Comparing Multiple Means - ANOVA

DS705

Video

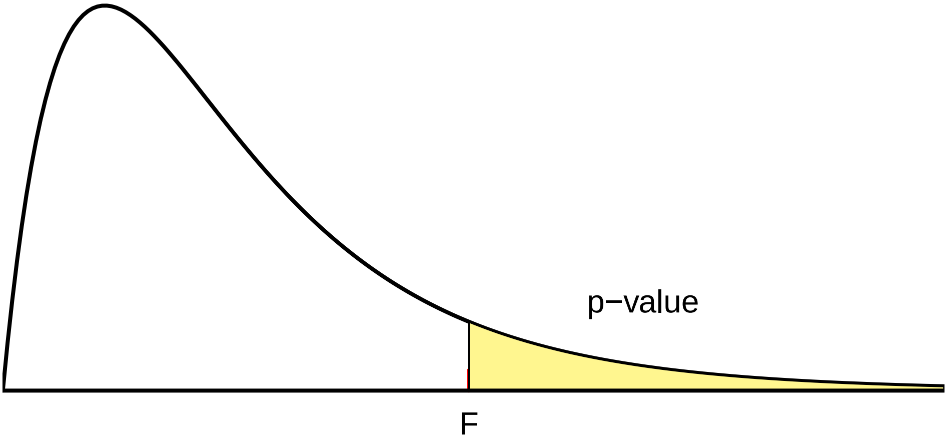
- use Slide02.mp4 here

Video

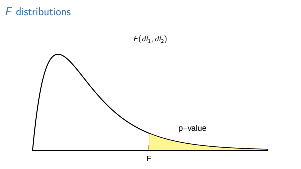
- use Slide 04.mp4 here

F distributions

$F(df_1, df_2)$



└*F* distributions



F distributions

- audio01.m4a
- F distributions are continuous, right-skewed distributions with non-negative values identified by two parameters called numerator degrees of freedom (labeled as dfN or df1) and denominator degrees of freedom (labeled as dfD or df2).
- The p-value in analysis of variance is the probability of seeing a value in the associated F distribution that is at least as big as the observed test statistic F.
- Large values of F provide more evidence against the null hypothesis. Notice, the larger F is, the smaller the p-value will be.

Video

- use video03.mp4

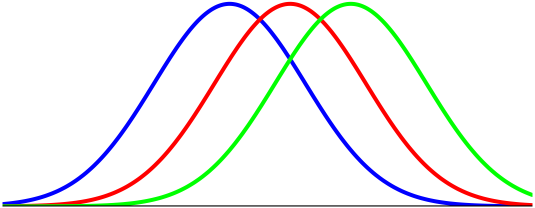
Video

- use video04.mp4

4 Self assessment slides

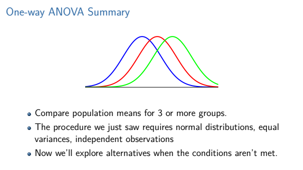
- use the 4 self assessment slides in ANOVA_self_assess.pptx

One-way ANOVA Summary



- Compare population means for 3 or more groups.
- The procedure we just saw requires normal distributions, equal variances, independent observations
- Now we'll explore alternatives when the conditions aren't met.

└─One-way ANOVA Summary



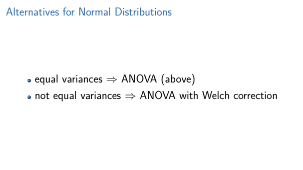
One-way ANOVA Summary

- Compare population means for 3 or more groups.
- The procedure we just saw requires normal distributions, equal variances, independent observations
- Now we'll explore alternatives when the conditions aren't met.

no audio

Alternatives for Normal Distributions

- equal variances $\Rightarrow$ ANOVA (above)
- not equal variances $\Rightarrow$ ANOVA with Welch correction



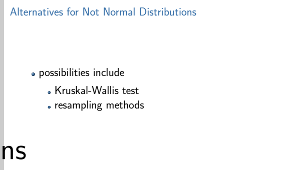
Alternatives for Normal Distributions

- equal variances ⇒ ANOVA (above)
- not equal variances ⇒ ANOVA with Welch correction

no audio

Alternatives for Not Normal Distributions

- possibilities include
  - Kruskal-Wallis test
  - resampling methods



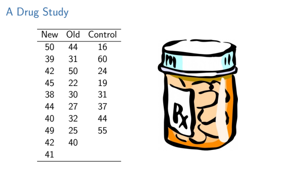
Alternatives for Not Normal Distributions

- possibilities include
  - Kruskal-Wallis test
  - resampling methods

- audio2.mp3
- Kruskall Wallis requires that the sampled distributions all have the same shape and scale and can be used to detect shifts between the populations, sometimes interpreted as a test of medians
- resampling is widely applicable, but isn't magic and may be useless if the samples are too small

A Drug Study

New	Old	Control
50	44	16
39	31	60
42	50	24
45	22	19
38	30	31
44	27	37
40	32	44
49	25	55
42	40	
41		



A Drug Study

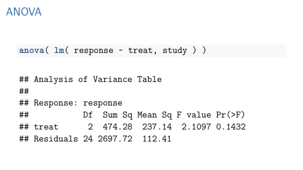
New	Old	Control
50	44	16
39	31	60
42	50	24
45	22	19
38	30	31
44	27	37
40	32	44
49	25	55
42	40	
41		

- audio03.mp3
- here is a motivating example as to why we need alternatives to standard one-way ANOVA
- suppose we have three groups for studying a drug study: new drug, old drug, and control
- we are measuring a response variable such as cholesterol level or blood pressure

ANOVA

```r
anova( lm( response ~ treat, study ) )
```

```
## Analysis of Variance Table
##
## Response: response
##           Df  Sum Sq Mean Sq F value Pr(>F)
## treat      2  474.28  237.14  2.1097 0.1432
## Residuals 24 2697.72  112.41
```



