

STAT 13 Homework 7 Solutions

1. In a study of [protein synthesis](#) in the [oocyte](#) (developing egg cell) of the frog [Xenopus laevis](#), a biologist injected individual oocytes with radioactively labeled [leucine](#). At various times after injection, he made radioactivity measurements and calculated how much of the leucine had been incorporated into protein. The results are given in the table below; each leucine value is the content of labeled leucine in two oocytes. All oocytes were from the same female. Assume time is measured in minutes after injection.

Index	Time	Leucine
1	0	0.02
2	10	0.25
3	20	0.54
4	30	0.69
5	40	1.07
6	50	1.50
7	60	1.74
Mean	30	0.83

Use linear regression to estimate the rate of incorporation of the labeled leucine. Compare your results against the [SOCR Regression Analysis](#) applet.

$$\hat{b} = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2}$$
$$= \frac{(0-30)(0.02-0.83) + (10-30)(0.25-0.83) + (20-30)(0.54-0.83) + (30-30)(0.69-0.83) + (40-30)(1.07-0.83) + (50-30)(1.50-0.83) + (60-30)(1.74-0.83)}{(0-30)^2 + (10-30)^2 + (20-30)^2 + (30-30)^2 + (40-30)^2 + (50-30)^2 + (60-30)^2}$$
$$= 0.02925$$

Now let's do it in SOCR. Go to the data tab and enter the values

	C1	C2	C3	C4
1:	0	0.02		
2:	10	0.25		
3:	20	0.54		
4:	30	0.69		
5:	40	1.07		
6:	50	1.50		
7:	60	1.74		
8:				
9:				
10:				

Then go to the mapping tab and map leucine as the dependent and time as independent variable.

The screenshot shows the 'MAPPING' tab of the software. On the left, a list of columns from C3 to C16 is visible. On the right, there are two sections: 'DEPENDENT' and 'INDEPENDENT'. Under 'DEPENDENT', 'C2' is selected, with 'ADD' and 'REMOVE' buttons below it. Under 'INDEPENDENT', 'C1' is selected, also with 'ADD' and 'REMOVE' buttons below it.

Then you click on the calculate button to get the result in the result tab

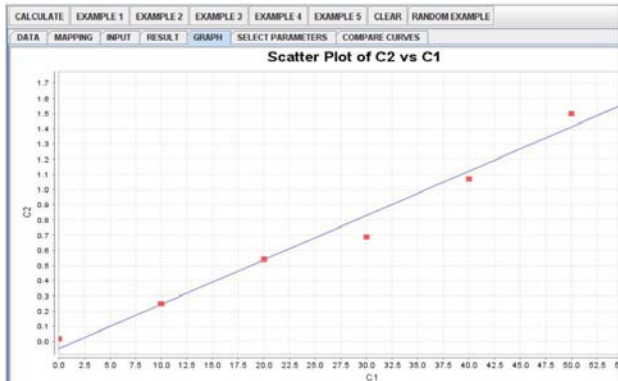
The screenshot shows the 'RESULT' tab of the software. It displays the following regression results:

- Independent Variable = C1**
- Simple Linear Regression Results:**
 - Mean of C1 = 30.000
 - Mean of C2 = .830
- Regression Line:**

$$C2 = -.047 + 0.02924999999999999 C1$$
- Correlation(C1, C2) = .993**
R-Square = .986
- Intercept:**
 - Parameter Estimate: -.047
 - Standard Error: .057
 - T-Statistics: -.831
 - P-Value: .444
- Slope:**
 - Parameter Estimate: .029
 - Standard Error: .002
 - T-Statistics: 18.440
 - P-Value: .000

Plot the data and explain the regression line.

Go to the graph tab



Calculate by hand the residual standard deviation.

Since residual is difference between the observed value and the fitted value,

$$\text{residual} = \text{Leucine} - \{(\text{intercept}) + (\text{slope} * \text{Time})\}.$$

Index	Time	Leucine	Residual
1	0	0.02	$0.02 - \{(-0.047) + (0.02925 * 0)\} = 0.067$
2	10	0.25	$0.25 - \{(-0.047) + (0.02925 * 10)\} = 0.0045$
3	20	0.54	$0.54 - \{(-0.047) + (0.02925 * 20)\} = 0.002$
4	30	0.69	$0.69 - \{(-0.047) + (0.02925 * 30)\} = -0.1405$
5	40	1.07	$1.07 - \{(-0.047) + (0.02925 * 40)\} = -0.053$
6	50	1.50	$1.50 - \{(-0.047) + (0.02925 * 50)\} = 0.0845$
7	60	1.74	$1.74 - \{(-0.047) + (0.02925 * 60)\} = 0.032$
Mean	30	0.83	0.0005

- The residuals (from index 1 to 7) are 0.067, 0.0045, 0.002, -0.1405, -0.053, 0.0845, and 0.032. The average residual is $(0.067+0.0045+0.002-0.1405-0.053+0.0845+0.032)/7=0.0005$.

