ANOVA

Comparing the means of more than two groups

Analysis of variance (ANOVA)

Like a *t*-test, but can compare more than two groups

Asks whether any of two or more means is different from any other.

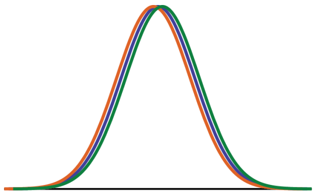
In other words, is the variance among groups greater than 0?

Null hypothesis for simple ANOVA

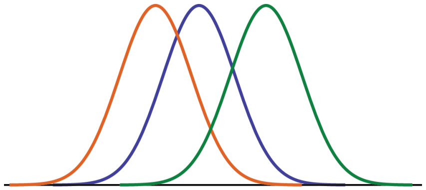
H_0 : Variance among groups = 0

OR

$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \ldots \mu_k$



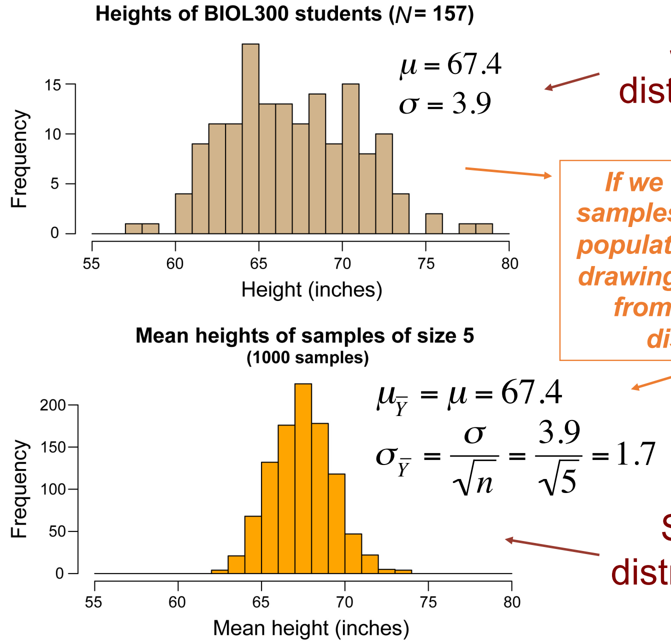
H_0: all populations have equal means



H_A: at least one population mean is different.

ANOVA's v. t-tests

An ANOVA with 2 groups is mathematically equivalent to a two-tailed 2-sample t-test.



$$\mu_{\bar{Y}} = \mu = 67.4$$

$$\sigma_{\bar{Y}} = \frac{\sigma}{\sqrt{n}} = \frac{3.9}{\sqrt{5}} = 1.7$$

Sampling distribution of Y

If we draw multiple samples from the same population, we are also drawing sample means from an expected distribution.

Sampling distribution of $\bar{\text{Y}}$

Under the null hypothesis, the sample mean of each group should vary because of sampling error.

The standard deviation of sample means, when the true mean is constant, is the standard error:

$$\sigma_{\bar{x}} = \frac{\sigma_x}{\sqrt{n}}$$

Squaring the standard error, the variance among groups due to sampling error should be:

$$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n}$$

In ANOVA, we work with variances rather than standard deviations.

$$\sigma_{\bar{x}}^2 = \frac{\sigma_x^2}{n} + Variance\left[\mu_i\right]$$

If the null hypothesis is <u>not</u> true, the variance among groups should be equal to the variance due to sampling error plus the real variance among population means.

With ANOVA, we test whether the variance among true group means is greater than zero.

We do this by asking whether the observed variance among groups is greater than expected by chance (assuming the null is true):

$$\sigma_{\bar{x}}^2 > \frac{\sigma_x^2}{n}?$$

$$\sigma_{\bar{x}}^2 > \frac{\sigma_x^2}{n}?$$

$$n\sigma_{\bar{x}}^2 > \sigma_x^2?$$

<u>Population parameters</u>		<u>Estimates from sample</u>
$n\sigma_{\bar{x}}^2$	is estimated by the "Mean Squares Group"	MS_{group}
σ_x^2	is the variance within groups, estimated by the "Mean Squares Error"	MS_{error}

Mean squares group

Abbreviation: MS_{group}

Estimates this parameter: $\sigma_{\bar{x}}^2 + n(Variance[\mu_i])$

Formula: $MS_{groups} = \frac{SS_{groups}}{df_{groups}}$

Mean squares group

$$SS_{group} = \sum n_i \left(\bar{X}_i - \bar{X}\right)^2$$

$\bar{X}_i$ is the mean of group i, and

$\bar{X} = \sum_i \sum_j X_{ij} / N$ is the overall mean.

