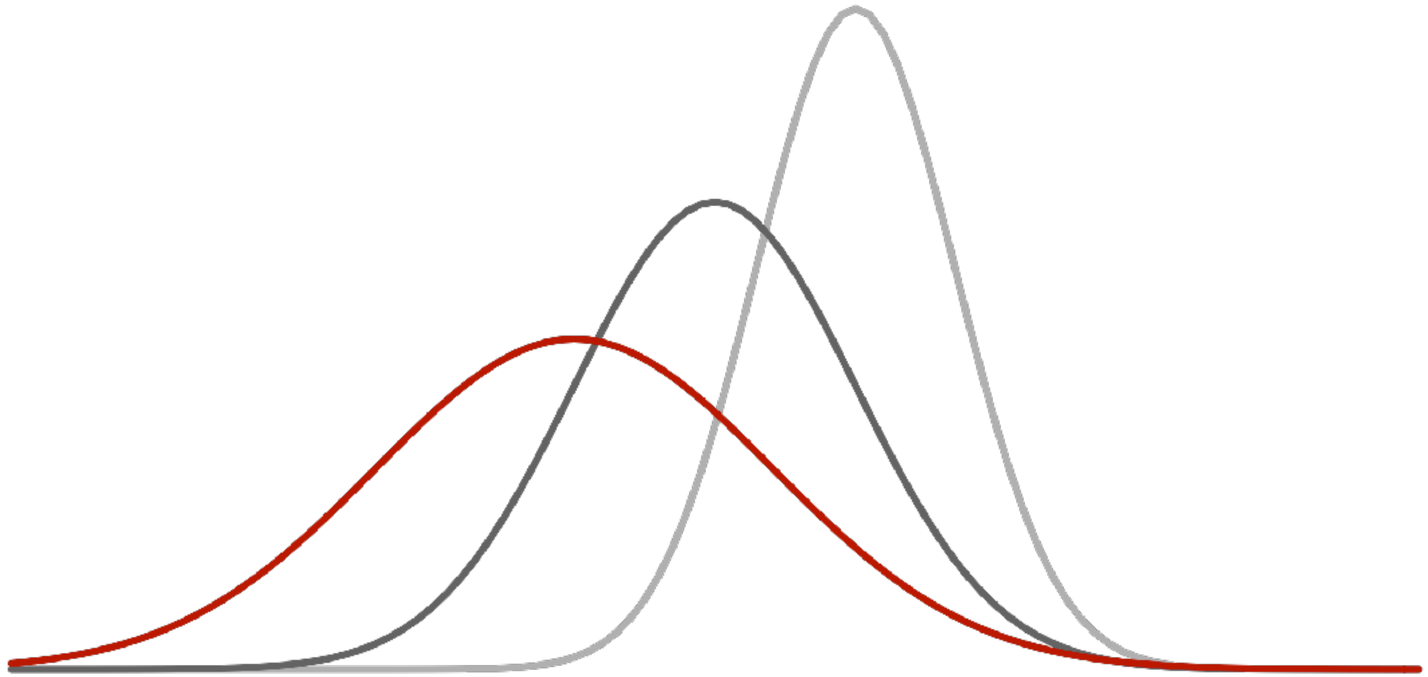


Introduction to Statistical Inference

STAT 3202 | OSU | Autumn 2018

Dalpiaz

Let’s take a trip back to STAT 3201…

Example: Assume that Professor Professorson receives emails according to a Poisson process with an average rate of two per day.

- What is the probability that Professor Professorson receives two emails on a particular day?
- What is the probability that Professor Professorson receives ten emails on a particular day?
- What is the probability that Professor Professorson receives at least three emails on a particular day?

Example: Assume that the amount of sleep obtained by undergraduate students at The Ohio State University follows a normal distribution with a mean of 7.5 hours and a variance of 0.5 hours.

- What is the probability that a randomly selected student sleeps more than 8 hours a night?

- What is the probability that the average sleep time of a randomly selected group of 25 students is greater than 8 hours?

Where’s the **data**?

3202!

Example: Professor Professorson received the following number of emails on each of the previous 40 days:

5 6 2 5 3 3 4 1 4 4 3 4 6 2 3 6 7 1 3 3

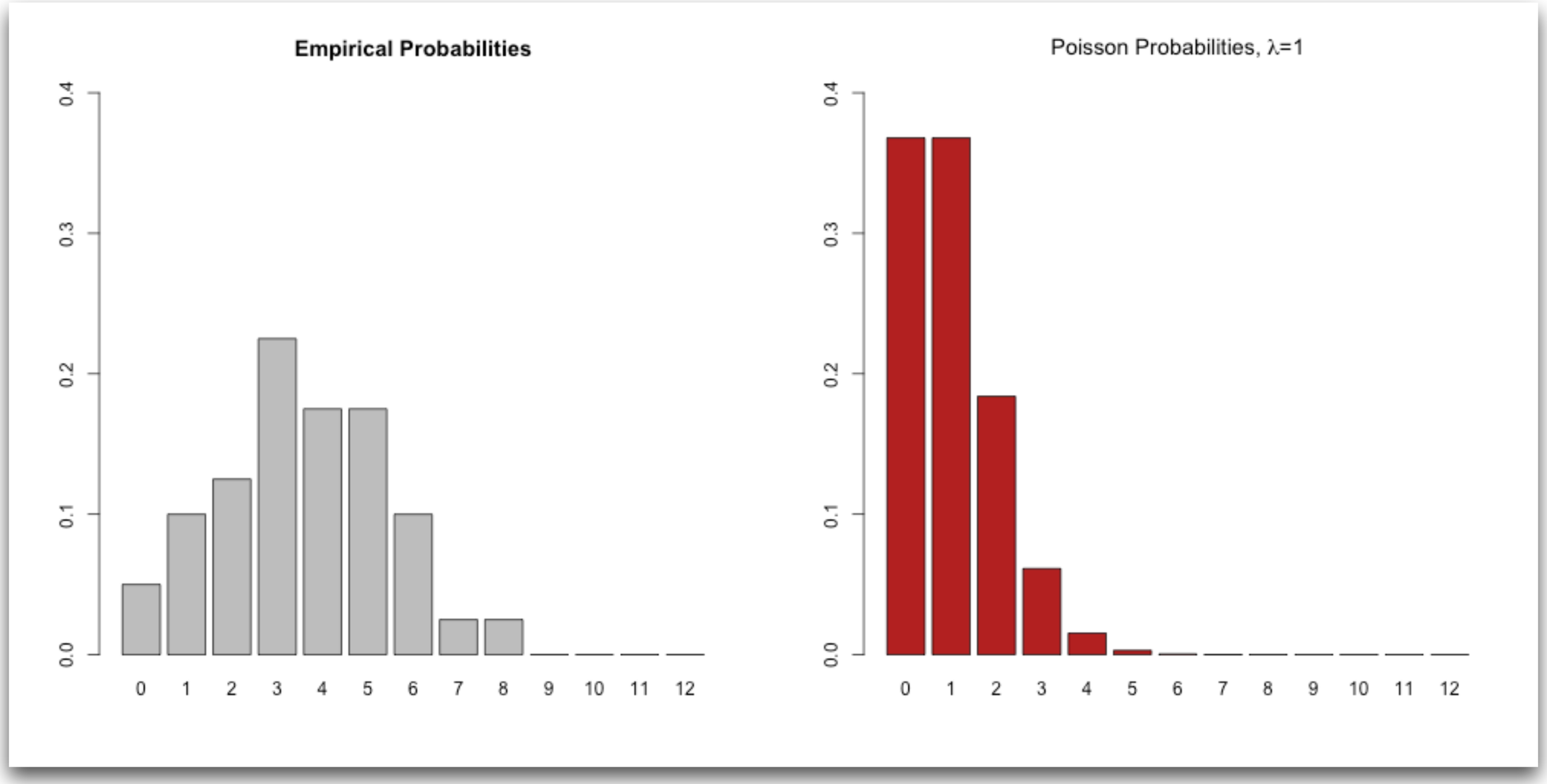
5 1 8 6 1 3 2 5 3 5 4 4 2 4 0 5 0 2 5 3

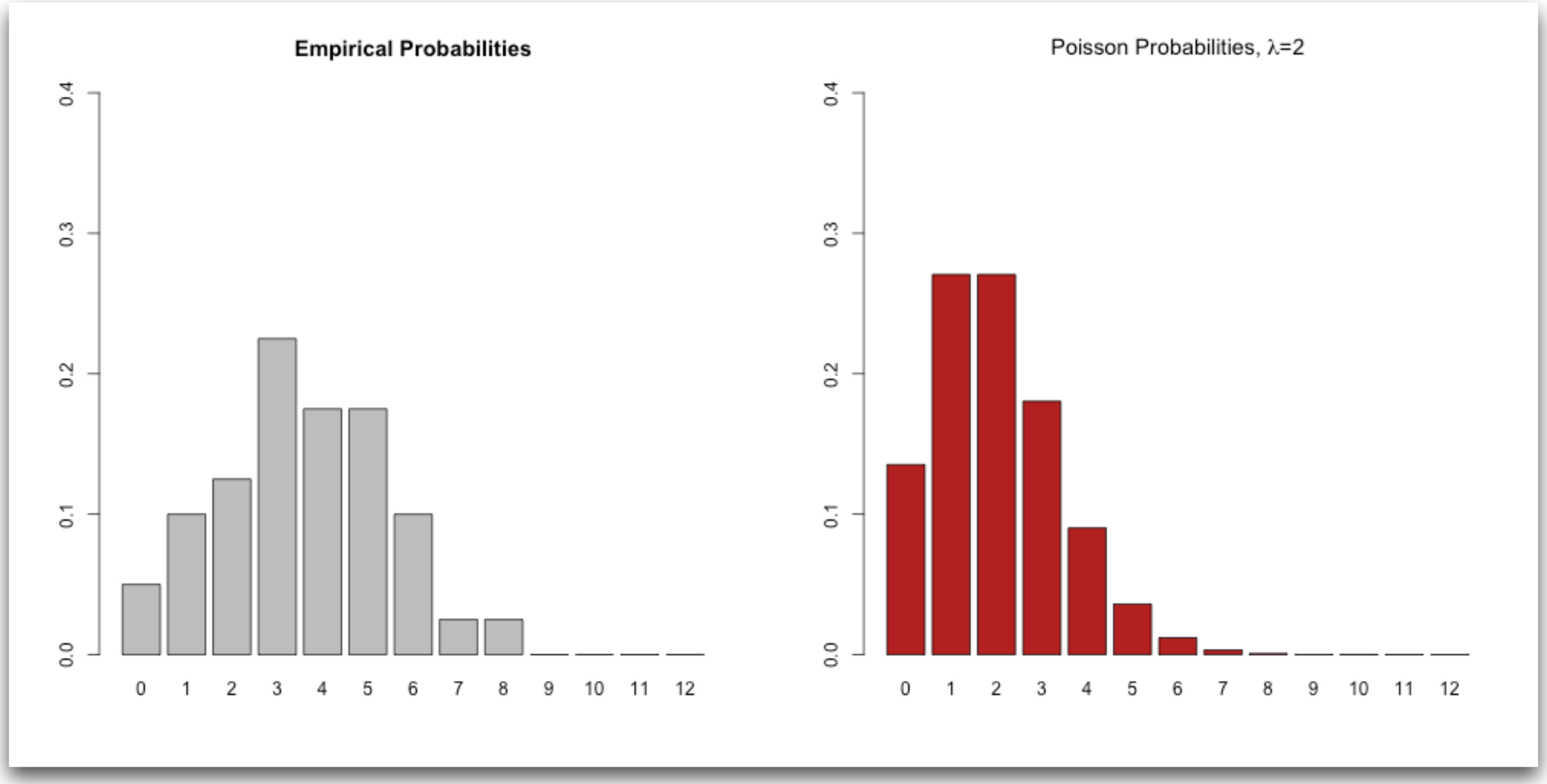
- What is the probability that