The standard deviation of the residuals is

$$\sqrt{\frac{(0.067 - 0.0005)^2 + (0.0045 - 0.0005)^2 + (0.002 - 0.0005)^2 + (-0.1405 - 0.0005)^2 + (-0.053 - 0.0005)^2 + (0.0845 - 0.0005)^2 + (0.032 - 0.0005)^2}{7 - 1}} = 0.0766$$

2. Twenty plots, each $10 \times 4 \text{ meters}^2$ were randomly chosen in a large field of corn. For each plot, the plant density (number of plants in the plot) and the mean cob weight (grams of grain per cob) were observed. The results are given in the table below.

Index	Plant_Density_X	Cob_Weight_Y
1	137	212
2	107	241
3	132	215
4	135	225
5	115	250
6	103	241
7	102	237
8	65	282
9	149	206
10	85	246
11	173	194

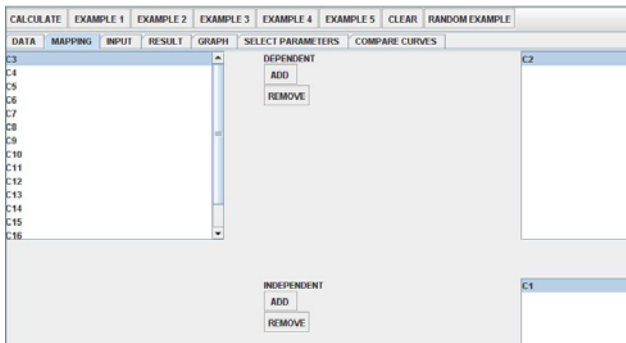
12	124	241
13	157	196
14	184	193
15	112	224
16	80	257
17	165	200
18	160	190
19	157	208
20	119	224
Mean	128.05	224.1

Calculate the linear regression of Y on X using the [SOCR Simple Linear Regression applet](#).

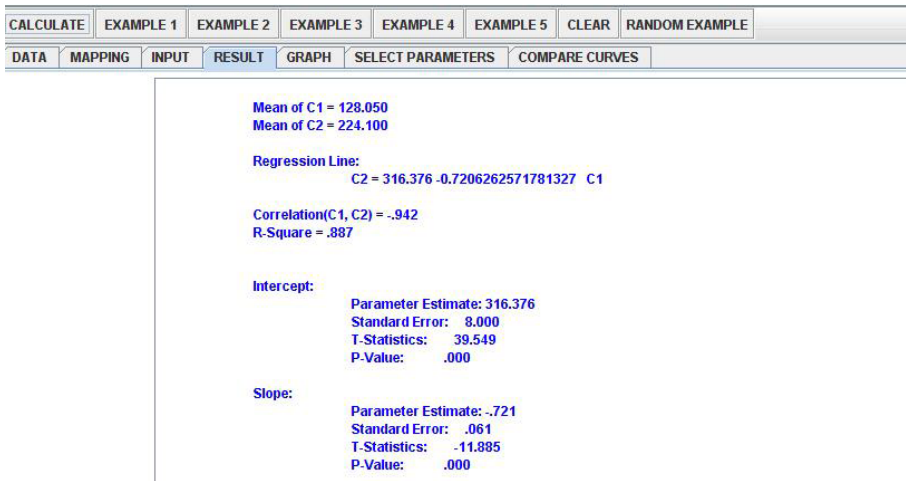
Go to the data tab and enter the numbers

	C1	C2	C3	C4	C5	C6	C7	C8	C9
1.	137	212							
2.	107	241							
3.	132	215							
4.	135	225							
5.	115	250							
6.	103	241							
7.	102	237							
8.	85	282							
9.	148	206							
10.	85	246							
11.	173	194							
12.	124	241							
13.	157	196							
14.	184	193							
15.	112	224							
16.	80	257							
17.	165	200							
18.	160	190							
19.	157	208							
20.	119	224							
21.									

