Analysis of Variance

Problem: what if you have more than two configurations to compare?

Assume you are the database administrator of a big information system. The database has just been installed and you are trying two tuning three configurations: Conf. **A**, Conf. **B** and Conf. **C**.

You use a given SQL package to test the execution time for each configuration.

After running several times the SQL package in both configu-rations you want to take a decision.

Conf. A exec. time	Conf. B exec. time	Conf. C exec. time
74	69	65
66	71	68
88	80	64
68	88	70
79	64	72
68	65	64
87	74	62
79	76	75
78	89	71
72	68	69
86	67	
85	72	
86		

Question: **what is the best configuration?**

Multiple comparisons

- One could be tempted to perform several two-sample t-tests for each pair of levels.
- But this leads to an increase of the (total) Type I error for H_0

$$\alpha_t = 1 - (1 - \alpha)^m$$

Assume $\alpha = 5\%$:

Number of samples	**Number of tests(m)**	**Type I error (α_t)**
2	**1**	**0.050**
3	**3**	**0.143**
4	**6**	**0.265**
5	**10**	**0.401**
10	**105**	**0.995**

One-Way ANOVA

- ANalysis Of Variance - **ANOVA**
- The one-way ANOVA is used to test the claim that three or more population means are equal
- This is an extension of the two independent samples t-test
- ANOVA tests the following hypotheses:
  - **H_0: $\mu_1 = \mu_2 = \mu_3 \ldots = \mu_k$ (the means of all the groups are equal)**
  - **H_1: Not all the means are equal**
- It does not say how or which ones differ. Need to follow up with multiple comparisons (*post-hoc* tests).

One-Way ANOVA

- **Dependent variable**: the variable you are comparing
- **Independent variable**: the factor variable being used to define the samples (groups)
- **Levels**: values of the independent variable selected to be studied. Each level will originate a sample (group)
- Example:
  - Dependent variable: **sorting time**
  - Independent variable: **elements size**
  - Levels: 5 levels
    - **10 chars** • **25 chars** • **50 chars**
    - **100 chars** • **200 chars**

Size of elements (no. characters)				
10	**25**	**50**	**100**	**200**
103	108	108	110	108
108	107	109	108	112
105	106	111	111	108
109	104	108	112	113
108	103	109	108	112
103	109	110	110	114

One-Way ANOVA

- Assumptions:
  - errors in the measurements are independent and normally distributed.
  - Variance in the measurements errors is the same for all levels
- ANOVA splits the total variation observed in all measurements:
  1. Variation observed within each system, assumed to be caused by measurement error
  2. Variation between levels

The goal is to understand whether the magnitude of component 2 of the variation **is significantly larger** that the magnitude of component 1. In words, are the differences among mean values for the levels due to the real differences between levels or due to measurement errors?

One-Way ANOVA

- Make *n* measurements on each of the *k* alternatives and organize the data collected in the following table.

Measurements	Levels						Overall mean
	1	2	…	j	…	k	
1	y_{11}	y_{12}	…	y_{1j}	…	y_{1k}	
2	y_{21}	y_{22}	…	y_{2j}	…	y_{2k}	
…	…	…	…	…	…	…	
i	y_{i1}	y_{i2}	…	y_{ij}	…	y_{ik}	
…	…	…	…	…	…	…	
n	y_{n1}	y_{n2}	…	y_{nj}	…	y_{nk}	
Column means	$\bar{y}_{.1}$	$\bar{y}_{.2}$	…	$\bar{y}_{.j}$	…	$\bar{y}_{.k}$	$\bar{y}_{..}$
Effects	α_1	α_2	…	α_j	…	α_k	

Example One-Way ANOVA

How does the size of the elements in an array affect the time necessary to sort all the elements of the array?

For economy of space, we will consider only 3 level of the independent variable size of elements: 3, 5 and 10 characters.

The table in the next slide shows the sorting times in milliseconds obtained with the Quicksort and an array of 10000 elements. 10 replications were performed for each size of elements.