ANOVA

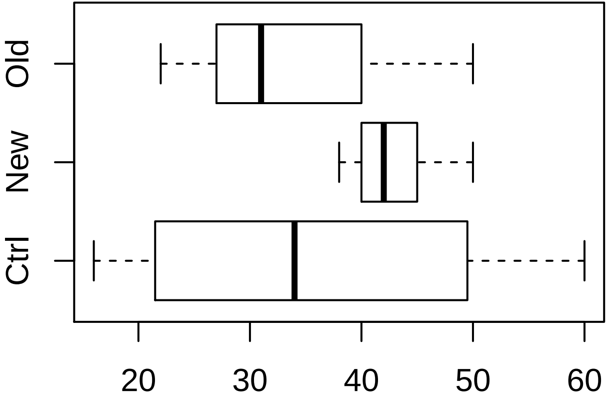
```
anova( lm( response ~ treat, study ) )
```

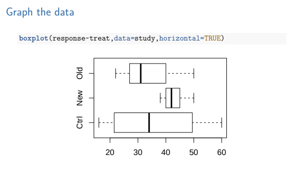
```
## Analysis of Variance Table
##
## Response: response
##           Df  Sum Sq Mean Sq F value Pr(>F)
## treat      2  474.28  237.14  2.1097 0.1432
## Residuals 24 2697.72  112.41
```

- audio04.mp3 -suppose we blindly apply ANOVA
- Since *P* is large, the ANOVA test suggests there are no significant differences between the average responses for the three treatements.
- This is wrong!
- Always EXPLORE the data first.

Graph the data

```
boxplot(response~treat,data=study,horizontal=TRUE)
```





Graph the data

```
boxplot(response~treat,data=study,horizontal=TRUE)
```

- audio05.mp3
- Never start your analysis with an inference procedure, always start by exploring the data
- We are looking at a response (on the horizontal axis) for three different treatments. Old, New, and Control.
- Notice that there are no outliers, and the boxes are symmetric ... indicating that the samples could have reasonably come from normal distributions.
- However, notice how different the scales in the boxes are ... the standard deviations are very different

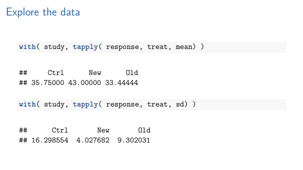
Explore the data

```r
with( study, tapply( response, treat, mean) )
```

```
##     Ctrl      New      Old
## 35.75000 43.00000 33.44444
```

```r
with( study, tapply( response, treat, sd) )
```

```
##      Ctrl       New       Old
## 16.298554  4.027682  9.302031
```



Explore the data

```
with( study, tapply( response, treat, mean) )
```

```
##     Ctrl      New      Old
## 35.75000 43.00000 33.44444
```

```
with( study, tapply( response, treat, sd) )
```

```
##      Ctrl       New       Old
## 16.298554  4.027682  9.302031
```

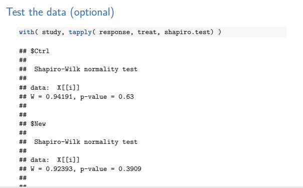
- audio06.mp3
- Notice the standard deviations differ by about a factor of 4 so the standard deviations are quite different for the groups.

Test the data (optional)

```r
with( study, tapply( response, treat, shapiro.test) )
```

```
## $Ctrl
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.94191, p-value = 0.63
##
##
## $New
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.92393, p-value = 0.3909
```

Test the data (optional)



Test the data (optional)

```
with( study, tapply( response, treat, shapiro.test) )
```

```
## $Ctrl
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.94191, p-value = 0.63
##
##
## $New
##
##  Shapiro-Wilk normality test
##
## data:  X[[i]]
## W = 0.92393, p-value = 0.3909
##
##
```

- audio07.mp3
- the results of the tests are cut off, but in each case *P* is large indicating the sample could plausibly have come from a normally distribution
- using this test probably isn't necessary even though its often standard advice
- if *n* is small, the test isn't powerful and will likely miss some departures from normality
- if *n* is large, the test will detect even a small departure from normality, but a small departure isn't important
- ANOVA procedures are robust to departures from normality, more about that below.

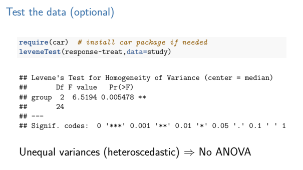
Test the data (optional)

```r
require(car)  # install car package if needed
leveneTest(response~treat,data=study)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value   Pr(>F)   
## group  2  6.5194 0.005478 **
##       24                    
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Unequal variances (heteroscedastic) ⇒ No ANOVA

└Test the data (optional)



Test the data (optional)

```
require(car)  # install car package if needed
leveneTest(response~treat,data=study)
```

```
## Levene's Test for Homogeneity of Variance (center = median)
##       Df F value   Pr(>F)
## group  2  6.5194 0.005478 **
##       24
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Unequal variances (heteroscedastic) ⇒ No ANOVA

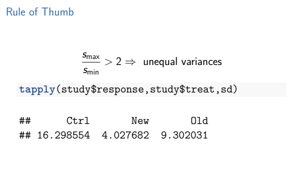
- audio08.mp3
- technically the test we've done here is called the Brown-Forsythe test which is the same as Levene execpt it computes variations about the medians and is the default in R
- Levene's test gives a small P which indicates the population variances are likely different
- again, testing for equal variances is standard advice and you can do so if you really want to
- but it's just like the Shaprio test
- if n is small, the test isn't powerful enough and will likely miss some unequal

Rule of Thumb

$$\frac{s_{\text{max}}}{s_{\text{min}}} > 2 \Rightarrow \text{ unequal variances}$$

```r
tapply(study$response,study$treat,sd)
```

```
##      Ctrl       New       Old
## 16.298554  4.027682  9.302031
```



