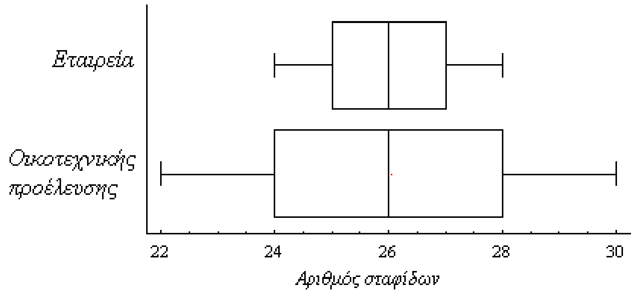
Εργαστήριο Μαθηματικών & Στατιστικής

2η Πρόοδος στο Μάθημα <u>Στατιστική</u> 24/01/2012
(Για τα Τμήματα Ε.Τ.Τ. και Γ.Β.)

1ο Θέμα [20] Επιλέξαμε 14 φακελάκια (της μισής ουγκιάς) που περιέχουν σταφίδες από την παραγωγή μιας εταιρείας συσκευασίας τροφίμων και άλλα 14 φακελάκια (της μισής επίσης ουγκιάς) οικοτεχνικής παραγωγής από την αποθήκη ενός συνεταιρισμού. Για κάθε φακελάκι μετρήσαμε τον αριθμό σταφίδων που περιέχει. Στο σχήμα που ακολουθεί φαίνεται το *θηκόκραμμα* της κατανομής καθενός από αυτά τα δύο δείγματα.



Για το δείγμα από την παραγωγή της εταιρείας τροφίμων δίνεται επίσης ότι, $\bar{x}_1 = 25.7$ & $s_1 = 1.3$ και για το δείγμα από το συνεταιρισμό ότι, $\bar{x}_2 = 26.2$ & $s_2 = 2.4$.
α) Βρείτε τα *τεταρτημόρια* της κατανομής καθενός από τα δύο δείγματα. **β)** Υπάρχουν ακραίες τιμές στα δύο δείγματα; **γ)** Να συγκρίνετε τα δύο δείγματα ως προς τη μεταβλητότητά τους. **δ)** Έστω $x_1 = 27$ ο αριθμός των σταφίδων που περιέχει ένα συγκεκριμένο φακελάκι από αυτά που επιλέξαμε από την εταιρεία τροφίμων, και $x_2 = 27$ ο αριθμός των σταφίδων που περιέχει ένα συγκεκριμένο φακελάκι από αυτά που επιλέξαμε από την αποθήκη του συνεταιρισμού. Τι μπορείτε να πείτε για τη θέση κάθε μιας από αυτές τις τιμές στην κατανομή του αντίστοιχου δείγματος;

2ο Θέμα [30] *(αναφέρεται στα δεδομένα του 1ου Θέματος)*
α) Βρείτε ένα *95% διάστημα εμπιστοσύνης* για το μέσο αριθμό σταφίδων που περιέχονται στα φακελάκια *(της μισής ουγκιάς)* της εταιρείας τροφίμων. Πώς αντιλαμβάνεσθε (πώς ερμηνεύετε) αυτό το διάστημα; **β)** Για να απαντήσετε στο ερώτημα (α), χρειάσθηκε να κάνετε κάποια παραδοχή; Αν ναι, να αναφέρετε έναν τρόπο ελέγχου της βασιμότητας αυτής της παραδοχής. Το *θηκόγραμμα* της κατανομής του δείγματος σας δίνει κάποια σχετική ένδειξη; **γ)** Αν από το ίδιο δείγμα κατασκευάσουμε για το μέσο αριθμό σταφίδων που περιέχονται στα φακελάκια της εταιρείας τροφίμων, ένα άλλο διάστημα εμπιστοσύνης με μεγαλύτερο συντελεστή εμπιστοσύνης, η <u>ακρίβεια</u> της εκτίμησης θα αυξηθεί, θα μειωθεί ή θα παραμείνει ίδια; **δ)** Γνωρίζουμε ότι η *δειγματική μέση τιμή* $\bar{X}$, είναι μια <u>*αμερόληπτη*</u> εκτιμήτρια της *μέσης τιμής* μ ενός πληθυσμού. Πώς αντιλαμβάνεσθε αυτή την ιδιότητα της $\bar{X}$;

