ANOVA – One Factor Analysis of Variance

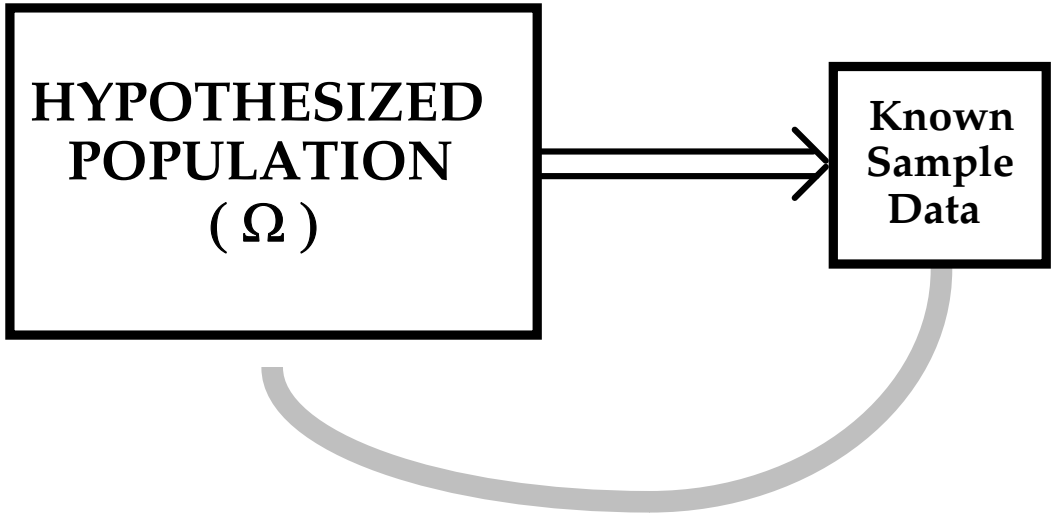
Aim

- To explain structure of *F-ratio*
- To make conclusions by the test named Analysis of Variance
- To construct the ANOVA table
- To realize the post hoc analysis

The Method of Hypothesis Testing

Pure Significance Tests

The pure significance test asks whether the sample result is extreme with respect to some hypothesized distribution.



Can this data reasonably be assumed to have been generated by Ω?

Is this data consistent with Ω ?

The pure significance test consists of the following components:

1. **Null Hypothesis**: H_0

– The null hypothesis expresses some belief about the nature of the population. It must be specified precisely enough to define a probability measure on the population.

2. **Sample Statistic**: $T(\underline{X})$

– The sample statistic is a function of the sample data drawn from the population. The choice of sample statistic is determined by the characteristics of the population's probability distribution with which the null hypothesis is concerned.

3. **Null Distribution**: $F_0(x)$

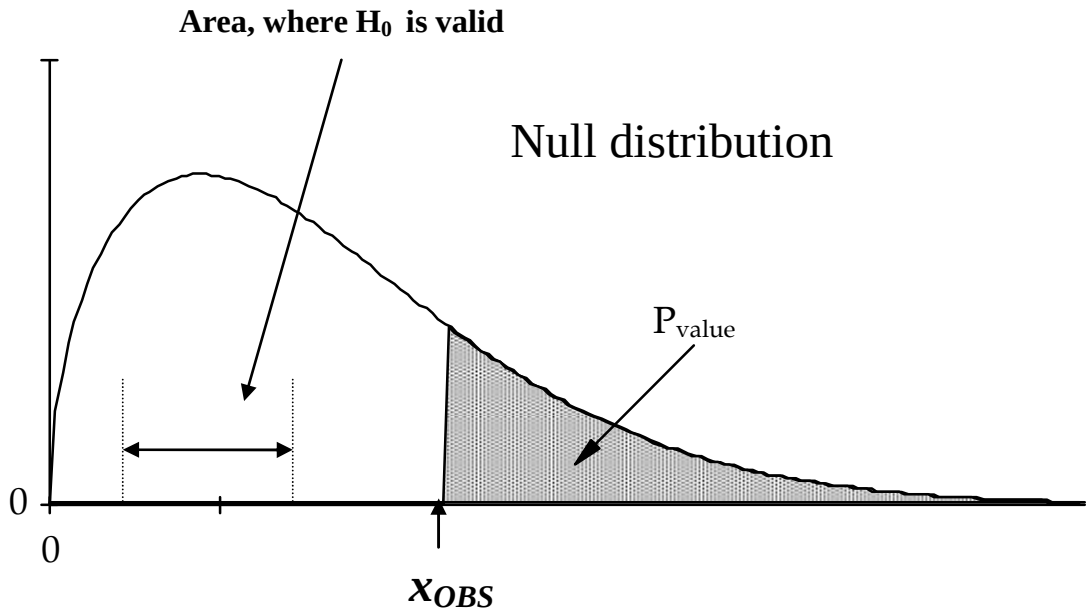
$$F_0(x) = P(T(\underline{X}) < x \mid H_0)$$

- The null distribution is the probability distribution of the sample statistic when the null hypothesis is correct. The null hypothesis must be specified precisely enough to determine the null distribution.

