Correlation and Regression HW questions.

Question 1:

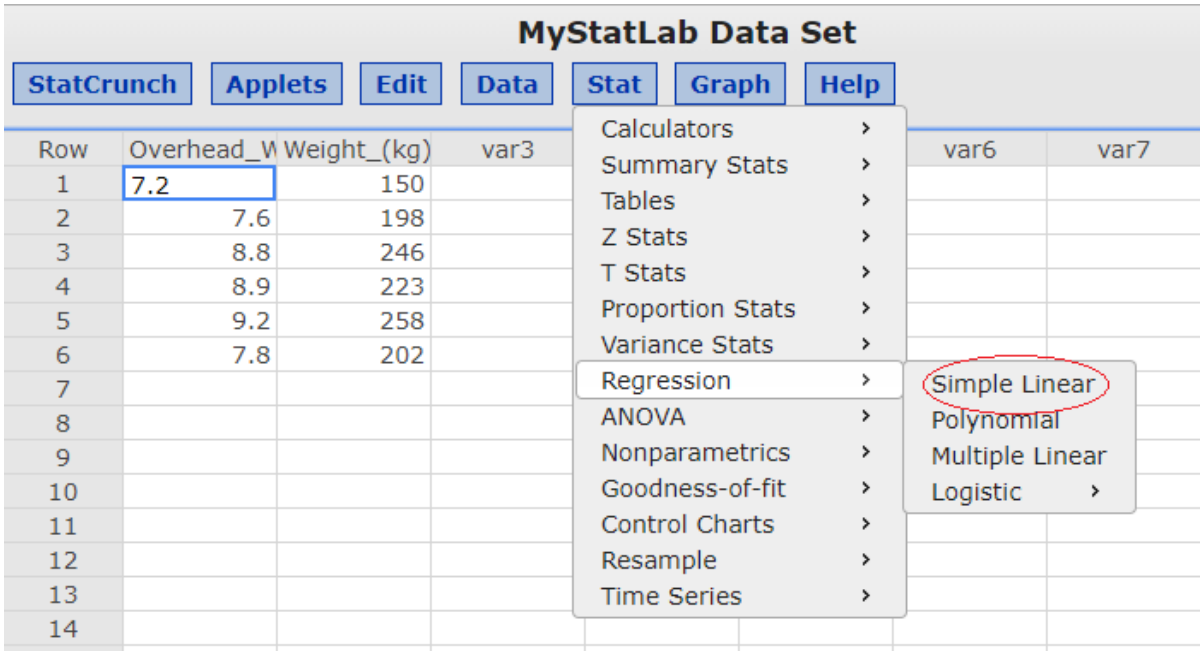
Find the regression equation, letting overhead width be the predictor (x) variable. Find the best predicted weight of a seal if the overhead width measured from a photograph is 2.1 cm. Can the prediction be correct? What is wrong with predicting the weight in this case? Use a significance level of 0.05.

