

## ΣΤ . ΑΝΑΛΥΣΗ ΔΙΑΣΠΟΡΑΣ (ANALYSIS OF VARIANCE - ANOVA)

### ΣΤ<sub>1</sub> . Ανάλυση Διασποράς κατά μία κατεύθυνση.

Όπως έχουμε δει στη παράγραφο Β<sub>2</sub>, όταν θέλουμε να ελέγξουμε, αν η μέση τιμή μιας ποσοτικής μεταβλητής διαφέρει ανάμεσα σε δύο ανεξάρτητα δείγματα, χρησιμοποιούμε το t-test. Για παράδειγμα, αν η μέση ηλικία των γυναικών από την Αθήνα ( $\mu_1$ ), διαφέρει από την μέση ηλικία των γυναικών από την Θεσσαλονίκη ( $\mu_2$ ). Αυτό, είναι ισοδύναμο με το να πούμε ότι *ελέγχουμε την διαφορά των μέσων τιμών μίας ποσοτικής μεταβλητής (ηλικία), ανάμεσα στις κατηγορίες μίας ποιοτικής (τόπος γέννησης)*. Στην περίπτωση όμως που η ποιοτική μεταβλητή, έχει περισσότερες από δύο κατηγορίες, δεν εφαρμόζουμε το t-test.

Προκειμένου να ελέγξουμε το αν διαφέρουν οι μέσες τιμές μίας ποσοτικής μεταβλητής, ανάμεσα στις κατηγορίες μιας ποιοτικής, όταν αυτή έχει περισσότερες από δύο κατηγορίες (έστω  $k$ ), χρησιμοποιούμε την *Ανάλυση Διασποράς μιας κατεύθυνσης (One-way ANOVA)*.<sup>1</sup> Εδώ δηλαδή, θέλουμε να ελέγξουμε τη:

**Μηδενική υπόθεση:**  $\mu_1 = \mu_2 = \dots = \mu_k$

έναντι της

**Εναλλακτικής υπόθεσης:**  $\mu_i \neq \mu_j$ ,  $i, j = 1, 2, \dots, k$  (τουλάχιστον ένα ζευγάρι διαφέρει).

#### Προϋποθέσεις<sup>2</sup> :

- Η ποσοτική μεταβλητή να κατανέμεται κανονικά, σε κάθε κατηγορία της ποιοτικής.
- Οι διασπορές της ποσοτικής μεταβλητής, σε κάθε κατηγορία της ποιοτικής, να είναι ίσες.
- Οι  $k$  - ομάδες ατόμων ( $k$  - δείγματα) να είναι ανεξάρτητες.

Αν οι προϋποθέσεις αυτές δεν ισχύουν, τότε ανατρέχουμε σε μη - παραμετρικό τεστ.

---

<sup>1</sup> Το t-test, δεν είναι τίποτα άλλο από υποπερίπτωση της ανάλυσης διασποράς, για δύο κατηγορίες.

<sup>2</sup> Αντίστοιχες του t-test αλλά για  $k$ -κατηγορίες!

## SPSS

### 1) *One – Way ANOVA (Παραμετρικό τεστ)*

Ας υποθέσουμε ότι θέλουμε να ελέγξουμε αν η μέση τιμή του βάρους των γυναικών, διαφέρει ανάλογα με την βαθμίδα εκπαίδευσης. Δηλαδή αν η μέση τιμή της wht, διαφέρει σε κάθε κατηγορία της edu\_cat.

Προκειμένου να ελέγξουμε το αν το βάρος κατανέμεται κανονικά, σε κάθε βαθμίδα εκπαίδευσης, θα χωρίσουμε τα δεδομένα μας αρχικά ανά κατηγορία της edu\_cat χρησιμοποιώντας την εντολή **split** :

**Data → Split File → Compare groups → Group based on:** και βάζουμε την μεταβλητή σύμφωνα με την οποία θα γίνει ο διαχωρισμός (edu\_cat) → **Ok**

Μετά εφαρμόζουμε τον έλεγχο των Kolmogorov –Smirnov.

