

STAT 305: Lecture 8

Describing Relationships

and

Fitting Mathematical Relationships

Course page: imouzon.github.io/stat305

Continuing from Lecture 7

Describing Relationships

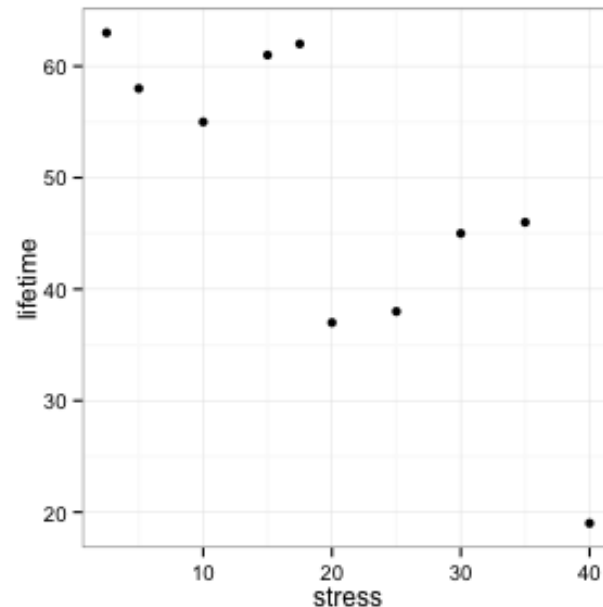
Recap

Last lecture example: Strain on Bars

We introduced data from an experiment:

stress (kg/mm ²)	2.5	5.0	10.0	15.0	17.5	20.0	25.0	30.0	35.0	40.0
lifetime (hours)	63	58	55	61	62	37	38	45	46	19

Judging from the plot, it seems reasonable to assume that there could be an underlying linear relationship between lifetime y and strain x



Describing Relationships

Recap

Last lecture example: Strain on Bars

stress (kg/mm ²)	2.5	5.0	10.0	15.0	17.5	20.0	25.0	30.0	35.0	40.0
lifetime (hours)	63	58	55	61	62	37	38	45	46	19

Theoretical Relationship

$$y = \beta_0 + \beta_1 \cdot x$$

Observed Relationship

$$\begin{aligned} y &= \beta_0 + \beta_1 \cdot x + \text{errors} \\ &= \text{signal} + \text{noise} \end{aligned}$$

Fitted Relationship:

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

Describing Relationships

Deriving the Least Squares Estimates

We can rewrite the target we want to minimize so that the variables are less tangled together:

Idea

Ex: Bars

Fitting Lines

Best Estimate

$$\begin{aligned}\sum_{i=1}^n (y_i - \hat{y}_i)^2 &= \sum_{i=1}^n (y_i - (b_0 + b_1 x_i))^2 \\&= \sum_{i=1}^n (y_i^2 - 2y_i(b_0 + b_1 x_i) + (b_0 + b_1 x_i)^2) \\&= \sum_{i=1}^n y_i^2 - \sum_{i=1}^n 2y_i(b_0 + b_1 x_i) + \sum_{i=1}^n (b_0 + b_1 x_i)^2 \\&= \sum_{i=1}^n y_i^2 - \sum_{i=1}^n (2y_i b_0 + 2y_i b_1 x_i) + \sum_{i=1}^n (b_0^2 + 2b_0 b_1 x_i + (b_1 x_i)^2) \\&= \sum_{i=1}^n y_i^2 - \sum_{i=1}^n 2y_i b_0 - \sum_{i=1}^n 2y_i b_1 x_i + \sum_{i=1}^n b_0^2 + \sum_{i=1}^n 2b_0 b_1 x_i + \sum_{i=1}^n b_1^2 x_i^2 \\&= \sum_{i=1}^n y_i^2 - 2b_0 \sum_{i=1}^n y_i - 2b_1 \sum_{i=1}^n y_i x_i + n b_0^2 + 2b_0 b_1 \sum_{i=1}^n x_i + b_1^2 \sum_{i=1}^n x_i^2\end{aligned}$$

Describing Relationships

Deriving the Least Squares Estimates (continued)

We set both partial derivatives equal to 0 and solve them at the same time:

Idea

Ex: Bars

Fitting Lines

Best Estimate

$$-2 \sum_{i=1}^n y_i + 2nb_0 + 2b_1 \sum_{i=1}^n x_i = 0$$

$$-2b_0 \sum_{i=1}^n y_i x_i + 2b_0 \sum_{i=1}^n x_i + 2b_1 \sum_{i=1}^n x_i^2 = 0$$

We can rewrite the first equation as:

$$\begin{aligned} b_0 &= \frac{1}{n} \sum_{i=1}^n y_i - b_1 \frac{1}{n} \sum_{i=1}^n x_i \\ &= \bar{y} - b_1 \bar{x} \end{aligned}$$

and then replace all b_0 in the second equation (there is some algebra type stuff along the way, of course)

Describing Relationships

Deriving the Least Squares Estimates (continued)

After a little simplification we arrive at our estimates:

Least Squares Estimates for Linear Fit

Idea

Ex: Bars

Fitting Lines

Best Estimate

$$b_0 = \bar{y} - b_1 \bar{x}$$

$$b_1 = \frac{\sum_{i=1}^n y_i x_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2}$$
$$= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

Wrap Up

- Don't try to memorize the derivation. I will never ask you to do that on an exam.
- Try to understand the simplification steps - the ones that moved constants out of summations for example.
- This is one rule - there are others, but **Least Squares Estimates** have some useful properties that will make it the obvious best choice as we continue the course.

