Model Assumptions

DASC 512

Last time…

We built an OLS model

$$y = \hat{\beta}_0 + \hat{\beta}_1 x + \epsilon$$

We talked about calculating the residual errors

$$\epsilon_i = y_i - \hat{\beta}_1 x_i - \hat{\beta}_0$$

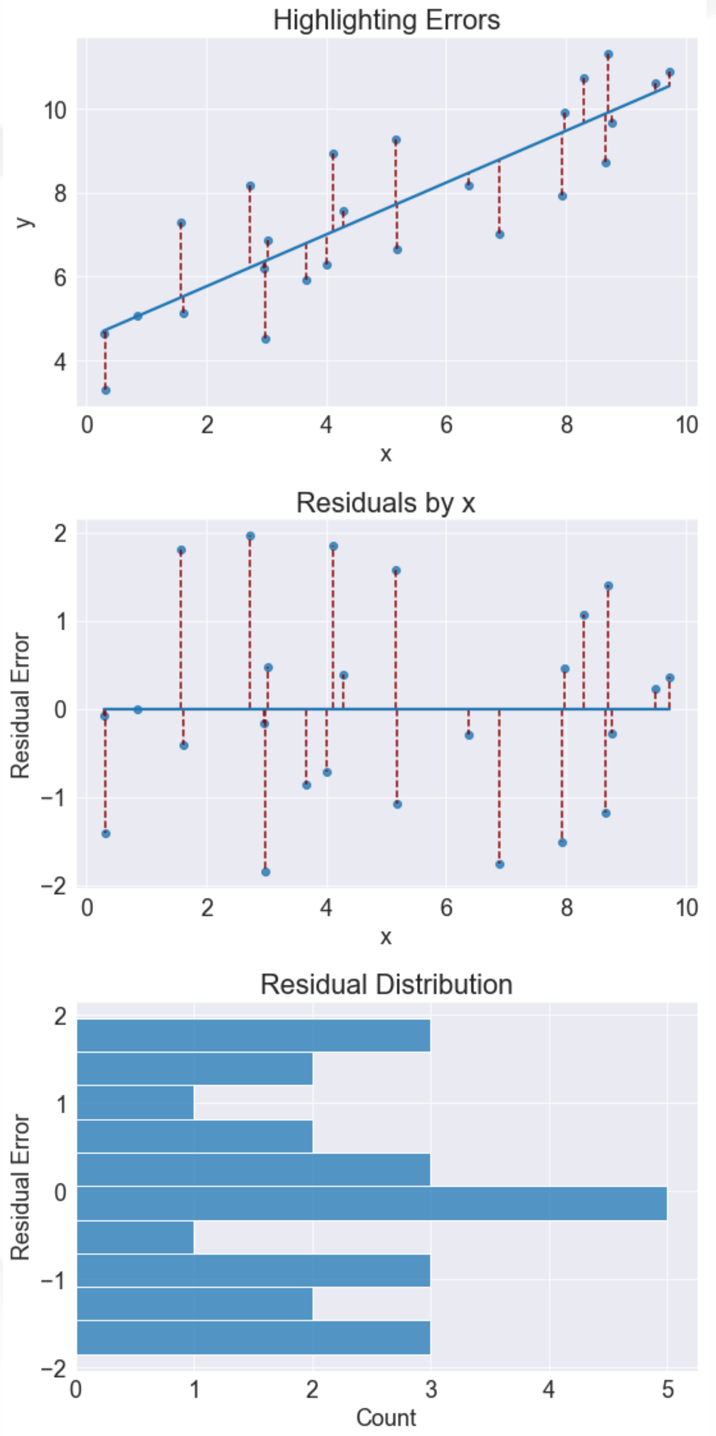
And we said that our linear model assumed that

