JMP INTRO® Lab Activities

Lab Activity – Inference for Regression

Data Set: Body Measurements.jmp

In *Gulliver's Travels*, the Lilliputians make an entire set of clothes for the (giant) Gulliver by taking only a few measurements from his body:
"The seamstresses took my measure as I lay on the ground, one standing at my neck, and another at my mid-leg, with a strong cord extended, that each held by the end, while a third measured the length of the cord with a rule of an inch long. Then they measured my right thumb, and desired no more; for by a mathematical computation, that twice round the thumb is once round the wrist, and so on to the neck and the waist, and by the help of my old shirt, which I displayed on the ground before them for a pattern, they fitted me exactly." (Swift, 1735)

In this lab, you are going to look at whether there is a relationship among the different parts of the body. You are going to work with a data set that contains measurements collected as part of a statistics project in Australia. These data are measurements from 22 randomly selected male subjects.

Open JMP INTRO, and from the **JMP Starter** menu, select **Open Data Table**. Open the **Body Measurements.jmp** file. The variable **Mass** is measured in kilograms, and all other variables are measured in centimeters. You will be constructing models that will attempt to predict the mass of a person based on other characteristics. You will conduct tests that will determine whether mass is associated with these other characteristics.

First you will examine body mass versus forearm length. Choose **Analyze** ® **Fit Y By X**. Enter the appropriate variables for the **Y,Response** and **X,Factor** locations, based upon the goal mentioned above. Choose **Fit Line** from the red triangle menu of **Bivariate Fit of Mass By Fore**. Examine the numbers that appear in the text box **Parameter Estimates**. For both the intercept and the slope of the line, there is an estimate of the parameter, its standard error, a t-ratio, and a corresponding p-value. You should also look at the residual plot to help determine if the conditions for using the regression test are reasonable. To examine the residual plot, from the red triangle **Linear Fit** menu in the **Bivariate Fit of Mass By Fore** window, choose **Plot Residuals**.

Repeat this procedure for body mass versus shoulder width.

Repeat the same procedure for body mass versus two other body measurements.

For each of the regression lines, conduct a hypothesis test using one of the commonly used significance levels to determine if there is a useful linear relationship between body mass and the other factor. Remember to include assumptions, hypotheses, sample statistic values, test statistic and p-value, and conclusions in the context of the problem. Include copies of the scatterplots and regression analysis from JMP INTRO for each pair of variables in your report. From the four regression analyses you have completed, determine which of these you consider to be the best choice for predicting body mass. Explain why.

Reference:

Swift, Jonathan (1735). *Gulliver's Travels*. Quote is from p. 44 of the *Norton Critical Edition*, (1961) Robert A. Greenwood, ed. New York: W. W. Norton & Co.

JMP INTRO® Lab Activities
Teacher Notes

Lab Activity – Inference for Regression

Data Set: Body Measurements.jmp

Objectives:
- Gain familiarity with some of the basic operational tools of JMP INTRO.
- Use JMP INTRO to conduct a hypothesis test for linear regression.
- Create a word processing document, incorporating displays from JMP INTRO.

Time Required: 45 minutes

Materials:
- Inference for Regression student activity directions
- Body Measurements.jmp data set

  Note: Body Measurements.jmp is a data set that is not included in the Sample Data folder for JMP INTRO. This data set can, however, be downloaded from SAS. Refer to the Teacher Notes from Hypothesis Testing – Two-Sample *t*-test for information on how to retrieve these data files.

Prerequisites:
- Students should have basic knowledge of how to use JMP INTRO.
- Students should have experience conducting tests of hypothesis by hand and/or with graphing calculators for linear regression.
- Students should have experience writing conclusions for hypothesis tests in the context of the problem.