One-Sample Kolmogorov-Smirnov Test

EDU_CAT			Βάρος (πριν)
Πρωτοβάθμια εκπαίδευση	N		89
	Normal Parameters <sup>a,b</sup>	Mean	67,72
		Std. Deviation	11,74
	Most Extreme Differences	Absolute	,108
		Positive	,108
		Negative	-,061
	Kolmogorov-Smirnov Z		1,022
Asymp. Sig. (2-tailed)		,247	
Δευτεροβάθμια εκπαίδευση	N		60
	Normal Parameters <sup>a,b</sup>	Mean	66,88
		Std. Deviation	10,15
	Most Extreme Differences	Absolute	,101
		Positive	,101
		Negative	-,054
	Kolmogorov-Smirnov Z		,785
Asymp. Sig. (2-tailed)		,569	
Τριτοβάθμια εκπαίδευση	N		25
	Normal Parameters <sup>a,b</sup>	Mean	61,92
		Std. Deviation	10,63
	Most Extreme Differences	Absolute	,128
		Positive	,128
		Negative	-,055
	Kolmogorov-Smirnov Z		,641
Asymp. Sig. (2-tailed)		,806	

a. Test distribution is Normal.

b. Calculated from data.

Η μεταβλητή βάρος κατανέμεται κανονικά και στις τρεις βαθμίδες της εκπαίδευσης. Μπορούμε να προχωρήσουμε λοιπόν στην ανάλυση διασποράς.

**Analyze → General linear model → Univariate → Dependent Variables:** βάζουμε την ποσοτική μεταβλητή, **Fixed factors:** βάζουμε την ποιοτική μεταβλητή

**Post hoc → Post hoc test for:** edu\_cat → ενεργοποιούμε τα: LSD, Bonferroni & Tukey → **Continue**

**Save → Residuals → ενεργοποιούμε τα: Unstandardized → Continue**

**Options → Display means for:** edu\_cat → **Display :** ενεργοποιούμε τα: Descriptive statistics, Estimates of effect size, Parameter estimates & Homogeneity tests → **Continue**

→ **Ok**

#### Levene's Test of Equality of Error Variances<sup>a</sup>

Dependent Variable: Βάρος (πριν)

F	df1	df2	Sig.
,930	2	171	,396

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+EDU\_CAT

#### σχόλιο:

Το Levene's Test of Equality of Error Variances κάτω από τη μηδενική υπόθεση:

οι διασπορές είναι ίσες ( $\sigma^2_{\text{πρωτοβ.}} = \sigma^2_{\text{δευτεροβ.}} = \sigma^2_{\text{τριτοβ.}}$ ),

δίνει p-value = 0.396. Κατά συνέπεια ισχύει και η προϋπόθεση της ισότητας των διασπορών.

#### Tests of Between-Subjects Effects

Dependent Variable: Βάρος (πριν)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Corrected Model	663,838 <sup>a</sup>	2	331,919	2,713	,069	,031
Intercept	568771,371	1	568771,371	4648,690	,000	,965
EDU_CAT	663,838	2	331,919	2,713	,069	,031
Error	20922,001	171	122,351			
Total	793320,000	174				
Corrected Total	21585,839	173				

a. R Squared = ,031 (Adjusted R Squared = ,019)

**σχόλιο:**

**F-test:** Ο έλεγχος γίνεται κάτω από την μηδενική υπόθεση ότι οι μέσες τιμές του βάρους είναι ίσες για κάθε βαθμίδα της εκπαίδευσης.

**Parameter Estimates**

Dependent Variable: Βάρος (πριν)

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval		Eta Squared
					Lower Bound	Upper Bound	
Intercept	61,920	2,212	27,990	,000	57,553	66,287	,821
[EDU_CAT=1]	5,799	2,504	2,316	,022	,857	10,741	,030
[EDU_CAT=2]	4,963	2,633	1,885	,061	-,234	10,161	,020
[EDU_CAT=3]	0 <sup>a</sup>	,	,	,	,	,	,

a. This parameter is set to zero because it is redundant.

