Ασκήσεις

Πίνακας 6.3: Βαθμολογίες μαθητών στα μαθηματικά				
	Σχολείο			
αα	**Α**	**Β**	**Γ**	**Δ**
1	19	11	13	13
2	15	13	13	13
3	08	20	15	18
4	11	13	09	07
5	13	15	20	14
6	15	11	17	14
7	14	09	16	18
8	18	09	10	11
9	20	18	20	10
10	13	14	12	10
ΜΟ	$\bar{x}_A = \boxed{14{,}6}$	$\bar{x}_B = \boxed{13{,}3}$	$\bar{x}_\Gamma = \boxed{14{,}5}$	$\bar{x}_\Delta = \boxed{12{,}8}$
ΤυπΑπ	$s_A = 3{,}7$	$s_B = 3{,}6$	$s_\Gamma = 3{,}8$	$s_\Delta = 3{,}5$
ΜΟ βαθμολογίας όλων των μαθητών : $\bar{x} = \boxed{13{,}8}$ *(* $s = 3{,}6$ *)*				

Ασκήσεις

Υποθέτουμε πως

- Οι βαθμολογίες είναι κανονικά κατανεμημένες.
- Οι τυπικές αποκλίσεις των δειγμάτων είναι παρόμοιες.
- Οι βαθμολογίες για κάθε δείγμα είναι ανεξάρτητες μεταξύ τους.

H_0 : η μέση επίδοση των σχολείων στα μαθηματικά είναι ίσες (H_0 : $\mu_1=\mu_2$),

H_1 : η μέση επίδοση των σχολείων είναι σημαντικά διαφορετικές (H_1 : $\mu_1 \neq \mu_2$).

Πίνακας 6.4: Απόκλιση ομάδων από τη μέση τιμή				
	Σχολείο			
	Α	**Β**	**Γ**	**Δ**
Απόκλιση	14,6 – 13,8 = 0,8	13,3 – 13,8 = -0,5	14,5 – 13,8 = 0,7	12,8 – 13,8 = -1
Τετράγωνο	0,64	0,25	0,49	1

Ασκήσεις

Θεωρώντας τη μέση βαθμολογία κάθε ενός δείγματος ως αντιπροσωπευτική όλων των μαθητών του δείγματος, υπολογίζουμε το άθροισμα των τετραγωνικών αποκλίσεων από τη μέση τιμή όλων των μαθητών (between-group" sum of squares):

$$SS_B = 10\left((\bar{x}_A - \bar{x})^2 + (\bar{x}_B - \bar{x})^2 + (\bar{x}_\Gamma - \bar{x})^2 + (\bar{x}_\Delta - \bar{x})^2\right) = 10\left(0{,}8^2 + (-0{,}5)^2 + 0{,}7^2 + (-1)^2\right) = 23{,}8$$

Μέση τετραγωνική απόκλιση μεταξύ των ομάδων $MSS_B = \dfrac{23{,}8}{4-1} = 7{,}93$

Πίνακας 6.5: Απόκλιση βαθμολογίας από το κέντρο του δείγματος				
	Σχολείο			
αα	**Α**	**Β**	**Γ**	**Δ**
1	4,4	-2,3	-1,5	0,2
2	0,4	-0,3	-1,5	0,2
3	-6,6	6,7	0,5	5,2
4	-3,6	-0,3	-5,5	-5,8
5	-1,6	1,7	5,5	1,2
6	0,4	-2,3	2,5	1,2
7	-0,6	-4,3	1,5	5,2
8	3,4	-4,3	-4,5	-1,8
9	5,4	4,7	5,5	-2,8
10	-1,6	0,7	-2,5	-2,8

