Standard errors, confidence intervals, hypothesis testing

IBIO 851

Sept 13 2016

Suggested reading:

Chapter 6 in **Biostatistical Design & Analysis Using R** (Logan)

Goals for today

- Review standard errors, confidence intervals, hypothesis testing: central tenants of biological inference

Sampling density of the mean

- Consider a population of trees in Hiawatha National Forest
 - Mean survival age is µ = 20

Exponential density with population mean age of survival = 20 yrs
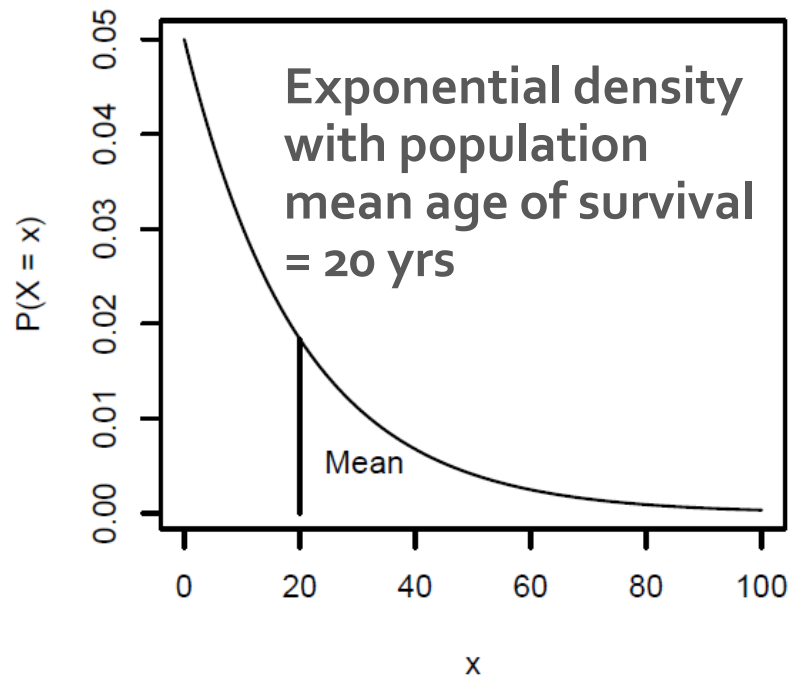
Sampling density of the mean

- Now let's age a random sample of 100 trees and calculate the sample mean $\overline{x}_1$

- Repeat the procedure 1000 times

- So we have 1000 means $\overline{x}_1 \dots \overline{x_{1000}}$
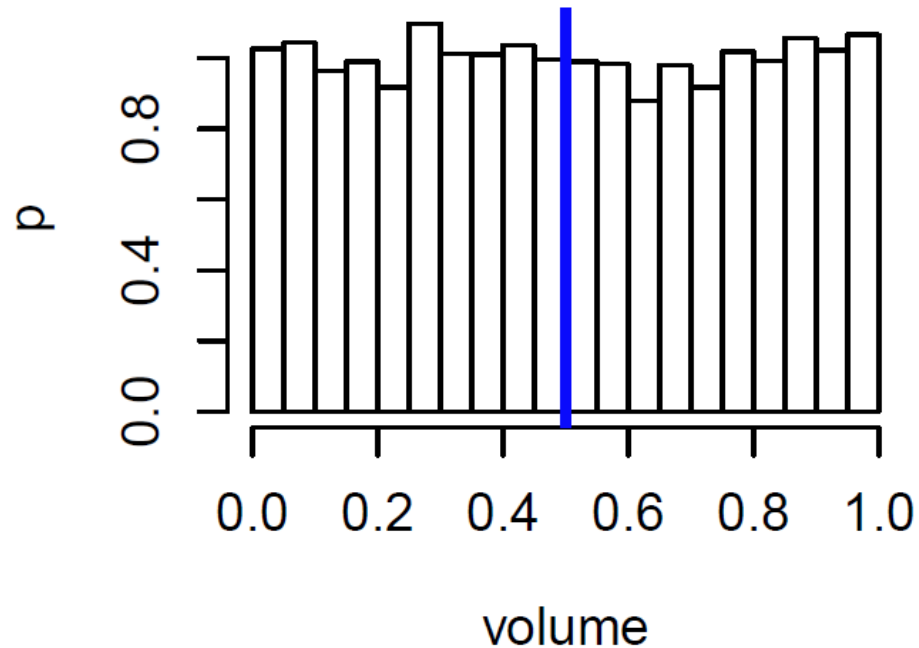
Sampling density of the mean

- True population mean is red
- Mean of sample means is black
- Thus the mean of the sampling means approximates the mean of the population
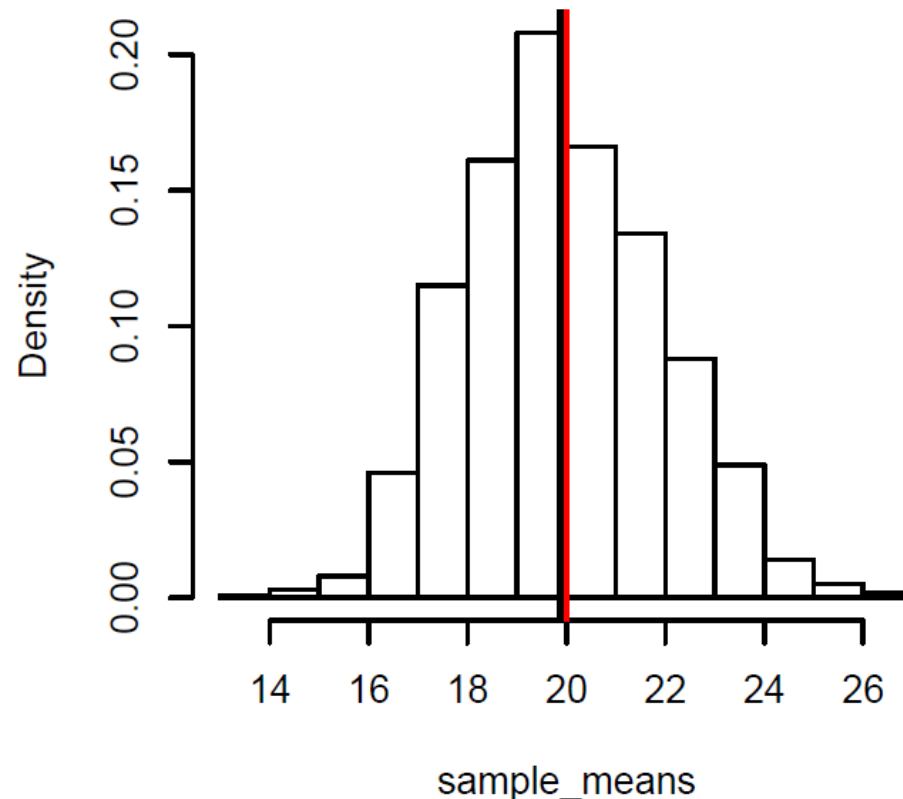
Standard error of the mean