The pure significance test - continuation

4. To determine whether the observed sample statistic $t=x_{OBS}$ is extreme with respect to the null distribution, a statistic known as the ***p-value*** is computed. The *p-value* has 3 definitions depending on the context of the null hypothesis, but in all cases, the interpretation of the *p-value* is the same.

$$P_{VALUE} = 1-F_0(x_{obs}),$$



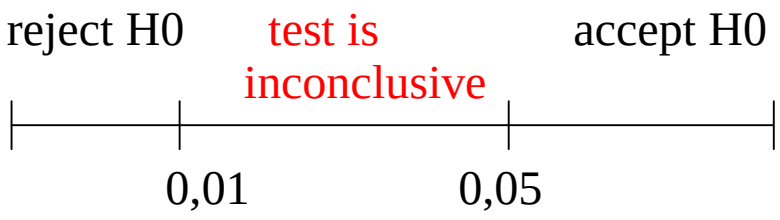
The pure significance test - continuation

5. Decision making based on P_{VALUE}

$P_{VALUE} < 0{,}01$ reject H_0

$0{,}01 < P_{VALUE} < 0{,}05$ test is inconclusive

$P_{VALUE} > 0{,}05$ accept H_0 (do not reject H_0)



Student's t test for difference of means

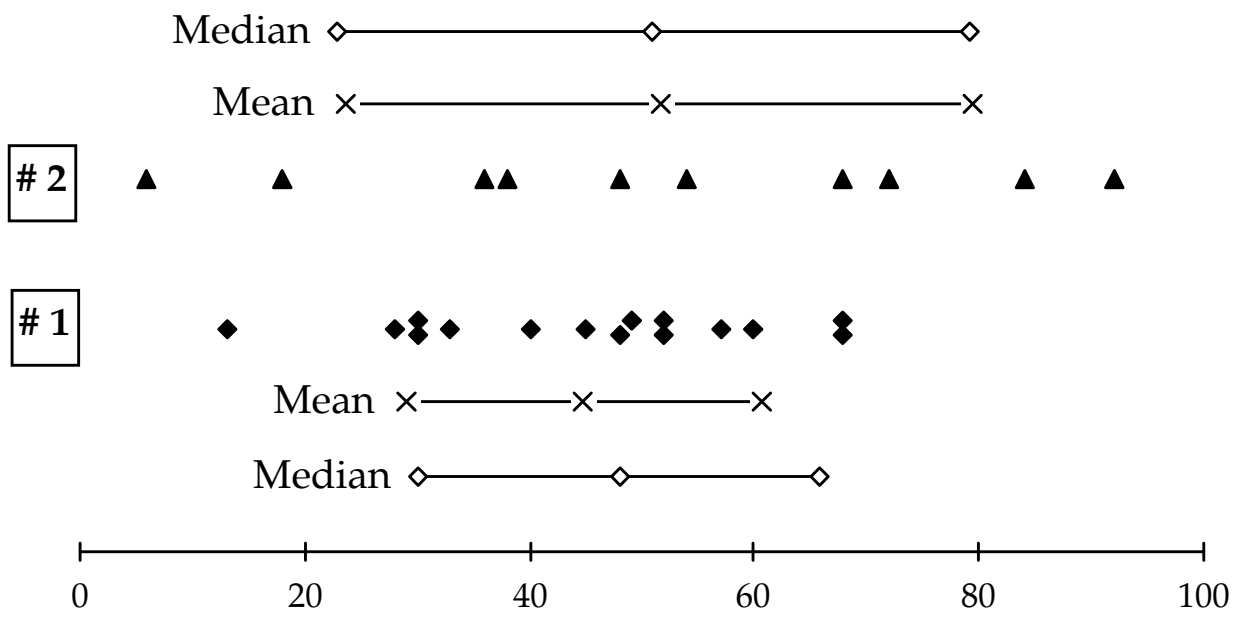
To determine the effectiveness of a new teaching method, a controlled experiment may be conducted in which one group of students, the control group, is taught by traditional methods and a second group by the experimental method. The research question in this case is whether the students taught by the experimental method attained higher results.

Sample $\underline{X}$ is from population #1

60	49	52	68	68
45	57	52	13	40
33	30	28	30	48

Sample $\underline{Y}$ is from population #2

38	18	68	84	72
48	36	92	6	54



The sample means and standard deviations are:

$$\#1: \overline{X} = 44.867; \quad S_1 = 15.77$$

$$\#2: \overline{Y} = 51.6; \quad S_2 = 27.93$$

We assume that both samples come from normal distributions:

$$X_i \to N(\mu_1, \sigma^2); \quad i = 1,\ldots,n_1 \quad Y_j \to N(\mu_2, \sigma^2); \quad j = 1,\ldots,n_2,$$

