

Στατιστική II

Γιώργος Τσιρογιάννης

Τμήμα Διοίκησης Επιχειρήσεων Αγροτικών
Προϊόντων και Τροφίμων,
Πανεπιστήμιο Πατρών



Διάλεξη 10η

- Ανάλυση διακύμανσης (ANOVA)
- Έλεγχος F

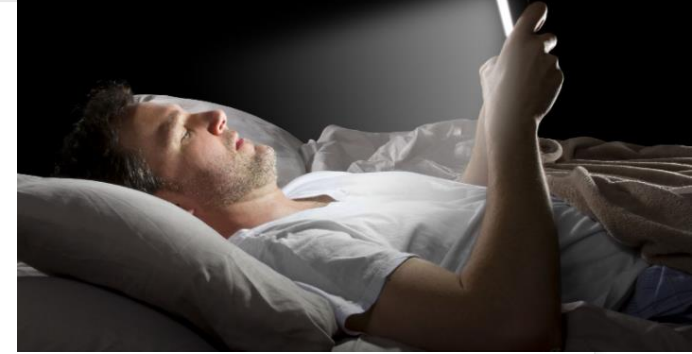


16^ο κεφάλαιο

Γενικά

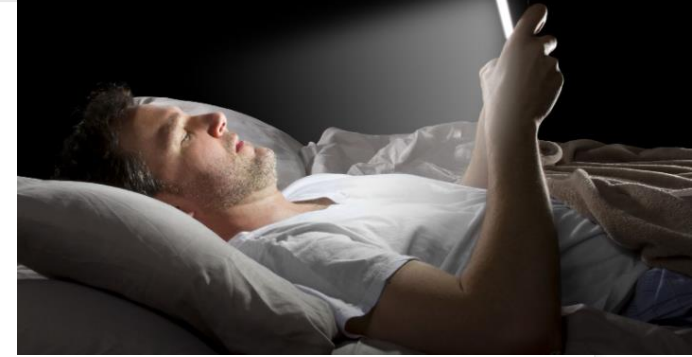
- Η ανάλυση διακύμανσης χρησιμοποιείται για να ανιχνεύει διαφορές μεταξύ δύο ή περισσότερων ομάδων βασισμένων σε μια ανεξάρτητη μεταβλητή
- Η μεταβλητότητα/διασπορά μεταξύ των παρατηρήσεων αναγνωρίζεται ως προερχόμενη από διάφορες πηγές και αν ελεγχθεί κατάλληλα μας δίνει πληροφορίες αναφορικά με τις διαφορές των ομάδων και κατά πόσο οφείλονται στην τυχαιότητα

Παράδειγμα



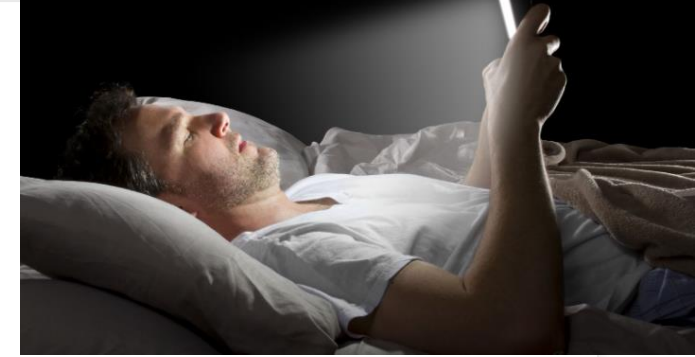
- Προκαλεί η έλλειψη ύπνου περισσότερο ή λιγότερο επιθετική συμπεριφορά;
- Ψυχολόγος εντάσσει τυχαία εθελοντές σε περιόδους έλλειψης ύπνου 0, 24 και 48 ωρών (παράγοντας ή ανεξάρτητη μεταβλητή)
- Μετράται η επιθετικότητά τους (εξαρτημένη μεταβλητή)
 - Ένας αριθμός που αποτυπώνει τον συνολικό αριθμό επιθετικής συμπεριφοράς (πχ προσβολές, διαφωνίες,...)

Παράδειγμα



- Μηδενική υπόθεση: τίποτα σημαντικό δεν συμβαίνει/καταγράφεται
 - Και οι τρεις ομάδες στέρησης ύπνου (0, 24 και 48 ωρών) κατά μέσο όρο έχουν την ίδια συμπεριφορά
 - $H_0: \mu_0 = \mu_{24} = \mu_{48}$
 - Η εναλλακτική υπόθεση δηλώνει ότι η H_0 είναι ψευδής
 - Οποιοσδήποτε συνδυασμός διαφορετικότητας
 - Αρκεί μια ισότητα να μην ισχύει
 - Ενδεχομένως για κάποιο ζεύγος να ισχύει η ισότητα

Παράδειγμα



- Τα δεδομένα είναι λίγα μόνο για την ευκολία στην παρουσίαση και δεν ανταποκρίνονται στην πραγματικότητα

Πείραμα 1

	OUTCOME A HOURS OF SLEEP DEPRIVATION			
	ZERO	TWENTY-FOUR	FORTY-EIGHT	
	3	4	2	
	5	8	4	
	7	6	6	
Group mean:	5	6	4	Grand mean = 5

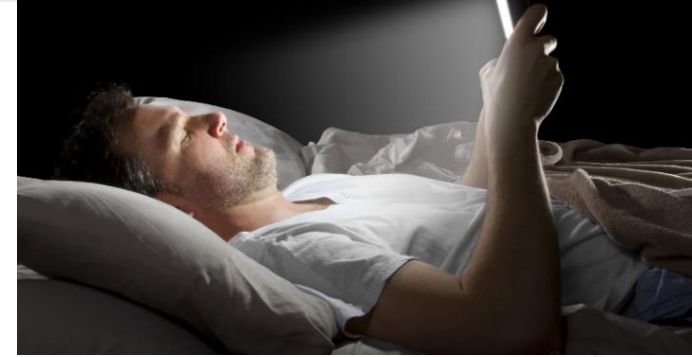
Διαισθητικά: Ποιο πείραμα θα απέρριπτε την μηδενική υπόθεση;

Πείραμα 2

	OUTCOME B HOURS OF SLEEP DEPRIVATION			
	ZERO	TWENTY-FOUR	FORTY-EIGHT	
	0	3	6	
	4	6	8	
	2	6	10	
Group mean:	2	5	8	Grand mean = 5

Ίδιος συνολικός μέσος

Παράδειγμα



OUTCOME A HOURS OF SLEEP DEPRIVATION

Πείραμα 1

	ZERO	TWENTY-FOUR	FORTY-EIGHT	
	3	4	2	
	5	8	4	
	7	6	6	
Group mean:	5	6	4	Grand mean = 5

OUTCOME B HOURS OF SLEEP DEPRIVATION

Πείραμα 2

	ZERO	TWENTY-FOUR	FORTY-EIGHT	
	0	3	6	
	4	6	8	
	2	6	10	
Group mean:	2	5	8	Grand mean = 5

Στην ANOVA ελέγχουμε πολλαπλές πηγές μεταβλητότητας

Ο ρόλος της μεταβλητότητας

OUTCOME A
HOURS OF SLEEP DEPRIVATION

ZERO	TWENTY-FOUR	FORTY-EIGHT
3	4	2
5	8	4
7	6	6

Group mean:

5	6	4
---	---	---

 Grand mean = 5

OUTCOME B
HOURS OF SLEEP DEPRIVATION

ZERO	TWENTY-FOUR	FORTY-EIGHT
0	3	6
4	6	8
2	6	10

Group mean:

2	5	8
---	---	---

 Grand mean = 5

Διαισθητικά

Μικρές διαφοροποιήσεις ακόμα και αν ο ύπνος δεν επηρεάζει την επιθετικότητα

Σημαντικές διαφοροποιήσεις που ενδεχομένως είναι πέρα της τυχαιότητας του δείγματος

Ορισμοί

- **Επίδραση πειραματικού χειρισμού:** Η ύπαρξη μιας τουλάχιστον διαφοράς μεταξύ των μέσων πληθυσμού.
- **Μεταβλητότητα μεταξύ ομάδων:** μεταβλητότητα μεταξύ αποτελεσμάτων αντικειμένων τα οποία όντας σε διαφορετικές ομάδες, υποβάλλονται σε διαφορετικούς πειραματικούς χειρισμούς
- **Μεταβλητότητα εντός ομάδων:** μεταβλητότητα μεταξύ αποτελεσμάτων αντικειμένων τα οποία όντας στην ίδια ομάδα, υποβάλλονται στον ίδιο πειραματικό χειρισμό
- **Τυχαίο σφάλμα:** οι συνδυασμένες επιδράσεις στα αποτελέσματα μεμονωμένων αντικειμένων όλων των ανεξέλεγκτων παραγόντων.
 - Ατομικές διαφοροποιήσεις αντικειμένων, μικρές διαφοροποιήσεις στην διενέργεια του πειράματος, λάθη στις μετρήσεις...

