Live Classroom
Module 2

Module 2 – Statistical Inference and Tests for Means

Module 2

This module extends the introduction to statistical inference started in the previous module.

What do we mean by statistical inference?

- Seek to draw conclusions about a population parameter (such as the population means) using information from a sample.
- Draw conclusions based on probabilities.
- Confidence intervals and tests of significance are two common types of statistical inference.

Confidence Intervals

Suppose a grocery store would like to estimate the average number of customers who shop at the store on Mondays. In other words, they want to estimate the population mean, μ.

- First, what might they do to obtain an estimate?
- Next, would it make sense for them to try and determine the accuracy of their estimate?

Confidence Intervals

Suppose a grocery store would like to estimate the average number of customers who shop at the store on Mondays. In other words, they want to estimate the population mean, μ.

- Let's say the grocery store takes a sample of 100 Mondays (i.e., for 100 Mondays, they count the number of customers that shop). Then, they take the average (i.e., the sample mean).
- Let's next assume the population standard deviation is 20 people.

- From the CLT, we know that the distribution of the sample mean should be centered on the true population mean. So, our sample mean should be close to the true mean, but it may not be exact.
  - Can we quantify how close the sample mean might be to the true population mean?

Confidence Intervals

Since $n \geq 30$, we can use the CLT.

- The sample mean will be normally distributed with a mean of μ and a standard deviation of $\frac{20}{\sqrt{100}} = 2$.

- Remember from last week, we learned that when data are normally distribution, 95% of the data falls within 2 standard deviations of the mean.

  - This implies that 95% of the sample means will fall within 2 standard deviations of the population mean.

  - This also means that 95% of the time, the population mean will be within approximately two standard deviations of the sample mean.

  - In this example, the sample mean and the population mean are within 4 people of each other 95% of the time.

Confidence Intervals

Confidence Intervals have the form:

Estimate ± margin of error

- The estimate is our guess for the value of the unknown parameter.

- The margin of error gives information about how far from the estimate we think the true value might be. This quantity is a reflection of the amount of variability in the sampling distribution of the estimate.

To calculate a confidence interval for a confidence level C for the population mean, we use the following formula:

$$\bar{X} \pm z \frac{\sigma}{\sqrt{n}}$$

Confidence Intervals

Values of the critical value, z, correspond with the points (− z and + z) on the standard normal curve that give a central area of size C under the normal curve.

Confidence level, C	90%	95%	99%
Critical Value, Z	1.645	1.960	2.576

Confidence Intervals

Back to our previous example…

Suppose a grocery store would like to estimate the average number of customers who shop at the store on Mondays. In other words, they want to estimate the population mean, μ.

- Let's say the grocery store takes a sample of 100 Mondays (i.e., for 100 Mondays, they count the number of customers that shop). Then, they take the average (i.e., the sample mean). Let's assume they get a sample average of 1000.

- Let's next assume the population standard deviation is 20 people.

To calculate a 99% confidence interval, we would do:

$$\bar{X} \pm z * \frac{\sigma}{\sqrt{n}} = 1000 \pm 2.576 \frac{20}{\sqrt{100}} = 1000 \pm 2.576 * 2 = (994.8, 1005.2)$$

We are 99% confident that the true mean number of customers on Mondays is between 995 and 1006 people.

Tests of Significance

Tests of significance are used to assess the evidence provided by the data from a sample about some claim concerning the population.

Example:

Suppose I claim that I can do a cartwheel on the balance beam, 95% of the time. Let's say that you are impressed, so you ask me to show you. I do 10 in a row, and I only complete 3 cartwheels successfully.

Would you be skeptical of my claim that I am successful 95% of the time?

Tests of Significance

Tests of significance are used to assess the evidence provided by the data from a sample about some claim concerning the population.

Example:

Suppose I claim that I can do a cartwheel on the balance beam, 95% of the time. Let's say that you are impressed, so you ask me to show you. I do 10 in a row, and I only complete 3 cartwheels successfully.

Would you be skeptical of my claim that I am successful 95% of the time?