3ο Θέμα [35] *(αναφέρεται στα δεδομένα του 1ου Θέματος)*
α) Δίνουν, σε επίπεδο σημαντικότητας 5%, τα ευρήματα στα δύο δείγματα στατιστικά σημαντικές αποδείξεις ότι ο μέσος αριθμός σταφίδων στα φακελάκια (της μισής ουγκιάς) της εταιρείας τροφίμων δεν είναι ίδιος με το μέσο αριθμό σταφίδων στα φακελάκια του συνεταιρισμού; **β)** Για να απαντήσετε στο ερώτημα (α), χρειάσθηκε να κάνετε κάποιες παραδοχές; Αν ναι, τα *θηκογράμματα* που δίνονται στο *1ο Θέμα* σας δίνουν κάποιες ενδείξεις για τη βασιμότητα αυτών των παραδοχών; **γ)** Με βάση την απάντησή σας στο (α), μπορείτε να συμπεράνετε αν τα ευρήματα στα δύο δείγματα δίνουν στατιστικά σημαντικές αποδείξεις ότι διαφέρουν οι δύο μέσοι σε επίπεδο σημαντικότητας 1%; **δ)** Μπορείτε να υπολογίσετε (ή γνωρίζετε) την

πιθανότητα το συμπέρασμά σας στο (α) να είναι λάθος; Εξηγείστε. **ε)** Η *P-τιμή* που δίνουν τα ευρήματα στα δύο δείγματα, θα μπορούσε να είναι ίση με 0.0026;

4ο Θέμα [20] Ένας υποψήφιος διδάκτορας, στο πλαίσιο της διδακτορικής του διατριβής που είχε ως αντικείμενο τη μελέτη του κύκλου ζωής της λευκής μύγας *(whitefly),* η οποία δημιουργεί πολύ σοβαρό πρόβλημα στις γεωργικές καλλιέργειες, σχεδίασε και εκτέλεσε στο εργαστήριο ένα πείραμα που σκοπό είχε τη διερεύνηση της επίδρασης στον κύκλο ζωής της λευκής μύγας δύο συγκεκριμένων παραγόντων, του επιπέδου της θερμοκρασίας και του είδους του φυτού που προσβάλλει. Συγκεκριμένα, σχεδίασε και εκτέλεσε το εξής πείραμα. Καθόρισε τρία επίπεδα θερμοκρασίας *(70^oF, 77^oF και 82^oF)* και δύο είδη φυτών *(βαμβάκι και αγγουριά)*. Στη συνέχεια, τοποθέτησε λευκές μύγες και στα δύο είδη φυτών υπό τις τρεις διαφορετικές θερμοκρασίες για όλους τους διαφορετικούς συνδυασμούς, *είδος φυτού-επίπεδο θερμοκρασίας*, και μέτρησε τον αριθμό των αυγών που γέννησαν. Για κάθε συνδυασμό *είδος φυτού-επίπεδο θερμοκρασίας* πήρε 5 παρατηρήσεις. Τα δεδομένα που συγκέντρωσε φαίνονται στον πίνακα που ακολουθεί.

		Επίπεδο θερμοκρασίας (Παράγοντας Β)		
		70^oF	*77^oF*	*82^oF*
Είδος φυτού (Παράγοντας Α)	*Βαμβάκι*	37 21 36 43 31	34 54 40 42 16	46 32 41 36 38
	Αγγουριά	50 53 25 37 48	59 53 31 69 51	62 71 49 43 59

Δίνεται επίσης ότι: *SSA=1512.3, SSB=487.5, SSE=2952.4, SSTot=5063.4,* $F_{2;9;0.05}=4.26$, $F_{4;9;0.05}=3.63$, $F_{2;24;0.05}=3.40$, $F_{2;6;0.05}=5.14$, $F_{6;2;0.05}=19.33$, $F_{1;24;0.05}=4.26$, $F_{4;12;0.05}=3.26$

