ΑΝΑΛΥΣΗ ΔΙΑΚΥΜΑΝΣΗΣ

Επικ. Καθ. Στέλιος Ζήμερας

Τμήμα Μαθηματικών
Κατεύθυνση Στατιστικής και Αναλογιστικά – Χρηματοοικονομικά Μαθηματικά

2015

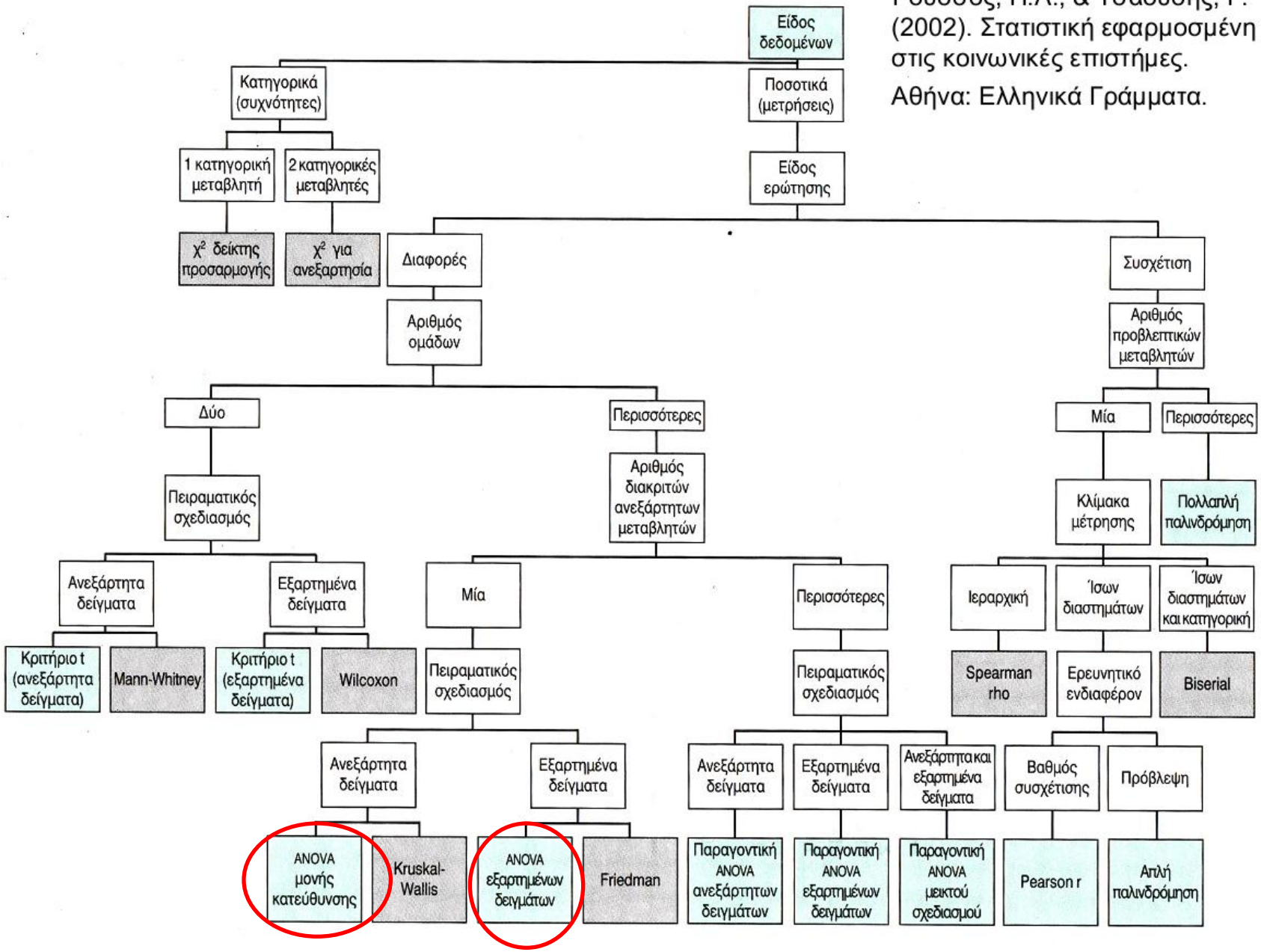
Ανάλυση Διακύμανσης

Η Ανάλυση Διακύμανσης είναι μία τεχνική που μας επιτρέπει να συγκρίνουμε *δύο ή περισσότερους* πληθυσμούς.

Η Ανάλυση Διακύμανσης είναι:

- μία διαδικασία που καθορίζει εάν *υπάρχουν διαφορές* μεταξύ των *μέσων των πληθυσμών*.
- μία διαδικασία η οποία δουλεύει με βάση την ύπαρξη *δειγματοληπτικής διακύμανσης*

Ανάλυση Διακύμανσης



Ρούσσος, Π.Λ., & Τσαούσης, Γ. (2002). Στατιστική εφαρμοσμένη στις κοινωνικές επιστήμες. Αθήνα: Ελληνικά Γράμματα.

Ανάλυση Διακύμανσης

Κάτω από το πλαίσιο αυτό μπορούμε να θεωρήσουμε την ANOVA σαν προέκταση της δοκιμασίας t για την σύγκριση των μέσων τιμών δύο πληθυσμών.