If I were telling the truth and really do complete successful cartwheels on the beam 95% of the time, then the probability of me only completing 3 out of 10 would be very small (it would be rare for this to happen!).

Tests of Significance

What would be the probability of only completing 3 out of 10 with a 95% success rate for each one?

We can use the binomial distribution to get our answer…

$$C(10,3).95^3\,.05^7 = 120 *.95^3\,.05^7 = 0.00000008038$$

Therefore, there is only a probability of 0.00000008038 of this happening by chance if my claim is actually true. This gives strong evidence against my claim. In fact, you should reject my claim!

Tests of Significance

This example demonstrates how statistical testing generally works.

A claim is refuted if the sample gives an outcome that would be exceedingly rare if the claim was actually true. When performing statistical testing, we start out with a particular claim about a population parameter.

We then gather evidence from a sample. If the evidence gathered from the sample would be exceedingly rare if the claim was true, we use this as evidence that the claim is not true.

We use probabilities to quantify how rare the outcome would be if the claim were true and then use this to make decisions about whether or not to formally reject the claim.

A Dietician's Experiment

A dietician is interested in seeing whether or not a new diet plan she has come up with is successful in helping people lose weight after just 6 weeks. It isn't possible to know the "true" population mean because she would have to gather every person in the world, put them on the diet plan, and then determine if they lose weight after 6 weeks. Instead, she takes a sample.

-She samples 50 people.
-She finds the sample mean weight loss to be 3.16 pounds.
-The population standard deviation is 9.53.

Key Question: Is the diet plan a success?

A Dietician's Experiment

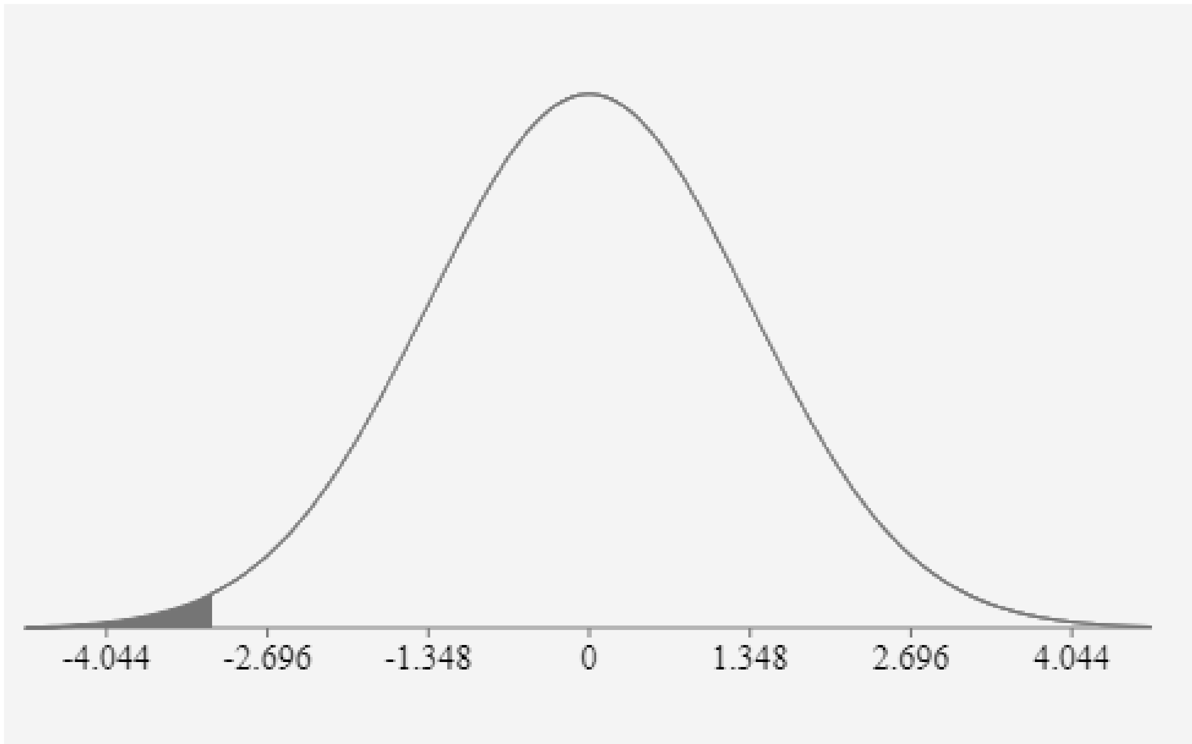
Ultimately, we want to make a claim and then test whether the claim is true or not. To test whether the claim is true, we need to decide if the data we have give evidence against it.