- The mean of the samples' means has some standard deviation
- Consider a population of 10000 *E. coli* that live in your stomach
  - Volume of smallest bacterium is zero and volume of largest is 1
  - Number of individuals in each small interval of volume is about equal: uniform distribution

Standard error of the mean

Uniform density of volumes of a population of 10000 individuals (μ = 0.5 blue line)

Standard error of the mean

- Take a random sample of 10 E. coli and compute its mean volume
- Repeat the process 10000 times
- The standard deviation of the 10000 means is called the **standard error** of the sampling distribution of means

$$SE = \frac{SD}{\sqrt{n}}$$

where SD is the standard deviation of the means & n is the number of means

Standard error of the mean

- As the number of samples (n) increases, the mean of the sample means approaches the mean of the population; thus $\bar{x} \rightarrow \mu$
- The SE of the distribution of the sample means approaches zero; thus SE $\rightarrow$ o

Standard error of the mean

mean of means

0.54

0.48

0   2000   6000   10000

number of samples of size 10
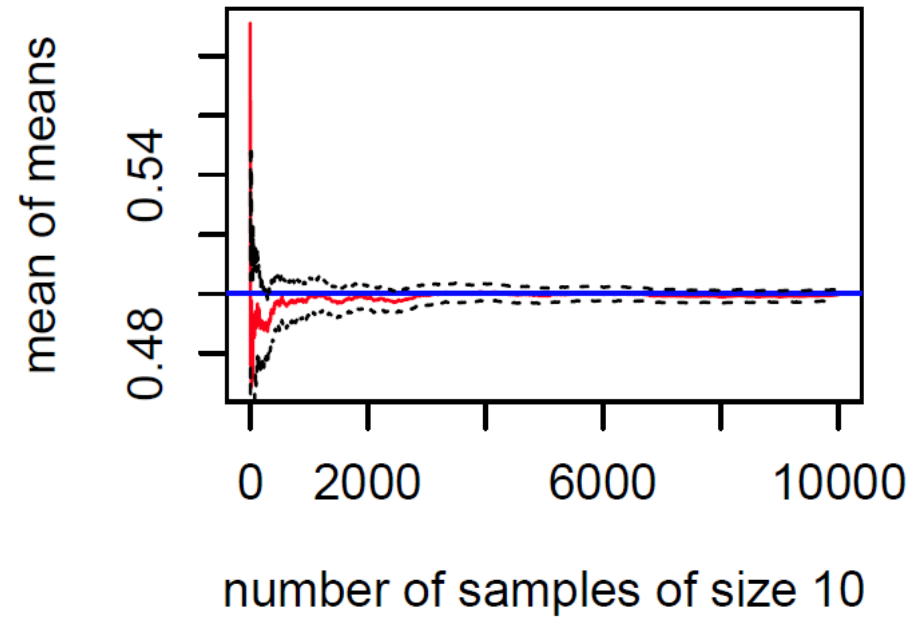
The red line is the sample mean; ± 2 x SE is in black; the population mean is in blue

Standard error by simulation

- Let's go to RStudio

Confidence intervals

- Interval estimation
  - Point estimate plus an interval that expresses the uncertainty or variability associated with the estimate

The t-distribution

- When you don't know anything about a population's behavior (i.e. just looking at sample data), you need to use the t-distribution to find the confidence interval
 - This is the case for the majority of situations since you don't know info about your 'true' population

The t-distribution
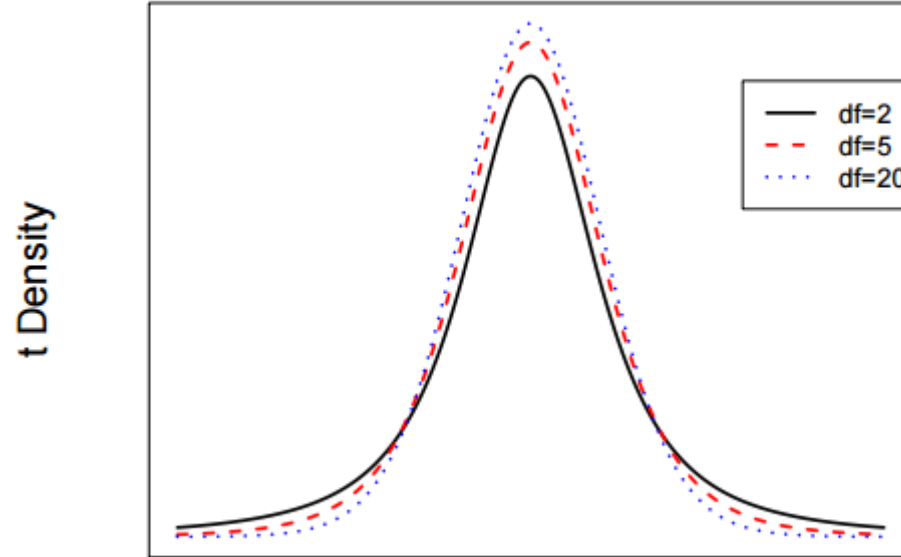
- **Properties of the t-distribution**
  - Mean=median=mode=zero
  - Symmetric about the mean
  - t ranges from –infinity to infinity
  - Family of distributions determined by n-1, the degrees of freedom
  - The t distribution approaches the normal distribution as n-1 approaches infinity

Comparing t with normal

- **Central Limit Theorem:** if the sample size is large enough, then the sample mean has an approximate normal distribution

