Analysis of Variance

Recall: ANalysis Of VAriance (ANOVA) refers to statistical procedures designed to analyze experiments with quantitative responses from more than two groups. The goal is to decide whether or not the population means for those groups are equal. The characteristic that differentiates the groups is called the FACTOR and the different treatments are referred to as the LEVELS of that factor.

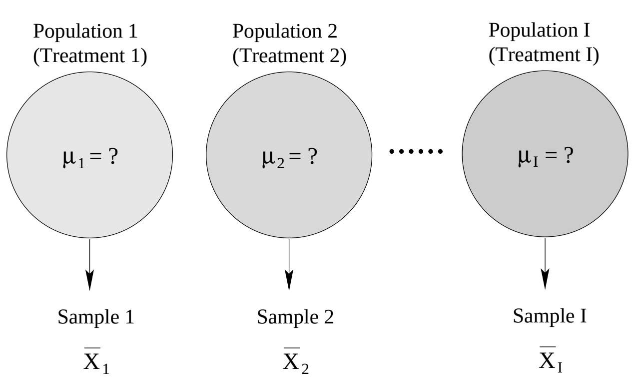
Example: For each of the following scenarios, name the factor, factor levels and the quantitative response.

- Compare the gasoline mileage (mpg) for ten Fords, Toyotas, and BMW.
- Compare the effect of a blood pressure lowering medication in patients from three different risk groups (healthy, pre-hypertension, hypertension).
- Allow groups of 20 students to take the same test within a different period of time (60, 90, 120 minutes).

Note: Single-factor ANOVA focuses on comparisons of more than two groups. If there are only two groups, you have learned how to use a t-test to decide whether there is a difference in population means.

Notation:

$$
\begin{aligned}
I &= \text{the number of groups being compared} \\
\mu_1 &= \text{the population mean of group 1} \\
\vdots &= \\
\mu_I &= \text{the population mean of group I}
\end{aligned}
$$



The Hypotheses

The null hypothesis tested in an ANOVA is

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_I$$

against the alternative

$$H_a : \text{ at least two of the } \mu_i\text{'s are different}$$

More Notation:

X_{ij} = the random variable representing the j^{th} measurement from group i.
x_{ij} = the actual measurement on the j^{th} individual in group i.

The dot notation: A dot in place of a subscript indicates that a sum was taken over all possible values of that subscript index. A horizontal bar (together with the dot notation) indicates that an average was taken, e.g.

$$\bar{X}_{i\cdot} = \frac{1}{J}\sum_{j=1}^{J} X_{ij}, \quad i = 1, \ldots, I, \qquad \bar{X}_{\cdot j} = \frac{1}{I}\sum_{i=1}^{I} X_{ij}, \quad j = 1, \ldots, J$$

The average over both indices is called the GRAND MEAN

$$\bar{X}_{\cdot\cdot} = \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} X_{ij}$$

The sample variances are denoted by

$$s_i^2 = \frac{1}{J-1}\sum_{j=1}^{J}\left(x_{ij} - \bar{x}_{i\cdot}\right)^2, \quad i = 1, \ldots, I$$

Example: Six samples from each of four types of cereal grain grown in a certain region were analyzed to determine thiamine content, resulting in the following data (μg/g).

	Wheat	Barley	Maize	Oats
	5.2	6.5	5.8	8.3
	4.5	8.0	4.7	6.1
	6.0	6.1	6.4	7.8
	6.1	7.5	4.9	7.0
	6.7	5.9	6.0	5.5
	5.8	5.6	5.2	7.2
$\bar{x}_{i\cdot}$				
s_i^2				

(a) Identify I, J and x_{24}.

(b) Compute $\bar{x}_{i\cdot}$ and s_i^2 for $i = 1, \ldots, I$ (write the results into the table).

(c) Compute the grand mean of the data set.

(d) Formulate the ANOVA null hypothesis and alternative in the context of this example.

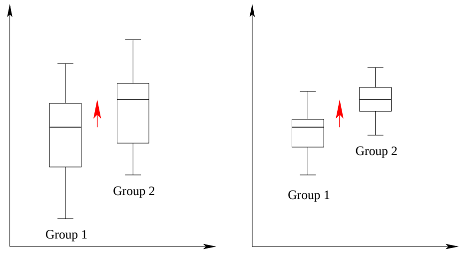
Assumptions: The I population or treatment distributions are all Normal with the same variance σ^2 (and possibly different means μ_i),

$$X_{ij} \sim \text{ Normal}(\mu_i, \sigma^2)$$

Remark: Do you think this assumption is reasonable in most practical applications? How can you check whether or not it is satisfied? What happens if it is not satisfied and you proceed with the ANOVA analysis anyway?