α) Σε *επίπεδο σημαντικότητας 5%*, υποστηρίζουν αυτά τα πειραματικά δεδομένα ότι υπάρχει στατιστικά σημαντική επίδραση στον αριθμό των αυγών που γεννά η λευκή μύγα που να οφείλεται ***i)*** στην αλληλεπίδραση *είδος φυτού-επίπεδο θερμοκρασίας* ***ii)*** στο *είδος φυτού* ***iii)*** στο *επίπεδο θερμοκρασίας* **β)** Ποιες παραδοχές χρειάσθηκε να κάνετε για να απαντήσετε στο ερώτημα (α);

5ο Θέμα [20] Οι απόγονοι που μπορεί να προκύψουν από κάποια συγκεκριμένη διασταύρωση νυφίτσας, διακρίνονται ως προς το χρώμα της γούνας, σε τέσσερις τύπους, *D, P, A και S* και σύμφωνα με ένα μοντέλο κληρονομικότητας, αυτοί πρέπει να βρίσκονται σε αναλογία 9:3:3:1, αντίστοιχα. Από ένα σχετικό πείραμα, προέκυψαν 102 απόγονοι. Στον πίνακα που ακολουθεί φαίνεται πώς αυτοί κατανέμονται στους τέσσερις τύπους *D, P, A* και *S.*

	Τύπος απογόνου			
	D	*P*	*A*	*S*
Συχνότητα	53	17	24	8

α) Σε επίπεδο σημαντικότητας 1%, τα συγκεκριμένα πειραματικά δεδομένα, δίνουν στατιστικά σημαντικές αποδείξεις εναντίον του μοντέλου κληρονομικότητας;
β) Για τον έλεγχο που κάνατε στο (α), η *P-τιμή* που δίνουν τα πειραματικά δεδομένα μπορεί να είναι αριθμός μικρότερος του 0.01;

Από τα θέματα 4 και 5 πρέπει να επιλέξετε ένα. Δηλαδή, πρέπει να απαντήσετε είτε στα θέματα 1, 2, 3 και 4 είτε στα θέματα 1, 2, 3 και 5. Για το άριστα (10) απαιτούνται 100 μόρια και για τη βάση (5), 50 μόρια.

<u>**Ενδεικτικές απαντήσεις**</u>

1ο Θέμα: α) Τα Q_1, Q_2, Q_3, για το δείγμα από την εταιρεία τροφίμων είναι αντίστοιχα, 25, 26 και 27 και για το δείγμα από το συνεταιρισμό, αντίστοιχα, 24, 26 και 28
β) Όχι γιατί αν υπήρχαν θα έπρεπε να σημειώνονται εκτός των κεραιών του αντίστοιχου θηκογράμματος. **γ)** Το δείγμα από την εταιρεία τροφίμων έχει μικρότερη μεταβλητότητα (φαίνεται και από το <u>*εύρος*</u> και από το <u>*ενδοτεταρτημοριακό εύρος*</u> αλλά και από τους <u>*συντελεστές μεταβλητότητας*</u>). **δ)** Η *z-τιμή* της $x_1 = 27$ είναι $z = \dfrac{27-25.7}{1.3} = +1$ δηλαδή, αυτή η τιμή είναι μία τυπική απόκλιση μεγαλύτερη από τη μέση τιμή της κατανομής του δείγματος στο οποίο ανήκει. Η *z-τιμή* της $x_2 = 27$ είναι $z = \dfrac{27-26.2}{2.4} \approx +0.33$ δηλαδή, αυτή η τιμή είναι 0.33 τυπικές αποκλίσεις μεγαλύτερη από τη μέση τιμή της κατανομής του δείγματος στο οποίο ανήκει και επομένως βρίσκεται πιο κοντά στη μέση τιμή του δείγματος στο οποίο ανήκει από ότι η $x_1 = 27$ στη μέση τιμή του δείγματος στο οποίο αυτή ανήκει. Πληροφορίες για τη θέση αυτών των τιμών παίρνουμε επίσης από τα τεταρτημόρια. Δηλαδή, ...