$$\epsilon \sim N(0, \sigma^2)$$

This leads us to the assumptions underlying an OLS model

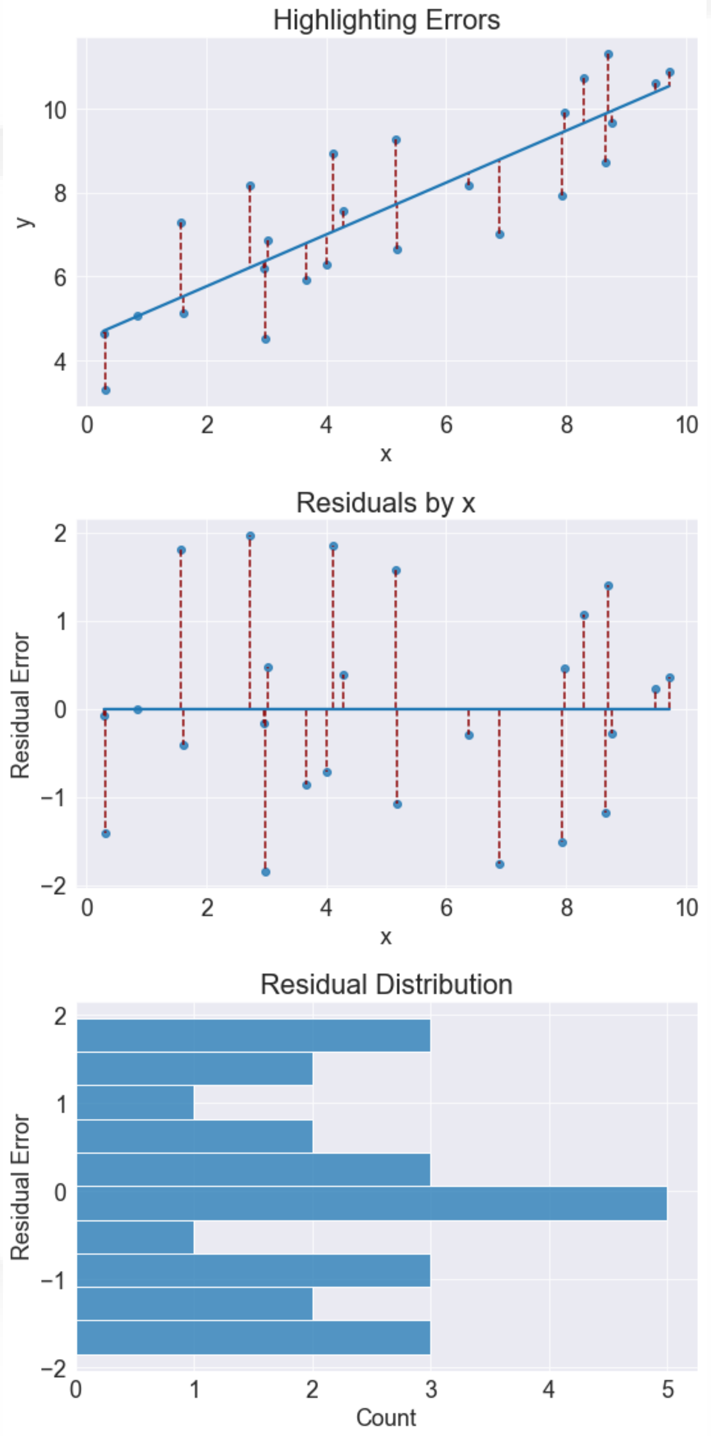
Assumptions

1. The mean value of ϵ is zero
2. The variance of ϵ (some σ^2), is constant for all values of x
3. ϵ is normally distributed
4. Each ϵ_i is iid (independent and identically distributed)



