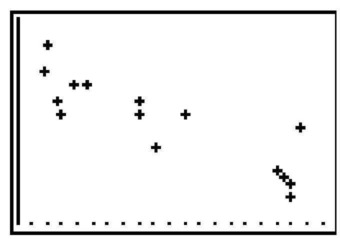
Chapter 13 – Inference for Regression

Computing a regression equation and looking at residuals plots is not the end of the story. We might want to know if the slope (or correlation) is meaningfully different from 0. It's not always apparent that a slope is meaningfully non-zero. Consider these two equations for the selling price of a house: $price = 25 + 0.061 * sqft$ and $price = 25000 + 61 * sqft$. At first blush one might look at the small value for the slope in the first equation and believe it's reasonable to say the true slope may in fact be 0; however the difference is in the units – the first has price measured in thousands of dollars, the second in dollars. They're really the same line. In addition, we'd like to (perhaps) make a confidence interval for a "true" slope just as we did for means and proportions as well as confidence intervals for the average value of *y* for a given *x* and prediction intervals for a new *y* observation for an *x* value. TI-83/84 calculators can perform the t-test on the slope as a native function. The other functions can be performed either using the calculator output and tables for the *t*-distribution or with a program which is included in this manual. TI-84 calculators have a built-in function to compute confidence intervals for the slope; the TI-89 can do those as well as confidence intervals for a response

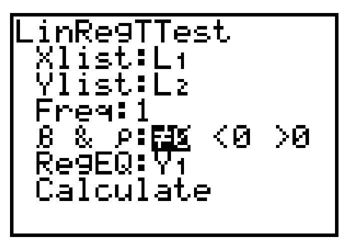
Returning to a problem considered before, here are advertised horsepower ratings and expected gas mileage for several model year 2007 vehicles.

Audi A4	200 hp	32 mpg	Honda Accord	166	34
BMW 328	230	30	Hyundai Elantra	138	36
Buick LaCrosse	200	30	Lexus IS 350	306	28
Chevy Cobalt	148	32	Lincoln Navigator	300	18
Chevy Trailblazer	291	22	Mazda Tribute	212	25
Ford Expedition	300	20	Toyota Camry	158	34
GMC Yukon	295	21	VW Beetle	150	30
Honda Civic	140	40			

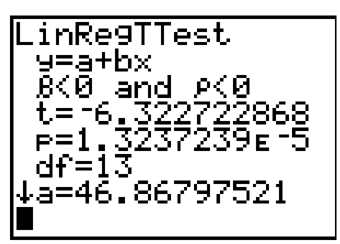
How is horsepower related to gas mileage? Recall the plot that was constructed for this data in Chapter 5. It is reproduced at right. The trend is decreasing. The residuals plots in Chapter 5 showed no overt pattern against *X* (horsepower) and the normal probability plot was reasonably straight. Inference for the regression is therefore appropriate.

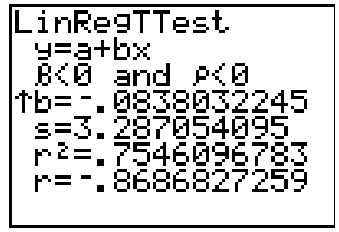


Press STAT, arrow to TESTS and select choice E:LinRegTTest. You tell the calculator which list contains the *x* (predictor variable) values, which contains the *y* (response) values. Freq is normally set to 1. Indicate the appropriate form of the alternate hypothesis. Notice there is an option to store the equation of the line. To store the equation as a function (here, Y_1), press VARS, arrow to Y-VARS, press ENTER to select Function, and ENTER to select Y_1. Finally, with the highlight on Calculate, press ENTER. The TI-89 input screen is similar, but there is an additional Draw option which, as usual shades the area under the *t* distribution corresponding to the p-value of the test.