2ο Θέμα: α) Με την υπόθεση ότι το δείγμα είναι **τυχαίο** και ότι προέρχεται από κανονικό πληθυσμό, ένα 95% διάστημα εμπιστοσύνης για τη μέση τιμή μ_1, του αριθμού των σταφίδων στα φακελάκια παραγωγής της εταιρείας τροφίμων είναι $\bar{x} \pm t_{n_1-1;0.025}\dfrac{s_1}{\sqrt{n_1}}$ ή $25.7 \pm t_{13;0.025}\dfrac{1.3}{\sqrt{14}}$ ή $25.7 \pm 2.16 \cdot 0.347$ ή 25.7 ± 0.75 ή [24.95, 26.45]. Αυτό σημαίνει ότι: Το διάστημα [24.95, 26.45], έχει 95% πιθανότητα να περιέχει την άγνωστη (πληθυσμιακή) μέση τιμή μ_1, του αριθμού των σταφίδων στα φακελάκια παραγωγής της εταιρείας τροφίμων. **β)** Ότι το δείγμα είναι **τυχαίο** και ότι προέρχεται από κανονικό πληθυσμό. Ένας τρόπος για να ελέγξουμε αν ευσταθεί η παραδοχή ότι το δείγμα προέρχεται από κανονικό πληθυσμό, είναι να κάνουμε X^2 έλεγχο καλής προσαρμογής των δεδομένων του δείγματος στην κανονική κατανομή με μέση τιμή και διασπορά που θα εκτιμήσουμε από το δείγμα. Επειδή στο θηκόγραμμα της κατανομής του δείγματος η διάμεσος είναι στο μέσο του ορθογωνίου, οι κεραίες είναι περίπου ίσου μήκους και δεν υπάρχουν ακραίες τιμές, η κατανομή του δείγματος συνηγορεί ότι αυτό προέρχεται από κανονικό πληθυσμό. **γ)** Θα μειωθεί η ακρίβεια της εκτίμησης (θα γίνει το διάστημα πιο φαρδύ). **δ)** Κατά μέσο όρο η εκτιμήτρια $\bar{X}$ εκτιμά σωστά την πληθυσμιακή μέση τιμή μ. Ούτε την υποεκτιμά, ούτε την υπερεκτιμά.

3ο Θέμα: Έστω μ_1 η (άγνωστη) μέση τιμή του αριθμού των σταφίδων στα φακελάκια παραγωγής της εταιρείας τροφίμων και μ_2 η (άγνωστη) μέση τιμή του αριθμού των σταφίδων στα φακελάκια οικοτεχνικής παραγωγής. Δίνονται: $\bar{x}_1 = 25.7$ με $s_1 = 1.3$, $\bar{x}_2 = 26.2$ με $s_2 = 2.4$ και $n_1 = n_2 = 14$. **α)** Θα κάνουμε τον έλεγχο της μηδενικής υπόθεσης $H_0 : \mu_1 = \mu_2$ έναντι της εναλλακτικής $H_1 : \mu_1 \neq \mu_2$ σε επίπεδο σημαντικότητας $\alpha = 0.05$. Με την υπόθεση ότι τα δείγματα είναι **τυχαία** και ότι προέρχονται από κανονικούς πληθυσμούς με ίσες διασπορές, η απορριπτική περιοχή του ελέγχου είναι: $t = \dfrac{|\bar{x}_1 - \bar{x}_2|}{s \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} > t_{n_1+n_2-2;0.025}$ ή $t = \dfrac{|\bar{x}_1 - \bar{x}_2|}{s \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} > t_{26;0.025}$ ή

$$t = \frac{|\bar{x} - \bar{y}|}{s \cdot \sqrt{\dfrac{1}{n_1} + \dfrac{1}{n_2}}} > 2.056$$

και επειδή, $s = \sqrt{\dfrac{(n_1 - 1) \cdot s_1^2 + (n_2 - 1) \cdot s_2^2}{n_1 + n_2 - 2}} = \sqrt{\dfrac{13 \cdot 1.3^2 + 13 \cdot 2.4^2}{26}} = 1.93$ έχουμε

$t = \dfrac{|-0.5|}{1.93 \cdot \sqrt{\dfrac{2}{14}}} = 0.6854$. Η τιμή αυτή δεν ανήκει στην απορριπτική περιοχή, και επομένως, η μηδενική υπόθεση, σε επίπεδο σημαντικότητας 5% δεν απορρίπτεται δηλαδή σε επίπεδο σημαντικότητας 5%, η διαφορά που παρατηρείται στους δύο δειγματικούς μέσους δεν είναι στατιστικά σημαντική.

