Least Square Method

- We covered this dude in linear algebra
- And intro to scientific computing
- And intro to data science
- And here too
- Estimate β_0, β_1 by mimimizing the function:

$$S(\beta_0, \beta_1) = \sum_{i=1}^{n} (Y_i - \beta_0 - \beta_1 x_i)^2$$

$$= \sum_{i=1}^{n} \epsilon_i^2$$

- Sow how to do that? Take partial derivatives and set to 0

$$\frac{\delta s}{\delta b_0} = -2\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\frac{\delta s}{\delta b_1} = -2\sum_{i=1}^{n}(Y_i - \beta_0 - \beta_1 x_i)x_i = 0$$

- So we have a system of linear equations. 2 equations, 2 unknowns, solve for β_0, β_1
- Denote this as $*$
- $(*)$ equations are known as the Normal Equations, for reasons we'll get into later (Oh yeah they're orthogonal which is also called normal that makes sense)
- We get:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

- Result: We get the statistics $\hat{\beta}_0, \hat{\beta}_1$ but they vary from sample to sample cuz they're statisticzzz
- So the least squares equation is given by:

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x$$

- Have we talked about σ^2? Nope. But we have to talk about σ^2. It's important
- The residuals:

$$\text{Residuals are } e_i = Y_i - \hat{Y}_i, i = 1, 2, \ldots, n$$

- Y_i is the observed value for $x_i, \hat{Y}_i$ is the predicted value based on the model you got from the least squares guy
- is ϵ_i the same as e_i? THEY ARE NOOOOT THE SAAAAAAAAAAME NOOOOO
- Recall: $\epsilon_i = y_i - \beta_0 - \beta_1 x_i$
- But $e_i = y_1 - \hat{\beta}_0 - \hat{\beta}_1 x_i$
- Similar but the x_is in the first one are not known to us. Unobservable random variables!
- But the e_is are observable because we are using estimates for the parameters
- So the e_is are good representatives for the unobservable ϵ dudes which we will neeeeeeeever knooooooooow