Lesson 6: Analysis of variance: One-way ANOVA

Victoria Blanes-Vidal
Manuella Lech Cantuaria
The Maersk Mc-Kinney Møller Institute
Applied AI and Data Science

Lesson	Week	Date	TOPICS	Teacher
1	35	1/Sep	Introduction to the course Descriptive statistics –Part I	MLC
2	36	8/sep	Descriptive statistics –Part II	MLC
3	37	15/Sep	Probability distributions	MLC
4	38	22/Sep	Hypothesis testing (one sample)	VBV
5	39	29/Sep	Hypothesis testing (two samples)	VBV
6	40	6/Oct	ANOVA one-way	VBV
7	41	13/Oct	R class (Introduction to R and descriptive statistics) Point giving activity	MLC
-	42	20/Oct	NO CLASS (Autum holidays)	
8	43	27/Oct	R class (hypothesis testing + ANOVA)	MLC
9	44	3/Nov	ANOVA two-way	VBV
-	45	10/Nov	NO CLASS	
10	46	17/Nov	Regression analysis	VBV
11	47	24/Nov	Multiple regression Point giving activity	MLC
12	48	1/Dec	Notions of experimental design and questions	VBV+MLC

VBV = Victoria Blanes-Vidal
MLC = Manuella Lech Cantuaria

About the point giving activity

- There are two point giving activities in this course: 13th of October and 24th of November.

- In the final exam you can get up to 100 points. Pointgiving activities serve as a **”bonus”** (if all responses of the two point giving activities are correct, we will add 10 points to the points you got at the final exam)

- Point giving activities are carried out in the classroom and individually, using itslearning.

- You can consult any material you like (**it will NOT be like that in the final exam**)

- The first point giving activity will be next week Thursday 13th of October from 13:10 to 14:00.

- In this first point giving activity there will be 5 multichoice questions, and each question will have a value of 1 point.

Analysis of variance: One-way ANOVA

1. **What is “Analysis of variance” (ANOVA)?**
2. **One-Way Analysis of Variance**
3. **The Least significant difference intervals**

An example

4. **The p-value**
5. **In-class exercise** 

What is “Analysis of variance” (ANOVA)?

In the previous lesson, we introduced a statistical method for comparing the means of two independent populations (Student t-test):

Hood closed

μ_c

Hood open

μ_o

However, many studies involve comparison of more than two independent populations. For example…

Example 1:

A magazine publisher wants to compare three different layouts for a magazine that will be displayed at supermarket checkout lines.

She is interested in whether there is a layout that better impresses shoppers and results in more sales.

μ_1 μ_2 μ_3

To investigate this, she randomly assigns 20 stores to each of the three layouts and records the number of magazines that are sold at each store, in a one-week period.

Example 2:

A marketing specialist needs to know how five Starbucks coffeehouses in the same city differ in the age of their customers.

μ_1

μ_2

μ_3

μ_4

μ_5

She asks 15 customers of each store to respond to a questionnaire, that includes a question about their age.

These two examples are similar in that:

The goal is to compare ***more than two*** populations:

Three magazine layouts.

Five coffeehouses.

There is ***a quantitative response variable***:

Number of magazines that are sold

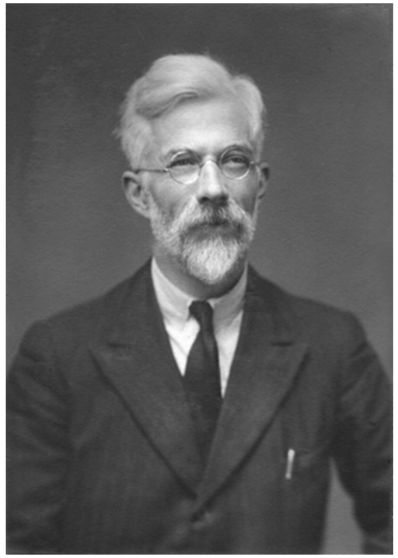
Age of the customers

A statistical test based on the **F-distribution (Fisher-Snedecor distribution)**, can be used **to compare the means of three or more populations.**

This technique is called ANOVA.

ANOVA is an abbreviation of the full name of the method:

ANalysis **O**f **VA**riance

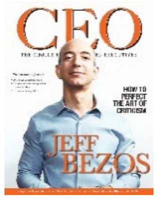


Sir Ronald Fisher (1890 –1962) was a British mathematician, statistician, geneticist, biologist and academic. He made hugely important contributions in many areas, including statistics and genetics. He's considered the father of modern statistics.

*Be aware that, although the method is called "analysis of **variance"** , the method is used to find out whether the **means** of three or more populations are different.*

(It is called analysis of variance because the theoretical foundation of the method is based on comparing variabilities in a "particular way")

Analysis of variance (ANOVA), is a statistical method that allows us to determine if differences in mean values between three or more sample groups are by chance, or if the populations are indeed significantly different.




$\bar{x}_1$ $\bar{x}_2$ $\bar{x}_3$ $\bar{x}_1$ $\bar{x}_2$ $\bar{x}_3$ $\bar{x}_4$ $\bar{x}_5$

? ?

μ_1 μ_2 μ_3 μ_1 μ_2 μ_3 μ_4 μ_5

Analysis of variance: One-way ANOVA

1. What is "Analysis of variance" (ANOVA)?
2. One-Way Analysis of Variance
3. The Least significant difference intervals

An example

4. The p-value
5. In-class exercise 

One-way analysis of variance

<u>One-way</u> analysis of variance is used with data that can be categorized according to **one** factor (sometimes also called, “treatment”).

The factor (or treatment) is a characteristic that allows us to distinguish the different populations from one another.

