Technology Insight 12 - Correlation, Regression, and Prediction Intervals

Refer to **Technology Insight 3** for step by step instructions on creating scatterplots and regression equations. These notes will look at determining if there is significant correlation, and how to calculate prediction intervals.

Example 1 The following data are randomly selected bill totals and tips.

a) Determine if there is a correlation between bill size and tip size.

b) Estimate the tip for a $75 bill.

c) Find a 95% confidence interval for the average tip ($\overline{y}$) for a $75 bill.

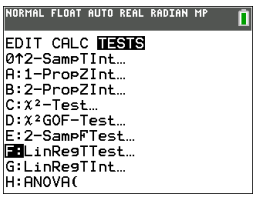
Bill	33.46	50.68	87.92	98.84	63.60	107.34
Tip	5.50	5.00	8.08	17.00	12.00	16.00

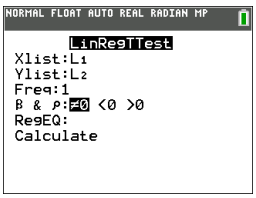
TI-84

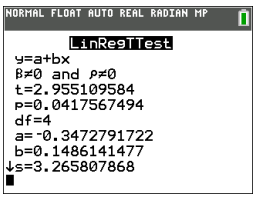
Step 1) Enter the data into L1 and L2.

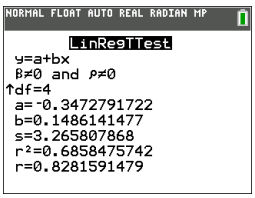
Step 2) Select **LinRegTTest** in the Stat TESTS menu.

Step 3) Use list L1 and L2, and set $\rho \neq 0$, and press **Calculate**







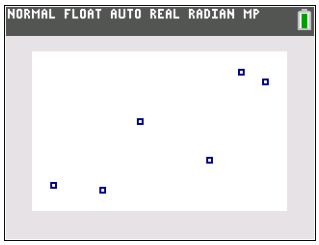


The last two screens show the results of the test, including the regression equation, correlation coefficients, the p-value for the test, and the standard error of regression, s (this is used in calculating a prediction interval.)

a) For the hypothesis test for correlation $\begin{cases} H_0: & \rho = 0 \\ H_1: & \rho \neq 0 \end{cases}$ and $\alpha = 0.05$, we see that the p-value is $p = 0.042$.

Since we reject H_0 this means there is statistical evidence of correlation with our calculated value $r = 0.828$. If everyone tipped at the exact same rate, $r = 1$.

A scatterplot of the data does show a possible linear trend, although there is a fair amount of variation.

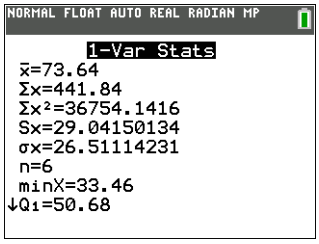


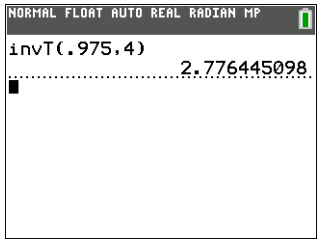
b) Using the regression equation $\hat{y} = -0.3473 + 0.1486\,x$ and substituting $x = 75$, we get

$$\begin{aligned} \hat{y} &= -0.3473 + 0.1486\,x \\ &= -0.3473 + 0.1486\,(75) \\ &= 10.7977 \end{aligned}$$

A point estimate for the tip is $10.80. (What is the significance of the slope of the regression equation?)

c) Using the prediction interval equation $\hat{y} - t_{\alpha/2} \cdot \text{SE}_{\hat{y}} \le \overline{y} \le \hat{y} + t_{\alpha/2} \cdot \text{SE}_{\hat{y}}$ where $SE_{\hat{y}} = s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{s_x^2 (n-1)}}$, we need $\overline{x}$, s_x, $t_{\alpha/2}$ using $d.f. = 4$, and also s, the standard error of the regression line, which was given in LinRegTTest above. (note: $x_0 = 75$)