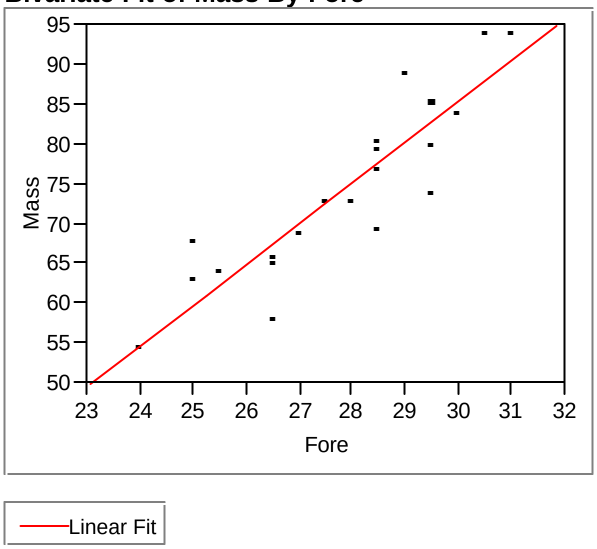
JMP INTRO Notes:
- JMP INTRO includes an extensive online help system. It contains a table of contents and/or can be used to search for a specific topic. In addition, JMP INTRO has context sensitive help. You can access it by selecting the help tool (?) from the tools toolbar and clicking inside a data table or report. JMP INTRO opens help specific to the clicked-on item.
- Students can convert measurements to the English system and use those values for the regression analysis.

JMP INTRO® Lab Activities
Answer Key

The following are suggested answers for Inference for Regression.

Scatterplot and Regression Analysis for Mass vs. Forearm Length

Bivariate Fit of Mass By Fore



Linear Fit

Mass = -68.6441 + 5.1336663 Fore

Summary of Fit

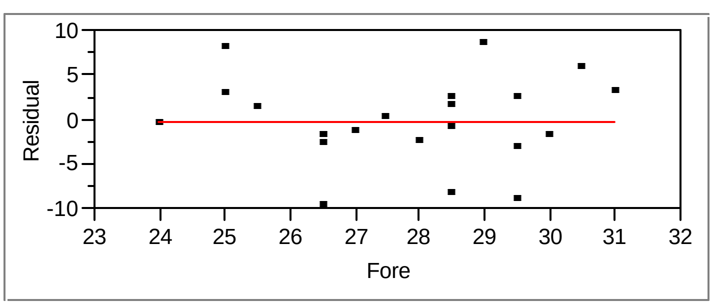
RSquare	0.807751
RSquare Adj	0.798138
Root Mean Square Error	4.925777
Mean of Response	73.93182
Observations (or Sum Wgts)	22

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2038.8822	2038.88	84.0316
Error	20	485.2655	24.26	Prob > F
C. Total	21	2524.1477		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-68.6441	15.58879	-4.40	0.0003
Fore	5.1336663	0.560024	9.17	<.0001



Assumptions: The scatterplot of the data shows a linear pattern, and the residual plot does not show a clear pattern. The variability of points does not appear to be changing with forearm lengths, and assuming that the distribution of errors at any given forearm length value is approximately normal, the assumptions of the simple linear regression model are appropriate.

Hypotheses: β = the true average change in kilograms of body mass with a one centimeter increase in forearm length.

$H_o: b = 0$ There is no linear relationship between body mass and forearm length.

$H_a: b \neq 0$ There is a linear relationship between body mass and forearm length.

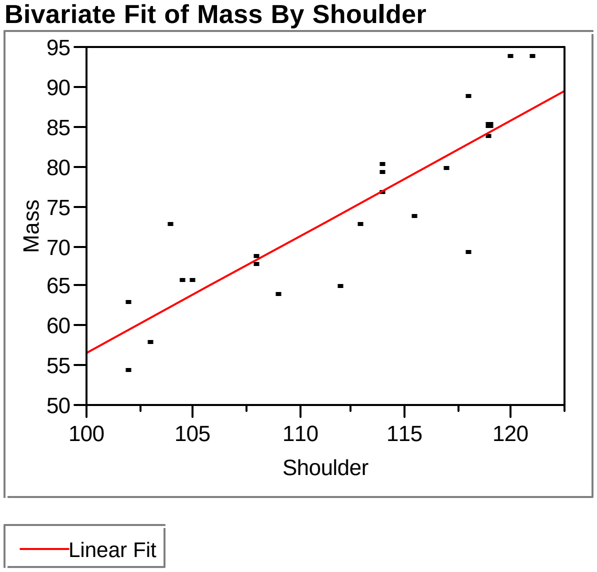
Test Mechanics: $$t = \frac{b - \text{hypothesizd value}}{s_b} = \frac{5.134}{0.56} = 9.167$$

$p < 0.0001$

Conclusion: At the $\alpha = 0.05$ level, because $p < 0.0001 < 0.05$, we can reject the null hypothesis. We have evidence that there is a linear relationship between body mass and forearm length.