Πλεονεκτήματα

1. συντομότερη διαδικασίας ανάλυσης
2. ακρίβεια της διάγνωσης.

Ανάλυση Διακύμανσης

αναφερθήκαμε σε στατιστικούς ελέγχους υποθέσεων για την τιμή της μέσης τιμής, μ, ή της διασποράς, σ^2, ενός πληθυσμού καθώς και σε στατιστικούς ελέγχους υποθέσεων για τη σύγκριση των μέσων τιμών, μ_1 και μ_2, ή των διασπορών σ_1^2 και σ_2^2, δύο πληθυσμών, αντίστοιχα.

Πως μπορούμε να συγκρίνουμε περισσότερες μέσες τιμές για πληθυσμούς πάνω από δύο, δηλ

H_0: $\mu_1 = \mu_2 = \ldots\ldots = \mu_k$ με $\kappa > 2$?

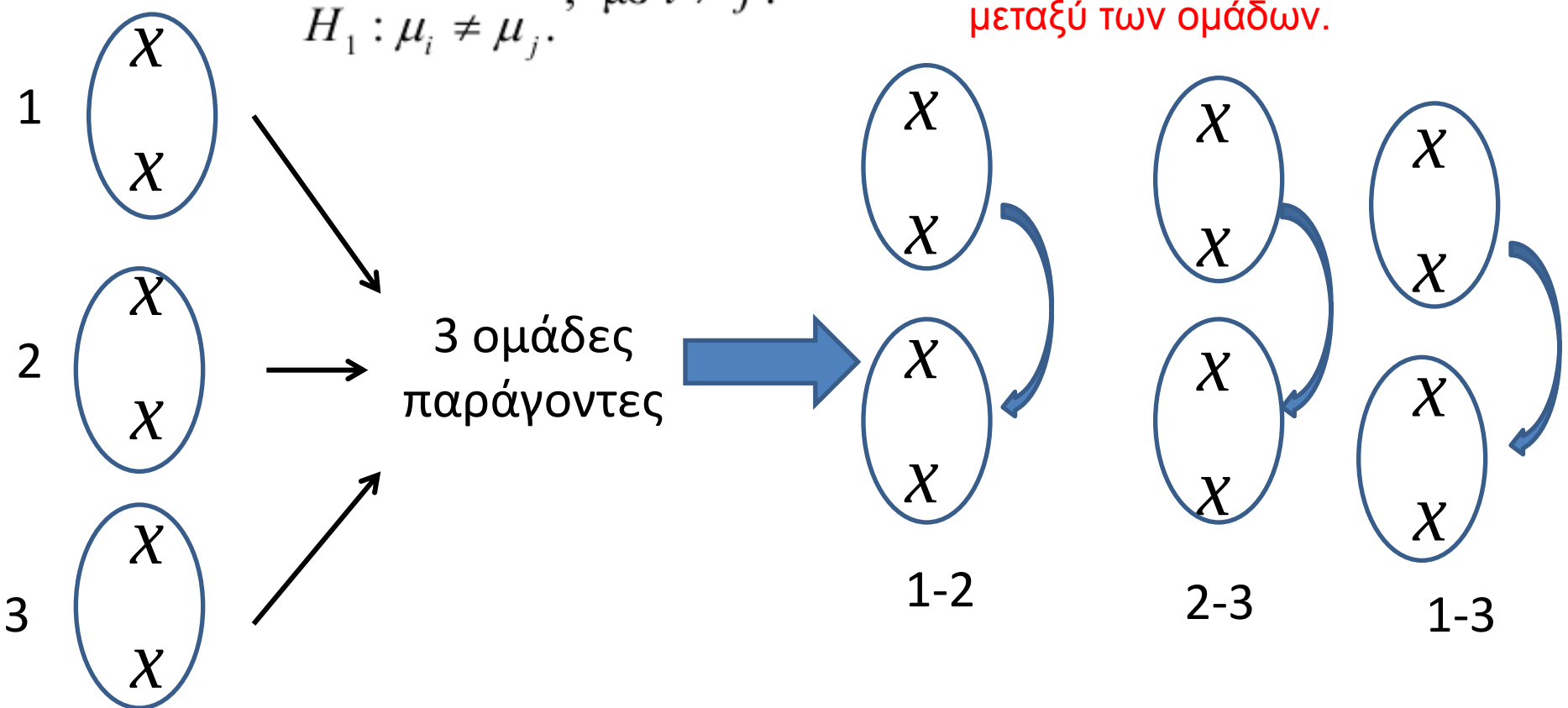
Ανάλυση Διακύμανσης

σύγκριση των k μέσων τιμών ανά δύο.

$$c = \binom{k}{2} = \frac{k \cdot (k-1)}{2}, \quad \text{ελέγχους}$$

$$H_0 : \mu_i = \mu_j$$
$$H_1 : \mu_i \neq \mu_j.$$
, με $i \neq j$.

Είναι η εκτίμηση κατά πόσο η διασπορά οφείλεται σε παράγοντες εντός των ομάδων ή μεταξύ των ομάδων.



