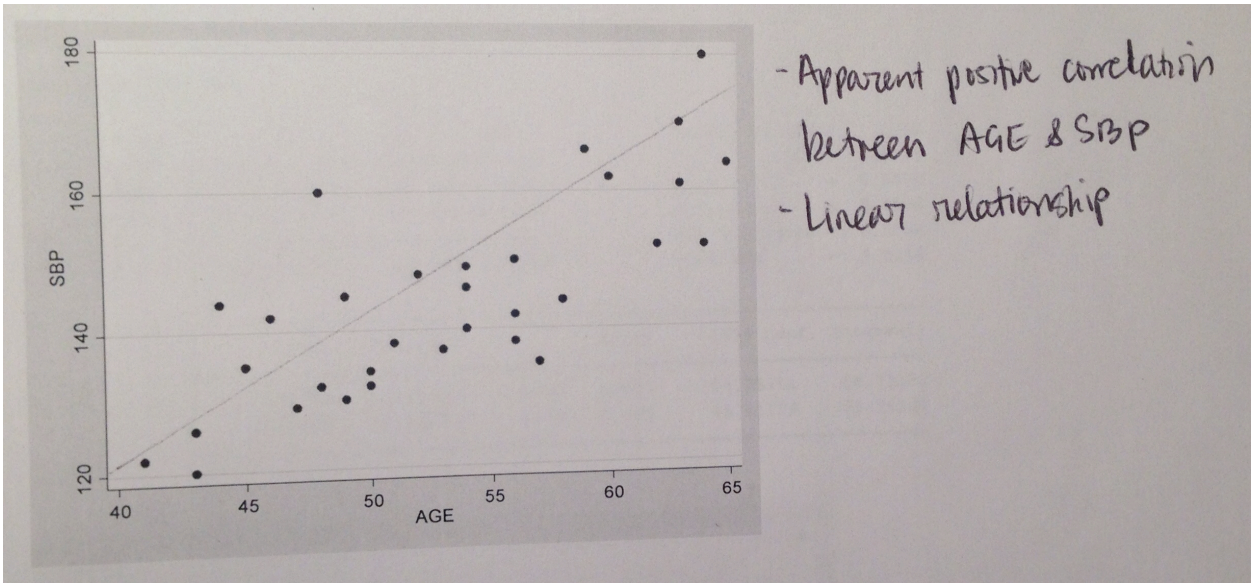
#1 – SBP vs. QUET



#3 – QUET vs. AGE



#4 – SBP vs. AGE


SBP
180
160
140
120
40
45
50
55
60
65
AGE
- Apparent positive correlation between AGE & SBP
- Linear relationship

Homework Solutions
Applied Regression Analysis

Exercise Three
Comparing Blood Pressure with Smoking History

1. **Determine the least-squares estimates of slope (β_1) and intercept (β_0) for the straight-line regression of SBP *(Y)* on SMK *(X)*.**

   We can determine the least squares estimates for the parameters in simple linear regression by regressing Y on X. In the command window, enter '.regress sbp smk'. This will produce the output below.

```
. regress sbp smk

  Source |       SS       df       MS                  Number of obs =      32
---------+------------------------------                F(  1,    30) =    1.95
   Model |  393.098162     1  393.098162                Prob > F      =  0.1723
Residual |  6032.87059    30  201.095686                R-squared     =  0.0612
---------+------------------------------                Adj R-squared =  0.0299
   Total |  6425.96875    31  207.289315                Root MSE      =  14.181

------------------------------------------------------------------------------
     sbp |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
     smk |   7.023529   5.023498      1.398   0.172      -3.235823    17.28288
   _cons |      140.8   3.661472     38.454   0.000       133.3223    148.2777
```

$$y = \beta_0 + \beta_1 x$$

$$= 140.8 + 7.02(\text{smk})$$

Note: When entering data into STATA, always list the dependant variable first (sbp, in this case) and then the independent variable (smk, in this case).

The slope for smk is determined by the value listed in the "Coef." column of the table above, and the value for the intercept is determined in a similar fashion in the _cons row.

How can you interpret the slope?

"Under this model, current or previous smokers have an average systolic blood pressure 7.02 mm Hg higher than that of non-smokers"

2. **Compare the value of $\hat{\beta}_0$ with the mean SBP for nonsmokers. Compare the value of $\hat{\beta}_0 + \hat{\beta}_1$ with the mean SBP for smokers. Explain the results of these comparisons.**

   To compare the mean of one variable across different categories of another variable in STATA, you must first sort the data by the second categorizing variable (in this case, smk).

   In the command window, enter '.sort smk'.

   You much then use the 'sum' command to get descriptive statistics on your variable of interest (in this case, sbp), but first you must use the 'by' command to split the results by smoking status.

   In the command window, enter '.by smk:sum sbp'.

   This will produce the output below.

```
. sort smk
. by smk:sum sbp

-------------------------------------------------------------------------------
-> smk = 0

    Variable |     Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         sbp |      15       140.8    12.90183       120        164

-------------------------------------------------------------------------------
-> smk = 1

    Variable |     Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         sbp |      17    147.8235    15.21198       126        180
```

The mean value of SBP for nonsmokers (140.8) is equal to the value of $\hat{\beta}_0$. The mean value of SBP for smokers is 147.82 which is equal to $\hat{\beta}_0 + \hat{\beta}_1$.

In simple linear regression, the intercept can be interpreted as the value for Y when X=0. Given that smoking is a binary variable, and is coded (0,1) (0 for non-smokers, 1 for smokers), then the intercept is the mean value for non-smokers (Y when X=0), and the intercept plus the slope is the mean value for smokers (Y when X=1).

3. **Test the hypothesis that the true slope (β_1) is 0.**

$H_o : \beta_1 = 0$
$H_A : \beta_1 \neq 0$

The null hypothesis cannot be rejected, $p = 0.172$. There is not sufficient evidence to conclude that the slope is significantly different from 0.

Note: You can test for the significance of the slope by looking at the p-value for the t-test in the table for the regression in problem 1. The p-value tells us that the probability of rejecting the null when the null is true is 17.2%, which exceeds 5%. Therefore there is insufficient evidence to reject the null.

4. **Is the test in part (3) equivalent to the usual two-sample *t* test for the equality of two population means assuming equal but unknown variances? Demonstrate your answer numerically.**

   To perform a t-test to compare the mean sbp across the populations in the different smoking categories, enter '.ttest sbp, by(smk)' into the command window.

   This will produce the output below.

```
. ttest sbp, by(smk)

Two-sample t test with equal variances

------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
       0 |      15       140.8    3.331237    12.90183    133.6552    147.9448
       1 |      17    147.8235    3.689448    15.21198    140.0022    155.6448
---------+--------------------------------------------------------------------
combined |      32    144.5313    2.545151    14.39755    139.3404    149.7221
---------+--------------------------------------------------------------------
    diff |           -7.023529    5.023498               -17.28288    3.235823
------------------------------------------------------------------------------
Degrees of freedom: 30

                    Ho: mean(0) - mean(1) = diff = 0

     Ha: diff < 0                 Ha: diff ~= 0                 Ha: diff > 0
       t =  -1.3981                 t =  -1.3981                  t =  -1.3981
   P < t =   0.0862           P > |t| =   0.1723           P > t =   0.9138
```

The t-test gives the same t-value and p-value as the test for the hypothesis that the true slope, β_1, is 0.

The p-value for question 3 is the same at the two-sided p-value for the two-sample t-test. In both tests, you are testing whether smoking has a significant impact on systolic blood pressure by determining if the sbp for smokers is significantly different than that of non-smokers.

Homework Solutions
Applied Regression Analysis

WEEK 2

Exercise Four

1. **Determine the least-squares estimates of slope and intercept for the straight-line regression of SBP *(Y)* on QUET *(X)*.**

   We can determine the least squares estimates for the parameters in simple linear regression by regressing Y on X.

   In the command window, enter '.regress sbp quet'.

