Question of the Day

What is the average mercury level of fish (Large Mouth Bass) in Florida lakes?

Mercury Levels in Fish

- The sample mean is 0.527 ppm
- In the US, the FDA action level is 1 ppm
  - Is this safely below the US limit?
- In Canada, the safety limit is 0.5 ppm
  - Is this clearly above the Canadian limit?
- We need a confidence interval!

Mercury in Fish

- The sample mean is 0.527 ppm
- In the US, the FDA action level is 1 ppm
  - Is this safely below the US limit?
- In Canada, the safety limit is 0.5 ppm
  - Is this clearly above the Canadian limit?
- We need a confidence interval!

Confidence Intervals

- Bootstrap distribution (3.3)
- 95% CI using standard error (3.3)
- Percentile method (3.4)

Statistics:
Unlocking the Power of Data

Mercury Levels in Fish

- The sample mean is 0.527 ppm
- In the US, the FDA action level is 1 ppm
  - Is this safely below the US limit?
- In Canada, the safety limit is 0.5 ppm
  - Is this clearly above the Canadian limit?
- We need a confidence interval!

Confidence Intervals

- Bootstrap distribution (3.3)
- 95% CI using standard error (3.3)
- Percentile method (3.4)
**Review**

- To create a plausible range of values for a parameter:
  - Take many random samples from the population, and compute the sample statistic for each sample.
  - Compute the standard error as the standard deviation of all these statistics.
  - Use statistic $\pm 2\times SE$.

- One small problem...

**Reality**

**Your best guess for the population is lots of copies of the sample!**

Sample repeatedly from this “population”

**Sampling with Replacement**

- It's impossible to sample repeatedly from the population...
- But we can sample repeatedly from the sample!
- To get statistics that vary, sample **with replacement** (each unit can be selected more than once).

**Bootstrap Sample**: Sample with replacement from the original sample, using the same sample size. **Remember: sample size matters!**

**M & Ms**

- How would you take a bootstrap sample from your sample of M&Ms?
### Bootstrap Sample

Your original sample has data values

18, 19, 19, 20, 21

Is the following a possible bootstrap sample?

18, 19, 20, 21, 22

a) Yes  
b) No

---

### Bootstrap Sample

Your original sample has data values

18, 19, 19, 20, 21

Is the following a possible bootstrap sample?

18, 19, 20, 21

a) Yes  
b) No

---

### Bootstrap Sample

Your original sample has data values

18, 19, 19, 20, 21

Is the following a possible bootstrap sample?

18, 18, 19, 20, 21

a) Yes  
b) No

---

### Mercury Levels in Fish

Create a bootstrap distribution with StatKey

---

### Why “bootstrap”?

“Pull yourself up by your bootstraps”

- Lift yourself in the air simply by pulling up on the laces of your boots
- Metaphor for accomplishing an “impossible” task without any outside help
**Sampling Distribution**

BUT, in practice we don't see the “tree” or all of the “seeds” – we only have ONE seed.

\[ \mu \]

**Bootstrap Distribution**

What can we do with just one seed?

Grow a NEW tree!

\[ \hat{x} \]

Estimate the distribution and variability (SE) of \( \hat{x} \)’s from the bootstraps.

**Golden Rule of Bootstrapping**

Bootstrap statistics are to the original sample statistic as the original sample statistic is to the population parameter.

**Center**

- The sampling distribution is centered around the population parameter.
- The bootstrap distribution is centered around the
  - a) population parameter
  - b) sample statistic
  - c) bootstrap statistic
  - d) bootstrap parameter

- Luckily, we don’t care about the center… we care about the **variability**!

**Standard Error**

- The variability of the bootstrap statistics is similar to the variability of the sample statistics.
- The standard error of a statistic can be estimated using the standard deviation of the bootstrap distribution!

**Confidence Intervals**

Margin of Error (ME) (95% CI: ME = 2×SE)

- Calculate statistic for each bootstrap sample
- Standard Error (SE): standard deviation of bootstrap distribution
Mercury and pH in Lakes

- For Florida lakes, what is the correlation between average mercury level (ppm) in fish taken from a lake and acidity (pH) of the lake?

$$ r = -0.575 $$

Give a 95% CI for $\rho$

$-0.575 \pm 2 \times 0.085$

$(-0.745, -0.405)$

We are 95% confident that the true correlation between mercury and pH level in Florida lakes is between -0.745 and -0.405.

