The concept

We are familiar with variance. $S^2 = \frac{\sum_{i=1}^{n} (Y_i - \bar{Y})^2}{n-1} = \frac{SS}{d.f.}$

We are familiar with the pooled variance $S_p^2 = \frac{\gamma_1 S_1^2 + \gamma_2 S_2^2}{\gamma_1 + \gamma_2} = \frac{SS_1 + SS_2}{(n_1 - 1) + (n_2 - 1)}$

We are familiar with the variance of the means. But we never get "multiple" estimates of the mean and calculate a variance from those. The calculation we use to get the variance of the means comes from statistical theory, $S_{\bar{Y}}^2 = \frac{S^2}{n}$. Could we actually get multiple estimates of the means and calculate a sum of squared deviations of the various means from an overall mean and get variance of the means from that?

Yes, we could, and using the formula $S_{\bar{Y}}^2 = \frac{S^2}{n} = \frac{\sum_{i=1}^{k} (\bar{Y}_{i.} - \bar{\bar{Y}}_{..})^2}{k-1}$ should give the same value.

Suppose we have some values from a number of different samples, perhaps taken at different sites. The values would be Y_{ij}, where the sites are i=1, 2, ..., k, and the observations from within the sites are j = 1, 2, 3, ..., n_i. For each site we calculate a value of the mean. We then take the various means (k different means) and calculate a variance among those. This would also give the "variance of the means".

The LOGIC

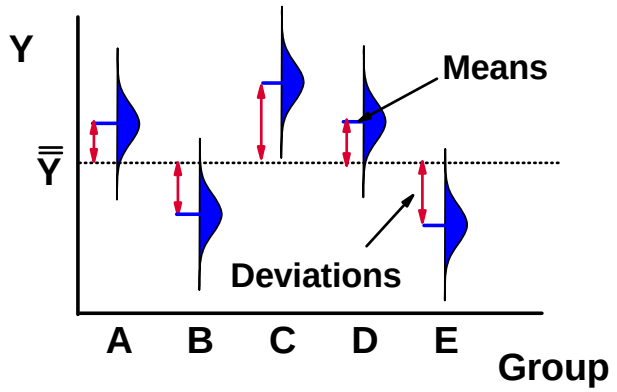
Remember, we want to test

H_0: $\mu_1 = \mu_2 = \mu_3 = ... = \mu_k$

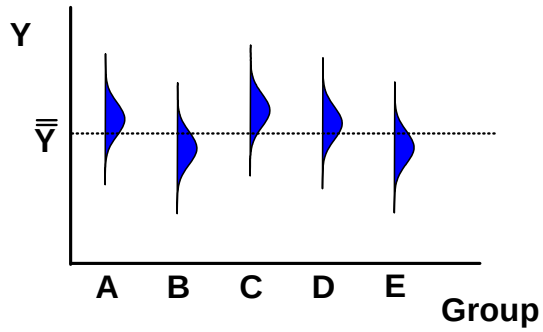
We have a bunch of means and we want to know if they were drawn from the same population or different populations. We also have a bunch of samples each with its own variance (S^2). If we can assume homogeneous variance (all variances equal) then we could POOL the multiple estimates of variance. So, to start with we will take the variances from each of the groups and pool them into one new & improved estimate of variance. This will be the very best estimate of variance that we will get (if the assumption is met).

$$S_p^2 = \frac{SS_1 + SS_2 + SS_3 + SS_4 + SS_5}{(n_1 - 1) + (n_2 - 1) + (n_3 - 1) + (n_4 - 1) + (n_5 - 1)}$$