Confidence intervals

Population → ? → Parameter: μ

Sampling

Sample → Estimation → Statistic: $\overline{x}$

Inference
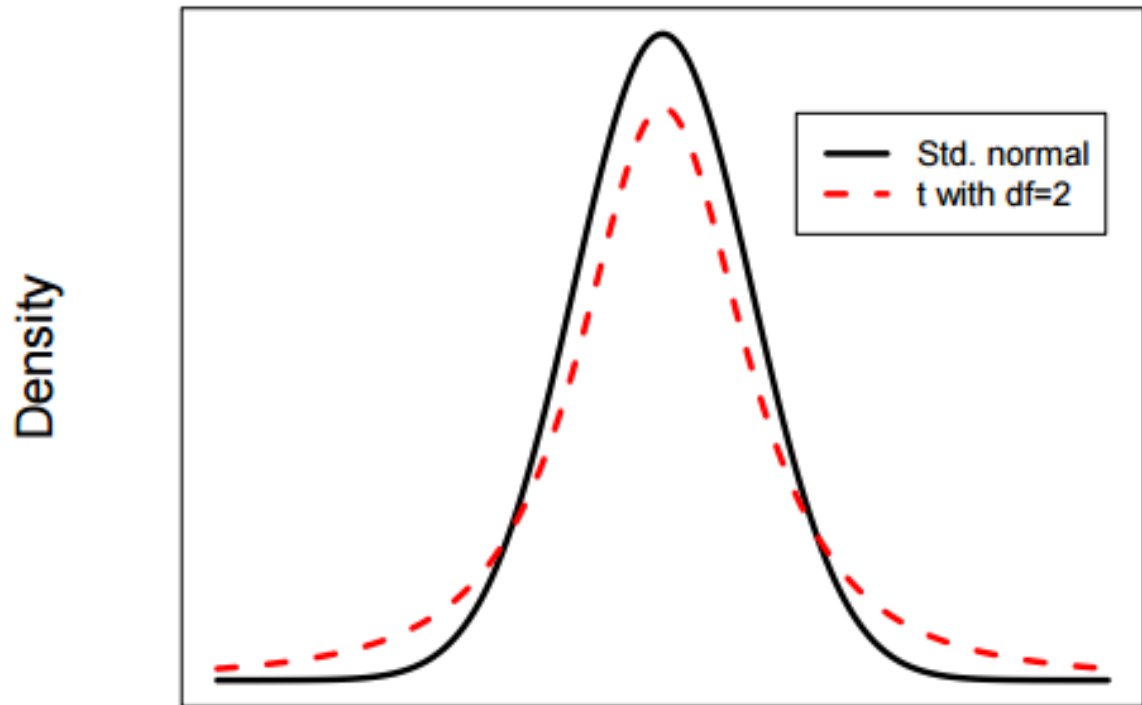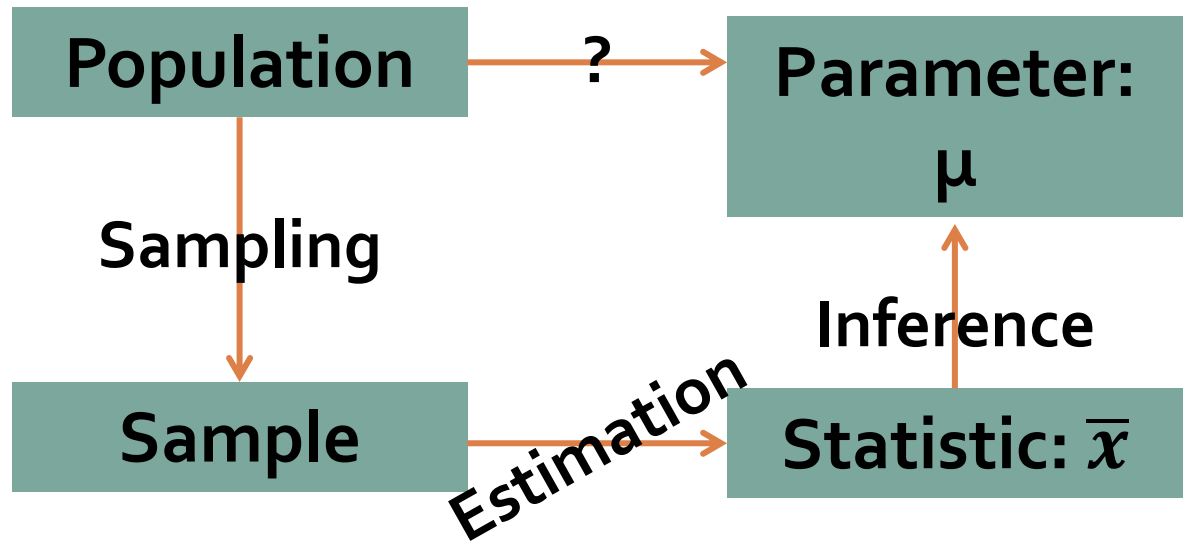
Calculating confidence intervals

- About 95% of the area under the curve is within 2 SEs of the mean
  - This applies when sample size > 20

- Formula for 95% CI:

95% Confidence interval

$$P(\overline{X} - 1.96 * S_{\overline{X}} \leq \theta \leq \overline{X} + 1.96 * S_{\overline{X}}) = 0.95$$

Where S_{xbar} = SE

Definition of confidence intervals

- There is a 95% chance that the true population mean μ occurs within the interval?

- 95% of the time upon repeated sampling, an interval will overlap with the true value of μ?