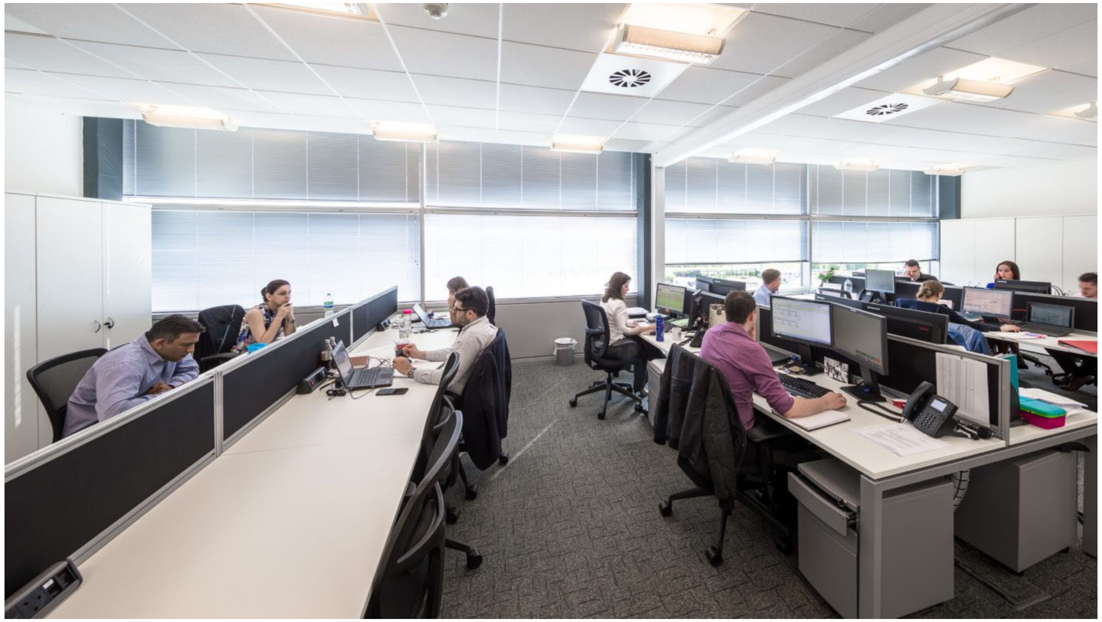
Factor: Layout of the magazines

Factor: Coffeehouse

Note: Later in the course, we will study the Two-way ANOVA

One-way ANOVA
An example: Computer mice

The employees of a company need to use a laptop and a mouse, in order to carry out their daily work.



An example: Computer mice

Currently, these employees can use one of these two kinds of laptop mouse:

Mouse type 1: Classic mouse

Mouse type 2: Touchpad

A 3rd company offers an alternative type of mouse (RollerMouse) which is more expensive, but the company claims that the rollermouse is better (that is, it takes less time to perform the same computer tasks) than those mice that are currently used at the company.

Mouse type 3: Rollermouse

In order to decide whether it is a good idea to buy this alternative type of mouse, the manager at the company decides to carry out one experiment.

He asks:

- 10 employees to perform a specific computer task using Mouse type 1 (Classic mouse)

- 10 other employees to perform that same specific computer task using Mouse type 2 (Touchpad)

- 10 other employees to perform the same specific computer task using Mouse type 3 (Rollermouse)

He then times how long it takes each of these 30 employees, to perform the computer task.

Note:

- The employees are assigned to each of these 3 groups, at random.

- Before starting the experiment, the employees have learned how to use the Mouse type 3 (Rollermouse), and they have used it for several hours a day, during one week prior to the experiment, so that they get familiar with it.

These are the results:



	Mouse type 1	Mouse type 2	Mouse type 3
1	23	35	23
2	28	36	26
3	21	29	21
4	27	40	23
5	95	43	26
6	41	49	32
7	37	51	23
8	30	28	26
9	32	50	25
10	36	52	31
Mean	37	41.3	25.6

Is there any difference in the time that it takes to perform the computer task, when the employees use Mouse type 1, 2 or 3?

ANOVA (hypothesis testing for comparison of means of >2 populations): 5 Steps

Step 1. Formulate the null hypothesis and the alternative hypothesis

Step 2: Select the statistical test

Step 3. Calculate the test statistic (F-ratio)

- **3.1. Calculate the total sum of squares, sum of squares of the treatments and residual sum of squares**
- **3.2. Calculate the degrees of freedom of the treatment and residuals (d.f.).**
- **3.3. Calculate the mean squares of the treatment and residuals**
- **3.4. Calculate the test statistic (F-ratio).**

Step 4. Find the critical value of F distribution in the table

Step 5: Make the decision to reject or accept the null hypothesis

Step 1. Formulate the null hypothesis and the alternative hypothesis

When comparing >2 populations:

The **null hypothesis (H0)** is that the three or more population means are not different from each other.

$$\mu_1 = \mu_2 = \mu_3 = \ldots \mu_k$$

The **alternative hypothesis (H1)** is that at least one mean is different from another one.

Null hypothesis:

$$\mu_1 = \mu_2 = \mu_3$$

Alternative hypothesis:

$\mu_1 \neq \mu_2$ **and/or**

$\mu_2 \neq \mu_3$ **and/or**

$\mu_1 \neq \mu_3$

"The three types of mice are not significantly different from each other", that is, "the time it takes to perform the task does not differ significantly when using the mouse 1, 2 or 3"

"At least mouse is different from another one", that is, "at least, the time it takes to perform the task using one of the types of mice is significantly different from the time it takes to perform the task when using one of the other types".