Rule of Thumb

$$\frac{s_{\text{max}}}{s_{\text{min}}} > 2 \Rightarrow \text{ unequal variances}$$

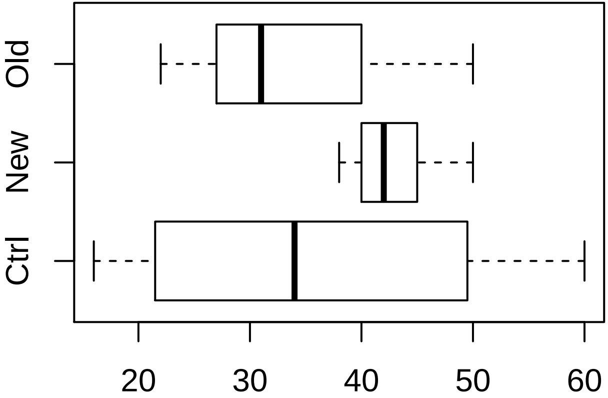
```
tapply(study$response,study$treat,sd)
```

```
##       Ctrl       New       Old
## 16.298554  4.027682  9.302031
```

- audio09.mp3
- instead of running a test of variances, just use this rule of thumb, or ALWAYS use the unequal variances ANOVA that is introduced below
- occassionally we want to do plain vanilla one-way ANOVA if we can because if we need to go under the hood, the math is easier, but usually the Welch corrected ANOVA is good enough

ANOVA Failed

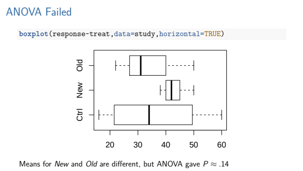
```
boxplot(response~treat,data=study,horizontal=TRUE)
```



Means for *New* and *Old* are different, but ANOVA gave $P \approx .14$

Comparing Multiple Means - ANOVA

└ANOVA Failed



ANOVA Failed

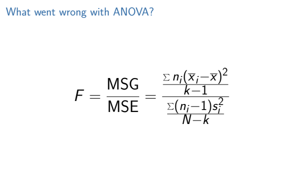
```
boxplot(response~treat,data=study,horizontal=TRUE)
```

Means for *New* and *Old* are different, but ANOVA gave $P \approx .14$

no audio

What went wrong with ANOVA?

$$F = \frac{\text{MSG}}{\text{MSE}} = \frac{\frac{\sum n_i(\overline{x}_i - \overline{x})^2}{k-1}}{\frac{\sum (n_i - 1)s_i^2}{N-k}}$$



What went wrong with ANOVA?

$$F = \frac{\mathrm{MSG}}{\mathrm{MSE}} = \frac{\frac{\sum n_i(\overline{x}_i - \overline{x})^2}{k-1}}{\frac{\Sigma(n_i-1)s_i^2}{N-k}}$$

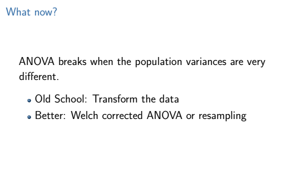
- audio10.mp3
- it is pretty clear that the responses for the NEW and OLD treatments are different
- the denominator is the estimated pooled variance of the pooled data
- since the variance for the Control group is so large ANOVA it overwhelms the difference between the NEW and OLD treatements giving a small F and a large P value.

What now?

ANOVA breaks when the population variances are very different.

- Old School: Transform the data
- Better: Welch corrected ANOVA or resampling

└What now?



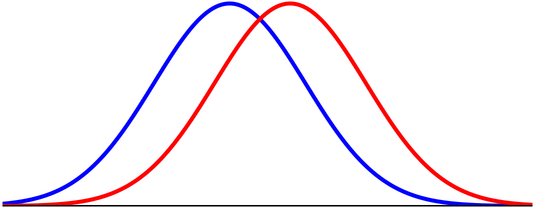
What now?

ANOVA breaks when the population variances are very different.

- Old School: Transform the data
- Better: Welch corrected ANOVA or resampling

- audio11.mp3
- its often possible to transform the data in such a way that the variances are similar, but transforming the data makes the interpretation more difficult
- The Kruskal-Wallis test, which we will talk about soon, is often touted as an alternative anytime ANOVA goes wrong, but it requires that all of the populations have the same shape and scales only possibly be shifts of each other

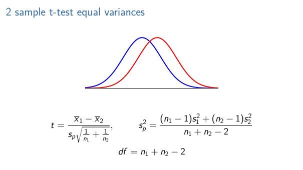
2 sample t-test equal variances



$$t = \frac{\overline{x}_1 - \overline{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \qquad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$df = n_1 + n_2 - 2$$

└2 sample t-test equal variances



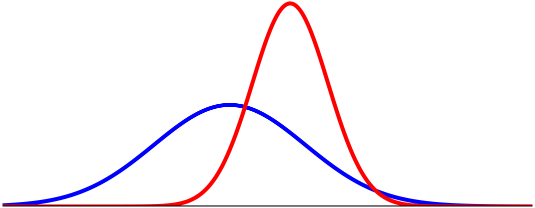
2 sample t-test equal variances

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s_p\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \qquad s_p^2 = \frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1 + n_2 - 2}$$

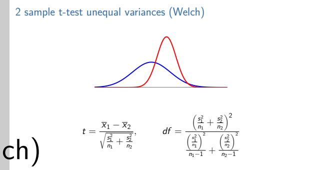
$$df = n_1 + n_2 - 2$$

- audio12.mp3
- what's the t-test for two population means doing here?
- this is the pooled t-test where the population variances are assumed to be equal
- the variance is estimated by pooling all of the data and computing the pooled variance
- just like ANOVA

2 sample t-test unequal variances (Welch)



