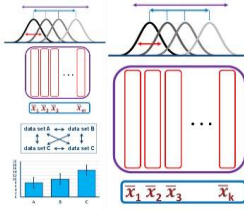


The one factor ANOVA: examples

Question: if any of 3+ population means differ?
 Approach: **ANALYSIS OF VARIANCE = ANOVA**
 Prerequisite = test for equality of population variances
 Do the ANOVA
 Conceptual hypotheses: $H_0: \mu_1 = \mu_2 = \dots = \mu_k$
 H_a : at least two differ
 Practical hypotheses: H_0 : MSA > MSW
 H_a : MSA < MSW
 Calculate SST, SSA, SSW, MSA, MSW
 Calculate F=MSA/MSW
 Create ANOVA table
 Determine p value and "reject" or "fail to reject" H_0
 If H_0 rejected: 1. Perform Bonferroni corrected t-tests
 2. Calculate MSD or HSD and compare



How to get SST, SSA, and SSW ?

SST: calculate sum of squares for all data values (comparing to overall mean).

SSA: calculate sum of squares of the \bar{x}_i values (comparing to overall mean), then multiply by group sample size, n.

SSW: calculate the sum of squares values separately for each of the k groups using the group means and sum them.

Degrees of freedom

df for among groups: $k - 1$
 df for within groups: $k(n - 1) = N - k$
 df for entire data set: $N - 1$

$$MSA = \frac{SSA}{k - 1}$$

$$MSW = \frac{SSW}{N - k}$$

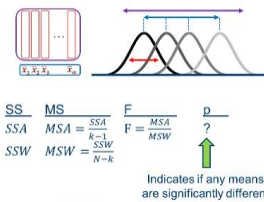
$$F_{calc} = \frac{MSA}{MSW}$$

Source	df	SS	MS	F	p
Among groups	k - 1	SSA	MSA	F = MSA/MSW	p
Within groups	N - k	SSW	MSW		
Total	N - 1				

Watch our intro to the F-test video for more about this!

The ANOVA Table

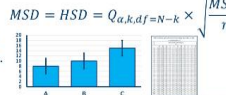
Data is usually presented in an ANOVA Table



After the ANOVA

Option 1: Bonferroni corrected t-tests
 Go back to data sets and do all pairwise t-tests, but with a smaller α value (i.e., less than 0.05) as the threshold for significance.
 Use $\alpha^* = \frac{0.05}{n}$ where n is the number of t-tests

Option 2: Tukey-Cramer comparison intervals
 Create $\pm 1/2$ MSD intervals around each sample mean. Non-overlapping intervals indicate differing means.



Example #1 - SVL of snakes

5. Calculate MSA, MSW, F value.

$$MSA = \frac{SSA}{k-1} = \frac{32}{3-1} = \frac{32}{2} = 16$$

$$MSW = \frac{SSW}{N-k} = \frac{28}{12-3} = \frac{28}{9} = 3.111$$

$$F_{calc} = \frac{16}{3.111} = 5.143$$

Ind.	siteA	siteB	siteC
1	32	34	35
2	34	35	35
3	32	36	38
4	30	31	36
mean	32	34	36
Var	2.667	4.667	2.000
Overall mean = 34			
SST=60, SSA=32, SSW=28			

Example #1 - SVL of snakes

4. Calculate SST, SSA, SSW.

$$SST = \sum (x_{ij} - 34)^2 = 60$$

$$SSA = 4[(32 - 34)^2 + (34 - 34)^2 + (36 - 34)^2] = 4[4 + 0 + 4] = 4 \times 8 = 32$$

$$SSW = \sum (x_{i,1} - 32)^2 + \sum (x_{i,2} - 34)^2 + \sum (x_{i,3} - 36)^2 = 8 + 14 + 6 = 28$$

Ind.	siteA	siteB	siteC
1	32	34	35
2	34	35	35
3	32	36	38
4	30	31	36
mean	32	34	36
Var	2.667	4.667	2.000
Overall mean = 34			

Example #1 - SVL of snakes

1. Calculate overall mean and variance, group means and variances.
 2. Conduct F_{max} (or other) test for equality of variances.