   This will produce the output below.

```
. regress sbp quet

  Source |       SS       df       MS                  Number of obs =      32
---------+------------------------------                F(  1,    30) =   36.75
   Model |  3537.94585     1  3537.94585                Prob > F      =  0.0000
Residual |   2888.0229    30  96.2674299                R-squared     =  0.5506
---------+------------------------------                Adj R-squared =  0.5356
   Total |  6425.96875    31  207.289315                Root MSE      =  9.8116

------------------------------------------------------------------------------
     sbp |      Coef.   Std. Err.       t     P>|t|       [95% Conf. Interval]
---------+--------------------------------------------------------------------
    quet |   21.49167   3.545147      6.062   0.000       14.25151    28.73182
   _cons |   70.57641   12.32187      5.728   0.000        45.4118    95.74102
------------------------------------------------------------------------------
```

$$\hat{\beta}_0 = 70.576$$

$$\hat{\beta}_1 = 21.492$$

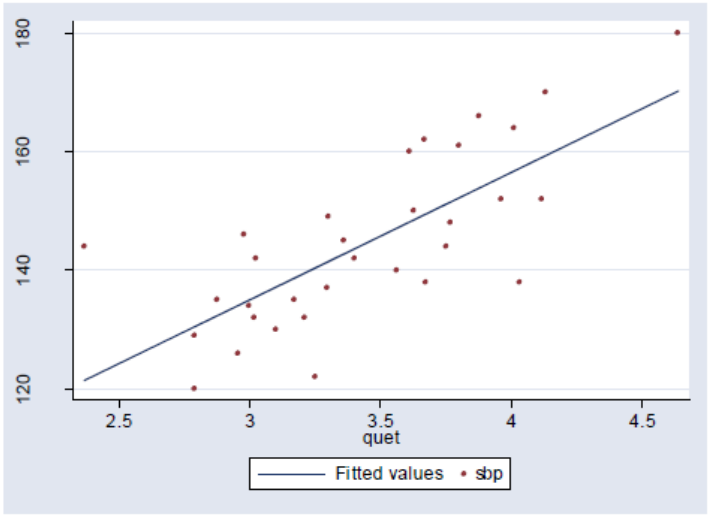
2. **Sketch the estimated regression line on the scatter diagram involving SBP and QUET.**

   In order to fit a regression line in STATA, you must first create a new variable in your dataset for of the predicted y given x under the regression model.

   You can do this simply by entering 'predict yhat' into the command window. Next, create a scatterplot with a line by entering 'scatter yhat sbp quet, c(1 .) s(i o)' into the command window.

   The commands 'c(1 .) s(i o)' specify that the yhat should be labeled with a line and data points with dots, respectively.

```
. predict yhat
(option xb assumed; fitted values)

. scatter yhat sbp quet, c(l .) s(i o)
```



3. **Test the hypothesis of zero slope.**

$$H_o : \beta_1 = 0$$
$$H_A : \beta_1 \neq 0$$

Reject the null hypothesis, $p < 0.001$. There is sufficient evidence to conclude that the slope is significantly different from 0.

Note: You can test for the significance of the slope by looking at the p-value for the t-test in the table for the regression in problem 1. The p-value tells us that the probability of rejecting the null when the null is true is less than 5%. Therefore there is sufficient evidence to reject the null.

4. **Find a 95% confidence interval for $\mu_{y|\bar{x}}$.**

   To calculate confidence intervals, you need to know the descriptive statistics for the variables, including their mean values and standard deviations.

   To get these values, use the 'sum' command by entering '.sum sbp quet age smk' into the command window.

   Next we can calculate $\mu_{y|\bar{x}}$ by entering the mean value for quet within the regression equation using our previously estimated parameters. The confidence limits about $\mu_{y|\bar{x}}$ can then be estimated using the mean value and standard deviation of x.

```
. sum sbp quet age smk

    Variable |     Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         sbp |      32    144.5313    14.39755       120        180
       quet |      32    3.441094    .4970781     2.368      4.637
         age |      32       53.25    6.956083        41         65
         smk |      32     .53125    .5070073         0          1
```

$$\hat{y}_{\bar{x}} = 70.57641 + 21.49167 * 3.44 = 144.508$$

$$s^2_{\hat{y}_{x_0}} = s^2_{y|x}\left(\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{(n-1)s_x^2}\right)$$

$$s^2_{\hat{y}_{\bar{x}}} = s^2_{y|\bar{x}}\left(\frac{1}{n}\right) = \frac{96.2674299}{32} = 3.008357$$

$$s_{\hat{y}_{\bar{x}}} = \sqrt{3.008357} = 1.7344616$$

$$95\%\ \text{CI: } \hat{y}_{\bar{x}} \pm t_{.975}(30)\, s_{\hat{y}_{\bar{x}}} = 144.508 \pm 2.042 \times 1.7344616 = (140.97, 148.05)$$

Interpretation: We are 95% confident that the true value for the mean value of y is between 140.97 and 148.05 mm Hg.

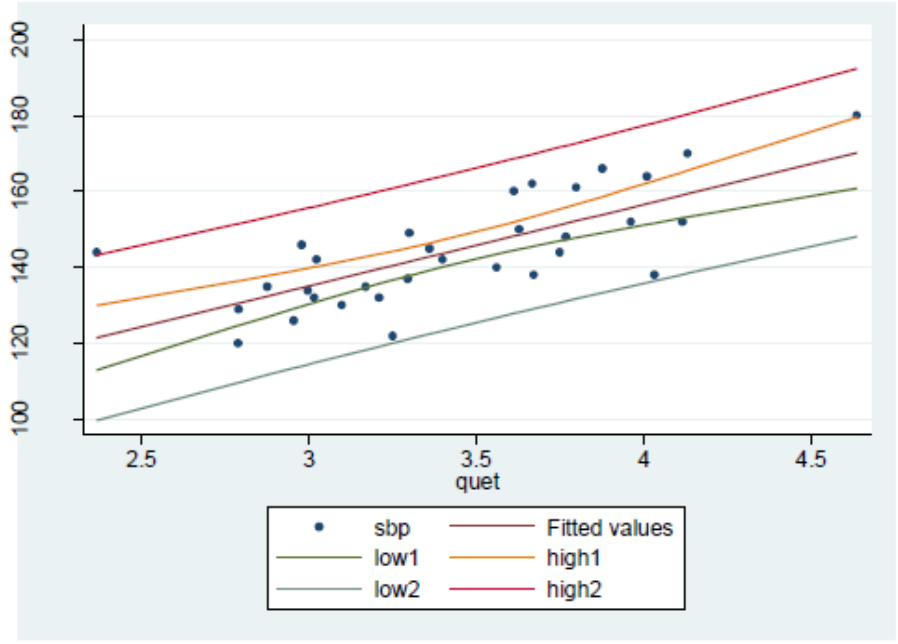
5. Calculate 95% prediction bands.

For this problem, we are asking for a plot of the prediction bands using STATA, not for hand-calculations. To do this, we must enter 'predict sepred, stdf' into the command window to generate a variable- 'sepred'- for the standard deviation used within the prediction interval.

Next, we can calculate the value for the lower limit of the prediction interval by entering 'generate low=yhat-invttail(30,0.025)*sepred' and the upper limit of the prediction interval by entering 'generate high=yhat+invttail(30,0.025)*sepred' (note: invttail(30,0.025)= $t_{0.975}(30)$).

From here you can create a plot of the prediction intervals with the regression line by entering 'scatter sbp yhat low high, sort connect (. 1 1 1) symbol (o i i i)'. The code and plot below includes both the confidence and prediction intervals, however you only need to graph the prediction intervals for this question.

```
. predict yhat
(option xb assumed; fitted values)

. predict seyhat, stdp

. display invttail(30,0.025)
2.0422724

. generate low1= yhat-2.0422724* seyhat

. generate high1= yhat+2.0422724* seyhat

. predict sepred, stdf

. generate low2= yhat-invttail(30,0.025)* sepred

. generate high2= yhat+invttail(30,0.025)* sepred

. scatter sbp yhat low1 high1 low2 high2 quet,sort connect(. l l l l l)
symbol(o i i i i i)
```



6. **Based on the above, would you conclude that blood pressure increases as body size increases?**

   Yes, because the fitted regression line, as well at the confidence and prediction band, appear to have an upward slope.