H$_0$: $\mu_1 = \mu_2$ **H$_A$:** $\mu_1 < \mu_2$

Null distribution is **Student's** distribution:

$$t_{n_1+n_2-2} = \frac{\bar{x} - \bar{y}}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where

$$s_p = \sqrt{\frac{\sum_{i=1}^{n_1} \left(x_i - \bar{x}\right)^2 + \sum_{j=1}^{n_2} \left(y_j - \bar{y}\right)^2}{n_1 + n_2 - 2}} = \sqrt{\frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}}$$

and p_{VALUE}:

$$s_p = \sqrt{\frac{14\left(15.77^2\right)+9\left(27.93^2\right)}{23}} = 21.37$$

$$t_{23} = \frac{44.867-51.6}{21.37\sqrt{\frac{1}{15}+\frac{1}{10}}} = -0.772$$

$$p_{VALUE} = t_{n_1+n_2-2}(-0.772) = 0.224$$

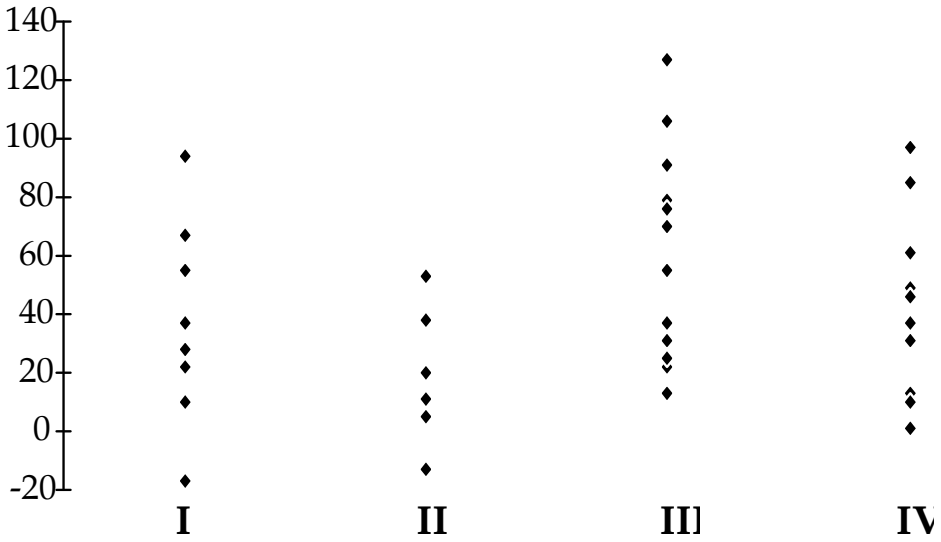
→ decision is

accept **H₀:** $\mu_1 = \mu_2$

Tests for three or more samples

Example:

Groups	I	II	III	IV
Data	67	20	106	13
	22	-13	127	49
	10	11	13	97
	55	5	79	85
	94	38	37	46
	-17	53	31	31
	37	5	22	37
	28		70	61
			76	10
			55	1
			91	
			25	
Sample size	8	7	12	10
Group means	37	17	61	43
Group St. Dev.	34.69	22.25	36.36	31.59



We face the problem of testing hypothesis H_0: $\mu_1 = \mu_2 = \ldots = \mu_k = \mu$?

One possibility is to realize two sample tests among each other. We have $k.(k-1)/2$ **tests in total.**

Observed values of Student's t test statistics:

Sample sizes		8 I	7 II	12 III	10 IV
8	I	0	-1.188	1.616	0.389
7	II	1.188	0	2.844	1.622
12	III	-1.616	-2.844	0	-1.292
10	IV	-0.389	-1.622	1.292	0

P-values (compare with 0.01):

Sample sizes		8 I	7 II	12 III	10 IV
8	I	0	0.11	0.079	0.35
7	II		0	0.0051	0.041
12	III			0	0.117
10	IV				0

Conclusion: Reject H_0, groups II a III have significantly different means.

Why is a special test necessary in this situation?

p_{12} = p-value of test comparing populations 1 and 2

p_{13} = p-value of test comparing populations 1 and 3

then the probability of either of the two tests having an observed p-value less than p is

Pr [min (p_{12}, p_{13}) < p],

$$\Pr\left[\min(p_{12}, p_{13}) < p\right] = \Pr\left[(p_{12} < p) \cup (p_{13} < p)\right] =$$
$$\Pr\left[p_{12} < p\right] + \Pr\left[p_{13} < p\right] - \Pr\left[(p_{12} < p) \cap (p_{13} < p)\right] = 2p - p^2$$

Note: the probability of either of the two tests having an observed p-value less than p is the complement of neither test having an observed p-value less than p.

Pr [min (p_{12}, p_{13}) < p]= 1 - $(1 - p)^2$ = $2p - p^2$

For n tests comparing n different treatments with a control, the probability of at least one having an observed p-value less than p is

$$\textbf{Prob [min (p1, p2, p3,…, pn) < p]} = 1 - (1 - p)^n = n.p - \ldots p^2 + \ldots p^3 \ldots \approx n.p$$

Since p-values are always very small, the square of the p-value becomes negligible and so on.

The probability of any one of three independent tests having an observed p-value less than p is

$$1 - (1 - p)^3 = 3p - 3p^2 + p^3 \approx 3.p$$

If n=20 then the probability is:

$$\textbf{Pr [min } (p_1, p_2, \ldots p_n) < 0{,}05 \textbf{]} \approx 20.\ 0{,}05 = 1$$

Therefore, the expected number of tests having an observed p-value less than .05 is 1!

