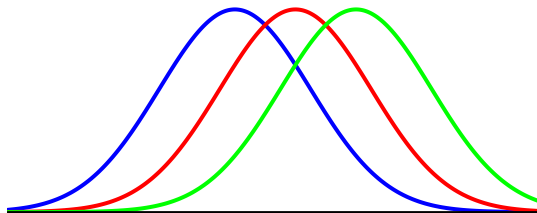


Multiple Comparison Hypothesis Tests

One-way ANOVA Review



- compare population means for 3 or more groups
- requires normal distributions, equal variances, independent observations
- This week = alternatives.

Alternatives for Normal Distributions

- equal variances \Rightarrow ANOVA (last week)
- not equal variances \Rightarrow ANOVA with Welch correction

Alternatives for Not Normal Distributions

- possibilities include
 - Kruskal-Wallis test
 - resampling methods

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		



ANOVA

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

```
## Analysis of Variance Table
```

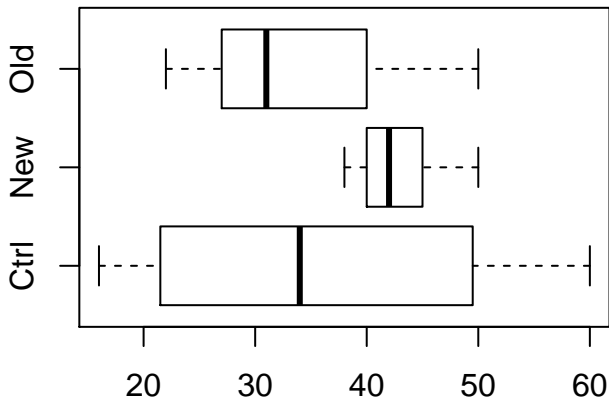
```
##
```

```
## 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		

Graph the data

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



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
```


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
```

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) \Rightarrow No ANOVA

Rule of Thumb

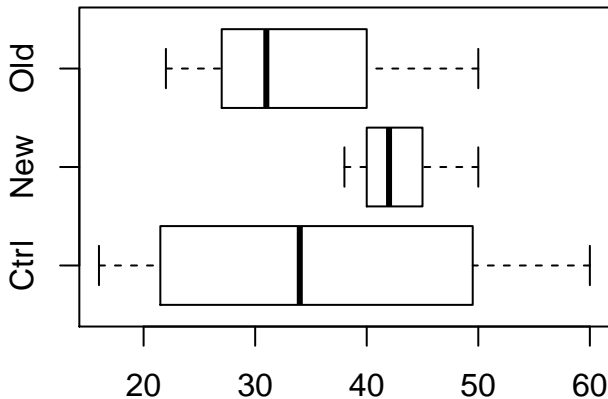
$$\frac{s_{\max}}{s_{\min}} > 2 \Rightarrow \text{unequal variances}$$

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

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

ANOVA Failed

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



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

What went wrong with ANOVA?

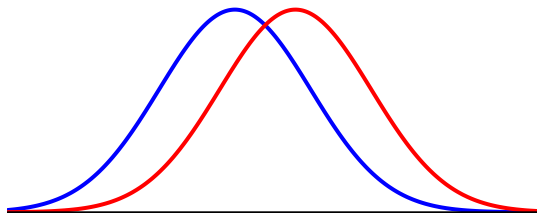
$$F = \frac{\text{MSG}}{\text{MSE}} = \frac{\frac{\sum n_i (\bar{x}_i - \bar{x})^2}{k-1}}{\frac{\sum (n_i - 1) s_i^2}{N-k}}$$

What now?

ANOVA breaks when the population variances are very different.

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

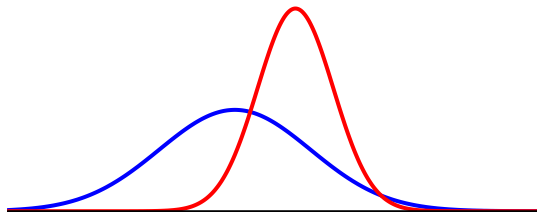
2 sample t-test equal variances



$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}, \quad s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$df = n_1 + n_2 - 2$$

2 sample t-test unequal variances (Welch)



$$t = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}},$$

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1-1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2-1}}$$

Welch's ANOVA

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

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 on Drug Study

```
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.

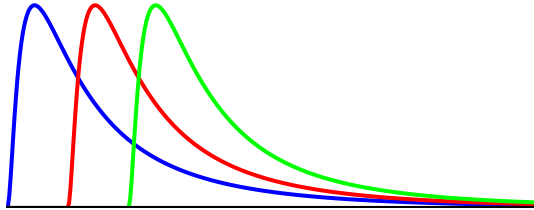
Effect Sizes for ANOVA

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

Non-normal distributions

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

Kruskal-Wallis Test



Generalization of Wilcoxon Rank Sum test to multiple samples

Kruskal-Wallis Idea

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

Misleading Hypotheses

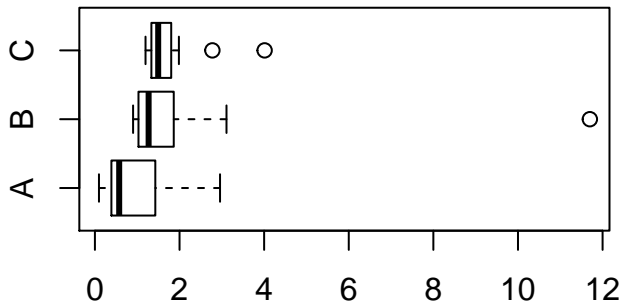
- H_0 : the population distributions are the same
- H_1 : the population distributions are not the same

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

Example



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
```

Summary so far

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

Bootstrapping

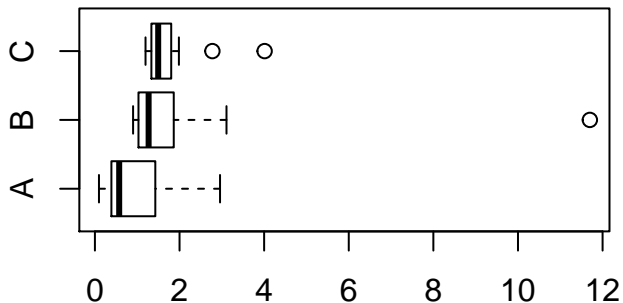
- no distributional requirements
- still very important that observations are independent
- may not work well for small samples (get more data!)

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 - 2

Same data as above for Kruskal-Wallis



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 - 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
```


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 replacement.
 - makes sense when residuals have similar distributions for all groups
2. Resample from within each set of residuals.
 - makes sense when residual distributions have different shapes, but requires larger group sizes

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
}
p.approx1 <- sum( Fstar1 > F.obs )/B; p.approx1
```

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

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
}
p.approx2 <- sum( Fstar2 > F.obs )/B; p.approx2
```

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

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"
```

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"
```

Which bootstrap?

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

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
```