Ανάλυση Διακύμανσης

Είναι φανερό, ότι μια τέτοια διαδικασία είναι **χρονοβόρα** ακόμη και όταν το k είναι μικρό

αν υποθέσουμε ότι έχουμε να συγκρίνουμε τις μέσες τιμές για 5 πληθυσμούς, τότε θα πρέπει να κάνουμε

$$\binom{5}{2} = \frac{5!}{2!(5-2)!} = \frac{5\cdot 4\cdot 3\cdot 2\cdot 1}{2\cdot 1\cdot 3\cdot 2\cdot 1} = 10$$

διαφορετικές ζευγαρωτές δοκιμασίες .

Ο έλεγχος υποθέσεων των μέσων

H_0:δεν υπάρχουν διαφορές μεταξύ των 5 μέσων

Ανάλυση Διακύμανσης

θα πρέπει να αποδεχθούμε και τις 10 ζευγαρωτές δοκιμασίες t.

Αν το επίπεδο σημαντικότητας κάθε ζευγαρωτής δοκιμασίας είναι $\alpha = 0.05$, τότε η πιθανότητα να αποδεχθούμε και τις 10 δοκιμασίες είναι $(0.95)^{10} = 0.5987$. Συνεπώς η πιθανότητα να απορρίψουμε τουλάχιστον μία ζευγαρωτή δοκιμασία θα είναι $1 - 0.5987 = 0.4013$ που σημαίνει ότι με τις 10 δοκιμασίες υποπίπτουμε σε σφάλμα τύπου Ι στις 40.13% των περιπτώσεων.

Ανάλυση Διακύμανσης

Αν ενδιαφερόμαστε να κάνουμε c ανεξάρτητους ελέγχους, σε επίπεδο σημαντικότητας α_{pc} (πιθανότητα σφάλματος τύπου I κατά σύγκριση/per comparison error rate), τότε η πιθανότητα λανθασμένης απόρριψης της μηδενικής υπόθεσης σε έναν τουλάχιστον από αυτούς, δηλαδή, η πιθανότητα, στους c ελέγχους, να συμβεί τουλάχιστον μια φορά σφάλμα Τύπου I (πιθανότητα σφάλματος τύπου I κατά πείραμα/per experiment error rate), είναι:

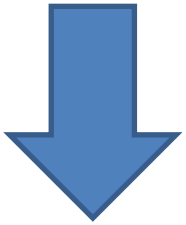
$$\alpha_{PE} = 1 - (1 - \alpha_{PC})^c .$$

Ανάλυση Διακύμανσης

αν κάνουμε c= 6, 10, 15, 45 ανεξάρτητους ελέγχους σε επίπεδο σημαντικότητας α_{PC}=0.05, τότε, η πιθανότητα, α_{PE} δίνεται από:

c	6	10	15	45
α_{PE}	0.26	0.40	0.54	0.90

α_{PE} αυξάνει τον αριθμό των ελέγχων c



πρόβλημα πολλαπλών συγκρίσεων (multiple comparisons problem),

Ανάλυση Διακύμανσης

Ελέγχουμε την ισότητα των μέσων μιας μεταβλητής με διαφορετικούς παράγοντες για k>2 πληθυσμούς.

Άρα σε επίπεδο σημαντικότητας α=0.05

H_0: $\mu_1 = \mu_2 = \ldots\ldots = \mu_k$ με κ>2

H_1: $\mu_i \neq \mu_j$ για τουλάχιστον ένα ζεύγος (i,j)

Έλεγχος μηδενικής υπόθεσης: ο λόγος F που ορίσαμε για τον έλεγχο της ισότητας δύο διασπορών μπορεί να χρησιμεύσει για τον έλεγχο της παραπάνω μηδενικής υπόθεσης.

Ανάλυση Διακύμανσης

Η ανάλυση διακύμανσης ενός παράγοντα χρησιμοποιείται για τον έλεγχο υποθέσεων σε μελέτες ανεξάρτητων δειγμάτων. Ο ερευνητής επιλέγει ένα δείγμα για κάθε διαφορετική τιμή της ανεξάρτητης μεταβλητής, και ελέγχει υποθέσεις που συγκρίνουν τις μέσες τιμές των δειγμάτων αυτών. Η μηδενική υπόθεση δηλώνει ότι δεν υφίσταται καμία διαφορά ανάμεσα σε όλες τις μέσες τιμές των δειγμάτων.

Ανάλυση Διακύμανσης

Είναι εμφανές λοιπόν, ότι στην περίπτωση που ο ερευνητής απορρίψει τελικά τη μηδενική υπόθεση, το μόνο που μπορεί να ισχυριστεί είναι ότι τα δείγματα διαφέρουν μεταξύ τους, αλλά δεν είναι σε θέση να γνωρίζει ποια συγκεκριμένα δείγματα διαφέρουν. Αν θέλει να εξακριβώσει τέτοιου είδους πληροφορίες μπορεί να χρησιμοποιήσει μετά την ανάλυση διακύμανσης ειδικά στατιστικά τεστ, τα οποία ονομάζονται post hoc tests,

Ανάλυση Διακύμανσης με ένα παράγοντα

Οι **προϋποθέσεις** που πρέπει να ισχύουν για να χρησιμοποιηθεί η ανάλυση διακύμανσης ενός παράγοντα είναι:

