Differences between more than two samples

Introduction to Quantitative Ecology
Fall 2018
Chris Sutherland
csutherland@umass.edu

ANOVA group exercise

Fill in the ANOVA table for the following data

- group A: 39, 43, 61, 46, 46, 62, 50, 32, 38, 41
- group B: 97, 89, 89, 86, 79, 103, 90, 65, 92, 80
- group C: 54, 63, 55, 58, 59, 48, 73, 67, 54, 78

Variation	SS	df	MS	F	significant?
Between-group	?	?	?	?	?
Within-group	?	?	?		

ANOVA group exercise

Fill in the ANOVA table for the following data

- group A: 39, 43, 61, 46, 46, 62, 50, 32, 38, 41
- group B: 97, 89, 89, 86, 79, 103, 90, 65, 92, 80
- group C: 54, 63, 55, 58, 59, 48, 73, 67, 54, 78

Variation	SS	df	MS	F	significant?
Between-group	8688.87	2	4344.43	44.69	YES!
Within-group	2624.50	27	97.20		

ANOVA group exercise - answers

```
#calculate the means
mean.A <- mean(group.A)
mean.A
[1] 45.8
```

```
mean.B <- mean(group.B)
mean.B
[1] 87
```

```
mean.C <- mean(group.C)
mean.C
[1] 60.9
```

```
mean.All <- mean(c(group.A,group.B,group.C))
mean.All
[1] 64.56667
```

ANOVA group exercise - answers

```
#between-groups sums of squares (SS)
bSS.A <- 10 * (mean.A - mean.All)^2
bSS.B <- 10 * (mean.B - mean.All)^2
bSS.C <- 10 * (mean.C - mean.All)^2
bSS <- sum(bSS.A, bSS.B, bSS.C)

#between-groups degrees of freedom (df)
bDF <- 3-1 #number of groups - 1

#between-groups mean squares (MS)
bMS <- bSS / bDF

c(bSS, bDF, bMS)
[1] 8688.867    2.000 4344.433
```

ANOVA group exercise - answers

```
#within-groups sums of squares (SS)
wSS.A <- sum((group.A - mean.A)^2)
wSS.B <- sum((group.B - mean.B)^2)
wSS.C <- sum((group.C - mean.C)^2)
wSS <- wSS.A + wSS.B + wSS.C

#within-groups degrees of freedom (df)
wDF <- 30-3 #number of observations - number of groups

#bwithin-groups mean squares (MS)
wMS <- wSS / wDF

c(wSS, wDF, wMS)
[1] 2624.5000   27.0000   97.2037
```

ANOVA group exercise - answers

```
#Calcualte the F statistic
Fstat <- bMS / wMS
Fstat
[1] 44.69411
```