Example One-Way ANOVA

Measurements	Levels			Overall mean
	1	2	3	
1	95	100	103	
2	93	100	108	
3	98	103	105	
4	95	103	109	
5	99	103	108	
6	100	100	103	
7	100	99	108	
8	95	105	108	
9	100	101	102	
10	102	105	103	
Column means	97.7	101.9	105.7	101.8
Effects	-4.07	0.13	3.94	

One-way ANOVA

- Each measurement y_{ij} is the sum of the mean of all measurements at alternative j, $\bar{y}_{.j}$, and a value e_{ij} that represents the deviation of measurement y_{ij} from the mean

$$y_{ij} = \bar{y}_{.j} + e_{ij}$$

- Each column mean $\bar{y}_{.j}$ is the sum of the overall mean, $\bar{y}_{..}$, and the deviation of the column mean from this overall mean, *effect* α_j

$$\bar{y}_{.j} = \bar{y}_{..} + \alpha_j$$

- Then, each measurement y_{ij} can be expressed as follows

$$y_{ij} = \bar{y}_{..} + \alpha_j + e_{ij}$$

One-way ANOVA

- Then, each measurement y_{ij} can be expressed as follows

$$y_{ij} = \bar{y}_{..} + \alpha_j + e_{ij}$$

- In this model, we split the total variation in all measurements into two components:
  1. Variation due to the effects of the levels
  2. Variation due to the errors

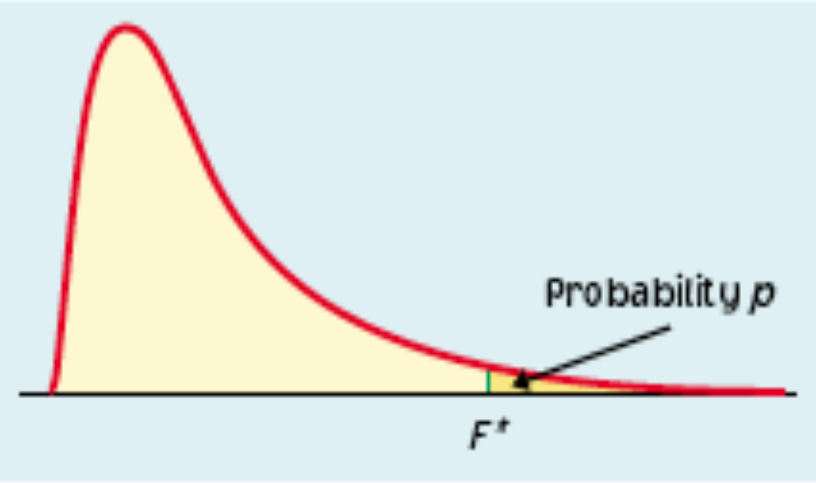
- Variation due to the effects: SSA = $n \sum_{j=1}^{k} (\bar{y}_{.j} - \bar{y}_{..})^2$
- Variation due to errors: SSE = $\sum_{j=1}^{k} \sum_{i=1}^{n} (y_{ij} - \bar{y}_{.j})^2$
- Total Variation: SST = SSA + SSE = $\sum_{j=1}^{k} \sum_{i=1}^{n} (y_{ij} - \bar{y}_{..})^2$

One-way ANOVA: the F statistics

- If the differences among levels are due to some real difference between levels, we expect SSA to be "larger" than SSE.
- The statistic for ANOVA is called the **F statistic**, which we obtain from the F Test.
- The F statistic determines if the variation between sample means is significant:

$$F = \frac{SSA/(k-1)}{SSE/(N-k)}$$

where N is the total number of observations



Obtaining the critical from F table

How to obtain the critical value from F tables for a given α? For example, for α = 0.05?