- τα δείγματα είναι αντιπροσωπευτικά και οι τιμές που τα απαρτίζουν οφείλονται σε ανεξάρτητες παρατηρήσεις,
- η κατανομή των τιμών των δειγμάτων είναι κανονική, και
- οι πληθυσμοί από τους οποίους έχουν επιλεγεί τα δείγματα έχουν την ίδια διακύμανση

Ανάλυση Διακύμανσης με ένα παράγοντα

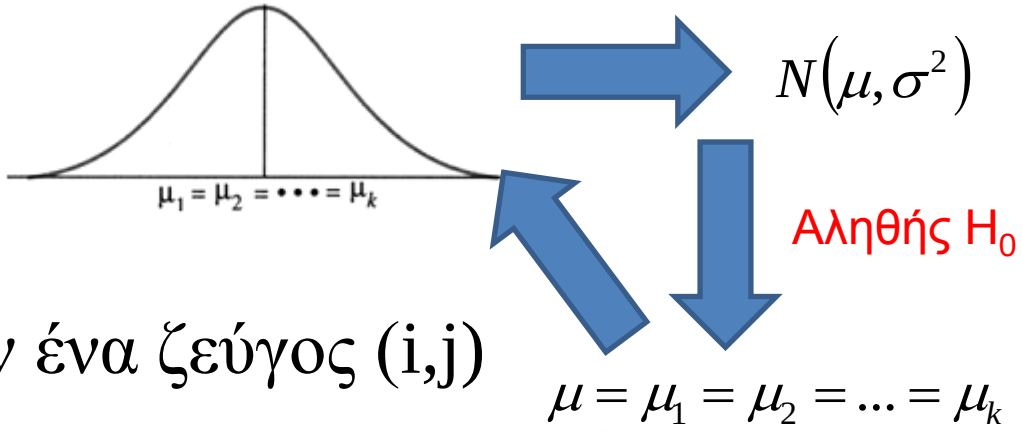
Εντελώς Τυχαιοποιημένο Σχέδιο (Completely Randomized Design)

- το πιο απλό πειραματικό σχέδιο
- εργαζόμαστε με k ανεξάρτητα τυχαία δείγματα, ένα από κάθε πληθυσμό
- αποτελεί ευθεία γενίκευση του σχεδίου για τον έλεγχο των μέσων, δύο κανονικών πληθυσμών με δύο ανεξάρτητα τυχαία δείγματα.

Ανάλυση Διακύμανσης με ένα παράγοντα

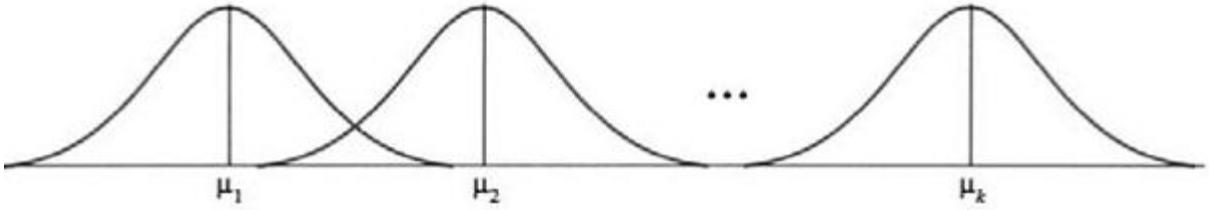
Ας θεωρήσουμε, ότι από καθέναν από k(>2) κανονικούς πληθυσμούς με κοινή διασπορά, σ^2 , και μέσες τιμές, αντίστοιχα, μ_1 , μ_2 ,……, μ_k παίρνουμε ένα τυχαίο δείγμα μεγέθους, αντίστοιχα, , n_1 , n_2 ,……, n_k για να κάνουμε, με βάση τα k δείγματα, τον έλεγχο, της μηδενικής υπόθεσης,

H_0: $\mu_1 = \mu_2 = \ldots\ldots = \mu_k$



H_1: $\mu_i \neq \mu_j$ για τουλάχιστον ένα ζεύγος (i,j)

κ=2⇒έλεγχο δύο μέσων με δύο ανεξάρτητα τυχαία δείγματα



Ανάλυση Διακύμανσης με ένα παράγοντα

Εντελώς Τυχαιοποιημένο Σχέδιο (Completely Randomized Design)

Πειραματικό σχέδιο

$$1^0 \text{ δείγμα} \Rightarrow Y_{11}, Y_{12}, \ldots, Y_{1n_1} \Rightarrow \mathrm{N}\left(\mu_1, \sigma^2\right)$$

$$2^0 \text{ δείγμα} \Rightarrow Y_{21}, Y_{22}, \ldots, Y_{2n_2} \Rightarrow \mathrm{N}\left(\mu_2, \sigma^2\right)$$

$$\vdots$$

$$\vdots$$

$$k^{\tau\acute{\alpha}\xi\varepsilon\omega\varsigma} \text{ δείγμα} \Rightarrow Y_{k1}, Y_{k2}, \ldots, Y_{kn_k} \Rightarrow \mathrm{N}\left(\mu_k, \sigma^2\right)$$

Ανάλυση Διακύμανσης με ένα παράγοντα

Υποθέτουμε ότι

$$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$$

Εάν κατατάξουμε τους πληθυσμούς ανάλογα με την "επίδραση" (κάθε"επίδραση" αντιστοιχει σε ένα πληθυσμό) θα μπορούμε να λέμε ότι οι k "επιδράσεις" μπορεί να αναφέρονται σε k διαφορετικά χαρακτηριστικά

Εστω ότι παίρνουμε k τυχαία δείγματα (ανεξάρτητα) μεγέθους $n_1,\ldots, n_\kappa$ αντίστοιχα, ένα από κάθε επίδραση.