Παράδειγμα

- Τα αντικείμενα για τις 3 παρακάτω ομάδες υπόκεινται σε 4 διαφορετικές δοκιμασίες. Για κάθε ένα από τα a, b, c και d ελέγξτε αν υπάρχει μεταβλητότητα μεταξύ/εντός ομάδων

(a)

<i>GROUP 1</i>	<i>GROUP 2</i>	<i>GROUP 3</i>
8	8	8
8	8	8
8	8	8
8	8	8

Μεταξύ ομάδων: όχι
Εντός ομάδων: όχι

(b)

<i>GROUP 1</i>	<i>GROUP 2</i>	<i>GROUP 3</i>
8	4	12
8	4	12
8	4	12
8	4	12

Μεταξύ ομάδων: ναι
Εντός ομάδων: όχι

(c)

<i>GROUP 1</i>	<i>GROUP 2</i>	<i>GROUP 3</i>
4	6	5
6	6	7
8	10	9
14	10	11

Μεταξύ ομάδων: όχι
Εντός ομάδων: ναι

(d)

<i>GROUP 1</i>	<i>GROUP 2</i>	<i>GROUP 3</i>
6	11	20
8	12	18
8	14	23
10	15	25

Μεταξύ ομάδων: ναι
Εντός ομάδων: ναι

Λόγος F

- Υλοποιεί τον λόγο της μεταβλητότητας μεταξύ των ομάδων (μέσω των δειγματικών μέσων), διαιρεμένο με την ομαδοποιημένη εκτίμηση της διασποράς (δηλ. μεταβλητότητα μεταξύ των ομάδων ή τυχαίο σφάλμα)

$$\text{λόγος F} = \frac{\text{μεταβλητότητα μεταξύ ομάδων}}{\text{μεταβλητότητα εντος ομάδων}}$$

Έλεγχος t

- Βασίζεται στην τυποποιημένη κατανομή δειγματοληψίας t
- Η τυπική απόκλιση του πρώτου και δεύτερου πληθυσμού δεν θεωρείται γνωστή, αλλά εκτιμάται
- Βασίζεται στον λόγο

$$t = \frac{(\text{διαφορα μεταξύ των δειγματικών μέσων}) - (\text{υποθετική διαφορα των μέσων πληθυσμού})}{\text{εκτιμωμενο τυπικό σφαλμα}}$$

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_{\bar{X}_1 - \bar{X}_2}}$$

Έλεγχος F

- Ένας έλεγχος F της μηδενικής υπόθεσης βασίζεται στο:
- Αν η μηδενική υπόθεση είναι αληθής, τότε ο αριθμητής και ο παρονομαστής του λόγου F θα έτειναν να είναι περίπου ίσοι,
- Αλλά αν η μηδενική υπόθεση είναι ψευδής, ο αριθμητής θα έτεινε να υπερτερεί

Λόγος F

- Υλοποιεί τον λόγο της μεταβλητότητας μεταξύ των ομάδων (μέσω των δειγματικών μέσων), διαιρεμένο με την ομαδοποιημένη εκτίμηση της διασποράς (δηλ. μεταβλητότητα μεταξύ των ομάδων ή τυχαίο σφάλμα)

$$\text{λόγος F} = \frac{\text{μεταβλητότητα μεταξύ ομάδων}}{\text{μεταβλητότητα εντος ομάδων}}$$

Έλεγχος t

• Βασίζεται στην υποθέτηση ομογενούς διασποράς

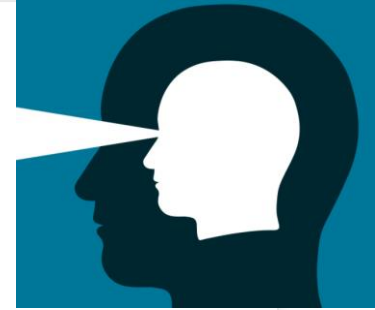
• Η τιμή του στατιστικού εξαρτάται από το αν οι δύο ομάδες έχουν ίση διασπορά

• Βασίζεται στο t

• Ο έλεγχος γίνεται με τη βοήθεια του στατιστικού λογισμικού

• Η τιμή του στατιστικού εξαρτάται από το αν οι δύο ομάδες έχουν ίση διασπορά

• Η τιμή του στατιστικού εξαρτάται από το αν οι δύο ομάδες έχουν ίση διασπορά



Έλεγχος F

Αν η μηδενική υπόθεση είναι αληθής, τότε δεν υπάρχει επίδραση του πειραματικού χειρισμού

$$F = \frac{\text{τυχαιο σφαλα}}{\text{τυχαιο σφαλμα}}$$

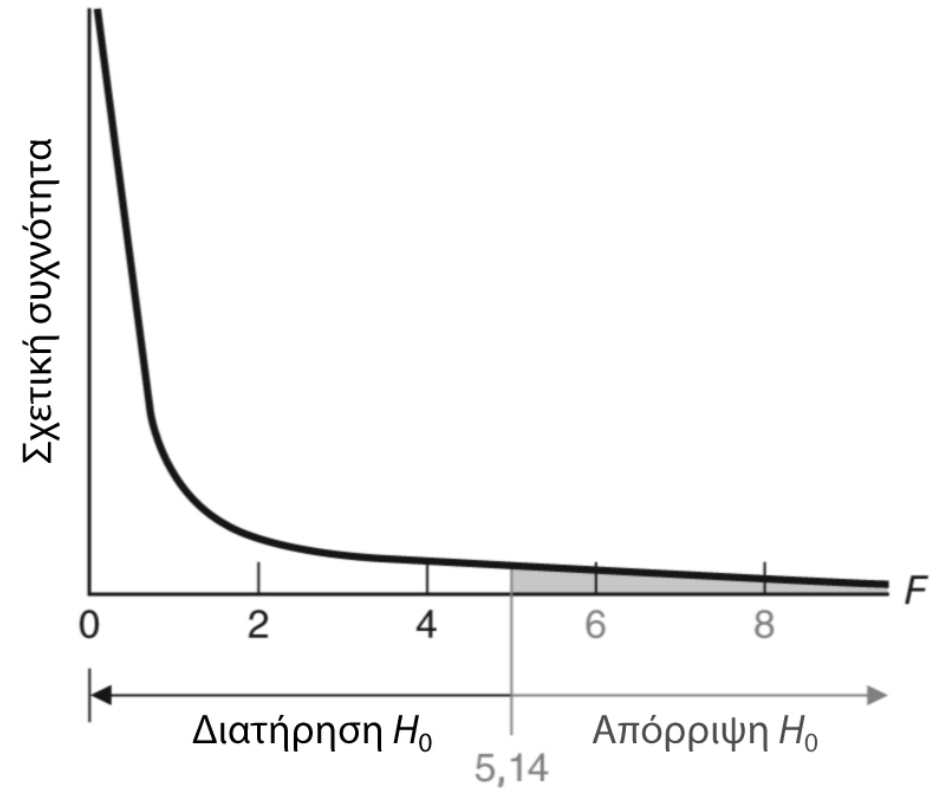
Αν η μηδενική υπόθεση είναι ψευδής, τότε δεν υπάρχει επίδραση του πειραματικού χειρισμού στον αριθμητή του λόγου και ο λόγος μεγαλύτερος από την μονάδα

$$F = \frac{\text{τυχαιο σφαλα} + \text{επιδραση πειραματικου χειρισμου}}{\text{τυχαιο σφαλμα}}$$

