**Chapter 15**

**Inference for Slope**

In the  test of independence, we are testing if there is an association between 2 categorical variables. However, in many cases we want to test if there is an association between 2 numerical variables.

When people have colds, common wisdom says to avoid dairy products since they produce extra mucus. However, this claim was challenged by researchers in Australia in 1990. They infected volunteers with a cold virus and assigned them at random to drink different amounts of milk. Meanwhile, all used tissues were weighed to measure nasal mucus secretions (based on Nutrition Action 12-04).

|  |  |
| --- | --- |
| Glasses of Milk | Ounces of Mucus |
| 0 | 0.7 |
| 0 | 0.3 |
| 0 | 0.6 |
| 1 | 0.5 |
| 1 | 0.8 |
| 1 | 1.3 |
| 2 | 0.2 |
| 2 | 0.9 |
| 2 | 1.1 |
| 3 | 0.5 |
| 3 | 0.6 |
| 3 | 1.2 |
| 4 | 0.5 |
| 4 | 0.8 |
| 4 | 1.3 |

1. How many factors were used in this experiment?
2. How many levels?
3. What are the treatments?
4. What are the experimental units?
5. How is randomization used? What is its purpose?
6. How is replication used? What is its purpose?
7. Sketch the scatterplot and least squares regression line.
8. Interpret the slope of the LSRL in context.
9. Interpret the value of r2 in context.
10. Interpret the value se = 0.35 in context.

In most studies, when we use a data set to calculate regression lines, the data we are using comes from a random sample or a randomized experiment. Thus, the values we find are really just estimates of the *true* values.

Notation:

 Estimated (sample) LSRL: ****  or ****

 True (population) LSRL:  or 

 Estimated (sample) correlation coefficient: r

 True (population) correlation coefficient: (“rho”)

 Estimated (sample) standard deviation of residuals: 

True (population) standard deviation of residuals: 

5 Steps:

1. At first glance, it appears that the true slope of the least squares regression line using x = glasses of milk and y = ounces of mucus () is greater than 0 since the sample slope (b) = .0567 > 0. However, it is possible that there is no association between the variables and we got a slope this high due to randomization variability. To decide I will conduct a linear regression t-test for  (=.05).

2. Ho: There is no association between milk consumption and mucus secretion ( = 0)

 Ha: There is a positive association ( > 0)

Note: Some questions will ask if there is a “useful linear relationship” between two variables. This is simply asking if there is a linear association that is useful for descriptions, predictions, etc. In this case, the alternate hypothesis would be  ≠ 0.

3. Conditions: Skip for now =)

4. Test statistic:  with n - 2 degrees of freedom

*  is the standard error of the slope. It describes how variable the sample slope is for the given sample size, etc. This is the variability we saw in our dotplot earlier.
* This is different than  which is the standard deviation of the residuals, although they are related:



* TI-83: LinRegTTest: enter the lists, freq = 1, Ha, and leave RegEQ blank.
* Note: The TI-83 and most statistics software only test for  = 0, so if you are testing for something else, you need to calculate t by hand.
* Note: on the TI-83, s is , not !

5.

On the AP Exam, you are usually provided computer output such as the following, so make sure you know how to interpret the output!

Predictor Coef SE Coef T P

Constant 0.6400 0.1569 4.08 0.001

Milk 0.05667 0.06407 0.88 0.393

S = 0.350933 R-Sq = 5.7% R-Sq(adj) = 0.0%

Note: the p-value for the slope provided by computer output is for Ha:  (two-sided).

Suppose that a researcher was studying the relationship between the outdoor temperature in ˚C (x) and the depth of a woodpeckers nest in cm (y). Based on a sample of 12 nests, a regression analysis was conducted and the results are shown below. Does this give evidence that woodpeckers dig deeper when it’s colder outside? You may assume the conditions for inference have been met.

Predictor Coef SE Coef T P

Constant 20.1251 0.9402 21.40 0.000

Temp -0.34504 0.06008 -5.74 0.000

S = 2.33378 R-Sq = 76.7% R-Sq(adj) = 74.4%

**CI’s for Slope**

The following is weather information for a random sample of SoCal cities:

Average High Temp (in ˚F): 70 71 73 74 76 76 77 72 72

Annual Precipitation (in inches): 15 13 9 12 8 10 7 11 12

Estimate the average change in precipitation associated with an increase of 1˚F in average high temperature for Southern California cities using a 95% confidence interval.

Predictor Coef SE Coef T P

Constant 76.44 14.43 5.30 0.001

Temperature -0.8940 0.1963 -4.55 0.003

S = 1.36349 R-Sq = 74.8% R-Sq(adj) = 71.2%

**Conditions for Inference for Slope**



Each value of x has its own distribution of y-values: y ~ N(+x, ) where +x is the mean value of y for that particular value of x.

Thus, when we do inference for regression, we will need to check that the linear model is appropriate, the variability of the residuals is the same for all x’s and that the residuals are approximately normally distributed (in other words, these are going to be some of the conditions we need to check).

Checking the Conditions for Inference for Slope:

a. Treatments assigned at random? Given.

Note: for observational studies, the conditions are: “random sample from population of interest?” and “sample less than 10% of population?”

b. The linear model is appropriate? Check with the residual plot:

c. The variability of the residuals stays constant for all values of x? Check with the residual plot:

d. The residuals are approximately normally distributed? Check with a NPP of the residuals:

Infants who cry easily may be more easily stimulated than others. This may be a sign of higher IQ. Child development researchers explored the relationship between the crying of infants four to ten days old and their later IQ test scores. A snap of rubber band on the sole of the foot caused infants to cry. The researchers recorded the crying and measured its intensity by the number of peaks in the most active 20 seconds. They later measured the children’s IQ at three years using the Standard-Binet IQ test. See the results below.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| *Crying*  | ***IQ*** |  | ***Crying***  | ***IQ*** |  | ***Crying***  | ***IQ*** |  | ***Crying***  | ***IQ*** |
| *10* | ***87*** |  | ***20*** | ***90*** |  | ***17*** | ***94*** |  | ***12*** | ***94*** |
| *12* | ***97*** |  | ***16*** | ***100*** |  | ***19*** | ***103*** |  | ***12*** | ***103*** |
| *9* | ***103*** |  | ***23*** | ***103*** |  | ***13*** | ***104*** |  | ***14*** | ***106*** |
| *16* | ***106*** |  | ***27*** | ***108*** |  | ***18*** | ***109*** |  | ***10*** | ***109*** |
| *18* | ***109*** |  | ***15*** | ***112*** |  | ***18*** | ***112*** |  | ***23*** | ***113*** |
| *15* | ***114*** |  | ***21*** | ***114*** |  | ***16*** | ***118*** |  | ***9*** | ***119*** |
| *12* | ***119*** |  | ***12*** | ***120*** |  | ***19*** | ***120*** |  | ***16*** | ***124*** |
| *20* | ***132*** |  | ***15*** | ***133*** |  | ***22*** | ***135*** |  | ***31*** | ***135*** |
| *16* | ***136*** |  | ***17*** | ***141*** |  | ***30*** | ***155*** |  | ***22*** | ***157*** |
| *33* | ***159*** |  | ***13*** | ***162*** |  |  |  |  |  |  |

1. Describe the relationship in a graph and by regression analysis.
2. Perform a test of significance for these data.
3. Construct a 95% confidence interval and interpret.