Ανάλυση Διακύμανσης με ένα παράγοντα

Επίδραση Μεταχείρισης (Treatment)

	1	2	…	i	…	k	
	Y_{11}	Y_{21}	…	Y_{i1}	…	Y_{k1}	
	Y_{12}	Y_{22}	…	Y_{i2}	…	Y_{k2}	
	.	.		.			
	.	.		.			
	.	.		.			
	Y_{1n_1}	Y_{2n_2}	…	Y_{in_i}	…	Y_{kn_k}	
Σύνολα	$Y_{1.}$	$Y_{2.}$	…	$Y_{i.}$	…	$Y_{k.}$	$Y_{..}$
Μέσοι	$\overline{Y}_{1.}$	$\overline{Y}_{2.}$	…	$\overline{Y}_{i.}$	…	$\overline{Y}_{k.}$	$\overline{Y}_{..}$

Ανάλυση Διακύμανσης με ένα παράγοντα

Υλοποίηση

Εκτιμάμε την κοινή διασπορά, σ^2 , των k κανονικών πληθυσμών (είναι άγνωστη συνήθως), με δύο τρόπους:

$1^{ος}$ τρόπος

Με το σταθμισμένο μέσο s_w^2 των k – δειγματικών διασπορών

$$s_1^2, s_2^2, \ldots, s_k^2$$

δηλ: με το σταθμισμένο μέσο των διασπορών εντός των δειγμάτων (ίδια με τον έλεγχο δύο μέσων με δύο ανεξάρτητα τυχαία δείγματα):

$$s_w^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2 + \ldots + (n_k - 1)s_k^2}{n - k}$$

$$n = n_1 + n_2 + \ldots + n_k$$

Ανάλυση Διακύμανσης με ένα παράγοντα

2ος τρόπος

Με την δειγματική διασπορά s_b^2

$$s_b^2 = \frac{\sum_{j=1}^{k} n_j \left(y_j - \bar{y}\right)^2}{k-1}$$

Μέσος όρος $n = n_1 + n_2 + \ldots + n_k$

Μέσος όρος του j διανύσματος

σχέση σύνδεσης της διασπορά του πληθυσμού, σ^2 , με τη διασπορά του δειγματικού μέσου, $\sigma_{\bar{y}}^2$

$$\sigma^2 = n\sigma_{\bar{Y}}^2$$

Διασπορά μεταξύ των δειγμάτων

Ανάλυση Διακύμανσης με ένα παράγοντα

Υποθέτουμε ότι

$$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$$

Εάν κατατάξουμε τους πληθυσμούς ανάλογα με την "επίδραση" (κάθε"επίδραση" αντιστοιχει σε ένα πληθυσμό) θα μπορούμε να λέμε ότι οι k "επιδράσεις" μπορεί να αναφέρονται σε k διαφορετικά χαρακτηριστικά

Εστω ότι παίρνουμε k τυχαία δείγματα (ανεξάρτητα) μεγέθους $n_1,\ldots, n_\kappa$ αντίστοιχα, ένα από κάθε επίδραση.

Ανάλυση Διακύμανσης με ένα παράγοντα

Η διαδικασία Ανάλυση Διακύμανσης περιλαμβάνει τα ακόλουθα στάδια:

- Βρίσκουμε τη διασπορά των μέσων τιμών των δειγμάτων για να εκτιμήσουμε τη διασπορά του πληθυσμού

διακύμανση κατά παράγοντες (between-groups variance)

- Εκτιμούμε τη διασπορά του πληθυσμού με βάση τις διασπορές στο εσωτερικό κάθε δείγματος **διακύμανση σφάλματος (within groups variance, error variance)**
- Συγκρίνουμε την πρώτη διασπορά με τη δεύτερη: αν η πρώτη είναι μικρή σε σχέση με τη δεύτερη, επαληθεύεται η μηδενική υπόθεση αλλιώς διαψεύδεται.

Ανάλυση Διακύμανσης με ένα παράγοντα

Αν η μηδενική υπόθεση είναι **αληθής**, τότε ο λόγος

$$\frac{s_b^2}{s_w^2} \to 1$$

Αν η μηδενική υπόθεση **δεν είναι αληθής (ψευδείς)**, τότε ο λόγος

$$\frac{s_b^2}{s_w^2} \to \text{αύξηση}$$

$$\frac{s_b^2}{s_w^2} = \frac{\text{μεταβλητότητα μεταξύ δειγμάτων}}{\text{μεταβλητότητα εντός δειγμάτων}} \quad \Rightarrow \quad F^* = \frac{s_b^2}{s_w^2}$$

