V36 - Simple Linear Regression-Part 1

❑ Introduction to Regression
   ❑ Definition & uses of

❑ Equation of a straight line

❑ Residual

Course: Statistical Testing & Regression
Dr. Renee Clark
Swanson School of Engineering
Industrial Engineering
University of Pittsburgh

Regression

- **Statistical method** that explores ~~Relationship~~ between 2 or more variables

- Examples of relationships:
  - Score on statistics exam vs. hours spent studying
  - One's happiness vs. one's wealth

- Has random or probabilistic aspect
  - People with same wealth may not be equally happy

"Views on" or "Uses of" Regression

"x, y" data

i	x	y
1	2.50	7.68
2	5.03	6.95
3	7.60	6.30
4	11.60	5.95
5	13.00	5.51
6	19.60	4.62
7	26.20	3.98
8	33.00	2.85
9	40.00	2.30
10	47.00	1.45
11	52.03	1.09

1. **Expresses** mathematical relationship between variables
   - **Equation or function** between x and y
2. **Model**
   - Abstraction of real-world process to represent reality
3. **Predict** or **estimate** values of one variable from values of another
4. **Measure** degree of relationship between variables
5. **Data mining** technique (extraction of patterns in data)

Goal in Linear Regression

Fit **linear model** to set of (x,y) data points

Model in this case is a ~~straight line~~

Equation of straight line?

$$y = mx + b$$
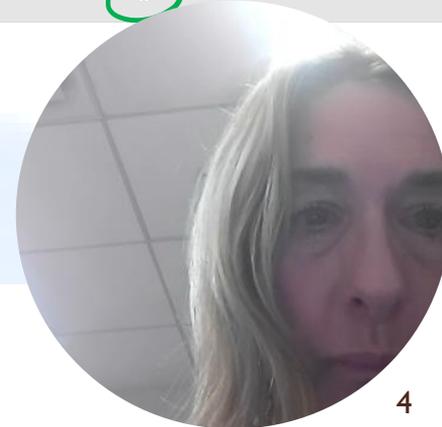
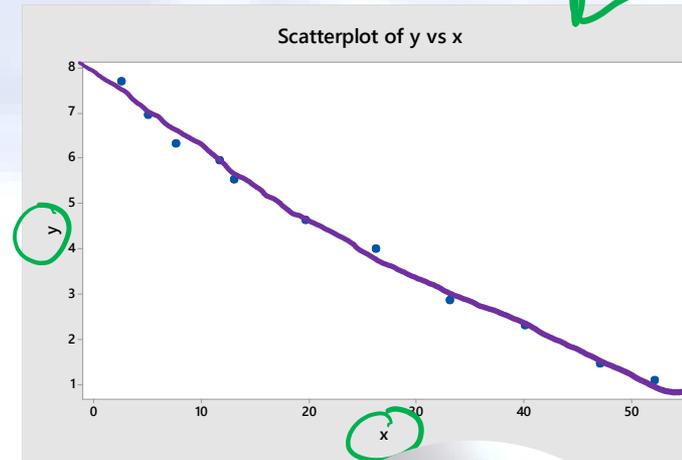(*m*: slope ; *b*: y-intercept)

Insert line →

Result w/ regression:

Summarization of (x,y) points via **equation** called

line of best-fit OR fitted line

Goal is "best fit" line (as close as possible to points)