The Test Statistic

Big Idea: In an ANOVA we will compare the variation in means between the treatment sample groups to the variation within the sample groups. If the variation *between* groups is large compared to the variation *within* groups, then we can conclude that there is a significant difference in means (and reject the ANOVA null hypothesis). So we need a way to quantify this variation...



DEFINITION:

The MEAN SQUARE FOR TREATMENTS is defined as

$$\text{MSTr} = \frac{J}{I-1}\sum_{i=1}^{I}\left(\bar{X}_{i\cdot} - \bar{X}_{\cdot\cdot}\right)^2.$$

This quantity is used to measure the variation in means *between* the groups. It compares each groups mean to the overall grand mean.

The MEAN SQUARE FOR ERROR is defined as

$$\text{MSE} = \frac{S_1^2 + S_2^2 + \cdots + S_I^2}{I}.$$

This quantity represents the variation *within* the groups through each groups standard deviation S.

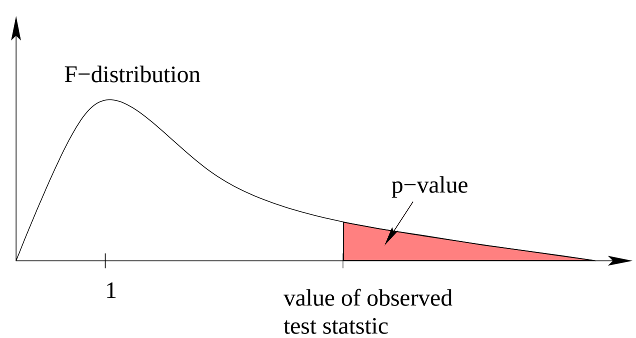
Test Statistic: For a single factor ANOVA, the test statistic is

$$F = \frac{\text{MSTr}}{\text{MSE}}.$$

If this quantity is large, i.e, when MSTr is much larger than MSE, we will reject H_0, and if this quantity is small, i.e., if MSTr $\simeq$ MSE, we fail to reject H_0.

The Distribution of the Test Statistic

Let F (defined as above) be the test statistic in a single-factor ANOVA problem with I populations and random samples of J individuals from each population. If the ANOVA assumptions are satisfied and H_0 is true, then F has an F-distribution with $\nu_1 = I - 1$ and $\nu_2 = I(J-1)$.



The null hypothesis of equal group means will be rejected for large values f of the test statistic F. Rejection region for significance level α: $f \geq F_{\alpha,I-1,I(J-1)}$. Even though some tables for F-distributions exist, the p-value is much more commonly found through software.

Example: (cont).

Refer to the thiamine in cereal grain example and use the values you have previously computed.

(a) Compute the value of the MSTr.

(b) Compute the value of the MSE.

(c) Compute the value of the test statistic F.

(d) Compare the value of the test statistic to $F_{0.05,3,20} = 3.1$ and draw a conclusion for the ANOVA hypothesis test.

ANOVA Tables

Usually, computations such as the ones in the previous example are not done by hand. Instead statistical computer programs, such as SPSS are used. These programs will produce numerical output in form of a table. While the format of the table may change slightly, depending on the program used, the entries remain fundamentally the same.

DEFINITIONS: Consider the following sums (not averages) of squares:

The total sum of squares (SST)

$$\text{SST} = \sum_{i=1}^{I}\sum_{j=1}^{J}(x_{ij} - \bar{x}_{..})^2 = \sum_{i=1}^{I}\sum_{j=1}^{J} x_{ij}^2 - \frac{1}{IJ}x_{..}^2$$

The treatment sum of squares (SSTr)

$$\text{SSTr} = \sum_{i=1}^{I}\sum_{j=1}^{J}(\bar{x}_{i\cdot} - \bar{x}_{..})^2 = \frac{1}{J}\sum_{i=1}^{I} x_{i\cdot}^2 - \frac{1}{IJ}x_{..}^2$$

The error sum of squares (SSE)

$$\text{SSE} = \sum_{i=1}^{I}\sum_{j=1}^{J}(x_{ij} - \bar{x}_{i\cdot})^2$$

FACT: SST = SSTr + SSE

DEFINITION: To convert a sum of squares into its respective *mean* square (in the sense of average square), divide by the associated degree of freedom.

$$\text{MSTr} = \frac{\text{SSTr}}{I-1}, \qquad \text{MSE} = \frac{\text{SSE}}{I(J-1)}, \qquad F = \frac{\text{MSTr}}{\text{MSE}}.$$