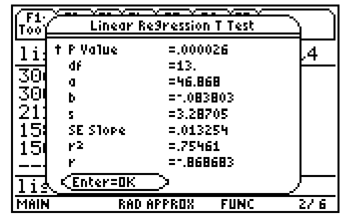


This is the first portion of the output (notice the ↓ at the bottom left). The first lines indicate the form of the regression so that you are reminded which quantity is the slope (b) and which the intercept (a) and the form of the alternate hypothesis in the test. The computed *t*-statistic for this regression is –6.32 and the p-value for the test is 0.00001, with 13 degrees of freedom. We will reject the null hypothesis and conclude not only that the slope is not zero; it is significantly negative. The intercept for the regression is 46.87.



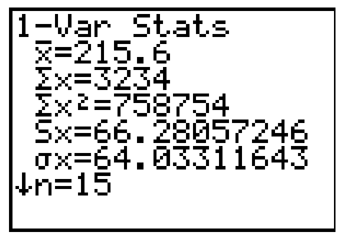


Pressing the down arrow several times we find the rest of the output. The slope is –0.084. We already know this is significantly different from zero even though its value seems small. The standard deviation of the data points around the line is 3.29. The relationship is strongly negative since r^2 = 75.5% and r = -0.869.



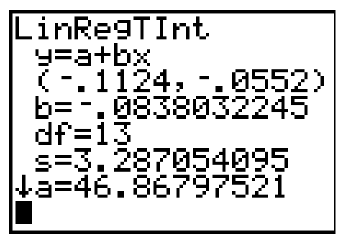
The output on a TI-89 includes all the quantities described above. It also gives the standard error of the slope (used in computing confidence intervals).

A CONFIDENCE INTERVAL FOR THE SLOPE



Confidence intervals (for any quantity) are always $estimate \pm (criticalvalue)(SE(estimate))$. In this case the critical value of interest will be a t statistic based on 12 degrees of freedom. From tables, we find this is 2.179 for 95% confidence. The standard error of the slope is $SE(b_1) = \frac{s(e)}{\sqrt{\sum (x-\bar{x})^2}} = \frac{s(e)}{\sqrt{n-1} * s(x)}$.

Using `1-Var Stats` for our horsepower data in `L1`, we find Sx is 66.281. We have all the pieces we need. $SE(b_1) = \frac{3.29}{\sqrt{14} * 66.281} = 0.0133$. When computing this, be sure to enclose the denominator in parentheses and close the parentheses for the square root. Putting all the pieces together, the 95% confidence interval for the slope is $-0.084 \pm 2.160 * 0.0133$ or (-0.113, -0.055). Based on this regression, we are 95% confident average gas mileage decreases between 0.055 and 0.113 miles per gallon for each horsepower in the engine.



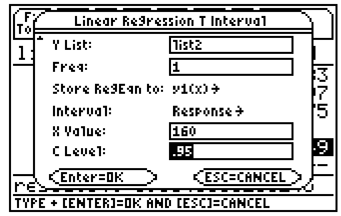
The TI-84 and -89 calculators can automatically compute the confidence interval for the slope. From the `STAT TESTS` menu, select `G:LinRegTInt`. This is option 7 on the `Ints` menu on the TI-89. The input screen is the same as for the t-test, except that it asks for the amount of confidence. From the output at right, we see we are 95% confident that gas mileage decreases between 0.055 and 0.112 miles per gallon for each horsepower in the engine. Notice this result is just slightly different from that obtained above due to rounding.

A CONFIDENCE INTERVAL FOR THE MEAN AT SOME *X*

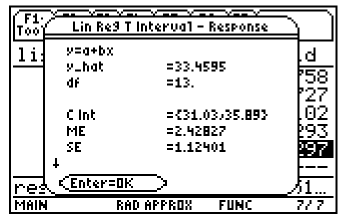
What should we predict as the average gas mileage for a vehicle with 160 horsepower? Evaluating the equation for 160 horsepower gives 33.46 miles per gallon. This is just a point estimate, however and is subject to uncertainty just as any mean is. Confidence intervals account for this uncertainty. In this case there are two sources – average variation around the line as well as uncertainty about the slope which makes estimation more "fuzzy" further away from the mean. Both of these are accounted for in the equation of the standard error, $SE(\hat{\mu}_v) = \sqrt{s^2(b_1) * (x_v - \bar{x})^2 + \frac{s^2(e)}{n}}$.

