Week Three Homework Exercise

Help Center

Homework Central / Week Three Homework /

For this week's homework, you will notice that the table of data is the same one from the **week two homework** from last week.

Use the data from the table to determine the ANOVA tables for some regressions specifically noted below. Scroll down below the table to view the exercise. Use the **homework forum** to discuss this exercise with your peers and share your discoveries.

To recall, the following 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

If desired, you may download the data for this exercise in this **CSV file**

For our homework, complete the following:

Exercise One

Determine the ANOVA tables for the following regressions:

1. SBP (Y) on SMK (X)
2. SBP (Y) on QUET (X)
3. QUET (Y) on AGE (X)
4. SBP (Y) on AGE (X)

Exercise Two

Use the ANOVA tables to perform the F-test for the significance of each straight-line regression.

Exercise Three

Interpret your results.
Use the **homework forum** to share your discoveries and discuss this homework with your peers.

You can download all homework assignments for week 3 here

For help and answers to this week's exercises, click the button below to visit the solutions page.

Solutions Page

Created Tue 10 Mar 2015 8:17 AM PDT

Last Modified Mon 6 Apr 2015 9:25 AM PDT

Week Three Homework Solutions

Help Center

Homework Central / Week Three Homework / Week Three Solutions

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

- **Exercise One**
- **Exercise Two**
- **Exercise Three** requires that you use our **Discussion Forums** to share your discoveries and discuss insights with your peers.

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 Thu 9 Apr 2015 12:38 PM PDT

Homework Solutions
Applied Regression Analysis

WEEK 3

Exercise One
Determine the ANOVA tables for the following regressions:

For each of the regressions, type "regress" followed by the Y and X variables, into the command box. From this command, you should see an ANOVA table, as well as a t-statistic for the slope coefficient below, and an F-statistic of model fit in the upper right corner *(See images below)*

1. SBP (Y) on SMK (X)

```
. 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
```

2. SBP (Y) on QUET (X)

```
. 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
------------------------------------------------------------------------------
```

3. QUET (Y) on AGE (X)

```
. regress quet age

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  1,    30) =   54.37
       Model |  4.93597216     1  4.93597216           Prob > F      =  0.0000
    Residual |  2.72371324    30  .090790441           R-squared     =  0.6444
-------------+------------------------------           Adj R-squared =  0.6326
       Total |   7.6596854    31  .247086626           Root MSE      =  .30131

------------------------------------------------------------------------------
        quet |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0573642   .0077799     7.37   0.000     .0414755    .0732529
       _cons |   .3864517   .4176903     0.93   0.362    -.4665857    1.239489
------------------------------------------------------------------------------
```

4. SBP (Y) on AGE (X)

```
. 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
```

Homework Solutions
Applied Regression Analysis

WEEK 3

Exercise Two

Use the ANOVA tables to perform the F test for the significance of each straight-line regression.

For each of the following, we are testing whether the slope coefficient is equal to zero. In other words, we are testing if the independent variable contributes significantly to the model. Notice that, because there is only one independent variable in the model, the F-test for the overall model yields the same significance as the t-test for the slope coefficient.

Looking at SBP (Y) on SMK (X)

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

```
. 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
```

With F=1.95 and p=0.1723, we fail to reject the null hypothesis. There is not sufficient evidence to conclude that there is a significant straight-line relationship between SBP and SMK.

Looking at SBP (Y) on QUET (X)

```
. 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
------------------------------------------------------------------------------
```

With F=36.75 and p<0.0001, we reject the null hypothesis. There is sufficient evidence to conclude that there is a significant straight-line relationship between SBP and QUET.

Week 3 Homework Assignment

Subscribe for email updates.

No tags yet. + Add Tag

Sort replies by: Oldest first Newest first Most popular

Amos B Robinson · 4 days ago

Effect of Smoking on Systolic Blood Pressure

Number of obs = 32
F(1, 30) = 1.95
Prob > F = 0.1723

Under the Naive Model the blood pressure is 144. So because the "F" statistic is so low, there is not convincing evidence that smoking has a statistical significance on SBP. This is confirmed on the "F" statistic because the "P" value for the "F" statistic is .1723, which is above the critical value of .05 for a 95% confidence level. Additionally, the R-squared for this regression is 0.0612, which further confirms that the statistical impact of smoking on blood pressure is minimal.