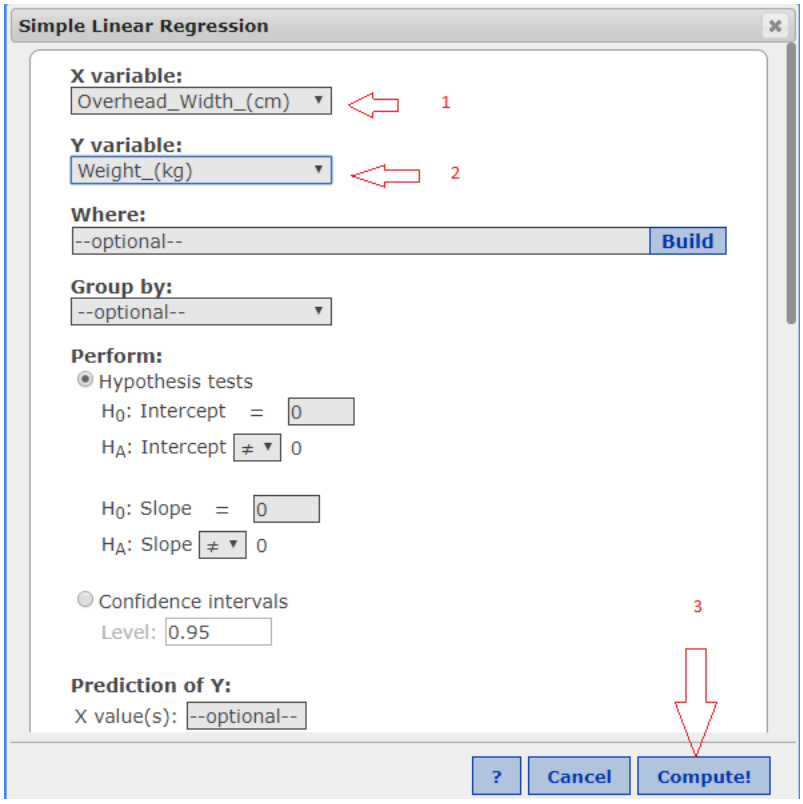
Overhead Width (cm)	7.2	7.6	8.8	8.9	9.2	7.8
Weight (kg)	150	198	246	223	258	202

Click to open StatCrunch

Click the icon to view the critical values of the Pearson correlation coefficient r.

The regression equation is $\hat{y} =$ ☐ + ☐x.
(Round to one decimal place as needed.)





Options (1 of 2)

Simple linear regression results:
Dependent Variable: Weight_(kg)
Independent Variable: Overhead_Width_(cm)
Weight_(kg) = -151.71585 + 44.187779 Overhead_Width_(cm) y = -151.7 + 44.2 x
Sample size: 6
R (correlation coefficient) = 0.93279154
R-sq = 0.87010006
Estimate of error standard deviation: 15.636456

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-value
Intercept	-151.71585	70.716816	≠ 0	4	-2.1453999	0.0985
Slope	44.187779	8.5367402	≠ 0	4	5.1761888	0.0066

Significant

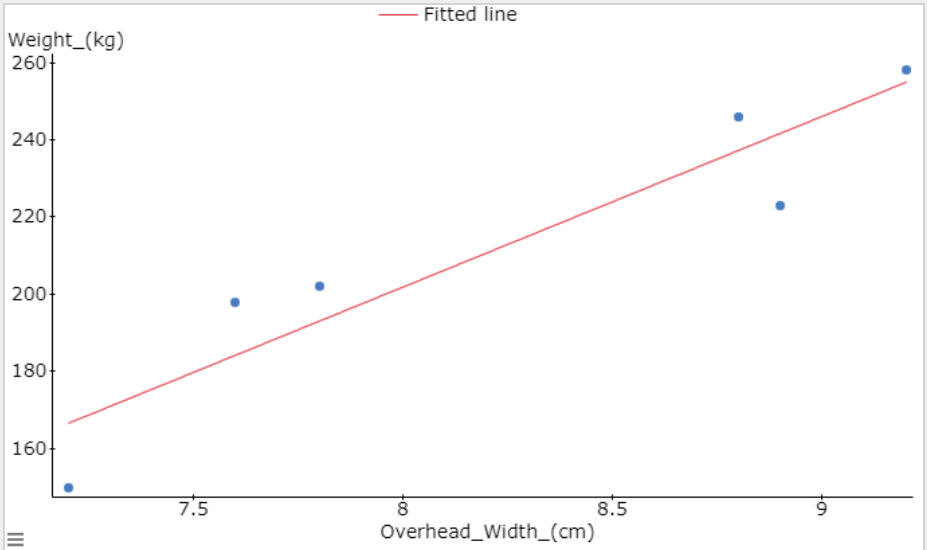
Analysis of variance table for regression model:

Source	DF	SS	MS	F-stat	P-value
Model	1	6550.8383	6550.8383	26.792931	0.0066
Error	4	977.99503	244.49876		
Total	5	7528.8333			

Scatterplot & regression line:

< >

Options (2 of 2)



< >

Find the regression equation, letting overhead width be the predictor (x) variable. Find the best predicted weight of a seal if the overhead width measured from a photograph is 2.1 cm. Can the prediction be correct? What is wrong with predicting the weight in this case? Use a significance level of 0.05.