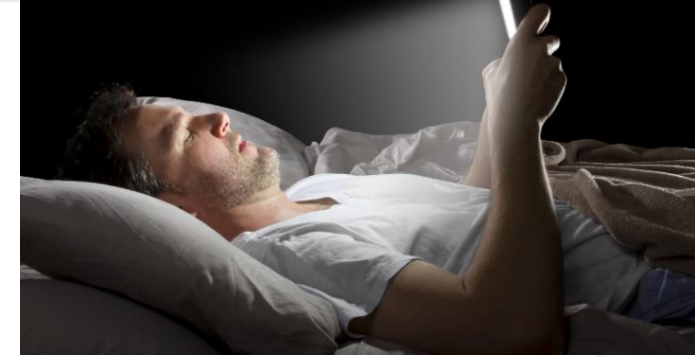
Στην πράξη κινούμαστε αντίστροφα: από τον λόγο να εξάγουμε συμπέρασμα για την H_0

Έλεγχος F

- Θεωρούμε ότι η μηδενική υπόθεση είναι αληθής
 - Ο λόγος F είναι κοντά στην μονάδα
 - Οι διαφορές είναι μικρές και προϊόν της τυχαιότητας δειγματοληψίας
 - Το F θα έπαιρνε ανάλογες κοινές τιμές στην υποθετική συνάρτηση κατανομής δειγματοληψίας
- Αν η H_0 είναι ψευδής τότε ο λόγος F αναμένεται να λαμβάνει σπάνιες τιμές στην υποθετική συνάρτηση κατανομής δειγματοληψίας



Παράδειγμα



Δεδομένα (πολύ λίγα μόνο για ευκολία παρουσίασης):

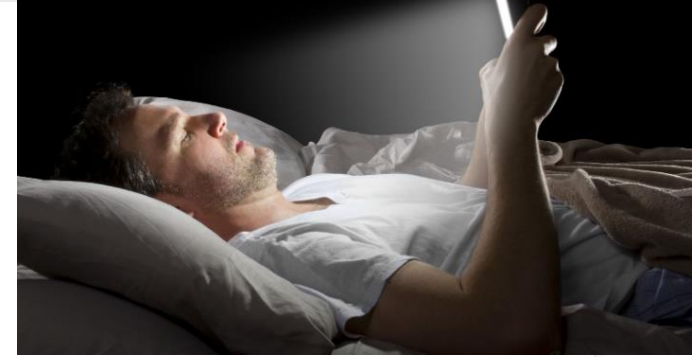
	HOURS OF SLEEP DEPRIVATION			
	ZERO	TWENTY-FOUR	FORTY-EIGHT	
	0	3	6	
	4	6	8	
	2	6	10	
Group mean:	2	5	8	Grand mean = 5

Στατιστικές υποθέσεις

$$H_0: \mu_0 = \mu_{24} = \mu_{48}$$

H_1 : η H_0 είναι ψευδής (κατά τουλάχιστον 1 συνδυασμό)

Παράδειγμα



- Υπολογίζουμε τον την κρίσιμη τιμή του λόγου F (ακολουθούν λεπτομέρειες των υπολογισμών)(δηλ. $F \geq 5.14$) για τους καταλλήλους βαθμούς ελευθερίας $df_{\text{μεταξυ}} = 2, df_{\text{εντος}} = 6$
- Υπολογίζουμε το λόγο F (δηλ. 7.36)
- Συγκρίνουμε με την κρίσιμη τιμή 5.14 και οπότε απορρίπτουμε την H_0
- Η φυσική ερμηνεία της απόφασης είναι ότι: οι ώρες έλλειψης ύπνου επηρεάζουν τους μέσους βαθμούς επιθετικότητας.

Λεπτομέρειες υπολογισμών

- Συμβολισμοί:
 - X =μη επεξεργασμένα δεδομένα
 - T =σύνολο ομάδας
 - n =μέγεθος δείγματος ομάδας
 - G =γενικό σύνολο
 - N : γενικό/συνδυασμένο μέγεθος δείγματος

Λεπτομέρειες υπολογισμών

- Άθροισμα τετραγώνων των αποκλίσεων γύρω από τον γενικό μέσο (grand)

$$SS_{total} = \sum (X - \bar{X}_{grand})^2 = \sum X^2 - \frac{G^2}{N}$$

- Άθροισμα τετραγώνων των αποκλίσεων των μέσων ομάδας (group) γύρω από τον γενικό μέσο (εναλλακτικό συμβολισμός $SS_{\text{μεταξύ/between}}$)

$$SS_{between} = n \sum (\bar{X}_{group} - \bar{X}_{grand})^2 = \sum \frac{T^2}{n} - \frac{G^2}{N}$$

- Άθροισμα τετραγώνων των αποκλίσεων γύρω από τον αντίστοιχο μέσο της ομάδας (εναλλακτικό συμβολισμός $SS_{\text{εντός/within}}$)

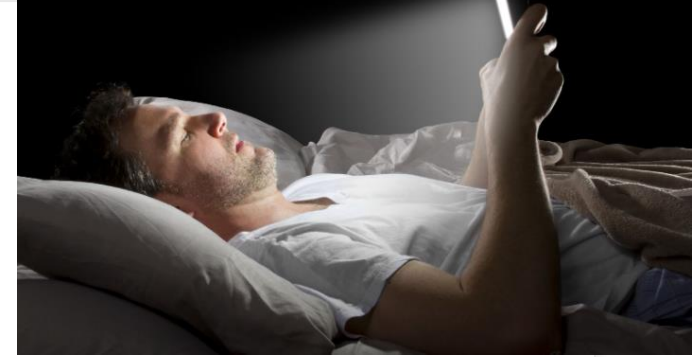
$$SS_{within} = \sum (X - \bar{X}_{group})^2 = \sum X^2 - \sum \frac{T^2}{n}$$

Λεπτομέρειες υπολογισμών

Εκτελούμε ακολουθιακά τα εξής βήματα:

- Εύρεση του συνόλου ομάδας T , του γενικού συνόλου G για όλες τις ομάδες (1)
- Υπολογισμός του $SS_{\text{μεταξύ}}$ (2)
- Υπολογισμός του $SS_{\text{εντός}}$ (3)
- Υπολογισμός του $SS_{\text{ολικό}}$ (4)

Παράδειγμα



Λεπτομέρειες υπολογισμών

- Συμβολισμοί:
 - X =μη επεξεργασμένα δεδομένα
 - T =σύνολο ομάδας
 - n =μέγεθος δείγματος ομάδας
 - G =γενικό σύνολο
 - N : γενικό/συνδυασμένο μέγεθος δείγματος

17

HOURS OF SLEEP DEPRIVATION

0	24	48
0	3	6
4	6	8
2	6	10

1 Group Totals (T) = 6 15 24 Grand Total (G) = 45

Λεπτομέρειες υπολογισμών

- Άθροισμα τετραγώνων των αποκλίσεων γύρω από τον γενικό μέσο (grand)

$$SS_{total} = \sum(X - \bar{X}_{grand})^2 = \sum X^2 - \frac{G^2}{N}$$

- Άθροισμα τετραγώνων των αποκλίσεων των μέσων ομάδας (group) γύρω από τον γενικό μέσο (εναλλακτικό συμβολισμός $SS_{\text{μεταξύ/between}}$)

$$SS_{between} = n\sum(\bar{X}_{group} - \bar{X}_{grand})^2 = \sum \frac{T^2}{n} - \frac{G^2}{N}$$