Describing Relationships

Recap

Last lecture example: Strain on Bars

stress (kg/mm ²)	2.5	5.0	10.0	15.0	17.5	20.0	25.0	30.0	35.0	40.0
lifetime (hours)	63	58	55	61	62	37	38	45	46	19

Estimating the best slope and intercept using least squares:

$$\begin{aligned}b_0 &= \bar{y} - b_1 \bar{x} \\b_1 &= \frac{\sum_{i=1}^n y_i x_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \\&= \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}\end{aligned}$$

In our case we have the following:

$$\sum_{i=1}^{10} y_i = 484, \sum_{i=1}^{10} x_i = 200, \sum_{i=1}^{10} x_i y_i = 8407.5, \sum_{i=1}^{10} x_i^2 = 5412.5,$$

Describing Relationships

Recap

Last lecture example: Strain on Bars

stress (kg/mm ²)	2.5	5.0	10.0	15.0	17.5	20.0	25.0	30.0	35.0	40.0
lifetime (hours)	63	58	55	61	62	37	38	45	46	19

$$\sum_{i=1}^{10} y_i = 484, \sum_{i=1}^{10} x_i = 200, \sum_{i=1}^{10} x_i y_i = 8407.5, \sum_{i=1}^{10} x_i^2 = 5412.5,$$

Using this we can estimate b_1 :

$$\begin{aligned} b_1 &= \frac{\sum_{i=1}^n y_i x_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} \\ &= \frac{8407.5 - 10 \left(\frac{200}{10} \right) \left(\frac{484}{10} \right)}{5412.5 - 10 \left(\frac{200}{10} \right)^2} \\ &= \frac{-1272.5}{1412.5} \\ &\approx -0.9009 \end{aligned}$$

Describing Relationships

Recap

Last lecture example: Strain on Bars

stress (kg/mm ²)	2.5	5.0	10.0	15.0	17.5	20.0	25.0	30.0	35.0	40.0
lifetime (hours)	63	58	55	61	62	37	38	45	46	19

$$\sum_{i=1}^{10} y_i = 484, \sum_{i=1}^{10} x_i = 200, \sum_{i=1}^{10} x_i y_i = 8407.5, \sum_{i=1}^{10} x_i^2 = 5412.5,$$

And using b_1 we can estimate b_0 :

$$\begin{aligned} b_0 &= \bar{y} - b_1 \bar{x} \\ &= \left(\frac{484}{10} \right) - b_1 \left(\frac{200}{10} \right) \\ &= 48.4 - \left(\frac{-1272.5}{1412.5} \right) 20.0 \\ &= 66.4177 \end{aligned}$$

Which gives us the **Fitted Relationship**:

$$\hat{y} = 66.4177 - 0.9009x$$

Describing Relationships

Recap

Using JMP

Topics to be covered in JMP

- Fitting linear relationships
- Describing quality of fit (correlation, R^2)
- Fitting relationships using multiple variables
- Fitting non-linear relationships

Example: Manufacturing Ball Bearings

Controlling surface roughness is an important part of the manufacture of bearing balls. The key step in this smoothing the balls involves the use of a spinning weighted disc. Two important aspects of this are the rotation speed of the disc and the weight applied to the disc. Since higher weights and higher rotation speed are all known to cause shorter lifetimes for the discs (which requires halts in production, costs of new discs, and so on), a team of engineers are attempting to better understand the relationship between the rotation speed, the weight, and the resulting surface roughness of the balls produced.

Experiment 1: Constant speed, changing applied weight

With the disc rotation speed locked at 50 rotations/second, the team of engineers created 60 batches of balls using differently weighted discs (0.025 g, 0.050 g, 0.075 g, 0.100 g, ..., 0.500 g) and randomly selected one ball from each batch. The results are recorded in the dataset "balls-001.csv" on the course page.

Experiment 2: Changing speed, constant applied weight

With an better understanding of the relationship between weight and surface roughness, the team turned their attention to rotation speed. This time the produced 3 batches for each of 15 rotation speeds (25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, and 95 rotations per second). The results are recorded in the dataset "balls-002.csv" on the course page.

Experiment 3: Changing speed changing applied weight

With a better understanding of the relationship between weight and surface roughness, the team turned their attention to rotation speed. This time the produced 3 batches for each combination of 20 weights (0.025 g, 0.050 g, 0.075 g, 0.100 g, ..., 0.500 g) and 15 rotation speeds (25, 30, 35, 40, 45, 50, 55, 60, 65, 70, 75, 80, 85, 90, and 95 rotations per second). The results are recorded in the dataset "balls-003.csv"