ANOVA Table

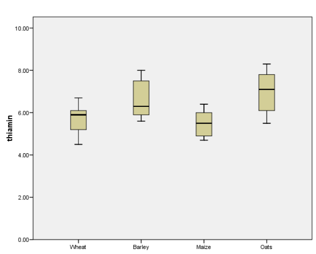
Source of Variation	**Sum of Squares**	**df**	**Mean Square**	f
Treatments	SSTr	$I-1$	MSTr = SSTr/(I-1)	MSTr/MSE
Error	SSE	$I(J-1)$	MSE = SSE/[I(J-1)]	
Total	SST	$IJ-1$		

Example: ANOVA table and boxplot for the cereal grain example as produced by SPSS:

ANOVA

thiamin

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	8.983	3	2.994	3.957	.023
Within Groups	15.137	20	.757		
Total	24.120	23			



Multiple Comparisons in ANOVA

When an F-test procedure for a single-factor ANOVA experiment rejects the null hypothesis this means that we conclude that the populations means are *not* all equal. In this case it is usually of interest to know which means actually differ. Because it would be necessary to look at all possible pairs of means, this further analysis is called a MULTIPLE COMPARISON PROCEDURE.

IDEA: Compute a confidence interval for each pairwise difference of means $\mu_i - \mu_j$. If an interval *does not* include zero, conclude that the two means are different. The trick here is to adjust the confidence level, so that it remains valid for all of the confidence intervals. This adjustment can be done in several ways, one of which is called TUKEY'S PROCEDURE (T-method).

Tukey's Procedure

Problem: If you compute a confidence interval at confidence level 95% and then another (unrelated) interval for a different parameter also at level 95%, then you are only 90.25% ($= 0.95^2 100\%$) certain that *both* intervals contain the true population parameter. Therefore, to compute simultaneous confidence intervals for all possible pairs of means in a single-factor ANOVA experiment, the confidence level in each computation needs to be adjusted.

PROPOSITION: Consider the simultaneous confidence intervals, each with confidence level $1 - \alpha$ defined by

$$\mu_i - \mu_j \in \left[\bar{X}_{i\cdot} - \bar{X}_{j\cdot} \pm Q_{\alpha,I,I(J-1)}\sqrt{\text{MSE}/J}\right]$$

$$\text{for } \textit{every } i \text{ and } j \in 1, \ldots, I \text{ with } i < j$$

where $Q_{\alpha,m,\nu}$ is the upper-tail α critical value of the Studentized range distribution (available in tables, or through software).

Then we can say with confidence $1 - \alpha$ that the true population mean differences $\mu_i - \mu_j$ are contained in *each* of those confidence intervals.

This allows us to make a decision for *every* pair of means simultaneously.

Example: (cont.)

For the cereal grain example SPSS produces the following output for the group means:

Report

thiamin

type	Mean	N	Std. Deviation
Barley	6.6000	6	.95079
Maize	5.5000	6	.66933
Oats	6.9833	6	1.04195
Wheat	5.7167	6	.76790
Total	6.2000	24	1.02406

(a) Use this information, together with the fact that the MSE in this example was 0.757 and that $Q_{0.05,4,20} = 3.96$ to obtain Tukey confidence intervals for the difference in average thiamine content between Barley and Maize.

(b) Decide which group means are significantly different.

Multiple Comparisons

Dependent Variable: thiamin
Tukey HSD

(I) type	(J) type	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Barley	Maize	1.100	.5023	.160	-.306	2.506
	Oats	-.383	.5023	.870	-1.789	1.022
	Wheat	.883	.5023	.321	-.522	2.289
Maize	Barley	-1.100	.5023	.160	-2.506	.306
	Oats	-1.483*	.5023	.036	-2.889	-.078
	Wheat	-.217	.5023	.972	-1.622	1.189
Oats	Barley	.383	.5023	.870	-1.022	1.789
	Maize	1.483*	.5023	.036	.078	2.889
	Wheat	1.267	.5023	.087	-.139	2.672
Wheat	Barley	-.883	.5023	.321	-2.289	.522
	Maize	.217	.5023	.972	-1.189	1.622
	Oats	-1.267	.5023	.087	-2.672	.139

(c) What does 95% confidence mean in this particular context?

An Alternative ANOVA Model

Instead of describing the observations as random variables with equal variance and (possibly different) means,

$$X_{ij} \sim \text{ Normal}(\mu_i, \sigma^2)$$

they can also be modeled as random deviations from a "true" population mean:

$$\boxed{X_{ij} = \mu_i + \epsilon_{ij}}$$

where the error terms ϵ_{ij} are assumed to be independent normally distributed random variables with mean zero and variance σ^2. This model is more easily generalized to the case of several factors than the random variable model.