$$F = \frac{SSA/(k-1)}{SSE/(N-k)}$$

***df* for the numerator**

Example: For 3 levels, *df* = k - 1 = 2

***df* for the denominator**

Example: For 30 observations
***df* = N - k = 30 - 3 = 27**

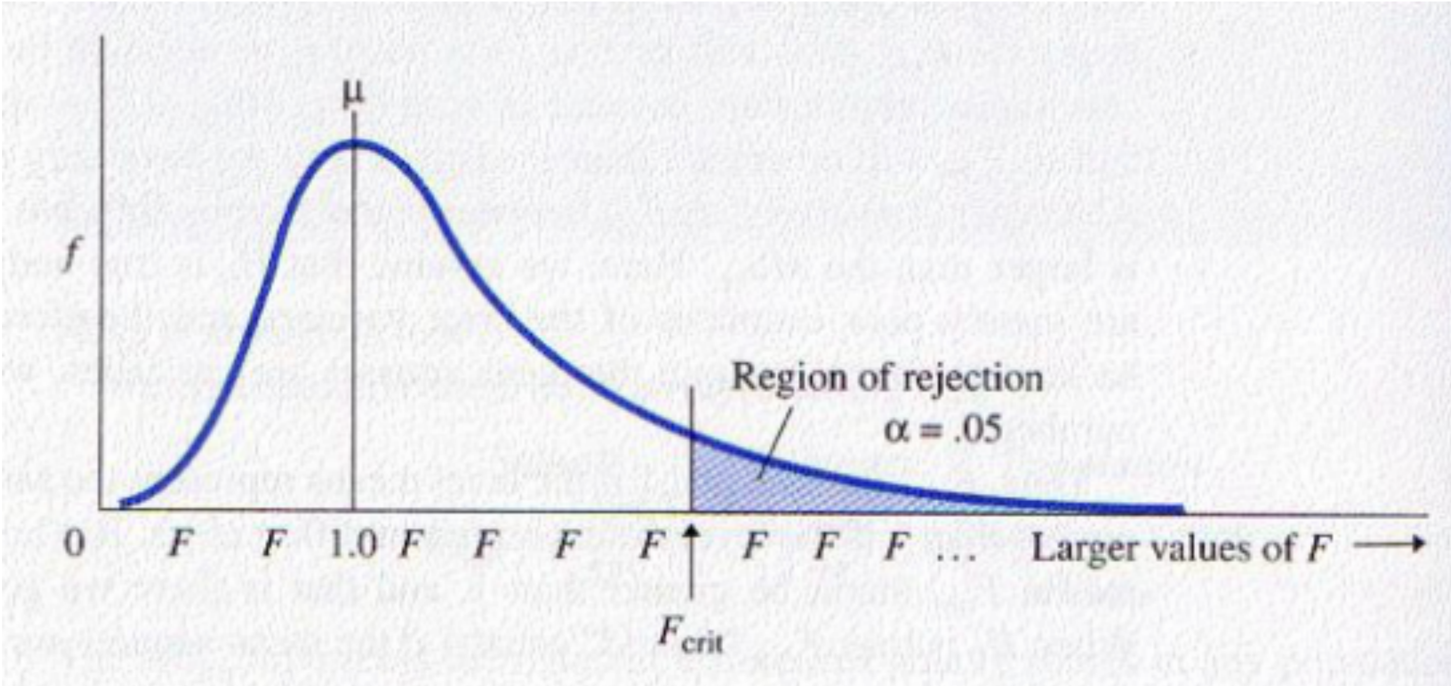
In the F table for α = 0.05, look for the critical value for $F_{(2,27)}$

Obtaining the critical from F table

How to obtain the critical value from F tables for a given α? For example, for $\alpha = 0.05$?

Critical value for $\alpha = 0.05$

and $F_{(2,27)} = 3.354$



In the F table for $\alpha = 0.05$, look for the critical value for $F_{(2,27)}$

The One-way ANOVA table

- The ANOVA table is a summary of all the elements needed for the calculation of the P-value

	SS	**df**	**MS**	**F**	**P**
Between	SSA	k-1	MSA = SSA/(k-1)	MSA/MSE	Tail are above F
Within	SSE	N-k	MSE = SSE/(N-k)		
Total	SST	N-1			

One-way ANOVA steps

Use the same general steps of hypothesis testing

1. State the hypothesis
2. Calculations (to compute the test statistic)
3. Obtain p value
4. Make a decision

Example One-Way ANOVA

Measurements	Levels			Overall mean
	1	2	3	
1	95	100	103	
2	93	100	108	
3	98	103	105	
4	95	103	109	
5	99	103	108	
6	100	100	103	
7	100	99	108	
8	95	105	108	
9	100	101	102	
10	102	105	103	
Column means	97.7	101.9	105.7	101.8
Effects	-4.07	0.13	3.94	

Example: one-way ANOVA
Step 1- State the hypothesis

Hypothesis

- **H_0: $\mu_1 = \mu_2 = \mu_3$** → All samples have equal means
- **H_1: $\exists(i, j)$: $\mu_i \neq \mu_j$** → At least two means are different
  - Does not say how or which ones differ

Example: one-way ANOVA
Step 2 – Calculations (cont.)