ANOVA group exercise - answers

```
#check working against R's anova functions
c(bSS, bDF, bMS)
[1] 8688.867    2.000 4344.433
```

```
c(wSS, wDF, wMS)
[1] 2624.5000   27.0000   97.2037
```

```
Fstat
[1] 44.69411
```

```
anova(aov(vals~groups))
Analysis of Variance Table

Response: vals
          Df Sum Sq Mean Sq F value    Pr(>F)
groups     2 8688.9  4344.4  44.694 2.714e-09 ***
Residuals 27 2624.5    97.2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

A question!

Having just analyzed some fish counts data in 16 lakes in Massachusetts, Thorsten found a significant ‘lake’ effect using an ANOVA, i.e., the mean number of fish was not the same in all lakes.

1. What would Thorsten do to find out *which lakes were different from eachother*?

A) A series of t-tests

B) A Tukey Honest Significant Difference test

C) A Kruskal-Wallis test

A question!

I am interested in testing whether there is a significant difference between the population sizes of fish in 30 low salinity lakes and 30 high salinity lakes:

1. Which statistical test should I use?

   A) A t-test
   B) A One-Way ANOVA
   C) A Chi-square test
   D) A Two-Way ANOVA

A question!

I am interested in testing whether there is a significant difference between the population sizes of fish in 30 low salinity lakes and 30 high salinity lakes:

2. Which is the test statistic for the test?

A) t

B) F

C) r

D) χ^2

A question!

I am interested in testing whether there is a significant difference between the population sizes of fish in 30 low salinity lakes and 30 high salinity lakes. In fact, I actually sampled 10 large, 10 medium, and 10 small lakes in each of the high and low salinity lakes. I want to explore whether there are differences in population size based on lake salinity and lake size.

3. Now which statistical test should I use?

A) A t-test
B) A One-Way ANOVA
C) A Chi-square test
D) A Two-Way ANOVA

A question!

I am interested in testing whether there is a significant difference between the population sizes of fish in 30 low salinity lakes and 30 high salinity lakes. In fact, I actually sampled 10 large, 10 medium, and 10 small lakes in each of the high and low salinity lakes. I want to explore whether there are differences in population size based on lake salinity and lake size.

4. Now Which is the test statistic for the test?

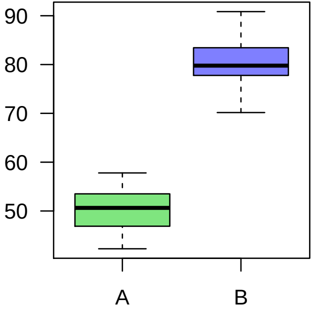
A) t
B) F
C) r
D) χ^2

Comparing differences - two samples

Two samples:

- what test?

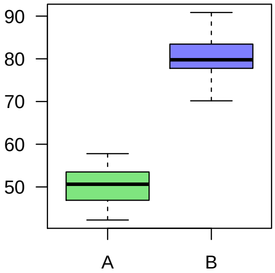
?



Comparing differences - two samples

Two samples:

- the t-test?
- test whether group means differ significantly
- H_0: there is no significant difference between the means
- H_1: there is a significant difference between the means



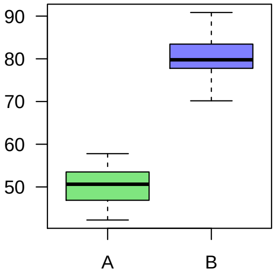
Comparing differences - two smaples

Two samples:

- the t-test?
- test whether group means differ significantly
- H_0: there is no significant difference between the means
- H_1: there is a significant difference between the means

Significance based on:

- t-statistic: $t = \frac{|\bar{x}_a - \bar{x}_b|}{\sqrt{\frac{s_a^2}{n_a} + \frac{s_b^2}{n_b}}}$
- degrees of freedom
- p-value



Comparing differences - more than two samples

What about if there are more than 2 samples?

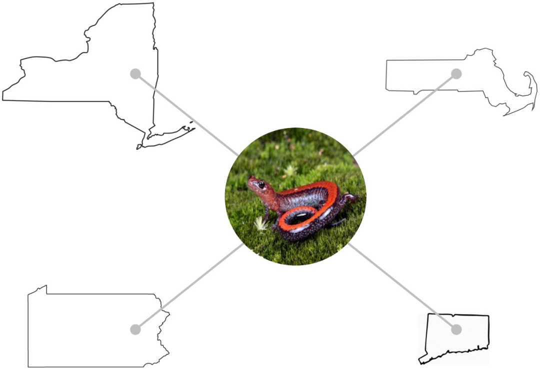
- can you think of any examples?

?

Comparing multiple groups - examples

Regional differences in salamander abundance:

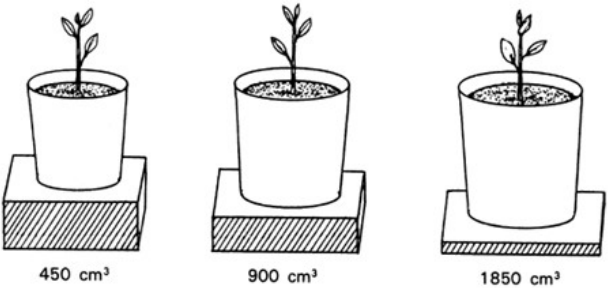
- comparing multiple populations
- quantify the differences between populations



Comparing multiple groups - examples

Plant growth related to available resources (pot size):

- comparing multiple treatments
- quantify the effects of resource availability



Comparing multiple groups - examples

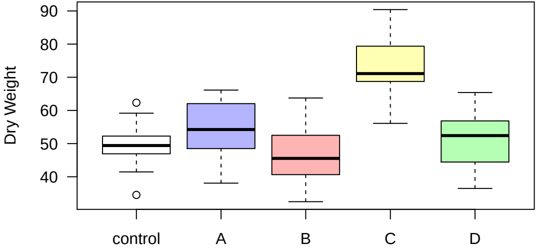
Plants productivity (dry mass in grams) related to fertilizer treatment

- do our treatments influence biomass production?
- is there a positive effect relative to a control?

Comparing multiple groups - examples

When there are more than 2 groups

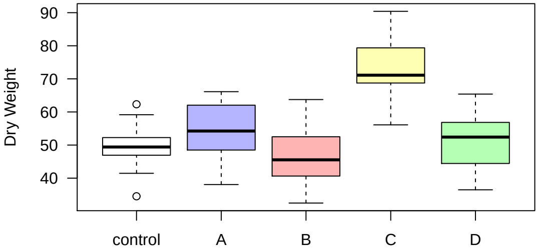
- t-test doesn't help
- need to do all possible pairs
- time consuming
- get spurious differences just by chance



Comparing multiple groups - ANOVA

Analysis of Variance (ANOVA):

- statistical test for testing for differences among >2 groups
- ANOVA and t-test are identical when there are 2 groups
