

The ANOVA for Dependent Groups -- Analysis of k-Within (or Matched)-Group Data with a Quantitative Response Variable

Application: This statistic has two applications that can appear very different, but are really just two variations of the same statistical question. In one application the same quantitative variable is measured under two or more different conditions from the same sample (or from two or more samples that have been matched on some important variable). In the other application, two or more comparable quantitative variables are measured from the same sample (usually at the same time). In both applications, the dependent ANOVA is used to compare the means of the quantitative variables. Thus, ANOVA is very much like a t-test, except that it can be applied to either 2 groups or to larger designs.

There are two specific versions of the H_0 ; depending upon whether one characterizes the data as representing a single population under two or more different conditions (e.g., comparing treated vs. not treated or comparing different treatments -- some consider this a representation of two or more different populations) or as representing comparable variables measured from a single population (as in the example below). Here are versions of the H_0 : statement for each of these characterizations.

H_0 : The population represented by the sample has the same mean on the quantitative variable under the "k" conditions.

To reject H_0 : is to say that the population represented by the sample has different means under the "k" conditions.

H_0 : The population represented by the sample has the same mean on the "k" different quantitative measures.

To reject H_0 : is to say that the means of the "k" variables are different in the population represented by the sample.

The data: In this analysis (which corresponds to the second application described above) the quantitative variable is the number of each of three types of animals (fish, reptiles, or mammals). From our database, we use three variables **reptnum** (number of reptiles on display), **fishnum** (number of fish on display) and **mamlnum** (number of mammals on display). These scores are shown for the 12 stores below (reptnum, fishnum, mamlnum).

12,32,34 14,41,38 15,31,45 12,38,32 7,21,12 4,13,11
10,17,22 4,22,9 14,24,20 4,11,8 5,17,19 10,20,8

Research Hypothesis: The data come from the Pet shop database. The researcher hypothesized that stores would tend to display more fish than other types of animals, fewer reptiles, and an intermediate number of mammals.

H_0 for this analysis : Pet stores display the same mean number of reptiles, fish and mammals.

Step 1 Rearrange the data so that scores from each participant are in the appropriate columns, one for each condition.

	reptnum (k_1)	fishnum (k_2)	mamlnum (k_3)
	X	X	X
	12	32	34
	14	41	38
15		31	45
12		38	32
7	21		12
4	13		11
10		17	22
4	22		9
14		24	20
4	11		8
5	17		19
10		20	8

Step 2 Compute the subject total for each participant by adding together their scores from the different conditions or variables. Then compute the square of this total.

reptnum (k_1)		fishnum (k_2)		mamlnum (k_3)		Subject total	
X		X		X		S	S ²
12		32		34		78	6084
14		41		38		93	8649
15		31		45		91	8281
12		38		32		82	6742
7		21		12		40	1600
4		13		11		28	784
10		17		22		49	2401
4		22		9		35	1225
14		24		20		58	3364
4		11		8		23	529
5		17		19		41	1681
10		20		8		38	1444

Step 3 Compute the square of each score (but not the Subject Total) and place it in an adjacent column.

reptnum (k_1)		fishnum (k_2)		mamlnum (k_3)		Subject total	
X	X ²	X	X ²	X	X ²	S	S ²
12	144	32	1024	34	1156	78	6084
14	196	41	1681	38	1444	93	8649
15	225	31	961	45	2025	91	8281
12	144	38	1444	32	1024	82	6742
7	49	21	441	12	144	40	1600
4	16	13	169	11	121	28	784
10	100	17	289	22	484	49	2401
4	16	22	484	9	81	35	1225
14	196	24	576	20	400	58	3364
4	16	11	121	8	64	23	529
5	25	17	289	19	361	41	1681
10	100	20	400	8	64	38	1444

Step 4 Compute ΣX and ΣX^2 for each condition or variable and the ΣS^2 (the sum of the squared Subject totals).