Our solution to these difficulties is to create a single test statistic whose distribution is known when all population means are equal and that is sensitive on any differences in population means. Then a single overall p-value can be used to determine whether experimental results are consistent with the null hypothesis.

ANOVA – Analysis of Variance

ANOVA … It enables compare any mean of independent random samples.

Assumptions:

Let's have *k*-random samples that are independent on each other. These samples have the standard distribution with the same variation:

$$\begin{array}{l} (X_{11}, X_{12}, ..., X_{1n_1}) \rightarrow N(\mu_1, \sigma^2) \\ (X_{21}, X_{22}, ..., X_{2n_2}) \rightarrow N(\mu_2, \sigma^2) \\ ... \\ (X_{k1}, X_{k2}, ..., X_{kn_k}) \rightarrow N(\mu_k, \sigma^2) \end{array},$$

Let n_i ... number of observations in i-th sample.

$$\sum_{i=1}^{k} n_i = N$$

Formulation of the problem:

The hypothesis of interest is H_0: $\mu_1 = \mu_2 = ... = \mu_k = \mu$

The alternate hypothesis is H_A: At least two μ_i's are different.

We have to find such a test statistics which is sensitive on H_0

Example

compare mathematical results of graduates of three different types of secondary school , see Table.

Gymnasium	SP	OU
55	54	47
54	50	53
58	51	49
61	51	50
52	49	46

Define the **total sum of squares** (or **total variability**) as

$$SS_{TOTAL} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \overline{X})^2 \text{, where } \overline{X} \text{ is the mean of all observations.}$$

- the total sum of squares is overall source of variability in the data

This total sum of squares we can separate into 2 components:

$$SS_{TOTAL} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \overline{X})^2 \Rightarrow SS_{TOTAL} = SS_W + SS_B \quad ,$$

where SS_W ... the within group variation (the sum of squares within groups) - is the raw variability within samples

$$SS_W = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)^2 = \sum_{i=1}^{k}(n_i - 1)S_i^2$$

s_i is a sample standard deviation of *i*-th random sample:

$$S_i = \sqrt{\frac{\sum_{j=1}^{n_i}\left(X_{ij} - \overline{X}_i\right)^2}{n_i - 1}} \text{ and } \qquad \overline{X}_i = \frac{\sum_{j=1}^{n_i} X_{ij}}{n_i}; \text{ is a sample mean in } i\text{-th sample.}$$

SS_B ... the between group variation (the sum of squares between groups) - is the raw variability between groups: $SS_B = \sum_{i=1}^{k} n_i \cdot (\overline{X}_i - \overline{X})^2$

Example – secondary schools

	Gymnázium	SPŠ	OU	
	55	54	47	
	54	50	53	
	58	51	49	
	61	51	50	
	52	49	46	
$\overline{X_i}$	56	51	49	$\overline{\overline{X}} = 52$
$\overline{X_i} - \overline{\overline{X}}$	4	-1	-3	$\sum_{i=1}^{3}\left(\overline{X_i} - \overline{X}\right) = 0$
$\left(\overline{X_i} - \overline{\overline{X}}\right)^2$	16	1	9	$\sum_{i=1}^{3}\left(\overline{X_i} - \overline{X}\right)^2 = 26$

$$SS_B = \sum_{i=1}^{k} n_i \left(\overline{X_i} - \overline{X}\right)^2 = 5 \cdot (56-52)^2 + 5 \cdot (51-52)^2 + 5 \cdot (49-52)^2 = 130{,}0$$

$$SS_W = \sum_{i=1}^{k}\sum_{j=1}^{n} (X_{ij} - \overline{X_i})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n} (X_{ij} - \overline{X_i})^2 =$$

$$= (55-56)^2 + (54-56)^2 + \ldots + (54-51)^2 + \ldots + (47-49)^2 + \ldots + (46-49)^2 = 94{,}0$$

Within and Between Variances

The within group variance (mean square within groups): $S_W^2 = \dfrac{SS_W}{N-k}$

The between group variance (mean squares between groups): $S_B^2 = \dfrac{SS_B}{k-1}$

Properties of these variances:

1. $ES_W^2 = \dfrac{1}{N-k} E(\sum_{i=1}^{k} (n_i - 1) S_i^2) = \dfrac{1}{N-k} \sum_{i=1}^{k} (n_i - 1) E(S_i^2) = \sigma^2$

   because $E(S_i^2) = \sigma^2$.

The within mean square is an unbiased estimate of the variance, independently on H_0.

2. $ES_B^2 = \sigma^2 + \dfrac{1}{k-1} \sum_{i=1}^{k} n_i (E\overline{X} - E\overline{X}_i)^2$

   $ES_B^2 = \sigma^2 \Leftrightarrow$ when H_0 is true

F-ratio

Therefore the ratio of the two sums of squared divided by their degrees of freedom will have an F distribution under the hypothesis of equal population means.

$$F = \frac{SS_B / k-1}{SS_W / N-k} = \frac{S_B^2}{S_W^2}$$

Definition:
We call this F statistic as ***F-ratio***.