If our goal is to see if participants on the program lost weight, then we claim that participants in general did not lose weight and then see if there is evidence to the contrary from the sample data.

$$H_0\!: \ \mu = 0$$

$$H_1\!: \ \mu < 0$$

IF the null hypothesis is true, then the distribution of the sample mean for a random sample of size 50 will be normally distributed with a mean of 0 (the same as the population mean under the null hypothesis) and a standard deviation of $\frac{\sigma}{\sqrt{n}} = \frac{9.53}{\sqrt{50}}$ = 1.348.



Given this value is pretty far out on the normal curve, it would be quite rare to observe this sample mean if the true population mean were actually equal to 0. Therefore, it is good evidence that the alternative hypothesis is true.

Let's be a bit more formal…

We want to standardize/generalize the process we went through on the previous slide….

So, instead of working with the sample mean, we work with what is called a "test statistic". The test statistic for this specific test (a test of the mean), is called the z-statistic.

$$\mathbf{z} = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$$

Test statistic for tests relating to the mean measure how far $\bar{x}$ is from the value of μ (under the null hypothesis) in standard deviation units. Z is normally distributed with a mean of 0 and a standard deviation of 1 (called the standard normal distribution).

5 Step Recipe

1) State the hypotheses (we did this already).

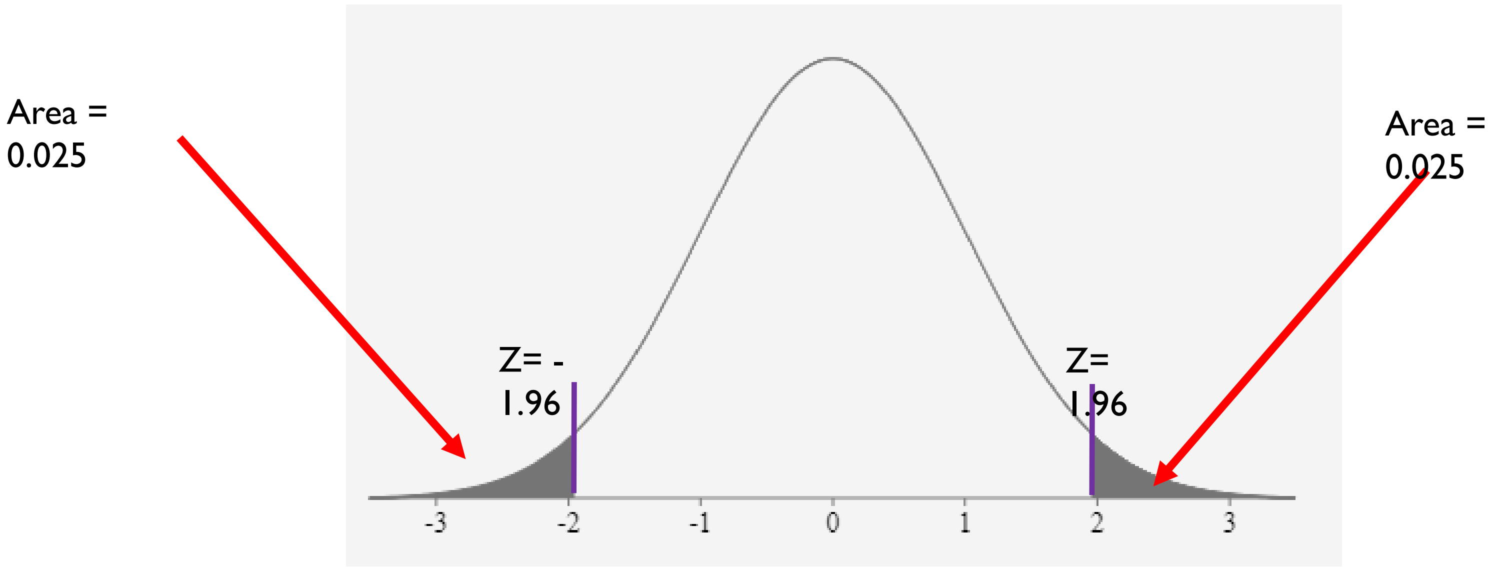
2) Select the appropriate test-statistic. In this case, it's the z- statistic!