Professor Professorson receives three emails on a particular day?

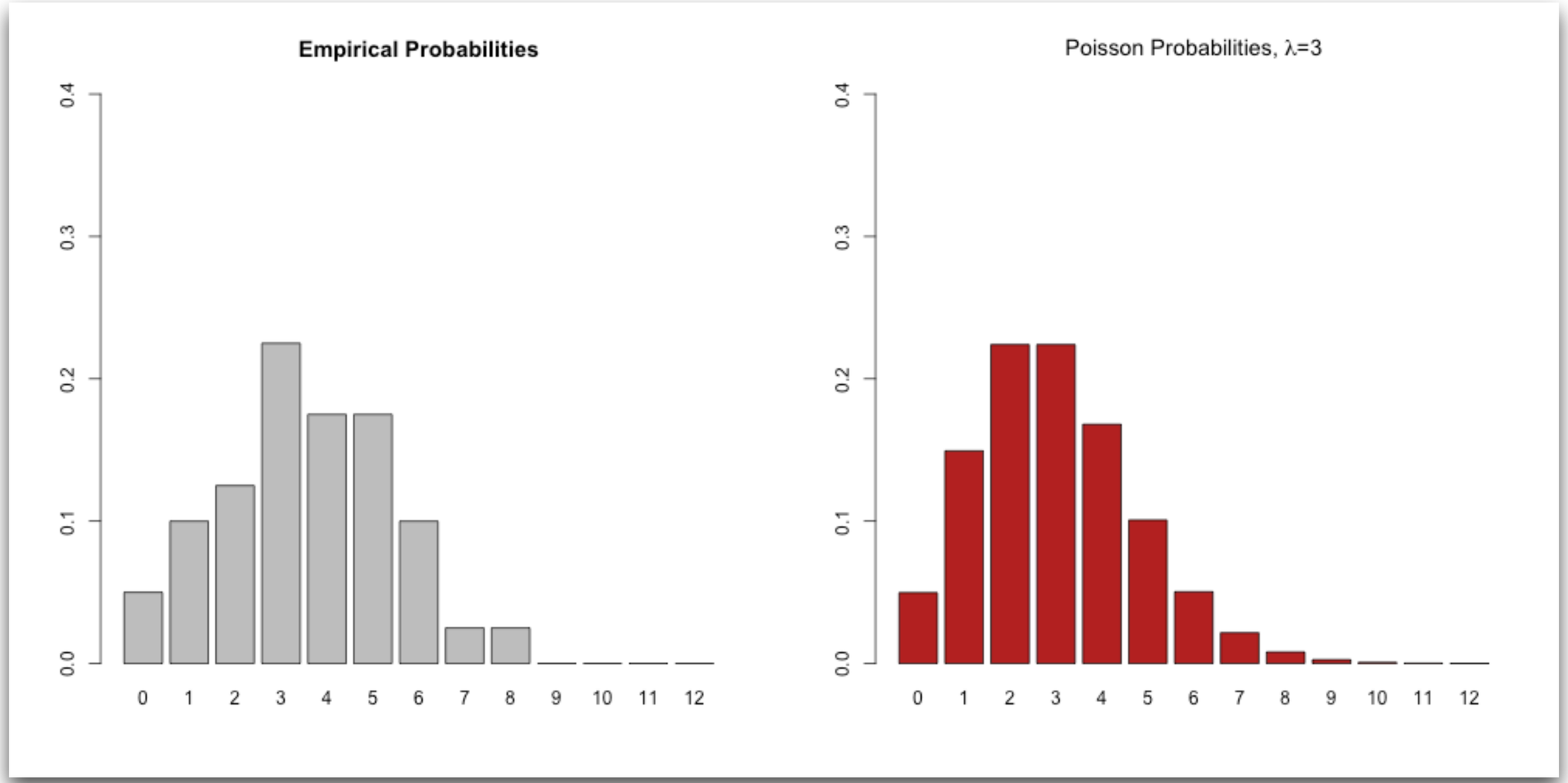
- What is the probability that Professor Professorson receives nine emails on a particular day?