Ανάλυση Διακύμανσης με ένα παράγοντα

Επομένως με βάση την σχέση

$$F^* = \frac{s_b^2}{s_w^2} = \frac{\text{μεταβλητότητα μεταξύ δειγμάτων}}{\text{μεταβλητότητα εντός δειγμάτων}}$$

μπορούμε να την εκφράσουμε ως:

$$F^* = \frac{s_b^2}{s_w^2} = \frac{\text{μέσο άθροισμα των τετραγώνων των επεμβάσεων } (\mathrm{MST_r} - \mathrm{MAT_E})}{\text{Μέσο άθροισμα τετραγώνων σφάλματος } (\mathrm{MSE} - \mathrm{MMA_\Sigma})}$$

Ανάλυση Διακύμανσης με ένα παράγοντα

Το Μέσο άθροισμα των τετραγώνων των επεμβάσεων MST_r εξαρτάται από το άθροισμα των τετραγώνων των επεμβάσεων SST_r όπου δίνεται από την σχέση

$$SST_r = \sum_j n_j \left(\bar{y}_j - \bar{y}\right)^2$$

με κ-1 βαθμούς ελευθερίας

$\bar{y} = \bar{y}_{..} \Rightarrow$ συνολικός μέσος

$\bar{y}_j \Rightarrow$ επιμέρους μέσοι για j - στήλες

Η μεταβλητότητα μεταξύ των δειγματοληπτικών μέσων μετράτε ως το άθροισμα των τετραγώνων των αποστάσεων κάθε μέσου με τον συνολικό μέσο.

Ανάλυση Διακύμανσης με ένα παράγοντα

$$SST_r = \sum_j n_j \left(\bar{y}_j - \bar{y}\right)^2$$

$\bar{y} = \bar{y}_{..} \Rightarrow$ συνολικός μέσος

$\bar{y}_j \Rightarrow$ επιμέρους μέσοι για j-στήλες

- Όταν οι δειγματικοί μέσοι είναι κοντά, οι αποστάσεις τους από τον συνολικό μέσο είναι μικρές, καταλήγοντας με ένα μικρό SST. Έτσι, μεγάλο SST υποδεικνύει μεγάλη διασπορά μεταξύ των δειγματικών μέσων, που υποστηρίζει H_1.

Ανάλυση Διακύμανσης με ένα παράγοντα

Το Μέσο άθροισμα των τετραγώνων των σφαλμάτων MSE εξαρτάται από το άθροισμα των τετραγώνων των σφαλμάτων SSE όπου δίνεται από την σχέση

$$SSE = \sum_{ij} \left(y_{ij} - \bar{y}_j \right)^2$$

με n-k βαθμούς ελευθερίας

Η μεταβλητότητα εντός του δείγματος μετράτε ως το άθροισμα των τετραγώνων των αποστάσεων κάθε παρατήρησης από τους επιμέρους μέσους .

Ανάλυση Διακύμανσης με ένα παράγοντα

Επομένως τα σφάλματα εντός και εκτός δειγμάτων δίνονται από τις σχέσεις:

$$MST_r = \frac{SST_r}{k-1} = \frac{\sum_j n_j \left(\bar{y}_j - \bar{y}\right)^2}{k-1}$$

$$MSE = \frac{SSE}{n-k} = \frac{\sum_{ij} \left(y_{ij} - \bar{y}_j\right)^2}{n-k}$$

μεταξύ

εντός

Επίδραση Μεταχείρισης (Treatment)

	1	2	…	i	…	k	
	Y_{11}	Y_{21}	…	Y_{i1}	…	Y_{k1}	
	Y_{12}	Y_{22}	…	Y_{i2}	…	Y_{k2}	
	.	.		.			
	.	.		.			
	.	.		.			
	Y_{1n_1}	Y_{2n_2}	…	Y_{in_i}	…	Y_{kn_k}	
Σύνολα	$Y_{1.}$	$Y_{2.}$	…	$Y_{i.}$	…	$Y_{k.}$	$Y_{..}$
Μέσοι	$\overline{Y}_{1.}$	$\overline{Y}_{2.}$	…	$\overline{Y}_{i.}$	…	$\overline{Y}_{k.}$	$\overline{Y}_{..}$

Εκτιμήτριες κοινής διασποράς

Ανάλυση Διακύμανσης με ένα παράγοντα

Αποδεικνύεται ότι:

$$\sum_j n_j \left(\bar{y}_j - \bar{y}\right)^2 + \sum_{ij} \left(y_{ij} - \bar{y}_j\right)^2 = \sum_{ij} \left(y_{ij} - \bar{y}\right)^2$$

$$SST_r + SSE = SST_{tot} \Rightarrow SST_{tot} = \sum_{ij} \left(y_{ij} - \bar{y}\right)^2$$

Ολικό άθροισμα τετραγώνων
n-1 βαθμοί ελευθερίας

εκφράζει την ολική μεταβλητότητα των παρατηρήσεων, ή αλλιώς, τη μεταβλητότητα όλων των παρατηρήσεων γύρω από το γενικό δειγματικό μέσο.