- one factor/group/category (*One-way ANOVA*)



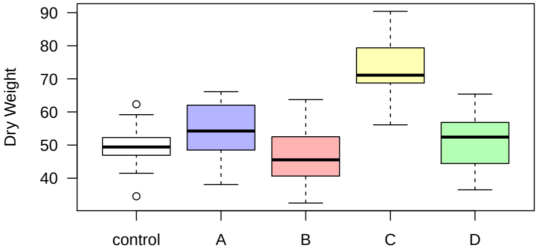
Comparing multiple groups - ANOVA

Analysis of Variance (ANOVA):

- statistical test for testing for differences among >2 groups
- ANOVA and t-test are identical when there are 2 groups
- one factor/group/category (*One-way ANOVA*)

Assumption:

- data are normally distributed



Comparing multiple groups - ANOVA

Analysis of Variance (ANOVA):

- statistical test for testing for differences among >2 groups
- ANOVA and t-test are identical when there are 2 groups
- one factor/group/category (*One-way ANOVA*)

Assumption:

- data are normally distributed

Hypotheses:

?

Comparing multiple groups - ANOVA

Analysis of Variance (ANOVA):

- statistical test for testing for differences among >2 groups
- ANOVA and t-test are identical when there are 2 groups
- one factor/group/category (*One-way ANOVA*)

Assumption:

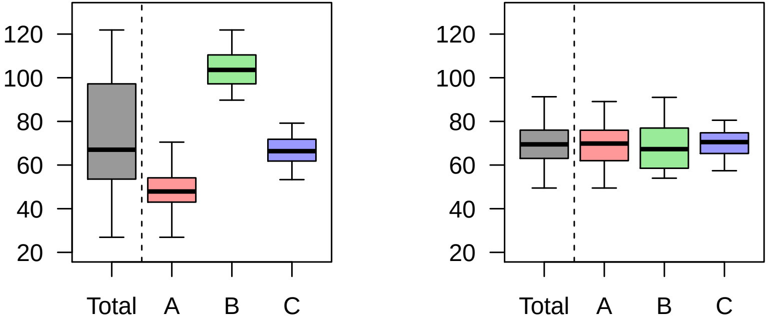
- data are normally distributed

Hypotheses:

- H_0: there are no significant differences between the means
  - all means are equal
- H_1: there are significant differences between the means
  - all means are not equal

ANOVA explained

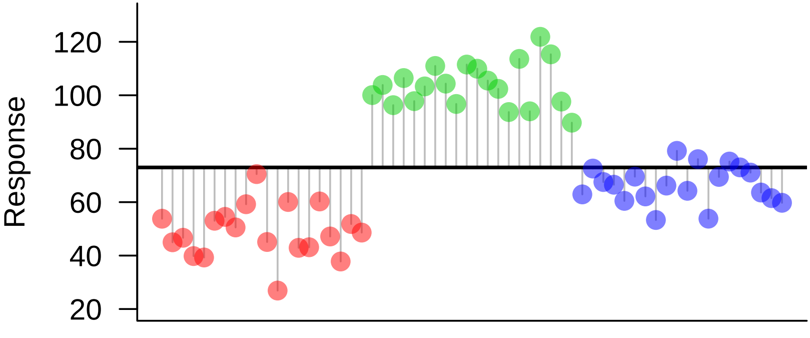
The ANOVA partitions the *total* variation into *within* sample variation with *between* sample variation to determine whether samples come from a single distribution or not.



ANOVA and the Sums of Squares

- *Total* sums of squares (SS_T)

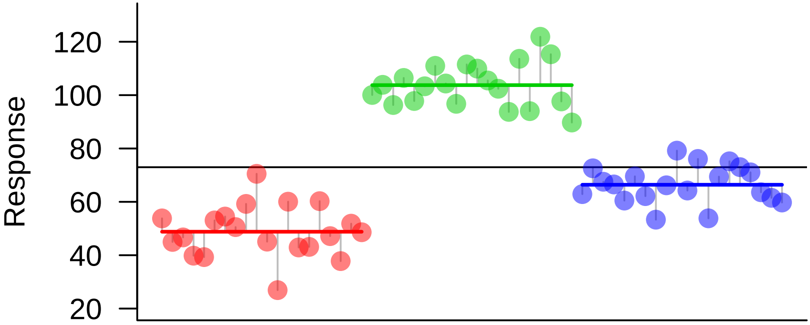
$$SS_T = \sum (x - \bar{x})^2$$



ANOVA and the Sums of Squares

- *Within-sample* sums of squares (SS_T)
- add up the within sample SS

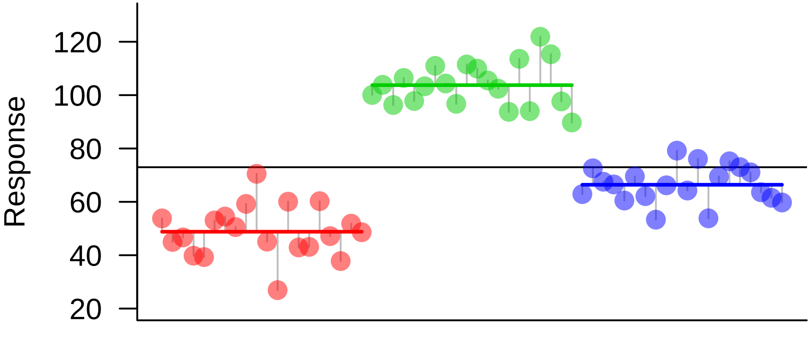
$$SS_W = \sum(x_1 - \bar{x}_1)^2 + \sum(x_2 - \bar{x}_2)^2 + \sum(x_3 - \bar{x}_3)^2$$



ANOVA and the Sums of Squares

- *Within-sample* sums of squares (SS_T)
- more generally (g is the number of groups)

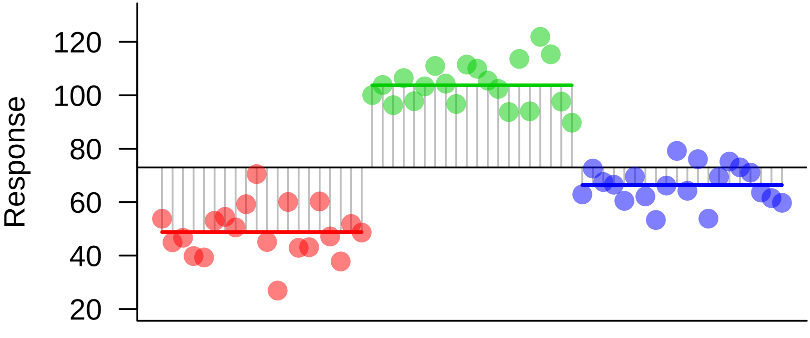
$$SS_W = \sum_g \sum_i (x_{ig} - \bar{x}_g)^2$$



ANOVA and the Sums of Squares

- *Between-sample* sums of squares (SS_T)
- add up the differences in the means

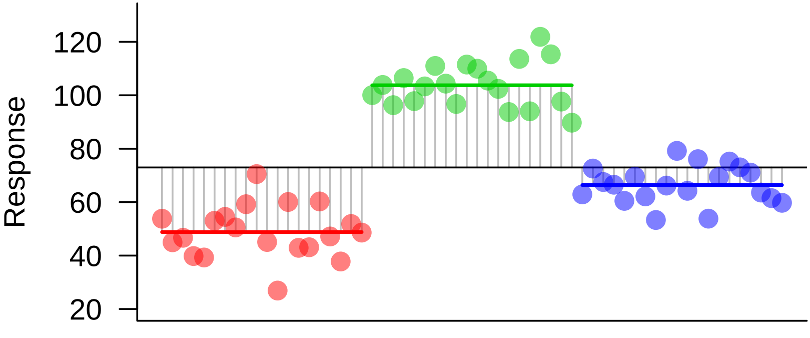
$$SS_B = n_1(\bar{x}_1 - \bar{x})^2 + n_2(\bar{x}_2 - \bar{x})^2 + n_3(\bar{x}_3 - \bar{x})^2$$



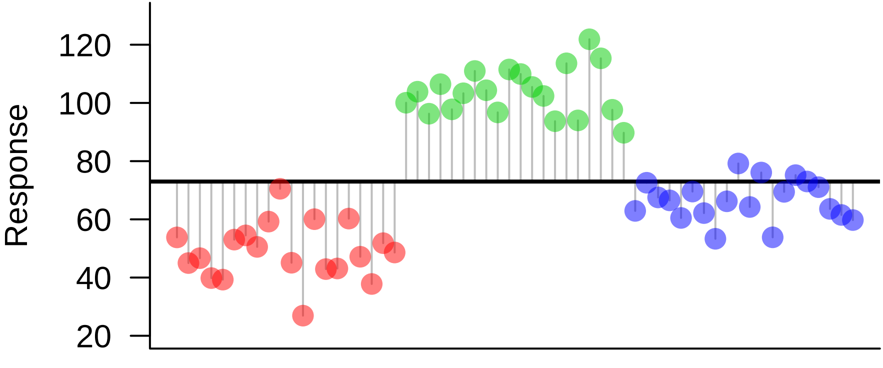
ANOVA and the Sums of Squares

- *Between-sample* sum of squares (SS_T)
- more generally (g is the number of groups)

$$SS_B = \sum_g n_g(\bar{x}_g - \bar{x})^2$$

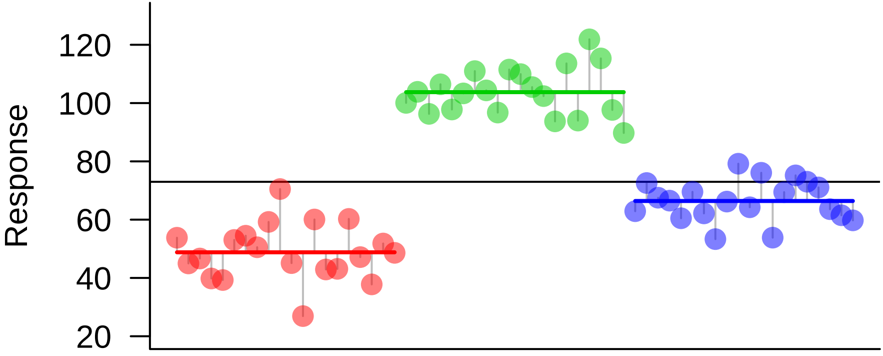


Total:



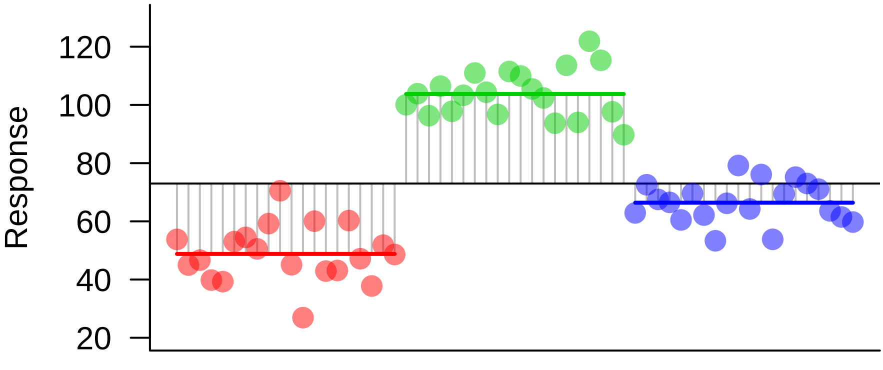
ANOVA and the Sums of Squares

Within group:



ANOVA and the Sums of Squares

Between group:

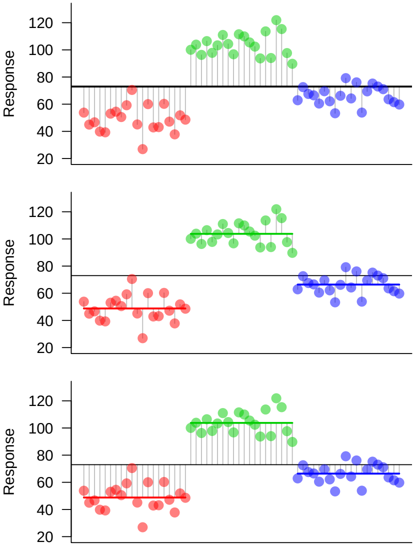


ANOVA and the Sums of Squares

$$SS_T = \sum (x - \bar{x})^2$$

$$SS_W = \sum_g \sum_i (x_{ig} - \bar{x}_g)^2$$

$$SS_B = \sum_g n_g (\bar{x}_g - \bar{x})^2$$



ANOVA degrees of freedom

If we define the following:

- n is the total sample size (number of observations)
- g is the number of groups/samples

ANOVA degrees of freedom

If we define the following:

- n is the total sample size (number of observations)
- g is the number of groups/samples

Then the degrees of freedom (df) are:

- Total: $df_T = n - 1$
- Within: $df_W = g - 1$
- Between: $df_B = n - g$

ANOVA the *mean square*

The mean square (MS) is the sum of squares divided by the degrees of freedom:

$$MS = SS/df$$

So:

- Total: $MS_T = SS_T/df_T$
- Within: $MS_W = SS_W/df_W$
- Between: $MS_B = SS_B/df_B$

ANOVA all the ingredients

	SS	df	MS
Total	$\sum (x - \bar{x})^2$	$n - 1$	SS_T / df_T
Within	$\sum_g \sum_i (x_{ig} - \bar{x}_j)^2$	$n - g$	SS_W / df_W
Between	$\sum_g n_g (\bar{x}_g - \bar{x})^2$	$g - 1$	SS_B / df_B

ANOVA the statistical test

ANOVA results are usually presented in an ANOVA table

Source of variation	SS	df	MS	F	p
Between	SS_B	df_B	MS_B		
Within	SS_W	df_W	MS_W		
Total	SS_T	df_T	–		

ANOVA the statistical test

ANOVA results are usually presented in an ANOVA table

Source of variation	SS	df	MS	F	p
Between	SS_B	df_B	MS_B		
Within	SS_W	df_W	MS_W		
Total	SS_T	df_T	–		

- F is the test statistic for the ANOVA

$$F = \frac{MS_B}{MS_W}$$

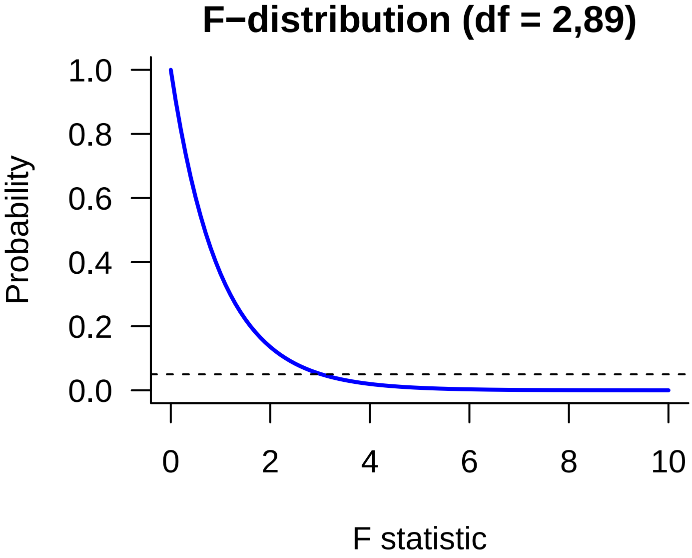
ANOVA the statistical test

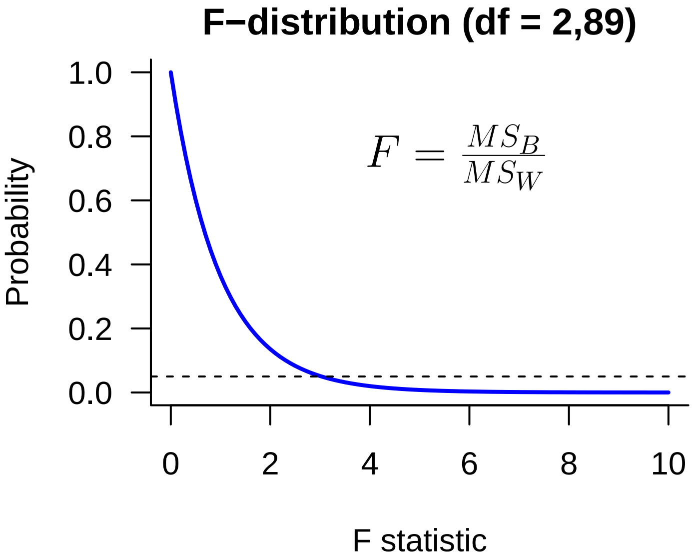
ANOVA results are usually presented in an ANOVA table

Source of variation	SS	df	MS	F	p
Between	SS_B	df_B	MS_B		
Within	SS_W	df_W	MS_W		
Total	SS_T	df_T	–		

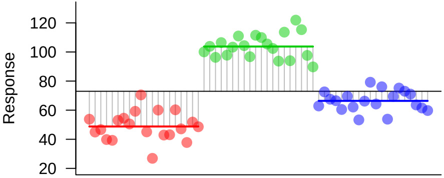