β) Ότι τα δείγματα είναι **τυχαία** και ότι προέρχονται από κανονικούς πληθυσμούς με ίσες διασπορές. Επειδή και στα δύο θηκογράμματα η διάμεσος είναι στο μέσο του ορθογωνίου, οι κεραίες είναι περίπου ίσου μήκους και δεν υπάρχουν ακραίες τιμές, οι κατανομές των δειγμάτων συνηγορούν ότι τα δείγματα αυτά προέρχονται από κανονικούς πληθυσμούς. Όμως για την τρίτη παραδοχή (ίσες διασπορές) οι ενδείξεις από τα θηκογράμματα δε συνηγορούν υπέρ της παραδοχής ότι οι διασπορές των αντίστοιχων πληθυσμών είναι ίσες (δείτε και τους αντίστοιχους συντελεστές μεταβλητότητας). Πρέπει επομένως να γίνει έλεγχος ισότητας των διασπορών.

γ) Προφανώς η μηδενική υπόθεση δεν απορρίπτεται και σε επίπεδο σημαντικότητας 1%, γιατί σε 1% απαιτούνται ισχυρότερες αποδείξεις εναντίον της μηδενικής.

δ) Όχι, γιατί για να υπολογισθεί επακριβώς η πιθανότητα σφάλματος τύπου ΙΙ πρέπει να γνωρίζουμε την τιμή στην εναλλακτική.

ε) Όχι, γιατί όπως διαπιστώσαμε στο ερώτημα (α), η μηδενική υπόθεση σε επίπεδο σημαντικότητας 5% **δεν** απορρίπτεται ενώ με βάση αυτή την *P-τιμή* θα έπρεπε να απορρίπτεται.

4^o Θέμα: Θα κάνουμε, σε *επίπεδο σημαντικότητας 5%,* τους εξής ελέγχους:

$H_{0\gamma}$: *οι παράγοντες Α και Β* ***δεν*** *αλληλεπιδρούν*

έναντι της,

$H_{1\gamma}$: *οι παράγοντες Α και Β αλληλεπιδρούν.*

$H_{0\alpha}$: *ο παράγοντας Α* ***δεν*** *επιδρά*

έναντι της,

$H_{1\alpha}$: *ο παράγοντας Α επιδρά.*

H_{0b} : *ο παράγοντας Β* ***δεν*** *επιδρά*

έναντι της,

H_{1b} : *ο παράγοντας Β επιδρά.*

Πηγή μεταβλητότητας	Β.Ε.	Άθροισμα τετραγώνων SS	Μέσο άθροισμα τετραγώνων MS	Κριτήριο F
Παράγοντας Α	1	1512.3	1512.3	12.29
Παράγοντας Β	2	487.5	243.75	1.98
Αλληλεπίδραση ΑΒ	2	111.2	55.6	0.45
Σφάλμα	24	2952.4	123.017	
Ολική	29	5063.4		

α) i) $F_{AB} = 0.45$ δεν ανήκει στην περιοχή απόρριψης διότι $F_{2,24;0.05} = 3.40$ άρα η $H_{0\gamma}$ σε *επίπεδο σημαντικότητας 5%* δεν απορρίπτεται, δηλαδή, τα πειραματικά δεδομένα, σε *επίπεδο σημαντικότητας 5%,* δε δίνουν στατιστικά σημαντικές αποδείξεις, ότι υπάρχει επίδραση στον *αριθμό των αυγών που γεννά η λευκή μύγα* που να οφείλεται στην *αλληλεπίδραση είδος φυτού – επίπεδο θερμοκρασίας.*