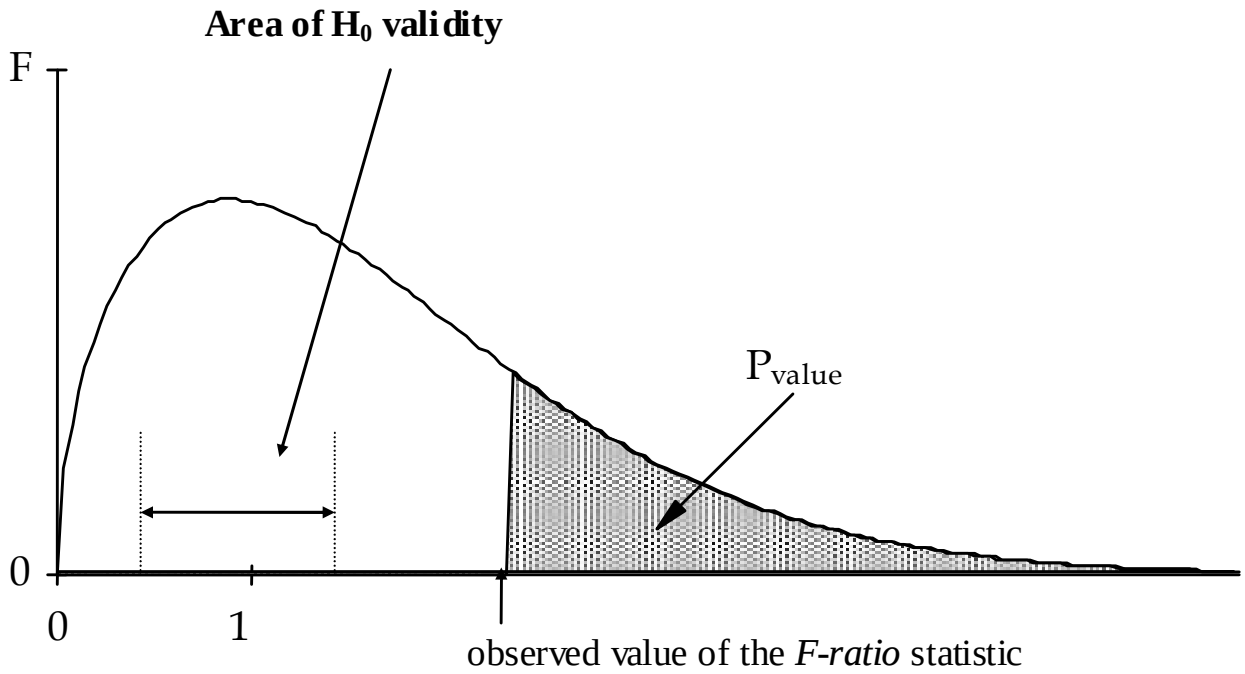
Why is useful use *F*-ratio as the test statistic?

i) If H_0 is true then *F-ratio* is a random number close to 1 ... $F \approx 1$.
ii) If H_0 is faslse then this number is markedly bigger than 1, see properties.

The statistic *F-ratio* is sensitive to validity of the hypothesis H_0.

If we know a *F-ratio* statistical behavior we can use it for analysis and determination of previously stated problem in H_0.

Following figure illustrated a usage of *F-ratio* to determine



ANOVA Table

We summarize the data in an ANOVA table:

Source	Sum of squares	Degrees of freedom	Mean squares	*F-ratios*	P-value
Total	$SS_{TOTAL} = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \overline{X})^2$	$N-1$			
Between	$SS_B = \sum_{i=1}^{k} n_i \cdot (\overline{X}_i - \overline{X})^2$	$k-1$	$S_B^2 = \frac{SS_B}{k-1}$	$F = \frac{S_B^2}{S_W^2}$	**p-value**
Within	$SS_W = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(X_{ij} - \overline{X}_i)^2$	$N-k$	$S_W^2 = \frac{SS_W}{N-k}$		

Solved examples

We assume three data sets for illustration of *F-ratio* statistical behavior. In each data set, the sample means are the same but the variations within groups differs. When the within group variation is small than the *F-ratio* is large. When the within group variation is large than the *F-ratio* is small. The examples illustrate three cases: small within group variation; normal within group variation; and large within group variation.