$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \qquad df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}}$$



2 sample t-test unequal variances (Welch)

$$t = \frac{\overline{x}_1 - \overline{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}, \qquad df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$$

- audio13.mp3
- the 2 sample unequal variances t-test does not pool the variances
- instead it uses a Welch correction to reduce the degrees of freedom so that the t distribution using the unpooled variances is approximately correct
- Welch corrected ANOVA does this same thing. It uses an unpooled variance estimate and adjusts the degrees of freedom so that F distribution is approximately correct

Welch's ANOVA

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k, \qquad H_a : \text{ not all the means are the same}$$

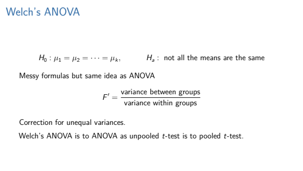
Messy formulas but same idea as ANOVA

$$F' = \frac{\text{variance between groups}}{\text{variance within groups}}$$

Correction for unequal variances.

Welch's ANOVA is to ANOVA as unpooled t-test is to pooled t-test.

└─Welch's ANOVA



Welch's ANOVA

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_k, \qquad H_a : \text{ not all the means are the same}$$

Messy formulas but same idea as ANOVA

$$F' = \frac{\text{variance between groups}}{\text{variance within groups}}$$

Correction for unequal variances.

Welch's ANOVA is to ANOVA as unpooled t-test is to pooled t-test.

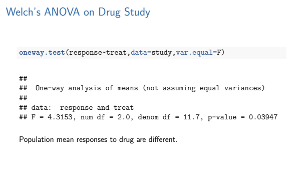
no audio

Welch's ANOVA on Drug Study

```r
oneway.test(response~treat,data=study,var.equal=F)
```

```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  response and treat
## F = 4.3153, num df = 2.0, denom df = 11.7, p-value = 0.03947
```

Population mean responses to drug are different.



Welch's ANOVA on Drug Study

```r
oneway.test(response~treat,data=study,var.equal=F)
```

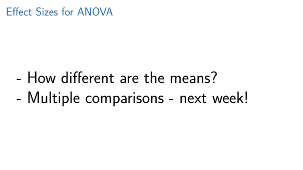
```
##
##  One-way analysis of means (not assuming equal variances)
##
## data:  response and treat
## F = 4.3153, num df = 2.0, denom df = 11.7, p-value = 0.03947
```

Population mean responses to drug are different.

- audio14.mp3
- The function oneway.test can also do regular ANOVA if the var.equal is set to TRUE

Effect Sizes for ANOVA

- How different are the means?
- Multiple comparisons - next week!



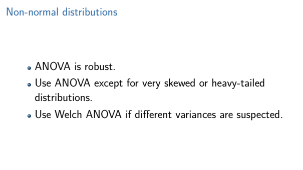
Effect Sizes for ANOVA

- How different are the means?
- Multiple comparisons - next week!

- audio15.mp3
- you should never report just a P value, a small P value is an indication of statistical significance, but practical significance is determined by looking at how big the effects are.
- there are some effect sizes for ANOVA, but they aren't very meaninful from a practical perspective
- ANOVA and the other tests this week are omnibus tests, that is, they test all of the means simultaneously to see if there may be a difference somewhere, but they don't tell you where or how big that difference is
- Next week we'll look at multiple comparisons which are procedures for determining which means are different and estimating how different they are — these

Non-normal distributions

- ANOVA is robust.
- Use ANOVA except for very skewed or heavy-tailed distributions.
- Use Welch ANOVA if different variances are suspected.

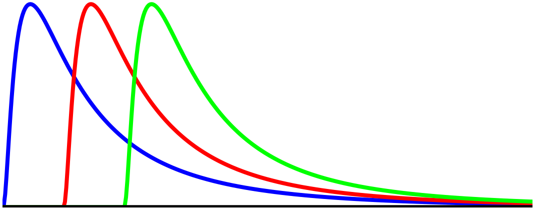


Non-normal distributions

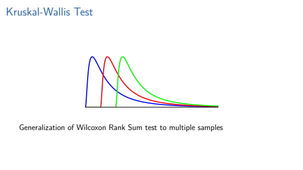
- ANOVA is robust.
- Use ANOVA except for very skewed or heavy-tailed distributions.
- Use Welch ANOVA if different variances are suspected.

- audio16.mp3
- we’ve seen how to fix ANOVA when the population variances are different, but what if the populations are not normally distributed
- ANOVA is robust to departures from normality as long as they aren’t really severe
- ANOVA is built on testing means and variances which are sensitive to extreme outliers
- extreme outliers, typical in heavy-tailed and very skewed distributions, cause ANOVA to be inaccurate
- we’ll meet some alternatives to ANOVA below

Kruskal-Wallis Test



Generalization of Wilcoxon Rank Sum test to multiple samples



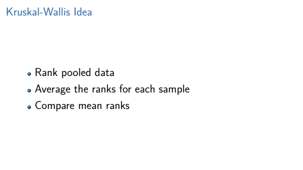
Kruskal-Wallis Test

Generalization of Wilcoxon Rank Sum test to multiple samples

- audio17.mp3 Under the right circumstances the Kruskal Wallis test can detect differences between the population medians of various groups

Kruskal-Wallis Idea

- Rank pooled data
- Average the ranks for each sample
- Compare mean ranks



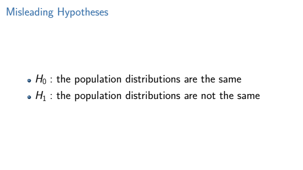
Kruskal-Wallis Idea

- Rank pooled data
- Average the ranks for each sample
- Compare mean ranks

- audio18.mp3
- formulas and examples may be found in your book
- We'll see how to do this in R in a minute

Misleading Hypotheses

- H_0 : the population distributions are the same
- H_1 : the population distributions are not the same



Misleading Hypotheses

- H_0 : the population distributions are the same
- H_1 : the population distributions are not the same

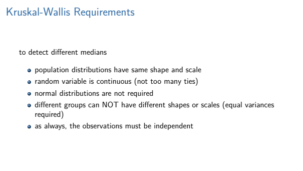
- audio19.mp3
- just as in your textbook these hypotheses are misleading
- If the populations all have same shape and scale but are possibly shifted relative to each other, then KW can determine if there are shifts, . . . are the medians different?
- However there are differenences in distributions that cannot be detected by the KW test. One such example is in your homework for this week.