$df_{group} = k-1$

$$MS_{groups} = \frac{SS_{groups}}{df_{groups}}$$

Mean squares error

Abbreviation: MS_{error}

Estimates this parameter: σ_x^2

Formula: $MS_{error} = \dfrac{SS_{error}}{df_{error}}$

Mean squares error

Error sum of squares =

$$SS_{error} = \sum df_i s_i^2 = \sum s_i^2 (n_i - 1)$$

Error degrees of freedom =

$$df_{error} = \sum df_i = \sum (n_i - 1) = N - k$$

$$MS_{error} = \frac{SS_{error}}{df_{error}} = \frac{\sum s_i^2 (n_i - 1)}{N - k}$$

MS_{error} is like the pooled variance in a 2-sample t-test:

$$s_p^2 = \frac{df_1 s_1^2 + df_2 s_2^2}{df_1 + df_2}$$

Test statistic: *F*

If H_0 is true, then $n\,\sigma_{\bar{x}}^2 = \sigma_x^2$

In other words: $F = \dfrac{n\,\sigma_{\bar{x}}^2}{\sigma_x^2} = 1$

But, the above refer to population parameters. We must estimate *F* from samples with: MS_{group} / MS_{error}

F if null hypothesis is false:

We test whether the F ratio is greater than one, as it would be if H_0 is false:

$$F = \frac{\sigma_{\bar{x}}^2 + n(Variance[\mu_i])}{\sigma_x^2} > 1$$

But we must take into account sampling error. Often, F calculated from data will be greater than one even when the null is true. Hence we must compare F to a null distribution.

ANOVA table

Source	SS	df	MS	F	*P*
Group					
Error					
Total					

An ANOVA table is a convenient way to keep track of the important calculations.

Scientific papers often report ANOVA results with ANOVA tables.

Example: Body temperature of squirrels in low, medium and hot environments

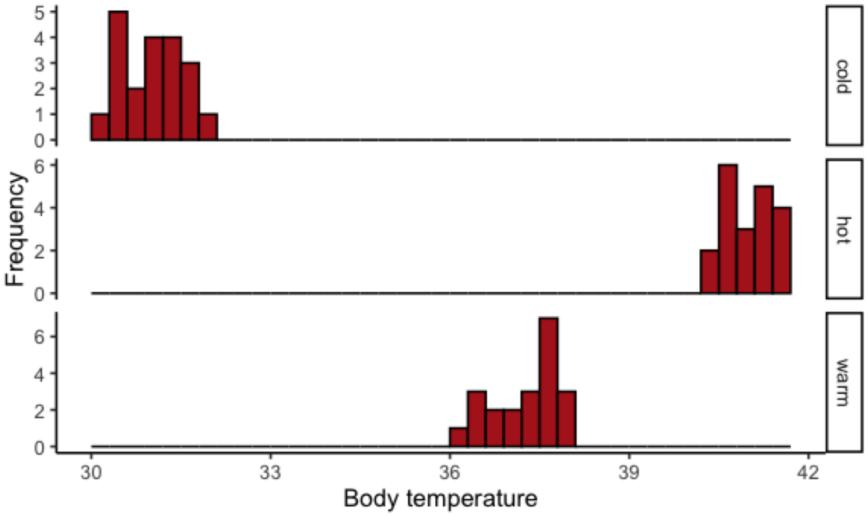
Wooden & Walsberg (2004) Body temperature and locomotor capacity in a heterothermic rodent. Journal of Experimental Biology 207:41-46.

Squirrel body temperature data (°C)

Cold: 30.4, 31.0, 31.2, 31.0, 31.5, 30.4, 30.6, 31.1, 31.3, 31.9, 31.4, 31.6, 31.5, 31.4, 30.3, 30.5, 30.3, 30.0, 30.8, 31.0

Warm: 36.3, 37.5, 36.9, 37.2, 37.5, 37.7, 37.5, 37.7, 38.0, 38.0, 37.6, 37.4, 37.9, 37.2, 36.3, 36.2, 36.4, 36.7, 36.8, 37.0, 37.7

Hot: 40.7, 40.6, 40.9, 41.1, 41.5, 40.8, 40.5, 41.0, 41.3, 41.5, 41.3, 41.2, 40.7, 40.3, 40.2, 41.3, 40.7, 41.6, 41.5, 40.5



Hypotheses

H_0: Mean body temperature is the same for all three groups of squirrels.