- p is the probability of observing the F statistic with a given degrees of freedom if the null hypothesis is true:
  - null hypothesis is 'no difference between the means'
  - based on the F-distribution

ANOVA the F distribution

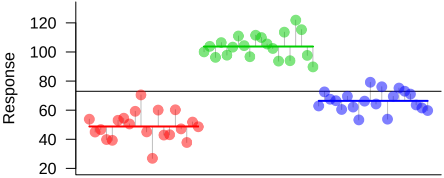



F−distribution (df = 2,89)
1.0
0.8
0.6
0.4
0.2
0.0
Probability
$F = \frac{MS_B}{MS_W}$
0
2
4
6
8
10
F statistic

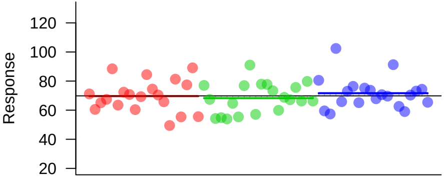
ANOVA and the Sums of Squares



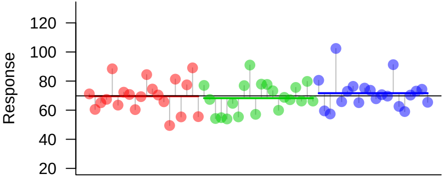
$$F = \frac{MS_B}{MS_W}$$



ANOVA and the Sums of Squares



$$F = \frac{MS_B}{MS_W}$$



ANOVA the p value

Hypotheses:

- H_0: there are no significant differences between the means
    - all means are equal
- H_1: there are significant differences between the means
    - all means are not equal

When do we reject or fail to reject the null hypothesis?

ANOVA the p value

Hypotheses:

- H_0: there are no significant differences between the means
  - all means are equal
- H_1: there are significant differences between the means
  - all means are not equal

When do we reject or fail to reject the null hypothesis?

- if F is large, then p is small
- if $p < 0.05$ we reject the null hypothesis
- if $p > 0.05$ we *fail to* reject the null hypothesis

Pairwise comparisons with ANOVA

The F statistic tells us whether there are differences, but *not* what the differences are:

- *Cannot* use t-tests to make pairwise comparisons
  - multiple t-tests will lead to significant results by chance

Pairwise comparisons with ANOVA

The F statistic tells us whether there are differences, but *not* what the differences are:

- Instead we conduct *Post-hoc* testing
  - Tukey Honest Significant Difference test (Tukey HSD)
  - accounts for multiple tests being conducted
  - calculation of a t-statistic
  - a pair, so degrees of freedom is 1
  - 5% critical value for $df = 1$ is 4.303
    - if $t > 4.303$ then $p < 0.05$

Pairwise comparisons with ANOVA

The F statistic tells us whether there are differences, but *not* what the differences are:

- Instead we conduct *Post-hoc* testing
  - Tukey Honest Significant Difference test (Tukey HSD)
  - accounts for multiple tests being conducted
  - calculation of a t-statistic
  - a pair, so degrees of freedom is 1
  - 5% critical value for $df = 1$ is 4.303
    - if $t > 4.303$ then $p < 0.05$

$$t_{a,b} = \frac{|\bar{x}_a - \bar{x}_b|}{\sqrt{\frac{MS_W\left(\frac{1}{n_a}+\frac{1}{n_b}\right)}{2}}}$$

Pairwise comparisons with ANOVA

The F statistic tells us whether there are differences, but *not* what the differences are:

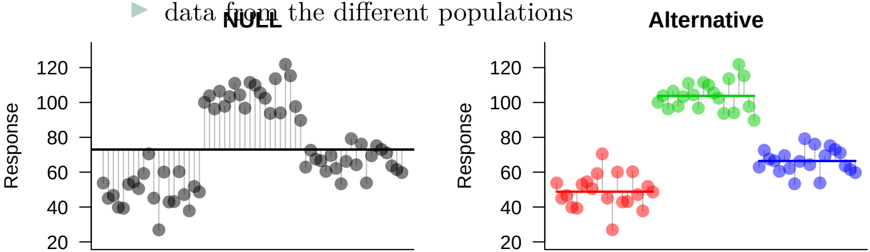
- Instead we conduct *Post-hoc* testing
  - Tukey Honest Significant Difference test (Tukey HSD)
  - accounts for multiple tests being conducted
  - calculation of a t-statistic
  - a pair, so degrees of freedom is 1
  - 5% critical value for $df = 1$ is 4.303
    - if $t > 4.303$ then $p < 0.05$

	A	B	C
A	-	$t_{A,B}$	$t_{A,C}$
B	-	-	$t_{B,C}$
C	-	-	-

ANOVA Recap

Comparing differences between >2 samples (groups) using ANOVA

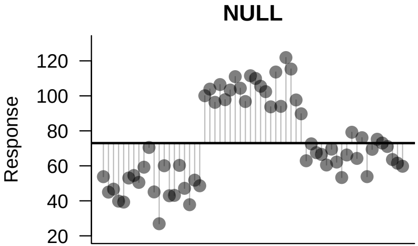
- null hypothesis:
  - no difference between the samples
  - data are from the same population
- alternative hypothesis:
  - sample means are different
  - data from the different populations

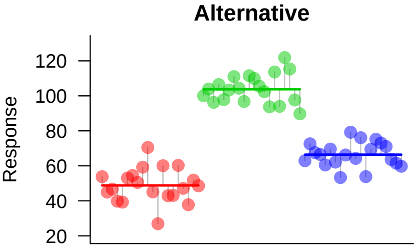


ANOVA Recap

Comparing differences between >2 groups using ANOVA

Source of variation	SS	df	MS	F	p
Between	SS_B	df_B	MS_B	$\frac{MS_B}{MS_W}$	
Within	SS_W	df_W	MS_W		
Total	SS_T	df_T	–		

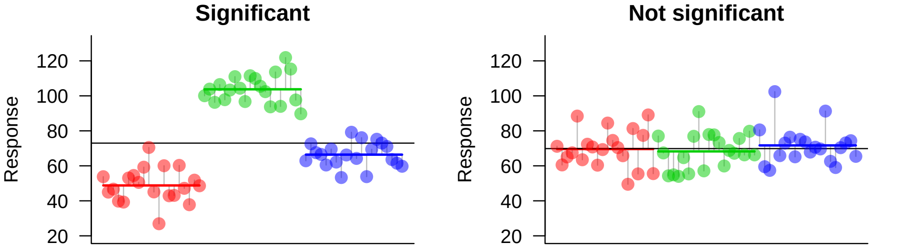




ANOVA Recap

Comparing differences between >2 groups using ANOVA

- Essentially comes down to:
  - a model with one mean *or* a model with a mean per group
  - which model best explains the data
  - which model significantly reduces the sums of squares



More than on factor with ANOVA

So far we have looked at multiple levels within a single factor

- factor: a single categorical predictor variable
- level: the categories within a factor

In some cases, we may be interested in >1 factor

- 2 factors: *two-way* ANOVA
- 3 factors: *three-way* ANOVA
- $\cdots$ multi-way ANOVA

Two-way ANOVA

Let's use a grazing example:

	Site	
Grazing Treatment	Top	Lower
Lo	9	7
Lo	11	6
Lo	6	5
Mid	14	14
Mid	17	17
Mid	19	15
Hi	28	44
Hi	31	38
Hi	32	37

Two-way ANOVA

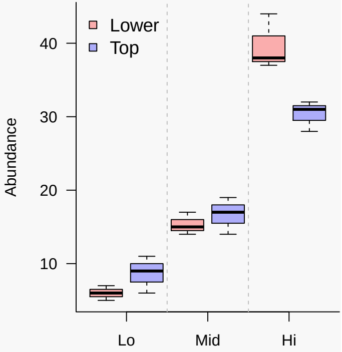
Lets use the example from the book (in `R` looks like this):

