Variance and Standard Deviation

Grinnell College

February 3, 2025

Review

Last time we ended with review of numerical summaries

- Measures of center
- Measures of dispersion

In particular, we considered two varieties: order and moment statistics

Variance

Today, we are going to take a closer look at variance:

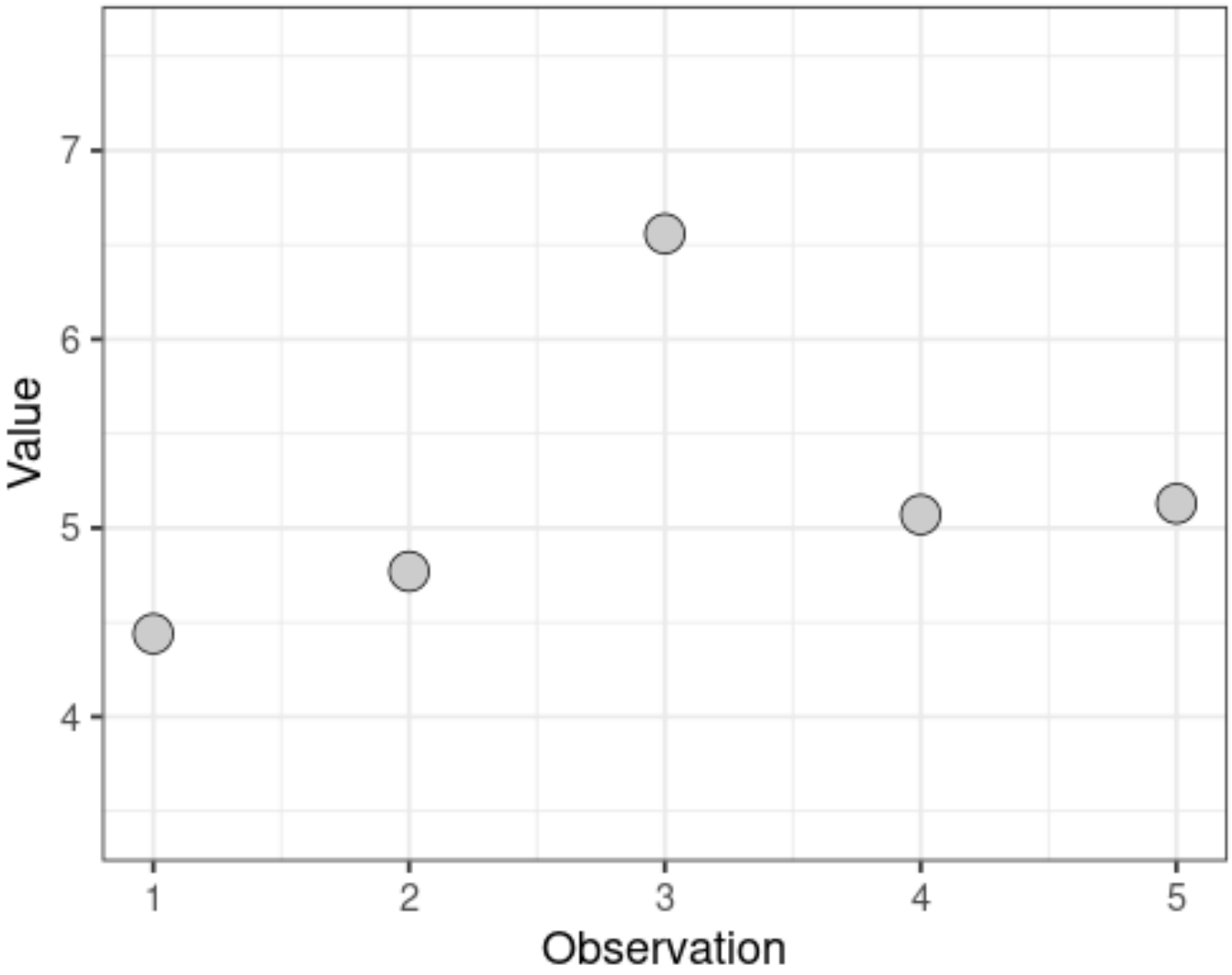
- How is it defined
- Relationship between variance and standard deviation
- What is it used for?
  - Dispersion
  - Uncertainty
  - Prediction

