

11 ANOVA: testing many means at once

This section is about a single test that asks whether all of a large number (well, more than two) of samples appear to have the same mean. I'll introduce it by way of an example.

Example 11.1 *A barmaid at the Bull's Head is working to pay her way through university. To get a little intellectual exercise while pulling pints, she keeps track of the drinking habits of the various UMIST departments whose staff drink in the pub. Here are her findings from the previous week:*

Physics	Maths.	Optometry
21	6	8
16	10	6
17	13	4
18	13	5
	4	2
	8	

Each entry represents a single individual's intake (in units of alcohol) over the course of a week and she wants to know whether these differences in drinking patterns are significant.

She could use some of the tests we already know— t -tests or Mann-Whitney (rank sum) tests, for example—on each of the pairs (Maths & Physics, Physics & Optometry, Maths & Optometry), but that requires a lot of tests. And the problem would only get worse if she started to think about Paper Science, Textiles and Chemical Engineering as well (6 groups would require 15 pairwise tests).

Instead she uses a single test, the Analysis of Variance (or ANOVA), that compares all the groups at once. The idea is to ask whether the variation between groups is large compared to the variation within them.

Recipe 11.1 (ANOVA) The ingredients are several lists of numbers, say r lists. The null hypothesis is that all of the numbers are drawn from the same (normal) distribution.

The data are usually arranged in columns, as in the table above:

Column 1	Column 2	...	Column r
$x_{1,1}$	$x_{1,2}$...	$x_{1,r}$
$x_{2,1}$	$x_{2,2}$...	$x_{2,r}$
\vdots	\vdots		\vdots
$x_{n_1,1}$	\vdots		\vdots
	\vdots		$x_{n_r,r}$
	$x_{n_2,2}$		

Notice that the lists need not have the same number of members; list j is the column $\{x_{1,j}, x_{2,j}, \dots, x_{n_j,j}\}$. It will prove helpful to have a symbol, N , for the total number of entries in the table:

$$N = \sum_{j=1}^r n_j.$$

Then one does the following.

- a) Compute the mean of each column separately:

$$m_j = \frac{\sum_{k=1}^{n_j} x_{k,j}}{n_j}.$$

- b) Compute the mean of all the entries in the table

$$\begin{aligned} m &= \frac{\sum_{j=1}^r \left[\sum_{k=1}^{n_j} x_{k,j} \right]}{N} \\ &= \frac{\sum_{j=1}^r n_j m_j}{N}. \end{aligned}$$

- c) Compute a measure of the variation *between* the columns. It is called SS_b , the “sum of squares between columns”, and defined by

$$SS_b = \sum_{j=1}^r n_j (m_j - m)^2.$$

- d) Compute a measure of the variation *within* the various columns. It is called SS_w , the “sum of squares within the columns”, and is given by

$$SS_w = \sum_{j=1}^r \left[\sum_{k=1}^{n_j} (x_{k,j} - m_j)^2 \right]$$

It’s messy, but not difficult, to show that the two sums-of-squares account for all the variation in the whole table:

$$SS_t = SS_b + SS_w = \sum_{j=1}^r \left[\sum_{k=1}^{n_j} (x_{k,j} - m)^2 \right]$$

Here the subscript ‘t’ stands for “total”; SS_t is the sum-of-squared-deviations from the mean for the entire table.

One then assembles these results into a standard table that looks like:

<i>Type of Variation</i>	<i>Sum of Squares</i>	<i>deg. of freedom</i>	<i>Average sum of squares (estimated variances)</i>
Between Samples	SS_b	$(r - 1)$	$SS_b/(r - 1)$
Within Samples	SS_w	$\sum_{j=1}^r (n_j - 1)$ $= (N - r)$	$SS_w/(N - r)$
Total	SS_t	$(N - 1)$	—

Finally, one does a test to see whether the mean variation between columns, $SS_b/(r - 1)$ is significantly bigger than the mean variation within columns $SS_w/(N - r)$. In the usual way one computes a statistic,

$$F = \frac{SS_b/(r - 1)}{SS_w/(N - r)},$$

then looks the value up (in special tables) to see if it is large enough to reject the null hypothesis. These tables depend on the number of degrees of freedom in the sums, so one consults the table for $F((r - 1), (N - r))$.

Putting the data above through the recipe gives $n_1 = 4$, $n_2 = 6$ and $n_3 = 5$, so $r = 3$ and $N = 15$. The steps of the recipe yield:

- a) The mean consumption of the physicists is 18 units per week, while that of the mathematicians is 9 and that of the optometrists is an abstemious 5.
- b) The mean of all the entries is ≈ 10.07 units per week.
- c) $SS_b \approx 386.9$
- d) $SS_w = 102$.

The ANOVA table is thus

<i>Type of Variation</i>	<i>Sum of Squares</i>	<i>deg. of freedom</i>	<i>Average sum of squares (estimated variances)</i>
Between Samples	386.9	2	193.5
Within Samples	102	12	8.5
Total	488.9	14	—

and the test stat. is $F \approx 22.8$. The attached table shows that if $F(2, 12)$ exceeds 3.89 we can reject the null with 95% confidence, so we can safely conclude that the members the various groups do not drink the same amount in the Bull's Head.