COOL SHOES 3: TWO-VARIABLES

PRELUDE

This is the third lesson in a series of *Cool Shoes* lessons. The first was originally published as a lesson on Line of Best Fit. It was inspired by the Bee Bopper Shoe Store problem from College Preparatory Mathematics (http://cpm.org). The second focuses on the 1-variable analysis of the same data set of heights and shoe sizes.

LESSON PLAN

The lesson is designed to be completed in two phases. Phase 1 (Day 1) is for completing the required 2-variable statistical measures as well as their interpretations on the Boys' Data. Phase 2 (Day 2) is intended to serve as practice or assessment by having the students conduct the same procedures on the Girls' Data.

For the Correlation Coefficient and the Line of Regression, the data should be stored in Lists in the graphing calculator. When the Regression Line is then computed, the r value (correlation coefficient) is also given. Note, it is suggested that the given values chosen to be used in this question be values that are not found in the original data set. Here those values will be 63″ for height and 15 for shoe size.

The Chi-Squared Test is done through the use of matrices on the graphing calculator. The students must first tally the data according to the intervals for each variable. These intervals are determined somewhat arbitrarily, with the understanding that each cell needs at least a count of at least 5, and that the data is conventionally spread somewhat equally among the intervals. Once the observed chart is generated, the values are placed in Matrix A, with expected values designated to Matrix B.

Concepts
Pearson's Correlation Coefficient, Regression Line Equation and Chi-Squared Test. HSS-ID.C.8,9

Time: 2 hours.

Materials
Student Handout, student data for height (inches) & shoe size from class or the provided data sheet, segregated by boys & girls.

Preparation
Students need a working knowledge of two-variable statistics. This lesson is an application of those tools. Having students work through the *Cool Shoes 1: Linear* and *Cool Shoes 2: One-Variable* lessons first is appropriate and helpful. A set of 40 points of data is recommended. Collect data from students in the class, and augment from the set provided here.

Technology
This lesson assumes access to technology, particularly a graphing calculator. All three calculations (Pearson's Coefficient, Regression Line, Chi-Squared) are done by the calculator. The lesson emphasizes the application and interpretation of these results. However, a teacher may have the students compute these values by hand.

It is strongly recommended that the teacher points out that the first value, Pearson's Coefficient, describes correlation, while the Chi-Squared test describes potential association (causation), between the two variables.

ANSWER KEY FOR BOYS' DATA (DAY 1)

1) The Correlation Coefficient equals 0.709, showing that there is a strong positive correlation.
2) The regression equation is y = 0.43x – 19.5, where *x* = height and *y* = shoe size.
   This slope implies that there is an increase of approximately half a shoe size for every inch in height.
   Given 63″ for a height not found in the data set, then y = 0.43(63) – 19.5 = 7.59. Size = 7.5
   Given Size 15 shoe not found in the data set, then 15 = 0.43x – 19.5; x = 80.2. Height = 80″ (6′8″)

COOL SHOES 3: 2-VARIABLES (CONTINUED)

3) H_0: Height and shoe size are independent.
H_A: Height and shoe size are dependent.
5% significance level given here.

CONTIGENCY TABLES

Observed			
	7-10	10.5-17	Totals
61-70	13	7	20
71-76	6	14	20
Totals	19	21	40

Expected			
	7-10	10.5-17	Totals
61-70	9.5	10.5	20
71-76	9.5	10.5	20
Totals	19	21	40

There appears to be a small deviation from the expected values. We can see a shift to the smaller sizes for shorter heights, and a shift to the larger sizes for taller heights.

Degree of Freedom = 1, Critical Value = 3.841, $\chi^2 = 4.91$, p = 0.027

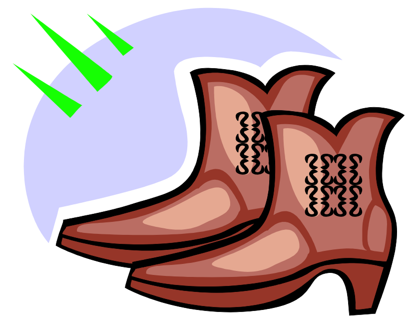
Since the Chi-Squared value is greater than the critical value, we reject the null hypothesis. The fact that the p-value is less than the significance level, also gives confidence to reject. Therefore, we accept the alternate hypothesis that height and shoe size are dependent.

SAMPLE DATA SETS

BOYS' DATA			
Height (in)	Shoe Size	Height (in)	Shoe Size
75	10.5	73	11
70	12	69	12
72	13	72	10
70	10	73	11
74	14	69	10
72	11.5	72	10
70	10.5	73	11
74	12	75	14
70	10.5	74	13
76	13	70	9
68	13	67	11
76	12.5	71	14
65	7	64	8
66	8	69	10
67	8.5	62	7
75	14	70	12
69	9.5	66	8
72	9	68	7.5
71	10	71	10
69	10	72	9

GIRLS' DATA			
Height (in)	Shoe Size	Height (in)	Shoe Size
59	5	60	7.5
60	7.5	66	7.5
61	7	65	7
66	9	64	8.5
63	7	68	9
66	7.5	65	8
63	6.5	63	8
67	6.5	68	8.5
62	7.5	64	7.5
62	6.5	63	6.5
68	11	68	8
69	10.5	63	7.5
72	11	66	8.5
68	8.5	69	9
63	8.5	67	8
68	9.5	68	10
67	8.5	65	7.5
66	9.5	66	8.5
67	7	64	6.5
69	9	72	12

COOL SHOES 3: TWO-VARIABLES



1. Determine Pearson's Correlation Coefficient: ______________________
   What does this say about the relationship between height and shoe size?

2. Find the least squares regression line. ______________________
   What does the slope imply about the relationship of shoe size to height?

   Using your regression equation, predict the shoe size for a given height of __________.

   Using your regression equation, predict the height for a given shoe size of __________.

3. Determine whether or not there exists an association between height and shoe size.

   Null Hypothesis (H_0):____

   Alternative Hypothesis (H_A):____

   Significance Level _______%

CONTIGENCY TABLES

Observed			

Expected			

What is your initial impression of the difference between the observed values and the expected values?

Should the null hypothesis be accepted or rejected? Why?

Degree of Freedom df = __________

Rejection Value (critical value) $\chi^2 >$ __________

Chi-Squared Value $\chi^2 =$ __________

p-value p = __________