

Math 107

Inference for a Single Mean
(Sections 6.4–6.6)

Central Limit Theorem

The sampling distribution of a sample mean is approximately normal and centered at μ , if n is sufficiently large.

$$\bar{x} \sim \mathcal{N}\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$

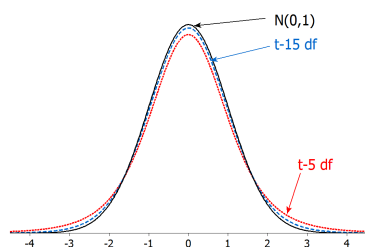
The normal approximation is usually good if $n \geq 30$.

t-distribution

t-distribution

- We don't know the population standard deviation, σ , so we would like to plug-in the sample mean, s
- This changes the distribution of the z-statistic from normal to t
- The t-distribution is very similar to the normal distribution, but has fatter tails to account for the added uncertainty

t-distribution



- The t -distribution is characterized by its **degrees of freedom (df)**
- **$df = n - 1$**
- The higher the degrees of freedom, the closer the t -distribution is to the standard normal

t-distribution

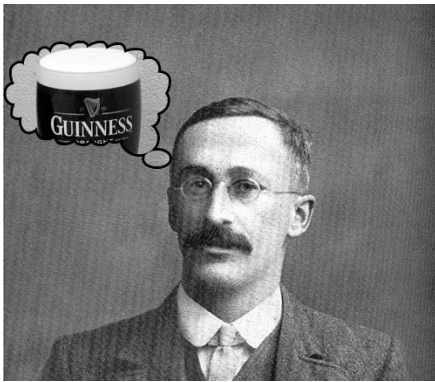
If a population with mean μ is approximately normal or if n is large ($n \geq 30$), the standardized statistic for a mean using the sample std. dev., s , follows a t -distribution with $n - 1$ degrees of freedom:

$$\frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t_{n-1}$$

Small Samples

- If sample sizes are small, only use the t -distribution if the data look reasonably symmetric and do not have any extreme outliers.
- Even then, remember that it is just an approximation!
- In practice/life, if sample sizes are small, you should just use simulation methods (bootstrapping and randomization)

Aside: William Sealy Gosset



Confidence Intervals

Confidence Interval Formula

General formula:

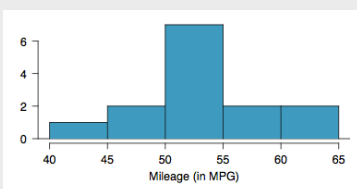
$$\text{sample statistic} \pm t^* \times \text{SE}$$

Formula for a mean:

$$\bar{x} \pm t^* \cdot \frac{s}{\sqrt{n}}$$

Prius Fuel Economy

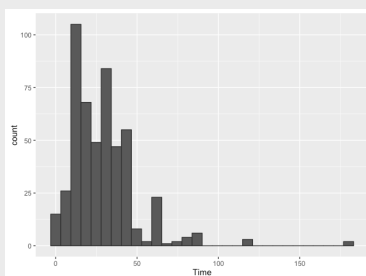
- Source: Fueleconomy.gov,
- Gas mileage (in MPG) from 14 users who drive a 2012 Toyota Prius
- Sample mean is 53.3 MPG
- Sample standard deviation is 5.2 MPG.



Find and interpret a 90% confidence interval for average gas mileage of a 2012 Prius by drivers who participate on fueleconomy.gov.

Commuting in Atlanta

- Source: American housing survey by US Census Bureau
- Commute times (min) for random sample of 500 commuters
- Sample mean is 29.11 min
- Sample standard deviation is 20.7 min

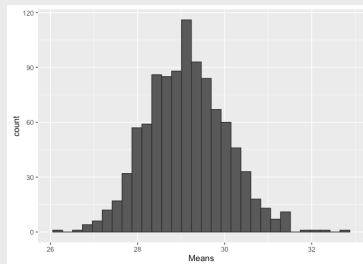


Find and interpret a 95% confidence interval for average commute time for someone living in Atlanta, GA

Commuting in Atlanta

Bootstrap results:

- mean = 29.1 min
- sd = 0.91



How does the 95% bootstrap confidence interval compare with the theory-based interval?

Using R

```
# Load the mosaic package
library(mosaic)

# Load the data
commute <- read.csv("data/CommuteAtlanta.csv")

# Constructing a 95% confidence interval
confint(t.test(~Time,
               data = commute,
               conf.level = 0.95))
```

Margin of Error

$$ME = t^* \cdot \frac{s}{\sqrt{n}}$$

You can choose your sample size in advance based on your desired margin of error!

$$n = \left(\frac{t^* s}{ME} \right)^2$$

Margin of Error

$$n = \left(\frac{t^* s}{ME} \right)^2$$

- Problem 1: For t^* , need to know n .
 - Solution: Use z^* instead of t^* (they are usually close)
- Problem 2: For s , need data.
 - Solution: estimate s .
 1. Use data from a previous study or similar population
 2. Take a small pre-sample to estimate s
 3. Estimate the range (max – min) and use $s \approx \text{range}/4$
 4. Make a reasonable guess.

Example

- Suppose we want to estimate the average GPA at Lawrence with a margin of error of 0.1 with 95% confidence.
- Determine how large a sample we would need to meet our requirements.

Hypothesis Tests

Hypothesis Test for μ

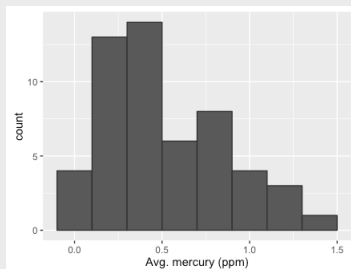
1. State the hypotheses
2. Check the conditions necessary to use the normal distribution
3. Compute the test statistic
4. Compute the p-value
5. Make a decision and state its implications in the context of the problem

Mercury Content in Fish

- Data: Average mercury levels for a sample of large mouth bass from 53 randomly selected lakes in FL
- Sample mean: 0.527
- Sample standard deviation: 0.341

Mercury Content in Fish

- Data: Average mercury levels (ppm) for a sample of large mouth bass from 53 randomly selected lakes in FL
- Sample mean: 0.527 ppm
- Sample standard deviation: 0.341 ppm



Mercury Content in Fish

- a) The Food and Drug Administration (FDA) has a limit on the mercury concentration in fish of 1.0 ppm. Is there evidence that the average mercury concentrations in FL lakes is greater than this cutoff?
- b) In Canada, the limit is 0.5 ppm. Is there evidence that the average mercury concentrations in FL lakes is greater than this cutoff?

Using R

```
# Load the data
lakes <- read.csv("data/FloridaLakes.csv")

# Running test for H_0: mu = 1 vs. H_a: mu < 1
result1 <- t.test(~ AvgMercury, data = lakes,
                  mu = 1, alternative = "less")

# Running test for H_0: mu = 0.5 vs. H_a: mu > 0.5
result2 <- t.test(~ AvgMercury, data = lakes,
                  mu = 1, alternative = "greater")
```