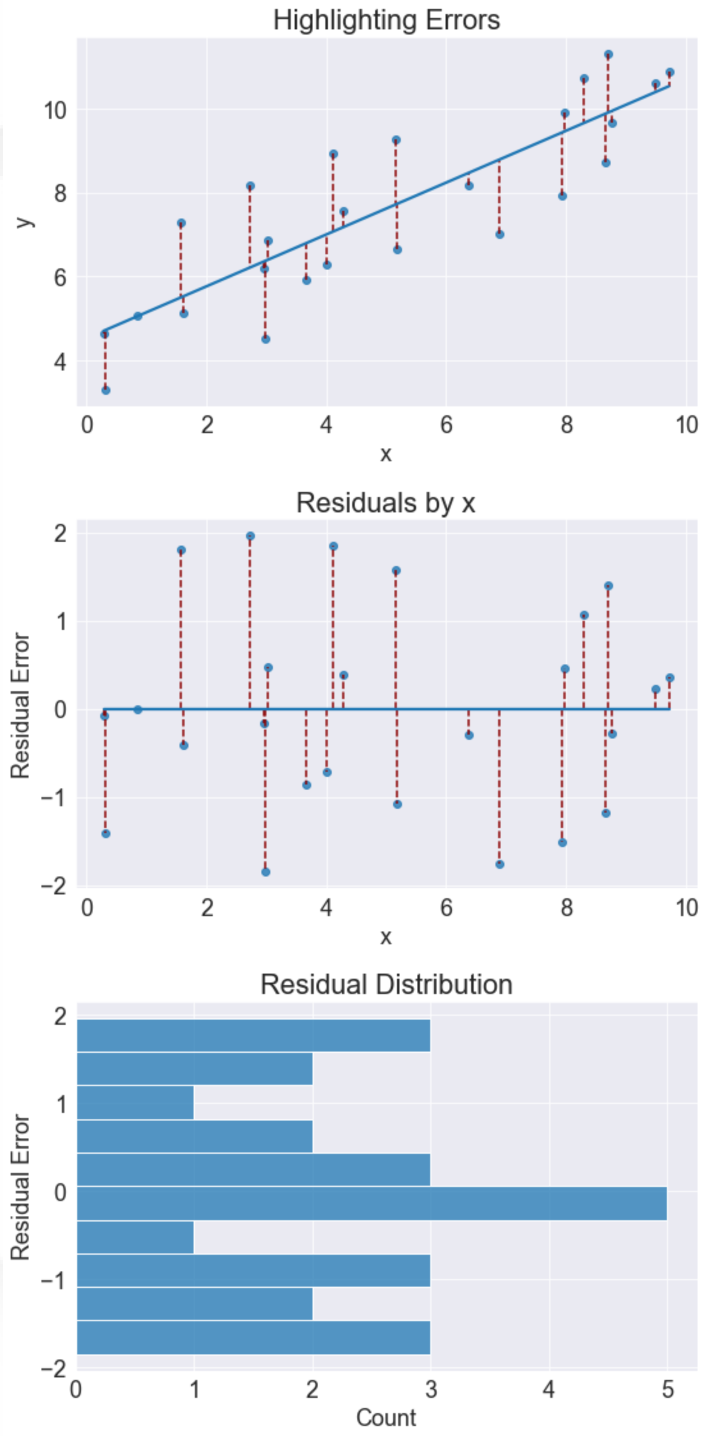
Assumptions

1. **The mean value of ϵ is zero**
2. The variance of ϵ (some σ^2), is constant for all values of x
3. ϵ is normally distributed
4. Each ϵ_i is iid (independent and identically distributed)



Assumptions

1. The mean value of ϵ is zero
2. **The variance of ϵ (some σ^2), is constant for all values of x**
3. ϵ is normally distributed
4. Each ϵ_i is iid (independent and identically distributed)



Checking variance

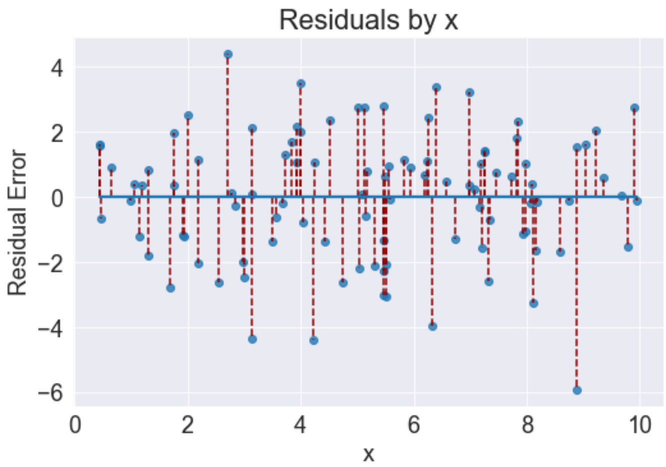
Graphically – look for patterns in the distribution of residuals over x

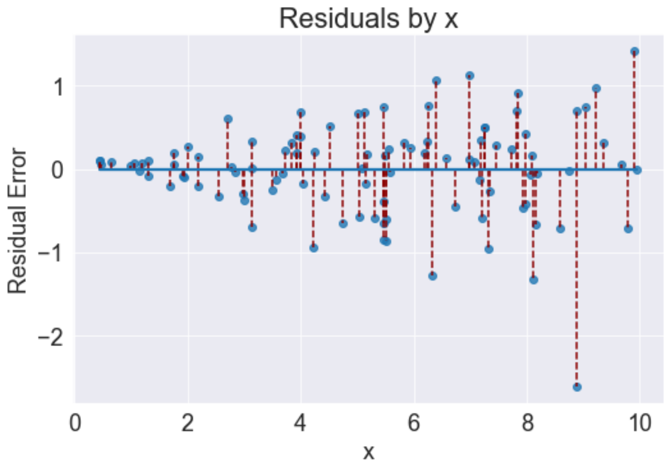
What is our estimate for variance?