Ασκήσεις

Υπολογίζουμε το άθροισμα των τετραγωνικών αποκλίσεων των παρατηρήσεων από τη μέση τιμή του δείγματος όπου ανήκουν :

$$SS_W = 4{,}4^2 + 0{,}4^2 + (-6{,}6)^2 + \ldots + (-2{,}8)^2 = 480{,}6$$

Μέση τετραγωνική απόκλιση στο εσωτερικό των ομάδων

$$MSS_W = \frac{480{,}6}{40-4} = 13{,}35$$

$$F = \frac{MSS_B}{MSS_W} \quad .$$

$$F = \frac{MSS_B}{MSS_W} = \frac{7{,}93}{13{,}35} = 0{,}59$$

Λόγος κοντά στο 0 δεν υπάρχει επίδραση του σχολείου στην βαθμολογία τω

Ασκήσεις

Ας υποθέσουμε ότι ο διευθυντής παραγωγής μιας βιομηχανίας η οποία παρασκευάζει και συσκευάζει σε κουτιά μίγμα δημητριακών σκέπτεται να αντικαταστήσει την παλιά μηχανή με μία σύγχρονη. Στην αγορά κυκλοφορούν τρεις τύποι μηχανών διαφορετικών προμηθευτών, το κόστος αγοράς και συντήρησης των οποίων είναι περίπου ίδιο. Κατά συνέπεια η επιλογή θα βασιστεί στην απόδοση των τριών μηχανών σε πραγματικές συνθήκες λειτουργίας. Για το σκοπό αυτό ο διευθυντής παραγωγής ζήτησε από τους τρεις προμηθευτές να του παραχωρήσουν από ένα μηχάνημα για μια περίοδο δοκιμής και σχεδίασε το ακόλουθο πείραμα για να προσδιορίσει αν υπάρχει σημαντική διαφορά στην απόδοση των τριών μηχανημάτων. Δεκαπέντε εργάτες της ίδιας πείρας, ηλικίας και των ίδιων γνώσεων επιλέγονται και τοποθετούνται τυχαία ανά πέντε σε κάθε ένα μηχάνημα ως χειριστές του. Μετά από ένα διάστημα εκπαίδευσης μετράται ο χρόνος (σε δευτερόλεπτα) που χρειάζεται ο κάθε εργάτης για να ολοκληρώσει τη συσκευασία ενός πακέτου χρησιμοποιώντας το αντίστοιχο μηχάνημα. Τα αποτελέσματα συγκεντρώνονται στον Πίνακα 1.

Πίνακας 1

	Μ Η Χ Α Ν Ε Σ		
	Α	Β	Γ
	25,40	23,40	20,00
	26,31	21,80	22,20
	24,10	23,50	19,75
	23,74	22,75	20,60
	25,10	21,60	20,40
Μέσοι	$\overline{Y}_{1.}=24{,}93$	$\overline{Y}_{2.}=22{,}61$	$\overline{Y}_{3.}=20{,}59$

οι τρεις μηχανές έχουν διαφορετικούς δειγματικούς μέσους χρόνους.

Ασκήσεις

Παρατηρούμε ότι πρόκειται για πλήρως τυχαιοποιημένο σχεδιασμό δεδομένου ότι οι 15 εργάτες τοποθετούνται τυχαία στις τρεις μηχανές. Ο «ένας παράγοντας», του οποίου την αντίδραση θέλουμε να προσδιορίσουμε, είναι η μηχανή με τρία «επίπεδα μεταχείρισης»: μηχανή Α, μηχανή Β, μηχανή Γ.

οι τρεις μηχανές έχουν διαφορετικούς δειγματικούς μέσους χρόνους.

Το ερώτημα που τίθεται είναι αν τα δειγματικά αυτά αποτελέσματα είναι σημαντικά διαφορετικά ώστε ο διευθυντής παραγωγής να οδηγηθεί στο συμπέρασμα ότι διαφέρουν σημαντικά και οι πληθυσμιακοί μέσοι όροι.

Για τον πειραματικό σχεδιασμό που περιγράψαμε η μηδενική και η εναλλακτική υπόθεση διατυπώνονται ως εξής:

H_o: Όλες οι μηχανές έχουν τους ίδιους μέσους.

ή H_o: $\mu_1 = \mu_2 = \mu_3$

H_1: Όλες οι μηχανές δεν έχουν τους ίδιους μέσους.

