

Appendix

Inference for Regression

Most statistical packages give the t statistics and P -values for testing $H_0: \beta_1 = 0$ against $H_a: \beta_1 \neq 0$ and $H_0: \beta_0 = 0$ against $H_a: \beta_0 \neq 0$ as part of the standard regression output. See the Chapter 2 Appendix for how to obtain those statistics. Note: The test of the intercept is usually not of interest because considering values for $x = 0$ would be extrapolation. The ANOVA table is also produced.

Regression is somewhat unusual in statistics because checking conditions for inference requires the regression be done first (so you have residuals). Check the conditions before using any inference results:

1. A line is the correct model (is the scatterplot of the data “linear”?). See the Chapter 2 Appendix on creating scatterplots if this is not produced as part of the regression output.
2. The residuals have a Normal distribution. Check a Normal quantile plot of the residuals (see the Chapter 1 Appendix or below).
3. The variation of the data points around the line is constant. Check a scatterplot of the residuals against the X (predictor) variable or fitted values (see the Chapter 2 Appendix or below). Also examine this plot and the original scatterplot for any outliers.

Inference for the Regression Slope and Assessing Conditions



1. Click **Data** → **Data Analysis**.
2. Select **Regression** from the menu box.
3. Click **OK**.
4. Either input or drag to highlight the Y range into its box and the X range into its box. If these ranges include labels, check that box.
5. For confidence intervals at other than 95% confidence, check the box and change from 95%.
6. To check the assumption of constant variance, check the box for **Residual Plots** (produces a plot of the residuals against fitted values). This also stores the residuals into the output.
7. Click **OK**.
8. To check the assumption that the residuals have a Normal distribution, produce a Normal plot of those as detailed in the Chapter 1 Appendix. The “Normal Probability Plot” option in the regression dialog produces a normal plot of the y variable—that is NOT an assumption. To check the assumption of a linear relationship, use the residual plot and/or a scatterplot of the data (see the Chapter 2 Appendix).

Output includes the t statistics, P -values, confidence intervals for the slope and intercept, as well as the ANOVA table.

For more information, see the Excel Video Technology Manual videos “CI for Slope” and “Inference and CI for Slope.”



There are two ways to perform a simple linear regression analysis in JMP. In addition to the method detailed below, you can also use **Help → Teaching Demos → Demonstrate Regression** and select the option to use your own data.

1. Click **Analyze → Fit Y by X**.
2. Select the dependent (y) variable into the “Y, Response” box and the predictor (x) variable into the “X, Factor” box.
3. Click **OK**. This produces a scatterplot of the data along with some summary statistics on each variable.
4. From the red triangle menu, select **Fit Line**. This gives the basic regression, t statistics, and two-sided P -values along with r^2 and s (labeled Root Mean Square Error) and the ANOVA table.
5. To assess the model with plots, click the red triangle next to Linear Fit. Select **Plot Residuals**. This produces a plot of the residuals against fitted values with a histogram of the residuals, a plot of the original y data against predicted \hat{y} values, a residual by row plot (useful in detecting whether order might be a concern), a plot of residuals against x -values, and a Normal quantile plot of the residuals.

An alternate method is to use **Applets, → Demonstrate Regression**.

1. Move the radio button to “My Sample.”
2. A box to select the appropriate data table will appear. Once you have selected that, a new box will appear to allow you to select the variables. Click to enter each into the appropriate box.
3. Click **OK**.
4. Use the red triangle next to “Sample Data” to select “Fit LS Line.”

JMP does not automatically compute a confidence interval for the slope. You can compute it given the estimated slope, standard error, and t_{n-2}^* . For the critical value, use either a table or **Applets → Distribution Calculator**. In the calculator, use the pull-down menu to select t distribution and enter the appropriate degrees of freedom. Move the radio button to “Input probability and calculate values.”

For more information, see the JMP Video Technology Manual videos “Linear Regression: Fitted Line Plots, Residual Plots and Prediction” and “Inference and CI for Slope.”



1. Click **Stat → Regression → Regression → Fit Regression Model**.
2. Select and enter the y (dependent) variable into the box labeled Responses and the x (predictor) variable into the box labeled Continuous predictors.
3. For plots to assess the model, click **Graphs**. It is easiest to click the radio button on this screen to select the “Four in one” plot, which produces all residuals plots of interest. Click **OK** to return to the main regression dialog.
4. To display the confidence interval for the slope and intercept, click **Results**. Change the drop-down at the top from “Simple tables” to “Expanded tables.” Click **OK**.

5. To change the confidence level for the confidence interval for the slope and intercept from the standard 95%, click **Options** and enter a value in the box labeled Confidence level for all intervals. Click **OK** to return to the main regression dialog.
6. Click **OK**. Residuals plots will appear below the regression output. Scroll down to assess the assumptions before using the regression model results.

The output displays the coefficients table first, then the model summary (r^2 and the standard error of the residuals, labeled S), then the ANOVA table.

For more information, see the Minitab Video Technology Manual video “Linear Regression: Inference for Slope.”



1. Click **Analyze** → **Regression** → **Linear**.
2. Select the dependent (y) variable and the independent (predictor, x) variable into their boxes.
3. For a confidence interval for the slope, click **Statistics**. Check the box next to “Confidence intervals.” If a confidence level other than the default 95% is desired, change that.
4. To produce a histogram and Normal probability plot of standardized residuals (these are the residuals divided by their estimated standard deviation, e_i/s), click **Plots**. Check the box labeled “Produce all partial plots” or click to place a check into the box by either plot.
5. To produce other plots with the residuals (like a plot of the residuals against the predictor variable), click **Save**. Check the box labeled “Unstandardized” in the Residuals section.
6. Click **OK**. The confidence interval (if selected) will be at the far right in the Coefficients table in the output.