7. **Are any of the assumptions for straight-line regression clearly not satisfied in this example?**

Simple Linear Regression Assumptions:

Linearity: SBP and SMK appear to be linearly related based on the above scatterplot
Independence: The study design does not suggest that the observations are not independent
Normality: The variables appear to be normally distributed (there are no significant outliers)
Equal Variance (homoscedasticity): The variances along the regression line appear to remain similar as you move across the line

There are no apparent violations of homoscedasticity, normality, or independence. Formal tests of these assumptions are possible but are not included here.

All results for: Week2 Homework

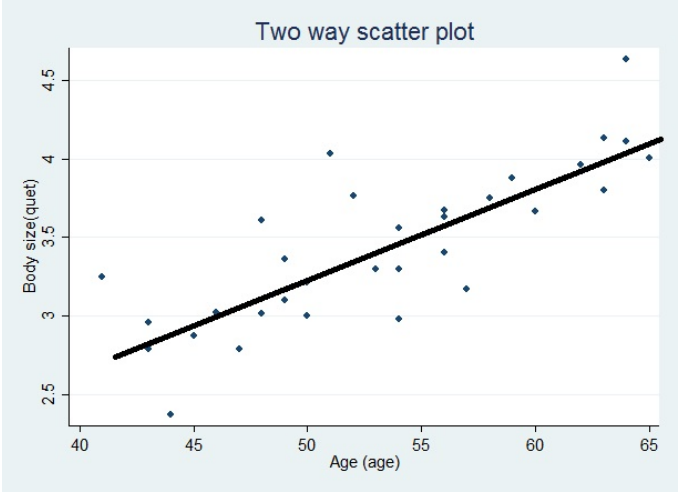
Week2 Homework | Search All Forums

Your search returned approximately **41** results

Ganesh Sharma · 10 days ago Posted in Week 2 homework exercise two

Week 2 homework exercise two

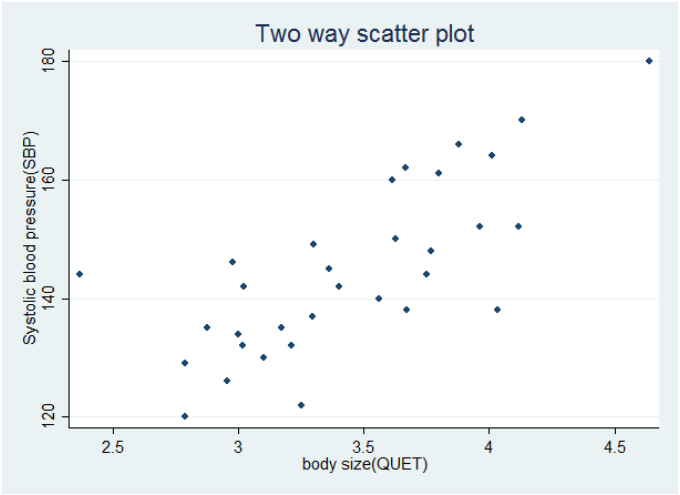
I used microsoft paint to draw lines in the scatter plot ...

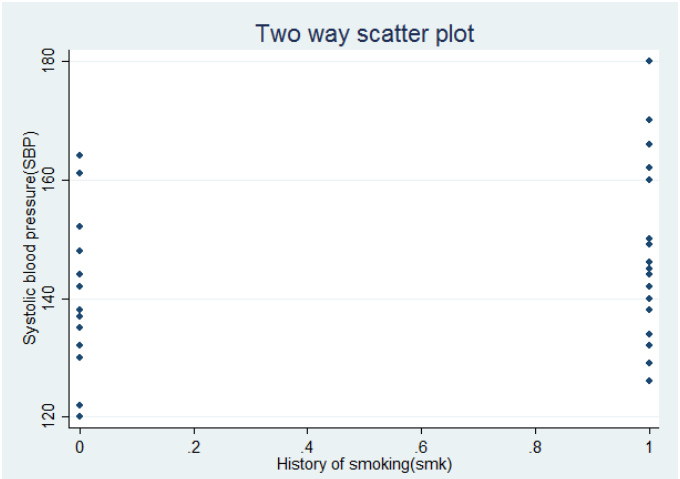


Ganesh Sharma · 10 days ago Posted in Week 2 homework exercise one...

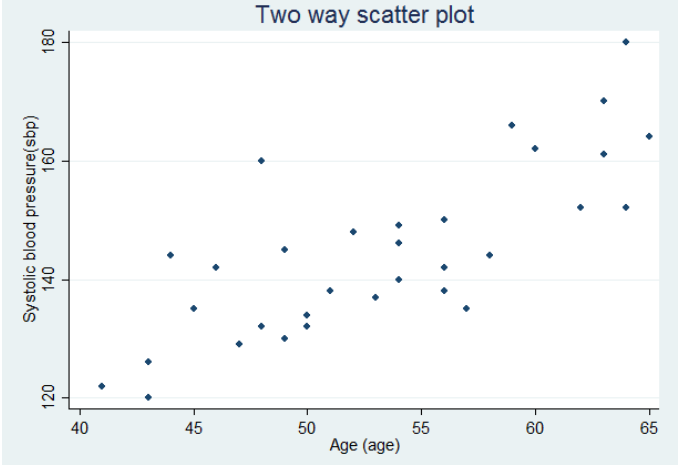
Dear all,

Week 2 homework exercise one..


Two way scatter plot
Systolic blood pressure(SBP)
120
140
160
180
2.5
3
3.5
4
4.5
body size(QUET)

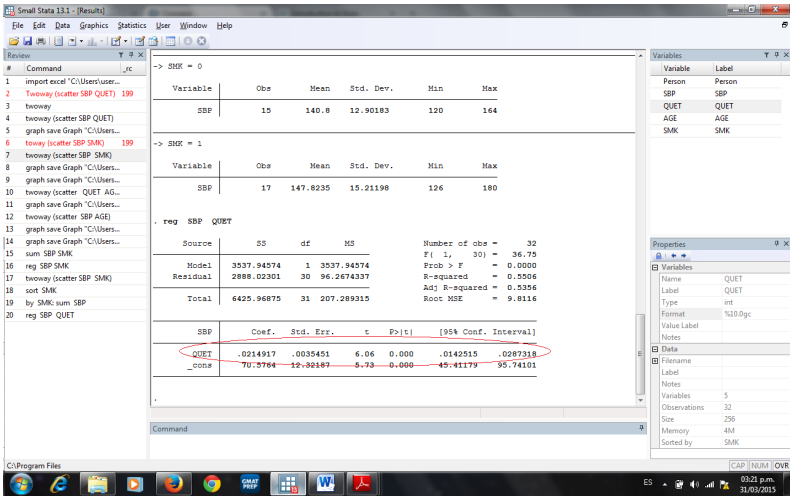

Two way scatter plot
Systolic blood pressure(SBP)
120
140
160
180
0
.2
.4
.6
.8
1
History of smoking(smk)


Two way scatter plot
Body size(quet)
2.5
3
3.5
4
4.5
40
45
50
55
60
65
Age (age)

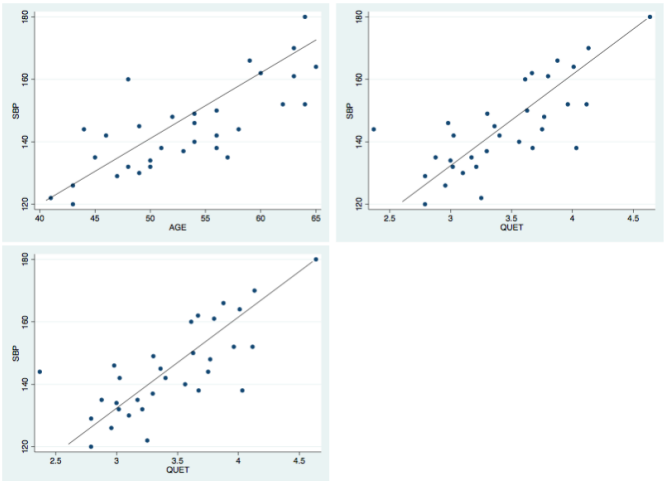

Two way scatter plot
Systolic blood pressure(sbp)
120
140
160
180
40
45
50
55
60
65
Age (age)

David Alejandro Ramirez Palacios · 13 days ago Posted in Week 2 HW Exercise 4

Hi, everyone.
I need a help with excercise 4 because I got a diffent answer in the excercise 4 that Homework solutions. The slope that I calculated was 0.0214917 , but Homework solutions have 21.4917.
Thanks



David C. Morris · 11 days ago Posted in Week Two Homework - Exercise 1



Positive and linear relationships.