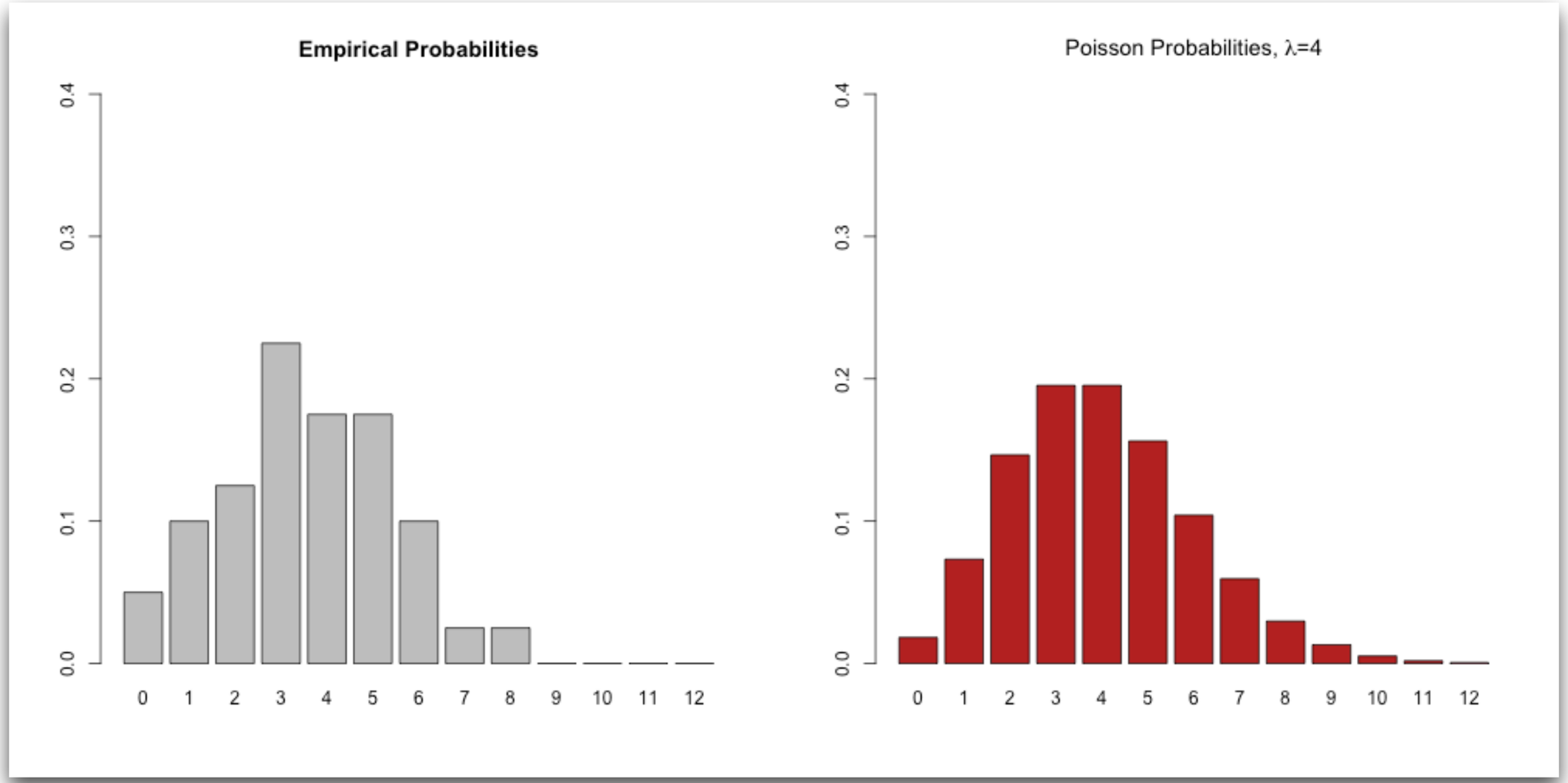
Idea: Use a Poisson distribution to model the number of emails received per day.