**σχόλιο:**

Το μοντέλο που αντιστοιχεί στον παραπάνω πίνακα είναι το:

$$Y_i = \mu + b_{EDU\_CAT=i}, \quad i = 1, 2 \text{ ή } 3$$

όπου

- στη θέση του  $i$ , βάζουμε το επίπεδο εκείνο της κατηγορικής μας, για τα άτομα του οποίου θέλουμε να υπολογίσουμε την αναμενόμενη μέση τιμή ( $Y_i$ )
- 
- $\mu$  είναι η σταθερά (intercept), η οποία αντιστοιχεί στην αναμενόμενη μέση τιμή του βάρους για τις γυναίκες που έχουν τελειώσει την τριτοβάθμια εκπαίδευση (επίπεδο αναφοράς)

$b_{EDU\_CAT=1} = 5.799 \rightarrow$  Η μέση τιμή του βάρους των γυναικών που έχουν τελειώσει την πρωτοβάθμια εκπαίδευση είναι κατά 5.8 κιλά μεγαλύτερη από τη μέση τιμή του βάρους των γυναικών που έχουν τελειώσει την τριτοβάθμια (επίπεδο αναφοράς). Το αντίστοιχο one sample t-test (p-value = 0.022) δείχνει ότι η διαφορά αυτή στις μέσες τιμές είναι στατιστικά σημαντική.

$b_{EDU\_CAT=2} = 4.963 \rightarrow$  Η μέση τιμή του βάρους των γυναικών που έχουν τελειώσει την δευτεροβάθμια εκπαίδευση είναι κατά 5 κιλά μεγαλύτερη από τη μέση τιμή του βάρους των γυναικών που έχουν τελειώσει την τριτοβάθμια (επίπεδο αναφοράς). Το αντίστοιχο one sample t-test (p-value = 0.061) δείχνει ότι η διαφορά αυτή στις μέσες τιμές δεν είναι στατιστικά σημαντική.

### Multiple Comparisons

Dependent Variable: Βάρος (πριν)

LSD

(I) EDU_CAT	(J) EDU_CAT	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Πρωτοβάθμια εκπαίδευση	Δευτεροβάθμια εκπαίδευση	,84	1,85	,652	-2,81	4,48
	Τριτοβάθμια εκπαίδευση	5,80*	2,50	,022	,86	10,74
Δευτεροβάθμια εκπαίδευση	Πρωτοβάθμια εκπαίδευση	-,84	1,85	,652	-4,48	2,81
	Τριτοβάθμια εκπαίδευση	4,96	2,63	,061	-,23	10,16
Τριτοβάθμια εκπαίδευση	Πρωτοβάθμια εκπαίδευση	-5,80*	2,50	,022	-10,74	-,86
	Δευτεροβάθμια εκπαίδευση	-4,96	2,63	,061	-10,16	,23

Based on observed means.

\*. The mean difference is significant at the ,05 level.

### σχόλιο:

Στον παραπάνω πίνακα μπορούμε να δούμε αναλυτικά τις διαφορές στις μέσες τιμές σε όλους τους δυνατούς συνδυασμούς ανά δύο των κατηγοριών της ποιοτικής. Το αστεράκι σε κάποια διαφορά, μας πληροφορεί ότι είναι στατιστικά σημαντική, στο επίπεδο του 5 %.

### Εναλλακτικός τρόπος για την εφαρμογή της One-way ANOVA

Analyze → Compare Means → One way anova → και δηλώνουμε την ποσοτική μεταβλητή (depended) και τη ποιοτική (Factor) → Post hoc → Equal variances Assumed και ενεργοποιούμε τα: LSD, Bonferroni & Tukey → Continue → Ok