3) Now we need a decision rule for when we are going to reject the null hypothesis…

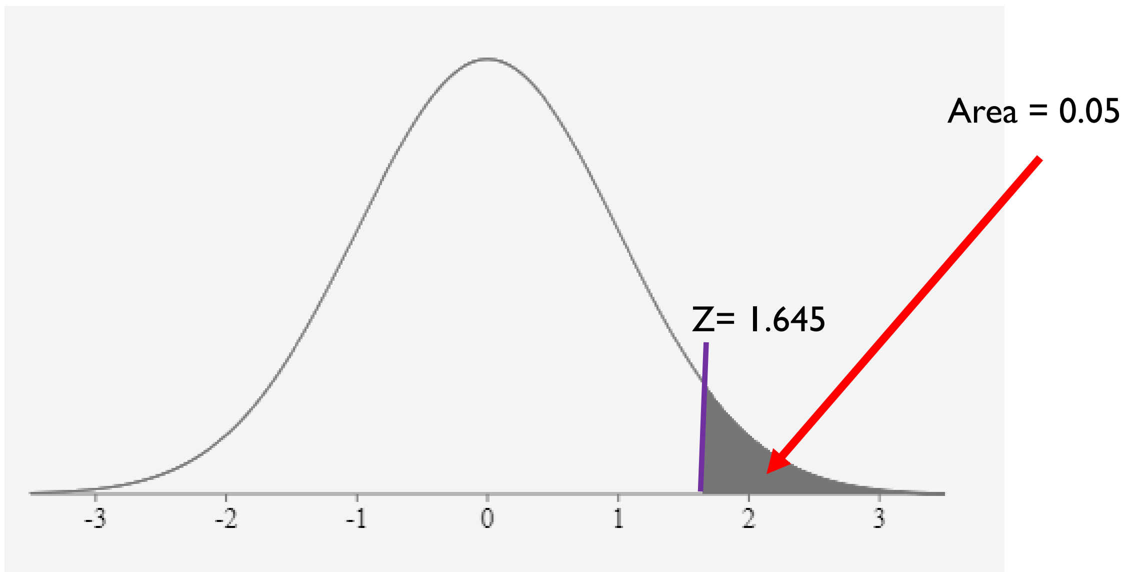
A decision rule is based on a significance level. The significance level is the probability of rejecting the null hypothesis given that it is actually true.

Usually the significance level is set at $\alpha = 0.05$.

Let's look at where this is on a standard normal curve…



For a two-sided test, this is what would represent an $\alpha = 0.05$.



For a one-sided test where the alternative hypothesis is that the μ >= some value, the area would be as described above. When the alternative hypothesis is in the other direction (like in our example), the shading would be on the left side of the curve.

5 Step Recipe

1) State the hypotheses (we did this already).

2) Select the appropriate test-statistic. In this case, it's the z- statistic!

3) Now we need a decision rule for when we are going to reject the null hypothesis

Critical Value Approach

Our decision rule for a significance level of 0.05 is that we will reject the null hypothesis if the z statistic is less than or equal to -1.645.

5 Step Recipe

1) State the hypotheses (we did this already).

2) Select the appropriate test-statistic. In this case, it's the z- statistic!

3) Now we need a decision rule for when we are going to reject the null hypothesis

P-value Approach

Our decision rule for a significance level of 0.05 is that we will reject the null hypothesis if the p-value we calculate is less than 0.05.

What is a p-value?

A p-value is the probability, assuming the null hypothesis is true, of observing the test statistic that you observed or one that is more extreme.

