

JEOPARDY!

Grinnell College

November 6, 2024

Rules

- ▶ Groups of 3-4
- ▶ Write answer with team name on note card for each question
- ▶ Daily double will be randomly determined by dice rolls
- ▶ Have R ready to go!
- ▶ Scores tabulated by total money at end
 1. First place is 4 extra credit points on exam
 2. Second place is 2 extra credit points on exam
 3. One point for all other participants

Probability

Probability 1

Of 1,000 people, 80% like peanut butter, 89% like jelly, and 78% like both.

Create a Venn Diagram that illustrates this relationship

Probability 2

Suppose I have a jar with 3 red marbles, 4 blue marbles, and 5 green marbles

If I draw three marbles in succession without replacement, what is the probability that I draw two red and one green?

Suppose that $P(A) = 0.2$ and $P(B) = 0.8$

1. If $P(A \text{ and } B) = 0.15$, are A and B independent?
2. If $P(A \text{ and } B) = 0.15$, what is $P(A|B)$?

Probability 4

Following an introductory statistics course, 80% of students are able to successfully construct a box plot. Of those who can construct a box plot, 86% received a passing grade. Of those who could not construct a box plot, only 65% of students passed

Given that a student received a passing exam, what is the probability that they could also construct a box plot?

Probability 5

Bob eats out every Friday evening to either a Mexican or Italian restaurant. When he goes to a Mexican restaurant, there is a 30% chance he eats chicken, a 15% chance he eats pork, and a 65% chance he has a vegetarian meal. When he goes to an Italian restaurant, there is a 40% chance he eats chicken, a 35% chance he eats pork, and a 25% chance he eats vegetarian. He eats at a Mexican restaurant 60% of the time.

Given that he is eating a vegetarian meal, what is the probability that he is eating at a Mexican restaurant?

Sampling Distributions

Sampling Distribution 1

What are the distributional parameters associated with:

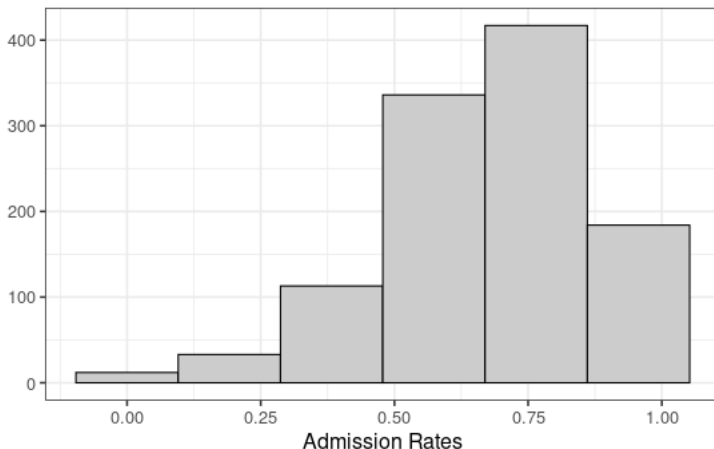
- ▶ The Normal Distribution
- ▶ The t -Distribution

Sampling Distribution 2

What are two factors that impact the amount of variability present in a sampling distribution?

Sampling Distribution 3

- ▶ What would I need to apply the CLT to find confidence intervals for the admission data shown below?
- ▶ For a given size n , how could I verify if the sampling distribution is approximately normal?



PLACE YOUR BETS

Randomly sample ONE value from a t distribution with $df = 15$

A. $|t| > 1$ (2 to 3 odds)

B. $|t| \leq 1$ (3 to 2 odds)

PLACE YOUR BETS

Sampling Distributions 5

Using the `penguins` dataset provided in the lab, do the following:

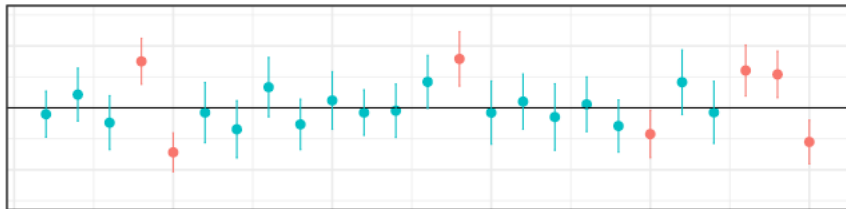
1. Subset the data to only include Adelie penguins
2. Find the standard error associated with the sampling distribution for the median of the variable `flipper_length_mm`

Confidence Intervals

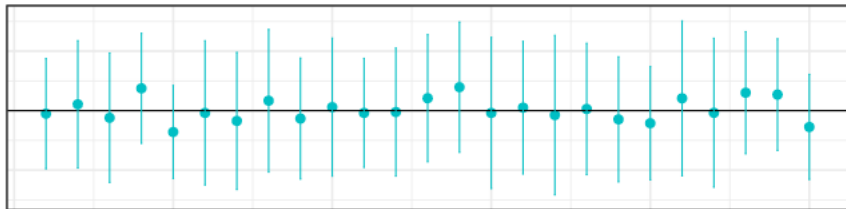
Confidence Intervals 1

Which of these is sampled from a distribution with a larger standard error?

Plot A



Plot B



PLACE YOUR BETS

Simulate 25 standard normal distribution with $C = 1.6$

Do you think:

- A. The number of times NOT covered will be ≥ 4
- B. The number fo times NOT covered will be < 4

PLACE YOUR BETS

120 seconds

Which of these would have a larger impact on the size of a confidence interval when $\hat{\sigma} = 25$ and $n = 100$

1. Changing n to 150
2. Changing $\hat{\sigma}$ to 20
3. Changing our confidence from a 99% interval to a 95% interval

90 seconds

Find the 70% confidence interval associated with a normal distribution with a mean value of $\mu = 75$ and standard error of $\sigma/\sqrt{n} = 10$

Confidence Intervals 5

Suppose that we have collected a sample with $\bar{x} = 22.5$, $\hat{\sigma} = 6.4$ and $n = 25$.

If $\mu_0 = 20$, would our null hypothesis fall within a 90% confidence interval centered around \bar{x} ?

General Topic

If we found the statistic $t = 2.14$, would our null hypothesis fall in a 95% confidence interval around \bar{x} for $n = 15$?

75 seconds

Using the `hawks` dataset, find the average wing length (`Wing`) of each species using the appropriate functions from `dplyr`

From the CLT for proportions, we have that

$$p \sim N \left(p, \sqrt{\frac{p(1-p)}{n}} \right)$$

Sampling marbles from an urn *with replacement*, we found that 24 of the marbles were red and 16 of them were blue.

Using a t-distribution, find a 90% confidence interval for the true proportion of red marbles in the urn

75 seconds

Using the bootstrap function, find an 80% confidence interval for the median admission rate in the college dataset

PLACE YOUR BETS

We will flip a fair coin 100 times

Do you think:

- A. The number of heads (1) will be between 45 and 55?
(3 to 2 odds)
- B. The number of heads will be less than 45 or greater than 55? (2 to 3 odds)

PLACE YOUR BETS