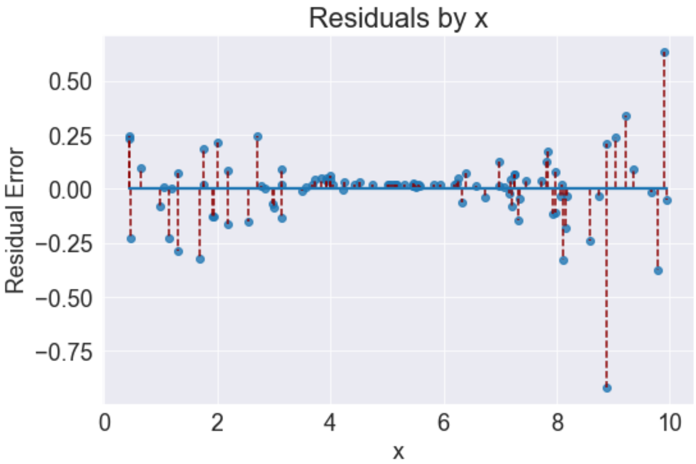
- Mean Squared Error (MSE)

$$s^2 = MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{n-2} = \frac{SS_{yy} - \hat{\beta}_1 SS_{xy}}{n-2}$$

$$SS_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

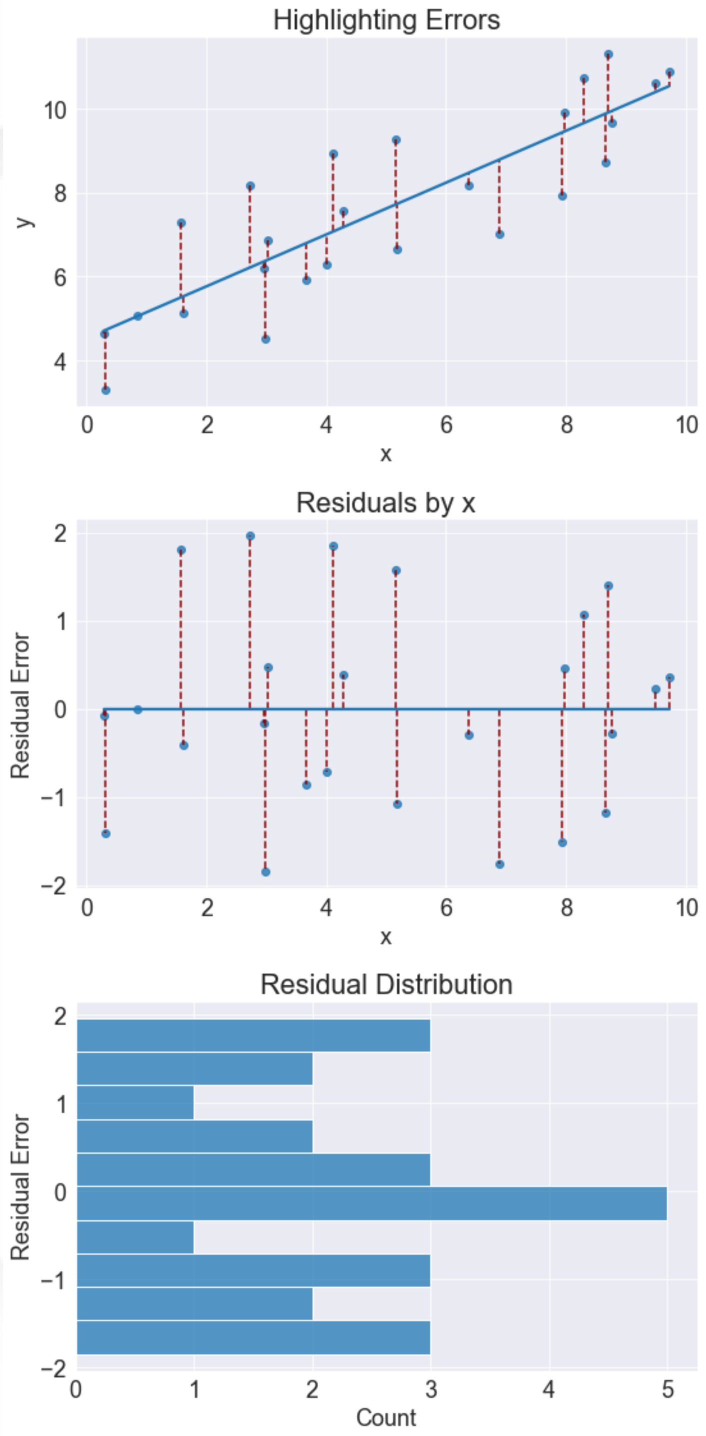