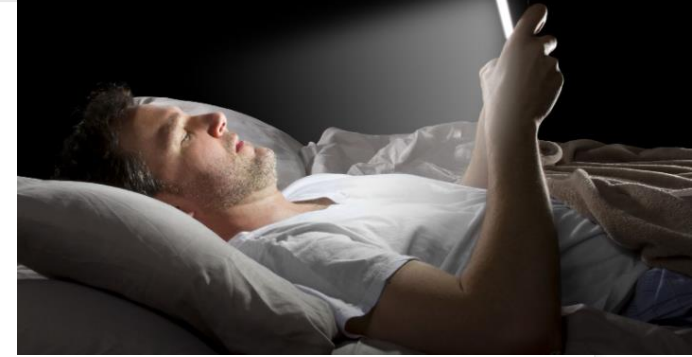
- Άθροισμα τετραγώνων των αποκλίσεων γύρω από τον αντίστοιχο μέσο της ομάδας (εναλλακτικό συμβολισμός $SS_{\text{εντός/within}}$)

$$SS_{within} = \sum(X - \bar{X}_{group})^2 = \sum X^2 - \sum \frac{T^2}{n}$$

18

$$\begin{aligned}
 2 \quad SS_{between} &= \sum \frac{T^2}{n} - \frac{G^2}{N} \\
 &= \left[\frac{(6)^2}{3} + \frac{(15)^2}{3} + \frac{(24)^2}{3} \right] - \frac{(45)^2}{9} = \left[\frac{36}{3} + \frac{225}{3} + \frac{576}{3} \right] - \frac{2025}{9} \\
 &= [12 + 75 + 192] - 225 = 279 - 225 = 54
 \end{aligned}$$

Παράδειγμα



Λεπτομέρειες υπολογισμών

- Συμβολισμοί:
 - X =μη επεξεργασμένα δεδομένα
 - T =σύνολο ομάδας
 - n =μέγεθος δείγματος ομάδας
 - G =γενικό σύνολο
 - N : γενικό/συνδυασμένο μέγεθος δείγματος

17

HOURS OF SLEEP DEPRIVATION

0	24	48
0	3	6
4	6	8
2	6	10

1 Group Totals (T) = 6 15 24 Grand Total (G) = 45

Λεπτομέρειες υπολογισμών

- Άθροισμα τετραγώνων των αποκλίσεων γύρω από τον γενικό μέσο (grand)

$$SS_{total} = \sum(X - \bar{X}_{grand})^2 = \sum X^2 - \frac{G^2}{N}$$

- Άθροισμα τετραγώνων των αποκλίσεων των μέσων ομάδας (group) γύρω από τον γενικό μέσο (εναλλακτικό συμβολισμός $SS_{\text{μεταξύ/between}}$)

$$SS_{between} = n \sum(\bar{X}_{group} - \bar{X}_{grand})^2 = \sum \frac{T^2}{n} - \frac{G^2}{N}$$

- Άθροισμα τετραγώνων των αποκλίσεων γύρω από τον αντίστοιχο μέσο της ομάδας (εναλλακτικό συμβολισμός $SS_{\text{εντός/within}}$)

$$SS_{within} = \sum(X - \bar{X}_{group})^2 = \sum X^2 - \sum \frac{T^2}{n}$$

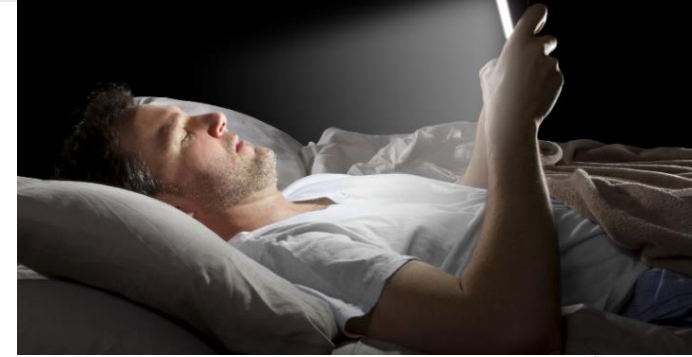
18

$$3 \quad SS_{within} = \sum X^2 - \sum \frac{T^2}{n}$$

$$= (0)^2 + (4)^2 + (2)^2 + (3)^2 + (6)^2 + (6)^2 + (6)^2 + (8)^2 + (10)^2 - \left[\frac{(6)^2}{3} + \frac{(15)^2}{3} + \frac{(24)^2}{3} \right]$$

$$= 0 + 16 + 4 + 9 + 36 + 36 + 36 + 64 + 100 - \left[\frac{36}{3} + \frac{225}{3} + \frac{576}{3} \right] = 301 - 279 = 22$$

Παράδειγμα



Λεπτομέρειες υπολογισμών

- Συμβολισμοί:
 - X =μη επεξεργασμένα δεδομένα
 - T =σύνολο ομάδας
 - n =μέγεθος δείγματος ομάδας
 - G =γενικό σύνολο
 - N : γενικό/συνδυασμένο μέγεθος δείγματος

17

HOURS OF SLEEP DEPRIVATION

0	24	48
0	3	6
4	6	8
2	6	10

1 Group Totals (T) = 6 15 24 Grand Total (G) = 45

Λεπτομέρειες υπολογισμών

- Άθροισμα τετραγώνων των αποκλίσεων γύρω από τον γενικό μέσο (grand)

$$SS_{total} = \sum(X - \bar{X}_{grand})^2 = \sum X^2 - \frac{G^2}{N}$$

- Άθροισμα τετραγώνων των αποκλίσεων των μέσων ομάδας (group) γύρω από τον γενικό μέσο (εναλλακτικό συμβολισμός $SS_{\text{μεταξύ/between}}$)

$$SS_{between} = n\sum(\bar{X}_{group} - \bar{X}_{grand})^2 = \sum \frac{T^2}{n} - \frac{G^2}{N}$$

- Άθροισμα τετραγώνων των αποκλίσεων γύρω από τον αντίστοιχο μέσο της ομάδας (εναλλακτικό συμβολισμός $SS_{\text{εντός/within}}$)

$$SS_{within} = \sum(X - \bar{X}_{group})^2 = \sum X^2 - \sum \frac{T^2}{n}$$

18

$$4 \quad SS_{total} = \sum X^2 - \frac{G^2}{N}$$

$$= (0)^2 + (4)^2 + (2)^2 + (3)^2 + (6)^2 + (6)^2 + (6)^2 + (8)^2 + (10)^2 - \frac{(45)^2}{9}$$

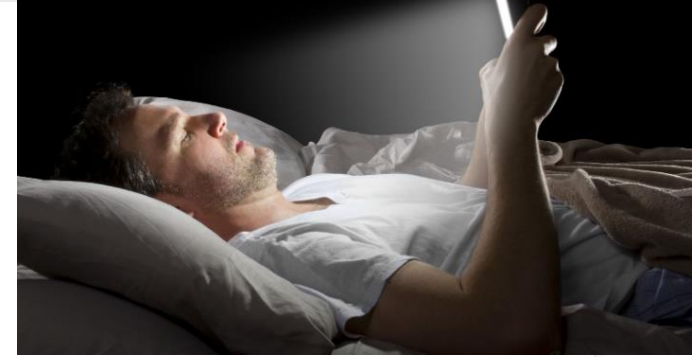
$$= 0 + 16 + 4 + 9 + 36 + 36 + 36 + 64 + 100 - \frac{2025}{9} = 301 - 225 = 76$$

Επαλήθευση

- Πρέπει να ισχύει:

$$SS_{total} = SS_{between} + SS_{within}$$

Παράδειγμα



$$SS_{total} = SS_{between} + SS_{within}$$

$$76 = 54 + 22$$

$$76 = 76$$

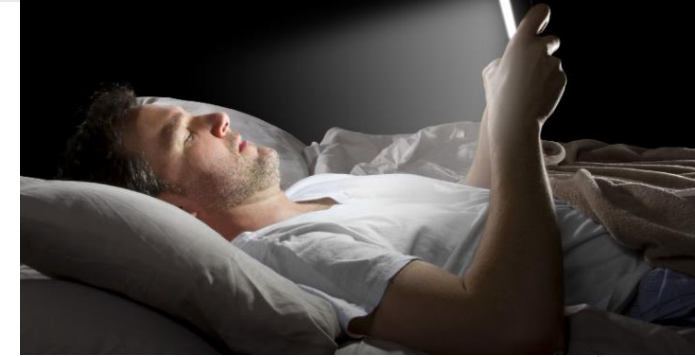


Εύρεση των βαθμών ελευθερίας

- Στις αποκλίσεις γύρω από τον ολικό μέσο συνεπάγεται απώλεια ενός βαθμού ελευθερίας
- Παρομοίως, απώλεια ενός βαθμού για τις αποκλίσεις για τις ομάδες και εντός των ομάδων

$$\begin{aligned}df_{total} &= N - 1, \\df_{between} &= k - 1, \\df_{within} &= N - k,\end{aligned}$$