ή H_1: $\mu_i \neq \mu_j$ για $i \neq j$, $i, j = 1, 2, 3$

Ασκήσεις

1ος τρόπος

$$\bar{Y}_{..} = \frac{\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}}{n} = 22{,}71$$

$$SST = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{..})^2 = 58{,}2172$$

$$SSTr = \sum_{i=1}^{k} ni(\bar{Y}_{i.} - \bar{Y}_{..})^2 = 47{,}164$$

$$SSE = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_{i})^2 = 11{,}0532$$

Ασκήσεις

2^{ος} τρόπος

Πίνακας 2

	ΜΗΧΑΝΕΣ		
	1	2	3
	25,40	23,40	20,00
	26,31	21,80	22,20
	24,10	23,50	19,75
	23,74	22,75	20,60
	25,10	21,60	20,40
$Y_{i.}$	$Y_{1.}$ = 124,65	$Y_{2.}$ = 113,05	$Y_{3.}$ = 102,95
n_i	5	5	5
$\overline{Y}_{i.}$	24,93	22,61	20,59

Ίδιος αριθμός υποδειγμάτων

Ασκήσεις

$n = n_1 + n_2 + n_3 = 5 + 5 + 5 = 15$

$Y_{..} = 124{,}65 + 113{,}05 + 102{,}95 = 340{,}65$

$$\frac{Y_{..}^{2}}{n} = \frac{(340{,}65)^2}{15} = 7736{,}1615$$

$$\sum_{i=3}^{3}\sum_{j=1}^{5} Y_{ij}^2 = 25{,}40^2 + 26{,}31^2 + \ldots + 23{,}40^2 + \ldots + 20{,}40^2 = 7794{,}3787$$

$$\sum_{i=1}^{3}\frac{Y_{i.}^2}{n_i} = \frac{124{,}65^2 + 113{,}05^2 + 102{,}95^2}{15} = 7783{,}3255$$

Ασκήσεις

$$SSTr = \sum_{i=1}^{3} \frac{Y_{i.}^2}{n_i} - \frac{Y_{..}^2}{n}$$

$$= 7783{,}3255 - 7736{,}1615 = 47{,}164$$

$$SST = \sum_{i=1}^{3} \sum_{j=1}^{5} Y_{ij}^2 - \frac{Y_{..}^2}{n}$$

$$= 7794{,}3787 - 7736{,}1615 = \boxed{58{,}2172}$$

$$SSE = \sum_{i=1}^{3} \sum_{j=1}^{5} Y_{ij}^2 - \sum_{i=1}^{3} \frac{Y_{i.}^2}{n_i}$$

$$= 7.794{,}3787 - 7.783{,}3255 = 11{,}0532$$

$$SST = \boxed{58{,}2172} = 47{,}164 + 11{,}0532 = SSTr + SSE$$

Ασκήσεις

$$MSTr = \frac{SSTr}{k-1} = \frac{47{,}164}{2} = 23{,}582$$

$$F_0 = \frac{23{,}5820}{0{,}9211} = \boxed{25{,}60}$$

$$MSE = \frac{SSE}{n-k} = \frac{11{,}0532}{12} = 0{,}9211$$

Πίνακας Ανάλυσης Διακύμανσης

Πηγή	S S	D F	M S	F
Μεταξύ Αγωγών	47,1640	2	23,5820	25,60
Εντός Αγωγών	11,0532	12	0,9211	
Σύνολο	58,2172	14		

Ασκήσεις

Για $\alpha = 0{,}01$ → $F_{2,12,0{,}99} = 6{,}93$

Επειδή $F_0 > F_{2,12,0{,}99}$ απορρίπτουμε την H_o.

Είναι προφανές ότι τα συμπεράσματα ισχύουν με την προϋπόθεση ότι:

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2$$

και ότι οι πληθυσμοί ακολουθούν την κανονική κατανομή.

Ασκήσεις

Ο Πίνακας 3 περιέχει τις βαθμολογίες μαθητών της Α΄ Λυκείου τριών σχολείων της Αθήνας οι οποίοι έλαβαν μέρος στον Πανελλήνιο Διαγωνισμό της Μαθηματικής Εταιρείας.