Definitions

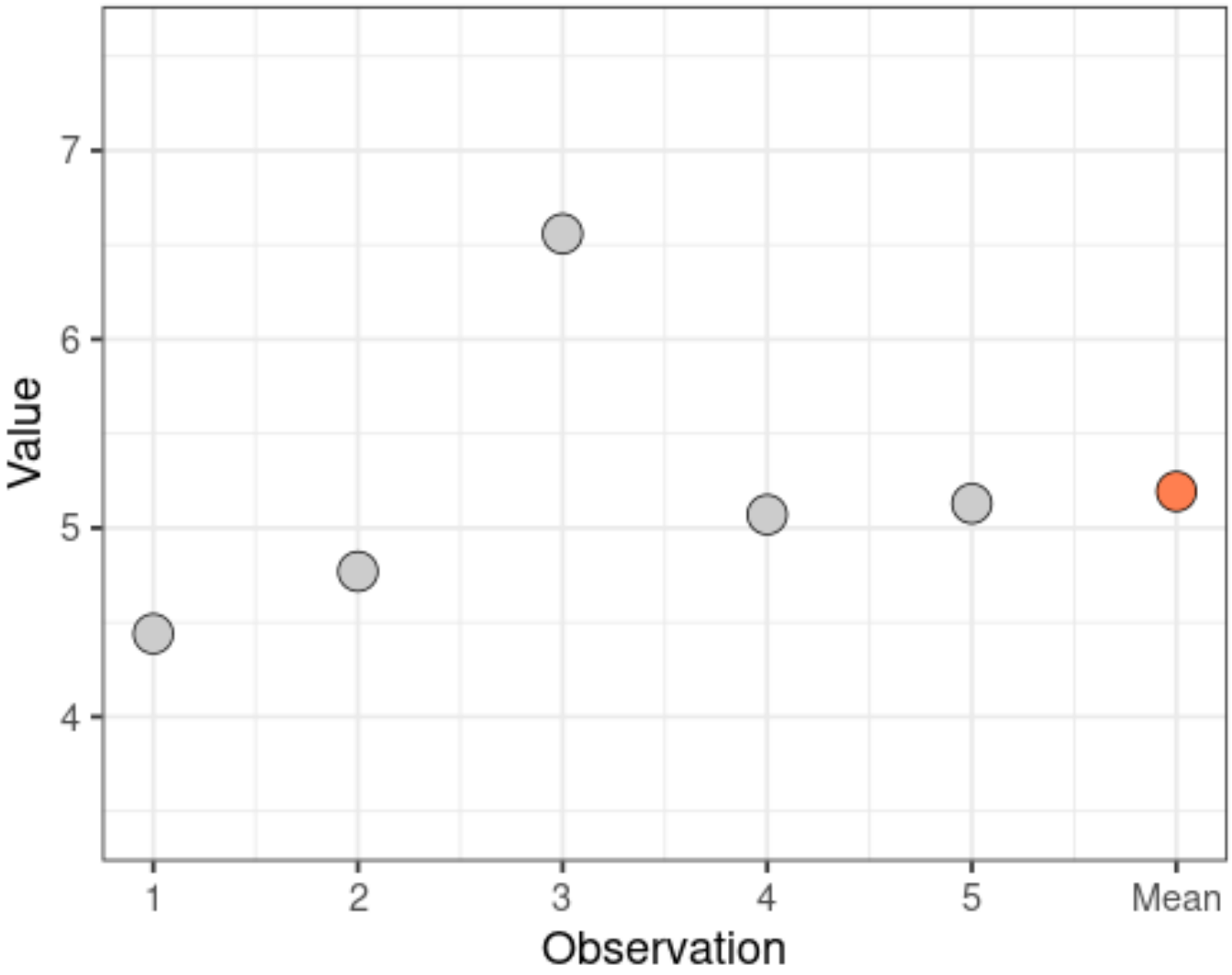
$$\sigma^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2$$

