The Logic of Inference With Confidence Intervals

School of Information Studies
Syracuse University

Learning Topics for This Week

Contrasting deduction and induction

- Reasoning with data rule number 1
- Reasoning with statistical inference

Comparing means of two independent samples

- Introducing the mtcars dataset
- Anatomy of a boxplot

A simulation of uncertainty: repetitive sampling to simulate distribution of mean differences

Our first inferential test

What is a confidence interval?

Interpreting the results of a confidence interval

Contrasting Deduction and Induction

Contrasting Deduction and Induction

Deduction	Induction
Begin with statements/premises that are treated as facts.	Begin with multiple observations of some phenomenon or set of occurrences.
Use principles of logic to combine the facts.	Generalize from the occurrences based on patterns that may or may not be universal.
Draw a conclusion that is definitely true, as long as the premises are true.	Draw a conclusion that is only provisionally (or probabilistically) true.
Example: All mammals are warm-blooded; this dog is a mammal; therefore, this dog is warm-blooded.	Example: Observe 14 cats jumping out of trees and landing on their feet plus one clumsy cat; conclude that cats generally land on their feet when jumping.

Reasoning With Data Rule Number 1

The reasoning with data rule number 1: *You cannot **prove** anything from samples or by using statistical inference.*

Reasoning With Statistical Inference

We want to draw provisional/probabilistic conclusions about a population (the totality of some group of interest, such as all bus drivers, all democracies, or all tablet computers).

We generally cannot access the whole population for practical reasons.

We use samples of data to "stand in" for the population from which they were drawn.

Sampling error is a veil between us and the population: We can never be quite sure how well a particular sample stands in for the population (although larger samples can often be more trustworthy than smaller ones).

We construct a "model" that quantifies and displays the extent to which sampling error may interfere with our conclusions.

We draw provisional/probabilistic conclusions about a population in the context of the model.

Comparing Two Independent Sample Means

Built-In Datasets in R

The data() command reveals which built-in datasets R has available for you.

There's a dataset for almost any type of analysis.

The str() command reveals the structure of a dataset.

Mtcars: A Small Dataset of Cars From 1974

	mpg	cyl	disp	hp	drat	wt	qsec	vs	am	gear	carb
Mazda RX4	21.0	6	160.0	110	3.90	2.620	16.46	0	1	4	4
Mazda RX4 Wag	21.0	6	160.0	110	3.90	2.875	17.02	0	1	4	4
Datsun 710	22.8	4	108.0	93	3.85	2.320	18.61	1	1	4	1
Hornet 4 Drive	21.4	6	258.0	110	3.08	3.215	19.44	1	0	3	1
Hornet Sportabout	18.7	8	360.0	175	3.15	3.440	17.02	0	0	3	2
Valiant	18.1	6	225.0	105	2.76	3.460	20.22	1	0	3	1
Duster 360	14.3	8	360.0	245	3.21	3.570	15.84	0	0	3	4
Merc 240D	24.4	4	146.7	62	3.69	3.190	20.00	1	0	4	2
Merc 230	22.8	4	140.8	95	3.92	3.150	22.90	1	0	4	2
Merc 280	19.2	6	167.6	123	3.92	3.440	18.30	1	0	4	4
Merc 280C	17.8	6	167.6	123	3.92	3.440	18.90	1	0	4	4
Merc 450SE	16.4	8	275.8	180	3.07	4.070	17.40	0	0	3	3
Merc 450SL	17.3	8	275.8	180	3.07	3.730	17.60	0	0	3	3
Merc 450SLC	15.2	8	275.8	180	3.07	3.780	18.00	0	0	3	3

Showing 1 to 15 of 32 entries

The variable “am” divides the dataset into two separate groups: am=0 refers to automatic-transmission cars; am=1 refers to manual-transmission cars.

Research question: Do cars with manual transmissions get better mileage than those with automatics?