Ανάλυση Διακύμανσης με ένα παράγοντα

Αποδεικνύεται ότι όταν η υπόθεση H_0: $\mu_1 = \mu_2 = \ldots\ldots = \mu_k$ είναι αληθής, η τ.μ. F^* ακολουθεί την F-κατανομή με (κ-1) και (n-k) βαθμούς ελευθερίας δηλ:

$$F^* = \frac{MST_r}{MSE} \sim F_{(k-1),(n-k)}$$

Άρα η υπόθεση απορρίπτεται σε επίπεδο σημαντικότητας α=5% όταν ισχύει για την κρίσιμη περιοχή ελέγχου

$$F^* = \frac{MST_r}{MSE} > F_{(k-1),(n-k),a}$$

Ανάλυση Διακύμανσης με ένα παράγοντα

Διευκολύνσεις υπολογισμών:

$$SST_{tot} = \sum_{ij} y_{ij}^2 - \frac{\sum_{ij} y_{ij}^2}{n}$$

$$SST_r = \frac{\sum_{j=1}^{k} y_j^2}{n_j} - \frac{\sum_{ij} y_{ij}^2}{n}$$

$$SSE = \sum_{ij} y_{ij}^2 - \sum_{i} \frac{y_{i.}^2}{n_i}$$

Ανάλυση Διακύμανσης με ένα παράγοντα

Πίνακας Ανάλυσης Διασποράς

Πίνακας Ανάλυσης Διασποράς *(για το εντελώς τυχαιοποιημένο σχέδιο)*					
Πηγή μεταβλητότητας	***Β.Ε.***	***Άθροισμα τετραγώνων SS***	***Μέσο άθροισμα τετραγώνων MS***	***Κριτήριο F***	***Περιοχή απόρριψης***
Επεμβάσεις (Treatments) ή Παράγοντας (factor) ή μεταξύ των δειγμάτων	$k-1$	$SSTr$	$MSTr = \frac{SSTr}{k-1}$	$F_{Tr} = \frac{MSTr}{MSE}$	$F_{Tr} > F_{k-1;\nu-k;\alpha}$
Σφάλμα (Error) ή εντός των δειγμάτων	$\nu-k$	SSE	$MSE = \frac{SSE}{\nu-k}$		
Ολική	$\nu-1$	$SSTot$			

Παραδείγματα

Παράδειγμα: Μια τάξη 20 μαθητών χωρίστηκε, με τυχαίο τρόπο σε 5 τμήματα με τον σκοπό να μελετηθεί η αποτελεσματικότητα 5 διαφορετικών μεθόδων διδασκαλίας της στατιστικής. Μετά από 6 εβδομάδες οι μαθητές έδωσαν ένα διαγώνισμα. Τα αποτελέσματα (οι βαθμοί) δίνονται παρακάτω. Να εξεταστεί αν οι βαθμοί αυτοί δίνουν κάποια ένδειξη στατιστικά σημαντικής διαφοράς των μεθόδων διδασκαλίας.

Πίνακας Ανάλυσης Διακύμανσης για τη Σύγκριση των Πέντε Μεθόδων Διδασκαλίας

	Μέθοδοι				
	1	2	3	4	5
Βαθμοί	93	73	75	89	59
	97	77	84	81	64
	92	67	80	76	55
	85	76	70	75	67
$y_{i.}$	$y_{1.}=367$	$y_{2.}=293$	$y_{3.}=309$	$y_{4.}=321$	$y_{5.}=245$
$\overline{y}_{i.}$	91.75	73.25	77.25	80.25	61.25
$\sum_i \sum_j^{n_i} y_{ij}^2$	33,747	22,523	23,981	25,883	15,091

Απάντηση

Πίνακας Ανάλυσης Διακύμανσης για τη Σύγκριση των Πέντε Μεθόδων Διδασκαλίας

	Μέθοδοι				
	1	2	3	4	5
Βαθμοί	93	73	75	89	59
	97	77	84	81	64
	92	67	80	76	55
	85	76	70	75	67
$y_{i.}$	$y_{1.}$=367	$y_{2.}$=293	$y_{3.}$=309	$y_{4.}$=321	$y_{5.}$=245
$\overline{y}_{i.}$	91.75	73.25	77.25	80.25	61.25
$\sum_i \sum_j^{n_i} y_{ij}^2$	33,747	22,523	23,981	25,883	15,091

$$SST_r = \frac{\sum_{j=1}^{k} y_{i.}^2}{n_i} - \frac{\sum_{ij} y_{ij}^2}{n} = \frac{479085}{4} - \frac{2356225}{20} = 1960$$

$$SSE = \sum_{ij} y_{ij}^2 - \sum_i \frac{y_{i.}^2}{n_i} = 121225 - \frac{479085}{4} = 453.75$$

Απάντηση

Πίνακας ANOVA
(Ανάλυση Διακύμανσης)

Αιτία Διασποράς	SS	Βαθμοί Ελευθερίας	MS	F
Μεταξύ Επιδράσεων	1960	4	4.90	16.20
Μέσα στις Επιδράσεις	453.75	15	30.25	
Σύνολο	2413.75	19		

Στο $\alpha = .10$

$$F_{4,\ 15,\ 0.90} = 2.36$$

Επειδή $F > F_{4,\ 15,\ 0.90}$ απορρίπτουμε την H_0.