Definition of confidence intervals

Demonstrating the meaning of confidence intervals
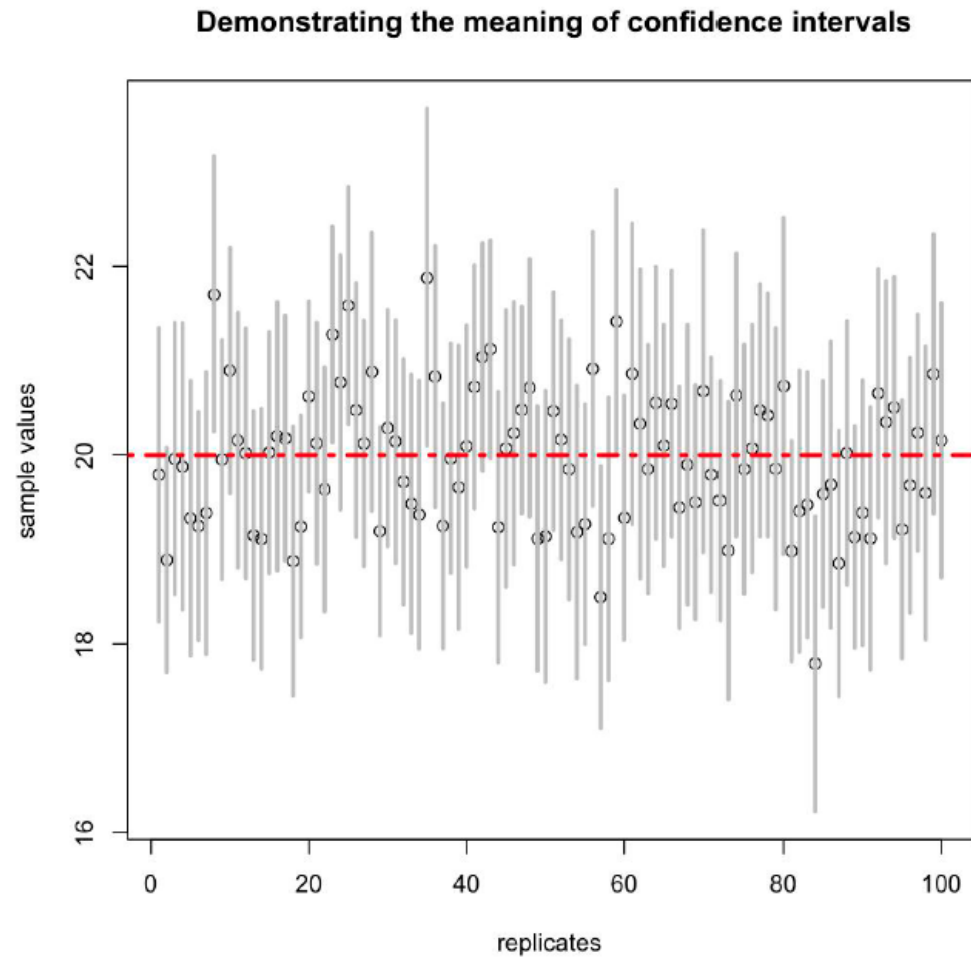
Confidence interval for a sample

- Example: A group of 10 juvenile elephants have a mean weight of 240 lbs. The sample SD = 25 lbs.
  - Note: sample size < 20 means we can't automatically use 1.96*SE
- Find a 95% confidence interval for the sample of the true mean weight of juvenile elephants in the population

Confidence interval for a sample

- STEP 1
  - Divide your SD by $\sqrt{n}$ = 25/$\sqrt{10}$ = 7.91
- STEP 2
  - Subtract 1 from your sample size to get degrees of freedom: 10-1=9
- STEP 3
  - Subtract confidence level from 1, then divide by 2: (1-0.95)/2=0.025
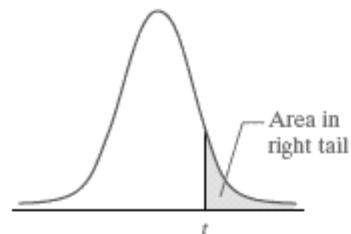- Use numbers from steps 2&3 to look up alpha level in t table

T-table

TABLE III
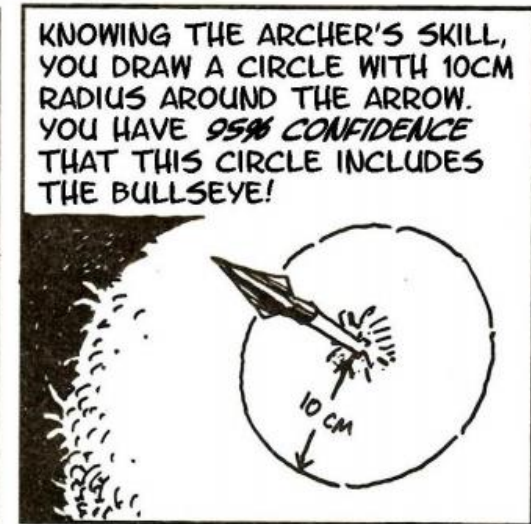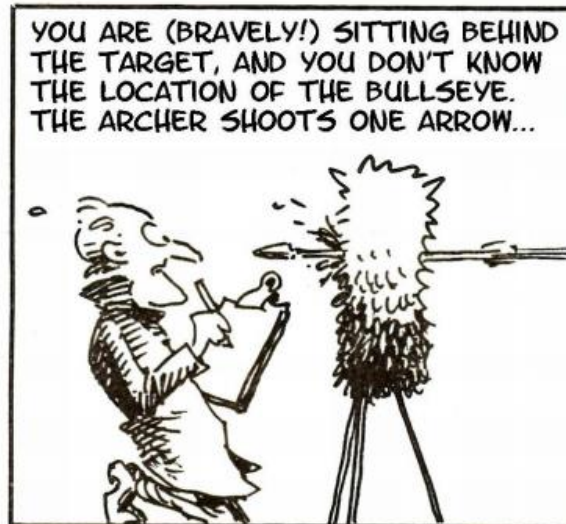
t-Distribution
Area in Right Tail

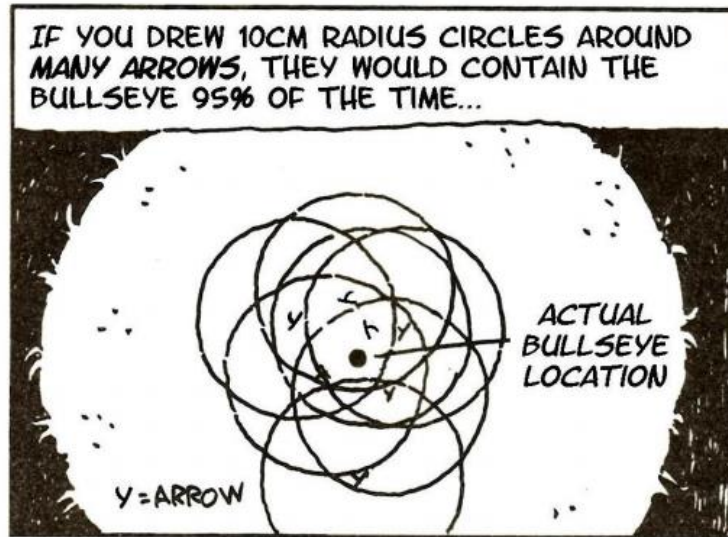| df	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005
1	1.000	1.376	1.963	3.078	6.314	12.706	15.894	31.821	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787