Scatterplot and Regression Analysis for Body Mass vs. Shoulder Width:

Bivariate Fit of Mass By Shoulder



Linear Fit

Mass = -90.60311 + 1.4714506 Shoulder

Summary of Fit

RSquare	0.734066
RSquare Adj	0.720769

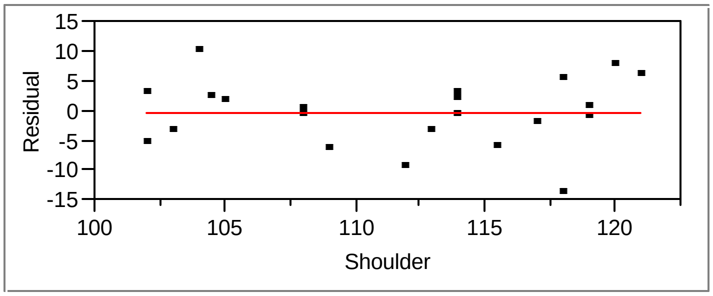
Root Mean Square Error	5.793345
Mean of Response	73.93182
Observations (or Sum Wgts)	22

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1852.8907	1852.89	55.2066
Error	20	671.2570	33.56	Prob > F
C. Total	21	2524.1477		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-90.60311	22.17875	-4.09	0.0006
Shoulder	1.4714506	0.198039	7.43	<.0001



Assumptions: The scatterplot of the data shows a linear pattern, and the residual plot does not show a clear pattern. The variability of points does not appear to be changing with shoulder width values, and assuming that the distribution of errors at any given shoulder width value is approximately normal, the assumptions of the simple linear regression model are appropriate.

Hypotheses: β = the true average change in kilograms of body mass with a one centimeter increase in shoulder width.

$H_o: b = 0$ There is no linear relationship between body mass and shoulder width.

$H_a: b \neq 0$ There is a linear relationship between body mass and shoulder width.

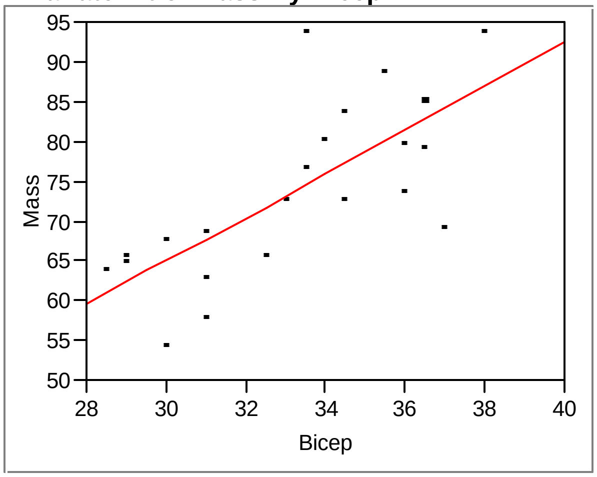
Test Mechanics: $t = \dfrac{b - \text{hypothesizd value}}{s_b} = \dfrac{1.4715}{0.198} = 7.43$

$p < 0.0001$

Conclusion: At the $\alpha = 0.05$ level, because $p < 0.0001 < 0.05$, we can reject the null hypothesis. We have evidence that there is a linear relationship between body mass and shoulder width.

Scatterplots and Regression Analysis for Body Mass vs. Other Variables:

Bivariate Fit of Mass By Bicep



Linear Fit

Linear Fit

Mass = -17.128 + 2.7423903 Bicep

Summary of Fit

RSquare	0.528354
RSquare Adj	0.504772
Root Mean Square Error	7.71526
Mean of Response	73.93182
Observations (or Sum Wgts)	22

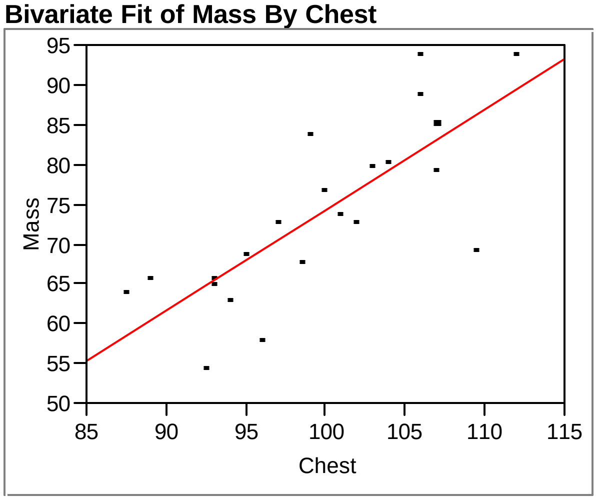
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1333.6431	1333.64	22.4047
Error	20	1190.5046	59.53	Prob > F
C. Total	21	2524.1477		0.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-17.128	19.30809	-0.89	0.3856
Bicep	2.7423903	0.579375	4.73	0.0001