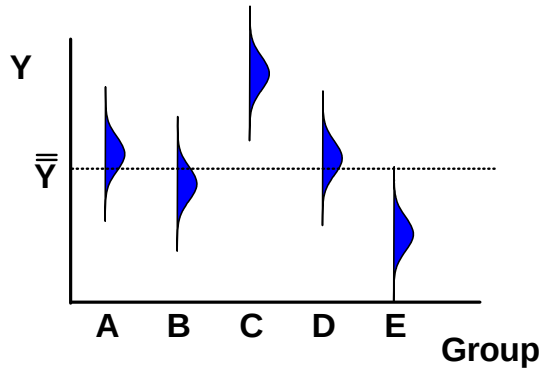
Now, think about the means. If the NULL HYPOTHESIS IS TRUE, then we could calculate the variance of the means from the multiple means. This would estimate $S_{\bar{Y}}^2$, the variance of the means. We would take the deviations of each $\bar{Y}_{i.}$ from the overall mean, $\bar{\bar{Y}}_{..}$, and get a variance from that.



If the null hypothesis is true, the means should be pretty close to the overall mean. They won't be exactly equal to the overall mean because of random sampling variation in the individual observations.



However, if the null hypothesis is false, then some mean will be different! At least one, maybe several.



So we take the Sum of squared deviations, divide by the degrees of freedom and we get an estimate of the variance of the means, $S^2_{\bar{Y}} = \frac{\sum_{i=1}^{k}(\bar{Y}_{i.} - \bar{\bar{Y}}_{..})^2}{k-1}$. But this does not exactly estimate the variance, it estimates the variance of the means, that is the variance divided by the sample size! The sample size is the number of observations in each mean. $S^2_{\bar{Y}} = \frac{\sum_{i=1}^{k}(\bar{Y}_{i.} - \bar{\bar{Y}}_{..})^2}{k-1} = \frac{S^2}{n}$. In order to estimate the variance we must multiply this estimate by *n*, the sample size, $nS^2_{\bar{Y}} = \frac{nS^2}{n} = S^2$, giving a second estimate of the variance. This is obviously easier if each sample size is the same (i. e. the experiment is balanced). We will usually use the calculations for a balanced design, but the analysis can readily be done if the data is not balanced. It's just a little more complicated.

The Solution

So what have we got?

One variance estimate that is pooled across all of the samples because the variances are equal (an assumption, sometimes testable). This is the best estimate of random error.

And another variance that should be the same IF the null hypothesis is TRUE.

The second mean (from the variances) may not be the same if the null hypothesis is false, depending on how great the departure from the null hypothesis. Not only will the second

variance from the mean not be the same, IT WILL BE LARGER! Why? Because when we are testing means for equality we will not consider rejecting if the means are too similar, only if they are too different and large differences in means yield large deviations which produce an overly large variance. So this will be a one tailed test.

And how to we go about testing these two variances for equality? Testing for equality of variances requires an F-test, of course.

If H_0: $\mu_1 = \mu_2 = \mu_3 = ... = \mu_k$ is true, then $S_p^2 = nS_{\bar{Y}}^2$

If H_1: some μ_i is different, then $S_p^2 < nS_{\bar{Y}}^2$

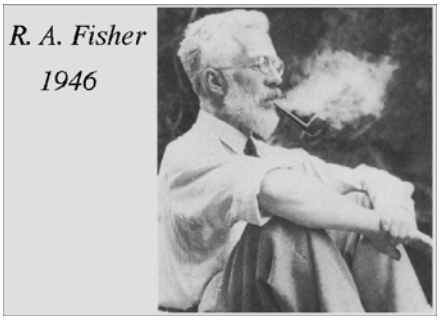
For a one tailed F test we put the ONE WE EXPECT TO BE LARGER IN THE NUMERATOR.

$$F = nS_{\bar{Y}}^2 \Big/ S_p^2$$

And that is Analysis of Variance.

We are actually testing means, but we are doing it by turning them into variances; one pooled variance from within the groups, called the "pooled within variance" and one variance from between groups or among groups called the "variance among groups" or "between group variance". If the variances are not significantly different as judged by the F test, then we cannot reject the null hypothesis. It is possible, as usual, that we make a Type II error with some unknown probability (β). If the variances are judged to not be the same, then the null hypothesis is probably not true. Of course we may have made a Type I error, with a known probability of α.