Putting everything together, we find $SE(\hat{\mu}_V) = \sqrt{0.0133^2 * (160 - 215.6)^2 + \frac{3.29^2}{15}} = 1.126$. The t critical value is still 2.16, so the confidence interval is 33.46± 2.16*1.126 or (31.03, 35.89). Based on this regression, we estimate with 95% confidence the average gas mileage for vehicles with 160 horsepower will be between 31.03 and 35.89 miles per gallon.

Using a TI-89, option `7:LinRegTInt` on the `Ints` menu will do all the calculations for us. Specify the lists containing the data, and select Response as the type of interval. Specify the x-value of interest, and the confidence level.



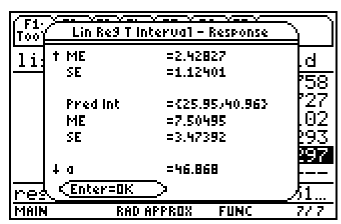
The first portion of output gives $\hat{y}$, the fitted value found by evaluation the equation at the x value of interest. We also see the confidence interval (31.03 to 35.89 miles per gallon is our 95% confidence estimate for average gas mileage for a vehicle with 160 horsepower, based on this sample), the margin of error for the estimate, the half width of the interval, and the standard error of the estimate.



A PREDICTION INTERVAL FOR A NEW OBSERVATION

What would we predict for gas mileage for a particular vehicle with 160 horsepower? The point estimate is still 23.76 miles per gallon, but we have some additional uncertainly because individual observations are more variable than means. The standard error becomes $SE(\hat{y}_V) = \sqrt{s^2(b_1) * (x_V - \overline{x})^2 + \frac{s^2(e)}{n} + s^2(e)}$ which for our data becomes $SE(\hat{\mu}_V) = \sqrt{0.0133^2 * (160 - 215.6)^2 + \frac{3.29^2}{15} + 3.29^2} = 3.477$. So the prediction interval is 33.46 ± 2.16*3.477 or (25.95, 40.97). Based on this regression we estimate with 95% confidence the gas mileage for a vehicle with 160 horsepower will be between 25.95 and 40.97 miles per gallon.

With the TI-89, this interval is found by scrolling down the output from the `LinRegTInt` described above. We see the calculator found the same interval as we computed "by hand."

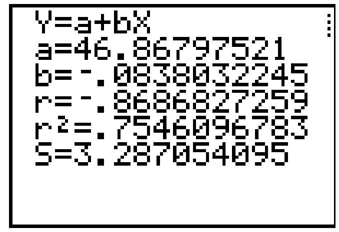


**A PROGRAM FOR REGRESSION INFERENCE

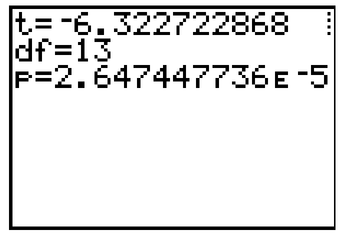
The author of this manual has written a program that performs these functions. (A listing is included and the program can also be obtained from the *Intro Stats* website which can be downloaded into a TI-83 or TI-84.) The program name is LSCINT. Once the program has been loaded into the calculator, to run the program, press [PRGM] and select that name from the list of programs. PgrmLSCINT is transferred to the home screen. Press [ENTER] to start the program.

You are prompted for the *X* list. enter its name and press ENTER. Do the same for the *Y* list.

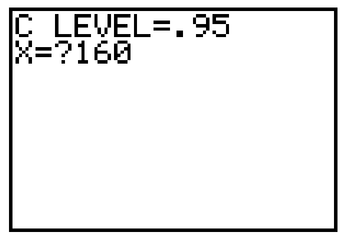
The next screen gives the coefficients in the equation, the correlation coefficient (*r*) and the coefficient of determination, r^2 and well as the standard deviation of the residuals (*s*). Press ENTER to continue.



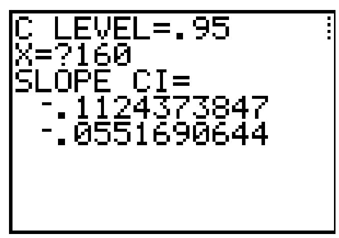
These are the results of the *t*-test for the slope. Press ENTER to continue.