Kruskal-Wallis Requirements

to detect different medians

- population distributions have same shape and scale
- random variable is continuous (not too many ties)
- normal distributions are not required
- different groups can NOT have different shapes or scales (equal variances required)
- as always, the observations must be independent

└Kruskal-Wallis Requirements



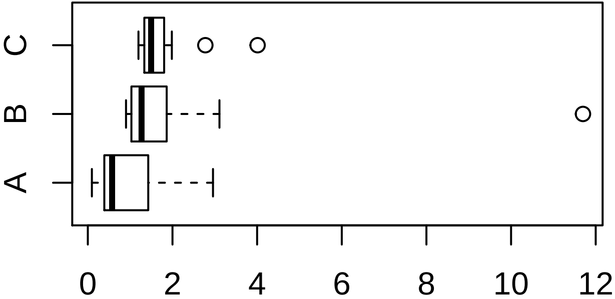
Kruskal-Wallis Requirements

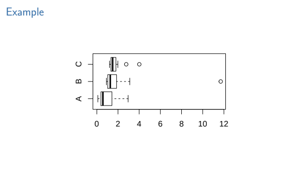
to detect different medians

- population distributions have same shape and scale
- random variable is continuous (not too many ties)
- normal distributions are not required
- different groups can NOT have different shapes or scales (equal variances required)
- as always, the observations must be independent

No audio.

Example





Example

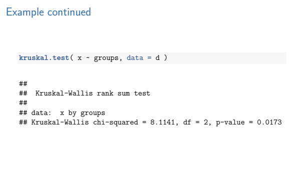
- audio20.mp3
- three samples of size 12 very skewed distribuitons
- can see that the shapes are similar in the boxplot, but there are shifts
- the extreme outliers can pose a problem for ANOVA, but OK for KW since average ranks are not sensitive to outliers

Example continued

```r
kruskal.test( x ~ groups, data = d )
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  x by groups
## Kruskal-Wallis chi-squared = 8.1141, df = 2, p-value = 0.0173
```

└Example continued



Example continued

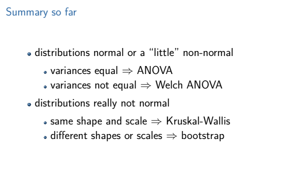
```
kruskal.test( x ~ groups, data = d )
```

```
##
##  Kruskal-Wallis rank sum test
##
## data:  x by groups
## Kruskal-Wallis chi-squared = 8.1141, df = 2, p-value = 0.0173
```

- audio21.mp3
- KW test is simple to run in R
- small *P* implies we should reject the null hypothesis
- since the samples appear to come from distributions with the same shape and scale we can conclude the population medians are different

Summary so far

- distributions normal or a “little” non-normal
  - variances equal ⇒ ANOVA
  - variances not equal ⇒ Welch ANOVA
- distributions really not normal
  - same shape and scale ⇒ Kruskal-Wallis
  - different shapes or scales ⇒ bootstrap



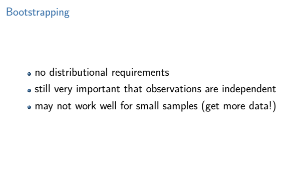
Summary so far

- distributions normal or a “little” non-normal
  - variances equal ⇒ ANOVA
  - variances not equal ⇒ Welch ANOVA
- distributions really not normal
  - same shape and scale ⇒ Kruskal-Wallis
  - different shapes or scales ⇒ bootstrap

- audio22.mp3
- what does a little non-normal mean? mound shaped is good, but even moderate skewness or mild outliers are not really a problem, extreme outliers or skewness can cause issues
- if you’re unsure what procedure to use, you can run multiple procedures
  - this doesn’t mean you get to keep trying until you get the result that you want
  - but if the procedures all agree, then you know you are on the right track
  - if different procedures produce different results, then you need to step back and think harder about your data and determine which procedure most closely matches what you know about your data . . .
- when we find ourselves having multiple groups that have different shapes or scales

Bootstrapping

- no distributional requirements
- still very important that observations are independent
- may not work well for small samples (get more data!)



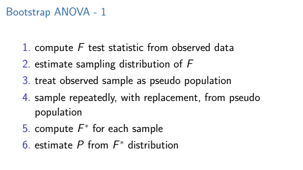
Bootstrapping

- no distributional requirements
- still very important that observations are independent
- may not work well for small samples (get more data!)

no audio

Bootstrap ANOVA - 1

1. compute F test statistic from observed data
2. estimate sampling distribution of F
3. treat observed sample as pseudo population
4. sample repeatedly, with replacement, from pseudo population
5. compute F^* for each sample
6. estimate P from F^* distribution



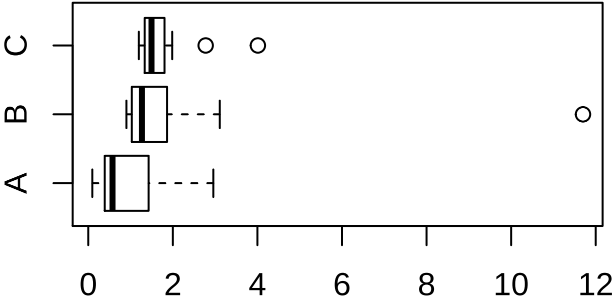
Bootstrap ANOVA - 1

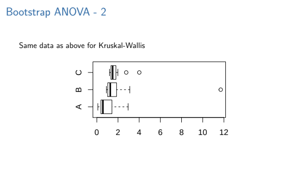
1. compute F test statistic from observed data
2. estimate sampling distribution of F
3. treat observed sample as pseudo population
4. sample repeatedly, with replacement, from pseudo population
5. compute F^* for each sample
6. estimate P from F^* distribution