Step 2: Select the statistical test

Since we want to compare the means of 3 or more independent populations, we select the statistical test:

One-way Analysis of Variance (One-way ANOVA)

If the null hypothesis (H_0) is true ($\boldsymbol{\mu_1 = \mu_2 = \mu_3}$), then, the test statistic F_{ratio} "follows" a F distribution ($\mathbf{F_{k-1, N-k}}$)

$$\boxed{F_{ratio}} \approx \boxed{F_{k-1,\, N-k}}$$

To make a decision, we compare the test statistic, with the F probability distribution.

Step 3. Calculate the test statistic (F-ratio)

3.1. Calculate the total sum of squares, sum of squares of the treatment and residual sum of squares

	Sum of Squares	**d.f.**	**Mean Squares**	***F-ratio***
Factor (or treatment, or between)	$SS_{between}$	k-1	$MS_{between}$	$MS_{between}/MS_{residual}$
Residual (or error, or within)	$SS_{residual}$	df_{total}- $df_{between}$	$MS_{residual}$	
Total	SS_{total}	*N-1*		

3.1. Calculate the total sum of squares, sum of squares of the treatment and residual sum of squares

Total sum of squares (SS_{total}): Total variability

Sum of squares of the deviations of each data to the total mean.

	Mouse 1	Mouse 2	Mouse 3
1	23	35	18
2	28	36	16
3	21	29	14
4	27	40	21
5	95	43	25
6	41	49	18
7	37	51	22
8	30	28	26
9	32	50	25
10	36	52	28
Mean	37	41.3	21.3

$$SS_{total} = \sum_{N} \left(x_i - total\ mean\right)^2$$

$$(23 - 33.2)^2 + (35 - 33.2)^2 + (18 - 33.2)^2 + \cdots$$
$$= 7247$$

Total mean = 33.2

N = 30

k = 3

3.1. Calculate the total sum of squares, sum of squares of the treatment and residual sum of squares

Sum of squares of the treatments ($SS_{between}$): Variability between treatments

Sum of squares of the deviations of the means of each treatment to the total mean.

	Mouse1	Mouse 2	Mouse 3
1	23	35	18
2	28	36	16
3	21	29	14
4	27	40	21
5	95	43	25
6	41	49	18
7	37	51	22
8	30	28	26
9	32	50	25
10	36	52	28
Mean	37	41.3	21.3

Total mean = 33.2

N = 30

k = 3

$$SS_{between} = N_1 \cdot \left(\overline{x_1} - total\ mean\right)^2 + N_2 \cdot \left(\overline{x_2} - total\ mean\right)^2 + N_3 \cdot \left(\overline{x_3} - total\ mean\right)^2$$

$$SS_{between} = 10 \cdot (37 - 33.2)^2 + 10 \cdot (41.3 - 33.2)^2 + 10 \cdot (21.3 - 33.2)^2 = 2217$$

3.1. Calculate the total sum of squares, sum of squares of the treatment and residual sum of squares

Residual sum of squares ($SS_{residual}$): Variability within treatments

$$SS_{residual} = SS_{total} - SS_{between}$$

$$SS_{residual} = 7247 - 2217 = 5030$$

3.2. Calculate the degrees of freedom of the treatment and residuals (d.f.).

	Sum of Squares	d.f.	Mean Squares	*F-ratio*
Factor (or treatment, or between)	2217	k-1	$MS_{between}$	$MS_{between}/MS_{residual}$
Residual (or error, or within)	5030	df_{total}-$df_{between}$	$MS_{residual}$	
Total	7247	N-1		

N = Total number of samples = 30

k = Number of levels or groups = 3

3.2. Calculate the degrees of freedom of the treatment and residuals (d.f.).

	Sum of Squares	d.f.	Mean Squares	*F-ratio*
Factor (or treatment, or between)	2217	3-1=2	$MS_{between}$	F-ratio
Residual (or error, or within)	5030	29-2=27	$MS_{residual}$	
Total	7247	30-1=29		

3.3. Calculate the mean squares of the treatment and residuals

	Sum of Squares	d.f.	Mean Squares	*F-ratio*
Factor (or treatment, or between)	2217	2	$MS_b=SS_b/df_b$	F-ratio = $MS_{between}/MS_{residual}$
Residual (or error, or within)	5030	27	$MS_r=SS_r/df_r$	
Total	7247	29		

3.3. Calculate the mean squares of the treatment and residuals

	Sum of Squares	d.f.	Mean Squares	*F-ratio*
Factor (or treatment, or between) Residual (or error, or within)	2217 5030	2 27	2217/2=1108 5030/27=186	F-ratio= $MS_{between}/MS_{residual}$
Total	7247	29		

3.4. Calculate the test statistic (F-ratio).

	Sum of Squares	**d.f.**	**Mean Squares**	***F-ratio***
Factor (or treatment, or between)	2217	2	1108	F-ratio= $MS_{between}/MS_{residual}$
Residual (or error, or within)	5030	27	186	
Total	7247	29		

3.4. Calculate the test statistic (F-ratio).

	Sum of Squares	d.f.	Mean Squares	*F-ratio*
Factor (or treatment, or between) Residual (or error, or within)	2217 5030	2 27	1108 186	1108/186= 5.94
Total	7247	29		

Step 4. Find the critical value of F distribution in the table

If the null hypothesis (H_0) is true ($\boldsymbol{\mu_1 = \mu_2 = \mu_3}$), then, the test statistic F_{ratio} "follows" a F distribution ($\mathbf{F_{k-1, N-k}}$)