### 2) Μη - Παραμετρικό τεστ

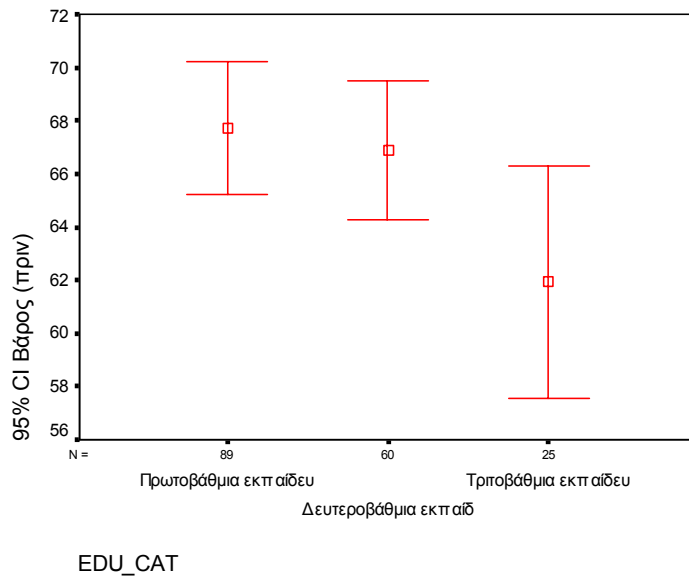
Στην περίπτωση που δεν ισχύουν οι προϋποθέσεις, καταφεύγουμε σε μη παραμετρικό έλεγχο. Τότε

### SPSS

Analyze → Non parametric tests → k Independent Samples test → Test variable list: την ποσοτική μεταβλητή, Grouping variable: την ποιοτική, στο Define Range δηλώνουμε το εύρος των τιμών της ποιοτικής μεταβλητής και επιλέγουμε το Kruskal– Wallis test → Ok.

## Γραφική Απεικόνιση

**Graphs** → **Error Bar** → **Simple** → **Define** → **Y axis**: βάζουμε την ποσοτική μεταβλητή, **X Axis**: βάζουμε την ποιοτική μεταβλητή → **Ok**



## ΣΤ<sub>2</sub> . Ανάλυση Διασποράς κατά δύο κατευθύνσεις.

Προκειμένου να ελέγξουμε το αν διαφέρουν οι μέσες τιμές μίας ποσοτικής μεταβλητής, ανάμεσα στις κατηγορίες όχι μίας ποιοτικής όχι αλλά δύο χρησιμοποιούμε την *Ανάλυση Διασποράς δύο κατευθύνσεων (Two-way ANOVA)*. Εδώ, για κάθε ποιοτική μεταβλητή ξεχωριστά θα γίνουν οι παρακάτω έλεγχοι:

**Μηδενική υπόθεση:**  $\mu_1 = \mu_2 = \dots = \mu_k$ .

έναντι της

**Εναλλακτικής υπόθεσης:**  $\mu_i \neq \mu_j$ ,  $i, j = 1, 2, \dots, k$  (τουλάχιστον ένα ζευγάρι διαφέρει), όπου η  $i$  στους δείκτες των μέσων τιμών δηλώνει ότι η άλλη μεταβλητή παραμένει σταθερή.

### Προϋποθέσεις :

- Οι παρατηρήσεις να προέρχονται από κανονική κατανομή.
- Οι διασπορές των συγκρινόμενων κανονικών κατανομών, να είναι ίσες.
- Οι προκύπτουσες ομάδες ατόμων να είναι ανεξάρτητες.

## SPSS

### *Two – Way ANOVA*

Ας υποθέσουμε ότι θέλουμε να ελέγξουμε αν η μέση τιμή του βάρους των γυναικών, διαφέρει ανάλογα με την βαθμίδα εκπαίδευσης και τον τόπο γέννησης. Δηλαδή αν η μέση τιμή της *wht*, διαφέρει σε κάθε κατηγορία της *edu\_cat* και της *bp*.