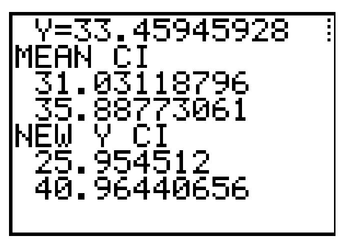
You will next be prompted for a confidence level (enter it as a decimal) and the value of *X* for which confidence and prediction intervals will be created. Press ENTER after inputting each value.



The calculator displays the confidence interval for the slope. Press ENTER to continue.



Now the calculator displays the *y*-value for the given *x* and confidence and prediction intervals. Press ENTER to finish the program.



WHAT CAN GO WRONG?

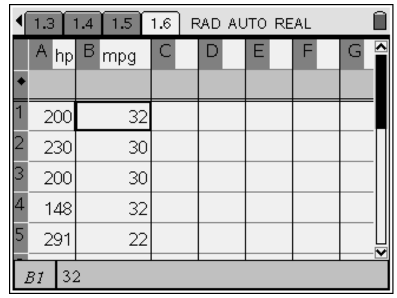
Assuming the lists are the same length, not much can go wrong that has not already been covered. One problem in doing many of these computations "by hand" comes from the compounding of round-off errors in intermediate computations. One is generally safest in using many digits in the interim and rounding only at the end. (Notice the "hand calculated" intervals are somewhat different from those obtained from the calculator. This is the reason.)

Program LSCINT listing (TI-83 Ascii version) This can also be found on the *Stats: Data and Modeling* website.

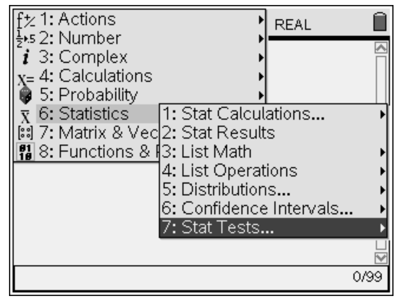
```
Input "X LIST=",LX
Input "Y LIST=",LY
FnOff
LinRegTTest LX,LY,0,Y1
σx^2nSTO▶V:sSTO▶S:ËSTO▶M:nSTO▶N
ClrHome
Output(1,2,"Y=a+bX"
Output(2,2,"a="
Output(2,4,a
Output(3,2,"b="
Output(3,4,b
Output(4,2,"r="
Output(4,4,r
Output(5,2,"r^2="
Output(5,5,r^2
Output(6,2,"S="
Output(6,4,S
Pause
dfSTO▶K
ClrHome
Output(1,1,"t="
Output(1,3,t
Output(2,1,"df="
Output(2,4,K
Output(3,1,"p="
Output(3,3,p
Pause
ClrHome
Input "C LEVEL=",C
Prompt X
K+1STO▶K:bSTO▶A
TInterval 0, √ (K),K,C
upperSTO▶T
A-T*S/√ (V)STO▶P
A+T*S/√ (V)STO▶Q
Output(3,1,"SLOPE CI="
Output(4,2,P
Output(5,2,Q
Pause
Y1-ST√(N^-1+(X-M)^2/V)STO▶P
Y1+ST√(N^-1+(X-M)^2/V)STO▶Q
Y1-ST√(1+N^-1+(X-M)^2/V)STO▶J
Y1+ST√(1+N^-1+(X-M)^2/V)STO▶U
ClrHome
Output(1,2,"Y="
Output(1,4,Y1
Output(2,1,"MEAN CI"
Output(3,2,P
Output(4,2,Q
Output(5,1,"NEW Y CI"
Output(6,2,J
Output(7,2,U
Pause
ClrHome
```

Commands for the TI-Nspire™ Handheld Calculator

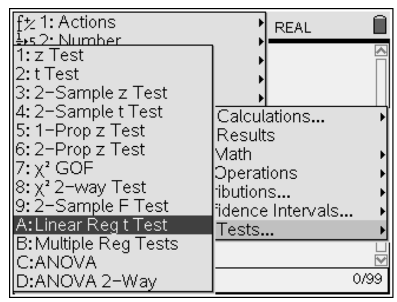
For inference for regression, create two lists of data. The automobile example is shown.