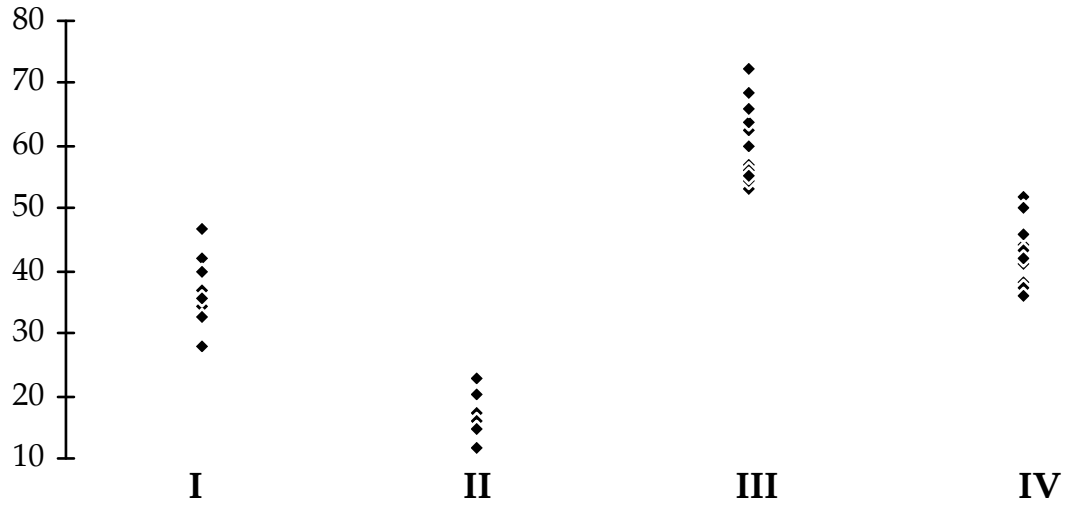
Example 1: **Small within group variance**

Groups	I	II	III	IV
Data	42 34.5 32.5 40 46.5 28 37 35.5	17.5 12 16 15 20.5 23 15	68.5 72 53 64 57 56 54.5 62.5 63.5 60 66 55	38 44 52 50 43.5 41 42 46 37.5 36
Sample size	8	7	12	10
Group means	37	17	61	43
Group standard deviations	5.78	3.71	6.06	5.27

ANOVA Table

	Degrees of freedom	Sum of squares	Mean squares	*F-ratio*
total	36	9872.7		
between	3	8902.7	2967.57	100.96
within	33	970	29.39	

P-value = 0.0000



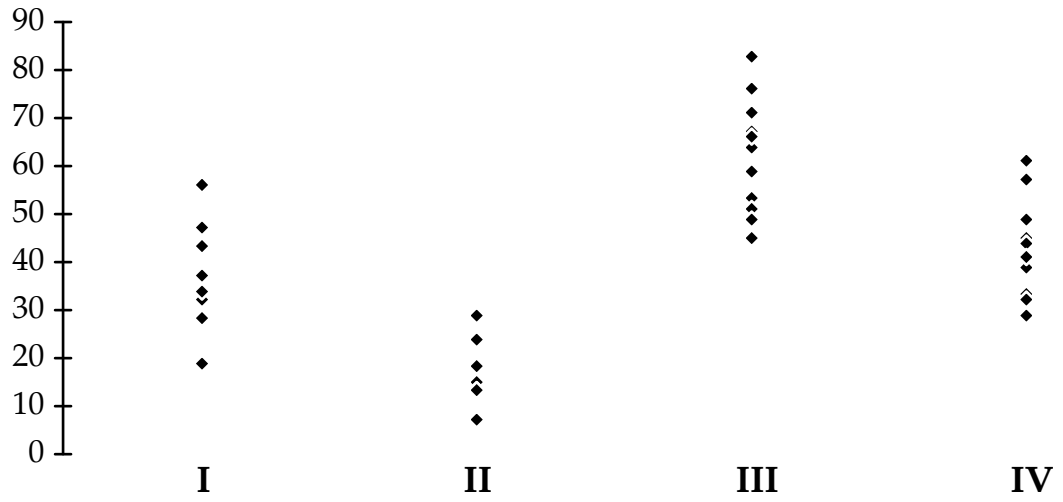
Example 2: Normal within group variance

Groups	I	II	III	IV
Data	47	18	76	33
	32	7	83	45
	28	15	45	61
	43	13	67	57
	56	24	53	44
	19	29	51	39
	37	13	48	41
	34		64	49
			66	32
			59	29
			71	
			49	
Sample size	8	7	12	10
Group means	37	17	61	43
Group standard deviations	11.56	7.42	12.12	10.53

ANOVA Table

	Degrees of freedom	Sum of squares	Mean squares	*F-ratio*
total	36	12782.7		
between	3	8902.7	2967.57	25.24
within	33	3880	117.58	