Effect of Body Size on Systolic Blood Pressure

Number of obs = 32
F(1, 30) = 36.75
Prob > F = 0.0000
R-squared = 0.5506

Under the Naive Model the blood pressure is 144. Here the "F" statistic is very large and the associated "P" value is < than .0000. So there is convincing evidence that body size has a statistically significant effect on systolic blood pressure. The high coefficient of determination "R-squared" shows that QUET explains about 55% of SBP.

Effect of AGE on Body Size

Number of obs = 32
F(1, 30) = 54.37
Prob > F = 0.0000
R-squared = 0.6444

The mean body size of this sample is 3.44. However, there is convincing evidence that age does effect body size. The "F" statistic is large at 54.37 and the associated "P" value is less than .0000. The R-squared shows that age explains 64.44% of body size.

Effect of Age on Systolic Blood Pressure

Number of obs = 32
F(1, 30) = 45.18
Prob > F = 0.0000
R-squared = 0.6009

The Naive Model for SBP is 144. The "F" statistic is 45.18, which is pretty large. The corresponding "P" value for this "F" statistic is less than .0000. So there is convincing evidence that age does have an impact on systolic blood pressure.

4 · flag

Kahsay Tadesse · 3 days ago

well done, great

0 · flag

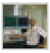
+ Comment

Kahsay Tadesse · 3 days ago

am not clear about the culculation on :"**Effect of Body Size on Systolic Blood Press**

0 · flag

+ Comment



KK Wong · 3 days ago

1. SBP vs SMK

```
. reg 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
```

Given relatively small F(1,30)=1.95, high Prob>F=0.1723 and low adj R-squared=0.0299, it suggests to reject the hypothesis that SBP has an association with SMK; ie, the association between SBP & SMK is approx 2.99% as suggested by the adj R-squared. It further supports by the fact that SMK 95% CI contains 0 and its P>|t|=0.172 which is significantly away from 0.

2. SBP vs QUET

```
. reg SBP QUET

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  1,    30) =   36.75
       Model |  3537.94574     1  3537.94574           Prob > F      =  0.0000
    Residual |  2888.02301    30  96.2674337           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 |    70.5764   12.32187     5.73   0.000     45.41179    95.74101
```

Given relatively high F(1,30)=36.75, significantly low Prob>F=0.0000 and high adj R-squared=0.5356, it suggests not to reject the hypothesis that SBP vs QUET; ie, the association between SBP & QUET is approx 53.56% as suggested by the adj R-squared. It further supports by the fact that QUET 95% CI does not contains 0 and its P>|t|=0.000.

3. QUET vs AGE

```
. reg QUET AGE

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  1,    30) =   54.37
       Model |  4.93597143     1  4.93597143           Prob > F      =  0.0000
    Residual |  2.72371329    30  .090790443           R-squared     =  0.6444
-------------+------------------------------           Adj R-squared =  0.6326
       Total |  7.65968472    31  .247086604           Root MSE      =  .30131

------------------------------------------------------------------------------
        QUET |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         AGE |   .0573642   .0077799     7.37   0.000     .0414755    .0732529
       _cons |   .3864519   .4176903     0.93   0.362    -.4665855    1.239489
```

Given relatively high F(1,30)=54.37, significantly low Prob>F=0.0000 and high adj R-squared=0.6326, it suggests not to reject the hypothesis that QUET vs AGE; ie, the association between QUET & AGE is approx 63.26% as suggested by the adj R-squared. It further supports by the fact that AGE 95% CI does not contains 0 and its P>|t|=0.000.

4. SBP vs AGE

```
. reg 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
```

Given relatively high F(1,30)=45.18, significantly low Prob>F=0.0000 and high adj R-squared=0.5876, it suggests not to reject the hypothesis that SMK vs AGE; ie, the association between SMK & AGE is approx 58.76% as suggested by the adj R-squared. It further supports by the fact that AGE 95% CI does not contains 0 and its P>|t|=0.000.

1 · flag

+ Comment

Juan C. Trujillo · 2 days ago

I am confused about this case. Could there be a possibility in which the F test turns out to be statistically significant, but the t test for the explanatory variable appears with a high p-value?

Please, explain.

0 · flag

Anonymous · 2 days ago

If you are testing the statistical significance of the estimate of a single regression coefficient, the inferential outcome will be the same whether you use the t-test or the F-test.

