**Algebraic Methods for Fitting a Least Squares Regression Line**

Name \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Steps

* Find the centroid (: Find the mean of the x and y values in your data set.
* Selecting another point to fit a line through the center of the data.
* Find the equation of the line that goes through those two points.
* Find the residuals
* Square the residuals
* Take the sum of the squares
* Compare this sum to the sum of another line that could fit the data. The lower the sum, the better the fit.

\*(Note: When describing a set of one-variable data, the mean is the most common predictor of a value in that data set. Therefore, the centroid is a logical choice for a point on the line of best fit because it uses the average of the x-values and the average of the y-values.)

Example

Below are season statistics from some of the players on NCSU’s 2012-2013 men’s basketball team.

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Minutes Played | 1135 | 1105 | 1189 | 1129 | 944 | 849 | 27 | 423 | 124 |
| Points Per Game | 528 | 445 | 441 | 408 | 425 | 292 | 15 | 118 | 17 |

Algebraically fit at least two linear models that can represent the situation described above. Calculate the sum of the squared residuals and determine which linear model is a better fit.



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  | ( |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  | ( |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Which line is a better fit? How do you know?

Independent Practice

Below are season rushing statistics from some of Appalachian State University’s 2006 football players.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Carries | 302 | 188 | 61 | 49 | 20 | 10 |
| Yards | 1676 | 1153 | 315 | 159 | 98 | 62 |

Algebraically fit at least two linear models that can represent the situation described above. Calculate the sum of the squared residuals and determine which linear model is a better fit.



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  | ( |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |



|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  |  |  |  | ( |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |

Which line is a better fit? How do you know?