First, calculate $\text{SE}_{\hat{y}}$

$$\begin{aligned}\text{SE}_{\hat{y}} &= s\sqrt{1 + \frac{1}{n} + \frac{(x_0 - \overline{x})^2}{s_x^2 (n-1)}} \\ &= 3.2658\sqrt{1 + \frac{1}{6} + \frac{(75-73.64)^2}{29.0415^2 \cdot (5)}} \\ &= 3.5281\end{aligned}$$

The 95% prediction interval is

$$\begin{aligned}\hat{y} - t_{\alpha/2} \cdot \text{SE}_{\hat{y}} \le \overline{y} \le \hat{y} + t_{\alpha/2} \cdot \text{SE}_{\hat{y}} \\ 10.8 - 2.7764\,(3.5281) \le \overline{y} \le 10.8 + 2.7764\,(3.5281) \\ 1.00 \le \overline{y} \le 20.60\end{aligned}$$

So, essentially, we are 95% confident the population mean tip for a \$75 bill will be between \$1.00 and \$20.60. This is not very useful, but remember it's calculated using only 6 data points which is an extremely small sample size. To get a better estimate we would need a much larger sample size.

EasyCalc

Select the **Correlation and Regression** page, and enter (or copy and paste) the data into var1 and var2.

Correlation and Regression

	Var1	Var2
1	33.46	5.5
2	50.68	5
3	87.92	8.08
4	98.84	17
5	63.6	12
6	107.34	16

Correlation r =	0.8282
r^2 =	0.6858
Regression: y = a + b x	
a =	-0.3473
b =	0.1486
degrees of freedom =	4
α =	0.05
critical r_{cr} =	± 0.8114
t =	2.9551
P-value =	0.0418
Predictions	
standard error of est.: s_e =	3.2658
x =	75
y-hat =	10.7988
95%-Prediction Interval	
1.0031	20.5945

With EasyCalc we get the same correlation coefficient and regression equation. The critical r value $r_c = 0.8114$ confirms that our correlation of $r = 0.8282$ is barely significant. The α-level determines both the critical value and the prediction interval.

We can also get a scatterplot along with the regression line (rev12 and later). Select the **Scatter Plots** page.

Scatter Plot

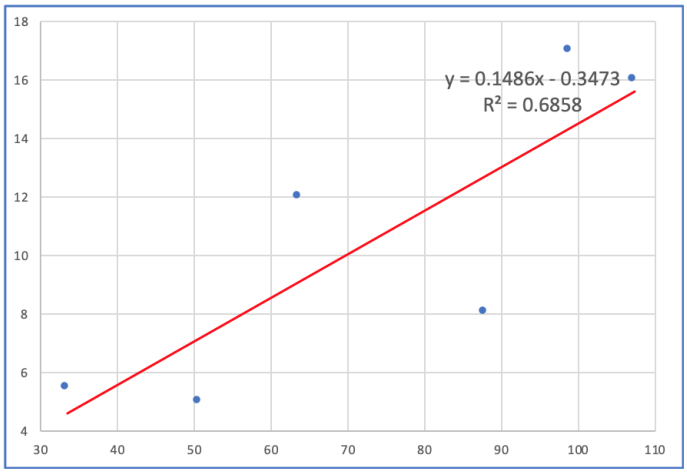
	x	y
1	33.46	5.5
2	50.68	5
3	87.92	8.08
4	98.84	17
5	63.6	12
6	107.34	16

	x	y
min	33.46	5
max	107.34	17

Axes Values		
min	30	4
max	110	18
ticks	10	2

Set Axes

Add Trendline



Set the minimum and maximum axes values using the calculated maximum and minimum values as a guide, and use appropriate values for the ticks (gridline spacing). Display the regression line by clicking the **Add Trendline** button. Clicking the **Set Axes** button will remove the trendline.