Mechanically, this is because the F-statistic associated with testing a single coefficient estimate is just squared t-statistic associated with that same estimate.

0 · flag

+ Comment

Varalakshmi · 2 days ago

Amos: I did the F statistic for SBP (Y) on SMK (X)
The Model with 1 d.f parameter because there is one independent variable SMK and the residual which is y and y predicted has 2 d.f one for intercept and slope. As you stated the R squared which is the explained variation to the Total variation is very little. P value so small suggests that there is a significant difference in Smoking history and SBP.

Finding the t-statistic = r/sqrt((1-r^2)/(n-2)) = 1.398459 or as is also given in the table below.
We can check that t^2 = 1.40^2 is the same as F statistic= 1.95

The lecture was quite informative!

Can someone help me to graph two scatter plots in one - Week 1 homework where we need to graph the residuals as well the given observations?

```
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-squa> red     =  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 % Con
>      f. Interval]
         smk |   7.023529   5.023498     1.40   0.172    -3.2 35823
>        17.28288
       _cons |      140.8   3.661472    38.45   0.000     133 .3223
>        148.2777
------------------------------------------------------------------
> -----------------

.
```

0 · flag

+ Comment

luca balestrini · a day ago

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

2)sbp on 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

3)quet on age

Source	SS	df MS		Number of obs	= 32	
		F(1, 30)		= 54.37		
Model	4.93597216	1 4.93597216		Prob > F	= 0.0000	
Residual	2.72371324	30 .090790441		R-squared	= 0.6444	
		Adj R-squared		= 0.6326		
Total	7.6596854	31 .247086626		Root MSE	= .30131	
quet	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
age	.0573642	.0077799	7.37	0.000	.0414755	.0732529
_cons	.3864517	.4176903	0.93	0.362	-.4665857	1.239489

4)sbp on 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

↑ 0 ↓ · flag

+ Comment

David C. Morris · a day ago

I got the same results as above. I'm not going to repost here. However, I was thinking about the four analyses we did. Two of them make sense to me: 1) Blood pressure (sbp) and Smoking (smk); and 2) Blood pressure (sbp) and Body size (quet). I know we're just doing the homework to learn how to run/interpret anova tables. However, it got me thinking about what 'quet' really is. The description says 'body size' but looking at the values (range from ~2 to ~4.6) I don't have any reference point for that. What is Quet? How is it measured? The reason I bring this up is I was a little surprise by the high correlation between Quet and Age. The range of age in the sample is 41 to 65. The correlation between them is R-squared .64. Looking at the scatter plot, as Age goes up, so does Quet. I would've expected it to taper off toward the older ages since people tend to lose muscle mass and usually weight the older they get. Does anyone know what QUET is? I did a Google search but only found other data sets with no description.

0 · flag

+ Comment

Emilija Nikolic-Djoric · 20 hours ago

I think that it is Quetelet's index defined as:
QUET=100*(weight/height^2)

1 · flag

+ Comment

Emilija Nikolic-Djoric · 19 hours ago

Week 3-Slide 17-at the bottom n-2 instead n-1?

0 · flag

+ Comment

ANCA MINCIU · 18 hours ago

I have done the test, with same results. To avoid re-posting the same information, I would just add one sentence to each interpretation.

1.**Effect of Smoking on Systolic Blood Pressure**
The confidence interval in this case contains zero, so smoking is not a good predictor for blood pressure.

2. Effect of Body Size on Systolic Blood Pressure
The confidence interval does not contain zero, so the body size influences the systolic blood pressure.

3. Effect of AGE on Body Size
The confidence interval does not contain zero, so age is a very important predictor for body size.

4. Effect of Age on Systolic Blood Pressure

Age is a very important predictor for systolic blood pressure, taking into consideration that the confidence interval does not contain zero.

0 · flag

+ Comment

Alina Denham · 11 hours ago

Hello, everybody!

Here are my observations:

- Let's test the null hypothesis (the slope equals zero) for SBP on SMK. F=1.95 (< F.95=4.20) and p=0.1723 (>.05). Therefore, we fail to reject the null hypothesis. This means that we do not have sufficient evidence to prove that there is a significant linear relationship between blood pressure and smoking history.