ii) $F_A = 12.29 > F_{1,24;0.05} = 4.26$, άρα η $H_{0\alpha}$ σε *επίπεδο σημαντικότητας 5%* απορρίπτεται, δηλαδή, τα πειραματικά δεδομένα, σε *επίπεδο σημαντικότητας 5%,* δίνουν στατιστικά σημαντικές αποδείξεις, ότι υπάρχει επίδραση στον *αριθμό των αυγών που γεννά η λευκή μύγα* που οφείλεται στο *είδος φυτού.*

iii) $F_B = 1.98$ δεν ανήκει στην περιοχή απόρριψης διότι $F_{2,24;0.05} = 3.40$ άρα η H_{0b} σε *επίπεδο σημαντικότητας 5%* δεν απορρίπτεται, δηλαδή, τα πειραματικά δεδομένα, σε *επίπεδο σημαντικότητας 5%,* δε δίνουν στατιστικά σημαντικές αποδείξεις, ότι υπάρχει επίδραση στον *αριθμό των αυγών που γεννά η λευκή μύγα* που οφείλεται στο *επίπεδο θερμοκρασίας.*

β) Για όλους τους συνδυασμούς, *είδος φυτού-επίπεδο θερμοκρασίας*, οι αντίστοιχες παρατηρήσεις προέρχονται από κανονικούς πληθυσμούς με ίσες διασπορές.

5^{o} Θέμα: α) Έστω p_1 η πιθανότητα ένας απόγονος να είναι τύπου D, p_2 η πιθανότητα ένας απόγονος να είναι τύπου P, p_3 η πιθανότητα ένας απόγονος να είναι τύπου A και p_4 η πιθανότητα ένας απόγονος να είναι τύπου S.

Σε επίπεδο σημαντικότητας $\alpha = 0.01$, θα ελέγξουμε τη μηδενική υπόθεση, $H_0 : p_1 = \dfrac{9}{16}$ και $p_2 = \dfrac{3}{16}$ και $p_3 = \dfrac{3}{16}$ και $p_4 = \dfrac{1}{16}$ έναντι της εναλλακτικής, $H_1 : p_1 \neq \dfrac{9}{16}$ ή $p_2 \neq \dfrac{3}{16}$ ή $p_3 \neq \dfrac{4}{16}$ ή $p_4 \neq \dfrac{1}{16}$.

Επειδή $102 \cdot p_{i0} \geq 5$, $i = 1,2,3,4$, θα κάνουμε X^2 έλεγχο καλής προσαρμογής. Με την υπόθεση ότι η H_0 είναι αληθής, θα υπολογίσουμε για κάθε τύπο απογόνων την αναμενόμενη (θεωρητική) συχνότητα, $E_i = 102 \cdot p_{i0}$, $i = 1,2,3,4$.

	D	P	A	S	Σύνολο
O_i	53	17	24	8	102
p_i	9/16	3/16	3/16	1/16	1.00
$E_i = 102 \cdot p_{i0}$	57.4	19.1	19.1	6.4	102

Η απορριπτική περιοχή του ελέγχου είναι: $X^2 = \sum_{i=1}^{4} \dfrac{(O_i - E_i)^2}{E_i} > \chi^2_{3;0.01}$ ή

$X^2 = \sum_{i=1}^{4} \dfrac{(O_i - E_i)^2}{E_i} > 11.34$ και επειδή,

$$X^2 = \frac{(53-57.4)^2}{57.4} + \frac{(17-19.1)^2}{19.1} + \frac{(24-19.1)^2}{19.1} + \frac{(8-6.4)^2}{6.4} = 2.23\,,$$

η μηδενική υπόθεση, σε επίπεδο σημαντικότητας 1% δεν απορρίπτεται και επομένως αυτά τα πειραματικά δεδομένα, σε επίπεδο σημαντικότητας 1%, δε δίνουν στατιστικά σημαντικές αποδείξεις εναντίον του μοντέλου κληρονομικότητας.

β) Όχι, γιατί αν *P-Τιμή* < 0.01, η μηδενική υπόθεση σε επίπεδο σημαντικότητας 1% θα έπρεπε να απορρίπτεται.