9 One-way analysis of variance

The topic will be introduced by two examples. We are interested in the effect of the day time (the early morning, before midday, the afternoon, the evening, the night) on the productivity of a manual worker (expressed in a total number of manufactured products).
Or we would like to know if the mother's education (university educ., secondary educ., elementary educ.) has effect on the length of nursing period (expressed in a number of weeks).
Generally the analysis of variance answers the question if the qualitative random variable (factor A) does have an effects on quantitative random variable X, or does not. In the first example the factor "day time" operates at 5 levels; in the second the factor "mother's education" operates at 3 levels.
To find out if the factor A does effect the random variable X, n_i independent observations of variable X at i-th factor level are acquired. Thus if the factor A has r levels then for each level one random sample is assigned and these r random samples are mutually independent:

factor A	Random sample
level 1	$X_{11}, \ldots, X_{1n_1} \sim N(\mu_1, \sigma^2)$
level 2	$X_{21}, \ldots, X_{2n_2} \sim N(\mu_2, \sigma^2)$
$\vdots$	$\vdots$
level r	$X_{r1}, \ldots, X_{rn_r} \sim N(\mu_r, \sigma^2)$

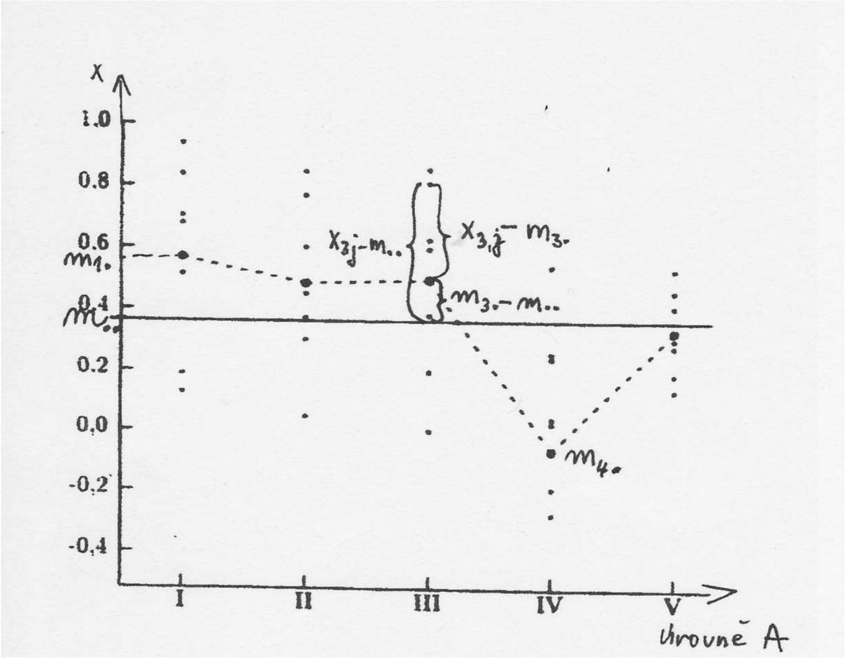
If the factor A does not effect the random variable X then the expected values $\mu_1,\ \mu_2, \ldots, \mu_r$ should be the same. Thus at the significance level α we are testing hypothesis:

$H_0: \quad \mu_1 = \mu_2 = \ldots = \mu_r$ versus
H_1 : At least two of the expected values are different.

This test can be thought of as an extension of the two-sample t-test. This leads to the question whether the hypothesis H_0 can not be accomplished by means of $\binom{r}{2}$ separate two-sample t-test each of them at the significance level α? If at least one t-test would reject the equality of two expected values, then H_0 about equality of all r expected values could be rejected. At the same time we would know which pair of expected values is different. But this procedure does not meet the condition that the type I error should be at most α. (The error would be substantially greater.) This condition is satisfied by ANOVA (Analysis of variance) method, which tests hypothesis about equality of all r population means. If this hypothesis is rejected at the significance level α then it is often desirable to know which factor level is responsible for the difference between population means. There are many different multiple comparison procedures that deal with this problem. We will illustrate these methods using Tukey's and Scheffe's multiple comparison method. First let us start with customary ANOVA notation.