- audio23.mp3
- if distributions aren't normal, then we don't have a theoretical sampling distribution for F
- approximate it by pretending the observed data is the population and repeatedly resampling from the *pseudo* population to simulate F, we call these simulated values F^*

Bootstrap ANOVA - 2

Same data as above for Kruskal-Wallis





Bootstrap ANOVA - 2

Same data as above for Kruskal-Wallis

no audio

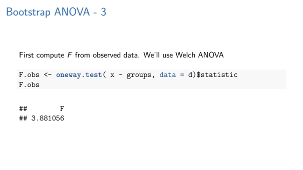
Bootstrap ANOVA - 3

First compute *F* from observed data. We'll use Welch ANOVA

```
F.obs <- oneway.test( x ~ groups, data = d)$statistic
F.obs
```

```
##        F
## 3.881056
```

└Bootstrap ANOVA - 3



Bootstrap ANOVA - 3

First compute *F* from observed data. We'll use Welch ANOVA

```
F.obs <- oneway.test( x ~ groups, data = d)$statistic
F.obs
```

```
##        F
## 3.881056
```

- audio24.mp3
- we use the Welch corrected ANOVA here since the large outlier gives group B a much larger standard deviation than the other groups
- in any case, using Welch anova all the time is not a bad idea

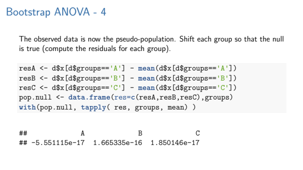
Bootstrap ANOVA - 4

The observed data is now the pseudo-population. Shift each group so that the null is true (compute the residuals for each group).

```r
resA <- d$x[d$groups=='A'] - mean(d$x[d$groups=='A'])
resB <- d$x[d$groups=='B'] - mean(d$x[d$groups=='B'])
resC <- d$x[d$groups=='C'] - mean(d$x[d$groups=='C'])
pop.null <- data.frame(res=c(resA,resB,resC),groups)
with(pop.null, tapply( res, groups, mean) )
```

```
##             A             B             C
## -5.551115e-17  1.665335e-16  1.850146e-17
```

└Bootstrap ANOVA - 4



Bootstrap ANOVA - 4

The observed data is now the pseudo-population. Shift each group so that the null is true (compute the residuals for each group).

```
resA <- d$x[d$groups=='A'] - mean(d$x[d$groups=='A'])
resB <- d$x[d$groups=='B'] - mean(d$x[d$groups=='B'])
resC <- d$x[d$groups=='C'] - mean(d$x[d$groups=='C'])
pop.null <- data.frame(res=c(resA,resB,resC),groups)
with(pop.null, tapply( res, groups, mean) )
```

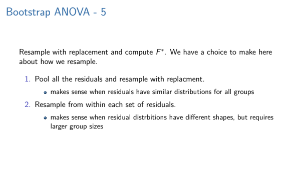
```
##             A             B             C
## -5.551115e-17  1.665335e-16  1.850146e-17
```

- audio25.mp3
- The mean of each group of residuals, also called errors, is 0 to nearly machine precision
- Now we a pseudopopulation consisting of three groups whose means are identially 0. This is a surrogate for the population our data might have come from if the null hypothesis of equal means were true.

Bootstrap ANOVA - 5

Resample with replacement and compute F^*. We have a choice to make here about how we resample.

1. Pool all the residuals and resample with replacment.
   - makes sense when residuals have similar distributions for all groups
2. Resample from within each set of residuals.
   - makes sense when residual distrbitions have different shapes, but requires larger group sizes



Bootstrap ANOVA - 5

Resample with replacement and compute F^*. We have a choice to make here about how we resample.

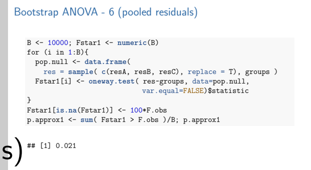
1. Pool all the residuals and resample with replacment.
   - makes sense when residuals have similar distributions for all groups
2. Resample from within each set of residuals.
   - makes sense when residual distrbitions have different shapes, but requires larger group sizes

- audio26.mp3
- Since the groups here all have the same approximate shape, except for an occasional outlier, it makes sense to pool the residuals. We'll use the first resampling approach as it seems to make the most sense, but we'll cycle back and try the second approach also for comparison.

Bootstrap ANOVA - 6 (pooled residuals)

```
B <- 10000; Fstar1 <- numeric(B)
for (i in 1:B){
  pop.null <- data.frame(
    res = sample( c(resA, resB, resC), replace = T), groups )
  Fstar1[i] <- oneway.test( res~groups, data=pop.null,
                            var.equal=FALSE)$statistic
}
Fstar1[is.na(Fstar1)] <- 100*F.obs
p.approx1 <- sum( Fstar1 > F.obs )/B; p.approx1
```

```
## [1] 0.021
```



Bootstrap ANOVA - 6 (pooled residuals)

```
B <- 10000; Fstar1 <- numeric(B)
for (i in 1:B){
  pop.null <- data.frame(
    res = sample( c(resA, resB, resC), replace = T), groups )
  Fstar1[i] <- oneway.test( res~groups, data=pop.null,
                            var.equal=FALSE)$statistic
}
Fstar1[is.na(Fstar1)] <- 100*F.obs
p.approx1 <- sum( Fstar1 > F.obs )/B; p.approx1
```

