Packet 7: More Testing

Chap 9.3 Analysis of Variance (ANOVA)

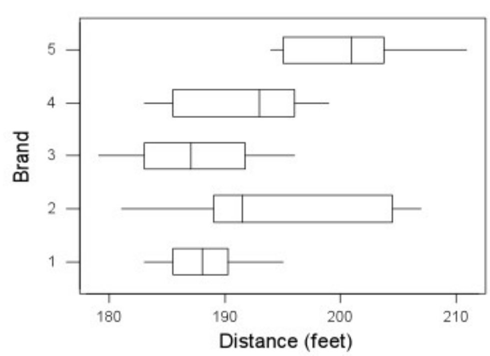
We could take a top-down approach by first presenting the theory of analysis of variance and then following it up with an example. We're not going to do it that way though. We're going to take a bottom-up approach, in which we first develop the idea behind the analysis of variance, and then present the general results.

Example 1: A researcher for an automobile safety institute was interested in determining whether or not the distance that it takes to stop a car going 60 miles per hour depends on the brand of the tire. The researcher measured the stopping distance (in feet) of ten randomly selected cars for each of five different brands. Here are the data resulting from his experiment:

Brand1	Brand2	Brand3	Brand4	Brand5
194	189	185	183	195
184	204	183	193	197
189	190	186	184	194
189	190	183	186	202
188	189	179	194	200
186	207	191	199	211
195	203	188	196	203
186	193	196	188	206
183	181	189	193	202
188	206	194	196	195

Do the data provide enough evidence to conclude that at least one of the brands is different from the others with respect to stopping distance?

Brand	N	MEAN	SD
1	10	188.20	3.88
2	10	195.20	9.02
3	10	187.40	5.27
4	10	191.20	5.55
5	10	200.50	5.44



It appears that the sample means differ quite a bit. For example, the average stopping distance of Brand3 is 187.4 feet, while the average stopping distance of Brand5 is 200.5 feet. A difference of 13 feet could mean the difference between getting into an accident or not.

The researcher needs to test the null hypothesis that the group population means are all the same against the alternative that at least one group population mean differs from the others.

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ v.s. H_1: at least one of the μ_i differs from the others.

We are going to learn how to use a method called analysis of variance to answer the researcher's question. The procedure is summarized in a so called analysis of variance table, such as this one:

Analysis of Variance
for comparing all 5 brands

Source	DF	SS	MS	F	P
Brand	4	1174.8	293.7	7.95	0.000
Error	45	1661.7	36.9		
Total	49	2836.5			

1. Source: the source of the variation in the data. The possible choices for a one-factor study, are Factor, Error, and Total. The factor is the characteristic that defines the populations being compared. In the tire study, the factor is the brand of tire.
2. DF: the degrees of freedom in the source.
3. SS: the sum of squares due to the source.
4. MS: the mean sum of squares due to the source.
5. F: the test-statistic which follows a F distribution under H_0.
6. P: the P-value.

To draw conclusions about the equality of two or more population means, we compare the P-value to α, our desired willingness to commit a Type I error. In this case, the researcher's P-value is very small, so he should reject his null hypothesis. That is, there is sufficient evidence, at even a 0.01 level, to conclude that the mean stopping distance for at least one brand of tire is different than the mean stopping distances of the others.

The Basic Idea Behind Analysis of Variance:

Analysis of variance involves dividing the overall (or "total") variability in observed data into two components:

1. the variability "between" groups
2. the variability "within" groups

H_0: $\mu_1 = \mu_2 = \mu_3 = \mu_4 = \mu_5$ v.s. H_1: at least one of the μ_i differs from the others.

Brand	N	MEAN	SD
1	10	188.20	3.88
2	10	195.20	9.02
3	10	187.40	5.27
4	10	191.20	5.55
5	10	200.50	5.44

1. SS(Factor) or SS(Treatment) or SS(between) is the sum of squares between the group means and the overall mean. As the name suggests, it quantifies the variability between the groups of interest.

2. SS(Error) or SS(Within) is the sum of squares between the data and the group means. It quantifies the variability within the groups of interest.

3. SS(Total) is the sum of squares between the n data points and the overall mean. As the name suggests, it quantifies the total variability in the observed data.

We'll show that: SS(Total) = SS(Between) + SS(Error), then provide the distribution of each SS, and finally provide the test statistic.

Let us now continue with *Example 1*. See also textbook *Examples 9.3-1 and 9.3-2*.

Brand	N	MEAN	SD
1	10	188.20	3.88
2	10	195.20	9.02
3	10	187.40	5.27
4	10	191.20	5.55
5	10	200.50	5.44

Analysis of Variance
for comparing all 5 brands

Source	DF	SS	MS	F	P
Brand	4	1174.8	293.7	7.95	0.000
Error	45	1661.7	36.9		
Total	49	2836.5			