

⚠ Lagunita is retiring and will shut down at 12 noon Pacific Time on March 31, 2020. A few courses may be open for self-enrollment for a limited time. We will continue to offer courses on other online learning platforms; visit <http://online.stanford.edu>.

Course > Inference: Relationships C→Q > ANOVA > ANOVA: Overview

🔖 Bookmark this page

ANOVA: Overview

Learning Objective: In a given context, carry out the inferential method for comparing groups and draw the appropriate conclusions.

Comparing More Than Two Means—ANOVA

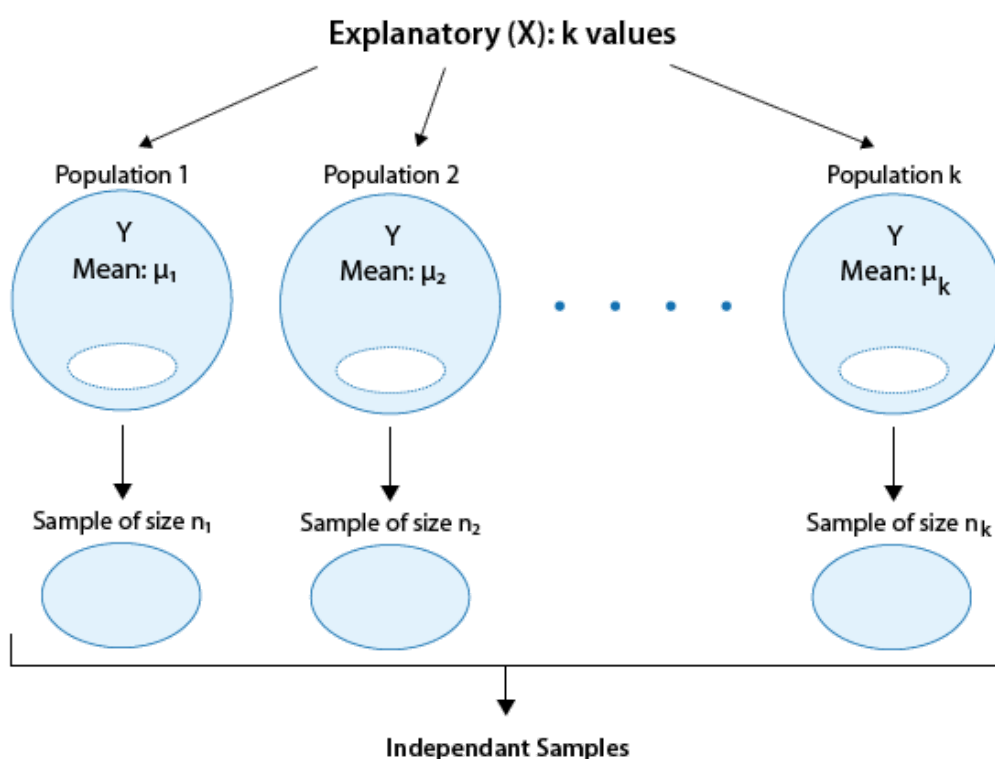
Overview

In this part, we continue to handle situations involving one categorical explanatory variable and one quantitative response variable, which is case C→Q in our role/type classification table:

		Response	
		Categorical	Quantitative
Explanatory	Categorical	C→C	C→Q
	Quantitative	Q→C	Q→Q

So far we have discussed the two samples and matched pairs designs, in which the categorical explanatory variable is two-valued. As we saw, in these cases, examining the relationship between the explanatory and the response variables amounts to comparing the mean of the response variable (Y) in two populations, which are defined by the two values of the explanatory variable (X). The difference between the two samples and matched pairs designs is that in the former, the two samples are independent, and in the latter, the samples are dependent.

We are now moving on to cases in which the categorical explanatory variable takes more than two values. Here, as in the two-valued case, making inferences about the relationship between the explanatory (X) and the response (Y) variables amounts to comparing the means of the response variable in the populations defined by the values of the explanatory variable, where the number of means we are comparing depends, of course, on the number of values of X. Unlike the two-valued case, where we looked at two sub-cases (1) when the samples are independent (two samples design) and (2) when the samples are dependent (matched pairs design, here, we are just going to discuss the case where the samples are independent. In other words, we are just going to extend the two samples design to more than two independent samples.



Comment

The extension of the matched pairs design to more than two dependent samples is called "Repeated Measures" and is beyond the scope of this course.

