|  |  |
| --- | --- |
| In this activity you will investigate the life expectancy of people over a 104 year time period from 1900 to 2004. You will do this by exploring the least squares line. |  |
| The spreadsheet and scatterplot displays the life expectancy for men, women, and people in the United States for every ten years from 1900 to 2000 and each year from 2001 to 2004. Let’s investigate the relationship between the year and the life expectancy for a person. | | |
| 1. Describe the relationship between the two variables. | | |
| **Manipulate the slope and y-intercept of the trendline and develop a linear model with your group that best fits the data.** **Record your model below and on your vertical surface.** | | |
| **Residuals:**  For each question, write your responses on a vertical surface before recording on your paper.  Select **Show Residual Squares**. A square will appear at each data point. Please note that each side of the square is equal to the vertical distance from the data point to the trendline. This vertical distance is known as the residual.   |  | | --- | | **Residual = Actual value – Predicted value** | | | |
| 1. Examine the data point such that the year is 1950. What is the residual for this data point? Interpret the residual for this point **in context**. 2. If a data point is above the regression line, is the value of the residual positive or negative? 3. If a data point is below the regression line, is the value of the residual positive or negative? | | |
| **Least squares regression line (LSRL):**  Move the line to minimize the size of the squares.   1. What is the smallest value you can find for the **sum of squares**? 2. `What do you notice about the distribution of the data points around the line (i.e., above vs. below the line, or equal spacing vs. clusters)? 3. To have the smallest sum, how do the points need to be distributed? 4. What was the benefit of using the residuals and squares in developing a model for the data? | | |

|  |
| --- |
| **Use technology to calculate a least squares regression line.**  Now, you will compare your line with the regression line created by technology. Select the box that will reveal the line that truly minimizes the sum of squares.   1. How much does the true LSRL differ from your line, and what is the value of this sum of squares? 2. Develop a definition for a LSRL: 3. In context, interpret the slope and y-intercept of this data’s least-squares regression line. 4. What do you predict the life expectancy of Americans in 2040 to be? Does this value make sense? |
|  |
| **Quick Notes:**  **Check Your Understanding**  Analyze the two additional sets of data in the spreadsheet by adding a line of best fit and then minimizing the sum of the squares. Check your answer each time by finding the linear regression model and the sum of the squares.   |  |  |  | | --- | --- | --- | |  | ***Women*** | ***Men*** | | ***your equation*** |  |  | | ***sum of squares*** |  |  | | ***regression equation*** |  |  | | ***sum of squares*** |  |  | | ***Correlation coefficient*** |  |  | |

1. Interpret the slope and y-intercept of each equation in context.
2. What are the correlation coefficients for both bivariate sets of data? What does this correlation tell you about the relationship between year vs women’s life span and year vs men’s life span?
3. What is the predicted life expectancy of **men** in 1950, and what is the residual of that value? Interpret that residual.
4. Use the linear regression model to predict the life expectancy for **women** in 2010. What would be the predicted life expectancy of **women** in 2200? Can you always extrapolate?

**Building Thinking Classrooms Quick Notes**