	reptnum (k_1)	fishnum (k_2)	mamlnum (k_3)
ΣX	111	287	258
ΣX^2	1227	7879	7368

$$\Sigma S^2 = 42766$$

Step 5 Compute SS_{Total} -- Please note: N = the total number of data values (not the number of participants)

$$\begin{aligned}
 SS_{\text{Total}} &= \frac{(\sum X_{k1}^2 + \sum X_{k2}^2 + \sum X_{k3}^2)}{N} - \frac{(\sum X_{k1} + \sum X_{k2} + \sum X_{k3})^2}{N} \\
 &= \frac{(1227 + 7879 + 7368)}{36} - \frac{(111 + 287 + 258)^2}{36} = \frac{16474}{36} - \frac{656^2}{36} \\
 &= \frac{430336}{36} - \frac{42766}{3} = 11953.78 - 14277.78 = 4520.22
 \end{aligned}$$

Step 6 Compute SS_{Cond} (also called SS_{Effect} or SS_{IV} for experimental data) -- Please note: n = the number of participants (data values in each condition). Also, you should notice that the “right side” of this step is the same as for Step 5.

$$\begin{aligned}
 SS_{\text{Cond}} &= \frac{(\sum X_{k1})^2}{n} + \frac{(\sum X_{k2})^2}{n} + \frac{(\sum X_{k3})^2}{n} - \frac{(\sum X_{k1} + \sum X_{k2} + \sum X_{k3})^2}{N} \\
 &= \frac{111^2}{12} + \frac{287^2}{12} + \frac{258^2}{12} - \frac{(111 + 287 + 258)^2}{36} = \\
 &= \frac{12321 + 82369 + 66564}{12} - \frac{42766}{3} = 161254 - 14277.78 = 146976.22
 \end{aligned}$$

Step 7 Compute SS_{subject} You should notice that the “right side” of this step is the same as in Steps 5 & 6.

$$\begin{aligned}
 SS_{\text{subject}} &= \frac{\sum S^2}{k} - \frac{(\sum X_{k1} + \sum X_{k2} + \sum X_{k3})^2}{N} = \frac{42766}{3} - 11953.78 \\
 &= 14255.33 - 11953.78 = 2301.55
 \end{aligned}$$

Step 8 Compute SS_{Error} (also called SS_{Residual})

$$SS_{\text{Error}} = SS_{\text{Total}} - SS_{\text{Cond}} - SS_{\text{Subject}} = 4520.23 - 1484.05 - 2301.55 = 734.63$$

Step 9 Compute df_{Cond} (also called df_a)

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

Step 10 Compute MS_{Cond}

$$MS_{\text{Cond}} = \frac{SS_{\text{IV}}}{df_{\text{IV}}} = \frac{1484.05}{2} = 742.03$$

Step 11 Compute df_{Total} Remember, N = total number of data values.

$$df_{\text{Total}} = N - 1 = 36 - 1 = 35$$

Step 12 Compute df_{Subjects} Remember, n = number of participants

$$df_{\text{Subjects}} = n - 1 = 12 - 1 = 11$$

Step 13 Compute df_{Error} (also called df_{Residual})

$$df_{\text{Error}} = df_{\text{Total}} - df_{\text{Cond}} - df_{\text{Subjects}} = 35 - 2 - 11 = 22$$

Step 14 Compute MS_{Error} (also called MS_{Residual})

$$MS_{\text{Error}} = \frac{SS_{\text{Error}}}{df_{\text{Error}}} = \frac{734.63}{22} = 33.39$$

Step 15 Compute the omnibus F (also called the obtained F)

$$F = \frac{MS_{\text{Cond}}}{MS_{\text{Error}}} = \frac{742.03}{33.39} = 22.22$$

Step 16 Use Table F to determine the critical value of F for $\alpha = .05$ and the appropriate degrees of freedom

$$\begin{aligned} \text{numerator degrees of freedom} &= df_{\text{Cond}} = 2 \text{ and denominator degrees of} \\ \text{freedom} &= df_{\text{Error}} = 22, \quad F(2, 22, \alpha = .05) = 3.44 \end{aligned}$$

Step 17 Compare the obtained F and critical F , and determine whether to reject or retain the null hypothesis.

-- if the obtained F is less than the critical F , then retain the null hypothesis -- conclude that the population represented by the sample has the same mean under the different conditions (or have the same mean on the different quantitative variables)

-- if the obtained F is greater than the critical F , then reject the null hypothesis -- conclude that the population represented by sample has different means under the different conditions (or have have different mean scores on the different quantitative variables)

For the example data, we would decide to reject the null hypothesis, because the obtained F value of 22.22 is larger than the critical F value of 3.44.