- Let's test the null hypothesis (the slope equals zero) for SBP on QUET. F=36.75 (>> F.95) and p=0.000 (<0.001). Therefore, we reject the null hypothesis. This means that we have sufficient evidence to prove that there is a significant linear relationship between blood pressure and body size.

-Let's test the null hypothesis (the slope equals zero) for QUET on AGE. F=54.37 (>>F.95) and p=0.000 (<0.001). Therefore, we reject the null hypothesis. This means that we have sufficient evidence to prove that there is significant linear relationship between body size and age.

- Let's test the null hypothesis (the slope equals zero) for SBP on AGE. F=45.18 (>>F.95) and p=0.000 (<0.001). Therefore, we reject the null hypothesis. This means that we have sufficient evidence to prove that there is significant linear relationship between blood pressure and age.

0 · flag

+ Comment

Lien-yu Yeh · 10 hours ago

Assume that,
H0:β1 equal to 0 vs H1:β1 not equal to 0

According to the rule of test with P-value,
if α>p-value,then we reject H0,and if α<p-value,we don't reject H0,where α is significant level
because p-value=Pr(reject H0 ︱ H0 is true)= probibility of type I error (with sample),
when p-value is large,H0 probably be true,because there is high probability to make mistake,so we don't reject H0
when p-value is small,the probability of type I error is too low,so we reject H0

I used above concept to test following question,and I assume that significant level is 0.05(α=0.05)

```
. 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      =  14.18

------------------------------------------------------------------------------
         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
------------------------------------------------------------------------------
```

significant level=0.05 < p-value=0.1723 ,we don't reject H0:β1 equal to 0,
so we *don't have* enough evidence to refer that SBP and SMK has significant relatonship.

```
. regress SBP QUET

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  1,    30) =   36.75
       Model |  3537.94574     1  3537.94574           Prob > F      =  0.0000
    Residual |  2888.02301    30  96.2674337           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 |    70.5764   12.32187     5.73   0.000     45.41179    95.74101
------------------------------------------------------------------------------
```

significant level=0.05 > p-value=0.0000 ,we reject H0:β1 equal to 0,
so we *have* enough evidence to refer that SBP and QUET has significant relatonship.

```
. regress QUET AGE

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  1,    30) =   54.37
       Model |  4.93597143     1  4.93597143           Prob > F      =  0.0000
    Residual |  2.72371329    30  .090790443           R-squared     =  0.6444
-------------+------------------------------           Adj R-squared =  0.6326
       Total |  7.65968472    31  .247086604           Root MSE      =  .30131
```

```
------------------------------------------------------------------------------
        QUET |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         AGE |   .0573642   .0077799     7.37   0.000     .0414755    .0732529
       _cons |   .3864519   .4176903     0.93   0.362    -.4665855    1.239489
------------------------------------------------------------------------------
```

significant level=0.05 > p-value=0.0000 ,we reject H0:β1 equal to 0,
so we *have* enough evidence to refer that QUET and AGE has significant relatonship.

```
. 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
------------------------------------------------------------------------------
```

significant level=0.05 > p-value=0.0000 ,we reject H0:β1 equal to 0,
so we *have* enough evidence to refer that SBP and AGE has significant relatonship.

0 · flag

+ Comment

Walter O. Augenstein · 2 hours ago

1. SBP(Y) vs. SMK(X)

```
. 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
------------------------------------------------------------------------------
```

invFtail(1, 30, 0.05) = 4.1708768, but F = 1.95 => we cannot reject the Null hypothesis.
The regression line explains 6% of the total squared variation.
We cannot establish a relationship of sbp to smk.

2. SBP(Y) vs. QUET(x)

```
. 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
------------------------------------------------------------------------------
```

invFtail(1, 30, 0.05) = 4.1708768, and F = 36.75 => we reject the Null hypothesis.
The regression line explains 55%% of the total squared variation.
There is a definite linear component to the regression of sbp on quet.

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

```
. regress quet age

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  1,    30) =   54.37
       Model |  4.93597216     1  4.93597216           Prob > F      =  0.0000
    Residual |  2.72371324    30  .090790441           R-squared     =  0.6444
-------------+------------------------------           Adj R-squared =  0.6326
       Total |   7.6596854    31  .247086626           Root MSE      =  .30131


------------------------------------------------------------------------------
        quet |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0573642   .0077799     7.37   0.000     .0414755    .0732529
       _cons |   .3864517   .4176903     0.93   0.362    -.4665857    1.239489
------------------------------------------------------------------------------
```

invFtail(1, 30, 0.05) = 4.1708768, and F = 54.37 => we reject the Null hypothesis.