```
graze
   graze  Site Abundance
1     Lo   Top         9
2     Lo   Top        11
3     Lo   Top         6
4    Mid   Top        14
5    Mid   Top        17
6    Mid   Top        19
7     Hi   Top        28
8     Hi   Top        31
9     Hi   Top        32
10    Lo Lower         7
11    Lo Lower         6
12    Lo Lower         5
13   Mid Lower        14
14   Mid Lower        17
15   Mid Lower        15
16    Hi Lower        44
17    Hi Lower        38
18    Hi Lower        37
```

Two-way ANOVA

Lets use the example from the book (in R looks like this):

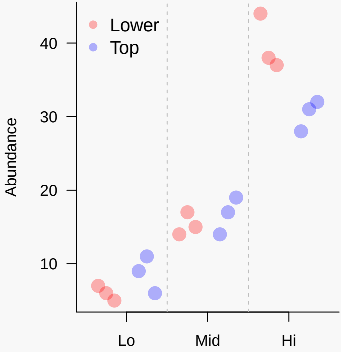
```
graze
   graze  Site Abundance
1     Lo   Top         9
2     Lo   Top        11
3     Lo   Top         6
4    Mid   Top        14
5    Mid   Top        17
6    Mid   Top        19
7     Hi   Top        28
8     Hi   Top        31
9     Hi   Top        32
10    Lo Lower         7
11    Lo Lower         6
12    Lo Lower         5
13   Mid Lower        14
14   Mid Lower        17
15   Mid Lower        15
16    Hi Lower        44
17    Hi Lower        38
18    Hi Lower        37
```



Two-way ANOVA

Lets use the example from the book (in `R` looks like this):

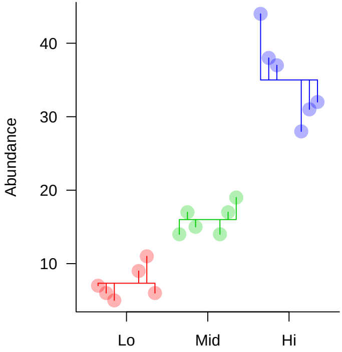
```
graze
   graze  Site Abundance
1     Lo   Top         9
2     Lo   Top        11
3     Lo   Top         6
4    Mid   Top        14
5    Mid   Top        17
6    Mid   Top        19
7     Hi   Top        28
8     Hi   Top        31
9     Hi   Top        32
10    Lo Lower         7
11    Lo Lower         6
12    Lo Lower         5
13   Mid Lower        14
14   Mid Lower        17
15   Mid Lower        15
16    Hi Lower        44
17    Hi Lower        38
18    Hi Lower        37
```



Conducting the ANOVA

Step one:

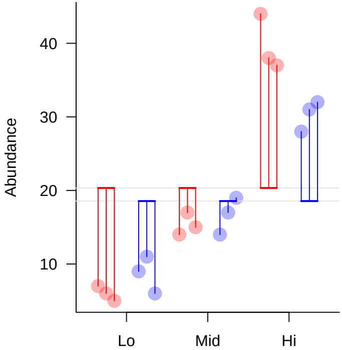
- SS for each factor
  - graze
  - site
- $SS_{graze} = \sum(x_{i,graze} - \bar{x}_{graze})^2$
- Ignore site grouping



Conducting the ANOVA

Step one:

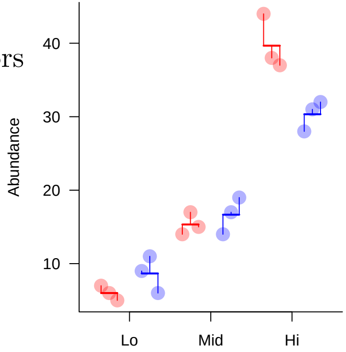
- SS for each factor
  - graze
  - site
- $SS_{site} = \sum(x_{i,site} - \bar{x}_{site})^2$
- Ignore graze grouping



Conducting the ANOVA

Step two:

- SS for each combinations of factors
- Treat all groupings as unique
- $SS_{within} = (x_{i,g} - \bar{x}_g)^2$



Conducting the ANOVA

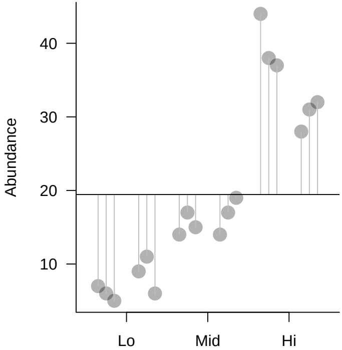
Step three:

- Sums of squares of both factors
- $SS_{both} = SS_{total} - SS_{graze} - SSsite - SS_{within}$

Conducting the ANOVA

Step four:

- Total sums of squares
- $SS_{total} = \sum(x_i - \bar{x})^2$
- the *null* model
- Ignore all group structure



Conducting the ANOVA - sums of squares

	SS	df	MS	F	p
Graze	SS_{graze}				
Site	SS_{site}				
Both factors(interaction)	SS_{both}				
Within group	SS_{within}				
Total	SS_{total}				

Degrees of freedom

In general:

- Factor 1 (F1): number of levels - 1
- Factor 2 (F2): number of levels - 1
- Within: n - (levels in F1 $\times$ levels in F2)
- Total: n - 1

Degrees of freedom

In general:

- Factor 1 (F1): number of levels - 1
- Factor 2 (F2): number of levels - 1
- Within: n - (levels in F1 $\times$ levels in F2)
- Total: n - 1

Grazing example:

- Graze: $3 - 1 = 2$
- Site: $2 - 1 = 1$
- Within: $18 - (3 \times 2) = 12$
- Total: $18 - 1 = 17$

Degrees of freedom

In general:

- Factor 1 (F1): number of levels - 1
- Factor 2 (F2): number of levels - 1
- Within: n - (levels in F1 $\times$ levels in F2)
- Total: n - 1

	SS	df	MS	F	p
Graze	SS_{graze}	df_{graze}			
Site	SS_{site}	df_{site}			
Both factors(interaction)	SS_{both}	df_{both}			
Within group	SS_{within}	df_{within}			
Total	SS_{total}	df_{total}			

Mean squares

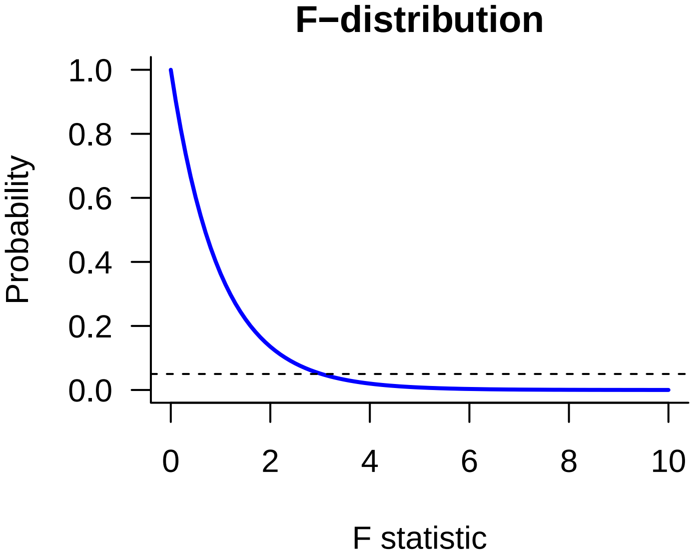
- the mean squares are calculated by dividing the sums of squares by the degrees of freedom for each element

	SS	df	MS	F	p
Graze	SS_{graze}	df_{graze}	$MS_{graze} = \frac{SS_{graze}}{df_{graze}}$		
Site	SS_{site}	df_{site}	$MS_{site} = \frac{SS_{site}}{df_{site}}$		
Both factors	SS_{both}	df_{both}	$MS_{both} = \frac{SS_{both}}{df_{both}}$		
Within group	SS_{within}	df_{within}	$MS_{within} = \frac{SS_{within}}{df_{within}}$		
Total	SS_{total}	df_{total}			

F statistic

- the F-statistic is calculated by taking the element of interest divided by the within group MS (the *error* term)

	SS	df	MS	F	p
Graze	SS_{graze}	df_{graze}	$MS_{graze} = \frac{SS_{graze}}{df_{graze}}$	$\frac{MS_{graze}}{MS_{within}}$	
Site	SS_{site}	df_{site}	$MS_{site} = \frac{SS_{site}}{df_{site}}$	$\frac{MS_{site}}{MS_{within}}$	
Both factors	SS_{both}	df_{both}	$MS_{both} = \frac{SS_{both}}{df_{both}}$	$\frac{MS_{both}}{MS_{within}}$	
Within group	SS_{within}	df_{within}	$MS_{within} = \frac{SS_{within}}{df_{within}}$		
Total	SS_{total}	df_{total}			


F−distribution
Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
2
4
6
8
10
F statistic

ANOVA in practice - Excel

