

SPSS

ΠΙΝΑΚΑΣ Α

Std-dev = τυπική απόκλιση

F = ανάλυση διακύμανσης

Dep.Var → ποσοτικές

Factor → ποιοτική/κατηγορική

p-value = Sig = στατιστική σημαντικότητα

H0 = Μηδενική Υπόθεση: Κανονική Κατανομή

H1 = Εναλλακτική Υπόθεση: ΟΧΙ Κανονική Κατανομή

→ Σύντομος για ερώτημα 1^ο

• Όταν μου ζητάει ευαρκτικό ποσοστό δείξετάων με εθμενεθίκα χαρτακθρηθίκα:

Analyze → Descriptive stat. → Frequencies
→ Ειθίκα τω μεταθρηθίκα τω επιθίκα → OK.

Δίτω δίωθίκα είκα εύκα διαθρηθίκα η απάνθίκα!

→ Σύντομος για ερώτημα 2^ο

• Ζητάει τα καθίκα περιθρηθίκα μέθρα για 2 ή 3 μεταθρηθίκα ΚΑΙ → Να απθρηθίκα τω επίκα εθμενεθίκα αθίκα και τω καταθίκα τω εθίκα και αν είκα σταθίκα εθμενεθίκα η διαθρηθίκα τω παραθίκα μεταθρηθίκα με mid αθίκα. (αυθίκα τωίκα)

• θίκα Anova: Analyze → Compare Means → One-Way Anova
→ Dependent: α, β, γ
→ factor: τω αθίκα μεταθρηθίκα.
→ Options → Descriptive → Homogeneity.

• Απο πίνακα Anova

Δίτω απθρηθίκα καθίκα σε καθίκα μεταθρηθίκα (περίθίκα) τω p-value θίκα τω Sig. αν είκα < 0,05

τότε αποθρηθίκα τω μηδενίκα υπόθίκα απθίκα

έκα εθμενεθίκα σταθίκα διαθρηθίκα.

σε απθίκα περίθίκα, εθμενεθίκα τω απθίκα.

!(να θίκα τω περιθρηθίκα και τω καταθίκα)!

descriptives

← nx Μίθος αυθίκα > Μίθος θίκα

1

ANOVA

$p < 0.05$ = Υπάρχει στατιστικά σημαντική διαφορά
[Απορρίπτεται η H_0 και ισχύει η H_1]
(Χρειάζεται Post-Hoc για 3 ή περισσότερες ομάδες)
 $p > 0.05$ = ΔΕΝ υπάρχει στατιστικά σημαντική διαφορά
Δεν χρειάζεται Post-Hoc

Homogeneity (Levene's Test)

$p < 0.05$ = Δεν είναι ίσες οι διασπορές
[επιλέγω στα post-HOC το Games-Howel]
 $p > 0.05$ = Είναι ίσες οι διασπορές
[επιλέγω στα post-HOC το LSD]

• Μέθοδος 1-Way ANOVA •

n ANOVA γίνεται για να συγκρίνω 3 και παραπάνω κατηγορίες (μεταβλητές)

Analyze → Compare Means → One-way ANOVA

→ Options → Descriptive, Homogeneity, Means Plot, exclude analysis
→ Post-HOC → LSD, Tamhane's

έστω H_0 : ο M.O. των 3 (n x pt) είναι ίσος.

αν $\text{sig} < 0,05$ απορρίπτεται H_0

είπα εννοιακά να δώσω όνομα τουλάχιστον κάποιου γέιμς διαφέρει

• Τα Post-hoc tests ελέγχω ανα 2 αν έχω ομοιότητα.

• Για να διαλέξω LSD ή Tamhane κοιτώ το Levene's test. •

Αν $\text{sig.} > 0,05$ → LSD

Αν $\text{sig.} < 0,05$ → Tamhane. = Games Howel

Επομένως, συγκρίνω πάντα τα sig. μεταξύ τους για να δω αν διαφέρουν οι M.O. μεταξύ τους.

• Όταν στην ANOVA έχω πολύ χαμηλό $\text{sig.} < 0,05$ σημαίνει ότι έχω στατιστικά διαφορά

Υπενθύμιση, στο pspp το Tamhane = Games Howel

• Small Analysis •

Συγκρίνω 2 κατηγορίες μεταξύ τους: n x. Αν οι αυτές/φωσικές καννίγαν/όχι
Analyze → Descriptive Statistics → Crosstabs → options → Chi-Square.

Κοιτώ το sig. (2-tailed), λέει, αν απορρίψω το H_0 (ότι καννίγαν το ίδιο)
τότε θα έχω 45% πιθανότητα να έχω λάθος απλ δεν απορρίπτω H_0 !