The smaller the p-value is, the more evidence the data shows against the null hypothesis. This is because a small p-value suggests that the observed result was unlikely to have occurred if the null hypothesis is in fact true.

Large p-values, on the other hand, do not give evidence against the null hypothesis.

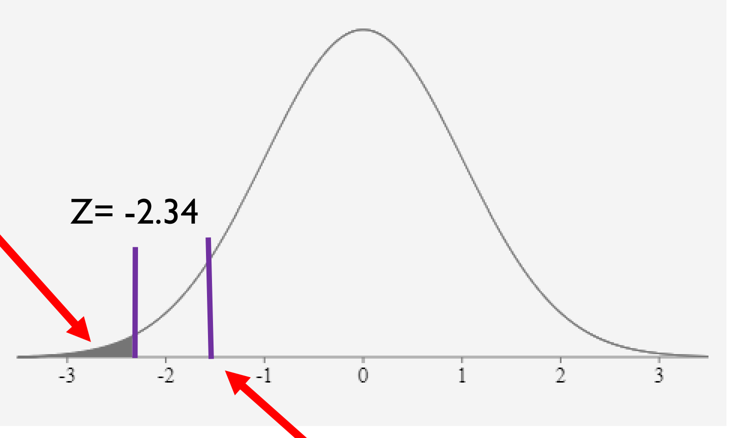
Back to our example…

Our z statistic calculation is as follows:

$$z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{-3.16 - 0}{9.53/\sqrt{50}} = -2.34$$

This area represents the p-value!

P=0.0096



Z= -1.645

The critical value for alpha=.05 for a one-sided test.

5 Step Recipe

1) State the hypotheses (we did this already).
2) Select the appropriate test-statistic. In this case, it's the z- statistic!
3) State decision rule for when we are going to reject the null hypothesis.
4) Calculate test statistic and p-value.
5) Give conclusion.

<u>Conclusion:</u> We reject the null hypothesis since the p-value was less than .05. The z statistic of -2.34 is also less than the critical value of -1.645, which again, leads us to reject the null hypothesis.

Therefore, we can conclude that the on average, those on the diet plan lose weight.

Important!

This was just an example of one type of test.

Module 2 also covers one-sample test of means using a t-test.

- When you don't know the population standard deviation AND when your sample size is less than 30.

- T distribution is used instead of Normal distribution. Remember to specify degrees of freedom!

Module 2 also covers two-sample test of means.

R Coding Topics

- Calculating quantities from the normal (z) and t distributions
- One sample testing and confidence intervals
  - z test (population standard deviation known or unknown)
  - t test
- Two sample t tests and confidence intervals
- Numerical and Graphical summaries for two sample tests – Review on your own!

Quantities from the t-distribution

- Calculating probabilities from z and t statistics
  - Use the pnorm(*[z statistic]*) function to calculate the area to the left of a given *z* statistic
  - Use the pt(*[t statistic]*, df = *[degrees of freedom]*) function to calculate the area to the left of a given *t*-statistic

- Calculating z and t statistics from probabilities
  - Use the qnorm(*[probability]*) function to find the z-statistic with the specified the area to the left
  - Use the qt(*[probability]*, df = *[degrees of freedom]*) function to find the t-statistic with the specified the area to the left

Table B. t-Distribution Critical Values

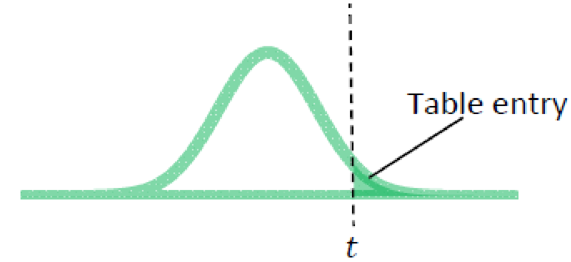


Table entry for p and C is the critical value t with probability p lying to its right and probability C lying between $-t$ and t