$$\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \overline{x})^2}$$

Just points

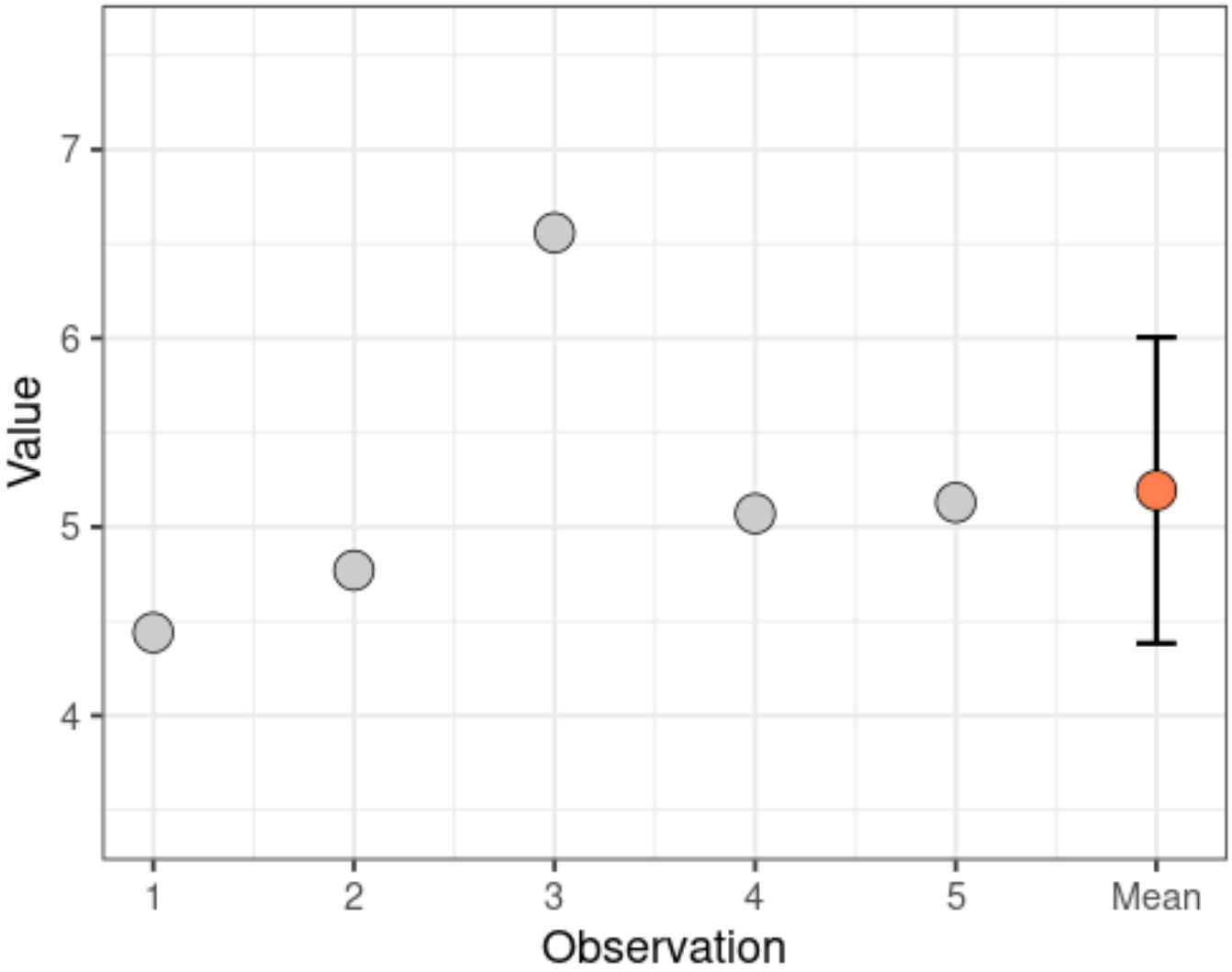


Just points



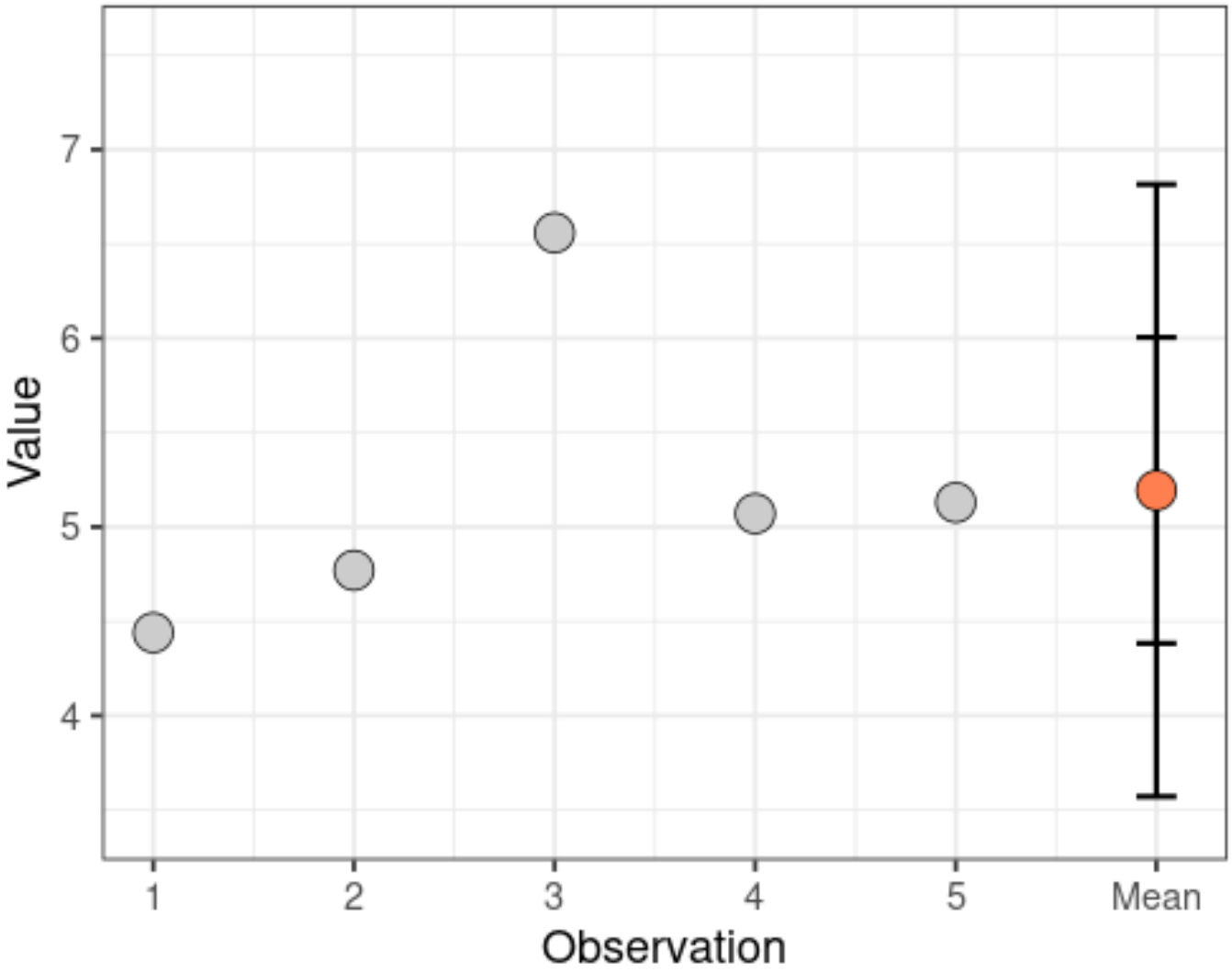
