

ANOVA

Comparing the means of more than two groups

Null hypothesis for simple ANOVA

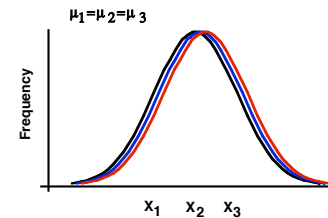
- H_0 : Variance among groups = 0

OR

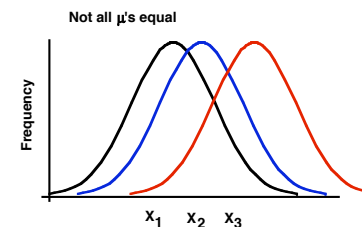
- $H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 = \dots \mu_k$

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?



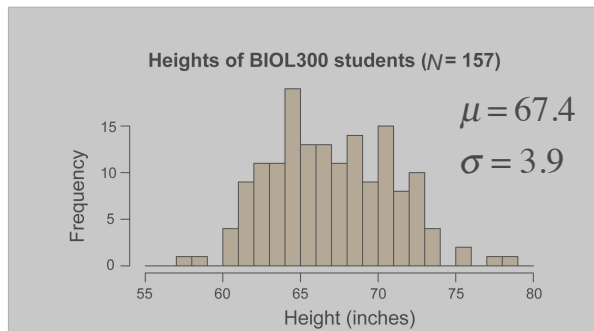
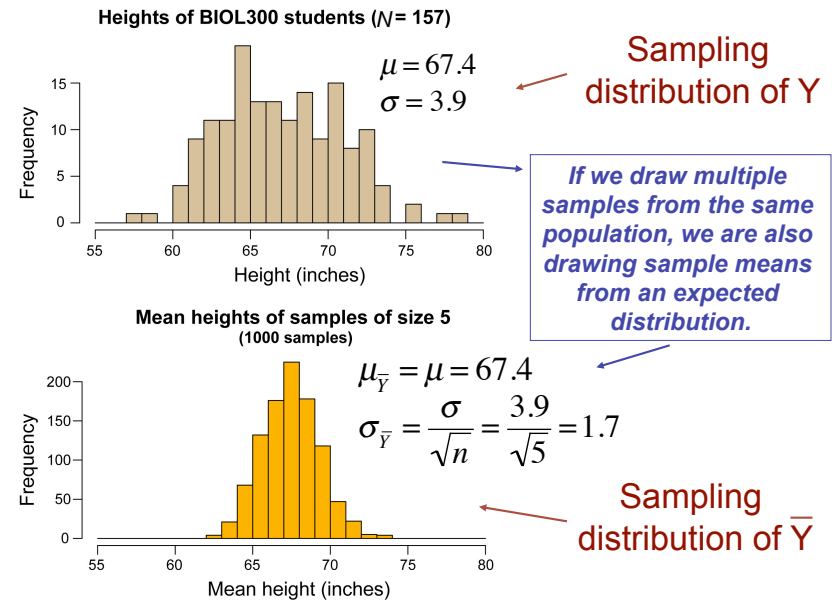
H_0 : all populations have equal means



H_A : at least one population mean is different.

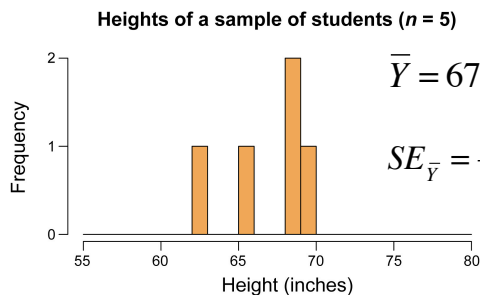
ANOVA's v. *t*-tests

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



In most cases, we don't know the real population distribution.

We only have a sample.



$$SE_{\bar{Y}} = \frac{s}{\sqrt{n}} = \frac{3.1}{\sqrt{5}} = 1.4$$

We use this as an estimate of $\sigma_{\bar{Y}}$

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:

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

(Note that we used standard errors in *t*-tests, e.g.: $t = \frac{\bar{Y}_1 - \bar{Y}_2}{SE_{\bar{Y}_1 - \bar{Y}_2}}$)

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.

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):

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

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

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

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

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

Population parameters

Estimates from sample

$$n \sigma_{\bar{x}}^2$$

is estimated by the
"Mean Squares Group"

MS_{group}

$$\sigma_x^2$$

is the variance within groups,
estimated by the
"Mean Squares Error"

MS_{error}

Mean squares group

Abbreviation: MS_{group}

Estimates this parameter: $n(\sigma_x^2 + \text{Variance}[\mu_i])$

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

Mean squares error

Abbreviation: MS_{error}

Estimates this parameter: σ_x^2

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

Mean squares group

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

\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

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: $F = \frac{n \sigma_{\bar{x}}^2}{\sigma_x^2} = 1$

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

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.