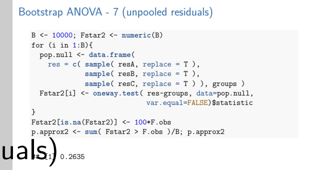
```
## [1] 0.021
```

- audio27.mp3
- BELOW SLIDE add note: The second line from the end guards against division by zero errors. A zero in the denominator of F indicates the between group variance is much larger than the variation within groups so we set F to a very large value.
- using a for loop is a very inefficient way to sample, but we're aiming for clarity here and not speed
- the *P* value is just the proportion of values of F-star that are greater than the original observed F, that is . . . F-star is our surrogate sampling distributiion
- notice that the P-value here is very similar to what KW yielded, both agree that the populations are not all the same

Bootstrap ANOVA - 7 (unpooled residuals)

```r
B <- 10000; Fstar2 <- numeric(B)
for (i in 1:B){
  pop.null <- data.frame(
    res = c( sample( resA, replace = T ),
             sample( resB, replace = T ),
             sample( resC, replace = T ) ), groups )
  Fstar2[i] <- oneway.test( res~groups, data=pop.null,
                            var.equal=FALSE)$statistic
}
Fstar2[is.na(Fstar2)] <- 100*F.obs
p.approx2 <- sum( Fstar2 > F.obs )/B; p.approx2
```

```
## [1] 0.2635
```



Bootstrap ANOVA - 7 (unpooled residuals)

```
B <- 10000; Fstar2 <- numeric(B)
for (i in 1:B){
  pop.null <- data.frame(
    res = c( sample( resA, replace = T ),
             sample( resB, replace = T ),
             sample( resC, replace = T ) ), groups )
  Fstar2[i] <- oneway.test( res~groups, data=pop.null,
                            var.equal=FALSE)$statistic
}
Fstar2[is.na(Fstar2)] <- 100*F.obs
p.approx2 <- sum( Fstar2 > F.obs )/B; p.approx2
```

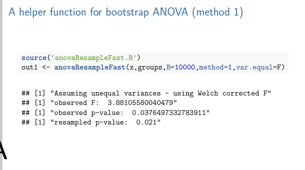
```
## [1] 0.2635
```

- audio28.mp3
- Whoa, that last approximated *P*-value is really different.
- Since we are sampling within each group we need larger samples to soften the effect of the extreme outlier in group B
- I'm more inclined to accept the result of the first method, particularly since it agrees with the KW test
- small samples, extreme outliers, averages and standard deviations do not play nicely with each other

A helper function for bootstrap ANOVA (method 1)

```r
source('anovaResampleFast.R')
out1 <- anovaResampleFast(x,groups,B=10000,method=1,var.equal=F)
```

```
## [1] "Assuming unequal variances - using Welch corrected F"
## [1] "observed F:  3.88105580040479"
## [1] "observed p-value:  0.0376497332783911"
## [1] "resampled p-value:  0.021"
```



A helper function for bootstrap ANOVA (method 1)

```
source('anovaResampleFast.R')
out1 <- anovaResampleFast(x,groups,B=10000,method=1,var.equal=F)
```

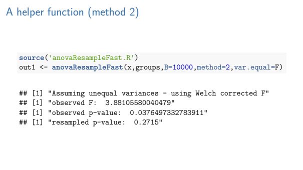
```
## [1] "Assuming unequal variances - using Welch corrected F"
## [1] "observed F:  3.88105580040479"
## [1] "observed p-value:  0.0376497332783911"
## [1] "resampled p-value:  0.021"
```

- in your download for the week you should find the anovaResampleFast function as well as a slower version that uses a for loop.
- the for loop version is there for you to study if you want to see how it works
- the fast version is much faster but at the cost of making the code harder to follow
- notice the result is quite similar to the simulated P above

A helper function (method 2)

```r
source('anovaResampleFast.R')
out1 <- anovaResampleFast(x,groups,B=10000,method=2,var.equal=F)
```

```
## [1] "Assuming unequal variances - using Welch corrected F"
## [1] "observed F:  3.88105580040479"
## [1] "observed p-value:  0.0376497332783911"
## [1] "resampled p-value:  0.2715"
```



A helper function (method 2)

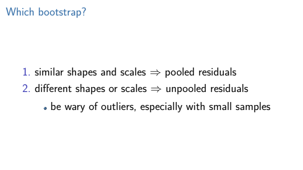
```
source('anovaResampleFast.R')
out1 <- anovaResampleFast(x,groups,B=10000,method=2,var.equal=F)
```

```
## [1] "Assuming unequal variances - using Welch corrected F"
## [1] "observed F:  3.88105580040479"
## [1] "observed p-value:  0.0376497332783911"
## [1] "resampled p-value:  0.2715"
```

no audio

Which bootstrap?

1. similar shapes and scales ⇒ pooled residuals
2. different shapes or scales ⇒ unpooled residuals
    - be wary of outliers, especially with small samples



Which bootstrap?