Just points

Here $n = 5$, $\overline{x} = 5.19$ and $\hat{\sigma} = 0.81$



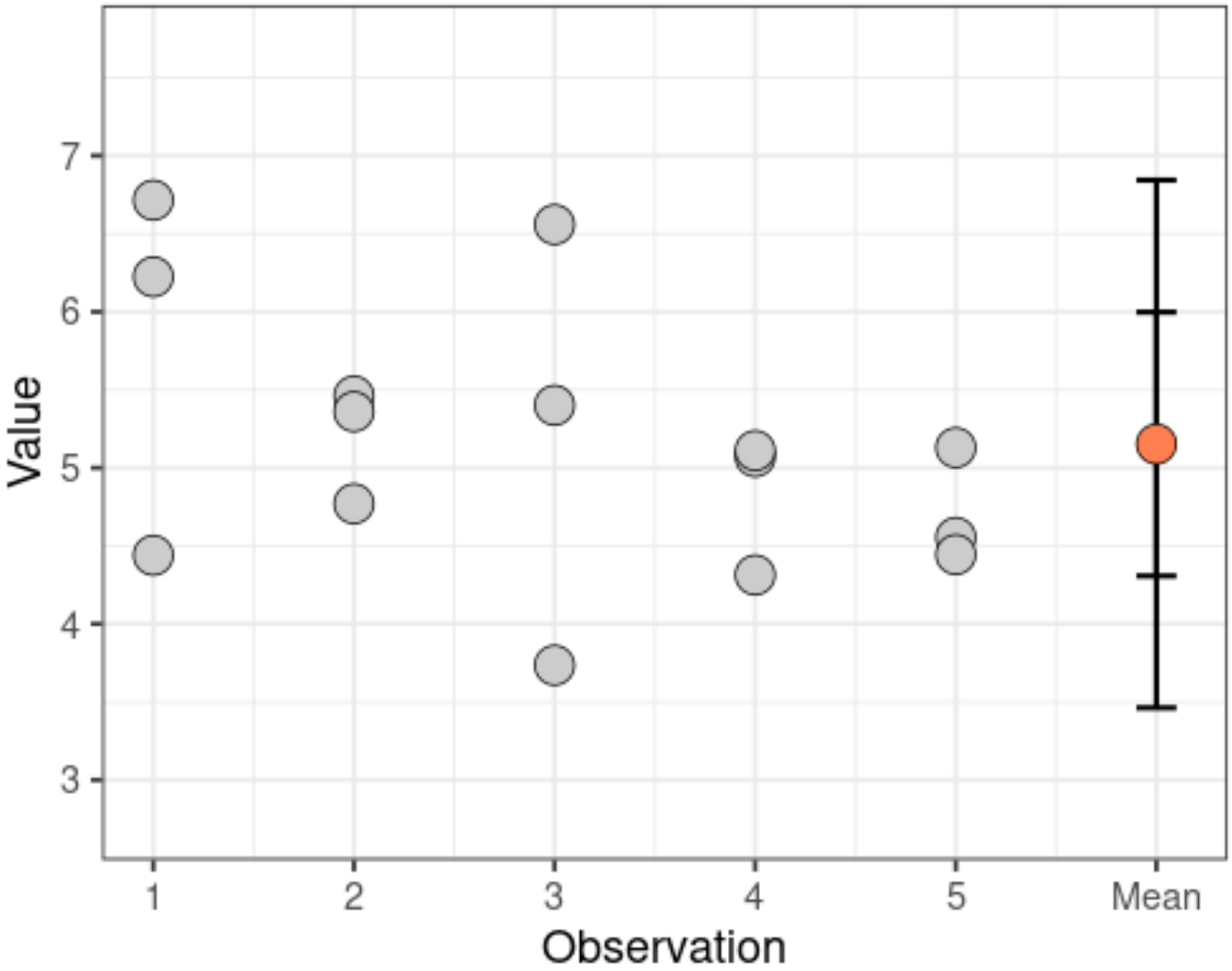
Just points

Here $n = 5$, $\overline{x} = 5.19$ and $\hat{\sigma} = 0.81$