Confidence interval for a sample

- STEP 4
  - Multiply step 1&3: 7.91*2.262=17.88
- STEP 5
  - For the lower end of the 95% CI, subtract step 4 from sample mean: 240-17.88=222.12
  - For the upper end of the 95% CI, add step 4 to sample mean: 240+17.88=257.88
- CONCLUSION
  - 95% of the time upon repeated sampling, interval of **222.12 to 257.88** will overlap true population mean

Another way of looking at confidence intervals

CONSIDER AN ARCHER SHOOTING AT A TARGET. SUPPOSE SHE AIMS AT THE 'BULLSEYE' (A SINGLE POINT) AND HITS WITHIN 10CM OF IT 95% OF THE TIME.

YOU ARE (BRAVELY!) SITTING BEHIND THE TARGET, AND YOU DON'T KNOW THE LOCATION OF THE BULLSEYE. THE ARCHER SHOOTS ONE ARROW...

KNOWING THE ARCHER'S SKILL, YOU DRAW A CIRCLE WITH 10CM RADIUS AROUND THE ARROW. YOU HAVE *95% CONFIDENCE* THAT THIS CIRCLE INCLUDES THE BULLSEYE!
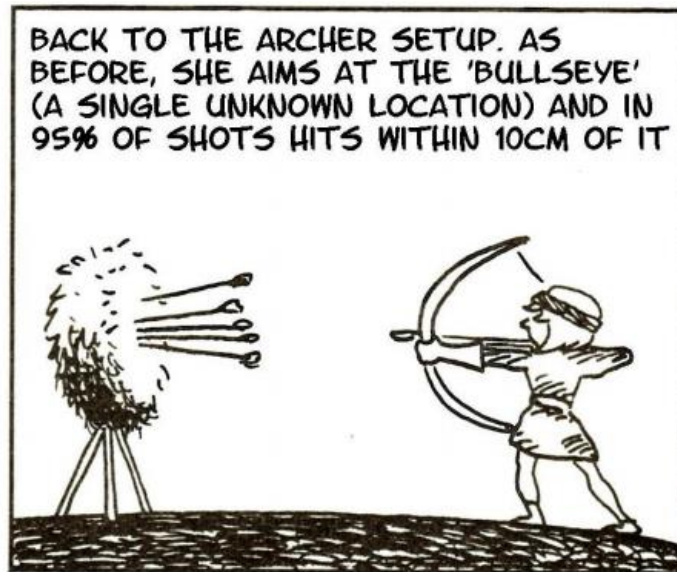
10 CM
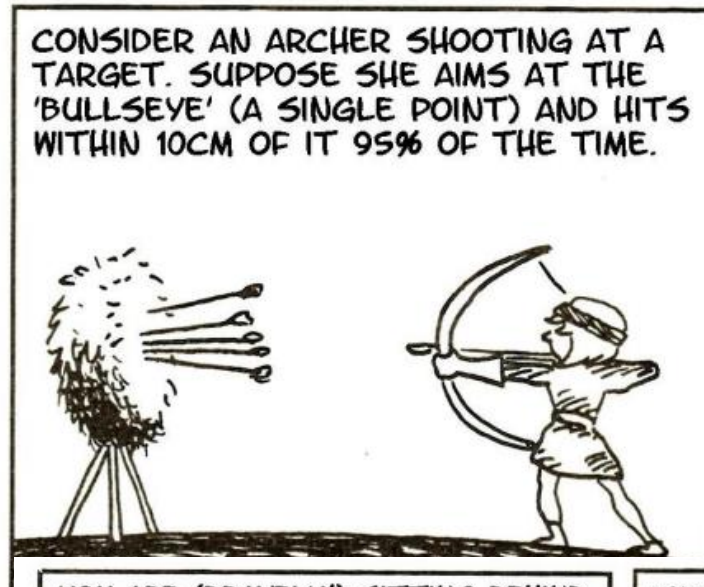
We 'trap' the truth with 95% confidence. Q. 95% of what?

Another way of looking at confidence intervals

IF YOU DREW 10CM RADIUS CIRCLES AROUND MANY ARROWS, THEY WOULD CONTAIN THE BULLSEYE 95% OF THE TIME...

ACTUAL BULLSEYE LOCATION

Y=ARROW

The interval traps the truth in 95% of experiments. To define anything frequentist, you *have to imagine* repeated experiments.

BACK TO THE ARCHER SETUP. AS BEFORE, SHE AIMS AT THE 'BULLSEYE' (A SINGLE UNKNOWN LOCATION) AND IN 95% OF SHOTS HITS WITHIN 10CM OF IT