Notation 9.1

$n = \sum\limits_{i=1}^{r} n_i$	the total size of all observations
$M_{i\cdot} = \frac{1}{n_i} \sum\limits_{j=1}^{n_i} X_{ij}$	the sample mean of the $i-$th sample
$M_{\cdot\cdot} = \frac{1}{n} \sum\limits_{i=1}^{r} \sum\limits_{j=1}^{n_i} X_{ij}$	the total sample mean of all n observationes
$S_T = \sum\limits_{i=1}^{r} \sum\limits_{j=1}^{n_i} (X_{ij} - M_{\cdot\cdot})^2$	the total sum of squares (the statistic S_T has $f_T = n - 1$ degrees of freedom)
$S_A = \sum\limits_{i=1}^{r} n_i \cdot (M_{i\cdot} - M_{\cdot\cdot})^2$	the factor (or between-groups) sum of squares (statistic S_A has $f_A = r - 1$ degrees of freedom)
$S_E = \sum\limits_{i=1}^{r} \sum\limits_{j=1}^{n_i} (X_{ij} - M_{i\cdot})^2$	the error (or within-groups) sum of squares (statistic S_E has $f_E = n - r$ degrees of freedom)



Remark 9.2

•The statistic S_T is nothing but the numerator of the expression for the variance of all n observations of the random variable X. It characterizes the total variation of all n observations X_{ij} around the total mean $M_{\cdot\cdot}$.

•Statistic S_A measures the proximity of r sample means to each other and brings out the factor impact on variability.

•Statistic S_E is a pooled measure of variation within the particular samples which is caused randomly, thus it is not caused by the factor.

Each sum of squares has its *degrees of freedom* which are given by the number of independent variables within considered group. Considering the statistic $\bullet S_T$, n observations correspond to n summands of the sum $\sum\limits_{i=1}^{r} \sum\limits_{j=1}^{n_i} (X_{ij} - M_{\cdot\cdot})$. But this summands are not totally arbitrary, they have to satisfy the condition $\sum\limits_{i=1}^{r} \sum\limits_{j=1}^{n_i} (X_{ij} - M_{\cdot\cdot}) = 0$. Thus S_T has $f_T = n - 1$ degrees of freedom. Analogously the statistic $\bullet S_A$ has r observations which have to satisfy the condition $\sum\limits_{i=1}^{r} n_i(M_{i\cdot} - M_{\cdot\cdot}) = 0$. Thus S_A has $f_A = r - 1$ degrees of freedom. The degrees of freedom for the statistic $\bullet S_E$ can be calculated from the formula $f_E = f_T - f_A$. (This formula follows from the relationship $S_T = S_A + S_E$, which is stated in the following theorem.) Thus $f_E = n - r$.

Theorem 9.3
Considering the notation in 9.1 it holds:

1.) $S_T = S_A + S_E$.

2.) $S_*^2 = \frac{S_E}{n-r}$, where S_*^2 is the weighted mean of sample variances.

3.) $\frac{S_E}{\sigma^2} \sim \chi^2(n-r)$. $[E(\frac{S_E}{n-r}) = \sigma^2]$

4.) Variables $\frac{S_E}{\sigma^2}$ and $\frac{S_A}{\sigma^2}$ are independent.

In addition if $H_0 : \quad \mu_1 = \mu_2 = \ldots = \mu_r$ is true then

5.) $\frac{S_A}{\sigma^2} \sim \chi^2(r-1)$. $[E(\frac{S_A}{r-1}) = \sigma^2$ if the assumption holds.]

9.4 Testing the hypothesis that the population means are equal

At the significance level α we are testing the hypothesis:

$H_0 : \quad \mu_1 = \mu_2 = \ldots = \mu_r$ versus
H_1 : At least two of the expected values are different.

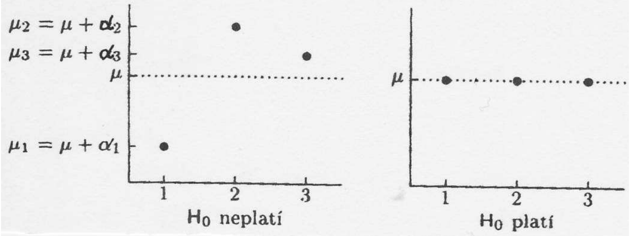
The random variable $X_{ij},\ i = 1, \ldots, r;\ j = 1, \ldots, n_i$ follows the normal distribution: $X_{ij} \sim N(\mu_i, \sigma^2)$. Thus X_{ij} can be expressed as follows:

$$\boxed{\begin{aligned} X_{ij} &= \mu_i + \varepsilon_{ij} \\ &= \mu + \alpha_i + \varepsilon_{ij}, \end{aligned}}$$

where ε_{ij} are independent random variables following the distribution$N(0, \sigma^2)$
μ is that part of an expected value $E(X)$, which is common to all r random samples
α_i is the population effect of the i-th factor level.