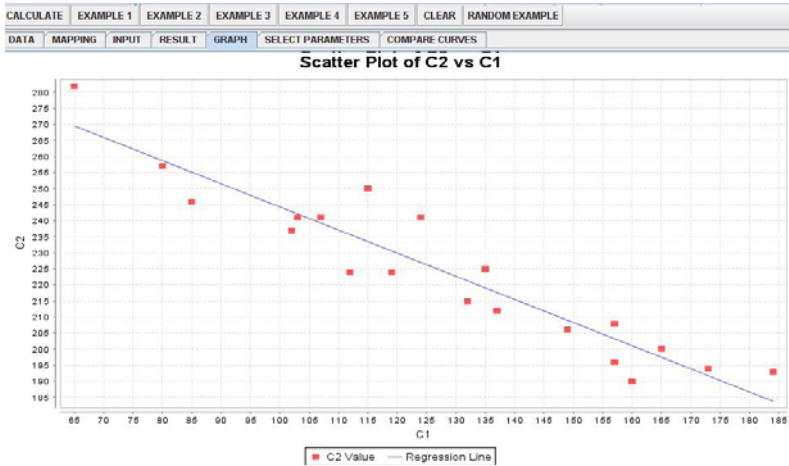
Go to the mapping tab and map the plant density as dependent and cob weight as independent variable.



Hit the calculate button and the result is in the result tab



Plot the data and draw the regression line on the graph.



Interpret the values of the regression slope, intercept and correlation in the context of this setting.

Slope estimate of -0.721 indicates that with every increase in a plant the cob weight is expected to decrease by 0.721 . Intercept of 316.376 indicates that when plant density is 0 cob weight is 316.376 which have no meaningful interpretation. This does not invalidate the model its self, it just means that the extrapolation of the model to 0 is meaningless. There is negative correlation between the plant density and cob weight.

3. Use the [SOCR Hot-dogs dataset](#) to the effects of the type of meat on the sodium (first) and calorie (second) content

of the hot-dogs. Use the [SOCR 1-Way ANOVA applet](#) to carry the calculations. Provide all results and explicitly state your inference.

Sodium

1. Setup the Null hypothesis

H0: There is not effect of meat types on the sodium content

H1: There is effect of meat types on the sodium content

2. Fit the ANOVA model

Go to the data tab and input the data

	C1	C2	C3	C4	C5	C6
1:	495	A				
2:	477	A				
3:	425	A				
4:	322	A				
5:	482	A				
6:	587	A				
7:	370	A				
8:	322	A				
9:	479	A				
10:	375	A				
11:	330	A				
12:	300	A				

Go to the mapping tab and map sodium(C1) as dependent variable

The screenshot shows the 'MAPPING' tab of the SOCR ANOVA applet. On the left side, there is a vertical list of columns labeled C3, C4, C5, C6, C7, C8, C9, C10, C11, C12, C13, C14, C15, and C16. On the right side, there are two main sections: 'DEPENDENT' and 'INDEPENDENT'. Under the 'DEPENDENT' section, column C1 is listed, and there are 'ADD' and 'REMOVE' buttons. Under the 'INDEPENDENT' section, column C2 is listed, and there are 'ADD' and 'REMOVE' buttons. The 'CALCULATE' button is visible at the top left, and 'EXAMPLE 1' through 'EXAMPLE 5', 'CLEAR', and 'RANDOM EXAMPLE' are at the top right.

Hit the calculate button and go to the result tab to see the result

CALCULATE	EXAMPLE 1	EXAMPLE 2	EXAMPLE 3	EXAMPLE 4	EXAMPLE 5	CLEAR	RANDOM EXAMPLE
DATA	MAPPING	INPUT	RESULT	GRAPH	SELECT PARAMETERS	COMPARE CURVES	
<p>Sample Size = 54</p> <p>Independent Variable = C2</p> <p>Dependent Variable = C1</p> <p>Results of One-Way Analysis of Variance:</p> <p>Model:</p> <p>Degrees of Freedom = 2</p> <p>Residual Sum of Squares = 31738.715</p> <p>Mean Square Error = 15869.357</p> <p>Error:</p> <p>Degrees of Freedom = 51</p> <p>Residual Sum of Squares = 455248.785</p> <p>Mean Square Error = 8926.447</p> <p>Corrected Total:</p> <p>Degrees of Freedom = 53</p> <p>Residual Sum of Squares = 486987.500</p> <p>F-Value = 1.778</p> <p>P-Value = 0.1793246682371903</p> <p>R-Square = .065</p>							

3. Analyze the result.

-F test

For $\alpha=0.05$, F-value is not in the critical region ($>F(1-\alpha,1,54-2)$), hence we fail to reject the null hypothesis and conclude that there is no effect of meat type on the sodium content.