**Analyze → General linear model → Univariate → Dependent Variables:** βάζουμε την ποσοτική μεταβλητή, **Fixed factors:** βάζουμε τις ποιοτικές μεταβλητές

**Model → Custom** και στο πλαίσιο **Model:** *edu\_cat*, *bp*, **Built terms:** Main Effects → **Continue**

**Post hoc → Post hoc test for:** *edu\_cat*, *bp* → ενεργοποιούμε τα: LSD, Bonferroni & Tukey → **Continue**

Save → Residuals → ενεργοποιούμε τα: Unstandardized → Continue

Options → Display means for: edu\_cat, bp → Display : ενεργοποιούμε τα: Descriptive statistics, Estimates of effect size, Parameter estimates & Homogeneity tests → Continue

→ Ok

### Levene's Test of Equality of Error Variances

Dependent Variable: Βάρος (πριν)

F	df1	df2	Sig.
1,188	5	168	,317

Tests the null hypothesis that the error variance of the dependent variable is equal across groups.

a. Design: Intercept+BP+EDU\_CAT

**σχόλιο:**

Το Levene's Test of Equality of Error Variances, κάτω από τη μηδενική υπόθεση της ισότητας των διασπορών, δίνει  $p$ -value = 0.317. Κατά συνέπεια ισχύει και η προϋπόθεση της ισότητας των διασπορών.

### Tests of Between-Subjects Effects

Dependent Variable: Βάρος (πριν)

Source	Type III Sum of Squares	df	Mean Square	F	Sig.	Eta Squared
Corrected Model	2033,316 <sup>a</sup>	3	677,772	5,893	,001	,094
Intercept	402851,666	1	402851,666	3502,606	,000	,954
BP	1369,477	1	1369,477	11,907	,001	,065
EDU_CAT	1008,597	2	504,298	4,385	,014	,049
Error	19552,523	170	115,015			
Total	793320,000	174				
Corrected Total	21585,839	173				

a. R Squared = ,094 (Adjusted R Squared = ,078)



**σχόλιο:**

**F(Corrected model):** Ο έλεγχος γίνεται κάτω από την μηδενική υπόθεση ότι η μέση τιμή της wht, διαφέρει σε κάθε κατηγορία της edu\_cat και της bp.

**F(Edu\_cat):** Ο έλεγχος γίνεται κάτω από την μηδενική υπόθεση ότι η μέση τιμή της wht, διαφέρει σε κάθε κατηγορία της edu\_cat, λαμβάνοντας υπόψη και τη μεταβλητή bp.

**F(Bp):** Ο έλεγχος γίνεται κάτω από την μηδενική υπόθεση ότι η μέση τιμή της wht, διαφέρει σε κάθε κατηγορία της bp, λαμβάνοντας υπόψη και τη μεταβλητή edu\_cat.

#### Parameter Estimates

Dependent Variable: Βάρος (πριν)

Parameter	B	Std. Error	t	Sig.	95% Confidence Interval		Eta Squared
					Lower Bound	Upper Bound	
Intercept	56,147	2,720	20,641	,000	50,778	61,517	,715
[BP=1]	6,560	1,901	3,451	,001	2,807	10,312	,065
[BP=2]	0 <sup>a</sup>	,	,	,	,	,	,
[EDU_CAT=1]	7,297	2,466	2,959	,004	2,429	12,165	,049
[EDU_CAT=2]	5,379	2,556	2,105	,037	,334	10,424	,025
[EDU_CAT=3]	0 <sup>a</sup>	,	,	,	,	,	,

a. This parameter is set to zero because it is redundant.

**σχόλιο:**

Το μοντέλο που αντιστοιχεί στον παραπάνω πίνακα είναι το:

$$Y_i = \mu + b_{BP=j} + b_{EDU\_CAT=i}, \quad j = 1 \text{ ή } 2 \text{ και } i = 1, 2 \text{ ή } 3$$

όπου

- στη θέση του  $i$ , βάζουμε το επίπεδο εκείνο της κατηγορικής μας, για τα άτομα του οποίου θέλουμε να υπολογίσουμε την αναμενόμενη μέση τιμή ( $Y_i$ )
- $\mu$  είναι η σταθερά (intercept), η οποία αντιστοιχεί στην αναμενόμενη μέση τιμή του βάρους για τις γυναίκες που έχουν τελειώσει την τριτοβάθμια εκπαίδευση και είναι από την Θεσσαλονίκη (επίπεδο αναφοράς).