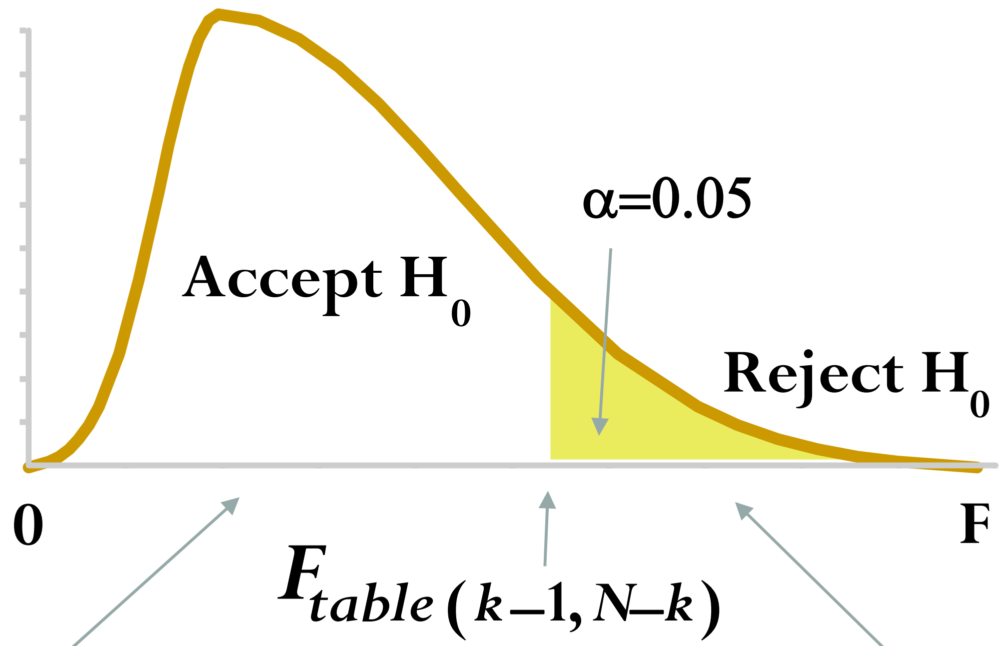
$$\boxed{F_{ratio}} \approx \boxed{F_{k-1,\, N-k}}$$

To make a decision, we compare the test statistic, with the F probability distribution.

5.94

Step 4. Find the critical value of F distribution in the table

ANOVA is "one-tailed" test.



If $\mathbf{F_{ratio}} < \mathbf{F_{table\ (k-1,\ N-k)}}$: Accept H_0

The three laptop mice are NOT significantly different from each other (that is, the time it takes to perform the task does not differ significantly when using the mouse 1, 2 or 3)

If $\mathbf{F_{ratio}} > \mathbf{F_{table\ (k-1,\ N-k)}}$: Reject H_0

At least one of the mice is significantly different from another one.