```
graze
   graze  Site Abundance
1     Lo   Top         9
2     Lo   Top        11
3     Lo   Top         6
4    Mid   Top        14
5    Mid   Top        17
6    Mid   Top        19
7     Hi   Top        28
8     Hi   Top        31
9     Hi   Top        32
10    Lo Lower         7
11    Lo Lower         6
12    Lo Lower         5
13   Mid Lower        14
14   Mid Lower        17
15   Mid Lower        15
16    Hi Lower        44
17    Hi Lower        38
18    Hi Lower        37
```

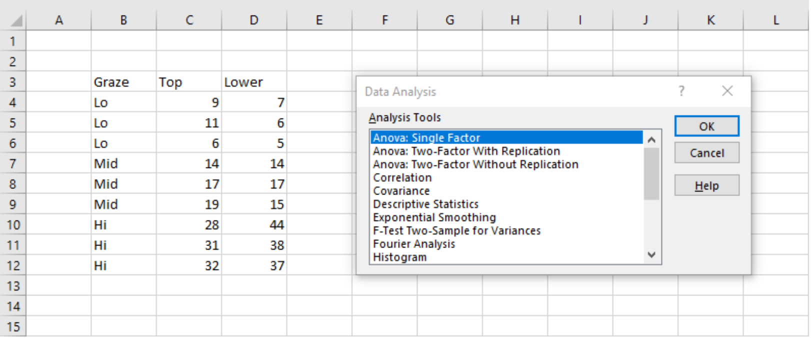
One-way ANOVA in practice - Excel

- Format data for specific test

Graze	Top	Lower
Lo	9	7
Lo	11	6
Lo	6	5
Mid	14	14
Mid	17	17
Mid	19	15
Hi	28	44
Hi	31	38
Hi	32	37

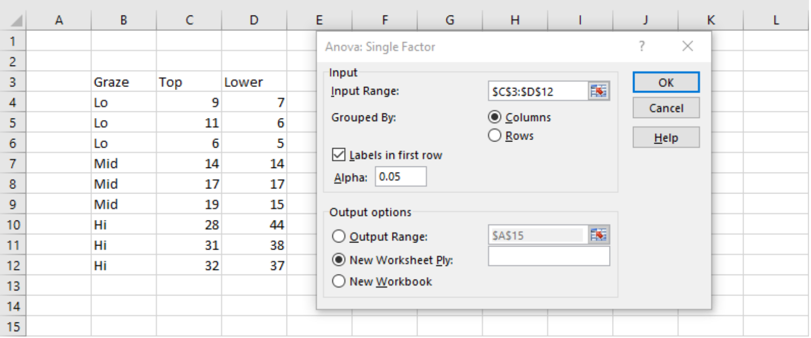
One-way ANOVA in practice - Excel

- Choose test from the *Analysis Toolpack*



One-way ANOVA in practice - Excel

- Select appropriate settings



One-way ANOVA in practice - Excel

- Interpret the output

	A	B	C	D	E	F	G
1	Anova: Single Factor						
2							
3	SUMMARY						
4	*Groups*	*Count*	*Sum*	*Average*	*Variance*		
5	Top	9	167	18.56	94.28		
6	Lower	9	183	20.33	231.00		
7							
8							
9	ANOVA						
10	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
11	Between Groups	14.22	1	14.22	0.09	0.77	4.49
12	Within Groups	2602.22	16	162.64			
13							
14	Total	2616.44	17				
15							

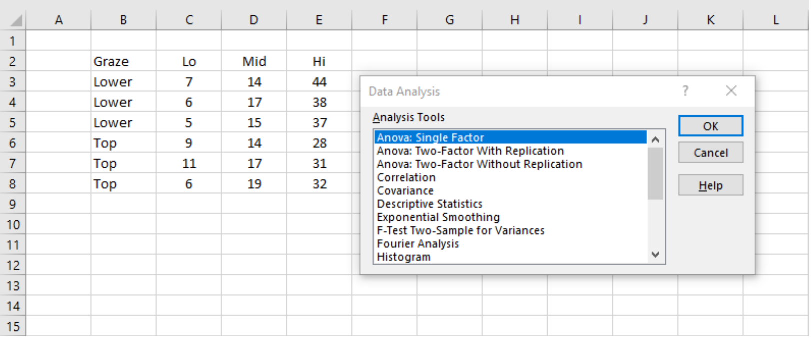
One-way ANOVA in practice - Excel

- Format data for specific test

Graze	Lo	Mid	Hi
Lower	7	14	44
Lower	6	17	38
Lower	5	15	37
Top	9	14	28
Top	11	17	31
Top	6	19	32

One-wayANOVA in practice - Excel

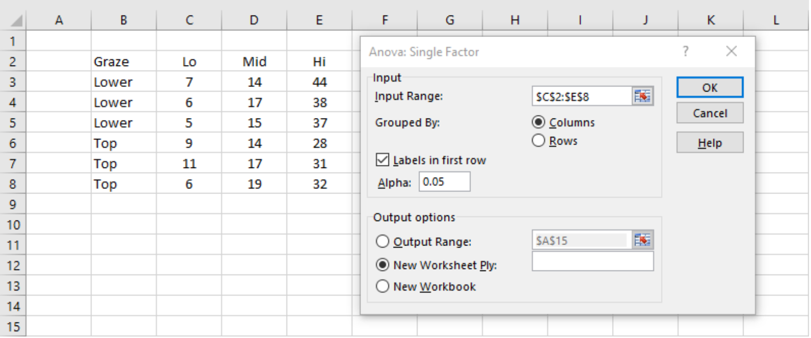
- Choose test from the *Analysis Toolpack*



	A	B	C	D	E
1					
2		Graze	Lo	Mid	Hi
3		Lower	7	14	44
4		Lower	6	17	38
5		Lower	5	15	37
6		Top	9	14	28
7		Top	11	17	31
8		Top	6	19	32

One-way ANOVA in practice - Excel

- Select appropriate settings



One-way ANOVA in practice - Excel

- Interpret the output

	A	B	C	D	E	F	G
1	Anova: Single Factor						
2							
3	SUMMARY						
4	*Groups*	*Count*	*Sum*	*Average*	*Variance*		
5	Lo	6	44	7.33	5.07		
6	Mid	6	96	16	4		
7	Hi	6	210	35	33.6		
8							
9	ANOVA						
10	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
11	Between Groups	2403.11	2	1201.56	84.48	0.000000007	3.68
12	Within Groups	213.33	15	14.22			
13							
14	Total	2616.44	17				
15							

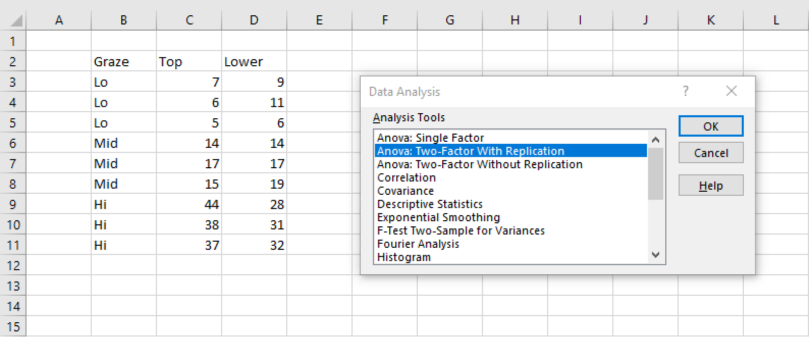
Two-way ANOVA in practice - Excel

- Format data for specific test

	A	B	C	D	E	F	G	H	I	J	K	L
1												
2												
3		Graze	Top	Lower								
4		Lo	9	7								
5		Lo	11	6								
6		Lo	6	5								
7		Mid	14	14								
8		Mid	17	17								
9		Mid	19	15								
10		Hi	28	44								
11		Hi	31	38								
12		Hi	32	37								
13												
14												
15												

Two-wayANOVA in practice - Excel

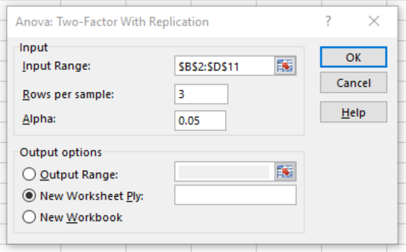
- Choose test from the *Analysis Toolpack*



Two-way ANOVA in practice - Excel

- Select appropriate settings

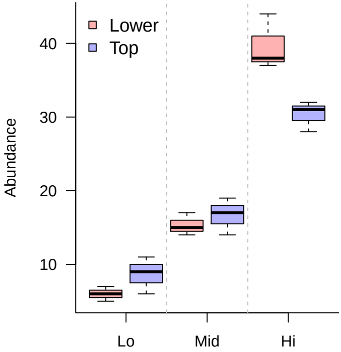
	A	B	C	D
1				
2		Graze	Top	Lower
3		Lo	7	9
4		Lo	6	11
5		Lo	5	6
6		Mid	14	14
7		Mid	17	17
8		Mid	15	19
9		Hi	44	28
10		Hi	38	31
11		Hi	37	32



Two-way ANOVA in practice - Excel

- Interpret the output

	A	B	C	D	E	F	G	H
1	**Anova: Two-Factor With Replication**							
2								
3	SUMMARY	Top	Lower	Total				
4	*Lo*							
5	Count	3	3	6				
6	Sum	18	26	44				
7	Average	6.00	8.67	7.33				
8	Variance	1.00	6.33	5.07				
9								
10	*Mid*							
11	Count	3	3	6				
12	Sum	46	50	96				
13	Average	15.33	16.67	16.00				
14	Variance	2.33	6.33	4.00				
15								
16	*Hi*							
17	Count	3	3	6				
18	Sum	119	91	210				
19	Average	39.67	30.33	35.00				
20	Variance	14.33	4.33	33.60				
21								
22	*Total*							
23	Count	9	9					
24	Sum	183	167					
25	Average	20.33	18.56					
26	Variance	231.00	94.28					
27								
28	ANOVA							
29	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*	
30	Sample	2403.11	2	1201.56	207.96	0.0000000005	3.89	
31	Columns	14.22	1	14.22	2.46	0.1426439913	4.75	
32	Interaction	129.78	2	64.89	11.23	0.0017827051	3.89	
33	Within	69.33	12	5.78				
34								
35	Total	2616.44	17					
36								



ANOVA in practice - R

- Read in the data as a data frame