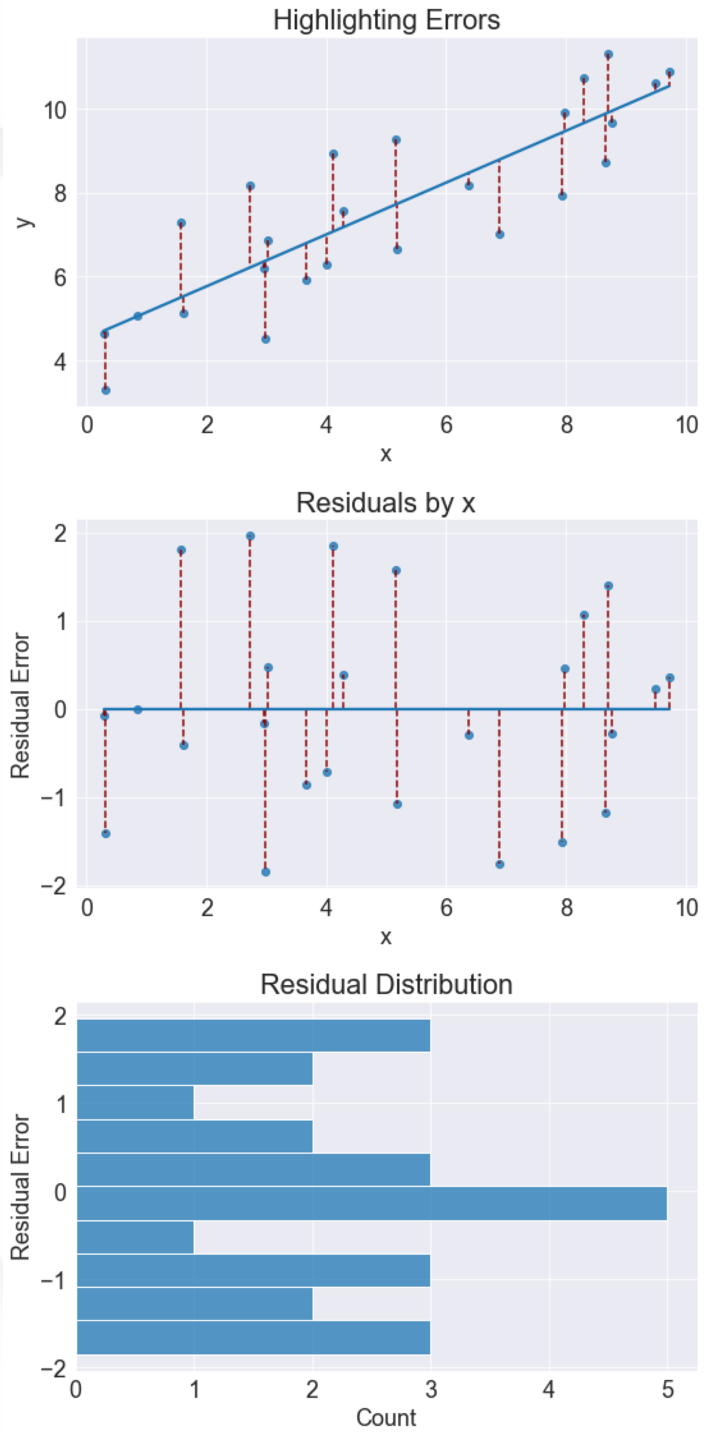
Assumptions

1. The mean value of ϵ is zero
2. The variance of ϵ (some σ^2), is constant for all values of x
3. **ϵ is normally distributed**
4. Each ϵ_i is iid (independent and identically distributed)