**σχόλιο:**

$b_{BP=1} = 6.560 \rightarrow$  Η μέση τιμή του βάρους των γυναικών που είναι από την Αθήνα, είναι κατά 6.5 κιλά μεγαλύτερη, από τη μέση τιμή του βάρους των γυναικών είναι από την Θεσσαλονίκη (επίπεδο αναφοράς), δεδομένου ότι έχουν την ίδια βαθμίδα

εκπαίδευσης. Το αντίστοιχο one sample t-test (p-value = 0.001) δείχνει ότι η διαφορά αυτή στις μέσες τιμές είναι στατιστικά σημαντική.

$b_{EDU\_CAT=1} = 7.297 \rightarrow$  Η μέση τιμή του βάρους των γυναικών που έχουν τελειώσει την πρωτοβάθμια εκπαίδευση είναι κατά 7.3 κιλά μεγαλύτερη από τη μέση τιμή του βάρους των γυναικών που έχουν τελειώσει την τριτοβάθμια (επίπεδο αναφοράς), δεδομένου ότι κατάγονται από την ίδια πόλη. Το αντίστοιχο one sample t-test (p-value = 0.004) δείχνει ότι η διαφορά αυτή στις μέσες τιμές είναι στατιστικά σημαντική.

$b_{EDU\_CAT=2} = 5.379 \rightarrow$  Η μέση τιμή του βάρους των γυναικών που έχουν τελειώσει την δευτεροβάθμια εκπαίδευση είναι κατά 5.4 κιλά μεγαλύτερη από τη μέση τιμή του βάρους των γυναικών που έχουν τελειώσει την τριτοβάθμια (επίπεδο αναφοράς), δεδομένου ότι κατάγονται από την ίδια πόλη. Το αντίστοιχο one sample t-test (p-value = 0.037) δείχνει ότι η διαφορά αυτή στις μέσες τιμές είναι στατιστικά σημαντική.

### Multiple Comparisons

Dependent Variable: Βάρος (πριν)

LSD

(I) EDU_CAT	(J) EDU_CAT	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Πρωτοβάθμια εκπαίδευση	Δευτεροβάθμια εκπαίδευση	,84	1,79	,641	-2,70	4,37
	Τριτοβάθμια εκπαίδευση	5,80*	2,43	,018	1,01	10,59
Δευτεροβάθμια εκπαίδευση	Πρωτοβάθμια εκπαίδευση	-,84	1,79	,641	-4,37	2,70
	Τριτοβάθμια εκπαίδευση	4,96	2,55	,054	-7,62E-02	10,00
Τριτοβάθμια εκπαίδευση	Πρωτοβάθμια εκπαίδευση	-5,80*	2,43	,018	-10,59	-1,01
	Δευτεροβάθμια εκπαίδευση	-4,96	2,55	,054	-10,00	7,62E-02

Based on observed means.

\*. The mean difference is significant at the ,05 level.

Για την προϋπόθεση της κανονικής κατανομής, ο συντομότερος τρόπος είναι να εξετάσουμε κατά πόσο κατανέμονται κανονικά τα κατάλοιπα (residuals)<sup>3</sup> που υπολογίστηκαν με την εφαρμογή της two-way-ANOVA. Αν τα κατάλοιπα κατανέμονται κανονικά, τότε όλες οι παρατηρήσεις προέρχονται από κανονική κατανομή.

<sup>3</sup> Αυτό θα μπορούσε να γίνει και στην περίπτωση της one-way ANOVA.

### One-Sample Kolmogorov-Smirnov Test

		Residual for WTA
N		174
Normal Parameters <sup>a,b</sup>	Mean	-3,29E-08
	Std. Deviation	10,6311
Most Extreme Differences	Absolute	,075
	Positive	,075
	Negative	-,046
Kolmogorov-Smirnov Z		,986
Asymp. Sig. (2-tailed)		,285

a. Test distribution is Normal.

b. Calculated from data.

### Γραφική Απεικόνιση

**Graphs → Error Bar → Clustered → Define → Variable:** βάζουμε την ποσοτική μεταβλητή, **Category Axis:** βάζουμε τη μια ποιοτική μεταβλητή και **Define clusters by** βάζουμε την άλλη ποιοτική μεταβλητή → **Ok**