```
graze
   graze  Site Abundance
1     Lo   Top         9
2     Lo   Top        11
3     Lo   Top         6
4    Mid   Top        14
5    Mid   Top        17
6    Mid   Top        19
7     Hi   Top        28
8     Hi   Top        31
9     Hi   Top        32
10    Lo Lower         7
11    Lo Lower         6
12    Lo Lower         5
13   Mid Lower        14
14   Mid Lower        17
15   Mid Lower        15
16    Hi Lower        44
17    Hi Lower        38
18    Hi Lower        37
```

Any ANOVA in practice - R

- Conduct *any* test using formula syntax

```
oneway.site <- aov(Abundance ~ Site, data = graze)
summary(oneway.site)
            Df Sum Sq Mean Sq F value Pr(>F)
Site         1   14.2   14.22   0.087  0.771
Residuals   16 2602.2  162.64
```

Any ANOVA in practice - R

- Conduct *any* test using formula syntax

```
oneway.site <- aov(Abundance ~ Site, data = graze)
summary(oneway.site)
            Df Sum Sq Mean Sq F value Pr(>F)
Site         1   14.2   14.22   0.087  0.771
Residuals   16 2602.2  162.64
```

	A	B	C	D	E	F	G
1	Anova: Single Factor						
2							
3	SUMMARY						
4	*Groups*	*Count*	*Sum*	*Average*	*Variance*		
5	Top	9	167	18.56	94.28		
6	Lower	9	183	20.33	231.00		
7							
8							
9	ANOVA						
10	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
11	Between Groups	14.22	1	14.22	0.09	0.77	4.49
12	Within Groups	2602.22	16	162.64			
13							
14	Total	2616.44	17				
15							

Any ANOVA in practice - R

- Conduct *any* test using formula syntax