The regression line explains 64% of the total squared variation.
There is a definite linear component to the regression of quet on age.

4. SBP(Y) vs AGE(X)

```
. 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
------------------------------------------------------------------------------
```

invFtail(1, 30, 0.05) = 4.1708768, and F = 45.18 => we reject the Null hypothesis.
The regression line explains 60% of the total squared variation.
There is a definite linear component to the regression of sbp on age.

0 · flag

+ Comment

New post

To ensure a positive and productive discussion, please read our forum posting policies before posting.

B I Link <code> Pic Math Edit: Rich Preview

Make this post anonymous to other students

Subscribe to this thread at the same time

Add post

Week Three Homework Assignment

Subscribe for email updates.

No tags yet. + Add Tag

Jesse Booher · a day ago

Regression 1

. 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
------------------------------------------------------------------------------
```

The ANOVA output for the first regression shows a weak linear relationships between blood pressure and smoking. A strong relationship would indicate a high F score. While higher than 0, an F score of 1.95 with 1 IV and 30 DF yields a P-Value of 0.172, which is much higher than the 0.05 value we would need to have confidence in this model. Further the R-Square indicates that even if the model were a statistically significant predictor of blood pressure, we would only be able to explain 6% of the variance.

Regression 2

. 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
------------------------------------------------------------------------------
```

The ANOVA Output for the second regression shows a strong linear relationship between blood pressure and body size. With an F Score of 36.75, 1 IV and 30 DF, we have a P-value much less than 0.05 at 0.000. Coupled with an R-Square score that explains 55% of the variance in blood pressure, we can be confident that the model is a statistically significant predictor of blood pressure.

Regression 3

. regress quet age

```
      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  1,    30) =   54.37
       Model |  4.93597216     1  4.93597216           Prob > F      =  0.0000
    Residual |  2.72371324    30  .090790441           R-squared     =  0.6444
-------------+------------------------------           Adj R-squared =  0.6326
       Total |   7.6596854    31  .247086626           Root MSE      =  .30131


------------------------------------------------------------------------------
        quet |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |   .0573642   .0077799     7.37   0.000     .0414755    .0732529
       _cons |   .3864517   .4176903     0.93   0.362    -.4665857    1.239489
------------------------------------------------------------------------------
```

The third ANOVA output shows an even stronger linear relationship between body size and age. With an F-Score of 54.37, 1 IV and 30 DF, we have a P-value much less than 0.05 at 0.000. Coupled with an R-Square score that explains 64% of variance in body size, we can be confident that the model is a statistically significant predictor of blood pressure.