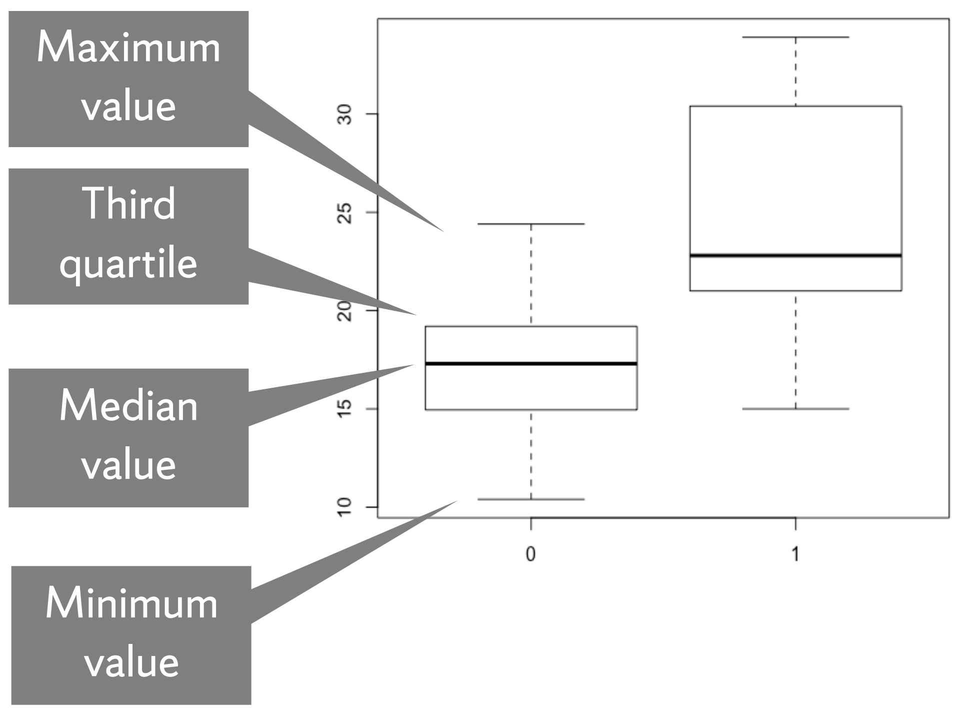
The Two Samples Differ

```
> mean( mtcars$mpg[ mtcars$am == 0 ] )     # Automatic transmissions
[1] 17.14737
> mean( mtcars$mpg[ mtcars$am == 1 ] )     # Manual transmissions
[1] 24.39231
> sd( mtcars$mpg[ mtcars$am == 0 ] )       # Automatic transmissions
[1] 3.833966
> sd( mtcars$mpg[ mtcars$am == 1 ] )       # Manual transmissions
[1] 6.166504
```

Anatomy of a Boxplot

> boxplot(mpg ~ am, data=mtcars)   # Boxplot of mpg, grouped by am

Formula notation



Simulating Uncertainty With Repetitive Sampling

A Simulation of Uncertainty: 100 Mean Differences Sampled From Our Original Data

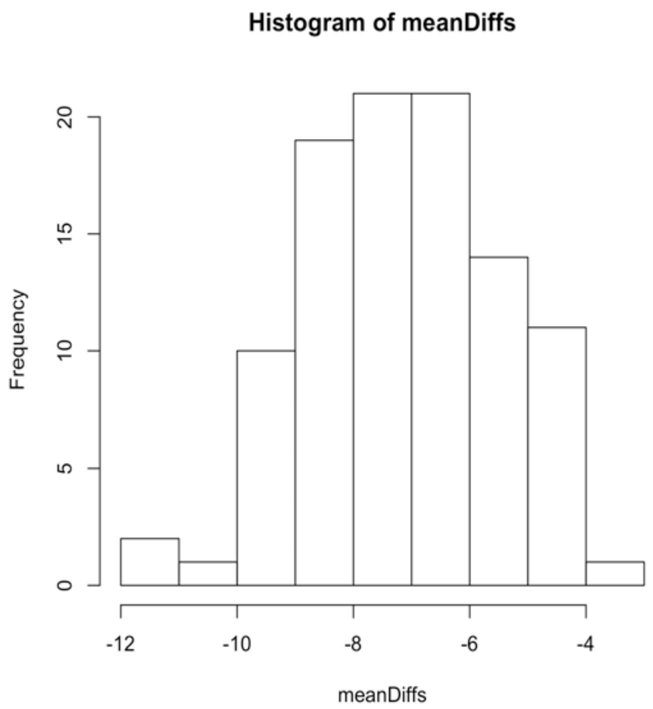
```
meanDiffs <- replicate(100,
   mean( sample(mtcars$mpg[ mtcars$am == 0 ], size=19,replace=TRUE) ) -
   mean( sample(mtcars$mpg[ mtcars$am == 1 ], size=13,replace=TRUE) ))
hist(meanDiffs)
```