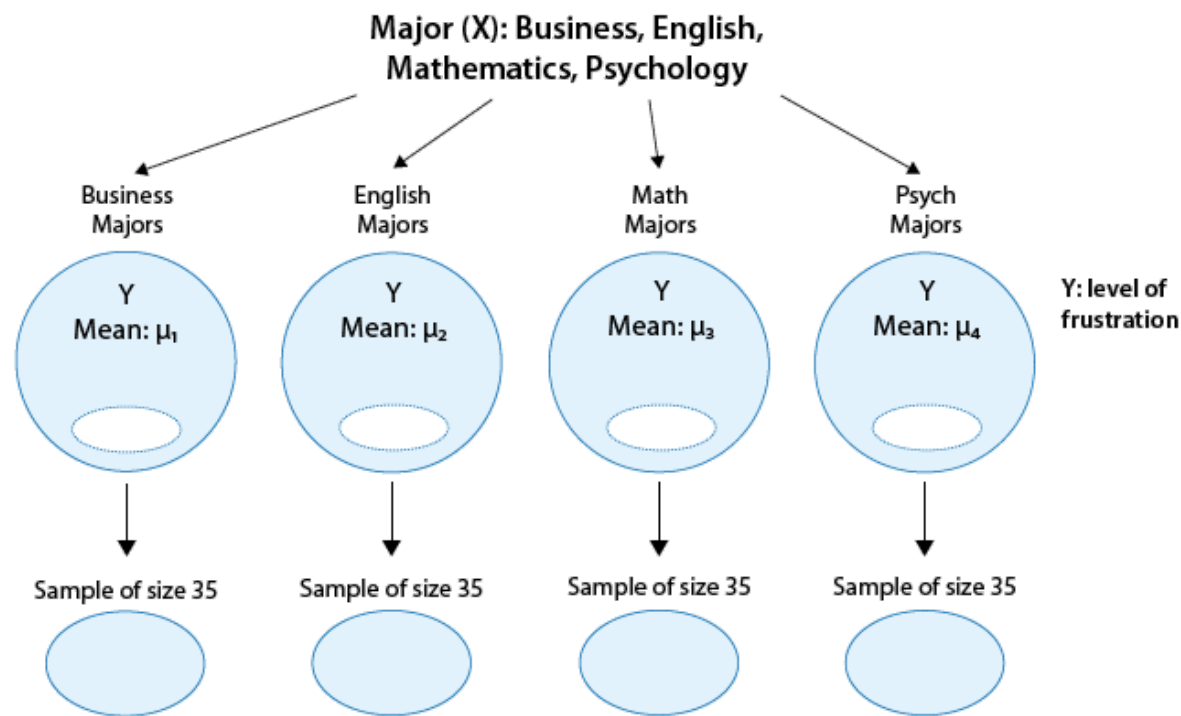
The inferential method for comparing more than two means that we will introduce in this part is called Analysis Of Variance (abbreviated as ANOVA), and the test associated with this method is called the ANOVA F-test. The structure of this part will be very similar to that of the previous two. We will first present our leading example, and then introduce the ANOVA F-test by going through its 4 steps, illustrating each one using the example. (It will become clear as we explain the idea behind the test where the name "Analysis of Variance" comes from.) We will then present another complete example, and conclude with some comments about possible follow-ups to the test. As usual, you'll have activities along the way to check your understanding, and learn how to use software to carry out the test.

Let's start by introducing our leading example.

Example

Is "academic frustration" related to major?

A college dean believes that students with different majors may experience different levels of academic frustration. Random samples of size 35 of Business, English, Mathematics, and Psychology majors are asked to rate their level of academic frustration on a scale of 1 (lowest) to 20 (highest).



The figure highlights what we have already mentioned: examining the relationship between major (X) and frustration level (Y) amounts to comparing the mean frustration levels ($\mu_1, \mu_2, \mu_3, \mu_4$) among the four majors defined by X. Also, the figure reminds us that we are dealing with a case where the samples are independent.

Comment

There are two ways to record data in the ANOVA setting:

- Unstacked: One column for each of the four majors, with each column listing the frustration levels reported by all sampled students in that major:

Business	English	Math	Psychology
11	11	9	11

6	9	16	19
6	14	11	13
etc.			

- Stacked: one column for all the frustration levels, and next to it a column to keep track of which major a student is in:

Frustration(Y)	Major(X)
9	Business
2	Business
9	Business
10	English
11	Psychology
13	English
13	Psychology
12	Math
etc.	

The "unstacked" format helps us to look at the four groups separately, while the "stacked" format helps us remember that there are, in fact, two variables involved: frustration level (the quantitative response variable) and major (the categorical explanatory variable).

Here are the first 50 cases that were loaded into a statistics software package:

	Business	English	Mathematics	Psychology	Frustration Score	Major
1	11	11	9	11	11	Business
2	6	9	16	19	6	Business
3	6	14	11	13	6	Business
4	4	13	11	10	4	Business
5	6	9	12	14	6	Business
6	9	12	17	10	9	Business
7	8	10	12	13	8	Business

8	3	12	14	14	3	Business
9	11	9	20	15	11	Business
10	12	11	12	17	12	Business
11	7	15	14	17	7	Business
12	5	13	15	14	5	Business
13	2	10	14	13	2	Business
14	11	10	14	9	12	Business
15	13	15	12	19	13	Business
16	9	11	15	17	9	Business
17	9	13	14	17	9	Business
18	9	9	12	11	9	Business
19	5	15	14	13	5	Business
20	8	17	14	19	8	Business
21	6	12	14	8	6	Business
22	3	10	14	20	3	Business
23	5	12	9	15	5	Business
24	10	9	14	14	10	Business
25	9	8	15	15	9	Business
26	6	14	12	15	6	Business
27	4	13	17	9	4	Business
28	8	13	13	16	8	Business
29	10	13	13	16	10	Business
30	2	11	11	11	2	Business
31	8	12	15	14	8	Business
32	4	11	14	13	4	Business
33	9	13	9	11	9	Business
34	9	12	15	16	9	Business

35	9	11	16	13	9	Business
36	NA	NA	NA	NA	11	English
37	NA	NA	NA	NA	9	English
38	NA	NA	NA	NA	14	English
39	NA	NA	NA	NA	13	English
40	NA	NA	NA	NA	9	English
41	NA	NA	NA	NA	12	English
42	NA	NA	NA	NA	10	English
43	NA	NA	NA	NA	12	English
44	NA	NA	NA	NA	9	English
45	NA	NA	NA	NA	11	English
46	NA	NA	NA	NA	15	English
47	NA	NA	NA	NA	13	English
48	NA	NA	NA	NA	10	English
49	NA	NA	NA	NA	10	English
50	NA	NA	NA	NA	15	English

Note that in the first 4 columns, the data are in unstacked format, and in the next two columns the data are stacked.

Open Learning Initiative [↗](#)



[↗](#) Unless otherwise noted this work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License [↗](#).

© All Rights Reserved