Παραδείγματα

Πρόβλημα-1: Ένας φοιτητής, στο πλαίσιο της πτυχιακής του εργασίας, προκειμένου να συγκρίνει τέσσερα είδη (A1, A2, A3 και A4, αντίστοιχα) πρόσθετης ύλης ζωοτροφών για αύξηση του βάρους νεογέννητων χοίρων, σχεδίασε και εκτέλεσε το εξής πείραμα. Επέλεξε 20 νεογέννητους χοίρους και με μια τυχαία διαδικασία αντιστοίχισε σε 5 από αυτούς την πρόσθετη ύλη A1, σε 5 άλλους την A2, σε 5 άλλους την A3 και σε 5 άλλους την A4 (δες το σχήμα που ακολουθεί). Δημιούργησε έτσι, 4 ομάδες των 5 χοίρων η κάθε μια. Αφού χορήγησε στους χοίρους κάθε ομάδας τροφή με την αντίστοιχη πρόσθετη ύλη για τρεις μήνες, κατέγραψε την αύξηση του βάρους κάθε χοίρου (σε pounds). Σε επίπεδο σημαντικότητας 5%, υποστηρίζουν αυτά τα πειραματικά δεδομένα ότι υπάρχουν στατιστικά σημαντικές διαφορές στη μέση αύξηση του βάρους των νεογέννητων χοίρων που να οφείλονται στα τέσσερα είδη πρόσθετης ύλης ζωοτροφών;

A1	A2	A3	A4
81	78	72	85
66	66	70	70
78	69	78	83
76	64	77	74
61	66	69	70

Απάντηση

το πείραμα σχεδιάσθηκε με βάση το εντελώς τυχαιοποιημένο σχέδιο.

Έχουμε στη διάθεσή μας k= 4 , ανεξάρτητα τυχαία δείγματα με παραδοχή ότι καθένα από αυτά προέρχεται από κανονικό πληθυσμό με κοινή διασπορά, σ^2 , και για τους τέσσερις πληθυσμούς,

Θα ελέγξουμε αν τα τέσσερα ανεξάρτητα τυχαία δείγματα υποστηρίζουν ή όχι, ότι η αύξηση του βάρους νεογέννητων χοίρων κατά τους τρεις πρώτους μήνες από τη γέννησή τους, επηρεάζεται από το είδος πρόσθετης ύλης (αν διαφοροποιείται από είδος σε είδος πρόσθετης ύλης).

Απάντηση

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4 ,$$

έναντι της *εναλλακτικής*,

$$H_1 : \mu_i \neq \mu_j , \text{ για ένα τουλάχιστον ζευγάρι } (i, j),\ i, j = 1,2,3,4 .$$

Τα τέσσερα δείγματα έχουν ίδιο μέγεθος, k=5, με συνολικό μέγεθος n=20

A1	A2	A3	A4
81	78	72	85
66	66	70	70
78	69	78	83
76	64	77	74
61	66	69	70
$T_1 = 362$	$T_2 = 343$	$T_3 = 366$	$T_4 = 382$

Απάντηση

A1	A2	A3	A4
81	78	72	85
66	66	70	70
78	69	78	83
76	64	77	74
61	66	69	70
$T_1 = 362$	$T_2 = 343$	$T_3 = 366$	$T_4 = 382$

$$G = \sum_{ji} y_{ij} = \sum_{j=1}^{4} T_j = 1453$$

$$SSTot = \sum_{ji} y_{ij}^2 - \frac{G^2}{\nu} = (81^2 + 66^2 + \ldots + 70^2) - \frac{1453^2}{20} = 838.55, \quad \text{με} \quad \nu - 1 = 19$$

βαθμούς ελευθερίας.

$$SSTr = \sum_{j=1}^{4} \frac{T_j^2}{n_j} - \frac{G^2}{\nu} = \frac{362^2 + 343^2 + 366^2 + 382^2}{5} - \frac{1453^2}{20} = 154.15, \quad \text{με} \quad k - 1 = 3$$

βαθμούς ελευθερίας.

$SSE = SSTot - SSTr = 838.55 - 154.15 = 684.40$, με $\nu - k = 16$ *βαθμούς ελευθερίας.*

Απάντηση

Πίνακας Ανάλυσης Διασποράς				
Πηγή μεταβλητότητας	***Β.Ε.***	***Άθροισμα τετραγώνων SS***	***Μέσο άθροισμα τετραγώνων MS***	***Κριτήριο F***
Παράγοντας (ή επεμβάσεις) *(είδος πρόσθετης ύλης)*	3	154.15	$MSTr = \frac{154.15}{3} = 51.38$	$F_{Tr} = \frac{51.38}{42.77} = 1.20$
Σφάλμα	16	684.40	$MSE = \frac{684.40}{16} = 42.77$	
Ολική	19	838.55		

Επειδή η τιμή $F_{Tr} = 1.20$ δε βρίσκεται στην *περιοχή απόρριψης* της *μηδενικής υπόθεσης* αφού, $F_{3;16;0.05} = 3.24$, συμπεραίνουμε ότι, σε *επίπεδο σημαντικότητας 5%*, τα πειραματικά δεδομένα **δεν υποστηρίζουν** ότι το *είδος πρόσθετης ύλης*, επηρεάζει την *αύξηση του βάρους νεογέννητων χοίρων κατά τους τρεις πρώτους μήνες από τη γέννησή τους*.