These are the same sizes as the original samples

A Simulation of Uncertainty: 100 Mean Differences Sampled From Our Original Data

Key features of this histogram:

- A plot showing 100 data points, where each point is a **difference between two sample means** (automatic minus manual)
- The largest difference is -12 (in favor of manuals); the smallest difference is -3
- The center is about -7, highly similar to the observed difference between the two groups



School of Information Studies
Syracuse University

First Inferential Test

Our First Official Inferential Test

```
t.test(mtcars$mpg[mtcars$am==0] ,
                    mtcars$mpg[mtcars$am==1])
```

Using square brackets to select all of group 0 (automatic)

A t-test: invented by William Sealy Gosset of Guinness Brewery

Using square brackets to select all of group 1 (manual)

Our First Official Inferential Test

Welch Two Sample t-Test

data: mtcars$mpg[mtcars$am == 0] and mtcars$mpg[mtcars$am == 1]

t = -3.7671, df = 18.332, p-value = 0.001374

Alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-11.280194 -3.209684

Sample estimates:

mean of x mean of y

17.14737 24.39231

This is the element of the output that we are interested in right now: the 95% confidence interval. Statisticians say that, if we could replicate our whole study 100 times, on average, in 95 of those replications the calculated confidence interval would contain the actual population mean difference.

Making Sense of Confidence Intervals

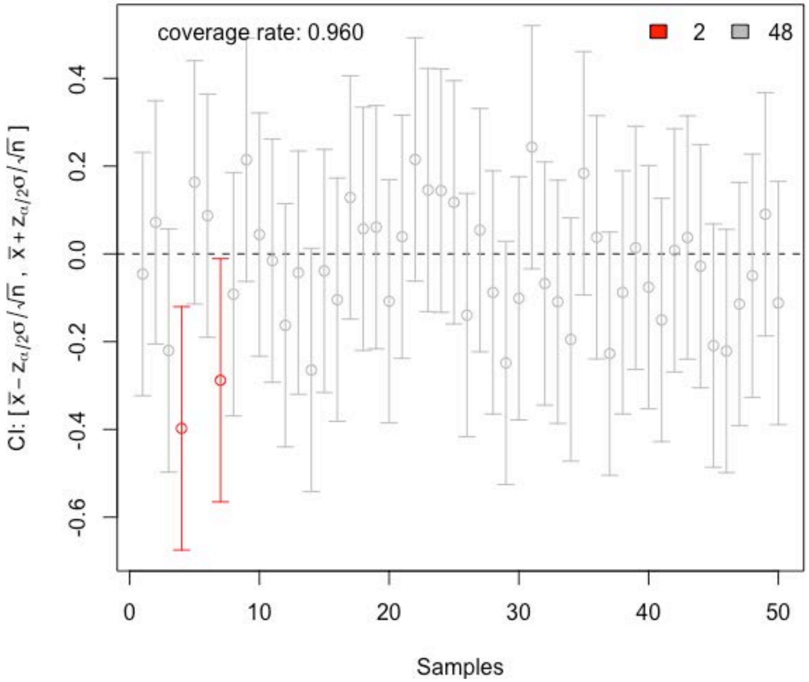
A Confidence Interval Animation

Run this code and watch the results:

```
> install.packages("animation")
> library(animation)
> conf.int(level=0.95)
```

The animation procedure repeatedly draws samples from the random normal distribution, which has a "true" population mean of zero and a standard deviation of 1. This graph shows only 50 repeats. In two of those cases, shown in red, the confidence interval does not overlap with the population mean.

Try it yourself!