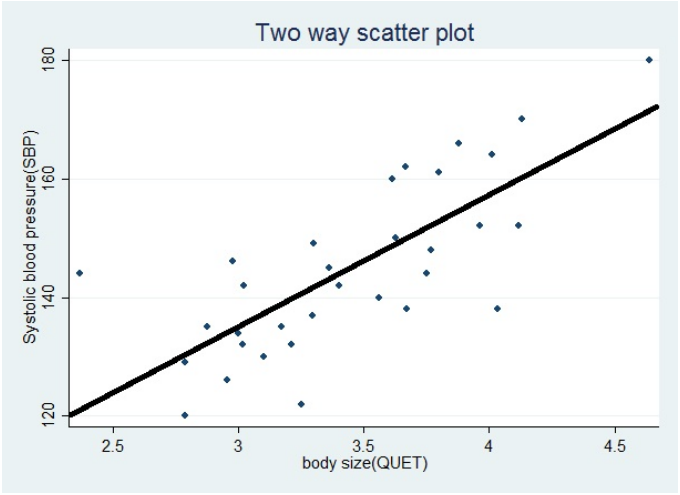
Thomas Evans STAFF · 7 days ago Posted in Different data for lectures & homework?

The TAs and professor can correct me if I'm wrong, but I believe that the dataset used in the homework is a bit different from the dataset demonstrated on video 2.1.

The video was demonstrating the process and then the homework had you practice the process in a similar style problem with different numbers. The differing data would most likely produce different plots.

Professor Lemeshow explains a bit about the week 2 homework in this **homework highlights video**.

Ganesh Sharma · 10 days ago Posted in Week 2 homework exercise two



Thomas Evans STAFF · 14 days ago Posted in homework data

Hmm... weird... I was able to copy/paste from Chrome to Excel (Mac) and it transferred properly. Regardless, I've made a CSV of the data and posted it as a download on the **Week Two Homework** page.

Hope this helps.

Cheers,

Tom

Ganesh Sharma · 7 days ago Posted in Week 2 homework exercise four.

Question 1. · Determine the least-squares estimates of slope and intercept for the straight-line regression of SBP (Y) on QUET (X).

Intercept(B0) = 70.57641

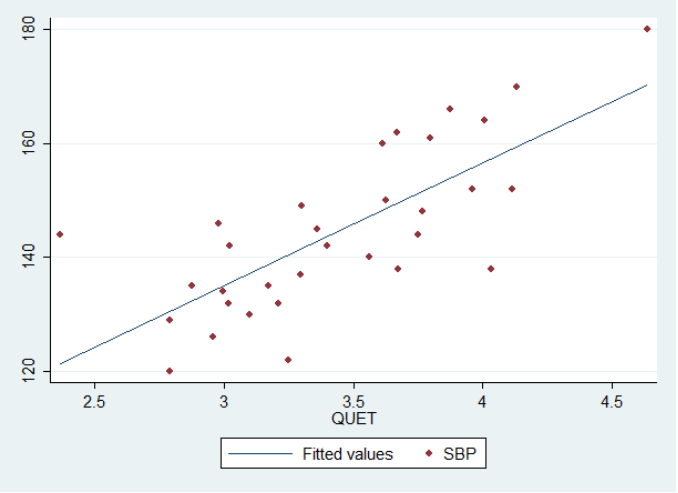
Slope(B1) = 21.49167

regress sbp quet

```
      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  1,    30) =   36.75
       Model |  3537.94585     1  3537.94585            Prob > F      =  0.0000
    Residual |   2888.0229    30  96.2674299            R-squared     =  0.5506
-------------+------------------------------           Adj R-squared =  0.5356
       Total |  6425.96875    31  207.289315            Root MSE      =  9.8116

------------------------------------------------------------------------------
         sbp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        quet |   21.49167   3.545147     6.06   0.000     14.25151    28.73182
       _cons |   70.57641   12.32187     5.73   0.000      45.4118    95.74102
------------------------------------------------------------------------------
```

Question 2

Sketch the estimated regression line on the scatter diagram involving SBP and QUET.



Question 3

We can reject the null hypothesis and accept the alternative hypothesis because p value is less than .05(i.e .000). It means there is significant relationship between body size and systolic blood pressure.

Question 4

Find a 95% confidence interval for $\mu y|\bar{x}$

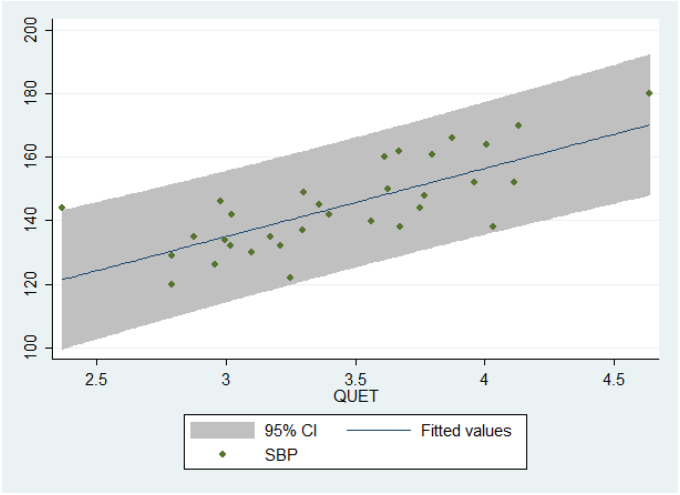
The confidence level for the equation is 140.5073, 148.5552. We are 95% sure that the true value lies between 140.5073, 148.5552.

ci sbp quet age smk yhat

Variable	Obs	Mean	Std. Err.	[95% Conf.	Interval]
sbp	32	144.5313	.545151	139.3404	149.7221
quet	32	3.441094	.0878718	3.261878	3.62031
age	32	53.25	1.229673	50.74206	55.75794
smk	32	.53125	.0896271	.3484544	.7140456
yhat	32	144.5312	1.973011	140.5073	148.5552

Question 5

Calculate 95% prediction bands



Question 6

Based on the above, would you conclude that blood pressure increases as body size increases

We are 95% confident that the systolic blood pressure (sbp) goes up when the body size(quet) increases. The regression including the shaded area goes upward right giving evidence to support the the statement.

Question 7. Are any of the assumptions for straight-line regression clearly not satisfied in this example

There are a lot of assumptions that still are not satisfied by the example prior to doing regression analysis. But in my previous submission exercise I mentioned quiet few of them. The assumptions that need to fulfill are independence, normality, linearity and homoedasticity. Besides that multi collinearity test is also required to skip insignificant variable/s in an attempt to find the causing factor limited to these datasets.

Ganesh Sharma · 8 days ago Posted in Week 2 homework exercise three

Question 1. Determine the least-squares estimates of slope $\beta 1$ and intercept $\beta 0$ for the straight-line regression of SBP (Y) on SMK (X).

The histogram check showed the SBP has almost in normal distribution pattern. Besides sapiro-wilk test and sapiro francia showed p>.05, indicating normality. Variance was same for SBP for any SMK(homoscedasticity). So, regression is applicable for the situation.