Mercury and pH in Lakes

- For Florida lakes, what is the correlation between average mercury level (ppm) in fish taken from a lake and acidity (pH) of the lake?

$$ r = -0.575 $$

Give a 95% CI for $\rho$

$-0.575 \pm 2 \times 0.085$

$(-0.745, -0.405)$

We are 95% confident that the true correlation between mercury and pH level in Florida lakes is between -0.745 and -0.405.

Confidence Interval

What if we want to be more than 95% confident?

A P% confidence interval contains the true parameter value for P% of all samples.

- P% is known as the confidence level

Percentile Method

- For a P% confidence interval, keep the middle P% of bootstrap statistics:

Generate samples with replacement
Calculate sample statistic
Repeat...

Same process for every parameter!

Estimate the standard error and/or a confidence interval for...

- proportion ($\hat{p}$)
- difference in means ($\mu_1 - \mu_2$)
- difference in proportions ($\hat{p}_1 - \hat{p}_2$)
- standard deviation ($\sigma$)
- correlation ($\rho$)

We are 95% confident that average mercury level in fish in Florida lakes is between 0.433 and 0.621 ppm.

Mercury and pH in Lakes

- For Florida lakes, what is the correlation between average mercury level (ppm) in fish taken from a lake and acidity (pH) of the lake?

$$ r = -0.575 $$

Give a 95% CI for $\rho$

$-0.575 \pm 2 \times 0.085$

$(-0.745, -0.405)$

We are 95% confident that the true correlation between mercury and pH level in Florida lakes is between -0.745 and -0.405.

Confidence Interval

What if we want to be more than 95% confident?

A P% confidence interval contains the true parameter value for P% of all samples.

- P% is known as the confidence level

Percentile Method

- For a P% confidence interval, keep the middle P% of bootstrap statistics:
Level of Confidence

Which is wider, a 90% confidence interval or a 95% confidence interval?

(a) 90% CI
(b) 95% CI

Mercury and pH: Correlation

We are 90% confident that the correlation between mercury in fish and pH levels in Florida lakes is between -0.693 and -0.429.

Average Mercury in Fish

We are 99% confident that the average mercury level of fish in Florida lakes is between 0.419 and 0.647.

Statistics:
Unlocking the Power of Data

Bootstrap CI

Option 1: Estimate the standard error of the statistic by computing the standard deviation of the bootstrap distribution, and then generate a 95% confidence interval by

\[
\text{statistic} \pm 2 \times \text{SE}
\]

Option 2: Generate a P% confidence interval as the range for the middle P% of bootstrap statistics.

Bootstrapping Cautions

- These methods for creating a confidence interval only work if the bootstrap distribution is smooth and symmetric.
- ALWAYS look at a plot of the bootstrap distribution!
- If the bootstrap distribution is skewed or looks “spiky” with gaps, you will need to go beyond intro stat to create a confidence interval.
Number of Bootstrap Samples

- When using bootstrapping, you may get a slightly different confidence interval each time. This is fine!
- The more bootstrap samples you use, the more precise your answer will be.
- For the purposes of this class, 1000 bootstrap samples is fine. In real life, you probably want to take 10,000 or even 100,000 bootstrap samples.

Summary

- The standard error of a statistic is the standard deviation of the sample statistic, which can be estimated from a bootstrap distribution.
- Confidence intervals can be created using the standard error or the percentiles of a bootstrap distribution.
- Increasing the number of bootstrap samples will not change the SE or interval (except for random fluctuation).
- Confidence intervals can be created this way for any parameter, as long as the bootstrap distribution is approximately symmetric and continuous.

EXAM 1

- In-class on Friday, 10/9
- Covers everything so far (Chapters 1 – 3)
- Bring a non-cell phone calculator and a one-sided page of notes (8 ½ x 11 paper)
- Monday's class will be synthesis and review
- Optional review problems posted on WileyPlus (doing lots of problems is the best way to study!)

To Do

- Read Sections 3.3, 3.4
- Do HW 3.3, 3.4 (due Friday, 10/9)
- Study for EXAM 1 (Friday, 10/9)