(Parameters μ and α_i are unknown and we require the validity of following equation $\sum\limits_{i=1}^{r} n_i\alpha_i = 0$)

The null hypothesis is true if the factor A has absolutely no effect on the random variable X. Thus the null could be rewritten to the form $H_0 : \quad \alpha_1 = \alpha_2 = \ldots = \alpha_r = 0$
Hence if H_0 is true then: $\boxed{X_{ij} = \mu + \varepsilon_{ij}}$



The statistic $M_{i\cdot}$ is a point estimator of the mean value μ_i
The statistic $M_{\cdot\cdot}$ is a point estimator of the mean value μ

The statistic $M_{i\cdot} - M_{\cdot\cdot}$ is a point estimator of the effect $\alpha_i = \mu_i - \mu$

To make decision about the null hypothesis, we compare the mean squares S_A/f_A and S_E/f_E, whose expected values should be the same under the true null hypothesis. This suggests that the test statistic $F_A = \frac{S_A/f_A}{S_E/f_E}$ follows (under true null) Fisher's distribution $F(f_A, f_E)$.

It is obvious that the big differences between $M_{i\cdot}$ and $M_{\cdot\cdot}$ bring evidence against the null; thus the statistic S_A is responsible for rejecting or not rejecting the null. The statistic S_E is instrumental towards estimation of the parameter σ^2. Therefore the null hypothesis is rejected at the significance level α if:

$$F_A = \frac{S_A/f_A}{S_E/f_E} \geq F_{1-\alpha}(r-1, n-r)$$

The results of the one-way analysis of variance are often displayed in a table similar to following one.

Sources of variability	sum of squares	degrees of freedom	mean squares	test statistic
factor	S_A	$f_A = r-1$	S_A/f_A	$F_A = \frac{S_A/f_A}{S_E/f_E}$
error	S_E	$f_E = n-r$	S_E/f_E	
total	S_T	$f_T = n-1$		

When the null hypothesis in the analysis of variance is rejected we are interested in comparing all pairs of expected values to find at least one pair of different expected values which caused the rejection. These pairs may be identified by *multiple comparison methods.*

9.5 Tukey's method

This method requires equal sample sizes, thus $p := n_1 = n_2 = \ldots = n_r$.
At the significance level α we are testing the hypothesis:

$H_0: \quad \mu_k = \mu_l \qquad$ versus
$H_1: \quad \mu_k \neq \mu_l$
The hypothesis about equality $\mu_k = \mu_l$ is rejected at the significance level α when

$$|M_{k\cdot} - M_{l\cdot}| \geq q_{1-\alpha}(r, n-r)\frac{S_*}{\sqrt{p}},$$

where the values $q_{1-\alpha}(r, n-r)$ are tabulated and are known as the *studentized range quantiles.*
This procedure identifies all pairs k, l, in which the expected values $\mu_k,\ \mu_l$ differs significantly at the significance level α.

9.6 Scheffe's method

The advantages of this method is simplicity and applicability to groups of unequal size. It is known to be relatively insensitive to departures from normality and homogenity of variance.
At the significance level α we are testing the hypothesis:

$H_0: \quad \mu_k = \mu_l \qquad$ versus
$H_1: \quad \mu_k \neq \mu_l$
The hypothesis about equality $\mu_k = \mu_l$ is rejected at the significance level α when

$$|M_{k\cdot} - M_{l\cdot}| \geq S_*\sqrt{(r-1)\left(\frac{1}{n_k} + \frac{1}{n_l}\right)F_{1-\alpha}(r-1, n-r)}$$

The following situation may occur: The hypothesis $H_0: \quad \mu_1 = \mu_2 = \ldots = \mu_r$ is rejected, but the multiple comparison methods do not identify any pair with significant difference of expected values. Then the more complicated combination of expected values known as *contrast* is significantly different.

Here let us recall the ANOVA assumptions and then we will follow with tests of these assumptions.

9.7 Assumptions of ANOVA

According to the established notation the random samples should have following properties:

1.) The normality: $X_{i1}, \ldots, X_{in_i} \sim N(\mu_i, \sigma^2),\ i = 1, \ldots, r.$

2.) The independence: The particular random samples are mutually independent.

3.) The homoskedasticity: The variances of particular samples are equal, thus $\sigma^2 := \sigma_1^2 = \ldots = \sigma_r^2$

The normality is either known or the tests of normality may be used. (Generally ANOVA is not too sensitive to departures from normality.) The independence should follow from the design of an experiment. In the end the homoskedasticity should be verified, thus we have to run a test that the r population variances are equal. The following tests may be used:

9.8 Leven's test

Leven's test is in fact one-way ANOVA formally applied on variables $|X_{ij} - M_{i\cdot}|$.
At the significance level α we are testing the hypothesis:

$H_0: \quad \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_r^2 := \sigma^2$ versus
H_1 : At least two of the variances are different.