$$F_{max} = \frac{4.667}{2.000} = 2.333, p > 0.05$$

Ind.	siteA	siteB	siteC
1	32	34	35
2	34	35	35
3	32	36	38
4	30	31	36
mean	32	34	36
Var	2.667	4.667	2.000
Overall mean = 34			

Example #1 - SVL of snakes

1. Calculate overall mean and variance, group means and variances.
2. Conduct F_{max} (or other) test for equality of variances.
3. Proceed with ANOVA or transform data values (retry step 2).
4. Calculate SST, SSA, SSW.
5. Calculate MSA, MSW, F value.
6. Assess significance of F value.
7. If $p < 0.05$ then do steps 7 & 8.
8. Do Bonferroni corrected t-tests
9. Interpret and present results.

Ind.	siteA	siteB	siteC
1	32	34	35
2	34	35	35
3	32	36	38
4	30	31	36
mean	32	34	36
Var	2.667	4.667	2.000

Example #1 - SVL of snakes

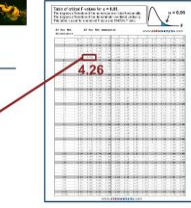
Consider a study that collects 4 snakes from each of 3 different locations and measures their lengths (SVL) to see if the mean SVL in any locations differ from the others.

Ind.	siteA	siteB	siteC
1	32	34	35
2	34	35	35
3	32	36	38
4	30	31	36

Example #1 - SVL of snakes

6. Assess significance of F value.

Source	df	SS	MS	F	p
Among	2	32	16	5.143	
Within	9	28	3.111		
Total	11	60			



Example #1 - SVL of snakes

6. Assess significance of F value.

Source	df	SS	MS	F	p
Among	2	32	16	5.143	0.0324
Within	9	28	3.111		
Total	11	60			



Example #1 - SVL of snakes

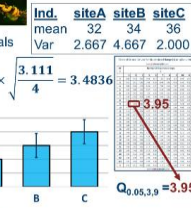
7. Do Bonferroni corrected t-tests

3 tests, $\frac{0.05}{3} = 0.01833$ is the corrected critical value threshold.
 A vs B: $p = 0.1937$ ← nonsignificant difference
 A vs C: $p = 0.0104$ ← significant difference
 B vs C: $p = 0.1801$ ← nonsignificant difference

Example #1 - SVL of snakes

8. Tukey-Cramer comparison intervals

$$MSD = Q_{\alpha,k,df=N-k} \times \sqrt{\frac{MSW}{n}} = 3.95 \times \sqrt{\frac{3.111}{4}} = 3.4836$$



Example #1 - SVL of snakes

9. Interpret results.

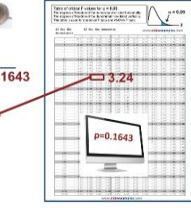
A one-factor ANOVA indicates that the mean SVL values differ significantly between the 3 snake populations ($F_{2,9} = 5.143$; $p = 0.0324$). Bonferroni corrected t-tests and the Tukey-Cramer comparison intervals indicate that the mean of population A is significantly smaller than the mean of population C.

Source	df	SS	MS	F	p
Among	2	32	16	5.143	0.0324
Within	9	28	3.111		
Total	11	60			

Example #2 - size of mice

6. Assess significance of F value.

Source	df	SS	MS	F	p
Among	3	25	8.333	2.222	0.1643
Within	16	60	3.750		
Total	19	85			



Example #2 - size of mice

5. Calculate MSA, MSW, F value.

$$MSA = \frac{SSA}{k-1} = \frac{25}{4-1} = \frac{25}{3} = 8.333$$

$$MSW = \frac{SSW}{N-k} = \frac{60}{20-4} = \frac{60}{16} = 3.750$$

$$F_{calc} = \frac{8.333}{3.750} = 2.222$$

Ind.	dietA	dietB	dietC	dietD
1	21	20	19	20
2	16	23	17	23
3	19	21	23	21
4	20	23	20	18
5	19	23	21	23
6	19	22	20	21
mean	3.50	2.00	5.00	4.50
var	3.50	2.00	5.00	4.50
overall mean = 20.5				
SST=85, SSA=25, SSW=60				