H_A: At least one of the three is different from the others.

Summary data

Group	$\overline{x}$	*s*	*n*
Cold	31.0	0.526	20
Warm	37.2	0.582	21
Hot	41.0	0.430	20

Total sample size: $N = \sum n = 20 + 21 = 20 = 61$

Error Mean square for squirrels

$$SS_{error} = \sum df_i s_i^2$$

$$= 19(0.526)^2 + 20(0.582)^2 + 19(0.430)^2$$

$$= 15.54$$

$$df_{error} = 19 + 20 + 19 = 58$$

$$MS_{error} = \frac{15.54}{58} = 0.27$$

ANOVA table – squirrel data

Source	**SS**	**df**	**MS**	**F**	***P***
Group					
Error	15.54	58	0.27		
Total					

Squirrel Mean Squares Group:

$$\bar{X} = \frac{20(30.96) + 21(37.21) + 20(40.96) +}{20 + 21 + 20} = 36.39$$

$$SS_{group} = \sum n_i \left(\bar{X}_i - \bar{X}\right)^2$$

$$SS_{group} = 20(30.96\text{-}36.39)^2 + 21(37.21\text{-}36.39)^2 + 20(40.96\text{-}36.39)^2$$
$$= 1021.66$$

Squirrel Mean Squares Group:

$$df_{group} = k - 1 = 3\text{-}1 = 2$$

$$MS_{groups} = \frac{SS_{groups}}{df_{grops}} = \frac{1021.66}{2} = 510.83$$

ANOVA table – squirrel data

Source	**SS**	**df**	**MS**	**F**	***P***
Group	1021.7	2	510.8		
Error	15.5	58	0.27		
Total	1037.2	60			

The test statistic for ANOVA is *F*

$$F = \frac{MS_{group}}{MS_{error}} = \frac{510.8}{0.27} = 1906.4$$

MS_{group} is always in the numerator;
MS_{error} is always in the denominator

Compare to $F_{\alpha(1),df\_group,df\_error}$

$$F_{0.05(1),2,58} = 3.15.$$

Since 1906 > 3.15, we know P < 0.05 and we can reject the null hypothesis.

The variance in sample group means is bigger than expected given the variance within sample groups.

Therefore, at least one of the groups has a population mean different from another group.

ANOVA table – squirrel data

Source	SS	df	MS	F	*P*
Group	1021.7	2	510.8	1906.4	<0.0001
Error	15.5	58	0.27		
Total	1037.2	60			