Chisq

αν p ή 2-tailedSig < 0.05 → Υπάρχει **συσχέτιση** μεταβλητών (δεν είναι ανεξάρτητες) και απορρίπτω H_0
αν p ή 2-tailedSig > 0.05 → ΔΕΝ υπάρχει **συσχέτιση** μεταβλητών (είναι ανεξάρτητες) και ισχύει H_0

Το p ή Sig επίσης εκφράζει την πιθανότητα να έχω λάθος αν απορρίψω την H_0

(όρα όταν αρκούν μεταβλητές όπως pH ή ναίμερα
(πάνω από 2 με κάποιον συγκεκριμένο τύπο
μεταβλ. χρησιμοποιώ ANOVA)

→ Σύντομα για ερώτημα 3^ο

- Ζητάει αν υπάρχει στατιστικά σημαντική διαφορά, αξιολογώντας τόσο το επίπεδο σημαντικότητας, αλλά και την κατεύθυνση της σχέσης:

i) Μεταξύ 2 μεταβλητών που μετράονται (ναίμερα μετρήσιμα, όχι κατηγορία)

Analyze → Correlate → Bivariate
→ είδη των 2 μεταβλητές
→ διαλέγω Pearson.

Bivariate Correlation

Ο **Pearson correlation coefficient (r)** μετράει κατά πόσο δύο αριθμητικές μεταβλητές σχετίζονται γραμμικά μεταξύ τους.

Τιμή r Σχέση

+1 Τέλεια θετική γραμμική συσχέτιση

0 Καμία γραμμική συσχέτιση

-1 Τέλεια αρνητική γραμμική συσχέτιση

- $r=0,10$ έως $0,29$ ή $r=-0,1$ έως $-0,29$ → μικρή συσχέτιση
- $r=0,30$ έως $0,49$ ή $r=-0,30$ έως $-0,49$ → μέτρια συσχέτιση
- $r=0,50$ έως $1,00$ ή $r=-0,50$ έως $-1,00$ → μεγάλη συσχέτιση

Τιμή Sig.

< 0.05 Υπάρχει στατιστικά σημαντική συσχέτιση

≥ 0.05 Δεν υπάρχει στατιστικά σημαντική συσχέτιση – μπορεί να είναι τυχαία

< 0.01 Πολύ ισχυρή ένδειξη – υψηλή στατιστική σημαντικότητα

Για 3. ii) ζητάει το ίδιο για σύγκριση
μίας αριθμητικής μεταβλητής και
μίας κατηγορικής αλλά συγκεκριμένα 2 κατηγορίες
από το σύνολο που εκφράζεται.

Ανάλυση → Compare Means → Independent T-test
 → βάζω Variable : των αριθμ. μεταβλητή.
 → βάζω Grouping : των κατηγορική
 → Define Groups → και φτιάχνω.

Από την **Independent T-test**

Περιμένουμε πίνακα με Sig από Lavene's Test και Sig 2-tailed. Για τα αποτελέσματα ισχύει ό,τι γνωρίζουμε από παραπάνω. (Βλέπε πίνακα Homogeneity & πίνακα Chisq)

πχ. **Lavene's Test Sig > 0.05** → H_0 και οι διασπορές είναι ίσες
Sig 2 tailed < 0.05 → Υπάρχει στατιστικά σημαντική συσχέτιση

Με βάση τον Mean (M.O) του πίνακα **Group Statistics**, βγάζουμε και την **κατεύθυνση**

Πίνακας Σύνοψης Τύπων Μεταβλητών:

Τύπος μεταβλητής	Ονομασία (EN/GR)	Έχει σειρά;	Μπορώ να κάνω πράξεις;	Παραδείγματα
Nominal	Ονομαστική	✗ Όχι	✗ Όχι	Φύλο (Α/Γ), Πόλη (Αθήνα, Πάτρα), Τύπος κρασιού
Ordinal	Τακτική / Ιεραρχική	✓ Ναι	✗ Όχι (δεν ξέρω πόση είναι η διαφορά)	Θέση σε αγώνα (1ος, 2ος), Εκπαίδευση (Λύκειο, Πτυχίο)
Scale (Interval)	Διαστημάτων	✓ Ναι	✓ Ναι (χωρίς απόλυτο μηδέν)	Θερμοκρασία °C, Βαθμολογία (0-10)
Scale (Ratio)	Αναλογική	✓ Ναι	✓ Ναι (με απόλυτο μηδέν)	Υψος, Βάρος, Έσοδα, Ποσότητα

ΠΧ Έχω λίστα με κρασιά (1:Red,2:White,3:Rose) και θέλω να φτιάξω μια μεταβλητή «Κρασία που πίνει η Άννα». Η Άννα πίνει μόνο κόκκινο και ροζέ οπότε: Transform → Recode into Different → επιλέγω winetype → name: Annas_wines → Old and New Values και ορίζω το 1 & το 3 να γίνουν 1 → αυτά που πίνει η Άννα και ELSE = 0 → οτιδήποτε άλλο ΔΕΝ το πίνει η Άννα → OK

Μπορώ μάλιστα να πάω στο Value Label και να ορίσω στην τιμή 1: Anna Likes that & 0: Anna Doesn't like that

• Όταν χρειαστεί να συζητήσω μια κατηγορία με μια άλλη και να φτιάξω νέα μεταβλητή κατηγορία.