School of Information Studies
Syracuse University

How the Confidence Interval Is Calculated

Confidence interval: $$Lower\ bound = (\bar{x}_1 - \bar{x}_2) - \ t^*\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

$$Upper\ bound = (\bar{x}_1 - \bar{x}_2) + \ t^*\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

The first half of each equation, a subtraction between two "x-bars," is simply the observed difference in sample means. In our mtcars example, that difference was 17.14 - 24.39 = -7.2.

The second part of the equation calculates the width of the confidence interval, in the top case subtracting it from the mean difference and in the bottom case adding it.

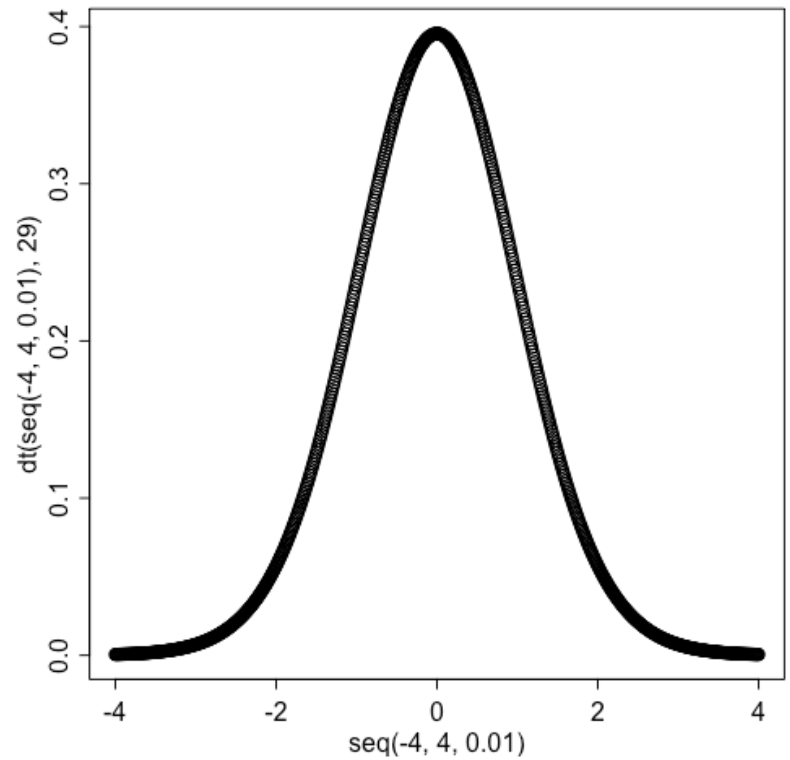
The width of the confidence interval starts with t*—this is a so-called "critical" value from the t-distribution. The critical value of t differs based on sample size and the desired confidence level.

Example t-Distribution for a Sample of 30 Observations

The t-distribution is very much like the normal distribution.

The t-distribution is a model of sample statistics, such as differences in means.

At smaller sample sizes, the tails are higher than the normal curve, representing greater uncertainty.