df	Upper Tail Probability, p											
	0.25	0.2	0.15	0.1	0.05	0.025	0.02	0.01	0.005	0.0025	0.001	0.0005
1	1.000	1.376	1.963	3.078	6.314	12.706	15.895	31.821	63.657	127.321	318.309	636.619
2	0.816	1.061	1.386	1.886	2.920	4.303	4.849	6.965	9.925	14.089	22.327	31.599
3	0.765	0.978	1.250	1.638	2.353	3.182	3.482	4.541	5.841	7.453	10.215	12.924
4	0.741	0.941	1.190	1.533	2.132	2.776	2.999	3.747	4.604	5.598	7.173	8.610
5	0.727	0.920	1.156	1.476	2.015	2.571	2.757	3.365	4.032	4.773	5.893	6.869
6	0.718	0.906	1.134	1.440	1.943	2.447	2.612	3.143	3.707	4.317	5.208	5.959
7	0.711	0.896	1.119	1.415	1.895	2.365	2.517	2.998	3.499	4.029	4.785	5.408

- pt(2.776, df=4)
  - R gives 0.975 (area to the left of t)
- qt(0.975, df = 4)
  - R gives back 2.776
- qt(0.025, df = 4)
  - R gives back -2.776 (since this is the t-value with 0.025 area to left)

One sample testing and confidence intervals

- **z-test**
  - n ≥ 30
  - Population standard deviation known or unknown
  - $z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$ or $z = \frac{\bar{x}-\mu}{s/\sqrt{n}}$
  - $\bar{x} \pm z_{CL} * {}^{\sigma}/_{\sqrt{n}}$ or $\bar{x} \pm z_{CL} * {}^{s}/_{\sqrt{n}}$
- **t-test**
  - n < 30
  - Population standard deviation unknown
  - $t = \frac{\bar{x}-\mu}{s/\sqrt{n}}$
  - $\bar{x} \pm t_{CL} * {}^{s}/_{\sqrt{n}}$

One sample testing and confidence intervals

- **z-test : one.sample.z function from the asbio package**
  - one.sample.z(

    null.mu=*[μ_0]*,

    xbar=mean(*data$variable*),

    sigma=*[value if population SD is known or sd(data$variable) if unknown]*,

    n=nrow(*data*),

    alternative=*["less", "greater" or "two.sided"])*

One sample testing and confidence intervals

- t-test: t.test function
  - t.test(
    data$variable, mu=*[$\mu_0$]*,
    alternative=*["less", "greater" or "two.sided"]*,
    conf.level=*[confidence level]*)

NOTE: must use two.sided option if you are calculating confidence intervals. Confidence intervals from other options are not correct

Example #1

A gym is interested in whether or not a 6 week weight loss training program they recently launched has been successful in helping their clients lose weight. In order to assess this, they took a sample of 30 participants and asked each to provide their weight at program initiation as well as their weight at completion. Suppose we know that for the general population, the standard deviation of changes in weights over a 6 week interval is 6 pounds. The sample mean of the change in weight for the 30 participants in the sample was -2.98 pounds. Perform a significance test to determine whether the weight loss training program they recently launched <u>has been successful in helping their clients lose weight</u> at the $\alpha = 0.05$ level of significance.

Example #1

(1) Set up the hypotheses and select the alpha level

- H_0: $\mu = 0$ (there is no effect on weight change of program participants)
- H_1: $\mu < 0$ (program participants lose weight on average)
- $\alpha = 0.05$

(2) Select the appropriate test-statistic

- $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$ since n ≥ 30 and population SD is known

Example #1

(3) State the decision rule

- Decision Rule: Reject H_0 if $p \le \alpha$
- Otherwise, do not reject H_0

Example #1

(4) Compute the test statistic and the associated p-value

- one.sample.z(null.mu=0, xbar=mean(data$wtchg),sigma=6, n=nrow(data), alternative="less", conf=.95)
- $z \approx -2.72$
- $p = P(Z \leq -2.72) = 0.0033$

(5) Conclusion

- Reject H_0 since $\mathrm{p} = 0.0033 < \alpha$. We have significant evidence at the $\alpha = 0.05$ level that $\mu < 0$ (null hypothesis is not true).
- We reject the null hypothesis that the weight loss program has no effect on weight change of program participants in favor of the alternative hypothesis that program participants lose weight on average (p = 0.0033). The mean weight loss of the sample is 2.98 pounds.