P-value = 0.0000



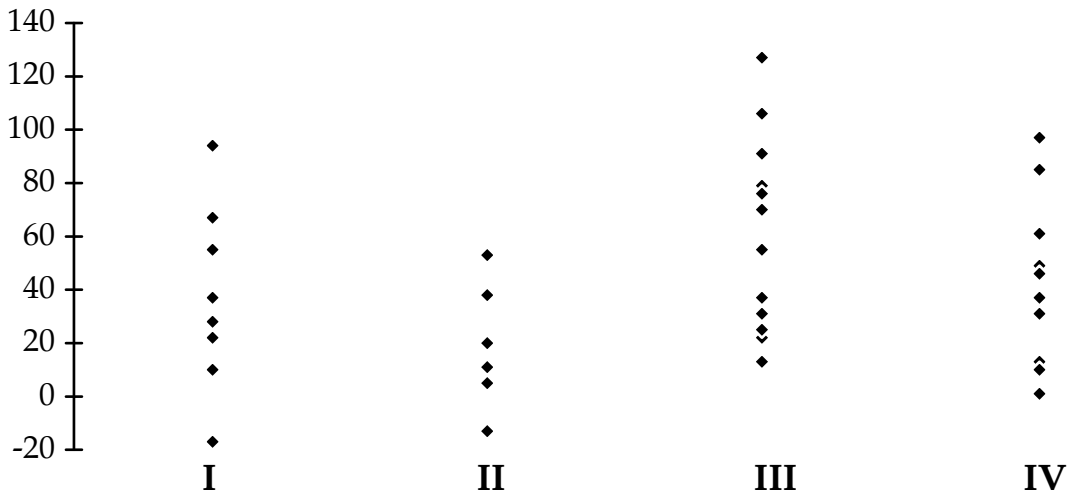
Example 3: Large within group variance

Groups	I	II	III	IV
Data	67	20	106	13
	22	-13	127	49
	10	11	13	97
	55	5	79	85
	94	38	37	46
	-17	53	31	31
	37	5	22	37
	28		70	61
			76	10
			55	1
			91	
			25	
Sample size	8	7	12	10
Group means	37	17	61	43
Group standard deviations	34.69	22.25	36.36	31.59

ANOVA Table

	Degrees of freedom	Sum of squares	Mean squares	*F-ratio*
total	36	43822.7		
between	3	8902.7	2967.57	2.804
within	33	34920	1058.18	

P-value = 0.0549



Post Hoc analysis

A large *F-ratio* indicates only that some differences exist among the group means, but not where those differences occur. If the *F-ratio* is large, our analysis would be incomplete without identifying which group means differ. This process is called **post hoc** analysis, and consists of comparing the means of all pairs of samples to determine if there is a difference of means.

Several methods are available for post hoc multiple comparisons. We will discuss the simplest method here, least significant differences. The Least Significant Difference or **LSD-method** consists of applying the two-sample *t* test to every pair of sample means. However, we make one adjustment and use the square root of mean square within rather than the pooled standard deviation from the two samples as our estimate of population standard deviation. Thus for any pair of sample means, we compute $(LSD)_{i,j}$ as,

$$(LSD)_{i,j} = \frac{\overline{X}_i - \overline{X}_j}{S_W \cdot \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}} \rightarrow t_{N-k}$$

where $S_W = \sqrt{S_W^2} = \sqrt{\frac{SS_W}{N-k}}$.

This statistic has Student distribution with *N-k* degrees of freedom.

The LSD method is illustrated for the three examples given previously:

Example 1: **Small within group variance**

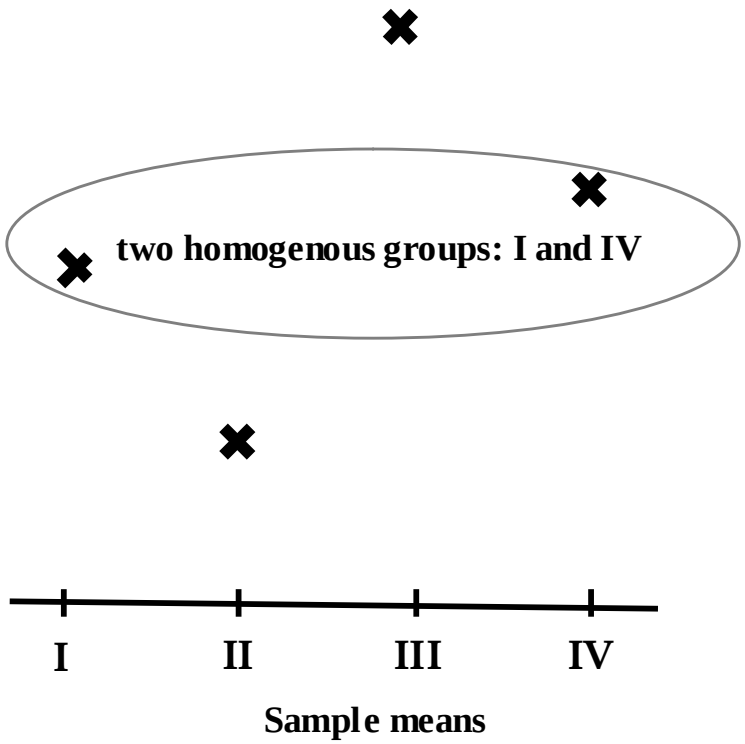
We determine $(LSD)_{i,j}$ for all pairs of given four groups and the obtained values we inscribe in the following table:

Sample sizes		8 I	7 II	12 III	10 IV
8	I	0	-7.128	9.698	2.333
7	II	7.128	0	17.064	9.731
12	III	-9.698	-17.064	0	-7.754
10	IV	-2.333	-9.731	7.7541	0

In this case, there is very strong evidence of differences between all groups except I and IV where the evidence is not as strong.