The output first displays the Model Summary (which includes r , r^2 , and the standard deviation of the residuals, labeled “Std. Error of the Estimate”), then the ANOVA table, and finally the regression coefficients with their confidence intervals (if requested) and the t statistics with P -values.

For more information, see the SPSS Video Technology Manual videos “CI for Slope” and “Inference and CI for Slope.”



1. Click **Statistics** → **Regression** → **Simple Linear**.
2. Use the drop-down boxes to select the dependent (y) and independent (x) variables.
3. Use the next drop-down to select the type of results to display (for a full analysis, you will need to use all four options):
 - a. “Numeric Results” gives the fitted equation, along with a table giving the standard errors and t statistics, along with P -values.
 - b. “Fitted Plot” produces a graph of the data with the regression line but no numeric results.
 - c. “Residuals Plot” produces a plot of the residuals against the predictor (x) variable.
 - d. “Residuals QQ Plot” produces a Normal probability plot of the residuals.
4. For a confidence interval for the slope, you must select “Numeric Results” in the drop-down and enter a confidence level in decimal form.
5. Click **Calculate**.

For more information (and an example), see the Crunchit! Help Video “Simple Linear Regression.”



TI-83/-84

For the hypothesis test, press arrow to Tests and select the LinRegTTest option (the number/letter varies with the model). Specify the data lists, the form of the alternate hypothesis, and a regression equation to store the results (press [VARS] [D] [ENTER] [ENTER] to select Y1) and select Calculate.

The calculator automatically stores the residuals in a list called RESID that can be accessed from the [2nd] [STAT] = LIST menu. You can then plot these against the predictor (x) variable or create a Normal probability plot using the methods described in Chapters 1 and 2 of this Appendix.

TI-84 calculators can calculate the confidence interval for the slope using LinRegTInt. Enter the data lists, the confidence level, and a Y-var to store the regression equation (if desired) and press Calculate.

For more information, see the TI Video Technology Manual videos “CI for Slope” and “Inference and CI for Slope.”



The basic regression command is of the form

> model <-lm(y~x)

To obtain the t statistics and P -values, follow this with the command

> summary(model)

For the confidence intervals, use the command

> confint(model)

Residuals are stored in a list named, using the above example, resid(model). This can be used in regular plotting commands. For example, to create a normal quantile (probability) plot of these, use the command

> qqnorm(resid(model))

For more information, see the R Video Technology Manual “Linear Regression: Inference for the Slope.”

Confidence and Prediction Intervals for Responses



Excel has no built-in function to calculate these intervals.



If the value of the explanatory variable (x) of interest is in the data set, you need do nothing. If not, enter that value at the bottom of the data spreadsheet with nothing for the response (that “observation” will be ignored when the regression is calculated).

For these intervals, click the red triangle next to Linear Fit just above the results of the regression from the **Fit Y by X** platform.

If needed, change the α level (1 – Confidence) using **Set α level**.

For the confidence interval for a mean response, click on **Mean Confidence Limit Formula**.

For a Prediction Interval, click on **Indiv Confidence Limit Formula**.

JMP will calculate these intervals for every data point and add two columns (corresponding to the lower and upper limits) to the spreadsheet.

For a video that shows how to use JMP with an example, see JMP Video Technology Manual: “Prediction and Confidence Intervals for $X = x$.”



After a regression has been fit, click **Stat → Regression → Regression → Predict**.

There are two methods to input the predictor value(s) for which you want intervals:

- a. Use the drop-down to select **Enter individual values** and enter the value(s) for the predictor variable in the spaces below, or
- b. select **Enter columns of values** and then specify the column name.

Use **Options** to change the confidence level.

For a video that shows how to use Minitab with an example, see Minitab Video Technology Manual: “Prediction and Confidence Intervals for $X = x$.”



If the value of the explanatory variable (x) of interest is in the data set, you need do nothing. If not, enter that value at the bottom of the data spreadsheet with nothing for the response (that will be ignored when the regression is calculated).

Before clicking OK to compute the regression, click **Save**. Check the box for either Prediction Intervals Mean (for the confidence interval) or Prediction Intervals Individual. Change the confidence level if needed.

For a video that shows how to use SPSS with an example, see SPSS Video Technology Manual: “Prediction and Confidence Intervals for $X = x$.”



Crunchit! only calculates Prediction Intervals for a single response. To obtain these intervals, enter the value of the predictor (x) variable in the box labeled “Predict” along with the confidence level for the prediction interval.



TI-83/-84

TI calculators do not have this capability as a built-in function.

For a video that shows how to do these calculators with a downloadable program (requires a connect cable) with an example, see TI Video Technology Manual: “Prediction and Confidence Intervals for $X = x$.”



To obtain these intervals, use a command of the form shown below where “prediction” can be replaced with “confidence.”

```
> predict(model,data.frame(x=10),interval="prediction")
```

These default to 95%. To change that, add “level=0.90” (for example) to that command.

For a video that shows how to use R with an example, see R Video Technology Manual: “Linear Regression: Prediction Intervals and Confidence Intervals for the Mean Response.”