Overhead Width (cm)	7.2	7.6	8.8	8.9	9.2	7.8
Weight (kg)	150	198	246	223	258	202

Click the icon to view the critical values of the Pearson correlation coefficient r.

The regression equation is $\hat{y}$ = −151.7 + 44.2 x.
(Round to one decimal place as needed.)

The best predicted weight for an overhead width of 2.1 cm is −58.9 kg.
(Round to one decimal place as needed.)

Can the prediction be correct? What is wrong with predicting the weight in this case?

- A. The prediction cannot be correct because there is not sufficient evidence of a linear correlation. The width in this case is beyond the scope of the available sample data.
- B. The prediction cannot be correct because a negative weight does not make sense. The regression does not appear to be useful for making predictions.
- ✔ C. The prediction cannot be correct because a negative weight does not make sense. The width in this case is beyond the scope of the available sample data.
- D. The prediction can be correct. There is nothing wrong with predicting the weight in this case.

Question is complete.

Notice that the slope p-value of 0.0066 < alpha of 0.05 (significance); therefore, we reject the Null hypothesis of no correlation between the variables –overhead width and weight—in favor of the alternative hypothesis that there is a correlation between the two variables. That is the reason why we use the regression eq. y = -151.7 + 44.2 x where x = 2.1 in order to calculate the "best predicted weight".

Question 2:

Find the regression equation, letting the first variable be the predictor (x) variable. Using the listed lemon/crash data, where lemon imports are in metric tons and the fatality rates are per 100,000 people, find the best predicted crash fatality rate for a year in which there are 500 metric tons of lemon imports. Is the prediction worthwhile?

Lemon Imports	227	260	360	460	535
Crash Fatality Rate	16	15.9	15.6	15.5	15.1

Find the equation of the regression line.

$\hat{y}$ = 16.604 + (−0.002671)x
(Round the constant three decimal places as needed. Round the coefficient to six decimal places as needed.)

The best predicted crash fatality rate for a year in which there are 500 metric tons of lemon imports is 15.3 fatalities per 100,000 population.
(Round to one decimal place as needed.)

Is the prediction worthwhile?

- ✔ A. Since common sense suggests there should not be much of a relationship between the two variables, the prediction does not make much sense.
- B. Since there appears to be an outlier, the prediction is not appropriate.
- C. Since all of the requirements for finding the equation of the regression line are met, the prediction is worthwhile.
- D. Since the sample size is small, the prediction is not appropriate.

Very important: Correlation does not imply causation.

Same procedure as question 1. See StatCruch output next page: again, the slope p value < alpha.

Options (1 of 2)

Simple linear regression results:
Dependent Variable: Crash Fatality Rate
Independent Variable: Lemon Imports
Crash Fatality Rate = 16.604081 - 0.0026712301 Lemon Imports Equation.
Sample size: 5
R (correlation coefficient) = -0.97703463
R-sq = 0.95459666
Estimate of error standard deviation: 0.087682944

Parameter estimates:

Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-value
Intercept	16.604081	0.12996614	≠ 0	3	127.75698	<0.0001
Slope	-0.0026712301	0.00033634471	≠ 0	3	-7.9419416	0.0042

Analysis of variance table for regression model:

Source	DF	SS	MS	F-stat	P-value
Model	1	0.4849351	0.4849351	63.074437	0.0042
Error	3	0.023064896	0.0076882986		
Total	4	0.508			

Question 3:

In this problem the p-value is not significant. Therefore, the regression equation, namely, y = 43.4 + 1.2 x is meaningless. The best predicted value in the left arm cannot be calculated using the equation; instead we use the mean value of the dependent variable, blood pressure left arm.