Example 2: Normal within group variance

Sample sizes		8 I	7 II	12 III	10 IV
8	I	0	-3.564	4.849	1.167
7	II	3.564	0	8.532	4.8656
12	III	-4.849	-8.532	0	-3.877
10	IV	-1.167	-4.866	3.877	0



In this case, even though the sample means are the same, there is no evidence of differences between the means of Groups I and IV.
Therefore **there are essentially three Groups: II; III; and I and IV together.**

Example 3: Large within group variance

Since the *F-ratio* for this example is very small, we would normally conclude that there is no evidence against the null hypothesis of equal group means and not proceed further. Any two-sample *t* test which produces small p-values should be regarded as spurious. However, for illustration, we have produced the table of least significant differences $(LSD)_{i,j}$:

Sample sizes		8 I	7 II	12 III	10 IV
8	I	0	-1.188	1.616	0.389
7	II	1.188	0	2.844	1.622
12	III	-1.616	-2.844	0	-1.292
10	IV	-0.389	-1.622	1.292	0

In this example, the only least significant difference which has a small **p-value** is between groups II and III. However, because the overall *F-ratio* was too small, this difference would be disregarded and we would conclude that no differences exist between the means of any of the groups.

Note

There exists other tests then LSD method which allows similar multiple comparisons what means a post hoc analysis. Also there were developed more flexible methods which are accessible thru the more advanced software (e.g. Duncan test, Tukey test for significant differences, Scheffe test and Bonferoni test). These tests are based on similar decision strategy and that's on setting of a critical difference requested for determination if two sample means from several groups are different. In many cases these tests are much more effective than LSD method.

Kruskal-Wallis test

- Described procedure ANOVA is sensitive to the assumption that the original observations are normally distributed.

- If this condition is not satistied then we must use nonparametric ***Kruskall-Wallis*** *rank test*.

- The test is based on ranks of original data values.

Kruskall – Wallis test

Formulation of H_0 a H_A:

H_0: $x_{0,5_I} = x_{0,5_{II}} = \ldots = x_{0,5_k}$

H_A: not correct H_0

Test statistics:

$$Q = \frac{12}{N \cdot (N+1)} \cdot \sum_{i=1}^{k} \frac{T_i^2}{n_i} - 3 \cdot (N+1) \rightarrow \chi^2_{k-1}$$

($N = \sum_{i=1}^{k} n_i$, T_i are sums of ranks for particular groups)

For the large within group variance data of the previous example the ranks for each group are listed in the following table:

Groups	I	II	III	IV
Ranks of original data	28	11	36	9.5
	12.5	2	37	23
	6.5	8	9.5	35
	25.5	4.5	31	32
	34	21	19	22
	1	24	16.5	16.5
	19	4.5	12.5	19
	15		29	27
			30	6.5
			25.5	3
			33	
			14	
Sample size	8	7	12	10
Mean rank	17.6875	10.7143	24.4167	19.35
St. dev.	11.1674	8.5919	9.6668	10.6538

The test statistic is a modification of calculating the *F-ratio* for the ranks. In this example, the test statistic and p-value are:

K-W test statistics = 7.24325

p-value = 0.0645

The p-value for the Kruskal-Wallis test is slightly higher than for the *F-test*, but the conclusions are the same in both cases. The null hypothesis of equal group medians is not rejected.

Let us take into account the example from the beginning of the Lecture:

Example:

Groups	I	II	III	IV
Data	67	20	106	13
	22	-13	127	49
	10	11	13	97
	55	5	79	85
	94	38	37	46
	-17	53	31	31
	37	5	22	37
	28		70	61
			76	10
			55	1
			91	
			25	
Sample size	8	7	12	10
Group means	37	17	61	43
Group St. Dev.	34.69	22.25	36.36	31.59

P-values (compare with 0.01):

Sample sizes		**8**	**7**	**12**	**10**
		I	**II**	**III**	**IV**
8	**I**	**0**	**0.11**	**0.079**	**0.35**
7	**II**		**0**	**0.0051**	**0.041**
12	**III**			**0**	**0.117**
10	**IV**				**0**

Conclusion: Reject H_0, groups II a III have significantly different means.

Now we compute ANOVA Table for this example:

ANOVA Table

	Degrees of freedom	Sum of squares	Mean squares	*F-ratio*
total	36	43822.7		
between	3	8902.7	2967.57	2.804
within	33	34920	1058.18	

P-value = 0.0549

P-value = 0.0549, so do not reject H_0,

Revised Conclusion: Groups are homogenous!!!

Conclusion

Since the *F-ratio* for this example is very small, we would normally conclude that there is no evidence against the null hypothesis of equal population means and not proceed further. Any two sample t test which produce small p-values should be regarded as **spurious**. However, for illustration, we have produced the table of least significant differences.

Sample sizes		8 I	7 II	12 III	10 IV
8	I	0	-1.188	1.616	0.389
7	II	1.188	0	2.844	1.622
12	III	-1.616	-2.844	0	-1.292
10	IV	-0.389	-1.622	1.292	0

In this example, the only least significant difference which has a small **p-value** is between groups II and III. However, because the overall *F-ratio* was too small, this difference would be disregarded (as **spurious**) and we would conclude that no differences exist between the means of any of the groups.