Bivariate Fit of Mass By Chest



Linear Fit

Linear Fit
Mass = -51.95744 + 1.2634871 Chest

Summary of Fit

RSquare	0.602151
RSquare Adj	0.582258
Root Mean Square Error	7.086008
Mean of Response	73.93182
Observations (or Sum Wgts)	22

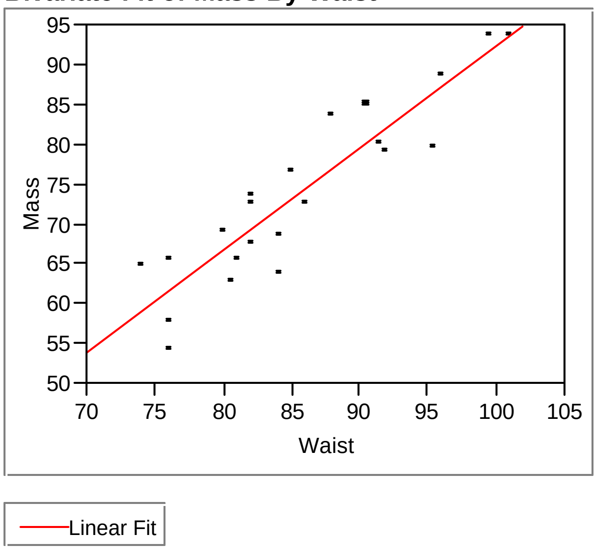
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1519.9175	1519.92	30.2703
Error	20	1004.2302	50.21	Prob > F
C. Total	21	2524.1477		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-51.95744	22.9311	-2.27	0.0347
Chest	1.2634871	0.229648	5.50	<.0001

Bivariate Fit of Mass By Waist



Linear Fit
Mass = -36.19109 + 1.2869609 Waist

Summary of Fit

RSquare	0.837369
RSquare Adj	0.829238
Root Mean Square Error	4.530477
Mean of Response	73.93182
Observations (or Sum Wgts)	22

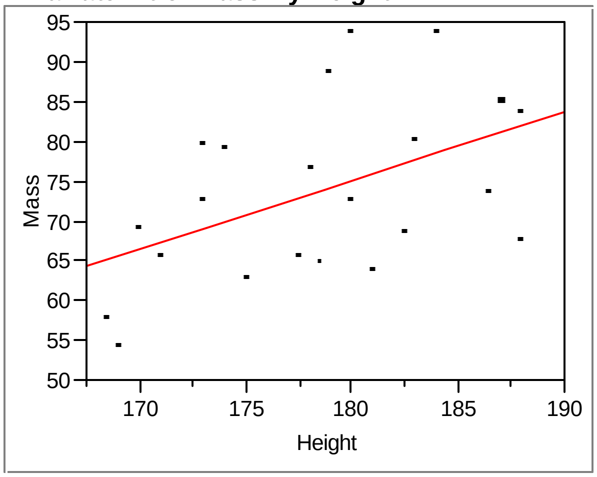
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	2113.6432	2113.64	102.9778
Error	20	410.5045	20.53	Prob > F
C. Total	21	2524.1477		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-36.19109	10.8948	-3.32	0.0034
Waist	1.2869609	0.126822	10.15	<.0001

Bivariate Fit of Mass By Height



Linear Fit

Linear Fit

Mass = -81.32534 + 0.8698988 Height

Summary of Fit

RSquare	0.242005
RSquare Adj	0.204105
Root Mean Square Error	9.780826
Mean of Response	73.93182
Observations (or Sum Wgts)	22

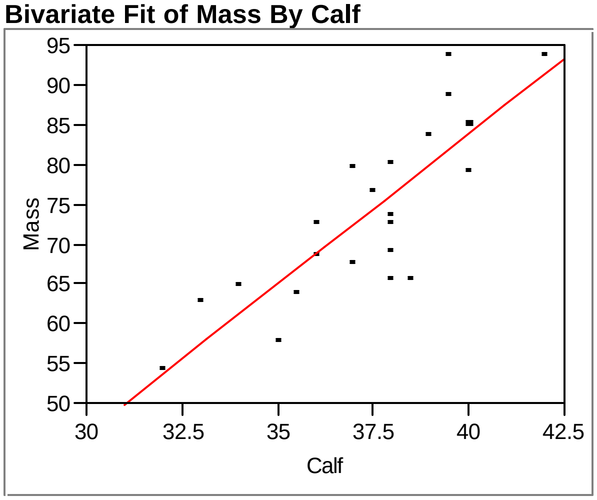
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	610.8568	610.857	6.3854
Error	20	1913.2910	95.665	Prob > F
C. Total	21	2524.1477		0.0200