Regression 4

. 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
------------------------------------------------------------------------------
```

The 4th ANOVA output shows a strong linear relationship between blood pressure and age. With an F-Score of 45.18, 1 IV and 30 DF, we have a P-value much less than 0.05 at 0.000. Coupled with an R-Square that explains 60% of he variance in blood pressure, we can be confident that the model is a statistically significant predictor of age.

As we progress into multiple IVs, I would expect a model with all IVs to show us that age and body size a statistically significant predictors of blood pressure.

0 · flag

New post

To ensure a positive and productive discussion, please read our forum posting policies before posting.

B I Link <code> Pic Math Edit: Rich Preview

Make this post anonymous to other students

Subscribe to this thread at the same time

Add post

Forums / Homework Exercises / Week Three Homework Help Center

Week 3 homework results and analysis

Subscribe for email updates.

No tags yet. + Add Tag

Sort replies by: Oldest first Newest first Most popular

W J Kinder · 5 days ago

SBP (Y) on SMK (X): With n=32, F=1.95 and p=.1723, we fail to reject the null hypothesis. There is not sufficient evidence to conclude that there is a significant linear relationship between SBP and SMK.

SBP (Y) on QUET (X): With n=32, F=36.75 and p=.0000 we reject the null hypothesis and conclude that there is a significant linear relationship between SBP and QUET. The model is SBP= 70.58+21.49*QUET . For each one unit increase in QUET, this model predicts SBP increases 21.49; CI(14.25 to 28.73)

QUET (Y) on AGE (X): With n=32, F= 54.37 and p=.0000 we reject the null hypothesis and conclude that there is a significant linear relationship between QUET and AGE. The model is QUET=.3864+.057*AGE. For evey one year increase in age, this model predicts an increase of .057 in QUET, CI(.04 to .07)

SBP (Y) on AGE (X): With n=32, F=45.18 and p=.0000 we reject the null hypothesis and conclude that there is a significant linear relationship between SBP and AGE. The model is SBP= 59.09+1.60*AGE. For each one year increase in age, this model predicts a sbp increase of 1.60, CI(1.11 to 2.09)

Other thoughts:
The dataset included ages from 41 to 65, mean age was 53, so we can not generalize the results to other ages. If a variable for gender would have been included and the sample size was larger, it would be interesting to run the analysis seperately by gender to see what differences by gender might exist.

1 · flag

Kathy Padkapayeva · 5 days ago

Nice post, W J Kinder! I'm wondering if professor Lemeshow has the information on gender for this sample, and could share it with us, so that we could see if there are differences by gender here.

Also, it is interesting what is the sum of the percentage of variance in SBP that is explained by the three models (taking the R-squared):
SMK explains about 6% of the variance in SBP
QUET explains about 55% of the variance

AGE explains about 64% of the variance
In total it gets to 125%. If we sum what we have from adjusted R-Squared, it still gets to about 115%. It could mean that I'm wrong in my interpretation of the results, or in my attempt to sum it up. Or maybe it just demonstrates that our independent variables SMK, QUET, and AGE are interrelated themselves? It would be helpful if anyone who is a more mature statistician could provide some explanation on it.
Thank you

0 · flag

W J Kinder · 5 days ago

We can't sum up the individual R-squares to get 115%. We can run a multiple regression model using all the variables and then examine the adjusted R square for that multivariable model. When I did that I got a model with adjusted r-square of .73, but QUET was not statistically significant in that model. I think we'll cover that in week5. And you are right the independent variables are correlated with each other and that can create challenges too and I think week 5 will be interesting.

0 · flag

+ Comment

New post

To ensure a positive and productive discussion, please read our forum posting policies before posting.

B I Link <code> Pic Math Edit: Rich Preview

Make this post anonymous to other students

Subscribe to this thread at the same time

Add post

Forums / Homework Exercises / Week Three Homework Help Center

Week 3 homework thread for R users

Subscribe for email updates.

R × homework × + Add Tag Sort replies by: Oldest first Newest first Most popular

Jai Broome · 6 days ago

Hi all, I thought I'd create a thread at the start of the week for those of us that are doing the analyses in R. Feel free to post your questions or solutions to this week's assignment

0 · flag

onur miskbay · 6 days ago

Hi folks, I have a question.

Is there a function for showing the summary statistics and the ANOVA table at the same time?

What I would like to have is a nice output like STATA but I need to punch in anova function and summary function seperately for thaat. The anova function does not have any arguments and summary function's arguments did not do what I wanted.

0 · flag

W J Kinder · 5 days ago

This is the R code that I've used so far for week3 homework. It gives same conclusions as STATA but I'd like to improve it.

```
# Read dataset
mydata<- read.csv("~/A stata MOOC/week3/week3-HW-data.csv")
# Plot SBP by AGE
plot(mydata$SBP~mydata$AGE)
# Same plot with Variable names
plot(SBP~AGE, data=mydata)
# Run linear regression model named model1
model1=lm(SBP~AGE, data=mydata)
#see what is included in model1
names(model1)
#Summarize model1 this gives nice output
summary(model1)
# ANOVA for this model
anova(model1)
#Show the Confidence interval for slope
```

```
confint(model1)
# Set view for 2 rows and 2 columns for diagnostic graphs
par(mfrow=c(2,2))
#plot diagnostics for regresson model1
plot(model1)
#reset view for 1 row 1 column graph
par(mfrow=c(1,1))
#plot residuals for model1
plot (model1$residuals)
#draw horizontal line at zero point
abline(h=0)
```

0 · flag

A post was deleted

+ Comment

New post

To ensure a positive and productive discussion, please read our forum posting policies before posting.

B I Link <code> Pic Math Edit: Rich Preview

☐ Make this post anonymous to other students

☑ Subscribe to this thread at the same time

Add post