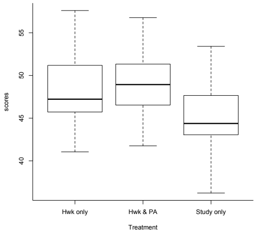
Comparing several groups of data

Does peer assessment (i.e. a student is grading other students' homework) enhance learning?

Students were randomized into three groups, each spending the same amount of time per week on either: homework only, homework and peer assessment, studying without doing homework.

The final exam scores were summarized by boxplots:



Is there sufficent evidence to conclude that the three treatments result in different outcomes?

Comparing several groups of data

H_0 = "nothing extraordinary is going on" = all group means are equal.

Recall that if we have only two groups, then we can compare them with the two-sample t-test:

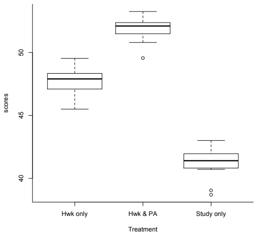
$$t = \frac{\text{difference in sample means}}{\text{SE of difference}}$$

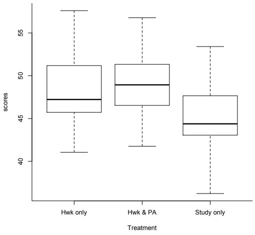
t compares the size of the difference in sample means to the size of the chance variability as measured by the SE.

ANOVA generalizes this idea.

Analysis of Variance

Consider these two hypothetical outcomes:





The differences between the sample means are large relative to the variability within the groups, suggesting that there is a difference in treatment means.

The differences between the sample means are small compared to the variability within the groups: It might well be due to chance variability.

Analysis of Variance

The key idea to make this precise is to compare the sample variance of the means to the sample variance within the groups.

That's why this methodology is called <u>An</u>alysis <u>o</u>f <u>Va</u>riance (ANOVA).

But recall that according to the square root law, the chance variability in the sample mean is smaller than the chance variability in the data. So the evidence against H_0 is not obvious from the boxplots. Rather, a computation is necessary.

Analysis of Variance

We have k groups and the jth group has n_j observations:

group 1	group 2	...	group k
y_{11}	y_{12}		y_{1k}
$\vdots$	$\vdots$		$\vdots$
$y_{n_1 1}$	$y_{n_2 2}$		$y_{n_k k}$

In total there are $N = n_1 + \ldots + n_k$ observations.

The sample mean of the jth group is $\bar{y}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} y_{ij}$.

The overall sample mean is $\bar{\bar{y}} = \frac{1}{N} \sum_{j=1}^{k} \sum_{i=1}^{n_j} y_{ij}$.

Analysis of Variance

The **treatment sum of squares**

$$\text{SST} = \sum_j \sum_i (\bar{y}_j - \bar{\bar{y}})^2$$

has $k-1$ degrees of freedom. The **treatment mean square**

$$\text{MST} = \frac{\text{SST}}{k-1}$$

measures the variability of the treatment means $\bar{y}_j$. The **error sum of squares**

$$\text{SSE} = \sum_j \sum_i (y_{ij} - \bar{y}_j)^2$$

has $N-k$ degrees of freedom. The **error mean square**

$$\text{MSE} = \frac{\text{SSE}}{N-k}$$

measures the variability within the groups.

The Analysis of Variance F-test

Since we want to compare the variation between the groups to the variation within the groups we look at the ratio

$$F = \frac{\mathsf{MST}}{\mathsf{MSE}}$$

Under the null hypothesis of equal group means this ratio should be about 1. It will not be exactly 1 due to sampling variability:

It follows a F**-distribution with** $k-1$ **and** $N-k$ **degrees of freedom.**

Large values of F suggest that the variation between the groups is unusually large. We reject H_0 if F is in the right 5% tail, i.e. when the p-value is smaller than 5%.

The ANOVA table

All the relevant information is summarized in the **ANOVA table**:

Source	df	Sum of Squares	Mean Square	F	p-value
Treatment	$k-1$	SST	MST	MST/MSE	
Error	$N-k$	SSE	MSE		
Total	$N-1$	TSS			

where TSS $= \sum_j \sum_i (y_{ij} - \bar{\bar{y}})^2$.

The ANOVA table

For the data of the boxplots in the right display we get a p-value of 0.097, so there is not enough evidence to reject H_0:

Source	df	Sum of Squares	Mean Square	F	p-value
Treatment	2	98.4	47.2	2.49	0.097
Error	38	723.8	19.1		
Total	40	822.2			

The one-way ANOVA model

The **one-way ANOVA** model behind this table is

$$y_{ij} \;=\; \mu_j + \epsilon_{ij}$$

where μ_j is the mean of the jth group and the ϵ_{ij} are independent random variables (e.g. measurement error) that follow the normal curve with mean 0 and common variance σ^2.

So the null hypothesis is $\mu_1 = \mu_2 = \ldots = \mu_k$.

Instead of looking at the group means μ_j it is helpful to look at the deviations τ_j from the overall mean μ: $\tau_j = \mu_j - \mu$.

So the model is

$$y_{ij} \;=\; \mu + \tau_j + \epsilon_{ij}$$

where τ_j is called the **treatment effect** of group j. Then the null hypothesis becomes

$\mathrm{H}_0 : \tau_1 = \tau_2 = \ldots = \tau_k = 0.$

The one-way ANOVA model

We estimate the overall mean μ by the 'grand mean' $\bar{\bar{y}}$. Then the estimate of $\tau_j = \mu_j - \mu$ becomes $\bar{y}_j - \bar{\bar{y}}$. The estimate of ϵ_{ij} is the residual $y_{ij} - \bar{y}_j$.

Corresponding to the model $y_{ij} = \mu + \tau_j + \epsilon_{ij}$ we can write y_{ij} as the sum of the corresponding estimates:

$$y_{ij} = \bar{\bar{y}} + (\bar{y}_j - \bar{\bar{y}}) + (y_{ij} - \bar{y}_j)$$

It turns out that such a decomposition is also true for the sum of squares:

$$\sum_j \sum_i (y_{ij} - \bar{\bar{y}})^2 = \sum_j \sum_i (\bar{y}_j - \bar{\bar{y}})^2 + \sum_j \sum_i (y_{ij} - \bar{y}_j)^2$$

$$\text{TSS} = \text{SST} + \text{SSE}$$

This splits the total variation TSS into two 'sources': SST and SSE.

More on ANOVA

- The F-test assumes that all the groups have the same variance σ^2. This can be roughly checked with side-by-side boxplots, and there are also formal tests.
- Another assumption was that the data are indendent within and across groups. This would be the case if the subjects were assigned to treatments at random.
- If the F-test rejects, then we can conclude that the group means are not all equal, but how do they differ?

  We can examine all pairs of means with a two-sample t-test using $s_{pooled} = \sqrt{\text{MSE}}$.

  But since that involves several tests, an adjustment such as the Bonferroni adjustment is necessary, see the next module.