Intercept ($\beta 0$) = 140.8

Slope (*β1*) = 7.023529

regress sbp smk

```
  Source |       SS           df       MS              Number of obs =      32
-------------+------------------------------           F(  1,    30) =    1.95
   Model |  393.098162     1      393.098162           Prob > F      =  0.1723
Residual | 6032.87059   30  201.095686           R-squared     =  0.0612
-------------+------------------------------           Adj R-squared =  0.0299
   Total |  6425.96875    31  207.289315           Root MSE      =  14.181

------------------------------------------------------------------------------
     sbp |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     smk |   7.023529   5.023498     1.40   0.172    -3.235823    17.28288
   _cons |      140.8   3.661472    38.45   0.000     133.3223    148.2777
```

There is no significant difference(confidence interval – 3.235823, 17.28288) in SBP between those with and without smoking history.

Question 2.
Question 2. Compare the value of *β*^0 with the mean SBP for nonsmokers. Compare the value of *β*^0+*β*^1 with the mean SBP for smokers.

```
. sort smk
  by smk: sum sbp

-------------------------------------------------------------------------------------------------
-> smk = 0

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         sbp |        15       140.8    12.90183       120        164
-------------------------------------------------------------------------------------------------
-> smk = 1

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         sbp |        17    147.8235    15.21198       126        180
```

The mean of non-smokers (coded as 0) is 140.8 as against smokers (coded as 1) is 147.8235. As non-smoker is coded as 0, it becomes base category, as intercept(because multiplying the coefficient with 0 becomes 0) and non-smoker includes intercept and slope(1)

While fitting the value of smoking history in straight line equation

i) If smoking = 0

Y = 140.8 + 7.023529*x = 140.8 + 7.023529 * 0 = 140.8

ii) If smoking = 1

Y = 140.8 + 7.023529*x = 140.8 + 7.023529 * 1 = 147.8

Question 3. Test the hypothesis that the true slope B1 is 0 .

Null hypothesis B1 = 0 (there is no linear relationship between SBP(Y) and SMK(X))

Alternate hypothesis B2 ≠ 0 (There is linear relationship between SBP(Y) and SMK(X))

There is no linear relationship (p = 0.172) between SBP(Y) and SMK(X)). I think this is non-directional (two sided) hypothesis checking in both positive and negative linear relationship.

Question 4. Is the test in question (3) equivalent to the usual two-sample t test for the equality of two population means assuming equal but unknown variances?

ttest sbp, by (smk)

```
   Two-sample t test with equal variances
    ------------------------------------------------------------------------------
 Group |     Obs    Mean    Std. Err.       Std. Dev.    [95% Conf. Interval]
    ---------+--------------------------------------------------------------------
    0 |      15       140.8       3.331237    12.90183    133.6552    147.9448
    1 |      17     147.8235    3.689448    15.21198    140.0022    155.6448
    ---------+--------------------------------------------------------------------
combined |32    144.5313    2.545151    14.39755    139.3404    149.7221
    ---------+--------------------------------------------------------------------
      diff |          -7.023529    5.023498            -17.28288    3.235823
    ------------------------------------------------------------------------------
  diff = mean(0) - mean(1)                                      t =  -1.3981
    Ho: diff = 0                     degrees of freedom =       30

     Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
     Pr(T < t) = 0.0862         Pr(|T| > |t|) = 0.1723          Pr(T > t) = 0.9138
```

think it is equivalent. Because, the hypothesis mentioned in question 3 is linear relationship. Linear relationship generated numerical values like 140.8 and 147.8.

In independent (two sample) t-test, the hypothesis is expressed like; There is no difference (means) in SBP between smoker and non-smokers (NULL) and there is significant difference as Alternate hypothesis.

Test in question (3) has one categorical and another numerical variable and that is the same for T-test.

Only the output formats are different. Regression covers more with % of variability explained by independent variable in dependent variable and shows more about residual sum of squares and model sum of squares. For t-test, it calculates mean difference between two groups.

Two sample test with equal variance

Pr(|T| > |t|) = 0.1723

Mario R. Melchiori · 13 days ago Posted in How to import data relavent to this course in Stata

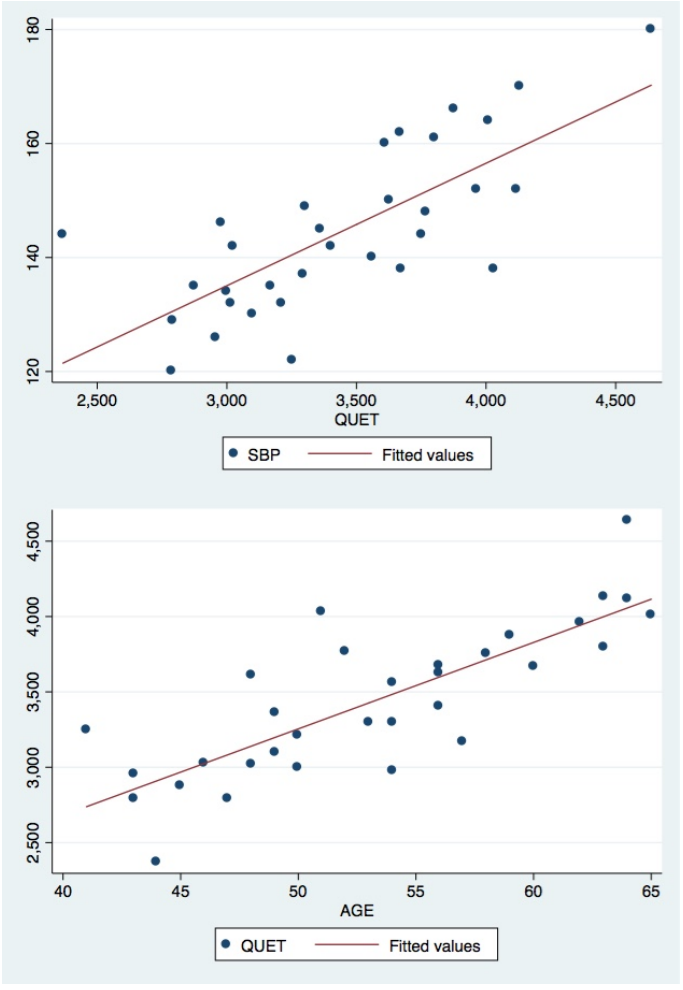
Hi,

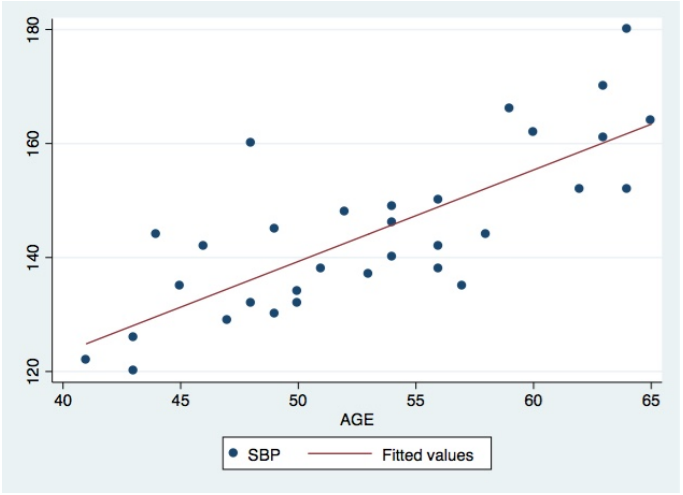
On the Quiz section, you will find out the data. Below, the link:

https://d396qusza40orc.cloudfront.net/appliedregression/Homeworks/week2-HW-data.csv

I hope it helps.