Let us denote $Z_{ij} = |X_{ij} - M_{i\cdot}|$. Then according to ANOVA notation we denote
$M_{Zi} = \frac{1}{n_i} \sum\limits_{j=1}^{n_i} Z_{ij}$
$M_Z = \frac{1}{n} \sum\limits_{i=1}^{r} \sum\limits_{j=1}^{n_i} Z_{ij}$
$S_{ZA} = \sum\limits_{i=1}^{r} n_i \cdot (M_{Zi} - M_Z)^2$
$S_{ZE} = \sum\limits_{i=1}^{r} \sum\limits_{j=1}^{n_i} (Z_{ij} - M_{Zi})^2$
If the null hypothesis about equal variances is true then the statistic

$$F_{ZA} = \frac{S_{ZA}/(r-1)}{S_{ZE}/(n-r)} \sim F(r-1, n-r)$$

The null hypothesis about equal variances is rejected at the significance level α
if $F_{ZA} \geq F_{1-\alpha}(r-1, n-r)$.

9.9 Bartlett's test

If the r population sizes are at least 7 the Bartlett's test about equality of variances may be used.
It's disadvantage is substantial sensitivity to the violence of assumption of normality.
If the null hypothesis about equal variances is true then the statistic

$$B = \frac{1}{C} \left((n-r) \ln S_*^2 - \sum_{i=1}^{r} (n_i - 1) \ln S_i^2 \right) \approx \chi^2(r-1), \text{ where}$$

$C = 1 + \frac{1}{3(r-1)}\left(\sum\limits_{i=1}^{r} \frac{1}{n_i-1} - \frac{1}{n-r}\right)$
$S_i^2 = \sum\limits_{j=1}^{n_i} \frac{1}{n_i-1}(X_{ij} - M_{i\cdot})^2$
$S_*^2 = \frac{1}{n-r}\sum\limits_{i=1}^{r}(n_i-1)S_i^2 = \frac{S_E}{n-r}$
The null hypothesis about equal variances is rejected at the significance level α if $B \geq \chi^2_{1-\alpha}(r-1, n-r)$.

9.10 The summing up
An outlie of the steps:

1.) We have to verify assumptions of ANOVA; to verify homoskedasticity use *Leven's test*, or *Bartlett's test.*

2.) Using ANOVA table we make decision about the null hypothesis stating that population means are equal.

3.) If the hypothesis about equal population means is rejected the multiple comparison methods may be used. These methods are aimed to identify pairs which caused the rejection. Use *Tukey's method*, or*Scheffe's method.*

Example 9.11
Considering four sorts of potatoes we are interested in the total weight od potatoes from one bunch. The results in kg are performed in the following table:

the sort	the weight
I.	0,9 0,8 0,6 0,9
II.	1,3 1,0 1,3
III.	1,3 1,5 1,6 1,1 1,5
IV.	1,1 1,2 1,0

Run a test at $\alpha = 5\%$ that the expected values of bunch weights are not effected by the sort. If you reject the null, find out at $\alpha = 5\%$ which pairs of sorts are different.

Solution
We assume the data to be realization of four mutually independent normally distributed random samplesα with equal population variance. Thus $X_{i1}, \ldots, X_{in_i} \sim N(\mu_i, \sigma^2);\ i = 1, 2, 3, 4.$
We are testing the hypothesis that all the four population means are equal:
$H_0: \quad \mu_1 = \mu_2 = \mu_3 = \mu_4$ versus
H_1 : At least two of the expected values are different.

First we have to determine the realization of needed statistics:
$m_{1\cdot} = 0,8; \quad m_{2\cdot} = 1,2; \quad m_{3\cdot} = 1,4; \quad m_{4\cdot} = 1,1; \quad m_{\cdot\cdot} = 1,14$
$S_E = 0,3; \quad S_A = 0,816; \quad S_T = 1,116$. Further $r = 4; \quad n = 15$

Source	sum of squares	degrees of freedom	mean squares	test statistic
factor	$S_A = 0,816$	$f_A = r - 1 = 3$	$S_A/3 = 0,272$	$F_A = \frac{S_A/f_A}{S_E/f_E} = 9,97$
error	$S_E = 0,3$	$f_E = n - r = 11$	$S_E/11 = 0,02727$	
total	$S_T = 1,116$	$f_T = n - 1 = 14$		

The critical region follows $W = \langle F_{0,95}(3,11); \infty) = \langle 3,59; \infty)$. The realization of the test statistic

$9,97 \in W$, thus H_0 about equal population means is rejected at α.
Using Scheffe's method we identify the pairs which caused the rejection at $\alpha = 0,05$.

Compared sorts	Differences $\lvert M_{k\cdot} - M_{l\cdot} \rvert$	The right side of an unequality
I., II.	0,4	0,41
I., III.	0,67$*$	0,36
I., IV.	0,3	0,41
II., III.	0,2	0,40
II., IV.	0,1	0,44
III., IV.	0,3	0,40

At $\alpha = 0,005$ the sorts I. and III. are different. (The asterisk in the table identifies the pair, in which the difference $\lvert M_{k\cdot} - M_{l\cdot} \rvert$ is significant.).