- Transform → Recode into Different
 - βάζω την κατηγορία που θέλω να αλλάξω
 - name: ορίζω νέο όνομα
 - Old and New values και ορίζω
 - Value: πάνω σε τι βάζω 1 και 2.

• Στην συνέχεια για να γίνει η σύγκριση:

- Analyze → Descriptive → Crosstabs
 - Rows: την μεταβλητή που βασική
 - Columns: την άλλη.
- Statistics → Chi-square

Κοιτάω στο Chi-Square το sig.2
και κεινω ανάλοφα για την εμπειρ. στατ. διαφ.

και στο Crosstabulation βλέπω την κατανομή

Linear → για συνεχείς εξαρτημένες

Binary Logistic → για δυαδική εξαρτημένη μεταβλητή (πχ ΝΑΙ, ΟΧΙ ή με ανθρακικό, χωρίς ανθρακικό)

→ Για Πολυπρόσβλητο.

Analyze → Regression → Linear → (statistics → collinearity)

Dependent: την κύρια μεταβλητή
Independent: όλες τις υπόλοιπες.

• Για χρυσή σχέση: κοιτάω το Coefficients

Συγκεκριμένα: Unstandardized

B και Std. error και η σχέση χτυπάει ως εξής:
$$q_c = \text{Constant} + 0,006 \cdot \text{pt} + \dots +$$

(error) (error) () ...

• Βαθμός Προσαρμογής: R-square στο Model Summary
 $R^2 = \dots$ όσο περισσότερο είναι στο 1 τόσο καλύτερα.

• Επίδραση μεταβλητών στην εξαρτημένη:

κοιτάω Coefficients στο Standardized Beta

η κατεύθυνση κινείται στο πρόσημο (βρίνω ποιο του μεγαλύτερη θετική και του μνη. αρνητική)

• Πολυσυγγραμμικότητα: κοιτάω Coefficients και σφίξι το VIF

όταν VIF → 1 είμαστε πολύ καλά!

$1 < VIF < 5$ αποδεκτή αλλά μετρία πολυσυγγραμμικότητα

$VIF > 5$ πρόβλημα.

Ερμηνεία Αποτελεσμάτων **Παλινδρόμησης**:

Y=εξαρτημένη μεταβλητή

X=ανεξάρτητη μεταβλητή

ΠΛΑΙΣΙΟ ΠΑΛΙΝΔΡΟΜΗΣΗΣ (για ερμηνεία):

1. Γραμμική Σχέση:

- $Y = (\text{σταθερά}) + B_1x_1 + B_2x_2 + \dots + e$

2. Κατεύθυνση επίδρασης:

- Αν $B > 0$: θετική επίδραση
- Αν $B < 0$: αρνητική επίδραση

3. Πιο ισχυρός παράγοντας:

- Κοιτάω **Standardized Beta**: όσο μεγαλύτερο (κατά απόλυτη τιμή), τόσο ισχυρότερη η επίδραση

4. Στατιστική σημαντικότητα:

- Κοιτάω **Sig. < 0.05** → στατιστικά σημαντική μεταβλητή

5. Πολυσυγγραμμικότητα:

- **VIF < 5** και **Tolerance > 0.1** → ✓ OK

6. R^2 → Πόσο καλά ταιριάζει το μοντέλο

- 0.7–0.9 = πολύ καλό, > 0.9 = εξαιρετικό

Γραμμική Σχέση

- Τύπος εξίσωσης:

$$Y = \alpha + \beta_1 \cdot X_1 + \beta_2 \cdot X_2 + \dots + e$$

Σύμβολο	Σημασία
Y	Εξαρτημένη μεταβλητή (π.χ. winetype)
α	Σταθερός όρος → (Constant) από τον πίνακα Coefficients
β_1, β_2, \dots	Συντελεστές (B) → πόσο επηρεάζει κάθε ανεξάρτητη μεταβλητή
X_1, X_2, \dots	Ανεξάρτητες μεταβλητές
e	Σφάλμα (error), δεν φαίνεται στο SPSS αλλά υπάρχει θεωρητικά

Πού κοιτάω στο SPSS / PSPP:

Τι ψάχνω	Πού το βρίσκω
α (constant)	1η γραμμή πίνακα Coefficients
β	Στήλη B (Unstandardized Coefficients)
Στατιστική σημαντικότητα	Στήλη Sig. (π.χ. $p < 0.05$ ✓)
Πρόβλημα collinearity	Τιμές VIF > 5 ή Tolerance < 0.1 ✗

Θυμήσου:

- Χρησιμοποιείς τη **στήλη B** (Unstandardized) για τη **γραμμική σχέση**.
- **Μόνο** οι μεταβλητές με **Sig. < 0.05** είναι **στατιστικά σημαντικές**.