Another way of looking at confidence intervals
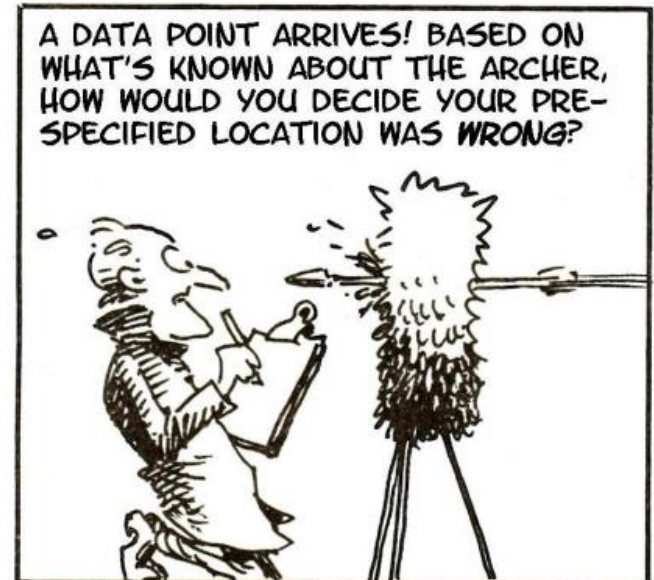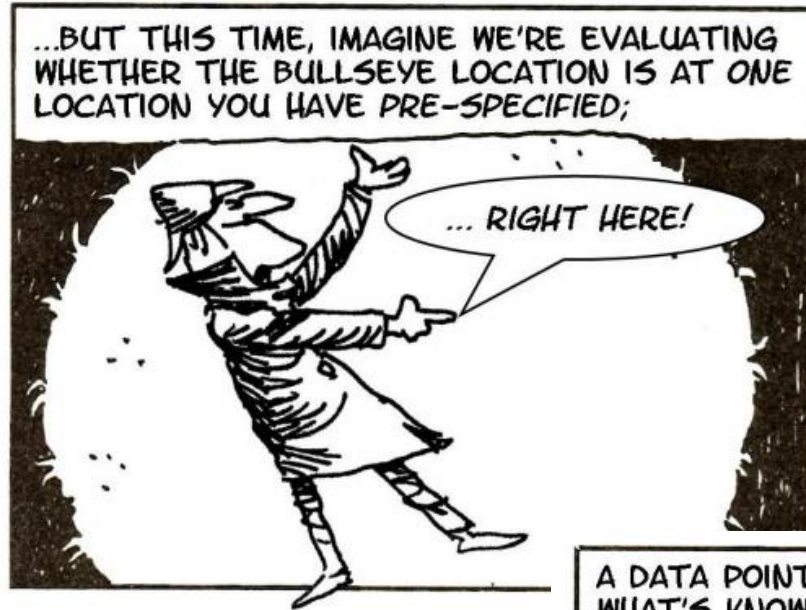
Another way of looking at confidence intervals
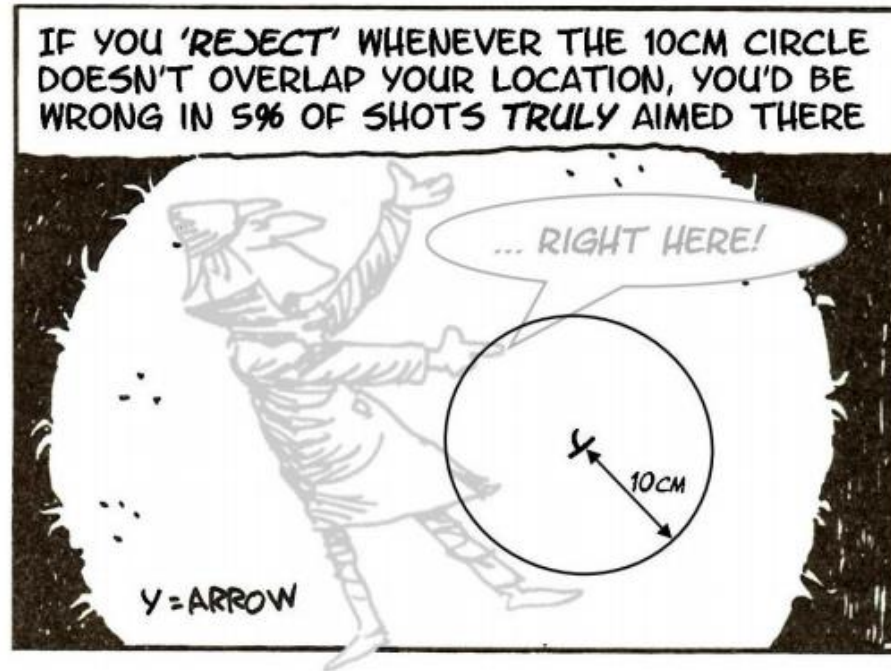
Keep in mind that this definition of a confidence interval does not say anything about whether your data is in the 95% or 5%. And it requires you to think about many other datasets, not just the one you have to analyze!

Stayed tuned…

…for the Bayesian inference interpretation!

Hypothesis testing

- A hypothesis test evaluates 2 mutually exclusive statements about a population to determine which statement is best supported by the sample data

- How do these tests really work and what does statistically significant actually mean?
  - Let's find out with a hypothetical scenario

Hypothesis testing

- An ecologist wants to know whether the monthly biofuel production of a region has changed from the previous year, when the mean production per month was 260 tons of plant matter

- The ecologist randomly samples 25 study plots and records the plant matter produced during the current year
  - The mean of the sample is 330.6 tons (SE=30.8; SD=154.2)

Hypothesis testing

- Why do we even need a hypothesis test? Our sample mean of 330.6 is different from 260 tons, right?
- But remember that this not necessarily descriptive of the entire population
  - Sampling error is the difference between a sample & the entire population
  - If we repeated the experiment, it's possible the 2nd sample would have a mean ~260
  - A hypothesis test helps assess the likelihood of that possibility

Hypothesis testing

- It's unlikely that the mean production of plant matter in the entire region (i.e. our 'population') is 330.6

- If we took multiple random samples of the same size from the same population (i.e. region), we could plot a distribution of the sample means

Hypothesis testing

- We can do this without taking many random samples
- Let's make a probability distribution plot using the t-distribution we covered, as well as the sample size & variability of our sample