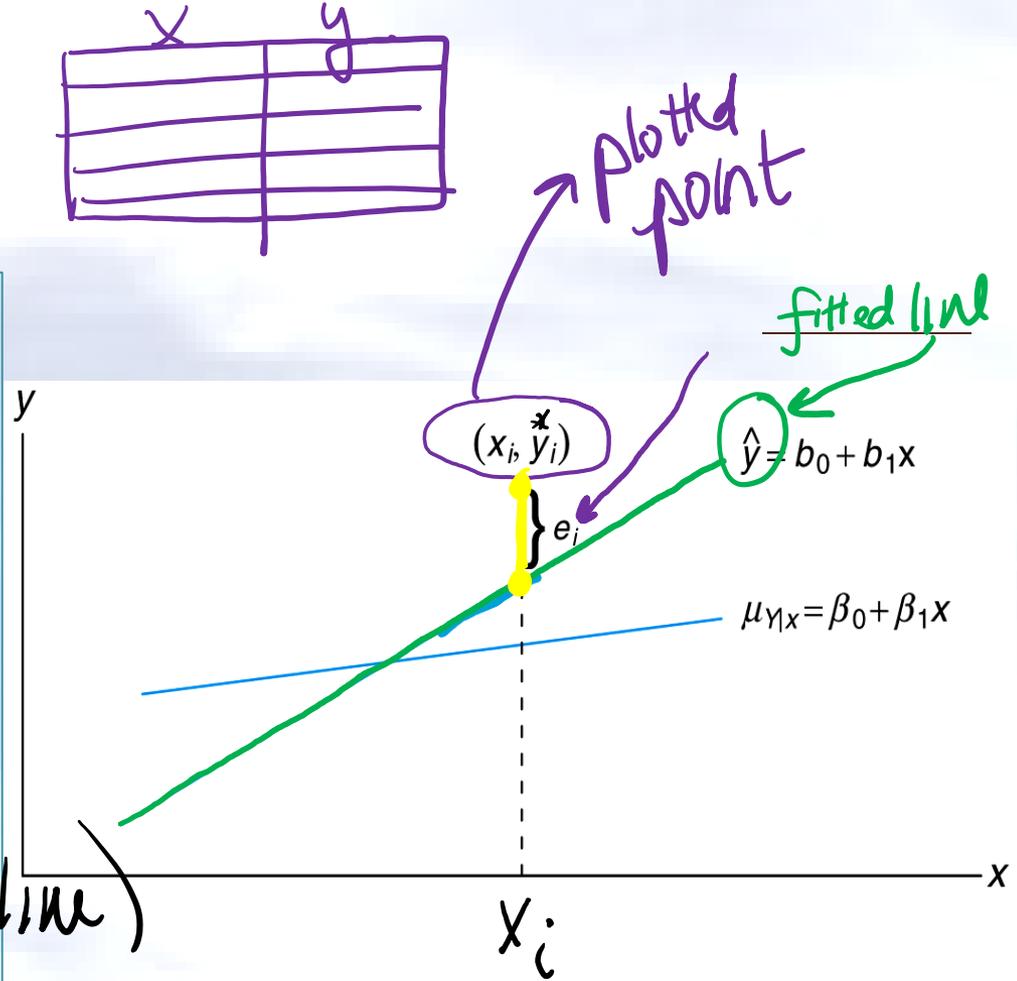| i	x	y
1	2.50	7.68
2	5.03	6.95
3	7.60	6.30
4	11.60	5.95
5	13.00	5.51
6	19.60	4.62
7	26.20	3.98
8	33.00	2.85
9	40.00	2.30
10	47.00	1.45
11	52.03	1.09

Scatterplot of y vs x

Residuals

Plotted point (x_i, y_i)

Residual e_i is vertical distance between plotted point & fitted line ($\hat{y}$)

Residual:
$$e_i = y_i - \hat{y}_i$$

Residual: error in fit of $\hat{y}$ (fitted line) to plotted point (x_i, y_i) (want small)

plotted point

fitted line

$(x_i, \overset{*}{y}_i)$

$\hat{y} = b_0 + b_1 x$

$\} e_i$

$\mu_{Y|x} = \beta_0 + \beta_1 x$

x_i

Acknowledgement

You are free

- to **Share** – to copy, distribute, display and perform the work
- to **Remix** – to make derivative works

Under the following conditions

- **Attribution** — You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).
- **Noncommercial** — You may not use this work for commercial purposes.
- **Share Alike** — If you alter, transform, or build upon work, you may distribute the resulting work only under same or similar license to this one.

THE END