Example #2 - size of mice

4. Calculate SST, SSA, SSW.

$$SST = \sum (x_{ij} - 20.5)^2 = 85$$

$$SSA = 5[(19 - 20.5)^2 + (22 - 20.5)^2 + (20 - 20.5)^2] = 5[2.25 + 2.25 + 0.25] = 5 \times 5.00 = 25$$

$$SSW = \sum (x_{i,1} - 19)^2 + \sum (x_{i,2} - 22)^2 + \sum (x_{i,3} - 20)^2 + \sum (x_{i,4} - 21)^2 = 14 + 8 + 20 + 18 = 60$$

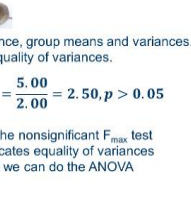
Ind.	dietA	dietB	dietC	dietD
1	21	20	19	20
2	16	23	17	23
3	19	21	23	21
4	20	23	20	18
5	19	23	21	23
6	19	22	20	21
mean	3.50	2.00	5.00	4.50
var	3.50	2.00	5.00	4.50
overall mean = 20.5				

Example #2 - size of mice

1. Calculate overall mean and variance, group means and variances.
 2. Conduct F_{max} (or other) test for equality of variances.

$$F_{max} = \frac{5.00}{2.00} = 2.50, p > 0.05$$

3. The nonsignificant F_{max} test indicates equality of variances and we can do the ANOVA



Example #2 - size of mice

Consider a study that tests 4 different diets for lab mice and has 5 mice in each treatment. The question is whether the mass at maturity differs for any of the diets.

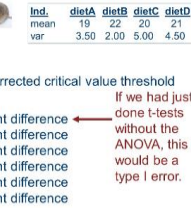
Source	df	SS	MS	F	p
Among	3	25	8.333	2.222	0.1643
Within	16	60	3.750		
Total	19	85			

Example #2 - size of mice

7. Do Bonferroni correct t-tests

$$6 \text{ tests, } \frac{0.05}{6} = 0.00833 \text{ is the corrected critical value threshold}$$

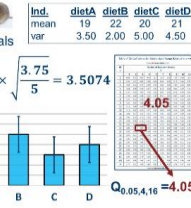
A vs B: $p = 0.0228$ ← significant difference
 A vs C: $p = 0.4658$ ← nonsignificant difference
 A vs D: $p = 0.1531$ ← nonsignificant difference
 B vs C: $p = 0.1363$ ← nonsignificant difference
 B vs D: $p = 0.4097$ ← nonsignificant difference
 C vs D: $p = 0.4889$ ← nonsignificant difference



Example #2 - size of mice

8. Tukey-Cramer comparison intervals

$$MSD = Q_{\alpha,k,df=N-k} \times \sqrt{\frac{MSW}{n}} = 4.05 \times \sqrt{\frac{3.75}{5}} = 3.5074$$



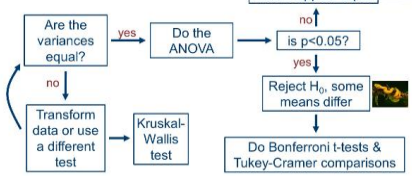
Example #2 - size of mice

9. Interpret results.

A one-factor ANOVA indicates that the mean sizes of the mice at maturity in the 4 diet groups do not differ significantly from one another ($F_{3,16} = 2.222$; $p = 0.1643$).

Source	df	SS	MS	F	p
Among	3	25	8.333	2.222	0.1643
Within	16	60	3.750		
Total	19	85			

ANOVA flowchart



Example #3 - self-check

Consider the data sets shown.
 ANOVA questions:
 1. What is the ANOVA table?
 2. What are the comparison intervals?
 3. Which means differ?

Ind.	setA	setB	setC	setD	setE
1	16	17	11	14	18
2	15	18	17	13	22
3	17	17	19	13	18
4	20	18	14	15	18
5	15	19	19	16	20
6	19	19	16	13	18

Answers at end of video