Example #2

- A scientist wishes to test the claim that great white sharks are on average 20 feet in length. To test this, he measures 10 great white sharks. Do the measurements provide evidence that great white sharks are longer than 20 feet in length at the $\alpha=0.10$ level of significance? The sample mean of the observations above is 22.27 and the sample standard deviation is (approximately) 3.31.

Example #2

(1) Set up the hypotheses and select the alpha level

- $H_0: \mu = 20$ (mean length is 20)
- $H_1: \mu > 20$ (mean length is greater than 20)
- $\alpha = 0.10$

(2) Select the appropriate test-statistic

- $t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ (since n < 30) with df = n-1

Example #2

(3) State the decision rule

- Decision Rule: Reject H_0 if $p \leq \alpha$
- Otherwise, do not reject H_0

Example #2

(4) Compute the test statistic and the associated p-value

  - t.test(*data$length*, mu=*20*, alternative=*"greater"*)
- $t \approx 2.17$, df = 9
- $p - value = 0.029$

(5) Conclusion

- Reject H_0 since $p = 0.029 < \alpha$. We have significant evidence at the $\alpha = 0.10$ level that $\mu > 20$ (null hypothesis is not true).
- We reject the null hypothesis that the average length of sharks is 20 feet in favor of the alternative hypothesis that shark lengths are greater than 20 feet on average (p = 0.029). The mean length in our sample was 22.27 feet. The 90% CI for the mean feet of sharks is 20.35 to 24.19 feet.

Two sample testing and confidence intervals

- **t-test**
  - 2 populations
  - Interested in testing if means are the same (H_0: $\mu_1 = \mu_2$)
- $t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$ $\qquad df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\left(\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}\right)}$
  - $(\bar{x}_1 - \bar{x}_2) \pm t * \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$

Two sample testing and confidence intervals

- t.test function
  - t.test(

    x,

    y,

    alternative=*[“less”, “greater” or “two.sided”]*,

    conf.level=*[confidence level]*)
  - *var.equal = FALSE* option (default) should always be used
  - For one sided tests, order that you put x and y matters! Be careful!
  - NOTE: must use two.sided option if you are calculating confidence intervals. Confidence intervals from other options are not correct

Example #3

- In order to assess how quickly polyester decays over time in landfills, a researcher buried strips of the material in the soil for different lengths of time and then tested the force required to break them (as a measure of decay). Test whether or not the breaking strengths of polyester strips buried for 2 weeks is greater than the breaking strengths of those buried for 16 weeks. Perform the test at the $\alpha = 0.10$ level of significance.

Example #3

(1) Set up the hypotheses and select the alpha level

- H_0: $\mu_1 = \mu_2$ (the mean breaking strengths for polyester are the same after 2 weeks versus after 16 weeks of decay)
- H_1: $\mu_1 > \mu_2$ (the mean breaking strengths for polyester are greater after 2 weeks versus after 16 weeks of decay), $\mu_1 - \mu_2 > 0$
- $\alpha = 0.10$

(2) Select the appropriate test-statistic

- $t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$ $\quad df = \dfrac{\left(\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}\right)^2}{\dfrac{\left(\dfrac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \dfrac{\left(\dfrac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$

Example #3

(3) State the decision rule

- Decision Rule: Reject H_0 if $p \leq \alpha$
- Otherwise, do not reject H_0

(4) Compute the test statistic and the associated p-value

- t.test(*data$brk[data$weeks==2], data$brk[data$weeks==16]*, alternative="*greater*", conf.level=0.*90*)
  - t = 0.9889, df = 4.651
  - p = 0.1857

(5) Conclusion

- Fail to reject H_0 since $p = 0.1857$ is greater than α. We do not have significant evidence at the $\alpha = 0.10$ level that the mean breaking strength is greater for polyester left to decay for 2 weeks as opposed to 16 weeks. The difference in sample means was 7.4 indicating a higher breaking strength for polyester strips left to decay 2 weeks as opposed to those left to decay for 16 weeks. We do not reject the null hypothesis that the mean breaking strength is the same between the two groups.