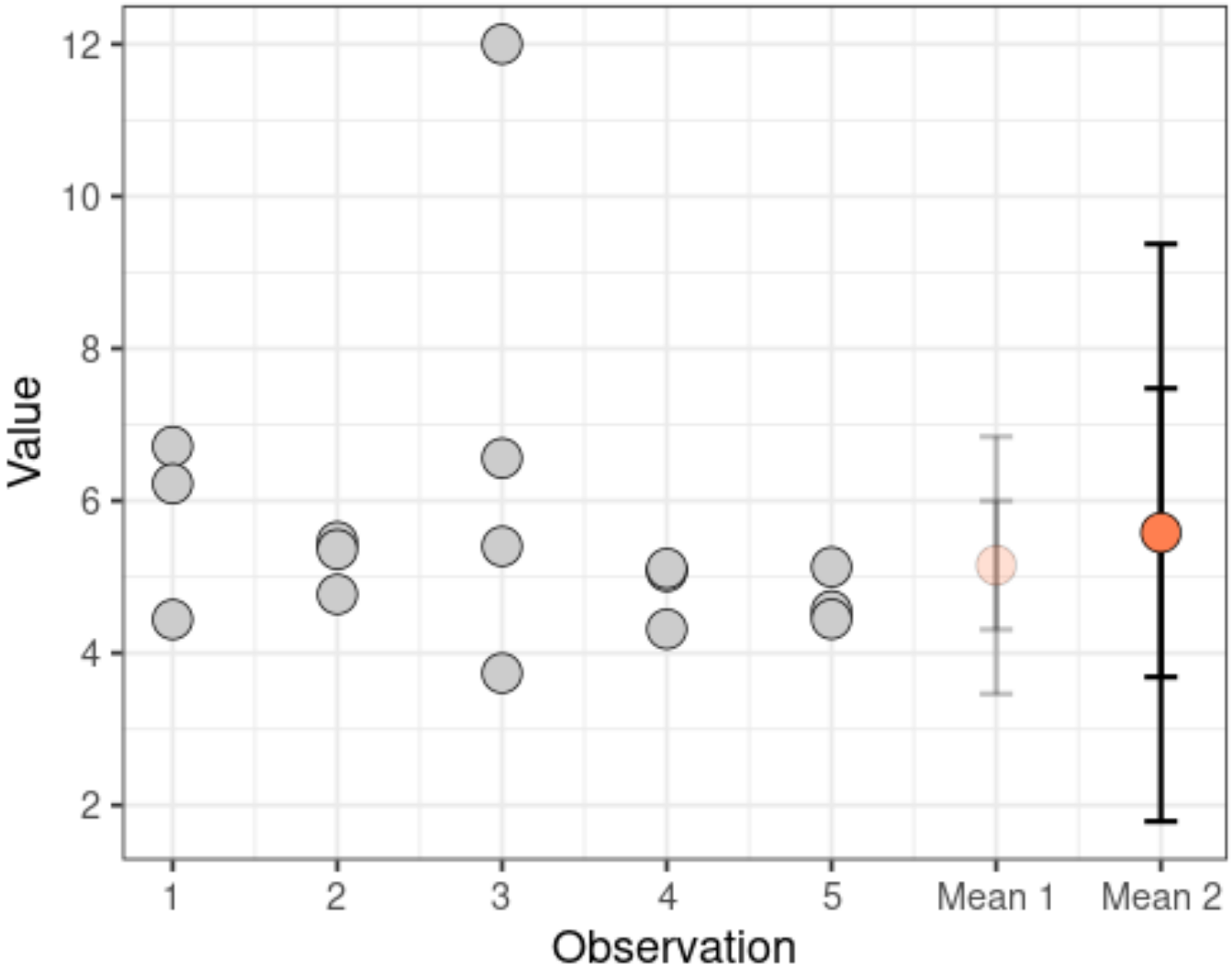
Just points

Note that it is not impacted by the number of observations. Here $n = 10$, $\overline{x} = 5.15$ and $\hat{\sigma} = 0.83$