Step 18 IF you reject the null hypothesis, determine whether the pattern of the mean differences completely supports, partially supports, or does not support the research hypothesis.

- IF you reject the null hypothesis, AND if the pattern of means agrees exactly with the research hypothesis, then the research hypothesis is completely supported.
- IF you reject the null hypothesis, AND if part of the pattern of means agrees with part of, but not all of the research hypothesis, then the research hypothesis is partially supported.
- IF you retain the null hypothesis, OR if you reject the null BUT no part of the pattern of means agrees with the research hypothesis, then the research hypothesis is not at all supported.

By the Way: To properly determine if the hypothesized pattern of mean differences was found, one should perform pairwise comparisons (using one of the procedures described in the next section); the report of the results given below are based on informal "inspection" of the means.

By the Way: Usually the researcher hypothesizes that there is a difference between the conditions (or quantitative variables). Sometimes, however, the research hypothesis is that there is NO difference between the conditions. If so, the research hypothesis and H0: are the same! When this is the case, retaining H0: provides support for the research hypothesis, whereas rejecting H0: provides evidence that research hypothesis is incorrect.

The researcher hypothesized that stores would tend to display more fish than other types of animals, fewer reptiles, and an intermediate number of mammals. This "translates" into an expected pairwise difference between all pairs of animal types fish > mammals > reptiles (with the implication that fish > reptiles).

Based on inspection of the means, we would conclude there is only PARTIAL SUPPORT for the research hypothesis. More fish than reptiles were displayed and also more mammals than reptiles were displayed. However, contrary to the research hypothesis, there was not much of a difference between the average number of fish and mammals displayed.

Step 19 Preparing the Summary Table (Also called a Source Table)

The Summary Table for a dependent groups design looks a bit different than the one for independent groups ANOVA that you learned about last time, because we have to add a row for the "subjects" source of variance. The values come from the ANOVA example. The table on the left shows which steps in the analysis produced each value, while the one on the right shows the actual values. df = degrees of freedom, SS = sum of squares, and MS = mean squares

Computational steps												
Source	df	SS	MS	F	p		Source	df	SS	MS	F	p
Cond	9	6	10	15	17		Cond	2	1484.05	742.03	22.22	< .05
Subj	12	7					Subj	11	2301.55			
Error	13	8	14				Error	22	734.63	33.39		
Total	11	5					Total	35	4520.22			

Step 20 Describing the results of the Dependent Groups Analysis of Variance-- be sure to include the following

- Name the quantitative variables and tell the mean and standard deviation for each
- The F-value, df (in parentheses) and p-value ($p < .05$ or $p > .05$).
- Describe the pattern of the data (which group(s) has the larger mean, if there is a significant difference)
- Whether or not the results support the research hypothesis

Please remember, this description of the data is based upon inspection of numerical differences among the means and not the formal "follow-up" analyses you will learn about later.

There was a significant difference among the means of the three types of animals, $F(2,22) = 22.22$, $p < .05$, $MSE = 33.39$. Inspection of the means revealed that, consistent with the research hypothesis, more fish ($M = 23.92$, $SD = 9.61$) than reptiles ($M = 9.25$, $SD = 4.27$) were displayed on average and also more mammals ($M = 21.50$, $SD = 12.87$) than reptiles were displayed on average. However, contrary to the research hypothesis, there was not much of a difference between the average number of fish and mammals displayed.

Here is the same thing using a Table to present the univariate statistics.

Table 1 summarizes the data for the numbers of animals displayed at the stores. There was a significant difference among the means of the three types of animals, $F(2,22) = 22.22$, $p < .05$, $MSE = 33.39$. Inspection of the means revealed that, consistent with the research hypothesis, more fish than reptiles were displayed on average and also more mammals than reptiles were displayed on average. However, contrary to the research hypothesis, there was not much of a difference between the average number of fish and mammals displayed.

Table 1
Number of animals of each type displayed in the pet stores.