Daniel Rojas · 5 days ago Posted in Homework-Week Two-Exercise Two



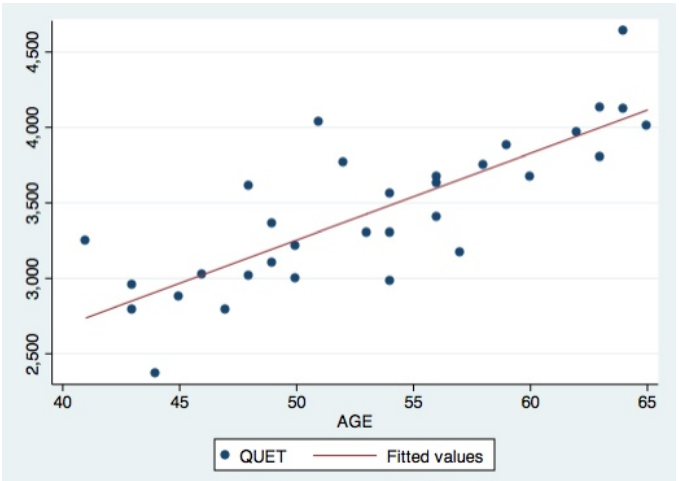


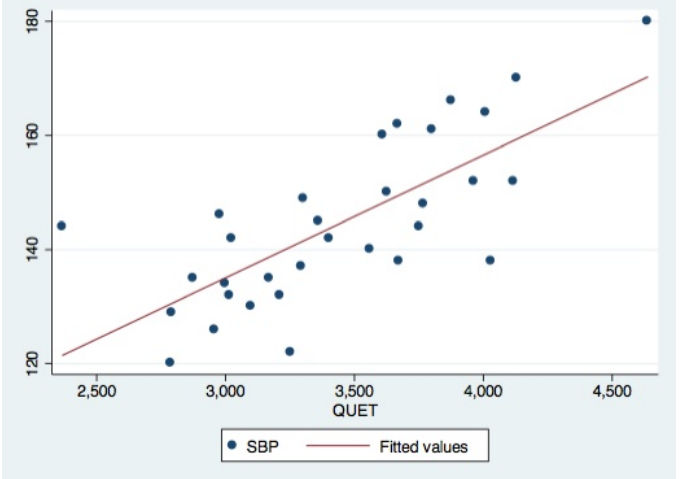
Pegando porte y la vara con la regresion lineal
Costa Rica representing
There is a lineal relationship in all these cases

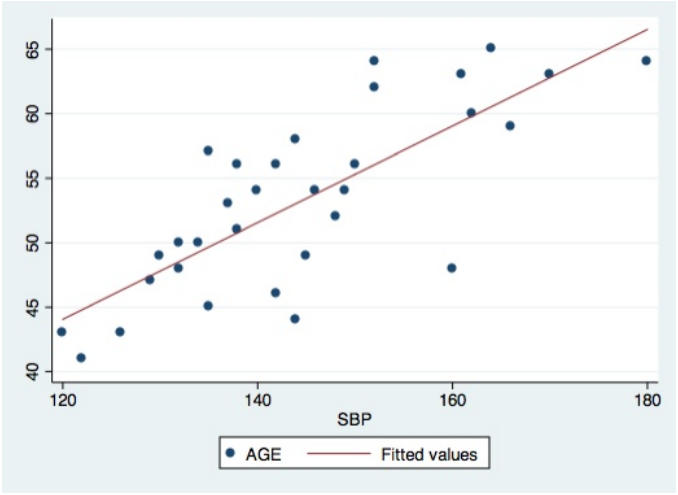
Susana Acosta Gonzalez · 5 days ago Posted in Week2 HW Exercise 2

A bit late but here's my results:

All of the graphs show a positive linear relationship between the variables.

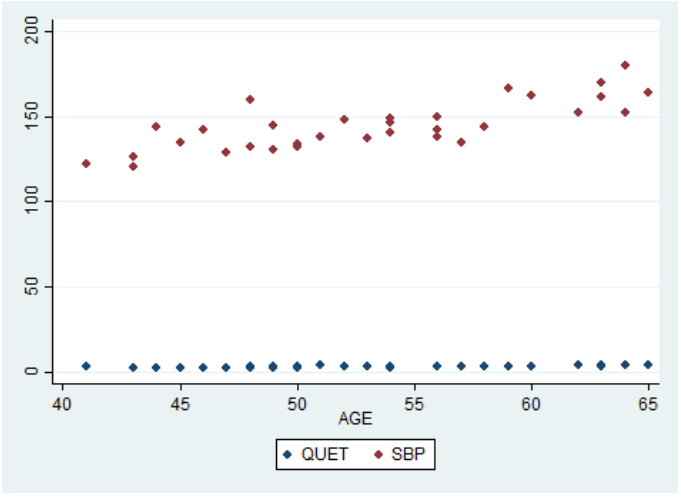






Hagai Levine · 13 days ago

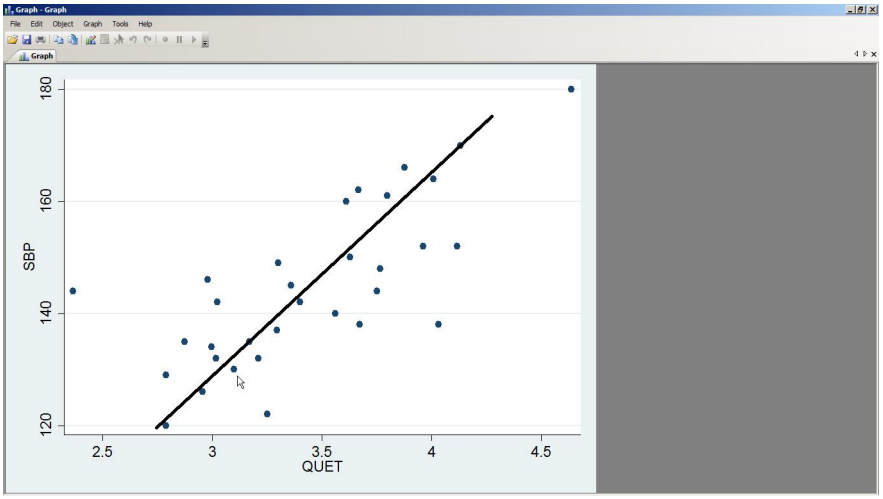
Posted in Week Two Homework - Exercise 1

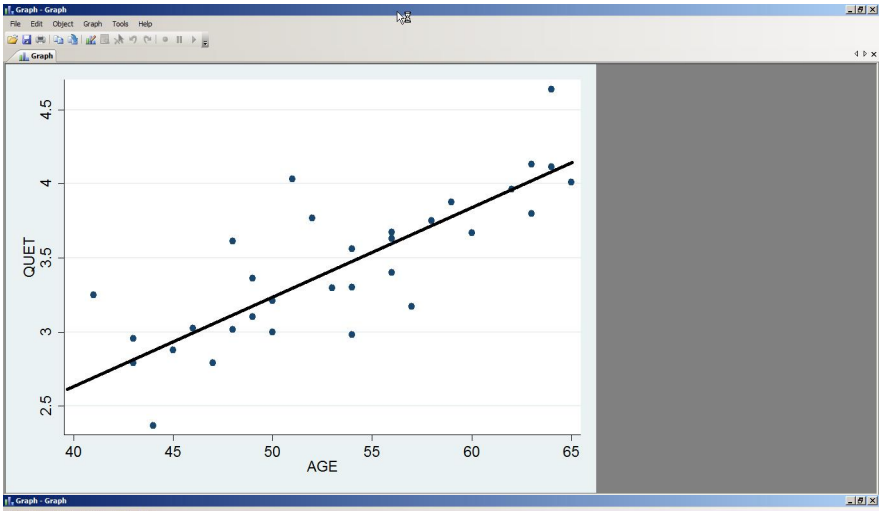


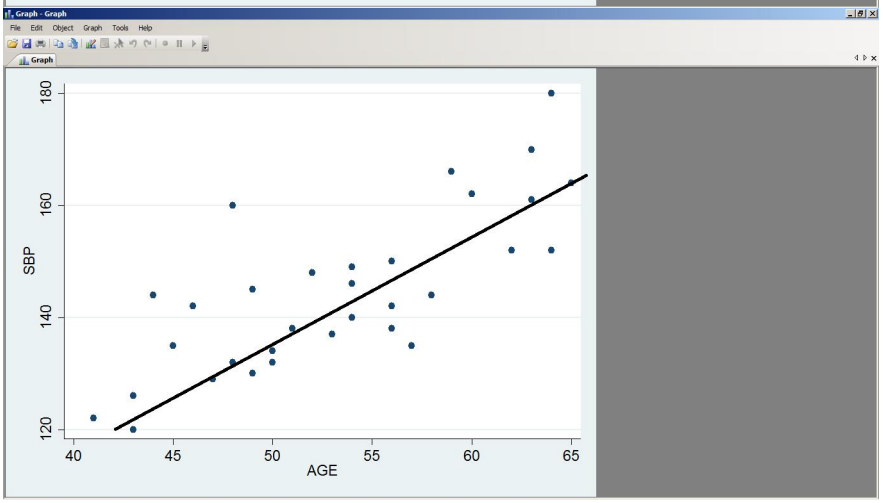
It's nice to see how scaling is important - see what happens if we show sbp and quet in the same graph

Gari de Jesus Jerez Aguero · 3 days ago

Posted in Week Two Homework - Exercise 1







Oscar Granados Cordero · 12 hours ago

Posted in Homework week 3

My results below:

SBP vs SMK

The regression is not significant. F=1.95. p-value = 0.1723 much greater than 0,05. r=0.2474, r^2=0.0612. Only 6% of SBP is explained by the regression with SMK.