<u>**Πίνακας 3**</u>

Βαθμολογία Μαθητών

(κλίμακα 0-6)

Μαθητές Σχολείου 1 (Y_{1i})	2	3	4				
Μαθητές Σχολείου 2 (Y_{2i})	4	5		6	4	5	6
Μαθητές Σχολείου 3 (Y_{3i})	2	3	3	4			

α. Να δημιουργηθεί ο Πίνακας Ανάλυσης Διακύμανσης.

β. Να ελεγχθεί σε επίπεδο σημαντικότητας α=0.05 η υπόθεση ότι η επίδραση των μαθητών στο συγκεκριμένο διαγωνισμό δεν επηρεάζεται από το σχολείο στο οποίο φοιτούν.

Ασκήσεις

ΜΑΘ. ΣΧΟΛ. 1	ΜΑΘ. ΣΧΟΛ. 2	ΜΑΘ. ΣΧΟΛ. 3	
2	4	2	
3	5	3	
4	5	3	
	6	4	
	4		
	5		
	6		
$Y_{1.}=9$	$Y_{2.}=35$	$Y_{3.}=12$	$Y_{..}=56$
$\overline{Y}_{1.}=3$	$\overline{Y}_{2.}=5$	$\overline{Y}_{3.}=3$	$\overline{Y}_{..}=4$

Διαφορετικός αριθμός υποδειγμάτων

n =14 (n_1=3, n_2=7, n_3=4), k=3, a=0.05

$$\sum_{i=1}^{k}\sum_{j=1}^{n_i} Y_{ij}^{\ 2} = 4+9+16+16+25+25+36+16+25+36+4+9+9+16=246$$

$$\frac{Y_{..}^{2}}{n} = \frac{56^2}{14} = \frac{3136}{14} = 224$$

$$\sum_{i=1}^{k}\frac{Y_{i.}^{2}}{n_i} = \frac{9^2}{3}+\frac{35^2}{7}+\frac{12^4}{4} = \frac{81}{3}+\frac{1225}{7}+\frac{144}{4} = 27+175+36 = 238$$

Ασκήσεις

$$SST_r = 238 - 224 \rightarrow SST_r = 14 \rightarrow MST_r = \frac{SST_r}{k-1} = \frac{14}{2} \rightarrow MST_r = 7$$

$$SSE = SST - SST_r \rightarrow SSE = 8 \rightarrow MSE = \frac{SSE}{n-k} = \frac{8}{11} \rightarrow MSE = 0.73$$

$$SST = 246 - 224 \rightarrow SST = 22$$

$$F_0 = \frac{7}{0.73} \rightarrow F_0 \cong 9.59$$

Ασκήσεις

Πίνακας Ανάλυσης Διακύμανσης

ΠΗΓΗ ΜΕΤΑΒΛΗΤΟΤΗΤΑΣ	SS	DF	MS	F_0
Μεταξύ Αγωγών	14	k-1=2	7	
Εντός Αγωγών	8	n-k=11	0.73	9.59
ΣΥΝΟΛΟ	22	n-1=13		

H_0: Η επίδοση των μαθητών δεν επηρεάζεται από το σχολείο

H_1: Η επίδοση των μαθητών επηρεάζεται από το σχολείο

Απορρίπτουμε την H_0 αν $F_0 \succ F_{k-1,n-k,1-a}$

$F_{k-1,\, n-k,\, 1-a} = F_{2,\, 11,\, 0.95} = 3.982$

Άρα $9.59 = F_0 > 3.982 = F_{2,\, 11,\, 0.95}$

Άρα απορρίπτουμε την H_0 και δεχόμαστε την H_1. Δηλαδή δεχόμαστε ότι η επίδοση των μαθητών επηρεάζεται από το σχολείο στο οποίο φοιτούν.

Ασκήσεις

Παρατηρούμε 4 ομάδες παιδιών που παίζουν με παιγνίδια διαφορετικού χρώματος και μετράμε πόσο χρόνο αφιερώνουν σε κάθε είδος παιγνιδιού.