Sampling Distribution for the Null Hypothesis for N=25
T, df=24
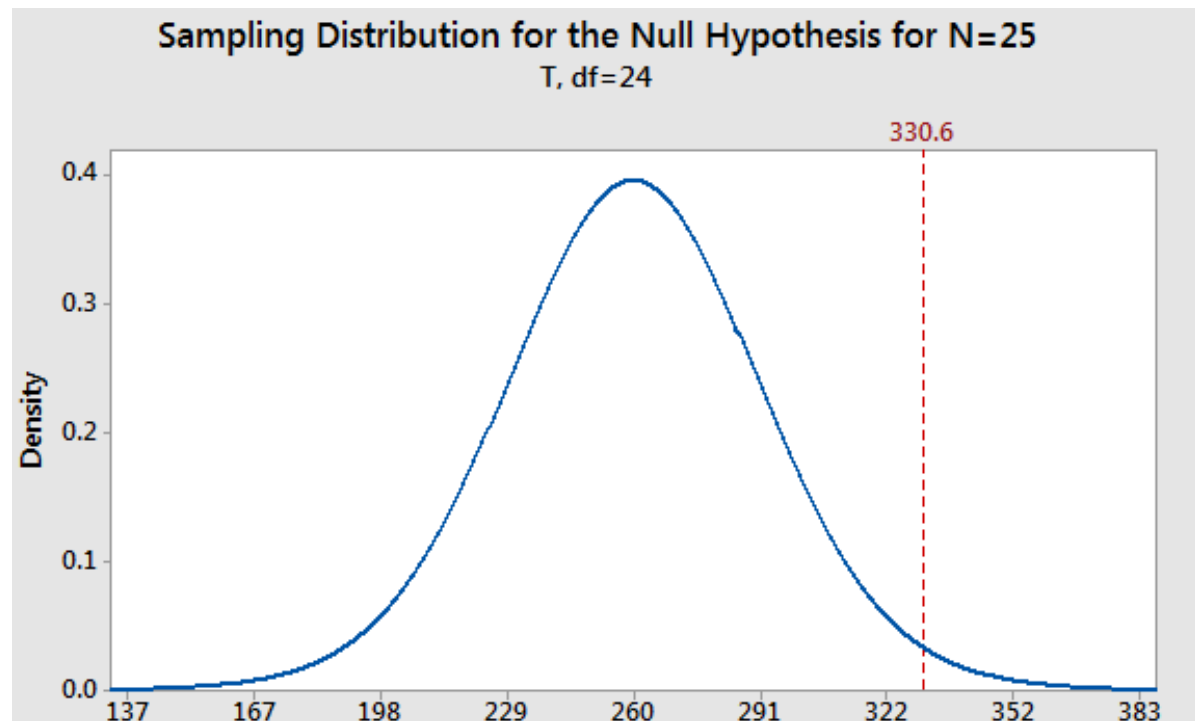
Hypothesis testing

- You can see the most probable mean is 260 (because we assume null hypothesis is true)

- But there's a reasonable probability of obtaining means ranging from 167 to 352

- While 330.6 is not probable, it's not outside the realm of possibility

- A hypothesis test allows us to quantify the probability that our sample mean is unusual

Hypothesis testing

- The decision-making tools we'll need are the significance level & p-value

- Null hypothesis: population mean = hypothesized mean (260)

- Alternative hypothesis: population mean differs from hypothesized mean

- Significance level (α) of 0.05: indicates a 5% risk of concluding that a difference exists when there's no actual difference

Hypothesis testing

Two-Tailed Critical Region for a Significance Level of 0.05
T, df=24
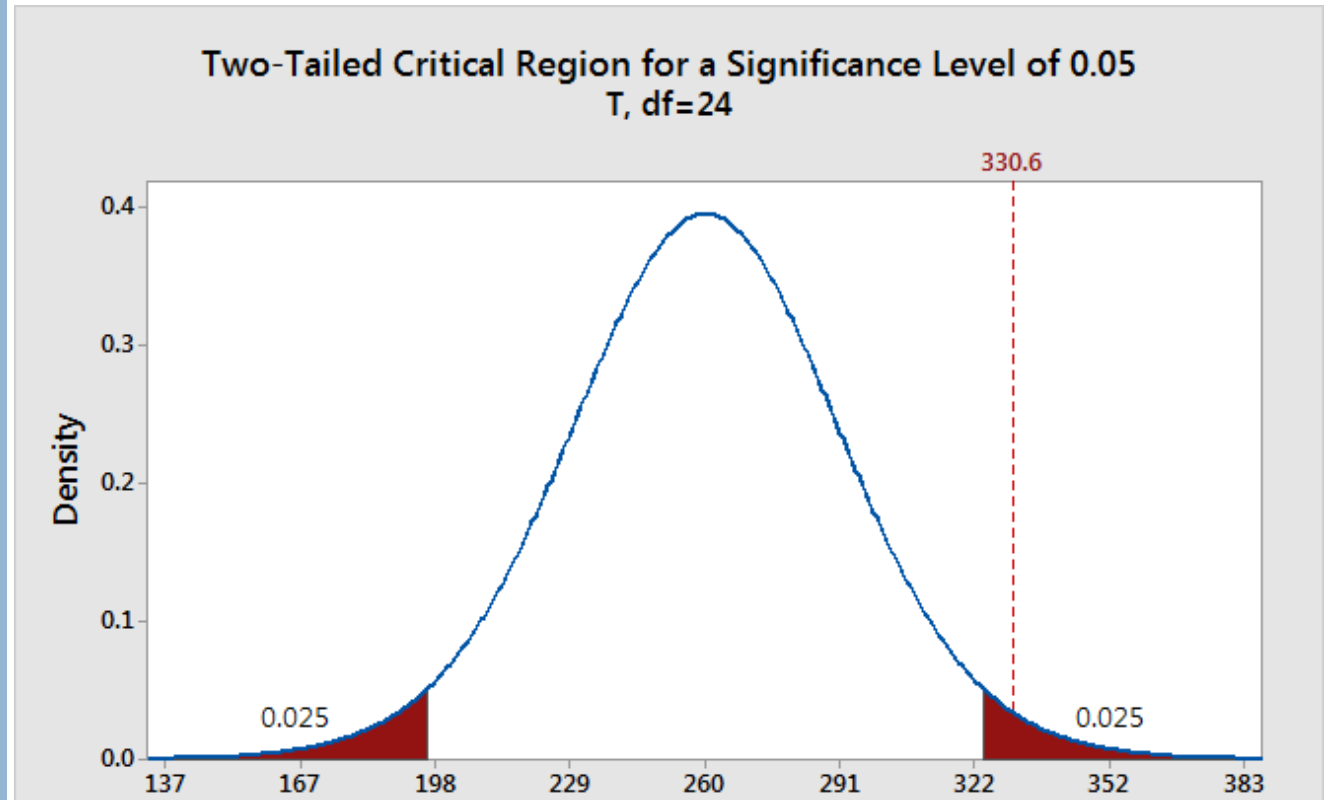
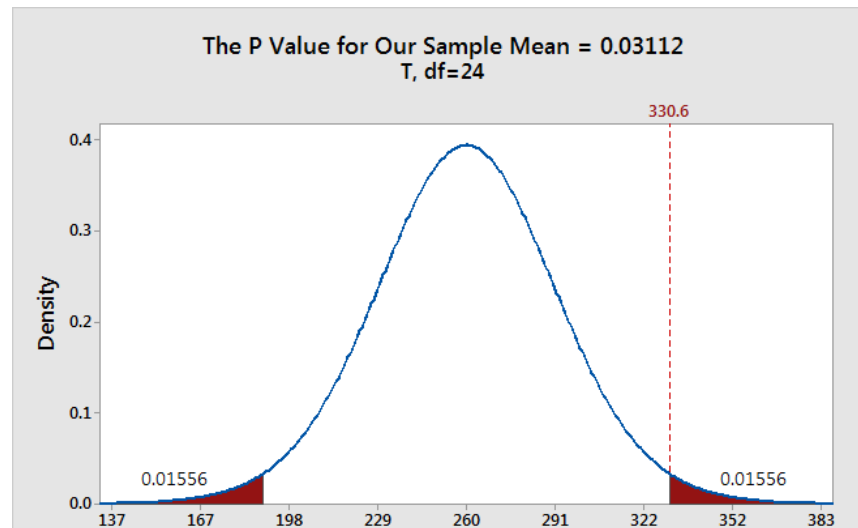## Hypothesis testing