1. similar shapes and scales ⇒ pooled residuals
2. different shapes or scales ⇒ unpooled residuals
    - be wary of outliers, especially with small samples

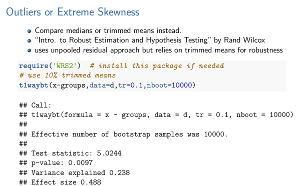
no audio

Outliers or Extreme Skewness

- Compare medians or trimmed means instead.
- “Intro. to Robust Estimation and Hypothesis Testing” by Rand Wilcox
- uses unpooled residual approach but relies on trimmed means for robustness

```r
require('WRS2')  # install this package if needed
# use 10% trimmed means
t1waybt(x~groups,data=d,tr=0.1,nboot=10000)
```

```
## Call:
## t1waybt(formula = x ~ groups, data = d, tr = 0.1, nboot = 10000)
##
## Effective number of bootstrap samples was 10000.
##
## Test statistic: 5.0244
## p-value: 0.0097
```



Outliers or Extreme Skewness

- Compare medians or trimmed means instead.
- “Intro. to Robust Estimation and Hypothesis Testing” by Rand Wilcox
- uses unpooled residual approach but relies on trimmed means for robustness

```
require('WRS2')  # install this package if needed
# use 10% trimmed means
t1waybt(x~groups,data=d,tr=0.1,nboot=10000)
```

```
## Call:
## t1waybt(formula = x ~ groups, data = d, tr = 0.1, nboot = 10000)
##
## Effective number of bootstrap samples was 10000.
##
## Test statistic: 5.0244
## p-value: 0.0097
## Variance explained 0.238
## Effect size 0.488
```

- audio30.mp3
- if you’ve got extreme outliers or skewness in the data, it probably isn’t a good idea to think about means anyway
- in the last decade or two a lot of attention has been paid to robust methods
- robust methods replace means with trimmed means or medians since those are not so sensitive to outliers and skewness
- the t1waybt command in package WRS2 performs bootstrap ANOVA using unpooled residuals and Welch’s correction, but it works on the trimmed means instead of the means
- a 10% trimmed mean takes 10% of the data from each end before calculating the

Fast Facts: ANOVA (equal variances)

Why: Hypothesis test - To compare two or more unknown population means

When: The following conditions are necessary for these procedures to be accurate and valid. Some may have to be assumed, but be careful in doing so.

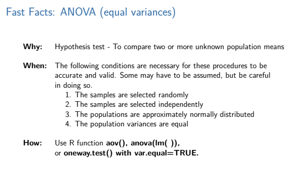
1. The samples are selected randomly
2. The samples are selected independently
3. The populations are approximately normally distributed
4. The population variances are equal

How: Use R function **aov(), anova(lm( )),** or **oneway.test() with var.equal=TRUE.**

2018-01-08

Comparing Multiple Means - ANOVA

└Fast Facts: ANOVA (equal variances)



Fast Facts: ANOVA (equal variances)

Why: Hypothesis test - To compare two or more unknown population means

When: The following conditions are necessary for these procedures to be accurate and valid. Some may have to be assumed, but be careful in doing so.

1. The samples are selected randomly
2. The samples are selected independently
3. The populations are approximately normally distributed
4. The population variances are equal

How: Use R function **aov(), anova(lm( )),**
or **oneway.test() with var.equal=TRUE.**

No audio

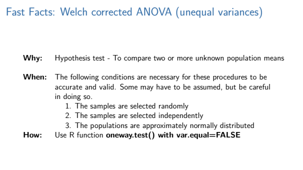
Fast Facts: Welch corrected ANOVA (unequal variances)

Why: Hypothesis test - To compare two or more unknown population means

When: The following conditions are necessary for these procedures to be accurate and valid. Some may have to be assumed, but be careful in doing so.

1. The samples are selected randomly
2. The samples are selected independently
3. The populations are approximately normally distributed

How: Use R function **oneway.test() with var.equal=FALSE**



Fast Facts: Welch corrected ANOVA (unequal variances)

Why: Hypothesis test - To compare two or more unknown population means

When: The following conditions are necessary for these procedures to be accurate and valid. Some may have to be assumed, but be careful in doing so.

1. The samples are selected randomly
2. The samples are selected independently
3. The populations are approximately normally distributed

How: Use R function **oneway.test() with var.equal=FALSE**

No audio

Fast Facts: Kruskal-Wallis

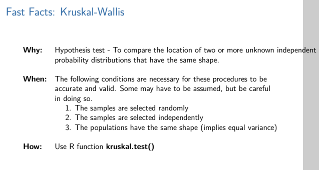
Why: Hypothesis test - To compare the location of two or more unknown indepen
probability distributions that have the same shape.

When: The following conditions are necessary for these procedures to be accurate and valid. Some may have to be assumed, but be careful in doing so.

1. The samples are selected randomly
2. The samples are selected independently
3. The populations have the same shape (implies equal variance)

How: Use R function **kruskal.test()**

└─Fast Facts: Kruskal-Wallis



Fast Facts: Kruskal-Wallis

Why: Hypothesis test - To compare the location of two or more unknown independent probability distributions that have the same shape.

When: The following conditions are necessary for these procedures to be accurate and valid. Some may have to be assumed, but be careful in doing so.

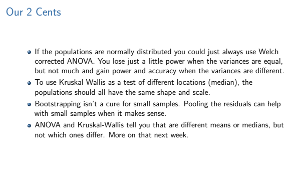
1. The samples are selected randomly
2. The samples are selected independently
3. The populations have the same shape (implies equal variance)

How: Use R function **kruskal.test()**

No audio

Our 2 Cents

- If the populations are normally distributed you could just always use Welch corrected ANOVA. You lose just a little power when the variances are equal, but not much and gain power and accuracy when the variances are different.
- To use Kruskal-Wallis as a test of different locations (median), the populations should all have the same shape and scale.
- Bootstrapping isn't a cure for small samples. Pooling the residuals can help with small samples when it makes sense.
- ANOVA and Kruskal-Wallis tell you that are different means or medians, but not which ones differ. More on that next week.



Our 2 Cents

- If the populations are normally distributed you could just always use Welch corrected ANOVA. You lose just a little power when the variances are equal, but not much and gain power and accuracy when the variances are different.
- To use Kruskal-Wallis as a test of different locations (median), the populations should all have the same shape and scale.
- Bootstrapping isn't a cure for small samples. Pooling the residuals can help with small samples when it makes sense.
- ANOVA and Kruskal-Wallis tell you that are different means or medians, but not which ones differ. More on that next week.

No audio