**Activity 7.1.2 – Inference on the Slope of a Linear Relationship**

**Examining Bivariate Relationships: Arm length vs. Foot Length**

The table and scatterplot below display the arm length and foot length data for a random sample of nine high school students in a local high school.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  |

|  |  |
| --- | --- |
| Foot length (cm) | Arm length (cm) |
| 24 | 164 |
| 24 | 166 |
| 24 | 171 |
| 25.5 | 179 |
| 24 | 175 |
| 22 | 156 |
| 21.5 | 161.5 |
| 28 | 181 |
| 20.32 | 172 |

 |

The sample data has a least-squares regression line of $\hat{y}=2.38x+113.15$, where $\hat{y}$ is the predicted arm length of a student and *x* is the student’s foot length.

Inference question:Does the sample provide evidence that a positive linear relationship exists between foot length and arm length for *all* students in the high school?

**Population:**

All students in the high school

**Parameter:** Slope of regression line

**Sample:**

9 high school students

**Statistic:** Slope of regression line

The regression line slope of the population data is a parameter. It is the slope of the regression line that summarizes the relationship between arm length and foot length for *all* students at the high school. This parameter is unknown. By conducting a *randomization test*, we can use the sample regression line slope to make an *inference* about the population regression line slope. We do so as follows.

Randomization Test for Population Regression Line Slope

* We assume the population regression line slope is zero. This means that there is no linear association between arm length and foot length.
* We treat the sample like a population, assuming there is no relationship between these two variables, and generate *randomization samples*.
* We find the likelihood of getting a sample regression line slope as extreme as the one we found by chance alone.
1. Your instructor has given you nine index cards. Write the arm length values on the cards. Write one value on each card. The arm lengths from the random sample are {164, 166, 171, 179, 175, 156, 161.5, 181, 172}.
2. Shuffle the index the cards well and stack the cards into a single pile. Then, in the random order that the arm lengths appear, copy them into the table below. Notice that the arm lengths have been shuffled while the foot lengths have remained in the same order. Calculate the slope of the regression line for this randomization sample. Round the slope to one decimal place.

Randomization Sample # 1

|  |  |
| --- | --- |
| Foot length (cm) | Arm length (cm) |
| 24 |  |
| 24 |  |
| 24 |  |
| 25.5 |  |
| 24 |  |
| 22 |  |
| 21.5 |  |
| 28 |  |
| 20.32 |  |

1. Repeat the process. Shuffle the index the cards well and stack the cards into a single pile. In the order that the arm lengths appear, copy them into the table below. Calculate the slope of the regression line for this randomization sample. Round the slope to one decimal place.

Randomization Sample # 2

|  |  |
| --- | --- |
| Foot length (cm) | Arm length (cm) |
| 24 |  |
| 24 |  |
| 24 |  |
| 25.5 |  |
| 24 |  |
| 22 |  |
| 21.5 |  |
| 28 |  |
| 20.32 |  |

1. Create a dotplot of all regression line slopes obtained from randomization samples in class.



The dot plot above is a *randomization distribution.* It was formed under the assumption that there is *no relationship* between foot length and arm length. The *variability* in the sample regression line slopes is due solely to random chance.

1. What do you notice about the distribution of regression line slopes from the randomization samples?
2. Does the distribution of sample statistics appear to come from a population in which there is no association between foot length and arm length?
3. Assuming there is no association between foot length and arm length, what is the probability of obtaining a random sample with a regression line slope greater than or equal to the one we found? Use the randomization distribution to answer this question. This probability is called a *P*-value (probability value).
4. Is the observed sample regression line slope statistically significant? Explain.
5. What inference can we make about the population regression line slope? What can we conclude about the relationship between foot length and arm length for all students at the high school?