Step 4. Find the critical value of F distribution in the table

$F_{\text{table }(k-1,\ N-k)}$

N = Total number of samples = 30

k = Number of levels or groups = 3

$F_{\text{table }(3-1,\ 30-3)}$

$F_{\text{table }(2,\ 27)}$

$F_{\text{table }(2,\ 27)} = 3.35$

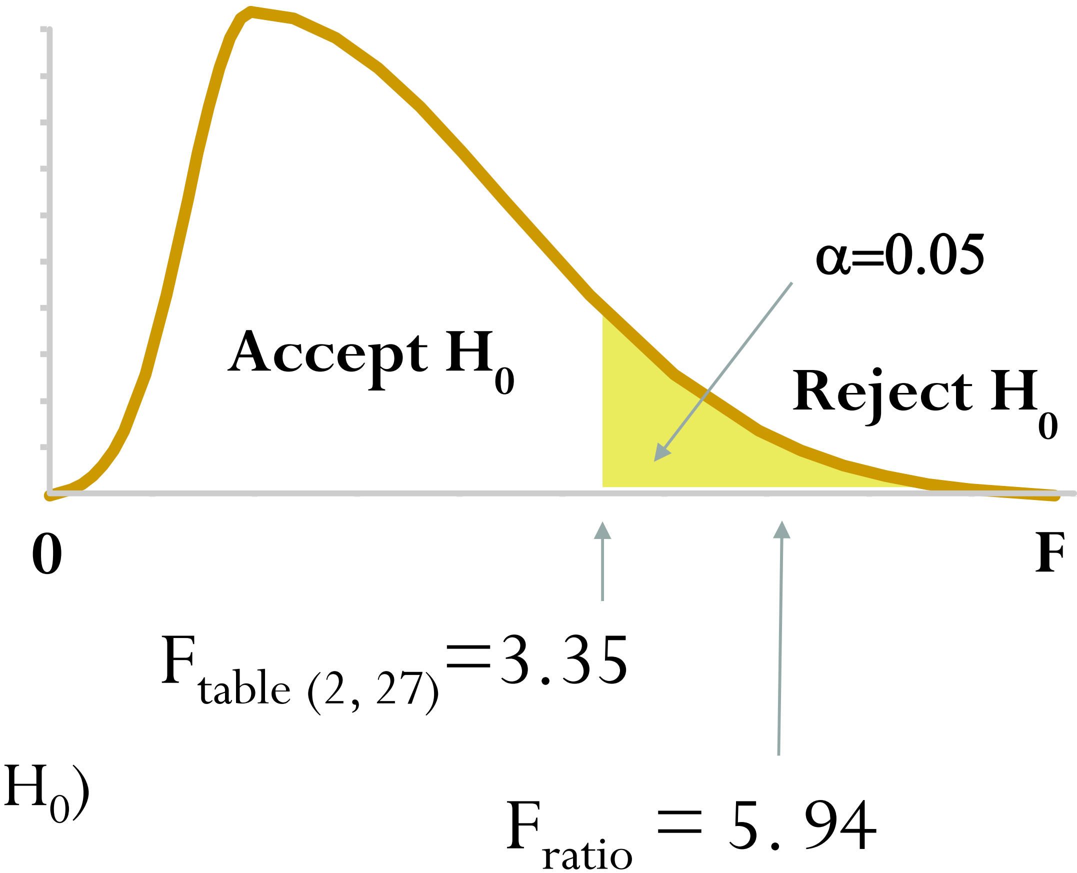
Table H *(continued)*

d.f.D.: degrees of freedom, denominator	α = 0.05										
	d.f.N.: degrees of freedom, numerator										
	1	2	3	4	5	6	7	8	9	10	12
1	161.4	199.5	215.7	224.6	230.2	234.0	236.8	238.9	240.5	241.9	243.9
2	18.51	19.00	19.16	19.25	19.30	19.33	19.35	19.37	19.38	19.40	19.41
3	10.13	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42
17	4.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16
26	4.23	3.37	2.98	2.74	2.59	2.47	2.39	2.32	2.27	2.22	2.15
27	4.21	3.35	2.96	2.73	2.57	2.46	2.37	2.31	2.25	2.20	2.13
28	4.20	3.34	2.95	2.71	2.56	2.45	2.36	2.29	2.24	2.19	2.12

Step 5: Make the decision to reject or accept the null hypothesis

$$F_{ratio} = 5.94$$
$$F_{table\,(2,\,27)} = 3.35$$

$$F_{ratio} > F_{table\,(2,\,27)}$$



We reject the null hypothesis (H_0)

Conclusion: At least one laptop mouse is significantly different from another one (regarding the time that takes to perform the task).

Conclusion from ANOVA: At least one laptop mouse is significantly different from another one.

Which one is different?

Mouse type 1: Classic mouse

Mouse type 2: Touchpad

Mouse type 3: Rollermouse

1 is different from 2 and 3.
2 and 3 are not different from each other. ?

2 is different from 1 and 3.
1 and 3 are not different from each other. ?

3 is different from 1 and 2.
1 and 2 are not different from each other. ?

All of them are different from each other. ?

This question cannot be answered with ANOVA. We need to use another test: A multiple comparison test

Analysis of variance: One-way ANOVA

1. What is “Analysis of variance” (ANOVA)?

2. One-Way Analysis of Variance

An example

3. The Least significant difference intervals

4. The p-value

5. In-class exercise 

The Least Significant Difference intervals

- The ANOVA test can only show whether or not a difference exists among the means, not where the difference lies.
- When the null hypothesis is rejected using ANOVA (i.e. “At least one mean is different from another one”), usually we need to know where the difference among the means is.
- The **least square difference intervals test** can be used (after the analysis of variance has been completed), to make pairwise comparisons between means.

Least Significant Difference (LSD) intervals

We build the LSD intervals for each factor, with the formula:

$$\overline{x}_k \pm \frac{\sqrt{2}}{2} \cdot t_{dfresiduals} \cdot \sqrt{\frac{MS_{residuals}}{n_k}}$$

LSD interval of Factor 1 (Mouse type 1: Classic Mouse):

	Mouse 1	Mouse 2	Mouse3
1	23	35	18
2	28	36	16
…	…	…	…
10	36	52	28
Mean	37	41.3	21.3

	Sum of Squares	d.f.	Mean Squares	*F-ratio*
Factor (or treatment, or between)	2217	2	1108	1108/186= 5.94
Residual (or error, or within)	5030	27	186	
Total	7247	29		

$$t_{27} = 2.052$$

$$\bar{x}_1 \pm \frac{\sqrt{2}}{2} \cdot t_{dfresiduals} \cdot \sqrt{\frac{MS_{residuals}}{n_k}} \qquad 37 \pm \frac{\sqrt{2}}{2} \cdot 2.052 \cdot \sqrt{\frac{186}{10}} \qquad 37 \pm 6.3$$

$$n_1 = 10$$

		LSD interval	
	Mean	Lower limit	Upper limit
Mouse 1	37	30.7	43.3
Mouse 2			
Mouse 3			

LSD interval of Factor 2 (Mouse type 2: Touchpad):

	Mouse 1	Mouse 2	Mouse3
1	23	35	18
2	28	36	16
…	…	…	…
10	36	52	28
Mean	37	41.3	21.3

	Sum of Squares	d.f.	Mean Squares	*F-ratio*
Factor (or treatment, or between)	2217	2	1108	1108/186= 5.94
Residual (or error, or within)	5030	27	186	
Total	7247	29		

$$t_{27} = 2.052$$

$$\bar{x}_2 \pm \frac{\sqrt{2}}{2} \cdot t_{dfresiduals} \cdot \sqrt{\frac{MS_{residuals}}{n_k}} \qquad 41.3 \pm \frac{\sqrt{2}}{2} \cdot 2.052 \cdot \sqrt{\frac{186}{10}} \qquad 41.3 \pm 6.3$$

$n_2 = 10$

		LSD interval	
	Mean	Lower limit	Upper limit
Mouse 1	37	30.7	43.3
Mouse 2	41.3	35.0	47.6
Mouse 3			

LSD interval of Factor 3 (Mouse type 3: Rollermouse):

	Mouse 1	Mouse 2	Mouse3
1	23	35	18
2	28	36	16
…	…	…	…
10	36	52	28
Mean	37	41.3	21.3