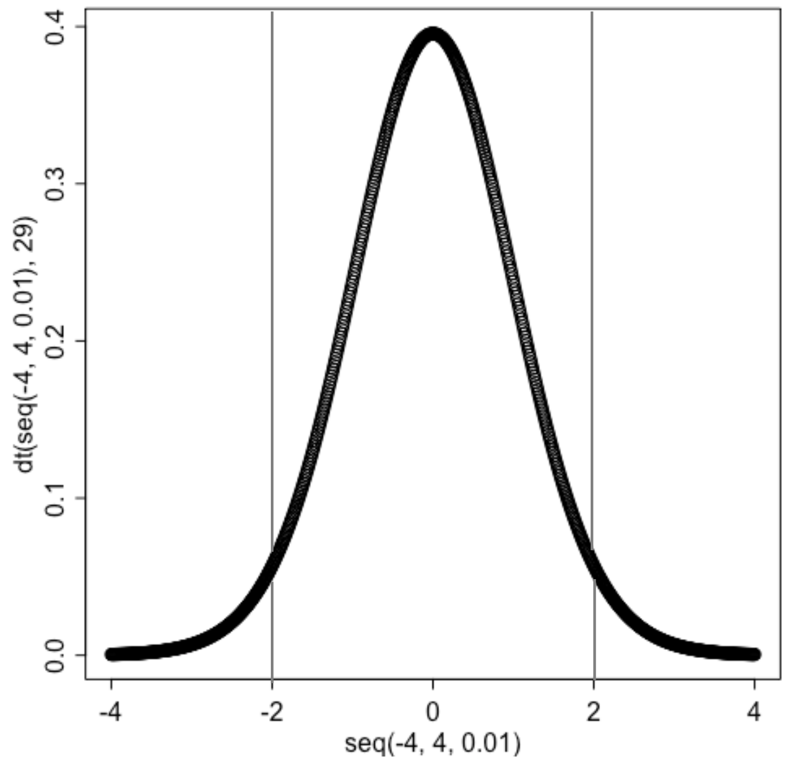


Critical Values of t Similar to 95% and 5% From Last Week

The critical values of t mark the upper and lower tails of the distribution.

95% of the area under the curve is in the central region.

5% of the area under the curve is in the tails.



Our First Official Inferential Test

Welch Two Sample t-Test

data: mtcars$mpg[mtcars$am == 0] and mtcars$mpg[mtcars$am == 1]

t = -3.7671, df = 18.332, p-value = 0.001374

Alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-11.280194 -3.209684

Sample estimates:

mean of x mean of y

17.14737 24.39231

The size of the space between the lower and upper bounds represents our uncertainty when discussing the difference in the means between the two groups. The point estimate—a mean difference of about -7.2—is at the center of that region of uncertainty.

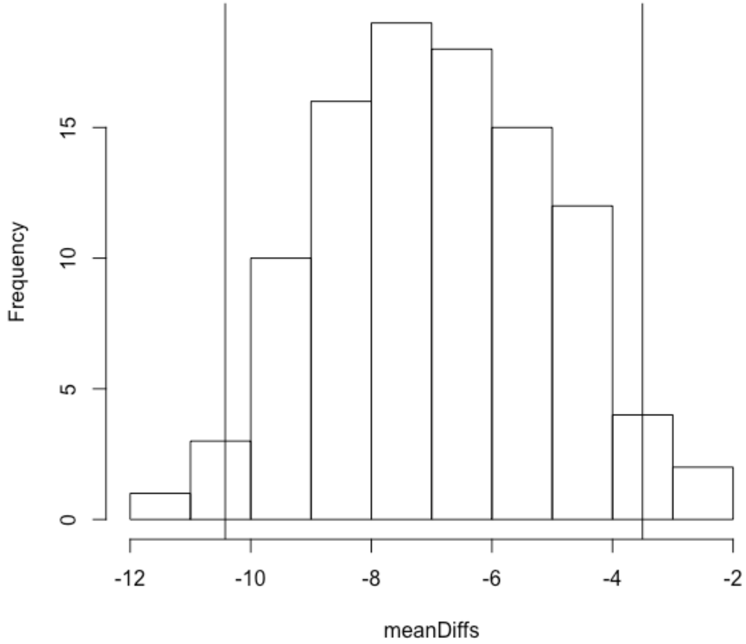
Interpreting the Results of a Confidence Interval

A Preview of Next Week's Episode

This histogram hearkens back to our "fake" simulation of the distribution of mean differences between automatic and manual transmission samples drawn from our original data.

The vertical lines mark the 0.025 quantile (close to -11) and the 0.975 quantile (close to -3); therefore, 95% of the simulated mean differences fall between those two lines.

The confidence interval is conceptually similar to those two vertical lines: an upper and lower boundary signifying the span of values where the true population difference may lie. There is no guarantee that it **will** lie anywhere in there, of course.



Remember Rule Number 1

The reasoning with data rule number 1: *You cannot **prove** anything from samples or by using statistical inference.*

The confidence interval ranging from -11.3 to -3.2 suggests a **reasonable possibility** of a difference in the fuel economy of 1974 cars when comparing automatic transmissions to manual transmissions. That is an inductive inference about the populations, based on data obtained from samples.

Running and Reporting Your Own Confidence Interval