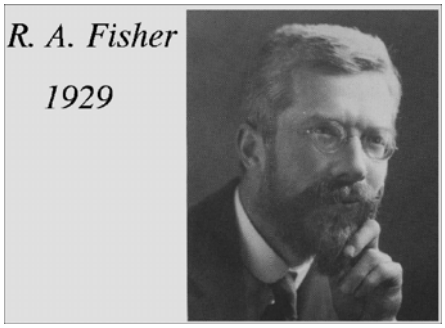
Some of the calculations later, but this is the basic idea.



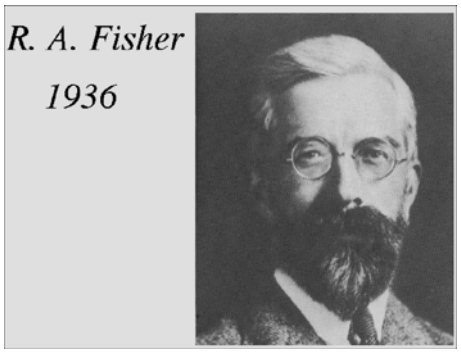
R. A. Fisher

Ronald Aylmer Fisher is sometimes called the father of modern statistics. Some of his major contributions include the development of the basics of design of experiments and Analysis of Variance.

Born in London 1890, he had very poor eyesight that prevented him from learning by electric light. He had to learn by having things read out to him. He developed the ability view problems geometrically and to figure mathematical equations in his head. In 1909 he won a scholarship to Cambridge.



He left an academic position teaching mathematics for a position at Rothamsted Agricultural Experiment Station. In this environment he developed many applied analyses for testing experimental hypotheses (Analysis of Variance, circa 1918), and provided much of the foundation for modern statistics.



We will see other analyses (in addition to ANOVA) developed by Fisher. Some other contributions by Fisher include the first use of the term "null hypothesis", development of the F distribution, of the Least Significant Difference, maximum likelihood estimation and contributed to the early use nonparametric statistics.

Terminology used in Analysis of Variance

Treatment – different experimental populations that are contained in an experiment and undergo some application or manipulation by the experimenter

Control or check – a "treatment" that receives no experimental manipulation

Experimental Unit – the unit to which a treatment is applied

Sampling Unit – the unit that is sampled or measured

The linear model is given by $Y_{ij} = \mu_i + \varepsilon_{ij}$ or $Y_{ij} = \mu + \tau_i + \varepsilon_{ij}$

where $\tau_i = (\mu_i - \mu_.)$ is estimated by $\hat{\tau}_i = (\bar{Y}_{i.} - \bar{Y}_{..})$

The calculation of treatment Sum of Squares for treatments is a sum of the squared treatment effects $SS_{Treatments} = n\sum_{i=1}^{t}(\bar{Y}_{i.} - \bar{Y}_{..})^2$.

The calculation of treatment Mean Square is a sum of squared effects divided by the degrees of freedom. A variance? $MS_{Treatments} = \frac{n\sum_{i=1}^{t}(\bar{Y}_{i.} - \bar{Y}_{..})^2}{t-1} = \frac{n\sum_{i=1}^{t}\tau_i^2}{t-1}$

A random treatment effect estimates a variance component. In order for treatments to be random, they should be a random selection from a large (theoretically ∞) number of treatments. Inferences developed from random treatments are for all the possible treatment levels.

Examples of random effects

The term used for the error in an experiment are always random. They represent random variation. This variation comes from the experimental unit and sometimes the sampling unit.

Compare production rice varieties, where rice varieties represent a random sample from the world's rice varieties.

Estimate the alcohol content of beer, where the beers tested are randomly sampled from all the beers in the population of interest (world, national).

Oxygen levels in bayous, where randomly selected bayous represent all bayous in the state.

A treatment is FIXED if all possible levels, or all levels of interest, are included in the experiment. The treatment levels are selected by the investigator and are probably not chosen from a very large number of possible values.

A fixed treatment estimates the sum of squared fixed effects for the treatments being investigated. This is NOT a variance, but the calculation is the same, $\sum_{i=1}^{t}\tau_i^2 \Big/ t-1$.