	Sum of Squares	d.f.	Mean Squares	*F-ratio*
Factor (or treatment, or between)	2217	2	1108	1108/186= 5.94
Residual (or error, or within)	5030	27	186	
Total	7247	29		

$t_{27} = 2.052$

$$\bar{x}_3 \pm \frac{\sqrt{2}}{2} \cdot t_{dfresiduals} \cdot \sqrt{\frac{MS_{residuals}}{n_k}} \qquad 21.3 \pm \frac{\sqrt{2}}{2} \cdot 2.052 \cdot \sqrt{\frac{186}{10}} \qquad 21.3 \pm 6.3$$

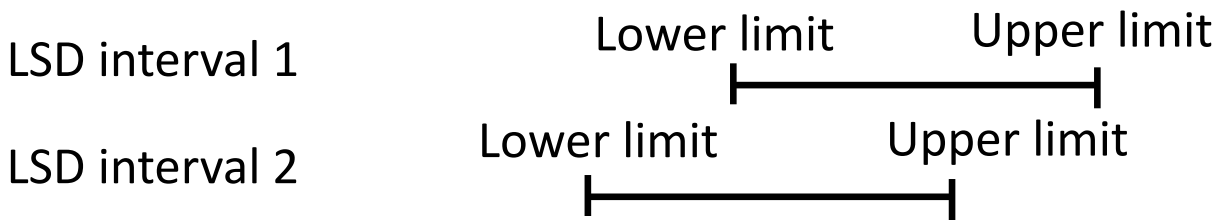
$n_3 = 10$

		LSD interval	
	Mean	Lower limit	Upper limit
Mouse 1	37	30.7	43.3
Mouse 2	41.3	35.0	47.6
Mouse 3	21.3	15.0	27.6

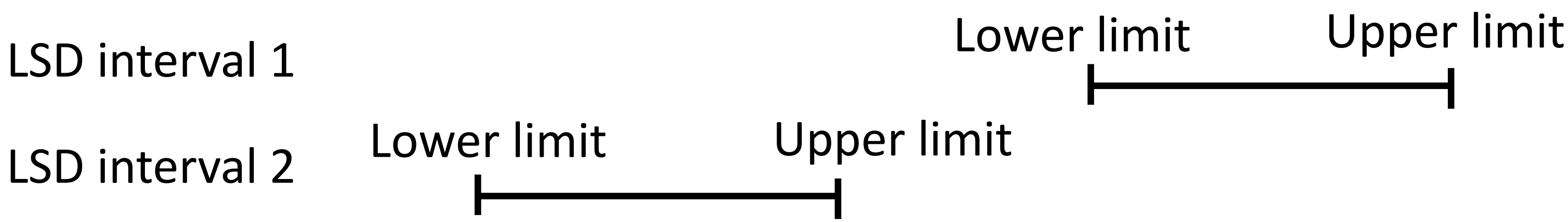
How do we interpret the LSD intervals?

		LSD interval	
	Mean	Lower limit	Upper limit
Mouse 1	37	30.7	43.3
Mouse 2	41.3	35.0	47.6
Mouse 3	21.3	15.0	27.6

- If the LSD intervals overlap, then there is NOT a significant difference between those two populations.



- If the LSD intervals DO NOT overlap, then there is a significant difference between those two populations.



How do we interpret the LSD intervals?

		LSD interval	
	Mean	Lower limit	Upper limit
Mouse 1	37	30.7	43.3
Mouse 2	41.3	35.0	47.6
Mouse 3	21.3	15.0	27.6

Conclusion

- Since the LSD interval of Mouse 3 does not overlap with the intervals of Mouse 1 or 2, the time required to perform the task when using Mouse 3 is significantly different from the time when using Mouse 1 or Mouse 2.

- Since the LSD intervals of Mouse 1 and 2 do overlap, there is NOT a significant difference between the Mouse 1 and 2.

Therefore, it is true that using the Rollermouse (instead of a Classic mouse or a Touchpad), decreases the time to perfom the task.

Analysis of variance: One-way ANOVA

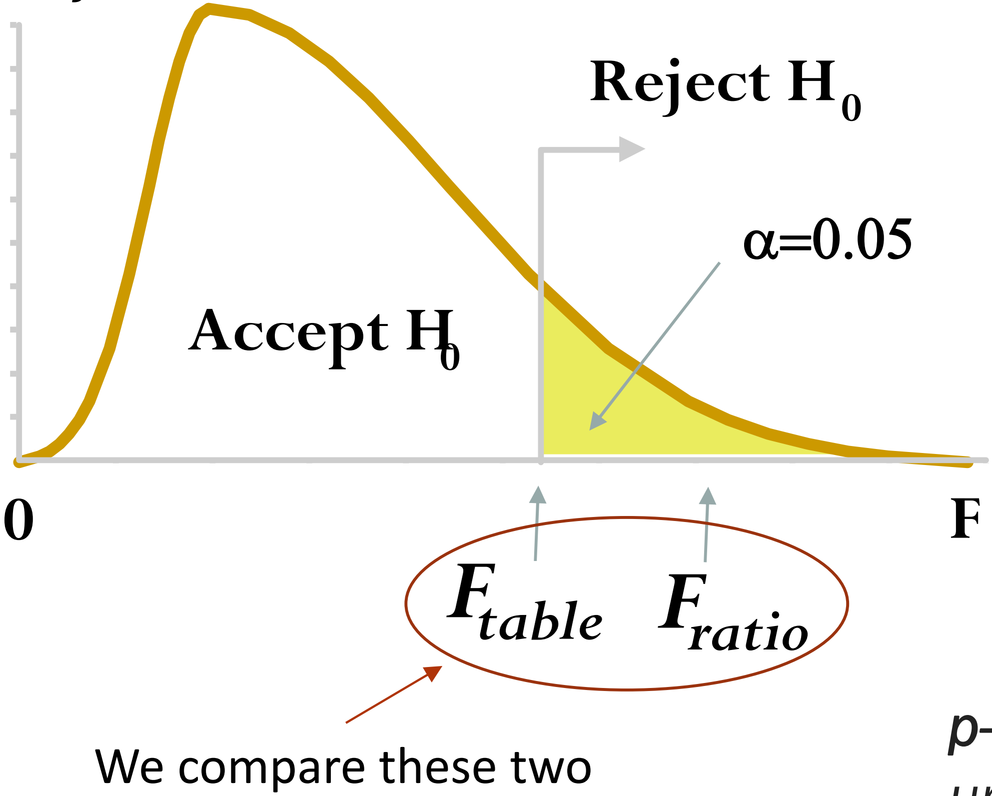
1. What is “Analysis of variance” (ANOVA)?
2. One-Way Analysis of Variance
3. The Least significant difference intervals