We can now fill the one-way ANOVA table

	SS	**Df**	**MS**	**F**	**P**
Between	320.3	2	160.1	22.625	1.695e-06
Within	191.1	27	7.078		
Total	511.4	29			

The P-Value for F = 22.625 is 1.695e-06 → **H_0 is Rejected** for α=0.05 !

Example: one-way ANOVA

Step 2 – Calculations (cont.)

We can now fill the one-way ANOVA table

	SS	Df	MS	F	P
Between	320.3	2	160.1	22.625	1.695e-06
Within	191.1	27	7.0		
Total	511.4	29			

So, where is the difference? We need to perform post-hoc analysis.

The P-Value for F = 22.625 is 1.695e-06 → **H_0 is Rejected** for α=0.05 !

Post-hoc analysis

- Once H_0 is rejected, we must find which pair(s) is statistically different.

Pairwise t-test with multiple comparisons correction.

- Perform t-test for each pair of levels. However, as several comparisons are performed, the p-value needs to be corrected (e.g Bonferroni, Holm procedures)

Example with Bonferroni:

Bonferroni correction multiplies the p-values by the number of tests. Holm's is less conservative.

	Level 1	**Level 2**
Level 2	0.0045	-
Level 3	9.7e-07	0.0107

Post-hoc analysis

- Once H_0 is rejected, we must find which pair(s) is statistically different.

Pairwise t-test with multiple comparisons correction.

- Perform t-test for each pair of levels. However, as several comparisons are performed, the p-value needs to be corrected (e.g Bonferroni, Holm procedures)

Example with Bonferroni:

Bonferroni correction multiplies the

For α=0.01, Level 1 differs from the remaining levels.

	Level 1	**Level 2**
Level 2	**0.0045**	-
Level 3	**9.7e-07**	0.0107

Post-hoc analysis

- Once H_0 is rejected, we must find which pair(s) is statistically different.

Tukey HSD (honest significant difference)

- Similar to the pairwise t-test but using a different distribution (*studentized range distribution*).

	Level 1	**Level 2**
Level 2	0.0041	-
Level 3	9.7e-07	0.0096

Post-hoc analysis

- Once H_0 is rejected, we must find which pair(s) is statistically different.

Tukey HSD (honest significant difference)

- Similar to the pairwise t-test but using a different distribution (*studentized range distribution*).

For α=0.01 all levels differ.

	Level 1	Level 2
Level 2	**0.0041**	-
Level 3	**9.7e-07**	**0.0096**

Assumptions of one-way ANOVA

The errors e_{ij} are independent and normally distributed, with mean 0 and constant variance.

$$e_{ij} = y_{ij} - \bar{y}_i .$$

Variance can be checked with **Bartlett Test of Homogeneity of Variances**

Normality of the errors (residuals) can be checked with **Shapiro test of normality**

Kruskal–Wallis one-way analysis of variance

- If ANOVA's assumptions are not met, it is possible to use the non-parametric test Kruskal–Wallis one-way analysis of variance, which extends Mann–Whitney's test for more than 2 samples.
- Post-hoc analysis can be performed with Dunn's test with multiple comparisons correction.

Kruskal Wallis one-way analysis of variance: p-value = 7.719e-05

Dunn's test with Holm's Correction:

	Level 1	**Level 2**
Level 2	0.0105	-
Level 3	0.0000	0.0207

Kruskal–Wallis one-way analysis of variance

- If ANOVA's assumptions are not met, it is possible to use the non-parametric test Kruskal–Wallis one-way analysis of variance, which extends Mann–Whitney's test for more than 2 samples.
- Post-hoc analysis can be performed with Dunn's test with multiple comparisons correction.

Kruskal Wallis one-way analysis of variance: p-value = 7.719e-05

Dunn's test with Holm's Correction: Level 1 differs only from level 3 for α=0.01

	Level 1	**Level 2**
Level 2	0.0105	-
Level 3	**0.0000**	0.0207

Example: One-Way ANOVA in R

Measurements	System		
	A	B	C
1	0.0952	0.1282	0.0966
2	0.0871	0.1432	0.1200
3	0.0969	0.1382	0.1152
4	0.1054	0.1330	0.1375
5	0.0874	0.1383	0.1298

Measurements of the time (in microseconds) required to perform a subroutine call and return on three different systems