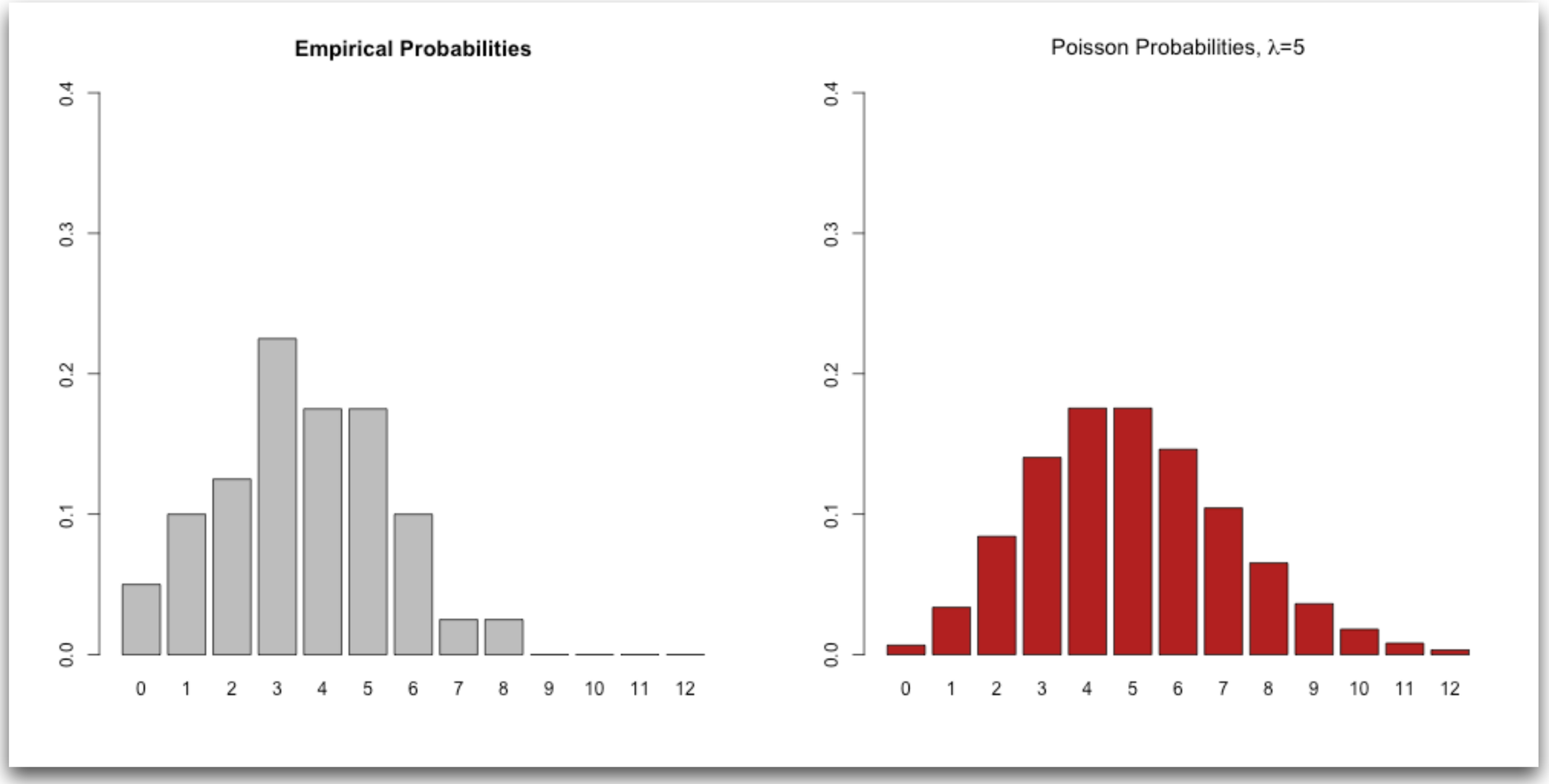
But which Poisson distribution?