$\left.\begin{array}{l}\text{2. One-Way Analysis of Variance}\\ \text{3. The Least significant difference intervals}\end{array}\right\}$ An example

4. **The p-value**
5. **In-class exercise** 

The P-value approach

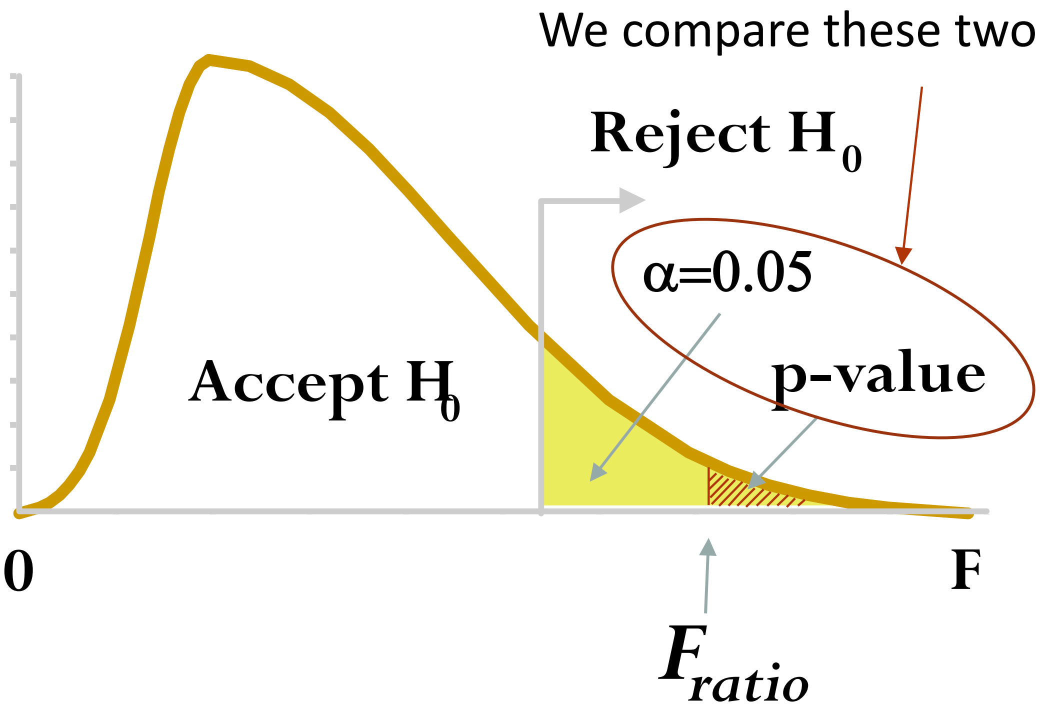
When we perform calculations by hand:



If $\mathbf{F_{ratio}} < \mathbf{F_{table}}$: Accept H_0

If $\mathbf{F_{ratio}} > \mathbf{F_{table}}$: Reject H_0

When we use R:



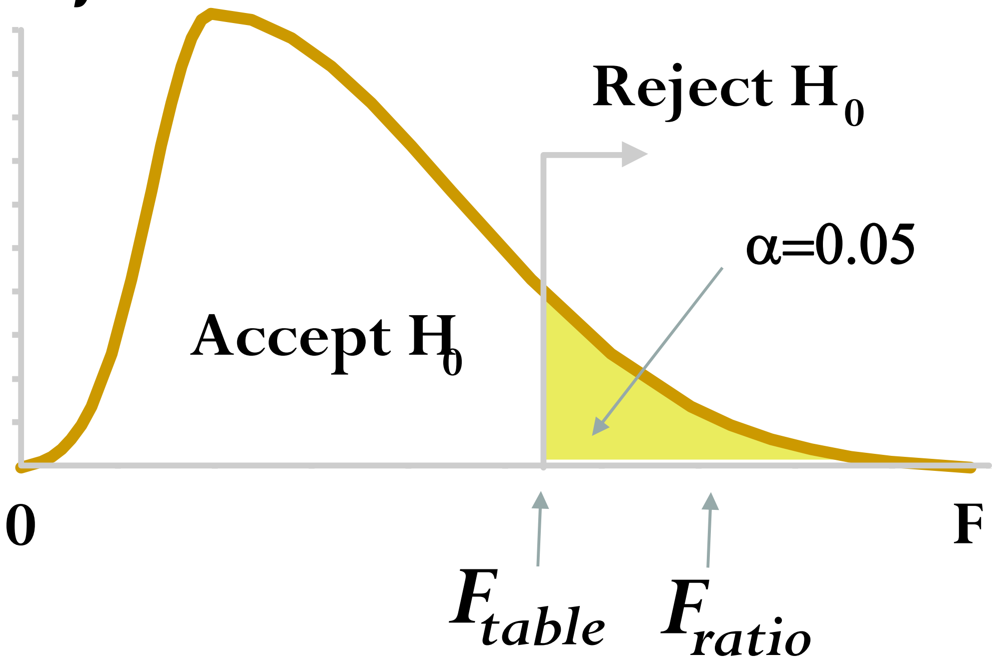
p-value is the cumulative probability (area under the curve) of the values to the right of the calculated test statistic (e.g. F ratio in the figure above).