Examples of fixed effects

- Experiment includes all of the 7 rice varieties commonly grown in Louisiana
- Beers are limited to the 5 micro-breweries in Anchorage, Alaska.

There are some treatments that are common and typically fixed.

For example, indicator variables that include all possible levels of a treatment.

- Sex (male, female)
- Class (Freshman, Sophomore, Junior, Senior)

Ordinal treatments, where data are categorized as large, medium & small or as deep & shallow or as a high level & a low level.

Before and After

Treatments with a control group

If the treatments are fixed, then inferences are limited to the treatments included in the experiment. Comparing the individual treatment levels is often of interest, since they are often specifically chosen by the investigator. If the treatments are random inferences are made to the whole population that the sample was taken from. Individual "treatment levels" are usually not of interest.

Some treatments can be either fixed or random.

Years – do they represent random variation or do we categorize as "wet & dry"

Months – randomly selected, or do they represent seasons

Sites or locations – random chosen or selected for certain characterisitics?

Example

A new insulin preparation is being compared to an older standard and a saline control. Ten rabbits are administered the preparations and blood sugar is measured after 20 minutes.

What are the treatments? Experimental units? Sampling units? Are the treatments fixed or random?

CRD – the Completely Randomized Design

The analysis of variance we have seen is called the Completely Randomized Design (CRD) because the treatments are assigned to the experimental units completely at random. The analysis is also called a "one-way analysis of variance". Later we will discuss the Randomized Block Design (RBD).

Key aspects of the analysis.

Everything is important, but there are some aspects that I consider more important. These are discussed below.

The 7 steps of hypothesis testing: Understand particularly the hypothesis being tested and the assumptions we need to make to conduct a valid ANOVA.

Understand the tests of the assumptions, particularly the HOV tests and evaluation of normality, particularly Shapiro-Wilks.

Calculations: We will primarily do the ANOVA using SAS. However, it is important to understand that the calculations, as originally derived by Fisher, were based on the marginal totals or means, averaging or summing over all observations in the treatment. This will take on additional significance when we talk about two-way ANOVA.

The ANOVA table: Understand the table usually used to express the results of an Analysis of Variance. This same table will also be used for regression.

Traditional ANOVA table

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	4	838.5976	209.6494	15.38	0.0001
Error	20	272.6680	13.6334		
Corrected Total	24	1111.2656			

SEE SAS OUTPUT

Expected Mean Square

What do we estimate when we calculate a pooled variance estimate (MSE) or the sum of squared treatment (SSTreatments) effects divided by its d.f.?

The MSE estimates σ^2, the random variation for individuals in the population.

If the null hypothesis is true, the MS for Treatments also estimate the same random variation, σ^2. The F value should only reject the null hypothesis α*100% of the time.

But what if the null hypothesis is NOT true? Then, the MSTreatments estimates σ^2, PLUS some additional component due to a treatment effect.

For a random effect this additional component would be called σ_τ^2. This is a variance.

For a FIXED effect the additional component is simple the sum of squared effects divided by the d.f., $\Sigma\tau_i^2 \Big/ t-1$. This is not a variance component.

The ANOVA source table with its d.f. and Expected mean squares (for a balanced design).

Note: 1 tailed test, n influences power

Source	d.f.	EMS Random	EMS Fixed
Treatment	t–1	$\sigma_\varepsilon^2 + n\sigma_\tau^2$	$\sigma_\varepsilon^2 + n\Sigma\tau_i^2 \Big/ (t-1)$
Error	t(n–1)	σ_ε^2	σ_ε^2
Total	tn–1		

We could also express our null hypothesis in terms of EMS [H_0: $\sigma_\tau^2 = 0$], particularly for the random effect since the variance component for treatments may be a value of interest.

Since for a fixed effect the individual means are usually of interest, the null hypothesis is usually expressed in terms of the means (H_0: $\mu_1 = \mu_2 = \mu_3 = ... = \mu_t$).