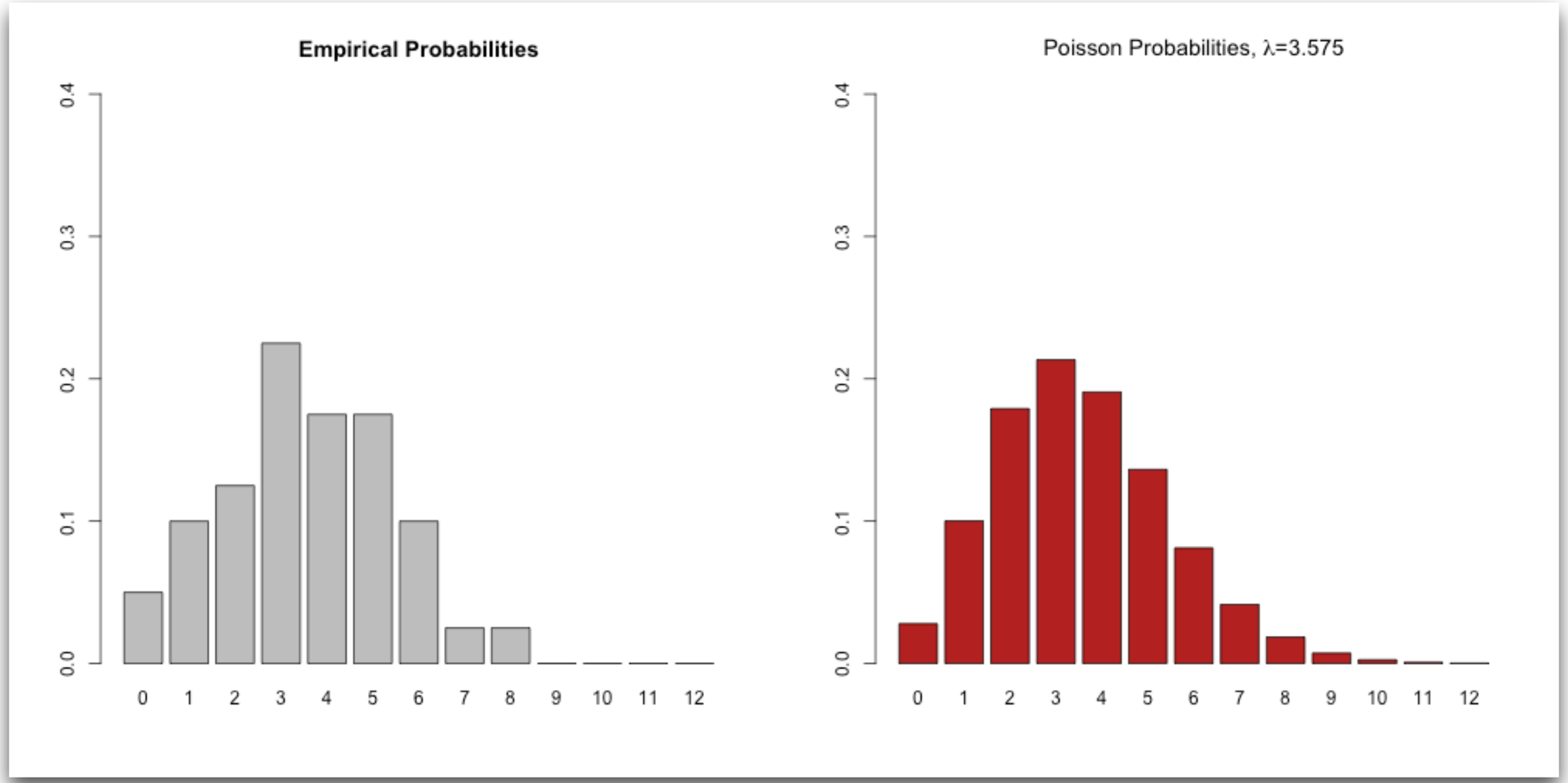

Empirical Probabilities
0.0
0.1
0.2
0.3
0.4
0 1 2 3 4 5 6 7 8 9 10 11 12
Poisson Probabilities, λ=1
0.0
0.1
0.2
0.3
0.4
0 1 2 3 4 5 6 7 8 9 10 11 12


Empirical Probabilities
0.0
0.1
0.2
0.3
0.4
0 1 2 3 4 5 6 7 8 9 10 11 12
Poisson Probabilities, λ=2
0.0
0.1
0.2
0.3
0.4
0 1 2 3 4 5 6 7 8 9 10 11 12


Empirical Probabilities
0.0
0.1
0.2
0.3
0.4
0 1 2 3 4 5 6 7 8 9 10 11 12
Poisson Probabilities, λ=3
0.0
0.1
0.2
0.3
0.4
0 1 2 3 4 5 6 7 8 9 10 11 12