Numerical summaries for two sample tests

- Use aggregate or tapply function to summarize data by population
  - aggregate(data$variable, by=list(data$group), FUN = *[function]*)
  - tapply(data$variable,data$group,mean)
  - Can be used for any function (summary, sd, hist)

Graphical summaries for two sample tests

- Side by Side box plots
  - boxplot(data$variable~data$group)
- Bar graph with confidence intervals
  - #install.packages("gplots")
  - library(gplots)
  - attach(data)
  - means<-tapply(variable,group,mean)
  - lower<-tapply(variable,group,function(v) t.test(v)$conf.int[1])
  - upper<-tapply(variable,group,function(v) t.test(v)$conf.int[2])
  - barplot2(means, plot.ci = TRUE, ci.l = lower, ci.u = upper, names.arg= [labels])

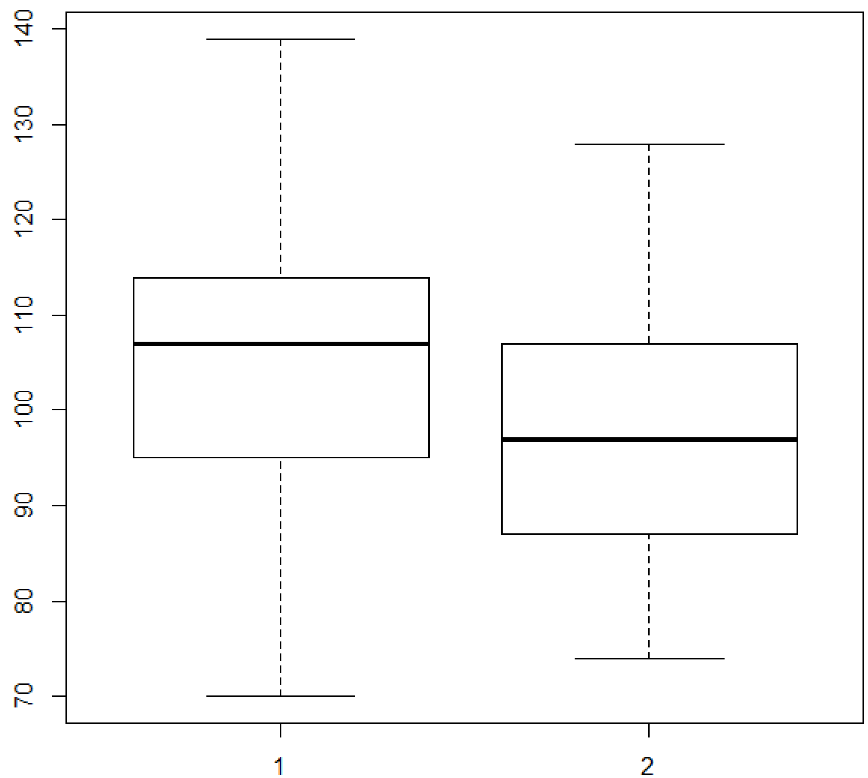
Example #4

- We are interested in whether or not students who volunteer have a different level of attachment to their friends. A study on data obtained from 74 students, 57 of which had reported volunteering in the last year. The response of interest was the score from an assessment that measured their attachment to their peers.

Group	n	Mean	Standard Deviation
1 – Volunteered in past year	57	105.32	14.68
2 – Did not Volunteer in past year	17	96.82	14.26

Example #4

- boxplot(data$score~data$group)



Example #4

- attach(data)
- means<-tapply(score,group,mean)
- lower<-tapply(score,group,function(v) t.test(v)$conf.int[1])
- upper<-tapply(score,group,function(v) t.test(v)$conf.int[2])
- barplot2(means, plot.ci = TRUE, ci.l = lower, ci.u = upper,
- names.arg= c("Yes","No"), xlab = "Volunteered in Past Year", main = "Mean Attachment Scores by Volunteer History", ylab = "Mean Attachment Score", col = "seagreen1",ylim=c(0,120))
- abline(h=0)