```
oneway.graze <- aov(Abundance ~ graze, data = graze)
summary(oneway.graze)
            Df Sum Sq Mean Sq F value   Pr(>F)
graze        2 2403.1  1201.6   84.48 6.84e-09 ***
Residuals   15  213.3    14.2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Any ANOVA in practice - R

- Conduct *any* test using formula syntax

```
oneway.graze <- aov(Abundance ~ graze, data = graze)
summary(oneway.graze)
            Df Sum Sq Mean Sq F value   Pr(>F)    
graze        2 2403.1  1201.6   84.48 6.84e-09 ***
Residuals   15  213.3    14.2                     
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

	A	B	C	D	E	F	G	H	I	J	K	L
1	Anova: Single Factor											
2												
3	SUMMARY											
4	*Groups*	*Count*	*Sum*	*Average*	*Variance*							
5	Lo	6	44	7.33	5.07							
6	Mid	6	96	16	4							
7	Hi	6	210	35	33.6							
8												
9	ANOVA											
10	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*					
11	Between Groups	2403.11	2	1201.56	84.48	0.000000007	3.68					
12	Within Groups	213.33	15	14.22								
13												
14	Total	2616.44	17									
15												

Any ANOVA in practice - R

- Conduct *any* test using formula syntax

```
twoway.additive <- aov(Abundance ~ Site + graze, data = graze)
summary(twoway.additive)
            Df Sum Sq Mean Sq F value   Pr(>F)
Site         1   14.2    14.2    1.00    0.334
graze        2 2403.1  1201.6   84.48 1.54e-08 ***
Residuals   14  199.1    14.2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Any ANOVA in practice - R

- Conduct *any* test using formula syntax

```
twoway.additive <- aov(Abundance ~ Site + graze, data = graze)
summary(twoway.additive)
            Df Sum Sq Mean Sq F value   Pr(>F)
Site         1   14.2    14.2    1.00    0.334
graze        2 2403.1  1201.6   84.48 1.54e-08 ***
Residuals   14  199.1    14.2
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

- Excel only *fits* the interaction model

Any ANOVA in practice - R

- Conduct *any* test using formula syntax

```
twoway.interaction <- aov(Abundance ~ Site * graze, data = graze)
summary(twoway.interaction)
            Df Sum Sq Mean Sq F value   Pr(>F)
Site         1   14.2    14.2   2.462  0.14264
graze        2 2403.1  1201.6 207.962 4.86e-10 ***
Site:graze   2  129.8    64.9  11.231  0.00178 **
Residuals   12   69.3     5.8
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Any ANOVA in practice - R

- Conduct *any* test using formula syntax

```
twoway.interaction <- aov(Abundance ~ Site * graze, data = graze)
summary(twoway.interaction)
             Df Sum Sq Mean Sq F value   Pr(>F)
Site          1   14.2    14.2   2.462  0.14264
graze         2 2403.1  1201.6 207.962 4.86e-10 ***
Site:graze    2  129.8    64.9  11.231  0.00178 **
Residuals    12   69.3     5.8
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

	A	B	C	D	E	F	G	H	I	J
1	*Total*									
2	Count	9	9							
3	Sum	183	167							
4	Average	20.33	18.56							
5	Variance	231.00	94.28							
6										
7	ANOVA									
8	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*			
9	Sample	2403.11	2	1201.56	207.96	0.0000000005	3.89			
10	Columns	14.22	1	14.22	2.46	0.1426439913	4.75			
11	Interaction	129.78	2	64.89	11.23	0.0017827051	3.89			
12	Within	69.33	12	5.78						
13										
14	Total	2616.44	17							
15										

Group Exercise - *salamANOVA*

We will conduct three analyses using the *salamANOVA*. We are interested in whether salamander snout-to-vent length (SVL) varies by sex and/or site. The data look like this:

```
str(sals)
'data.frame':   48 obs. of  3 variables:
 $ Site: Factor w/ 4 levels "P1A","P1B","P2A",..: 1 1 1 1 1 1 2 2 2 2 ...
 $ Sex : Factor w/ 2 levels "F","M": 1 1 1 1 1 1 1 1 1 1 ...
 $ SVL : int  36 42 42 41 44 40 35 39 38 44 ...
```

- `Site`: there are four sites (`P1A`, `P1B`, `P2A`, `P2B`)
- `Sex`: `M` (male) and `F` (female)
- `SVL`: the snout-to-vent length in mm

Group Exercise - *salamANOVA*

Analysis 1: Does SVL vary by sex?

- What is the null hypothesis?
- Make a plot to visualize the hypothesis.
- What statistical test will you use to test H_0?
- What is the:
  - test statistic for this particulat test (e.g., t, F, etc)
  - degrees of freedom (calculate this)
  - significance level
- Conduct the analysis:
  - what is the value of the test statistic
  - what the p-value
- Write a short paragraph reporting the conclusion, use values from the statistical test to suppo, supported by the results from the test.

Group Exercise - *salamANOVA*

Analysis 2: Does SVL vary by site?

- What is the null hypothesis?
- Make a plot to visualize the hypothesis.
- What statistical test will you use to test H_0?
- What is the:
  - test statistic for this particulat test (e.g., t, F, etc)
  - degrees of freedom (calculate this)
  - significance level
- Conduct the analysis:
  - what is the value of the test statistic
  - what the p-value
- Write a short paragraph reporting the conclusion, use values from the statistical test to suppo, supported by the results from the test.

Group Exercise - *salamANOVA*

Analysis 3: Does SVL vary by sex and/or site?

- What is the null hypothesis?
- Make a plot to visualize the hypothesis.
- What statistical test will you use to test H_0?
- What is the:
  - test statistic for this particulat test (e.g., t, F, etc)
  - degrees of freedom (calculate this)
  - significance level
- Conduct the analysis:
  - what is the value of the test statistic
  - what the p-value
- Write a short paragraph reporting the conclusion, use values from the statistical test to suppo, supported by the results from the test.

Group Exercise - *salamANOVA*

Assignment: Statistical analysis of variation in salamnder SVL.

- Write a report with four sections:
  1. Analysis 1
  2. Analysis 2
  3. Analysis 3
  4. Reflection: how does analysis 3 compare to analyses 1 and 2?
- Sections 1 to 3 sould report on each of the prompts in the previous slides.
- Section 4 is an opportunity to demonstrate your undertanding of the material covered over the previous weeks.
- Assignment due: 11.55pm Tuesday November $20^t h$