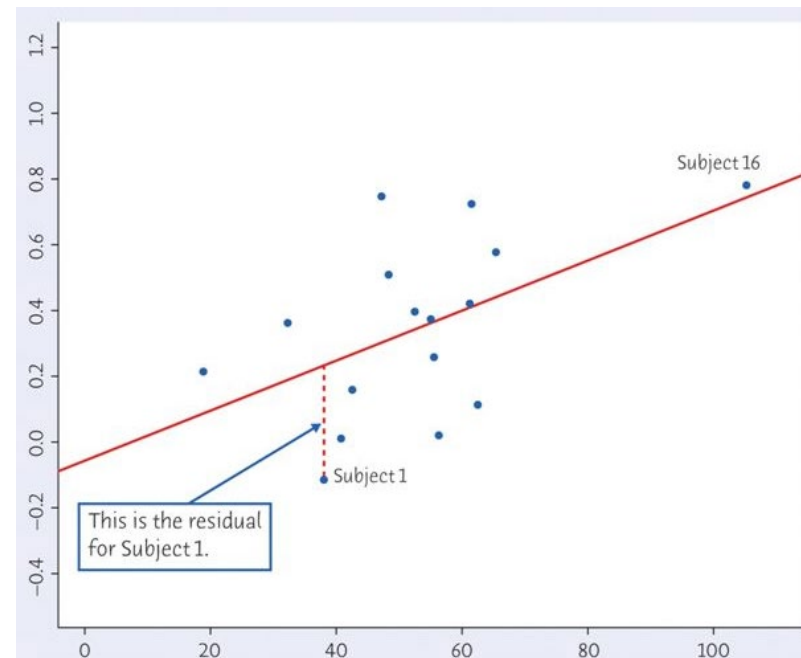


# Module 3c – Regression

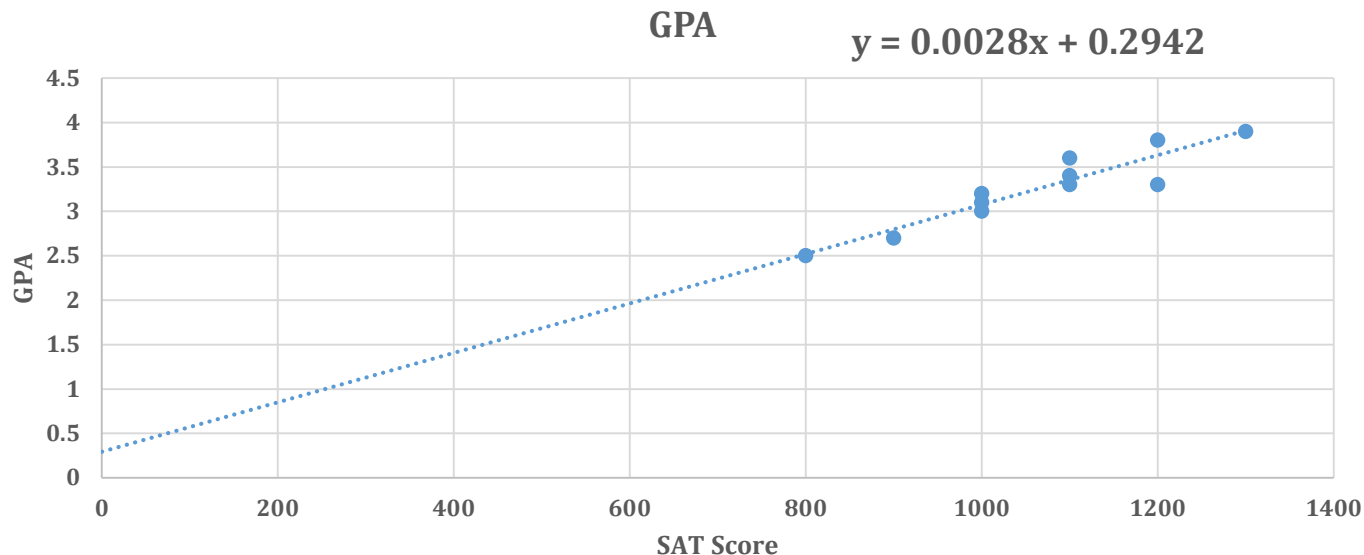
# Regression Line

- Regression Line
  - **Line of best fit:** The line passing through points on a scatterplot that is closest to all points.
  - **Prediction Line/Trend Line:** Shows the pattern of an relationship/association
  - **Least squares regression line:** The line that reduces the total *residuals* to the smallest possible value
- Residuals (Error terms)
  - The distance between the prediction line and the actual values.



# Regression Line

- Formula for regression line:
  - $y = a + bx$
  - $y$  = response/dependent variable value
  - $x$  = explanatory/independent variable value
  - $b$  = slope (rise/run)
    - For every one unit increase in  $x$ , there is a predicted  $b$  change in  $y$
  - $a$  =  $y$ -intercept
    - The value of  $y$  when  $x = 0$  (where the line crosses  $y$ -axis)



# Regression Line

- You can use this line for predicting values
- For someone with an SAT score of 1200, we would predict a GPA of what?
  - $y = .0028(1200) + .2942$
  - Predicted GPA = 3.65
- Beware of ***extrapolation*** (predicting beyond the scope of the results)