-P value

P-value is moderate thus we do not have strong evidence against the null hypothesis that there is no effect of meat type on the sodium content from this study.

Calories

1. Setup the Null hypothesis

H0: There is not effect of meat types on the calorie content

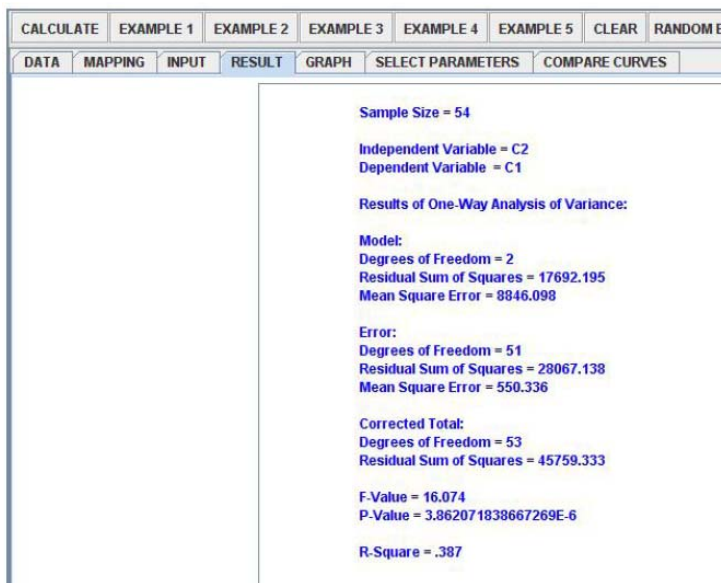
H1: There is effect of meat types on the calorie content

2. fit the ANOVA model

Go to the data tab and input the data

CALCULATE	EXAMPLE 1	EXAMPLE 2	EXAMPLE 3	EXAMPLE 4	EXAMPLE 5	CLEAR	RANDOM EXAMPLE
DATA	MAPPING	INPUT	RESULT	GRAPH	SELECT PARAMETERS	COMPARE CURVES	
		C1	C2	C3	C4	C5	C6
1:		186	A				
2:		181	A				
3:		176	A				
4:		149	A				
5:		184	A				
6:		190	A				
7:		158	A				
8:		139	A				
9:		175	A				
10:		148	A				
11:		152	A				
12:		111	A				

Map calorie as the dependent variable, hit calculate, and go to the result tab.



-F test

For $\alpha=0.05$, F-value is in the critical region ($>F(1-\alpha,1,54-2)$), we reject the null hypothesis and conclude that there is effect of meat type on the calorie content.

-P value

P-value is very small indicating strong evidence against the null hypothesis that there is no effect of meat type on the calorie content.

4. Use the [SOCR 1981-2007 Consumer Price Index Dataset](#) to explore the effect of the CPI item on the CPI-value, i.e., you can take the *item* as the categorical variable and the *values* as the observations (across months or/and years). Use the [SOCR 1-Way ANOVA applet](#) to carry the calculations. Provide all results and explicitly state your inference.

1. Setup the Null hypothesis

H0: There is not effect of CPI item on the CPI-value

H1: There is effect of CPI item on the CPI-value

2. fit the ANOVA model

Repeat the same process as problem 3.

3. Interpret the result

Sample Size = 48

Independent Variable = C2

Dependent Variable = C1

Results of One-Way Analysis of Variance:

Model:

Degrees of Freedom = 3

Residual Sum of Squares = 211590.919

Mean Square Error = 70530.306

Error:

Degrees of Freedom = 44

Residual Sum of Squares = 668.348

Mean Square Error = 15.190

Corrected Total:

Degrees of Freedom = 47

Residual Sum of Squares = 212259.267

F-Value = 4643.289

P-Value = < 1E-15

R-Square = .997

-F test

For $\alpha=0.05$, F-value is in the critical region ($>F(1-\alpha,1,48-2)$), we reject the null hypothesis and conclude that there is effect of CPI item on the CPI value.

-P value

P-value is very small indicating strong evidence against the null hypothesis that there is no effect of CPI item on CPI value.