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:

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

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.

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

Summary data

Group	\bar{x}	s	n
Cold	31.0	0.551	20
Warm	37.2	0.582	21
Hot	41.0	0.430	20

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

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.

Error Mean square for squirrels

$$\begin{aligned}SS_{error} &= \sum df_i s_i^2 \\&= 19(0.551)^2 + 20(0.582)^2 + 19(0.430)^2 \\&= 16.1\end{aligned}$$

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

$$MS_{error} = \frac{16.1}{58} = 0.277$$

Squirrel Mean Squares Group:

$$\bar{X} = \frac{20(31.0) + 21(37.2) + 20(41.0)}{20 + 21 + 20} = 36.4$$

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

$$SS_{group} = 20(31.0 - 36.4)^2 + 21(37.2 - 36.4)^2 + 20(41.0 - 36.4)^2 \\ = 1015.7$$

The test statistic for ANOVA is
 F

$$F = \frac{MS_{group}}{MS_{error}} = \frac{507.9}{0.277} = 1834.7$$

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

Squirrel Mean Squares Group:

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

$$MS_{groups} = \frac{SS_{groups}}{df_{groups}} = \frac{1015.7}{2} = 507.9$$

Compare to $F_{\alpha(1), df_{group}, df_{error}}$

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

Since $1835 > 3.15$, 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	1015.7	2	507.9	1834.7	<0.0001
Error	16.1	58	0.277		
Total	1031.8	60			

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



"rayed"

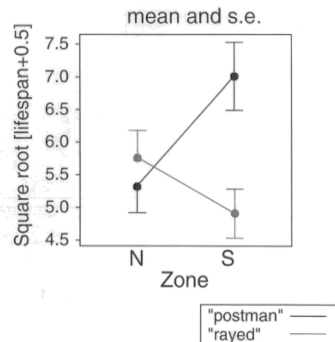


Heliconius erato

"postman"

This experiment uses two "morphs":
the rayed morph from the "north",
and the postman morph from the "south."

Testing multiple hypotheses



- H_0 : Mean lifespans are the same in both geographical zones.
- H_0 : Mean lifespans are the same for both morphs.
- H_0 : There is no interaction between geographical zone and morph.

Heliconius ANOVA table

Source of variation	SS	df	MS	F	P
Zone	9.1	1	9.1	0.96	0.327
Morph	34.6	1	34.6	3.68	0.056
Zone*Morph	80.5	1	80.5	8.59	0.004
Error	1837.9	196	9.38		

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%.

"Bonferroni correction" for multiple comparisons

Uses a smaller α 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.



Test for 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

$$H_0 : \mu_1 = \mu_2$$

$$H_0 : \mu_1 = \mu_3$$

$$H_0 : \mu_2 = \mu_3$$

Results

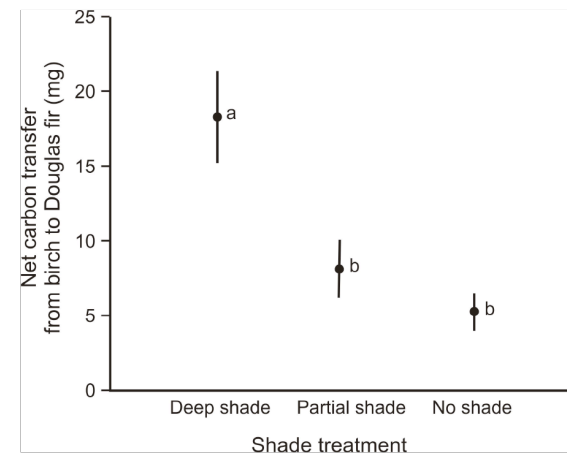
Group <i>i</i>	Group <i>j</i>	$\bar{Y}_i - \bar{Y}_j$	SE	Test statistic <i>q</i>	Critical value $q_{0.05,3,12}$	Conclusion
Deep	No	13.12	3.273693	4.008	2.67	Reject H_0
Deep	Partial	10.04	3.273693	3.067	2.67	Reject H_0
Partial	No	3.08	3.273693	0.941	2.67	Do not reject H_0

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.



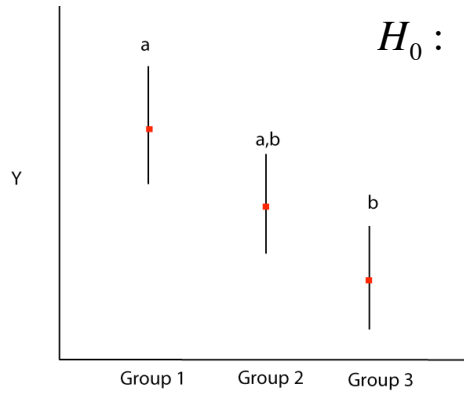
Groups which cannot be distinguished share the same letter.

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 .

