Analysis of variance (ANOVA)

Analysis of Variance (often abbreviated to ANOVA) is a technique for comparing the mean levels of response arising in several samples. Where just two samples are involved it might be appropriate to use a two-sample t-test or a paired t-test.

In the case of several samples, we might consider using t-tests to compare each pair of samples. However, this is a poor approach for the following reasons:

- It can be very time consuming;
- It can be misleading.

What is needed is a single test of the extended null hypothesis

$$H_0 : \mu_1 = \mu_2 = \mu_3 \cdots = \mu_t$$

where t is the number of samples, i.e. no differences among the population means. This is what ANOVA sets out to do.

Here is a sample markdown image:

Why do observations vary?

There are two reason why observations vary:

- Variation in response, from group to group;
- Variation between individuals within groups.

We can break down (explain) the total variability of responses into these two components. We calculate variance estimates for each source and compare them using a variance-ratio test (or F-test).

The sums of squares can be calculated as follows,

$$
\begin{aligned}
TSS &= \Sigma\Sigma\left(x^2\right) - \frac{G^2}{N} \\
BTSS &= \Sigma\left(\frac{T_i^2}{n_i}\right) - \frac{G^2}{N} \\
RSS &= TSS - BTSS,
\end{aligned}
$$

where N is the number of observations, $G = \Sigma\Sigma x$ is the sum of all of the observations, T_i is the sum of observations in group i and n_i is the number of observations in group i.

ANOVA

The resulting variance estimates (referred to in ANOVA as mean squares) can be compared using the variance-ratio test (F-test).

The standard Analysis of Variance is a parametric test and is therefore dependent (like the t-test) on the assumptions that the observations are Normally distributed, independent and have a common standard deviation σ (often abbreviated to Normality, Independence, Homogeneity).

Example

In an educational experiment on teaching children to read, the children were divided into three groups and listened to while they read. The teacher was instructed to respond as follows: for Group A, praise as much as possible and do not criticise; for Group B, criticise errors and do not praise; for Group C, comment as little as possible. The children were assessed at the start and end of the experiment and a response measurement of the degree of improvement (0-50 scale) was recorded.

A (praised)	46	35	34	39	43	40	47	39	37
B (criticised)	37	31	40	28	36	39	42	26	36
C (ignored)	34	26	37	34	32	31	35	40	28

We wish to test whether the different methods have made any difference to the mean levels of response.

R Code

```
Response = c(46,35,34,39,43,40,47,39,37,37,31,40,28,36,39,
             42,26,36,34,26,37,34,32,31,35,40,28)
Group = factor(c('A','A','A','A','A','A','A','A','A',
                 'B','B','B','B','B','B','B','B','B',
                 'C','C','C','C','C','C','C','C','C'))
df = data.frame(Response, Group)
fit <- aov(Response ~ Group, data=df)
summary(fit)
```