#1	#2	#3	#4
κόκκινο	κίτρινο	πράσινο	μπλε
.... 10 τιμές ανά ομάδα			
$n_1=10$	$n_2 = 10$	$n_3 = 10$	$n_4 = 10$
$\langle x_1 \rangle=3.4$	$\langle x_2 \rangle = 5.0$	$\langle x_3 \rangle = 2.4$	$\langle x_4 \rangle = 2.5$
$s_1^2=4.5$	$s_2^2=5.6$	$s_3^2=1.2$	$s_4^2=1.8$

$N = 40$

$\langle x \rangle_{ολικό} = 3.3$

$s^2_{ολικό} = 4.1$

'Ισο μέγεθος

Ασκήσεις

Η διακύμανση κατά παράγοντες επομένως είναι αυτή που υπολογίζουμε από τις μέσες τιμές κάθε ομάδας, δηλαδή η διακύμανση των αριθμών 4.5, 5.6, 1.2 και 1.8.

Άρα, η διακύμανση κατά παράγοντες είναι χαρακτηριστικό της *κατανομής δειγματοληψίας* και σχετίζεται με την κατανομή του πληθυσμού όπως ορίζει το *Κεντρικό Οριακό Θεώρημα*, δηλαδή:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

δε γνωρίζουμε τη διασπορά της διακύμανσης της κατανομής δειγματοληψίας του πληθυσμού όλων των δυνατών παραγόντων, οπότε βασιζόμαστε σε μια εκτιμήτρια αυτής, το $s_{\bar{x}}$ που θα υπολογίσουμε από τους 4 αριθμούς, βάσει της σχέσης όπου διαιρούμε με το $n_{\bar{x}} - 1$ = 4-1 = 3.

Άρα, αντικαθιστούμε το $\sigma_{\bar{x}}$ με το $s_{\bar{x}}$.

Σύμφωνα με τη μεθοδολογία της ανάλυσης διασποράς θα προβούμε σε *διαφορετικές εκτιμήσεις της διασποράς του πληθυσμού* τις οποίες θα συγκρίνουμε μεταξύ τους, άρα αντικαθιστούμε επίσης, το σ με το s.

Στον παρονομαστή υπεισέρχεται το μέγεθος δείγματος, άρα εδώ n = 10.

Ασκήσεις

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \Rightarrow s^2 = n\, s_{\bar{x}}^2$$

$$\mu_{\bar{x}} = 3.425$$

Επομένως οι διαφορές από τη μέση τιμή και τα τετράγωνά τους είναι

0,07500	0,00562
1,67500	2,80563
-0,92500	0,85563
-0,82500	0,68063

Αθροίζουμε τα τετράγωνα και διαιρούμε διά 3 για να βρούμε 1.45 και πολλαπλασιάζουμε επί μέγεθος δείγματος n = 10 για να βρούμε
$s^2 = 10\,(1.45) = 14.5$
<u>Αυτή είναι η διασπορά κατά παράγοντες</u>.

Η διασπορά του σφάλματος για κάθε παράγοντα εκτιμάται αν πάρουμε τα τετράγωνα των τυπικών αποκλίσεων (διασπορές) για κάθε ομάδα, οπότε έστω ότι βρίσκουμε
$s_1^2 = 4.5$, $s_2^2 = 5.6$, $s_3^2 = 1.2$, $s_4^2 = 1.8$
Τώρα, για να βρούμε από αυτές μια δεύτερη εκτίμηση της διασποράς όλου του πληθυσμού, μπορούμε να πάρουμε το μέσο όρο των παραπάνω τιμών για να βρούμε s = 3.28

Ασκήσεις

F = διασπορά κατά παράγοντα : διασπορά λόγω σφάλματος

$$F = 14.5 / 3.28 = 4.42$$

Από πίνακες, για επίπεδο σημαντικότητας 1% βρίσκουμε ότι η κρίσιμη περιοχή είναι: $F >= 4.38$.