Listed below are systolic blood pressure measurements (in mm Hg) obtained from the same woman. Find the regression equation, letting the right arm blood pressure be the predictor (x) variable. Find the best predicted systolic blood pressure in the left arm given that the systolic blood pressure in the right arm is 95 mm Hg. Use a significance level of 0.05.

Right Arm	100	99	93	78	78
Left Arm	175	169	144	142	143

Click the icon to view the critical values of the Pearson correlation coefficient r

The regression equation is $\hat{y}$ = 43.4 + 1.2 x.
(Round to one decimal place as needed.)

Given that the systolic blood pressure in the right arm is 95 mm Hg, the best predicted systolic blood pressure in the left arm is 154.6 mm Hg.
(Round to one decimal place as needed.)

Options (1 of 2)

Simple linear regression results:
Dependent Variable: Left Arm
Independent Variable: Right Arm
Left Arm = 43.407376 + 1.2409891 Right Arm The equation is meaningless since there is no corr between the vars.
Sample size: 5
R (correlation coefficient) = 0.84502255
R-sq = 0.7140631
Estimate of error standard deviation: 9.9043131

Parameter estimates:

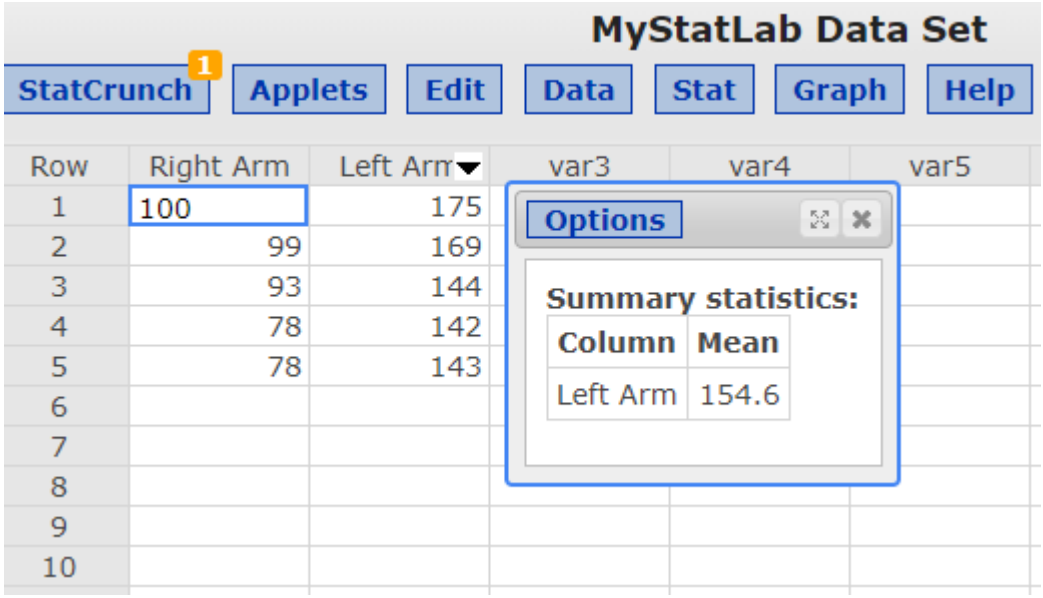
Parameter	Estimate	Std. Err.	Alternative	DF	T-Stat	P-value
Intercept	43.407376	40.864709	≠ 0	3	1.0622216	0.3661
Slope	1.2409891	0.4533923	≠ 0	3	2.7371199	0.0715

Analysis of variance table for regression model:

Source	DF	SS	MS	F-stat	P-value
Model	1	734.91375	734.91375	7.4918254	0.0715
Error	3	294.28625	98.095418		

P value greater than alpha. No significant. That is, we fail to reject the hypothesis of no corr between the two variables.

In order to calculate the mean of a column of values, click Stat, then Summary Statistics, choose column left arm, select mean, and calculate:



The mean blood pressure left arm is the best predicted value.