Empirical Probabilities
0.0
0.1
0.2
0.3
0.4
0 1 2 3 4 5 6 7 8 9 10 11 12
Poisson Probabilities, λ=4
0.0
0.1
0.2
0.3
0.4
0 1 2 3 4 5 6 7 8 9 10 11 12


Empirical Probabilities
0.0
0.1
0.2
0.3
0.4
0 1 2 3 4 5 6 7 8 9 10 11 12
Poisson Probabilities, λ=5
0.0
0.1
0.2
0.3
0.4
0 1 2 3 4 5 6 7 8 9 10 11 12


Empirical Probabilities
0.0
0.1
0.2
0.3
0.4
0 1 2 3 4 5 6 7 8 9 10 11 12
Poisson Probabilities, λ=3.575
0.0
0.1
0.2
0.3
0.4
0 1 2 3 4 5 6 7 8 9 10 11 12

The previous “analysis” is flawed.

- The previous 40 days is **not** a random sample.
    - Why do we need a random sample?
- We were just guessing and checking.
    - How do we know if Poisson was a reasonable distribution?
    - How do we pick a good lambda given some *data*?
    - How do we know if our method for picking is good?

What is **Statistics**?

What is **Statistics**?

- Technically: The study of **statistics**.
- Practically: The science of collecting, organizing, analyzing, interpreting, and presenting **data**.

What is a **statistic**?

What is a **statistic**?

- Technically: A function of (sample) data.

Terminology

- **Population:** The entire group of interest.
- **Parameter:** A (usually unknown) numeric value associated with the population.
- **Sample:** A subset of individuals taken from the population.
- **Statistic:** A numeric value computed using the sample data.

What is an **estimator**?

- A **statistic** that attempts to provide a good guess for an unknown population parameter.

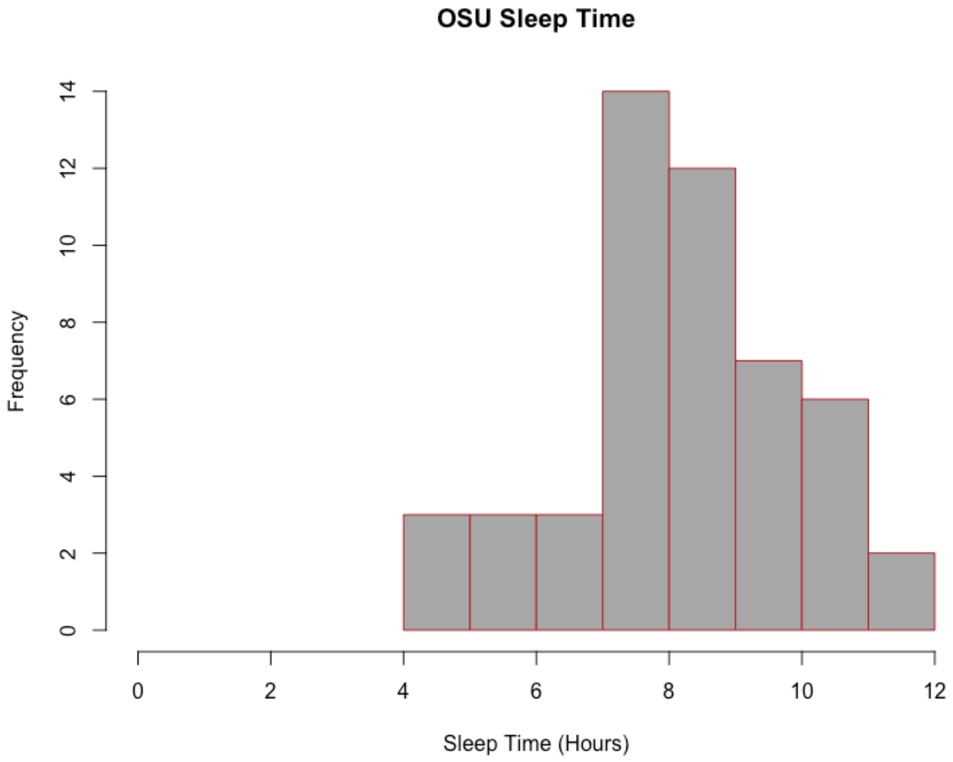
“The objective of Statistics is to make an inference about a population based on information contained in a sample from that population and to provide an associated measure of goodness for the inference.”

–WMS

What is **statistical inference**?

Making evidence based claims about a population by using data.

Example: An administrator claims that undergraduate students at Ohio State are extremely healthy. In particular, she claims that they sleep at least 8 hours a night on average. To test this claim, a random sample of 50 students is selected to report on the amount of sleep they obtained the previous night. They slept on average 8.15 hours, with a standard deviation of 1.63 hours. Do you believe the administrator's claim?