Παράδειγμα



HOURS OF SLEEP DEPRIVATION

	0	24	48
0	0	3	6
4	4	6	8
2	2	6	10

1 Group Totals (T) = 6 15 24 Grand Total (G) = 45

Εύρεση των βαθμών ελευθερίας

- Στις αποκλίσεις γύρω από τον ολικό μέσο συνεπάγεται απώλεια ενός βαθμού ελευθερίας
- Παρομοίως, απώλεια ενός βαθμού για τις αποκλίσεις για τις ομάδες και εντός των ομάδων

$$\begin{aligned}df_{total} &= N - 1, \\df_{between} &= k - 1, \\df_{within} &= N - k,\end{aligned}$$

25

$$\begin{aligned}df_{total} &= N - 1 = 9 - 1 = 8 \\df_{between} &= k - 1 = 3 - 1 = 2 \\df_{within} &= N - k = 9 - 3 = 6\end{aligned}$$

Επαλήθευση

- Πρέπει να ισχύει:

$$df_{total} = df_{between} + df_{within}$$

Παράδειγμα



Επαλήθευση

- Πρέπει να ισχύει:

$$df_{total} = df_{between} + df_{within}$$

27

Παράδειγμα



HOURS OF SLEEP DEPRIVATION

0	24	48
0	3	6
4	6	8
2	6	10

1 Group Totals (T) = 6 15 24 Grand Total (G) = 45

Εύρεση των βαθμών ελευθερίας

- Στις αποκλίσεις γύρω από τον ολικό μέσο συνεπάγεται απώλεια ενός βαθμού ελευθερίας
- Παρομοίως, απώλεια ενός βαθμού για τις αποκλίσεις για τις ομάδες και εντός των ομάδων

$$\begin{aligned}df_{total} &= N - 1 \\df_{between} &= k - 1 \\df_{within} &= N - k\end{aligned}$$

$$\begin{aligned}df_{total} &= N - 1 = 9 - 1 = 8 \\df_{between} &= k - 1 = 3 - 1 = 2 \\df_{within} &= N - k = 9 - 3 = 6\end{aligned}$$

26

$$\begin{aligned}df_{total} &= N - 1 = 9 - 1 = 8 \\df_{between} &= k - 1 = 3 - 1 = 2 \\df_{within} &= N - k = 9 - 3 = 6\end{aligned}$$



Τετραγωνικός μέσος

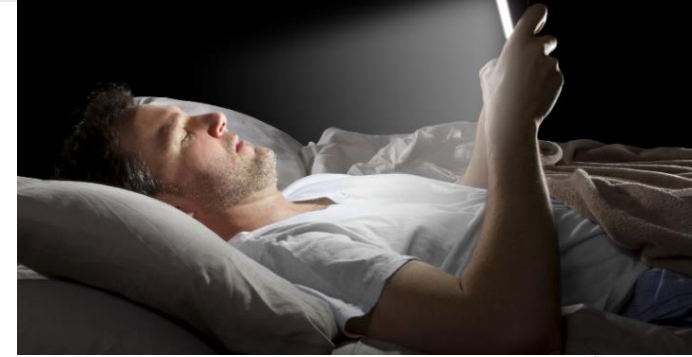
- Είναι μια εκτίμηση της διασποράς με την γενική μορφή:

$$MS = \frac{SS}{df} \quad SS = \sum (X - \bar{X})^2$$

- Τετραγωνικός μέσος μεταξύ των ομάδων και εντός είναι αντίστοιχα:

$$MS_{between} = \frac{SS_{between}}{df_{between}} \quad MS_{within} = \frac{SS_{within}}{df_{within}}$$

Παράδειγμα



Παράδειγμα



Λεπτομέρειες υπολογισμών

- Στοιβάδα
- Κρυφά επισημασμένα δεδομένα
- Τυχαία σειρά
- Συμμετρικά δεδομένα ομάδων
- Ομογενή σφάλμα
- Η γραμμικοποιημένη ομάδα δεδομένων

HOURS OF SLEEP DEPRIVATION

0	24	48
0	3	6
4	6	8
2	6	10

1 Group Totals (T) = 6 15 24 Grand Total (G) = 45

Λεπτομέρειες υπολογισμών

- Μονοπλά επιπέδου των αποτελεσμάτων είναι από τον γενικό μέσο (Grand)
- Μονοπλά επιπέδου των αποτελεσμάτων των ομάδων (Group) για από τον γενικό μέσο (αποδοτικότητα του δεδομένου, $SS_{between}$)
- Μονοπλά επιπέδου των αποτελεσμάτων για από τον αντίστοιχο μέσο της ομάδας (αποδοτικότητα του δεδομένου, SS_{within})

$$2 \text{ } SS_{between} = \sum \frac{T^2}{n} - \frac{G^2}{N}$$

$$= \left[\frac{(6)^2}{3} + \frac{(15)^2}{3} + \frac{(24)^2}{3} \right] - \frac{(45)^2}{9} = \left[\frac{36}{3} + \frac{225}{3} + \frac{576}{3} \right] - \frac{2025}{9}$$

$$= [12 + 75 + 192] - 225 = 279 - 225 = 54$$

20

$$MS_{between} = \frac{SS_{between}}{df_{between}} = \frac{54}{2} = 27$$

Παράδειγμα



HOURS OF SLEEP DEPRIVATION

0	24	48
0	3	6
4	6	8
2	6	10

1 Group Totals (T) = 6 15 24 Grand Total (G) = 45

Εύρεση των βαθμών ελευθερίας

- Στις αποκλίσεις είναι από τον ομαλό μέσο ανεξάρτητα από την ενός βαθμού ελευθερίας
- Παρομοίως, από την ενός βαθμού για τις αποκλίσεις για τις ομάδες και εντός των ομάδων

$$df_{total} = N - 1 = 9 - 1 = 8$$

$$df_{between} = k - 1 = 3 - 1 = 2$$

$$df_{within} = N - k = 9 - 3 = 6$$

26

Παράδειγμα



Παράδειγμα

Λεπτομέρειες υπολογισμών

- Σφάλματα:
- Στις υπολογιστικές βήματα
- Τριγωνικά αριθμούς
- Συναρτήσεις δύναμης, ρίζας
- Σύνταξη κελιών
- % γενική ένδειξη για κελιά με δεκαδικά

HOURS OF SLEEP DEPRIVATION

0	24	48
0	3	6
4	6	8
2	6	10

1 Group Totals (T) = 6 15 24 Grand Total (G) = 45

3 $SS_{within} = \sum X^2 - \sum \frac{T^2}{n}$

$$= (0)^2 + (4)^2 + (2)^2 + (3)^2 + (6)^2 + (6)^2 + (6)^2 + (8)^2 + (10)^2 - \left[\frac{(6)^2}{3} + \frac{(15)^2}{3} + \frac{(24)^2}{3} \right]$$

$$= 0 + 16 + 4 + 9 + 36 + 36 + 36 + 64 + 100 - \left[\frac{36}{3} + \frac{225}{3} + \frac{576}{3} \right] = 301 - 279 = 22$$

21

$$MS_{within} = \frac{SS_{within}}{df_{within}} = \frac{22}{6} = 3.67$$

Παράδειγμα



Εύρεση των βαθμών ελευθερίας

- Στις αποκλίσεις γύρω από τον ομαλό μέσο συνιστάται απώλεια ενός βαθμού ελευθερίας
- Παρομοίως, απώλεια ενός βαθμού για τις αποκλίσεις για τις ομάδες και εντός των ομάδων

HOURS OF SLEEP DEPRIVATION

0	24	48
0	3	6
4	6	8
2	6	10

1 Group Totals (T) = 6 15 24 Grand Total (G) = 45

$$df_{total} = N - 1 = 9 - 1 = 8$$

$$df_{between} = k - 1 = 3 - 1 = 2$$

$$df_{within} = N - k = 9 - 3 = 6$$

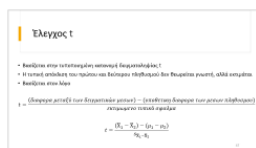
26

Λόγος F

Λόγος F

- Υλοποιεί τον λόγο της μεταβλητότητας μεταξύ των ομάδων (μέσω των δειγματικών μέσων), διαιρεμένο με την ομαδοποιημένη εκτίμηση της διασποράς (δηλ. μεταβλητότητα μεταξύ των ομάδων ή τυχαίο σφάλμα)

$$\text{λόγος } F = \frac{\text{μεταβλητότητα μεταξύ ομάδων}}{\text{μεταβλητότητα εντος ομάδων}}$$



