

CHAPTER 23 Appendix

Inference for Regression

This chapter is primarily about inference for the slope of a linear regression. Most statistical packages give the t -statistic and P -value for testing $H_0: \beta_1 = 0$ against $H_a: \beta_1 \neq 0$ as part of the standard regression output. The appendices for Chapters 3 and 4 explain how to obtain those statistics.

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 3 appendix on creating scatterplots if this graph is not produced as part of the regression.
2. The residuals have a Normal distribution. Check a Normal quantile plot of the residuals (see the Chapter 11 Appendix or the remainder of this appendix).
3. The variation of the data points around the line is constant. Check a scatterplot of the residuals against the x variable or fitted values (see the Chapter 3 Appendix or the remainder of this appendix). Also examine this plot and the original scatter plot for any outliers or “fan” shapes.

The chapter also includes instructions for making predictions about the value of the response (y) variable, given a value of the predictor (x) variable. Predictions must account for sampling variability; we have uncertainty about the population means of the two variables, as well as uncertainty about the true population regression line. There are two situations here that require slightly different formulas:

1. We want to know the average value of the response, so we calculate a *confidence interval* for the mean response.
2. We want to know the possible values for one particular (new) observation, so we calculate a *prediction interval*. Prediction intervals are wider than confidence intervals because we must also account for the variability of observations around the line.

Note: Some of these technologies do not have built-in functions to calculate these intervals. If you are using Excel, TI calculators, or CrunchIt!, you will have to calculate the intervals using the formulas.

Inference for the Regression Slope and Assessing Conditions



Excel

1. **Data → Data Analysis**
2. Select **Regression** from the menu box.
3. **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 a level other than 95% confidence, check the box and change the value 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 option also stores the residuals into the output.
7. **OK**

TA23-1

8. To check the assumption that the residuals have a Normal distribution, produce a Normal plot of them as detailed in the Chapter 11 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 3 Appendix).

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 here, you can use **Help → Teaching Demos → Demonstrate Regression** and select the option to use your own data.

1. **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. **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”).
5. To assess the model with plots, click the Red Triangle next to **Linear Fit**. Select **Plot Residuals**. It 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.

JMP does not automatically compute a confidence interval for the slope. You can compute this interval given the estimated slope, standard error, and t_{n-2}^* . For the critical value, either use a table or select **Help → Teaching Demos → Distribution Calculator**.

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



Minitab

1. **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**. You will see the residuals plots before the regression output; either close or minimize that window after examining the plots for any violations of the assumptions.

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



1. **Analyze → Regression → Linear**
2. Select the dependent (y) variable and the independent (predictor, x) variable into the appropriate 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 value. Click **Continue**.
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**. Click **Continue**.
5. To produce other plots with the residuals (such as a plot of the residuals against the predictor variable), click **Save**. Check the box labeled **Unstandardized** in the Residuals section. Click **Continue**.
6. **OK**. The confidence interval (if selected) will appear at the far right in the Coefficients table in the output.

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



1. **Statistics → Regression → Simple Linear**
2. Use the drop-down boxes to select the response (y) and predictor (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. To obtain a confidence interval for the slope, select **Numeric Results** in the drop-down and enter a confidence level in decimal form.
5. **Calculate**

For more information (and an example), see the CrunchIt! Help Video *Simple Linear Regression*.



TI-83/-84

For the hypothesis test, press $\boxed{\text{STAT}}$, arrow to **Tests**, and select the **LinRegTTest** option (the number/letter varies with the model). Specify the data lists, the form of the alternative hypothesis, a regression equation in which to store the results (press $\boxed{\text{VARS}} \rightarrow \boxed{\text{ENTER}} \rightarrow \boxed{\text{ENTER}}$ to select Y1), and select **Calculate**.

The calculator automatically stores the residuals in a list called RESID that can be accessed from the $\boxed{2\text{nd}} \boxed{\text{STAT}} = \text{LIST}$ menu. You can then plot these data against the predictor (x) variable or create a Normal probability plot using the methods described in the appendices for Chapters 3 and 11.

TI-84 calculators can calculate the confidence interval for the slope using option G:LinRegTInt from the Stat Tests menu. Enter the data lists, the confidence level, a y variable 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 has 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 list 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

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



If the value of the explanatory variable (*x*) of interest is included in the data set, you do not need to do anything. 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).

For all the following commands, click the red triangle next to **Linear Fit** just above the results of the regression.

For the actual \hat{y} values, click **Save Predicted**. A new column will be added to the worksheet.

If needed, check the α level ($1 - \text{Confidence}$) 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 $X = x$* .



Minitab

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

You can use either of two methods to input the predictor value(s) for which you want intervals:

1. Use the drop-down to select **Enter individual values** and enter the value(s) for the predictor variable in the spaces below.
2. Select **Enter columns of values** and then specify the column name.
3. Use **Options** to change the confidence level. Click **OK** to return to the main dialog.
4. **OK**

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 included in the data set, you do not need to do anything. 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, then click **Continue**.

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 perform 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 following form, where “prediction” can be replaced with “confidence”:

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

These intervals default to a 95% confidence level. To change that level, add **level=0.90** (for example) to the 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*.