Συμπεραίνουμε ότι ο λόγος F για το συγκεκριμένο παράδειγμα είναι (οριακά) μέσα στην κρίσιμη περιοχή, άρα η παρατηρούμενη διασπορά λόγω διαφορετικών παραγόντων είναι σημαντική.
Αν η μηδενική υπόθεση ήταν ότι οι παράγοντες δεν ευθύνονται για τη διασπορά, τότε αυτή πρέπει να απορριφθεί.

Ασκήσεις

	#1		#2		#3		#4	
	x	x^2	x	x^2	x	x^2	x	x^2
	.... n_i τιμές ανά ομάδα i							
Sums	25	143	35	253	41	257	24	146
	n_1=5		$n_2 = 6$		$n_3 = 7$		$n_4 = 4$	
Συνολικά αθροίσματα:								
N = 22	$(\Sigma x)_{ολικό} = 125$				$(\Sigma x_T^2)_{ολικό} = 799$			

‘Ανόμοιο μέγεθος

Η λογική της μεθοδολογίας είναι παρόμοια, με τη διαφορά ότι οι μέσες τιμές και διασπορές πρέπει να *σταθμιστούν με βάση το μέγεθος των δειγμάτων*.

Ασκήσεις

Ολικό Διπλό Άθροισμα Τετραγώνων =
n φορές το (Άθροισμα Τετραγώνων Διασπορών κατά Παράγοντα)
+ Διπλό Άθροισμα Τετραγώνων Τυχαίων Διασπορών Ομάδων

ταυτότητα του Αθροίσματος Τετραγώνων για Ανάλυση Διασποράς

Για ίσο μέγεθος δειγμάτων η ταυτότητα γράφεται:

$$\sum_{i=1}^{g}\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{..}\right)^2 = n\sum_{i=1}^{g}\left(\bar{y}_{i.}-\bar{y}_{..}\right)^2+\sum_{i=1}^{g}\sum_{j=1}^{n}\left(y_{ij}-\bar{y}_{i.}\right)^2$$

όπου η τελεία συμβολίζει άθροισμα ως προς το δείκτη τον οποίο αντικαθιστά και η μπάρα παριστάνει διαίρεση αυτού του αθροίσματος με το αντίστοιχο πλήθος.

$$SS_T = SS_t + SS_E$$

Ασκήσεις

Για άνισα δείγματα, οι παραπάνω ποσότητες τροποποιούνται ως εξής:

$$SS_T = \sum_{i=1}^{g} \sum_{j=1}^{n} y_{ij}^2 - \frac{y_{\cdot\cdot}^2}{N} \quad , \quad SS_t = \sum_{i=1}^{g} \frac{y_{i\cdot}^2}{n} - \frac{y_{\cdot\cdot}^2}{N} \quad \text{και} \quad SS_E = SS_T - SS_t$$

όπου g σημαίνει “groups”, ομάδες.

$$\sum_{i=1}^{g} \sum_{j=1}^{n} y_{ij}^2 = 799 \quad , \quad y_{\cdot\cdot}^2 = 125^2 = 15625 \quad ,$$

$SS_T = 799 - (125)^2/22 = 88.8$

$SS_t = (25)^2/5 + (35)^2/6 + (41)^2/7 + (24)^2/4 - (125)^2/22 = 125+204.2 +240.1+144-710.2 = 3.1$

$SS_E = SS_T - SS_t = 88.8 - 3.1 = 85.7$

Ασκήσεις

Πηγή	Άθροισμα	df	Μέσο Τετράγωνο (εκτίμηση διασποράς)	F_0
Σύνολο	88.8	21	-----	
Παράγοντες	3.1	3	1.03	
Σφάλμα	85.7	18	4.76	

Οπότε το στατιστικό ελέγχου είναι $F = 1.03 / 4.76 = 0.22$ που είναι μικρότερο από τις τιμές της κατανομής του F τόσο για 1% όσο και για 5% (3.16 και 5.09, αντίστοιχα), άρα οι διαφορές οφείλονται σε τυχαίους και όχι στους εξεταζόμενους παράγοντες.