Έλεγχος F

Αν η μηδενική υπόθεση είναι αληθής, τότε δεν υπάρχει επίδραση του πειραματικού χειρισμού

$$F = \frac{\text{τυχαίο σφάλμα}}{\text{τυχαίο σφάλμα}}$$

Αν η μηδενική υπόθεση είναι ψευδής, τότε δεν υπάρχει επίδραση του πειραματικού χειρισμού στον αριθμητή του λόγου και ο λόγος μεγαλύτερος από την μονάδα

$$F = \frac{\text{τυχαίο σφάλμα} + \text{επίδραση πειραματικού χειρισμού}}{\text{τυχαίο σφάλμα}}$$

Στην πράξη κινούμαστε αντίστροφα: από τον λόγο να εξάγουμε συμπέρασμα για την H0

13

$$F = \frac{MS_{between}}{MS_{within}}$$

Τετραγωνικός μέσος

- Είναι μια εκτίμηση της διασποράς με την γενική μορφή:

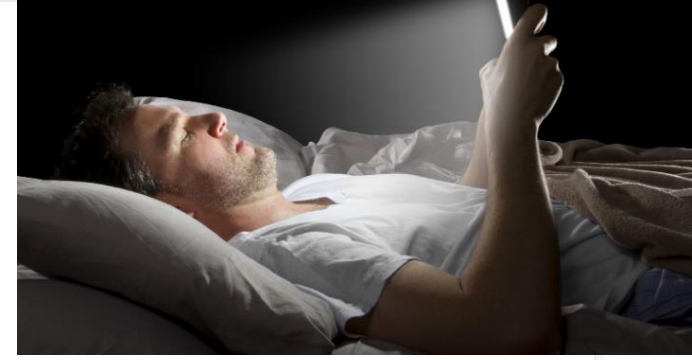
$$MS = \frac{SS}{df} \quad SS = \sum (X - \bar{X})^2$$

- Τετραγωνικός μέσος μεταξύ των ομάδων και εντός είναι αντίστοιχα:

$$MS_{between} = \frac{SS_{between}}{df_{between}} \quad MS_{within} = \frac{SS_{within}}{df_{within}}$$

29

Παράδειγμα

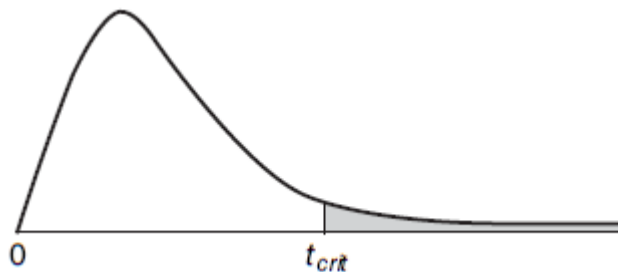


$$F = \frac{MS_{between}}{MS_{within}}$$
$$MS_{between} = \frac{SS_{between}}{df_{between}} = \frac{54}{2} = 27$$
$$MS_{within} = \frac{SS_{within}}{df_{within}} = \frac{22}{6} = 3.67$$

$$F = \frac{MS_{between}}{MS_{within}} = \frac{27}{3.67} = 7.36$$

Αξιολόγηση του λόγου F

- Απαιτείται η σύγκριση του υπολογισμένου λόγου F , με μια κρίσιμη τιμή που ποσοτικοποιεί την σπανιότητα ώστε βάσει της σύγκρισης να απορρίψουμε την μηδενική υπόθεση.
- Προεπιλέγουμε ένα επίπεδο σημαντικότητας (συνήθως 0.05, 0.01)
- Συμβουλευόμαστε τους πίνακες της κατανομής F αναζητώντας την εγγραφή με το κατάλληλο ζεύγος βαθμών ελευθερίας



Χρήση των πινάκων της κατανομής F

Γραμμές με έντονα ψηφία αντιστοιχούν σε σημαντικότητα 0.01
 Γραμμές με κανονικά ψηφία αντιστοιχούν σε σημαντικότητα 0.05

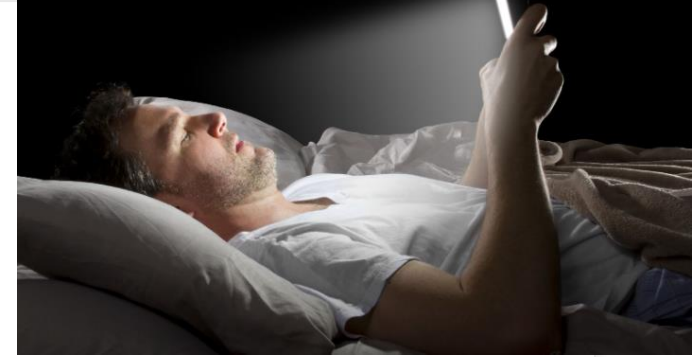
DEGREES OF FREEDOM IN NUMERATOR
 Εντοπίζουμε την στήλη που ισούται $df_{\text{μεταξύ}}$