- To get our p-value for this hypothesis test, we need to first determine the distance between the sample mean & null hypothesis value
  - 330.6-260=70.6
- Next, graph probability of obtaining a sample mean that is at least as extreme in both tails of the distribution
  - 260 +/- 70.6

Hypothesis testing

The P Value for Our Sample Mean = 0.03112
T, df=24

- 2 shaded areas have a probability of 0.01556 (total=0.03112)

- This probability represents likelihood of obtaining a sample mean that is at least as extreme as our sample mean in both tails of the distribution if the population mean is 260. That's our p-value.

- We can conclude that the average biofuel production per month for the population/region is greater than 260

Hypothesis testing: steps

- Formulate the null hypothesis
- Choose an appropriate statistic
- Decide on a significance level (typically 0.05)
- Compute a p-value for your test and determine if it is significant
- Reject or fail to reject null hypothesis

One sample t-test

- Tests if the mean of a population, μ, is equal to a particular value m
  - Like our biofuel plant production example
- Null hypothesis: Ho: μ = m
- Alternative hypothesis: Ha: μ ≠ m
- Test statistic t

One sample t-test

- A researcher is interested in determining whether nutrient supplementation affects plant height

- The independent variable, nutrient supplementation, is administered to a sample of plants (n=9) in an attempt to determine its affect on the dependent variable, height

- Based on previous data, the population mean μ for height is 24 cm

- Sample mean $\bar{x}$ is 26 cm with SD=4

One sample t-test

State the hypotheses

- NULL hypothesis:
  - No difference in means
  - $\bar{x}$ - μ = zero

- ALTERNATIVE hypothesis:
  - Difference in means
  - $\bar{x}$ - μ ≠ zero

One sample t-test

Set criteria

- Need degrees of freedom, which is n-1 (where n is the sample size)
  - In this case, df = 8

- We want a confidence level (α) of 0.05 (or a 95% confidence)

One sample t-test

Calculate test statistic

- t = $\dfrac{\bar{x} - \mu}{S\bar{x}}$

where xbar is the mean we estimated; μ is the population mean we are comparing to; S_{xbar} is the SE of the mean

- Calculate the SE: SD/$\sqrt{n}$
  - 4/$\sqrt{9}$ = 1.33
- t = 26 − 24/ 1.33 = 1.503
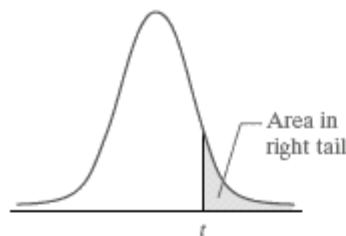
One sample t-test

Look up value in t table

Area in right tail

t

TABLE III

t-Distribution
Area in Right Tail

df	0.25	0.20	0.15	0.10	0.05	0.025	0.02	0.01	0.005
1	1.000	1.376	1.963	3.078	6.314	12.706	15.894	31.821	63.657
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499
8	0.706	0.889	1.108	1.397	1.860	2.306	2.449	2.896	3.355
9	0.703	0.883	1.100	1.383	1.833	2.262	2.398	2.821	3.250
10	0.700	0.879	1.093	1.372	1.812	2.228	2.359	2.764	3.169
11	0.697	0.876	1.088	1.363	1.796	2.201	2.328	2.718	3.106
12	0.695	0.873	1.083	1.356	1.782	2.179	2.303	2.681	3.055
13	0.694	0.870	1.079	1.350	1.771	2.160	2.282	2.650	3.012
14	0.692	0.868	1.076	1.345	1.761	2.145	2.264	2.624	2.977
15	0.691	0.866	1.074	1.341	1.753	2.131	2.249	2.602	2.947
16	0.690	0.865	1.071	1.337	1.746	2.120	2.235	2.583	2.921
17	0.689	0.863	1.069	1.333	1.740	2.110	2.224	2.567	2.898
18	0.688	0.862	1.067	1.330	1.734	2.101	2.214	2.552	2.878
19	0.688	0.861	1.066	1.328	1.729	2.093	2.205	2.539	2.861
20	0.687	0.860	1.064	1.325	1.725	2.086	2.197	2.528	2.845
21	0.686	0.859	1.063	1.323	1.721	2.080	2.189	2.518	2.831
22	0.686	0.858	1.061	1.321	1.717	2.074	2.183	2.508	2.819
23	0.685	0.858	1.060	1.319	1.714	2.069	2.177	2.500	2.807
24	0.685	0.857	1.059	1.318	1.711	2.064	2.172	2.492	2.797
25	0.684	0.856	1.058	1.316	1.708	2.060	2.167	2.485	2.787

Model for t-test

- Y(i) = α + β*x(i) + error(i)

- Error(i) ~ normal(0, σ^2)
  - Note that the error structure may differ between groups (this would be a t-test with unequal variance)

Hypothesis testing

- Let's do some t-tests in R