SBP vs QUET

The regression is significant. F=36.75. p-value = 0.0 less than 0,05 and 0.01. r=0.7420, r^2=0.5506. 55% de SBP is explained by the regression with QUET.

QUET vs AGE

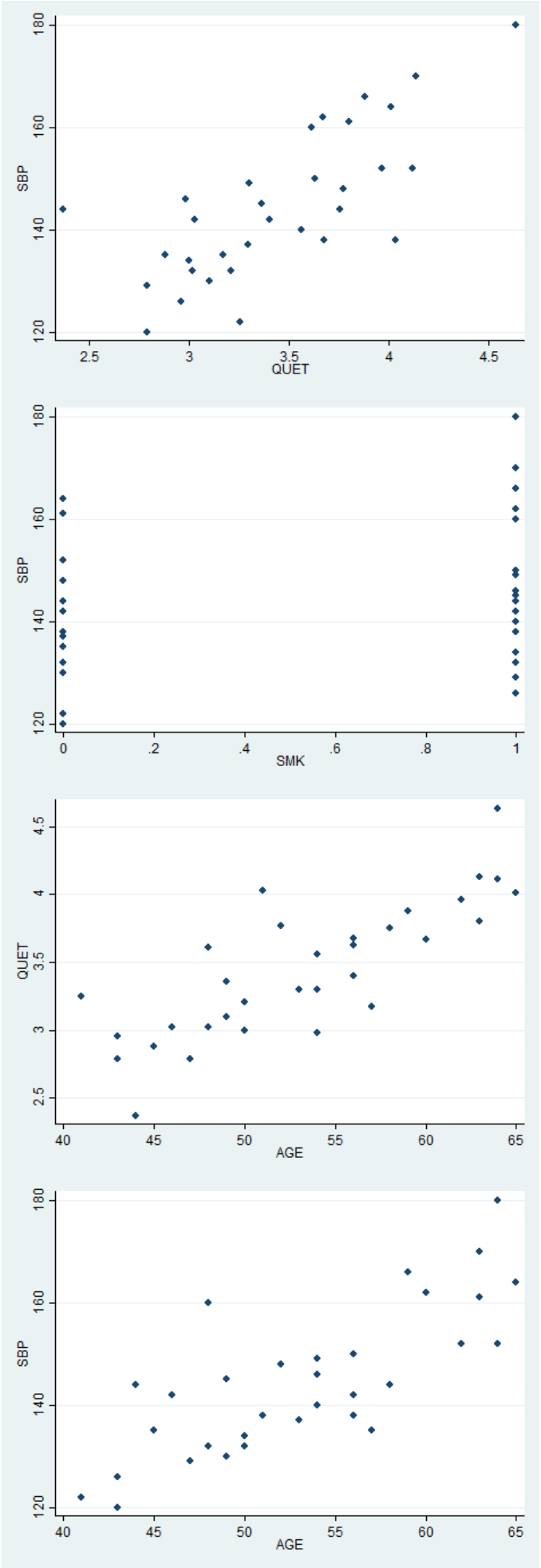
The regression is significant. F=54.37. p-value = 0.0 less than 0,05 and 0.01. r=0.8027, r^2=0.6444. 64.4% de QUET is explained by the regression with AGE.

SBP vs AGE

The regression is significant. F=45.18. p-value = 0.0 less than 0,05 and 0.01. r=0.7752, r^2=0.6009. 60% de SBP is explained by the regression with AGE.

luca balestrini · 4 days ago

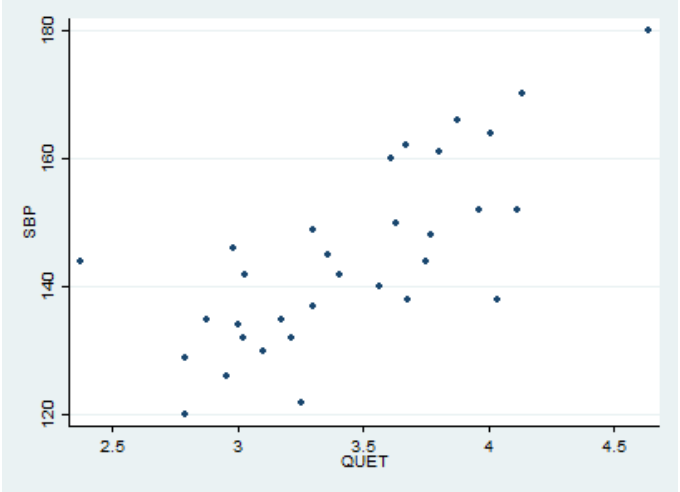
Posted in Week Two Homework - Exercise 1



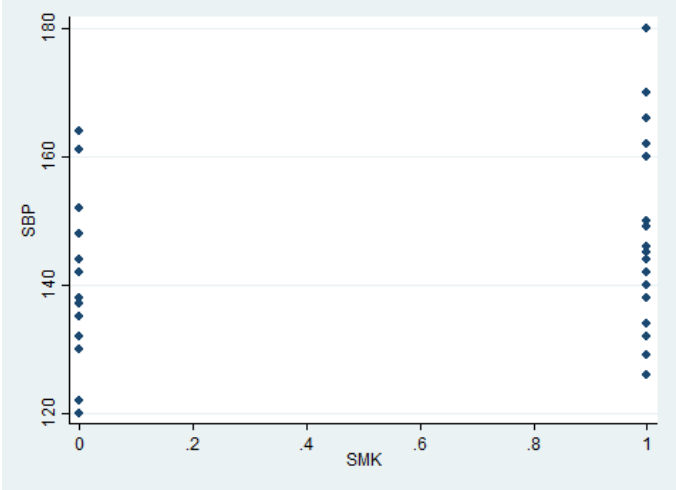
Viorel NITA · 14 days ago

Posted in Week Two Homework - Exercise 1

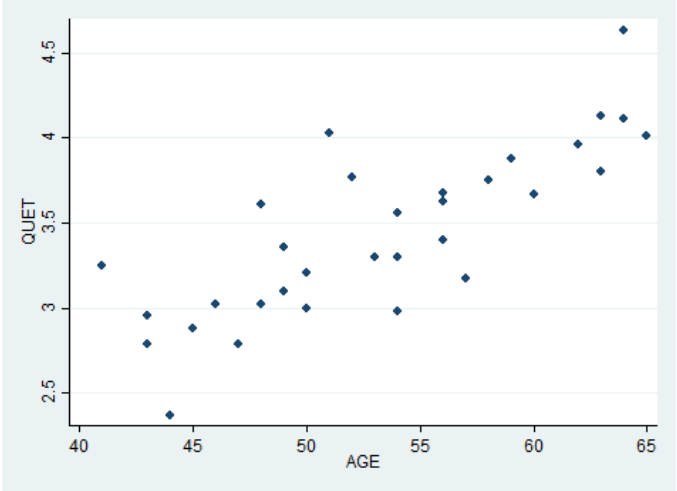
1.