DEGREES OF FREEDOM IN DENOMINATOR	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
1	161 4,052	200 4,999	216 5,403	225 5,625	230 5,764	234 5,859	237 5,928	239 5,981	241 6,022	242 6,056	243 6,082	244 6,106	245 6,142	246 6,169	248 6,208	249 6,234	250 6,258	251 6,286	252 6,302	253 6,323	253 6,334	254 6,352	254 6,361	254 6,366
2	18.51 98.49	19.00 99.00	19.16 99.17	19.25 99.25	19.30 99.30	19.33 99.33	19.36 99.34	19.37 99.36	19.39 99.38	19.40 99.40	19.41 99.41	19.42 99.42	19.43 99.43	19.44 99.44	19.45 99.45	19.46 99.46	19.47 99.47	19.48 99.48	19.49 99.48	19.49 99.49	19.49 99.49	19.50 99.49	19.50 99.50	19.50 99.50
3	10.13 34.12	9.55 30.82	9.28 29.46	9.12 28.71	9.01 28.24	8.94 27.91	8.88 27.67	8.84 27.49	8.81 27.34	8.78 27.23	8.76 27.13	8.74 27.05	8.71 26.92	8.69 26.83	8.66 26.69	8.64 26.60	8.62 26.50	8.60 26.41	8.58 26.35	8.57 26.27	8.56 26.23	8.54 26.18	8.54 26.14	8.53 26.12
4	7.71 21.20	6.94 18.00	6.59 16.69	6.39 15.98	6.26 15.52	6.16 15.21	6.09 14.98	6.04 14.80	6.00 14.66	5.96 14.54	5.93 14.45	5.91 14.37	5.87 14.24	5.84 14.15	5.80 14.02	5.77 13.93	5.74 13.83	5.71 13.74	5.70 13.69	5.68 13.61	5.66 13.57	5.65 13.52	5.64 13.48	5.63 13.46
5	6.61 16.26	5.79 13.27	5.41 12.06	5.19 11.39	5.05 10.97	4.95 10.67	4.88 10.45	4.82 10.27	4.78 10.15	4.74 10.05	4.70 9.96	4.68 9.89	4.64 9.77	4.60 9.68	4.56 9.55	4.53 9.47	4.50 9.38	4.46 9.29	4.44 9.24	4.42 9.17	4.40 9.13	4.38 9.07	4.37 9.04	4.36 9.02
6	5.99 13.74	5.14 10.92	4.76 9.78	4.53 9.15	4.39 8.75	4.28 8.47	4.21 8.26	4.15 8.10	4.10 7.98	4.06 7.87	4.03 7.79	4.00 7.72	3.96 7.60	3.92 7.52	3.87 7.39	3.84 7.31	3.81 7.23	3.77 7.14	3.75 7.09	3.72 7.02	3.71 6.99	3.69 6.94	3.68 6.90	3.67 6.88
7	5.59 12.25	4.47 9.55	4.35 8.45	4.12 7.85	3.97 7.46	3.87 7.19	3.79 7.00	3.73 6.84	3.68 6.71	3.63 6.62	3.60 6.54	3.57 6.47	3.52 6.35	3.49 6.27	3.44 6.15	3.41 6.07	3.38 5.98	3.34 5.90	3.32 5.85	3.29 5.78	3.28 5.75	3.25 5.70	3.24 5.67	3.23 5.65
8	5.32 11.26	4.46 8.65	4.07 7.59	3.84 7.01	3.69 6.63	3.58 6.37	3.50 6.19	3.44 6.03	3.39 5.91	3.34 5.82	3.31 5.74	3.28 5.67	3.23 5.56	3.20 5.48	3.15 5.36	3.12 5.28	3.08 5.20	3.05 5.11	3.03 5.06	3.0 5.00	2.98 4.96	2.96 4.91	2.94 4.88	2.93 4.86
9	5.12 10.56	4.26 8.02	3.86 6.99	3.63 6.42	3.48 6.06	3.37 5.80	3.29 5.62	3.23 5.47	3.18 5.35	3.13 5.26	3.10 5.18	3.07 5.11	3.02 5.00	2.98 4.92	2.93 4.80	2.90 4.73	2.86 4.64	2.82 4.56	2.80 4.51	2.77 4.45	2.76 4.41	2.73 4.36	2.72 4.33	2.71 4.31
10	4.96 10.04	4.10 7.56	3.71 6.55	3.48 5.99	3.33 5.64	3.22 5.39	3.14 5.21	3.07 5.06	3.02 4.95	2.97 4.85	2.94 4.78	2.91 4.71	2.86 4.60	2.82 4.52	2.77 4.41	2.74 4.33	2.70 4.25	2.67 4.17	2.64 4.12	2.61 4.05	2.59 4.01	2.56 3.96	2.55 3.93	2.54 3.91
11	4.84 9.65	3.98 7.20	3.59 6.22	3.36 5.67	3.20 5.32	3.09 5.07	3.01 4.88	2.95 4.74	2.90 4.63	2.86 4.54	2.82 4.46	2.79 4.40	2.74 4.29	2.70 4.21	2.65 4.10	2.61 4.02	2.57 3.94	2.53 3.86	2.50 3.80	2.47 3.74	2.45 3.70	2.42 3.66	2.41 3.62	2.40 3.60
12	4.75 9.33	3.88 6.93	3.49 5.95	3.26 5.41	3.11 5.06	3.00 4.82	2.92 4.65	2.85 4.50	2.80 4.39	2.76 4.30	2.72 4.22	2.69 4.16	2.64 4.05	2.60 3.98	2.54 3.86	2.50 3.78	2.46 3.70	2.42 3.61	2.40 3.56	2.36 3.49	2.35 3.46	2.32 3.41	2.31 3.38	2.30 3.36
13	4.67 9.07	3.80 6.70	3.41 5.74	3.18 5.20	3.02 4.86	2.92 4.62	2.84 4.44	2.77 4.30	2.72 4.19	2.67 4.10	2.63 4.02	2.60 3.96	2.55 3.85	2.51 3.78	2.46 3.67	2.42 3.59	2.38 3.51	2.34 3.42	2.32 3.37	2.28 3.30	2.26 3.27	2.24 3.21	2.22 3.18	2.21 3.16

Εντοπίζουμε την γραμμή που ισούται $df_{\text{εντός}}$

Γραμμές με έντονα ψηφία αντιστοιχούν σε σημαντικότητα **0.01**

Γραμμές με κανονικά ψηφία αντιστοιχούν σε σημαντικότητα **0.05**



Παράδειγμα

$$df_{total} = N - 1 = 9 - 1 = 8$$

$$df_{between} = k - 1 = 3 - 1 = 2$$

$$df_{within} = N - k = 9 - 3 = 6$$

- Για επίπεδο σημαντικότητας $\alpha = 0.05$

$$F = \frac{MS_{between}}{MS_{within}} = \frac{27}{3.67} = 7.36 > 5.14$$

Απορρίπτουμε την H_0

DEGREES OF FREEDOM IN NUMERATOR

Εντοπίζουμε την στήλη που ισούται $df_{\text{μεταξύ}}$

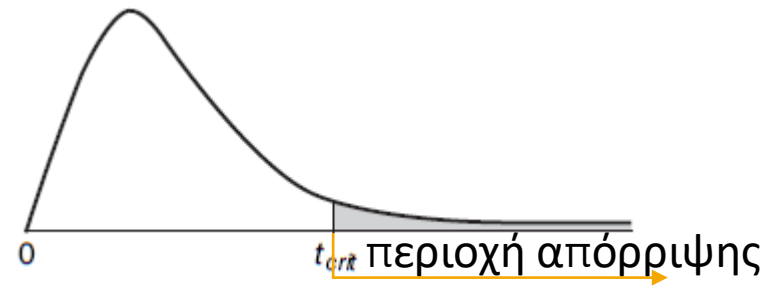
DEGREES OF FREEDOM IN DENOMINATOR	1	2	3	4	5	6	7	8	9	10	11	12	14	16	20	24	30	40	50	75	100	200	500	∞
1	161	200	216	225	230	234	237	239	241	242	243	244	245	246	248	249	250	251	252	253	253	254	254	254
2	4.052	4.999	5.403	5.625	5.764	5.859	5.928	5.981	6.022	6.056	6.082	6.106	6.142	6.169	6.208	6.234	6.258	6.286	6.302	6.323	6.334	6.352	6.361	6.366
3	10.13	9.55	9.28	9.12	9.01	8.94	8.88	8.84	8.81	8.78	8.76	8.74	8.71	8.69	8.66	8.64	8.62	8.60	8.58	8.57	8.56	8.54	8.54	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.93	5.91	5.87	5.84	5.80	5.77	5.74	5.71	5.70	5.68	5.66	5.65	5.64	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.78	4.74	4.70	4.68	4.64	4.60	4.56	4.53	4.50	4.46	4.44	4.42	4.40	4.38	4.37	4.36
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.03	4.00	3.96	3.92	3.87	3.84	3.81	3.77	3.75	3.72	3.71	3.69	3.68	3.67
7	5.59	4.47	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.63	3.60	3.57	3.52	3.49	3.44	3.41	3.38	3.34	3.32	3.29	3.28	3.25	3.24	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.34	3.31	3.28	3.23	3.20	3.15	3.12	3.08	3.05	3.03	3.0	2.98	2.96	2.94	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.13	3.10	3.07	3.02	2.98	2.93	2.90	2.86	2.82	2.80	2.77	2.76	2.73	2.72	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.97	2.94	2.91	2.86	2.82	2.77	2.74	2.70	2.67	2.64	2.61	2.59	2.56	2.55	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.86	2.82	2.79	2.74	2.70	2.65	2.61	2.57	2.53	2.50	2.47	2.45	2.42	2.41	2.40
12	4.75	3.88	3.49	3.26	3.11	3.00	2.92	2.85	2.80	2.76	2.72	2.69	2.64	2.60	2.54	2.50	2.46	2.42	2.40	2.36	2.35	2.32	2.31	2.30
13	4.67	3.80	3.41	3.18	3.02	2.92	2.84	2.77	2.72	2.67	2.63	2.60	2.55	2.51	2.46	2.42	2.38	2.34	2.32	2.28	2.26	2.24	2.22	2.21
	9.07	6.70	5.74	5.20	4.86	4.62	4.44	4.30	4.19	4.10	4.02	3.96	3.85	3.78	3.67	3.59	3.51	3.42	3.37	3.30	3.27	3.21	3.18	3.16

Εντοπίζουμε την γραμμή που ισούται $df_{\text{εντός}}$

Συνοπτικοί πίνακες ANOVA

SOURCE	SS	df	MS	F
Between	54	2	$\left(\frac{54}{2} =\right) 27$	$\left(\frac{27}{3.67} =\right) 7.36^*$
Within	22	6	$\left(\frac{22}{6} =\right) 3.67$	
Total	76	8		