OSU Sleep Example

- **Population:** All 45,946 OSU undergraduate students.
- **Parameter:** The mean, μ, the average sleep time for OSU undergraduates.
- **Sample:** The 50 students chosen at random.
- **Statistic:** The sample mean and sample standard deviation.

Random Sample

- A **random sample** is a sample where each individual in the population is equally likely to be included.

Why do we care?

Statistics are Random Variables

- **A statistic:** A function that tells us what calculation we will perform after taking a sample from the population.

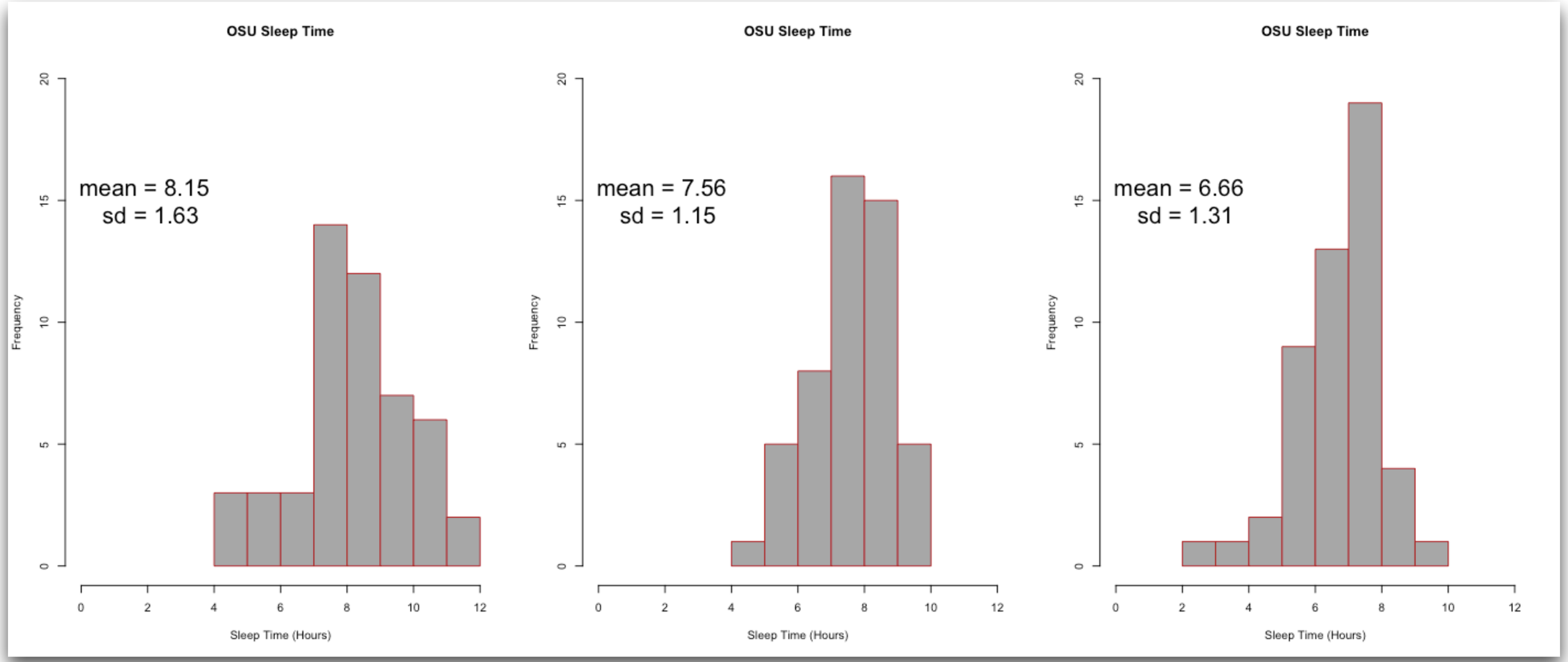
- **The value of a statistic:** The numeric result of performing the calculation on a *particular* sample of data.

X vs x

Sampling Distributions

Because statistics are *random variables*, they have a **distribution**!

What if…


OSU Sleep Time
mean = 8.15
sd = 1.63
Frequency
Sleep Time (Hours)
OSU Sleep Time
mean = 7.56
sd = 1.15
Frequency
Sleep Time (Hours)
OSU Sleep Time
mean = 6.66
sd = 1.31
Frequency
Sleep Time (Hours)

Where are we going?

- Near Future: **Estimation**
    - How do we decide if an estimator is good?
    - How do we create estimators?
    - How to we account for variability in estimation?
- Later: **Testing**
    - How do we evaluate a statistical hypothesis?
    - How do we assess how well a testing procedure works?
    - How to we translate a real world hypothesis into a statistical hypothesis?

Questions?

Comments? Concerns?


That's all Folks!