	Type of Animal		
	Fish	Mammals	Reptiles
<u>M</u>	23.92	21.50	9.25
<u>SD</u>	9.61	12.87	4.27

Table F: ANOVA (F-tests) Critical values of F for $\alpha = .05$ & $\alpha = .01$

Denominator		Numerator df						
df	α	1	2	3	4	5	6	
1	.05	161	200	216	225	230	234	
2	.05	18.5	19.0	19.2	19.2	19.3	19.3	
	.01	98.5	99.0	99.2	99.2	99.3	99.3	
3	.05	10.1	9.55	9.28	9.12	9.01	8.94	
	.01	34.1	30.8	29.5	28.7	28.2	27.9	
4	.05	7.71	6.94	6.59	6.39	6.26	6.16	
	.01	21.2	18.0	16.7	16.0	15.5	15.2	
5	.05	6.61	5.79	5.41	5.19	5.05	4.95	
	.01	16.3	13.3	12.1	11.4	11.0	10.7	
6	.05	5.99	5.14	4.76	4.53	4.39	4.28	
	.01	13.7	10.9	9.78	9.15	8.75	8.47	
7	.05	5.59	4.74	4.35	4.12	3.97	3.87	
	.01	12.2	9.55	8.45	7.85	7.46	7.19	
8	.05	5.32	4.46	4.07	3.84	3.69	3.58	
	.01	11.3	8.65	7.59	7.01	6.63	6.37	
9	.05	5.12	4.26	3.86	3.63	3.48	3.37	
	.01	10.6	8.02	6.99	6.42	6.06	5.80	
10	.05	4.96	4.10	3.71	3.48	3.33	3.22	
	.01	10.0	7.56	6.55	5.99	5.64	5.39	
11	.05	4.84	3.98	3.59	3.36	3.20	3.09	
	.01	9.65	7.21	6.22	5.67	5.32	5.07	
12	.05	4.75	3.89	3.49	3.26	3.11	3.00	
	.01	9.33	6.93	5.95	5.41	5.06	4.82	
13	.05	4.67	3.81	3.41	3.18	3.03	2.92	
	.01	9.07	6.70	5.74	5.21	4.86	4.62	
14	.05	4.60	3.74	3.34	3.11	2.96	2.85	
	.01	8.86	6.51	5.56	5.04	4.69	4.46	
15	.05	4.54	3.68	3.29	3.06	2.90	2.79	
	.01	8.68	6.36	5.42	4.89	4.56	4.32	
16	.05	4.49	3.63	3.24	3.01	2.85	2.74	
	.01	8.53	6.23	5.29	4.77	4.44	4.20	
17	.05	4.45	3.59	3.20	2.96	2.81	2.70	
	.01	8.40	6.11	5.18	4.67	4.34	4.10	
18	.05	4.41	3.55	3.16	2.93	2.77	2.66	
	.01	8.29	6.01	5.09	4.58	4.25	4.01	

Numerator df		Denominator						
df	α	1	2	3	4	5	6	
19	.05	4.38	3.52	3.13	2.90	2.74	2.63	
	.01	8.18	5.93	5.01	4.50	4.17	3.94	
20	.05	4.35	3.49	3.10	2.87	2.71	2.60	
	.01	8.10	5.85	4.94	4.43	4.10	3.87	
22	.05	4.30	3.44	3.05	2.82	2.66	2.55	
	.01	7.95	5.72	4.82	4.31	3.99	3.76	
24	.05	4.26	3.40	3.01	2.78	2.62	2.51	
	.01	7.82	5.61	4.72	4.22	3.90	3.67	
26	.05	4.23	3.37	2.98	2.74	2.59	2.47	
	.01	7.72	5.53	4.64	4.14	3.82	3.59	
28	.05	4.20	3.34	2.95	2.71	2.56	2.45	
	.01	7.64	5.45	4.57	4.07	3.75	3.53	
30	.05	4.17	3.32	2.92	2.69	2.53	2.42	
	.01	7.56	5.39	4.51	4.02	3.70	3.47	
40	.05	4.08	3.23	2.84	2.61	2.45	2.34	
	.01	7.31	5.18	4.31	3.83	3.51	3.29	
60	.05	4.00	3.15	2.76	2.53	2.37	2.25	
	.01	7.08	4.98	4.13	3.65	3.34	3.12	
120	.05	3.92	3.07	2.68	2.45	2.29	2.17	
	.01	6.85	4.79	3.95	3.48	3.17	2.96	
200	.05	3.89	3.04	2.65	2.42	2.26	2.14	
	.01	6.76	4.71	3.88	3.41	3.11	2.89	
∞	.05	3.84	3.00	2.60	2.37	2.21	2.10	
	.01	6.65	4.61	3.78	3.32	3.02	2.80	