Γιατί η περιοχή απόρριψης βρίσκεται μόνο στο δεξιό μέρος;



$$SS_{between} = n \sum (\bar{X}_{group} - \bar{X}_{grand})^2 = \sum \frac{T^2}{n} - \frac{G^2}{N}$$

$$SS_{within} = \sum (X - \bar{X}_{group})^2 = \sum X^2 - \sum \frac{T^2}{n}$$

$$SS_{total} = \sum (X - \bar{X}_{grand})^2 = \sum X^2 - \frac{G^2}{N}$$

Όλες οι διαφορές είναι υψωμένες στο τετράγωνο και το συσσωρευμένο αποτέλεσμα είναι επίσης θετικό. Ο έλεγχος F είναι μη-κατευθυντικός

Εκτίμηση μεγέθους επίδρασης

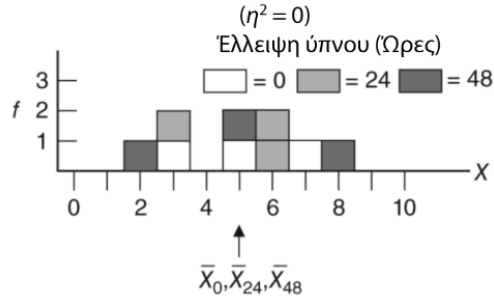
- Εκφράζει την αναλογία της αιτιολογημένης διασποράς (explained variance) και ορίζεται ως ο λόγος

$$\eta^2 = \frac{SS_{between}}{SS_{total}}$$

- Γνωστή ως τετραγωνική καμπυλόγραμμη συσχέτιση

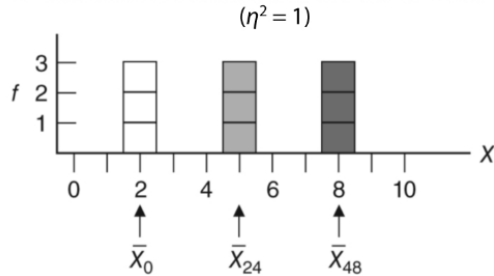
Τετραγωνική καμπυλόγραμμή συσχέτιση

I ΚΑΜΙΑ ΜΕΤΑΒΛΗΤΟΤΗΤΑ ΜΕΤΑΞΥ ΟΜΑΔΩΝ



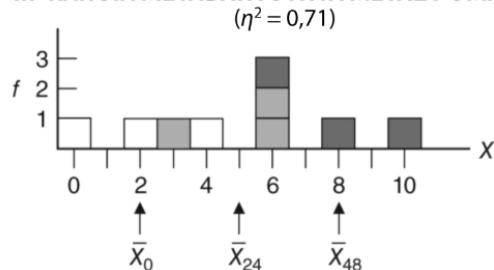
$$\eta^2 = \frac{SS_{between}}{SS_{total}} = \frac{SS_{between}}{SS_{between} + SS_{within}} = \frac{0}{0 + SS_{within}} = \frac{0}{SS_{within}} = 0$$

II ΟΛΗ Η ΜΕΤΑΒΛΗΤΟΤΗΤΑ ΜΕΤΑΞΥ ΟΜΑΔΩΝ



$$\eta^2 = \frac{SS_{between}}{SS_{total}} = \frac{SS_{between}}{SS_{between} + SS_{within}} = \frac{SS_{between}}{SS_{between} + 0} = \frac{SS_{between}}{SS_{between}} = 1$$

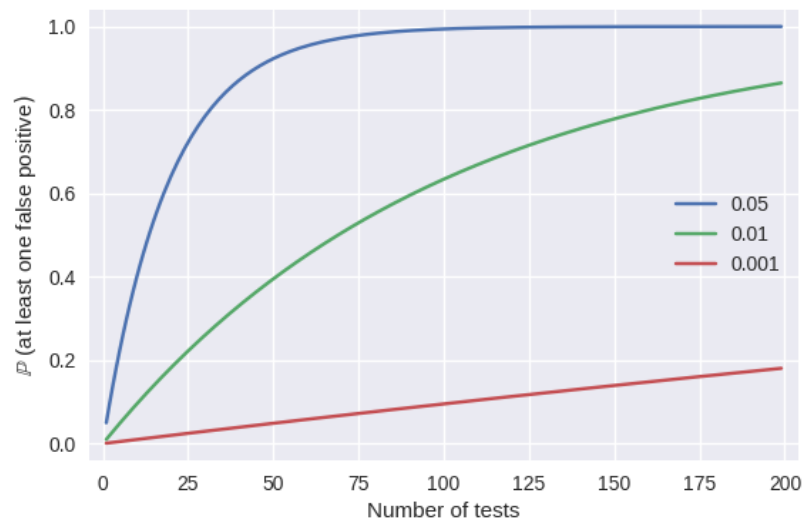
III ΚΑΠΟΙΑ ΜΕΤΑΒΛΗΤΟΤΗΤΑ ΜΕΤΑΞΥ ΟΜΑΔΩΝ



$$\eta^2 = \frac{SS_{between}}{SS_{total}} = \frac{SS_{between}}{SS_{between} + SS_{within}} = \frac{54}{54 + 22} = \frac{54}{76} = .71$$

Γιατί όχι πολλαπλές συγκρίσεις;

- Οι στατιστικοί έλεγχοι έχουν δημιουργηθεί ώστε να εκτελούνται μόνο μια φορά
- Αν αποφασίσουμε να κάνουμε χρήση σε πολλαπλά ζεύγη υπάρχει κίνδυνος αύξησης του σφάλματος τύπου I
- Συνιστάται η χρήση τεχνικής διόρθωσης πχ Bonferroni



Υποθέσεις

- Όλοι οι υποκείμενοι πληθυσμοί κατανέμονται κανονικά
- Μεγέθη ομάδας > 10

Σύνδεσμοι σε online συγγράμματα

- <http://www-personal.umich.edu/~gonzo/coursenotes/file2.pdf>

Lecture Notes #2: Introduction to Analysis of Variance

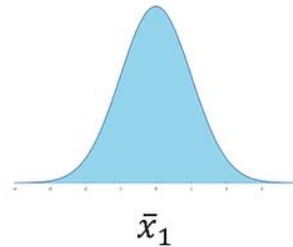
2-1

Richard Gonzalez
Psych 613
Version 2.8 (9/2019)

Σύνδεσμοι σε online video

- <https://www.youtube.com/watch?v=0Vj2V2qRU10>

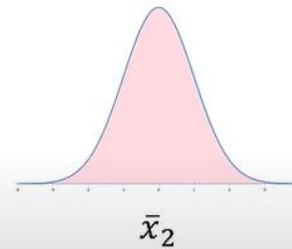
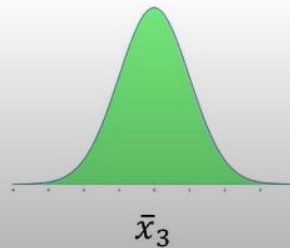
Suppose we want to compare
THREE sample means to see
if a difference exists
somewhere among them.



Is one mean so far away from
the other two that it is likely
not from the same
population?

What we are asking is:

*Do all three of these means come from a
common population?*



Or are all three so far apart that
they ALL likely come from
unique populations?

R code

- <https://www.scribbr.com/statistics/anova-in-r/>

ANOVA in R: A step-by-step guide

Published on March 6, 2020 by [Rebecca Bevans](#). Revised on January 19, 2021.

ANOVA is a [statistical test](#) for estimating how a [quantitative dependent variable](#) changes according to the levels of one or more categorical independent variables. ANOVA tests whether there is a difference in [means](#) of the groups at each level of the independent variable.



Back up



Έλεγχος t

- Βασίζεται στην τυποποιημένη κατανομή δειγματοληψίας t
- Η τυπική απόκλιση του πρώτου και δεύτερου πληθυσμού δεν θεωρείται γνωστή, αλλά εκτιμάται
- Βασίζεται στον λόγο

$$t = \frac{\text{(διαφορα μεταξύ των δειγματικών μεσων)} - \text{(υποθετικη διαφορα των μεσων πληθυσμου)}}{\text{εκτιμωμενο τυπικό σφαγμα}}$$

$$t = \frac{(\bar{X}_1 - \bar{X}_2) - (\mu_1 - \mu_2)}{S_{\bar{X}_1 - \bar{X}_2}}$$