Adapted from the example in Chapter 5.2 from Lilja's book

Example: One-Way ANOVA in R
Step 1- State the hypothesis

Hypothesis

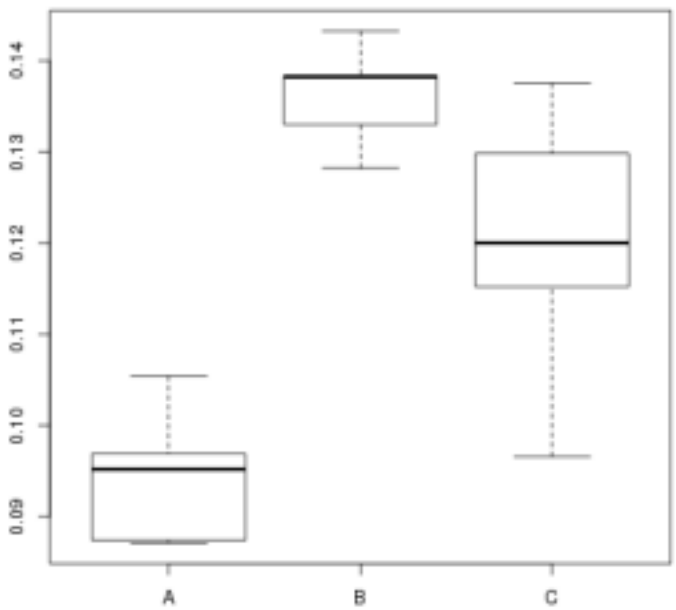
- **H_0: $\mu_1 = \mu_2 = \mu_3$** → All samples have equal means
- **H_1: $\exists(i, j)$: $\mu_i \neq \mu_j$** → At least two means are different

Example: One-Way ANOVA in R
Exploratory data analysis

Data file

```
Time System
0.0952 A
0.0871 A
0.0969 A
0.1054 A
0.0874 A
0.1282 B
0.1432 B
0.1382 B
0.1330 B
0.1383 B
0.0966 C
0.1200 C
0.1152 C
0.1375 C
0.1298 C
```

Check Boxplot




```
D = read.table("data.in",header=TRUE)

boxplot( Time ~ System, data=D)
```

Note: use letters to denote different non-numeric levels of a factor

Example: One-Way ANOVA in R
Step 2 – Calculations

We can now fill the one-way ANOVA table

```
> aov.out = aov(Time ~ System, data = D)
> summary(aov.out)

             Df   Sum Sq   Mean Sq F value   Pr(>F)
System        2 0.004432 0.0022162   19.93 0.000154 ***
Residuals    12 0.001335 0.0001112
---
```

Rejection of H_0 for a significance level of 5%

There is at least one difference between groups.

But we need to check assumptions

Example: One-Way ANOVA in R
Check assumptions

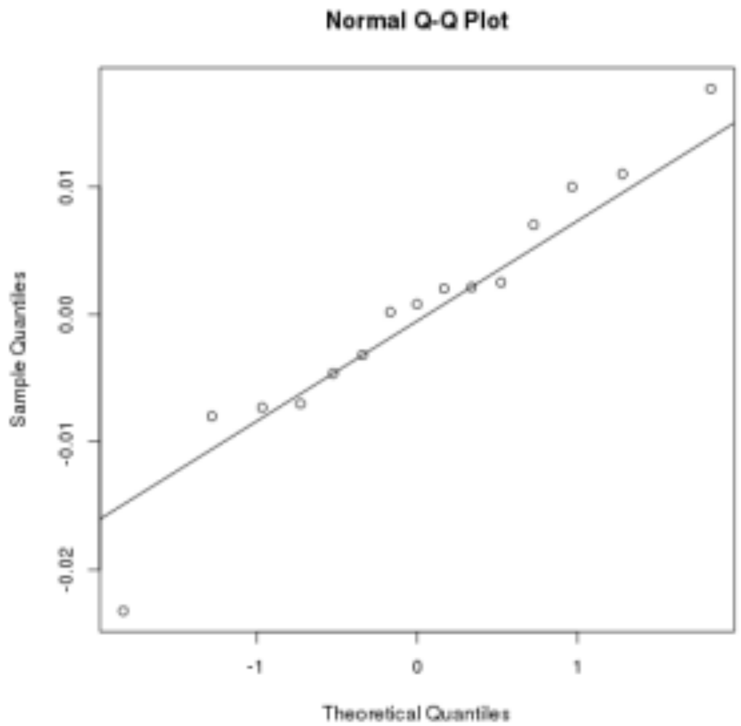
Test normality of the residuals

```
> qqnorm(aov.out$res)
> qqline(aov.out$res)

> shapiro.test(aov.out$res)

  Shapiro-Wilk normality test

W = 0.9599, p-value = 0.69
```