ANOVA in R

```
squirrelANOVA <- lm(bodyTemp ~ outsideTemp,
                 data = squirrelData)
anova(squirrelANOVA)
```

```
Analysis of Variance Table

Response: bodyTemp
            Df  Sum Sq Mean Sq  F value   Pr(>F)    
outsideTemp  2 1021.66  510.83   1906.4 < 2.2e-16 ***
Residuals   58   15.54    0.27
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Assumptions of ANOVA

(1) Random samples

(2) Normal distributions for each population

(3) Equal variances for all populations. (Homoscedasticity)

Kruskal-Wallis test

A non-parametric test similar to a single factor ANOVA

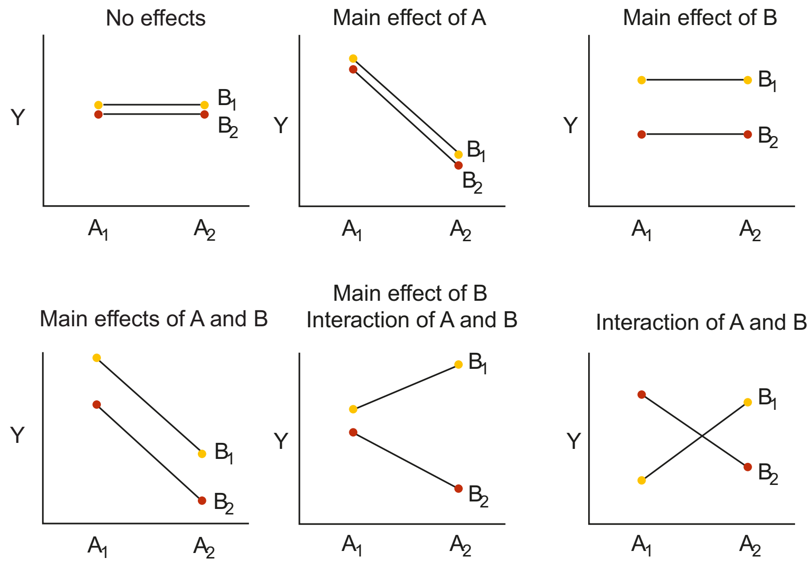
Uses the ranks of the data points

Multiple-factor ANOVA

A *factor* is a categorical variable.

ANOVAs can be generalized to look at more than one categorical variable at a time.

Not only can we ask whether each categorical variable affects a numerical variable, but also do they interact in affecting the numerical variable.



Fixed vs. random effects

1. *Fixed effects*: With fixed effects, the treatments are chosen by the experimenter. They are not a random subset of all possible treatments.
   (e.g., specific drug treatments, specific diets, season...)

2. *Random effects*: With random effects, the treatments are a random sample from all possible treatments.
   (e.g., family, location, ...)

For single-factor ANOVAs, there is no difference in the statistics for fixed or random effects.

2-factor ANOVA: Example

Plants taken from two Regions:
one south; one nozrth

Two watering treatments:
wet; dry

Work from UBC undergrad Rylan McCallum working with Haley Branchand Amy Angert

© Jason Hollinger, CC-BY

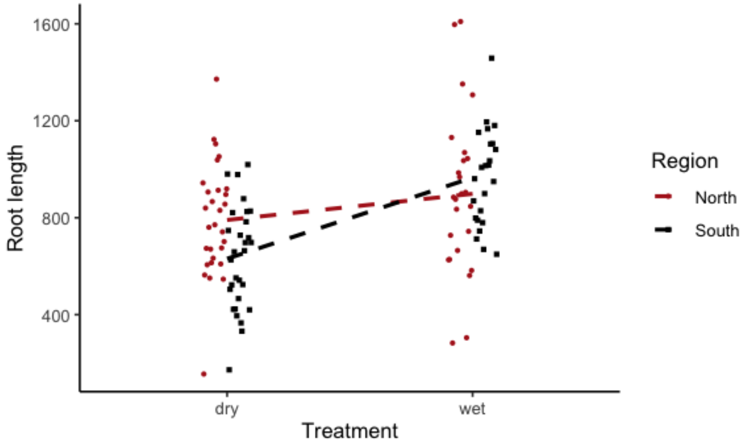
Testing multiple hypotheses

H_0: Mean root length is the same for all regions.

H_0: Mean root length is the same for all watering treatments.

H_0: There is no interaction between region and treatment for determining mean root length .

Testing multiple hypotheses



Heliconius ANOVA table

Source of variation	SS	df	F	P	
Region	57645	1	0.9	0.33	
Watering treatment	1300090	1	20.9	0.00001	
Region * treatment	341372	1	5.5	0.02	
Error	6401586	103			

Multiple comparisons

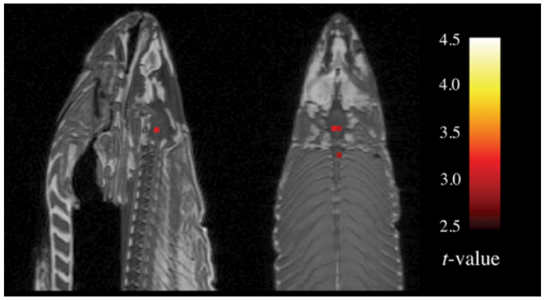
Probability of a Type I error in *N* tests =

$$1-(1-\alpha)^N$$

For 20 tests, the probability of at least one Type I error is ~65%.

MRI on dead salmon:

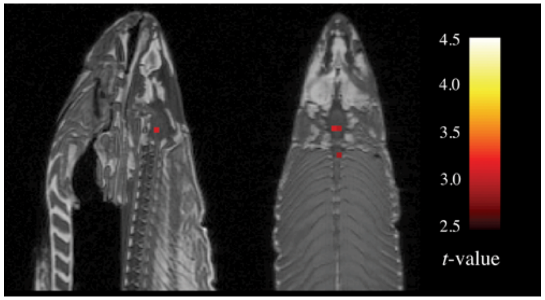
Dead fish shows emotional reactions to photos of human activities



“The salmon was shown a series of photographs depicting human individuals in social situations with a specified emotional valence. The salmon was asked to determine what emotion the individual in the photo must have been experiencing.”

MRI on dead salmon:

Dead fish shows emotional reactions to photos of human activities



“The salmon was shown a series of photographs depicting human individuals in social situations with a specified emotional valence. The salmon was asked to determine what emotion the individual in the photo must have been experiencing.”

(Each MRI had 130,000 voxels to compare.)

"Bonferroni correction" for multiple comparisons

Uses a smaller *a* value:

$$\alpha' = \frac{\alpha}{\text{number of tests}}$$

Which groups are different?

After finding evidence for differences among means with ANOVA, sometimes we want to know:

Which groups are different from which others?

One method for this: the Tukey-Kramer test

The Tukey-Kramer test

Done after finding variation among groups with single-factor ANOVA.

Compares all group means to all other group means

The wood-wide web

Trees (and other plants) are often connected by roots via mycorrhizae, which allow the exchange of resources.

Carbon transfer between birch and Douglas fir: Comparing effects of shading on fir

Net amount of carbon transferred from birch to fir

Shade treatment	Sample mean	Sample standard deviation	Sample size
Deep shade	18.33	6.98	5
Partial shade	8.29	4.76	5
No shade	5.21	3.00	5

Simard et al. (1997) **Net transfer of carbon between ectomycorrhizal tree species in the field.** *Nature* 388:579-582.

ANOVA results

Source of variation	SS	df	MS	F	P
Groups (treatments)	470.704	2	235.352	8.784	0.004
Error	321.512	12	26.792		
Total	792.216	14			

Order the group means

No shade	Partial shade	Deep shade
$\bar{Y}_3$	$\bar{Y}_2$	$\bar{Y}_1$
5.21	8.29	18.33

Null hypotheses for Tukey-Kramer

$$H_0: \mu_1 = \mu_2$$

$$H_0: \mu_1 = \mu_3$$

$$H_0: \mu_2 = \mu_3$$

Why not use a series of two-sample *t*-tests?

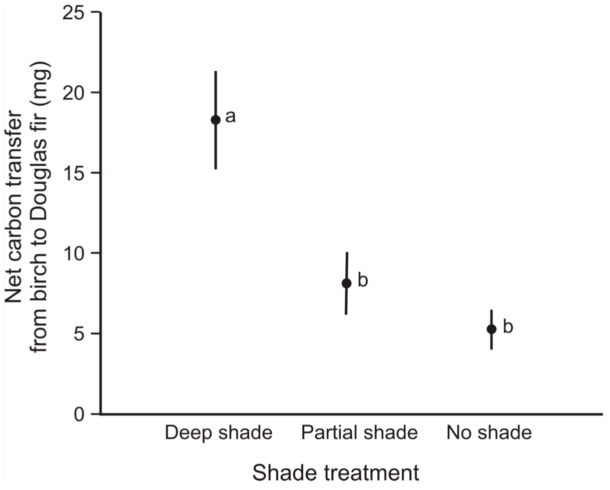
Multiple comparisons would cause the *t*-tests to reject too many true null hypotheses.

Tukey-Kramer adjusts for the number of tests.

Tukey-Kramer also uses information about the variance within groups from all the data, so it has more power than a *t*-test with a Bonferroni correction.

Results

Group *i*	**Group *j***	$\bar{Y}_i - \bar{Y}_j$	**SE**	**Test statistic** q	**Critical value** $q_{0.05,3,12}$	**Conclusion**
Deep	No	13.12	3.273693	4.008	2.67	Reject H_0
Deep	Partial	10.04	3.273693	3.067	2.67	Reject H_0
Partial	No	3.08	3.273693	0.941	2.67	Do not reject H_0



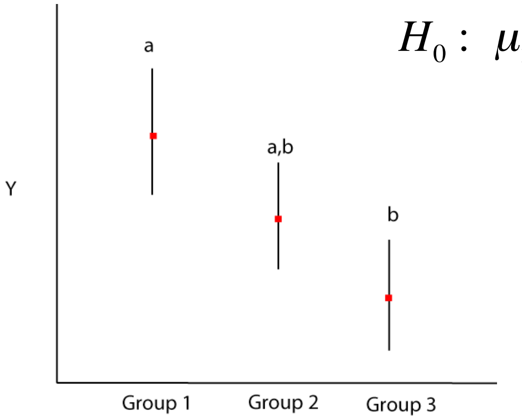
Groups which cannot be distinguished share the same letter.

Another imaginary example:

$$H_0: \mu_1 = \mu_2$$ Cannot reject

$$H_0: \mu_1 = \mu_3$$ Reject

$$H_0: \mu_2 = \mu_3$$ Cannot reject



With the Tukey-Kramer method, the probability of making at least one Type 1 error throughout the course of testing all pairs of means is no greater than the significance level α.