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-81.32534	61.47626	-1.32	0.2008
Height	0.8698988	0.34425	2.53	0.0200

Bivariate Fit of Mass By Calf



Linear Fit

Linear Fit
Mass = -66.98282 + 3.7737334 Calf

Summary of Fit

RSquare	0.692226
RSquare Adj	0.676838
Root Mean Square Error	6.23244
Mean of Response	73.93182
Observations (or Sum Wgts)	22

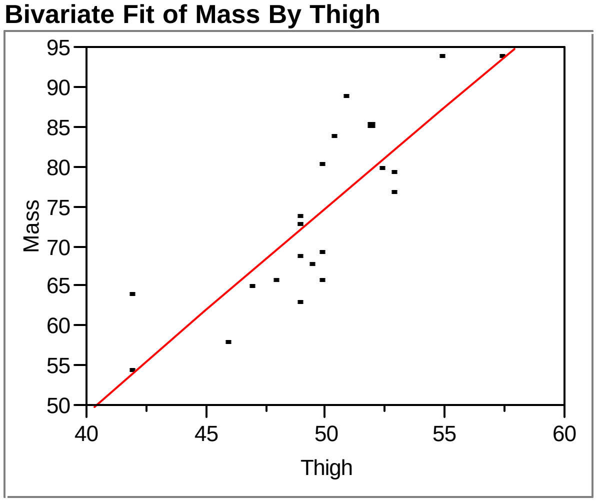
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1747.2815	1747.28	44.9828
Error	20	776.8663	38.84	Prob > F
C. Total	21	2524.1477		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-66.98282	21.0523	-3.18	0.0047
Calf	3.7737334	0.562662	6.71	<.0001

Bivariate Fit of Mass By Thigh



Linear Fit

Linear Fit
Mass = -53.8098 + 2.5688442 Thigh

Summary of Fit

RSquare	0.709435
RSquare Adj	0.694906
Root Mean Square Error	6.055699
Mean of Response	73.93182
Observations (or Sum Wgts)	22

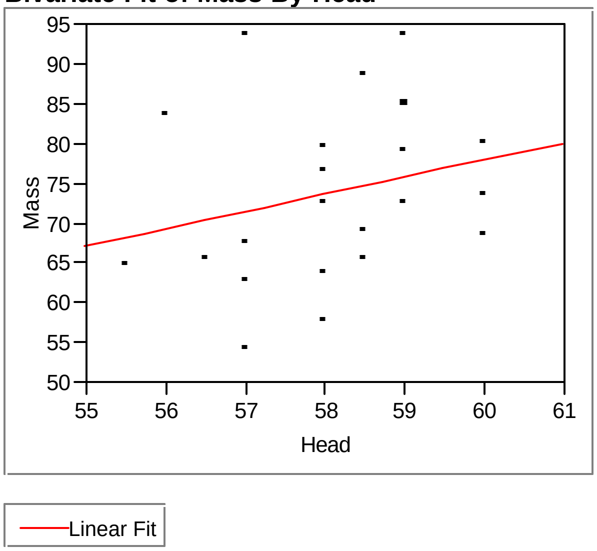
Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	1790.7180	1790.72	48.8313
Error	20	733.4298	36.67	Prob > F
C. Total	21	2524.1477		<.0001

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-53.8098	18.32583	-2.94	0.0082
Thigh	2.5688442	0.367611	6.99	<.0001

Bivariate Fit of Mass By Head



Linear Fit

Mass = -51.06308 + 2.152554 Head

Summary of Fit

RSquare	0.060848
RSquare Adj	0.013891
Root Mean Square Error	10.88705
Mean of Response	73.93182
Observations (or Sum Wgts)	22

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	1	153.5896	153.590	1.2958
Error	20	2370.5581	118.528	Prob > F
C. Total	21	2524.1477		0.2684

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-51.06308	109.8294	-0.46	0.6470
Head	2.152554	1.890965	1.14	0.2684