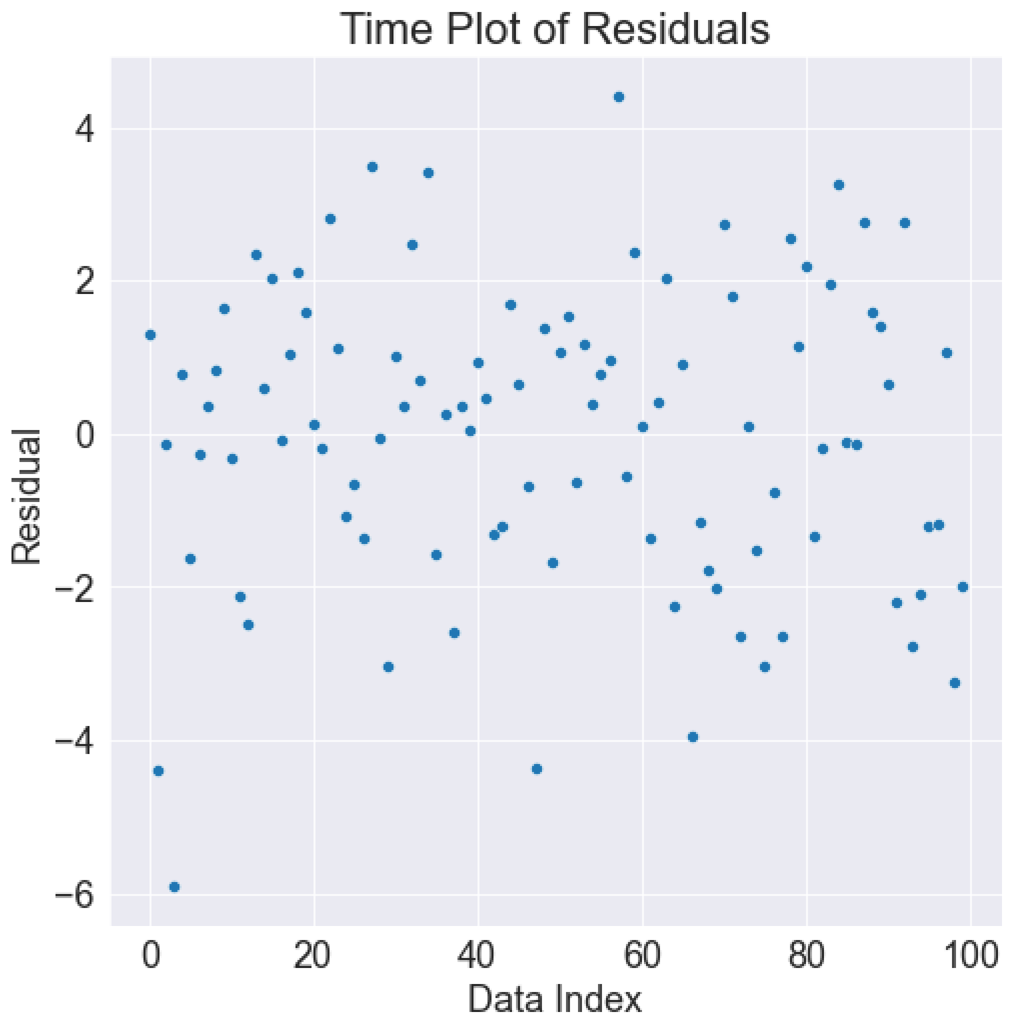
Assumptions

1. The mean value of ϵ is zero
2. The variance of ϵ (some σ^2), is constant for all values of x
3. ϵ is normally distributed
4. **Each ϵ_i is iid (independent and identically distributed)**



Checking independence

- Violations of independence can be very difficult to identify
- One good method is to plot residuals in the order that data was collected and look for patterns
- For this class, you may assume data is iid unless otherwise specified
- For more information, take a class on time-series modeling or forecasting



Example: Rocket Propellant

$$y = 2628 - 37.15x + \epsilon$$

Continuing our example for the last lesson, let’s look at the residuals

Equations Recap

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

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}}$$

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

$$s^2 = MSE = \frac{\sum_{i=1}^{n}(y_i - \hat{y})^2}{n-2}$$

$$= \frac{SS_{yy} - \hat{\beta}_1 SS_{xy}}{n-2}$$

$$SS_{xx} = \sum_{i=1}^{n}(x_i - \bar{x})^2$$

$$SS_{xy} = \sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

$$SS_{yy} = \sum_{i=1}^{n}(y_i - \bar{y})^2$$

Next time…

Hypothesis Tests about the Slope (β_1)