SBP
120
140
160
180
2.5
3
3.5
4
4.5
QUET

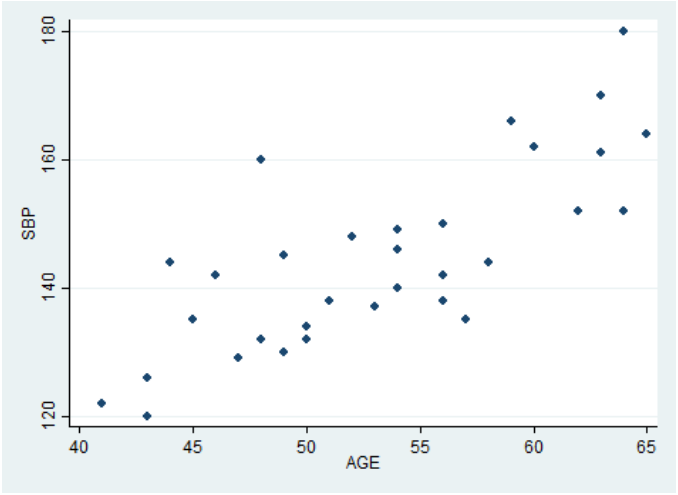
2.


SBP
120
140
160
180
0
.2
.4
.6
.8
1
SMK

3.

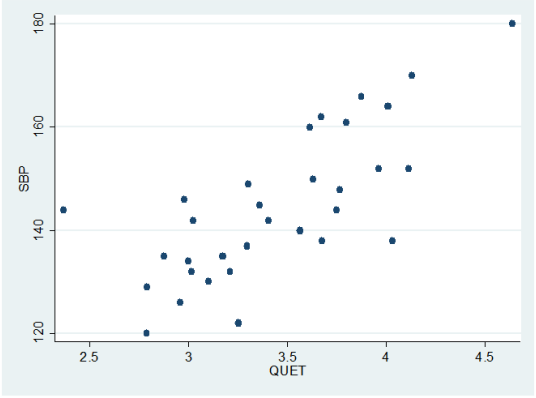

QUET
2.5
3
3.5
4
4.5
40
45
50
55
60
65
AGE

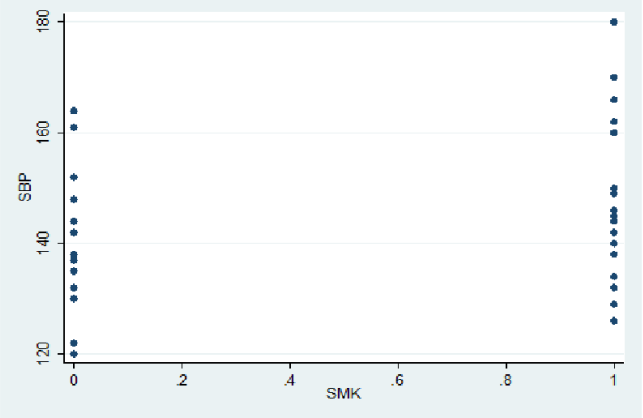
4.

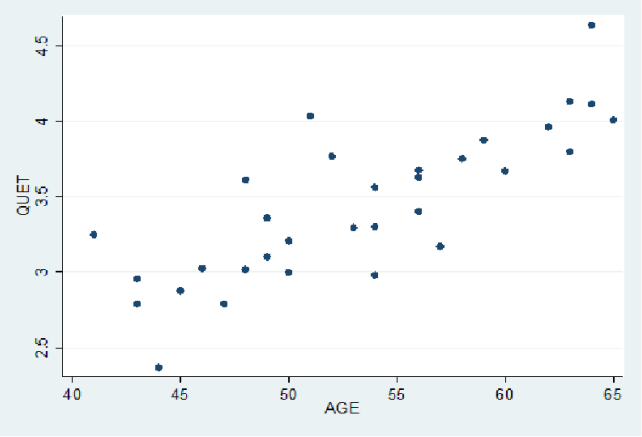


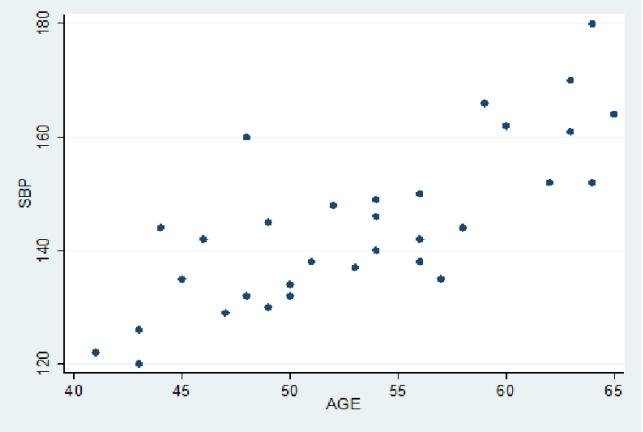
Musa Hasen Ahmed · 12 days ago

Posted in Week Two Homework - Exercise 1









Viorel NITA · 2 days ago

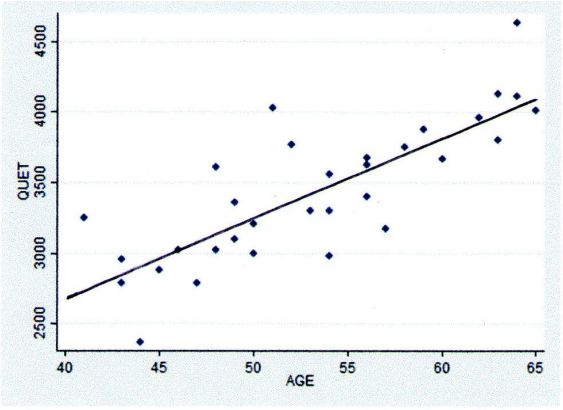
Posted in week 3 homework

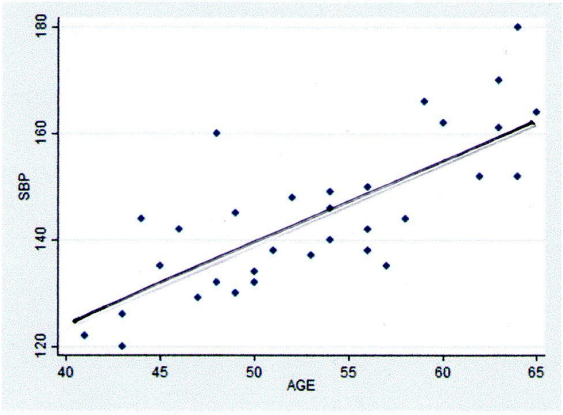
. regress SBP AGE

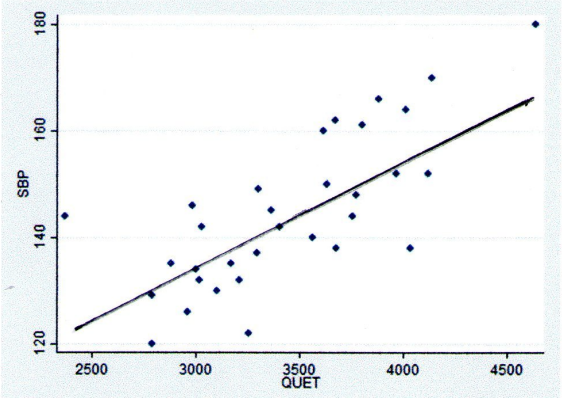
```
      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  1,    30) =   45.18
       Model |  3861.63037     1  3861.63037            Prob > F      =  0.0000
    Residual |  2564.33838    30  85.4779458            R-squared     =  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     6.72   0.000     1.116977    2.092023
       _cons |   59.09162   12.81626     4.61   0.000     32.91733    85.26592
------------------------------------------------------------------------------
```

Jorge Marques Pontes · 11 days ago

Posted in Week2 HW Exercise 2







I can perceive that there is positive linear relation between the variables in the three graphs..

1 2 3 → Next