- We assume normality

Example: One-Way ANOVA in R

Check assumptions

Test homogeneity of variances

```
> bartlett.test(Time~System, data=D)

        Bartlett test of homogeneity of variances

data:  Time by System
Bartlett's K-squared = 3.9389, df = 2, p-value = 0.1395
```

- We assume that variances are similar.

Example: One-Way ANOVA in R

Post-hoc analysis

Pairwise t-test with Bonferroni correction

```
 > pairwise.t.test(D$Time, D$System, p.adjust.method="bonf",
alternative="two.sided")

        Pairwise comparisons using t tests with pooled SD

data:  D$Time and D$System

  A         B
B 0.00012   -
C 0.00744   0.09130
```

Level A is significant at 5% significant level.

Example: One-Way ANOVA in R
Post-hoc analysis

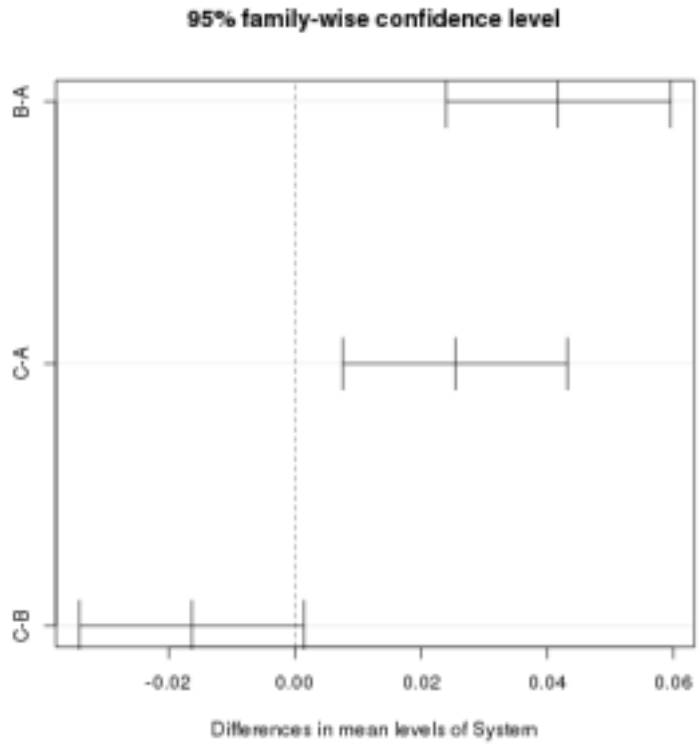
Tukey HSD

```
> t = TukeyHSD(aov.out,alternative="two.sided")
> plot(t)

 Tukey multiple comparisons of means
    95% family-wise confidence level

Fit: aov(formula = Time ~ System, data = D)

$System
          diff          lwr         upr     p adj
B-A  0.04178  0.023986089 0.059573911 0.0001142
C-A  0.02542  0.007626089 0.043213911 0.0064758
C-B -0.01636 -0.034153911 0.001433911 0.0727016
```



Level A is significant at 5% significant level.

Kruskal–Wallis rank sum test

- When the ANOVA assumptions are not met, it is possible to use the non-parametric test Kruskal–Wallis rank sum test.

```
> > kruskal.test(Time~System, data = D)

        Kruskal-Wallis rank sum test

data:  Time by System
Kruskal-Wallis chi-squared = 10.14, df = 2, p-value = 0.006282
```

There exists at least a significant difference at 5% significant level.

Example: One-Way ANOVA in R

Post-hoc analysis

Dunn test (with Bonferroni's correction)

```
> library(dunn.test)
> dunn.test(D$Time,g=D$System,method="bonferroni")

Col Mean-|
Row Mean |            A            B
---------+----------------------
       B |     3.181980
         |       0.0022
         |
       C |     1.697056  -1.484924
         |       0.1345     0.2063
```

Only levels A and B are significant at 5% significant level.