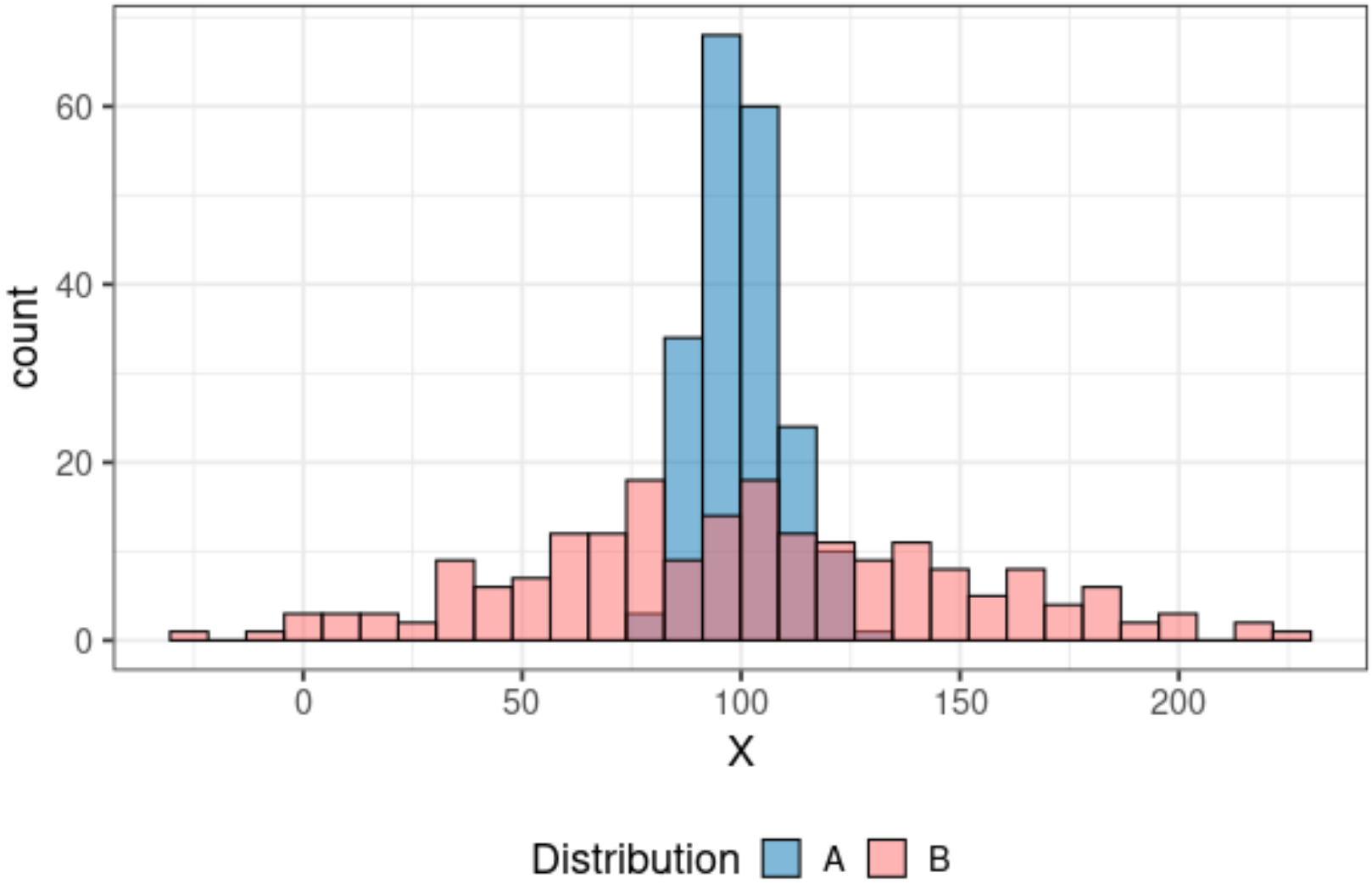
Outlier

Now $n = 11$, $\overline{x} = 5.6$ and $\hat{\sigma} = 1.9$

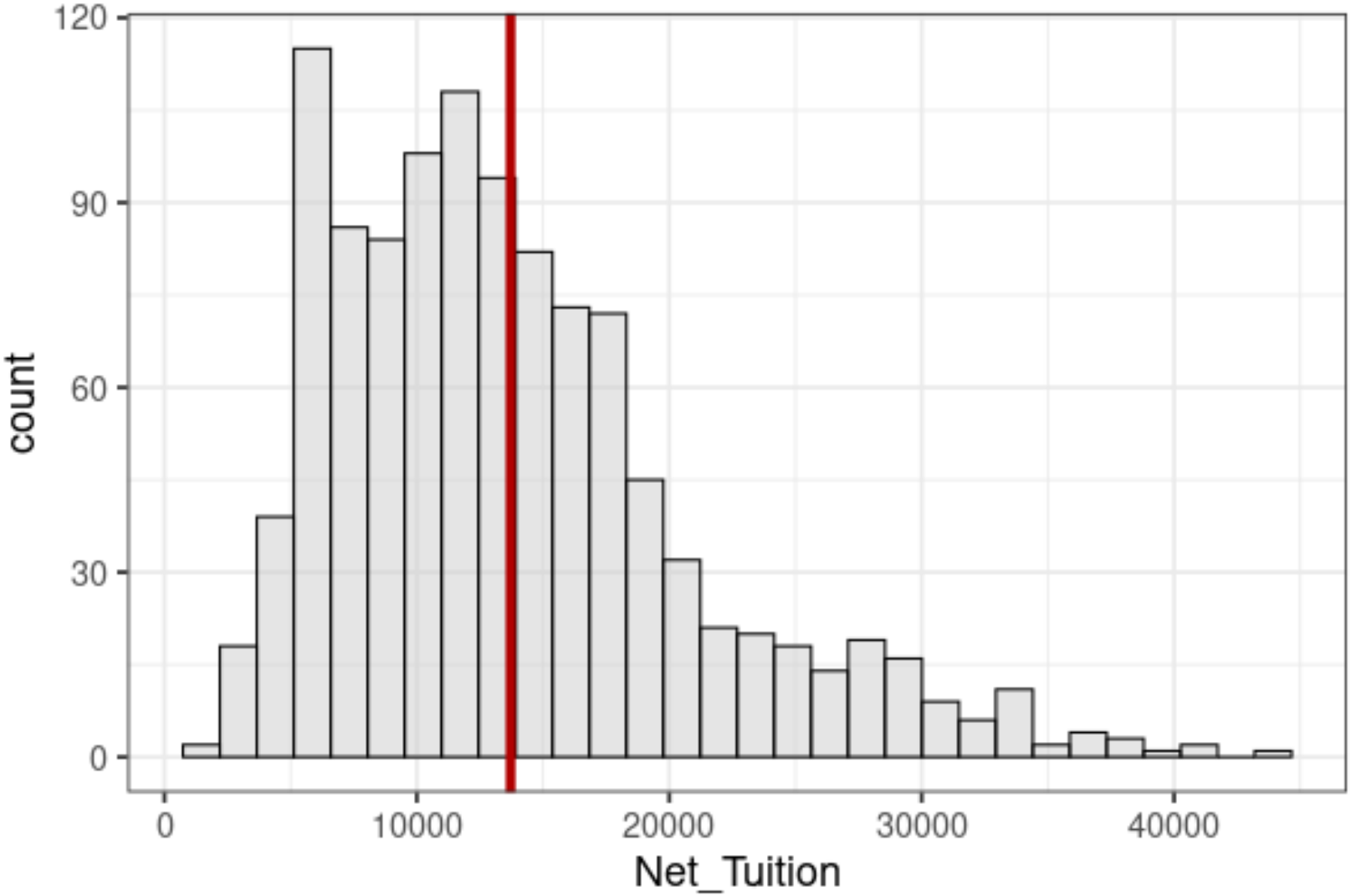


Dispersion

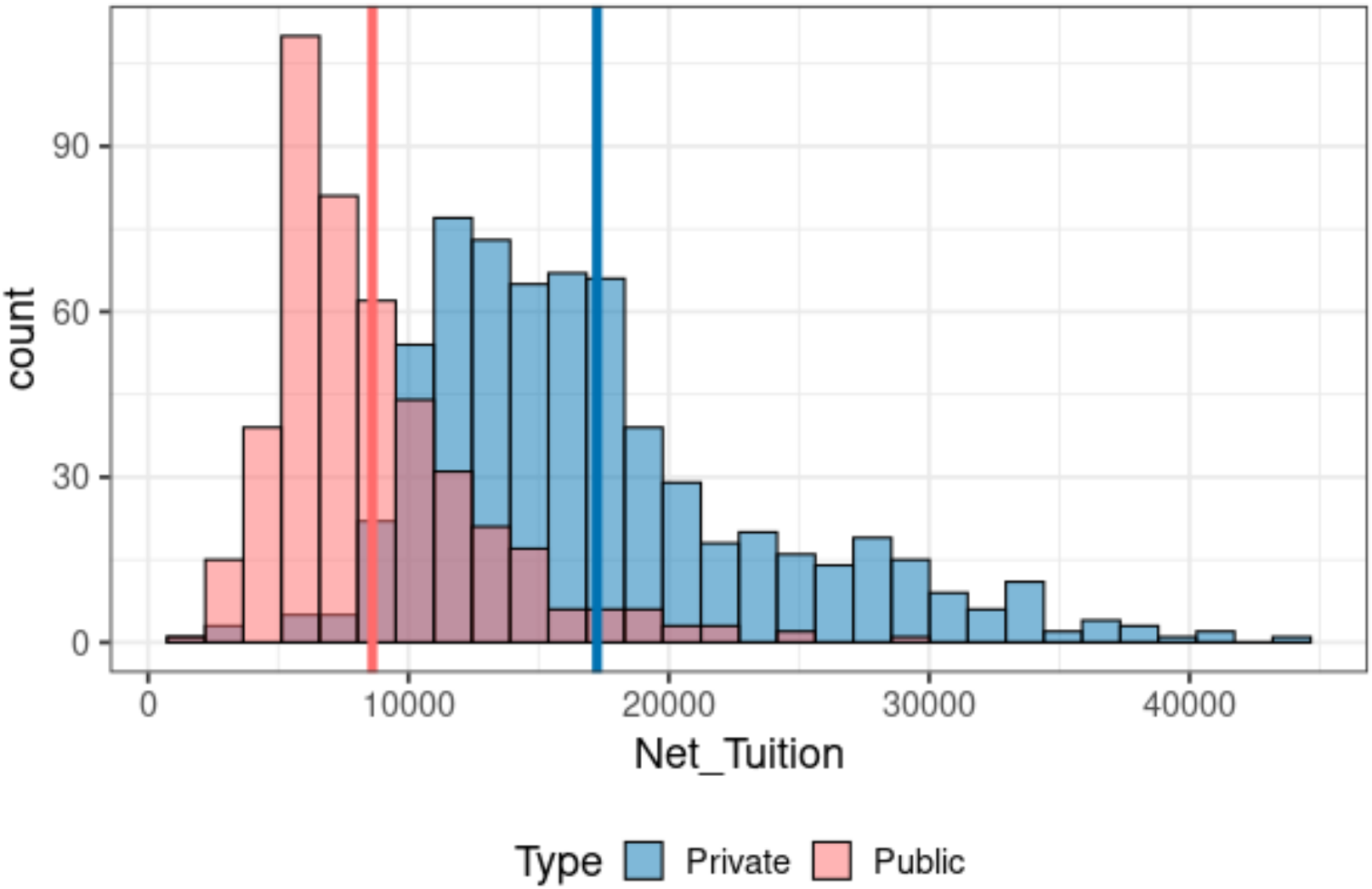
Both of these have $\mu = 100$



Better Centers?



Better Centers?



Main Takeaways

Variance and standard deviation are metrics of dispersion
Tell us how far things are from mean
Identify outliers
Allows us to see uncertainty based on a point estimate
Allows us to compare different centers to see if they offer improvement
we will never have to do by hand

Advantages and Disadvantages

Order Statistics

Advantages:

- Robust to outliers
- More "correct" center for skew

Disadvantages:

- Discards most data
- No nice math properties

Moment Statistics

Advantages:

- Very useful math properties for inference
- Utilizes all of the data

Disadvantages:

- Sensitive to outliers
- Sensitive to skew