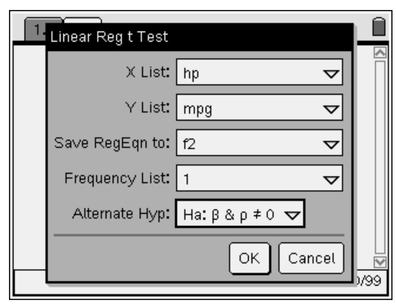
Then open a Calculator page, press menu , select Statistics, and Stat Tests.

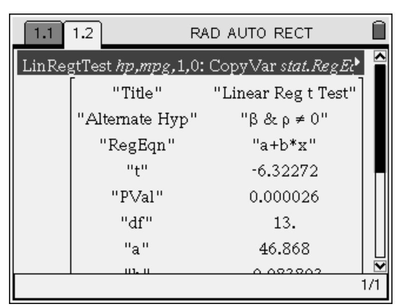


Select Linear Reg t Test.

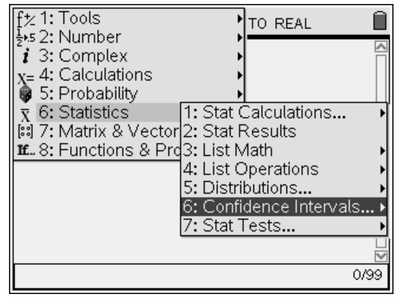


Select the list names.

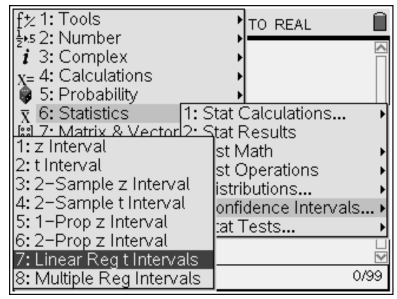




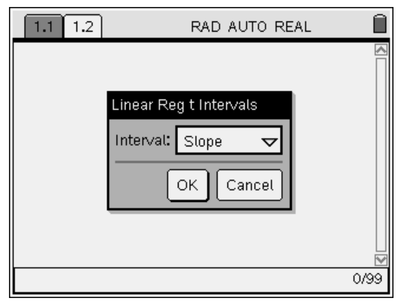
To construct a confidence interval for the slope, press menu, select Statistics, and Confidence Intervals.



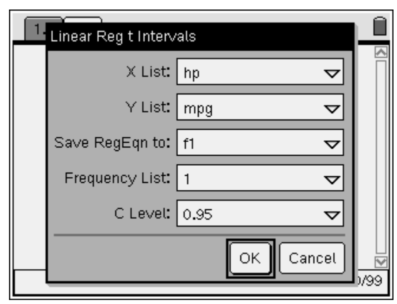
Select Linear Reg t Intervals.

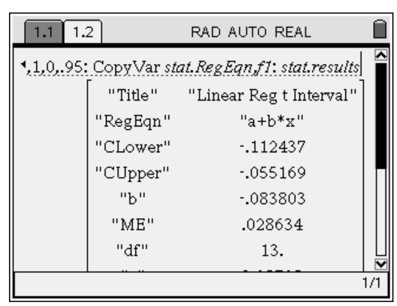


Select Slope.

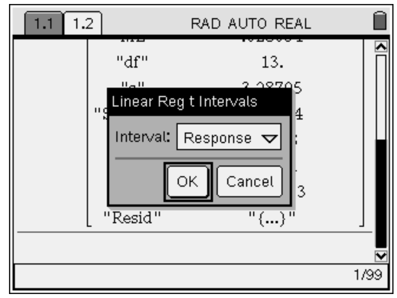


Select the list names and the level.





To construct the interval for the average value of the response variable, begin in the same way and select Response, rather than slope.



Select the list names as before. In this case also enter the value for the x variable. In this case *hp* = 160.