Yet another alternative model writes each treatment mean μ_i as $\mu + \alpha_i$ where μ is the same overall mean for all groups and the α_i represent the group effects.

$$\boxed{X_{ij} = \mu + \alpha_i + \epsilon_{ij}}$$

Here, the overall mean μ can be found through

$$\mu = \frac{1}{I} \sum_{i=1}^{I} \mu_i,$$

and the sum of the deviations from the overall mean is subsequently zero:

$$\sum_{i=1}^{I} \alpha_i = \sum_{i=1}^{I} (\mu_i - \mu) = 0.$$

The ANOVA Hypothesis

In the alternative model notation, the ANOVA null hypothesis and alternative can be rephrased as

$$H_0 : \alpha_1 = \alpha_2 = \cdots = \alpha_I = 0$$

and

$$H_a : \text{ at least one } \alpha_i \neq 0$$

Power of the F-test

Recall: Statistical Power is the ability to correctly reject a false null hypothesis (or the ability to detect an existing effect). It can be written as $1-\beta$, where β is the probability to commit a type II error.

The power of an F-test depends on the formulation of the alternative hypothesis only through $\sum \alpha_i^2$. The quantity $J\sum \alpha_i^2/\sigma^2$ is called the NONCENTRALITY PARAMETER for a one-way ANOVA, because if H_0 is false, then the test-statistic has a *non-central* F distribution with this as one of its parameters.

Usually, power computations are very complicated. But for F-tests tables exist, from which β can be read for specific values of I, J and the noncentrality parameter.

Unequal Sample Sizes

So far we have only discussed the case, where we had the same number of observations at each factor level (e.g., for each grain type we had 6 observations). If the sample sizes for each population are not equal, the ANOVA table needs to be slightly modified:

Let $J_1, J_2, \ldots, J_I$ denote the I sample sizes and let $n = \sum J_i$ denote the total number of observations. Then

$$\text{SST} = \sum_{i=1}^{I}\sum_{j=1}^{J_i}(X_{ij}-\bar{X}_{\cdot\cdot})^2, \qquad df = n-1$$

$$\text{SSTr} = \sum_{i=1}^{I}\sum_{j=1}^{J_i}(\bar{X}_{i\cdot}-\bar{X}_{\cdot\cdot})^2, \qquad df = I-1$$

$$\text{SSE} = \sum_{i=1}^{I}\sum_{j=1}^{J_i}(X_{ij}-\bar{X}_{i\cdot})^2 = \text{SST - SSTr}, \qquad df = n-I$$

Test statistic value: $f = \dfrac{\text{MSTr}}{\text{MSE}}$ reject H_0 for $f \geq F_{\alpha,I-1,n-I}$

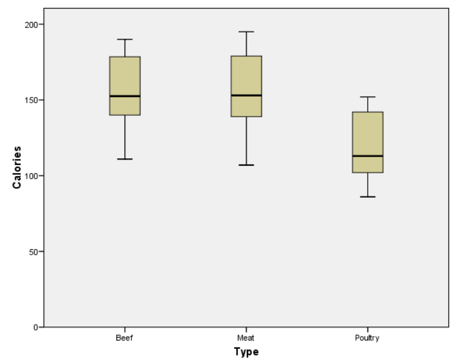
Example: People who are concerned about their health may prefer hot dogs that are low in calories. The "Hot Dog" data file contains data on the calories (and sodium) contained in each of 54 major hot dog brands. The hot dogs are classified by type: (1) beef, (2) poultry, and (3) meat (mostly pork and beef, but up to 15% poultry meat).

SPSS is used to run a one-way ANOVA with unequal sample sizes and produces the following output:

Report

Calories

Type	Mean	N	Std. Deviation
Beef	156.85	20	22.642
Meat	158.71	17	25.236
Poultry	118.76	17	22.551
Total	145.44	54	29.383



(a) Formulate the null hypothesis and alternative for this problem.

(b) Identify I, the J_i's and n.

(c) Identify SST, SSTr and SSE as well as MSTr and MSE from the ANOVA output:

ANOVA

Calories

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	17692.195	2	8846.098	16.074	.000
Within Groups	28067.138	51	550.336		
Total	45759.333	53			

(d) Identify the value of the test statistic f. What is the distribution of this test statistic?

(e) Formulate a conclusion for the hypothesis test.