If **p-value > 0.05** : Accept H_0

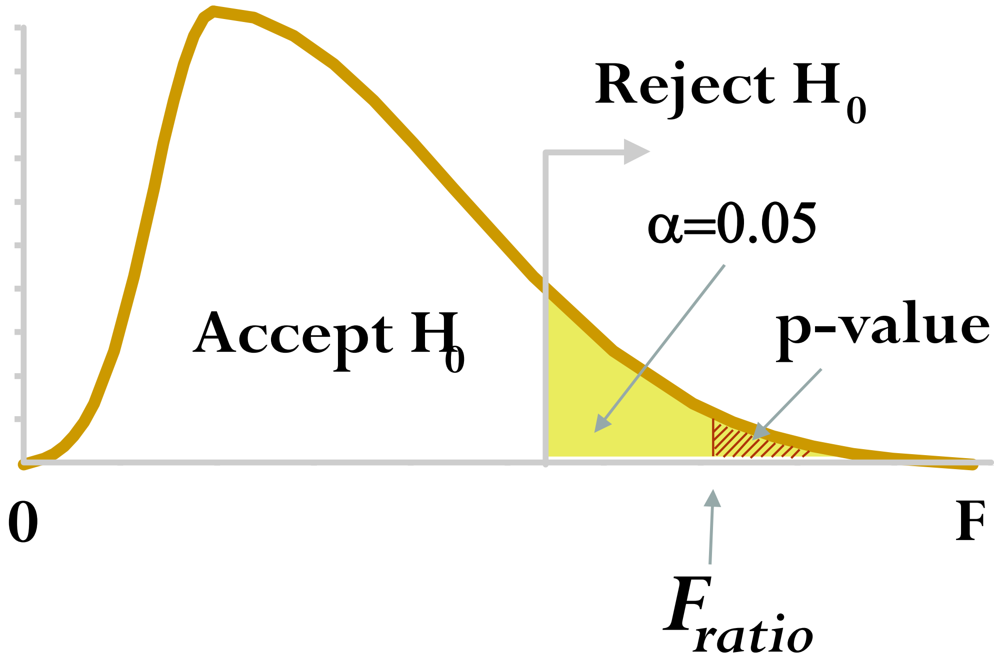
If **p-value < 0.05:** Reject H_0

The P-value approach

When we perform calculations by hand:



When we use R:



If $\mathbf{F_{ratio} < F_{table}}$: Accept H_0

If $\mathbf{F_{ratio} > F_{table}}$: Reject H_0

If **p-value > 0.05 :** Accept H_0

If **p-value < 0.05:** Reject H_0

$F_{ratio} = 5.94$

$F_{table} = 3.35$

$F_{ratio} > F_{table} \rightarrow$ Reject H_0

p-value = 0.007

p-value < 0.05 $\rightarrow$ Reject H_0

In-class exercise

The manager of an electronics store wants to improve the selling strategy of his shop.

In order to do that, he needs to find out how five cell phones models differ in the age of their buyers.

μ_1 μ_2 μ_3 μ_4 μ_5

9 buyers of each cell phone respond to a questionnaire, that includes a question about their age.

Based on the data collected, we carry out an ANOVA and obtain the following table:

ANOVA				
	SS	df	MS	Fratio
Between Groups	8223.2	4	2055.8	36.4
Within Groups	2259.1	40	56.48	
Total	10482.3	44		

Is there a significant difference in the age of buyers of different cell phones?

A. Since the F_{table} = 5.72 (at α=0.05), then we conclude that there is not a significant difference in the age of buyers of different cell phones.

B. Since the F_{table} = 5.72 (at α=0.05), then we conclude that there is a significant difference in the age of buyers of different cell phones.

C. Since the F_{table} = 2.61 (at α=0.05), then we conclude that there is not a significant difference in the age of buyers of different cell phones.

D. Since the F_{table} = 2.61 (at α=0.05), then we conclude that there is a significant difference in the age of buyers of different cell phones.

We then obtain the LSD intervals:

	Age_customers	LSD Lower limit	LSD Upper limit
Cell_phone_1	21.22	16.16	26.29
Cell_phone_2	32.11	27.05	37.17
Cell_phone_3	23.22	18.16	28.29
Cell_phone_4	41.55	36.49	46.62
Cell_phone_5	58.11	53.04	63.17

What can you conclude?

A. All cell phones differ from each other in the age of their buyers. Buyers of cell phone 5 are the oldest, followed (in decreasing order), by buyers of cell phone 4, 2, 3 and 1.

B. Buyers of cell phone 5 are the oldest. The second oldest are buyers of cell phones 2 and 4 (these two are not significantly different from each other). The youngest customers are those of cell phones 1 and 3 (these two are not significantly different from each other).

C. Buyers of cell phone 5 and 4 are the oldest (these two are not significantly different from each other). The second oldest are buyers of cell phone 2. The youngest customers are those of cell phones 1 and 3 (these two are not significantly different from each other).

D. None of the previous is correct.