Εργαστήριο Μαθηματικών & Στατιστικής

Γραπτή Εξέταση Περιόδου Φεβρουαρίου 2011 για τα Τμήματα Ε.Τ.Τ. και Γ.Β. στη <u>Στατιστική</u>

25/02/2011

1. [20] Η ποσότητα, έστω *X*, ενός συντηρητικού που περιέχεται σε φιάλες αναψυκτικού (ορισμένου μεγέθους) είναι τυχαία μεταβλητή η οποία, σύμφωνα με την εταιρεία παραγωγής, ακολουθεί κανονική κατανομή με μέση τιμή 40*mgr* και τυπική απόκλιση 1*mgr* (ανά φιάλη). Με την υπόθεση ότι ο ισχυρισμός της εταιρείας παραγωγής είναι αληθής, ποια είναι η πιθανότητα, **α)** μια τυχαία επιλεγμένη φιάλη να περιέχει ποσότητα του συγκεκριμένου συντηρητικού μεγαλύτερη από 42*mgr* **β)** από τέσσερις τυχαία επιλεγμένες φιάλες, μία το πολύ να περιέχει ποσότητα του συγκεκριμένου συντηρητικού μεγαλύτερη από 42*mgr* **γ)** η **μέση** ποσότητα του συγκεκριμένου συντηρητικού που περιέχεται σε τέσσερις τυχαία επιλεγμένες φιάλες να είναι μεγαλύτερη από 42*mgr*.

2. [10] Ι. Έστω δύο ενδεχόμενα *Α, Β* ενός δειγματοχώρου *Ω*, με $P(A) = 0.4$, $P(B) = 0.3$ και $P(A/B) = 0.45$. **α)** Τα *Α, Β* είναι ανεξάρτητα ή εξαρτημένα; **β)** Μπορεί η πιθανότητα $P(B/A)$ να είναι ίση με $P(B)$; **ΙΙ.** Δώστε την έννοια του τυχαίου πειράματος ή φαινομένου.

3. [40] Ένας φοιτητής του *Τμήματος Επιστήμης & Τεχνολογίας Τροφίμων*, στο πλαίσιο της πτυχιακής του εργασίας μελέτησε, μεταξύ άλλων, την ποσότητα κορεσμένων λιπαρών, έστω *X*, που περιέχεται στην κατσικίσια γραβιέρα που παράγει μια γεωργοκτηνοτροφική μονάδα παραγωγής προϊόντων βιολογικής γεωργίας στην Αρκαδία. Οι μετρήσεις που πήρε από εννέα «μερίδες» γραβιέρας που επέλεξε, σύμφωνα με ένα σχέδιο τυχαίας δειγματοληψίας, από την παραγωγή της συγκεκριμένης γεωργοκτηνοτροφικής μονάδας έδωσαν $\bar{x} = 7.65$ *gr/μερίδα* με τυπική απόκλιση $s = 1$*gr/μερίδα.* **α)** Βρείτε, με βάση το δείγμα που πήρε ο φοιτητής, ένα 95% διάστημα εμπιστοσύνης για τη μέση ποσότητα κορεσμένων λιπαρών που περιέχεται στην κατσικίσια γραβιέρα που παράγει η συγκεκριμένη γεωργοκτηνοτροφική μονάδα. Πώς αντιλαμβάνεσθε (ερμηνεύετε) αυτό το διάστημα εμπιστοσύνης; Για να το υπολογίσετε χρειάσθηκε να κάνετε κάποια παραδοχή; **β)** Αν με βάση το ίδιο δείγμα κατασκευάσετε για τη μέση ποσότητα κορεσμένων λιπαρών που περιέχεται στην κατσικίσια γραβιέρα της συγκεκριμένης γεωργοκτηνοτροφικής μονάδας, ένα άλλο διάστημα εμπιστοσύνης με <u>μεγαλύτερο</u> συντελεστή εμπιστοσύνης, η <u>ακρίβεια</u> της εκτίμησης θα αυξηθεί, θα μειωθεί ή θα παραμείνει ίδια; **γ)** Σύμφωνα με τις προδιαγραφές της γεωργοκτηνοτροφικής μονάδας, η μέση ποσότητα κορεσμένων λιπαρών που περιέχεται στην κατσικίσια γραβιέρα που παράγει είναι 7.5*gr/μερίδα*. Να ελέγξετε, σε επίπεδο σημαντικότητας 5%, αν τα δεδομένα του δείγματος που πήρε ο φοιτητής δίνουν στατιστικά σημαντικές αποδείξεις ότι η μέση ποσότητα κορεσμένων λιπαρών που περιέχεται στην κατσικίσια γραβιέρα που παράγει η γεωργοκτηνοτροφική μονάδα είναι μεγαλύτερη από 7.5*gr/μερίδα;* **δ)** Με βάση το συμπέρασμά σας στο (γ), μπορείτε να συμπεράνετε αν τα δεδομένα του δείγματος δίνουν, σε επίπεδο σημαντικότητας 1%, στατιστικά σημαντικές αποδείξεις ότι η μέση ποσότητα κορεσμένων λιπαρών που περιέχεται στην κατσικίσια γραβιέρα που παράγει η γεωργοκτηνοτροφική μονάδα είναι μεγαλύτερη από 7.5*gr/μερίδα,*; **ε)** Μπορείτε να υπολογίσετε την πιθανότητα το συμπέρασμά σας στο (γ) να είναι λάθος; Εξηγείστε. **στ)** Για τον έλεγχο που πρέπει να γίνει στο (γ) μπορεί το δείγμα να δίνει $P - \tau\iota\mu\acute{\eta} = 0.0324$;

4. [20] Ένας μεταπτυχιακός φοιτητής του *Τμήματος Γεωπονικής Βιοτεχνολογίας*, στο πλαίσιο μιας εργαστηριακής άσκησης, μελέτησε την περιεκτικότητα σε σωματιδιακές προσμείξεις ενός φαρμακευτικού σκευάσματος που χορηγείται ενδοφλεβίως σε ταύρους και το οποίο παράγεται από τρεις διαφορετικές εταιρείες Α1, Α2 και Α3. Για το σκοπό αυτό, έλεγξε 6 σκευάσματα από κάθε εταιρεία (τυχαία επιλεγμένα) και κατέγραψε τον αριθμό σωματιδίων ανά λίτρο. Οι μετρήσεις που πήρε δίνονται στον πίνακα που ακολουθεί.

		Αριθμός σωματιδίων *(ανά λίτρο)*					
Σκεύασμα εταιρείας	A1	255	264	342	331	234	217
	A2	105	288	98	275	221	240
	A3	577	515	214	413	401	260

Σε επίπεδο σημαντικότητας 5%, δίνουν αυτά τα δεδομένα στατιστικά σημαντικές αποδείξεις ότι υπάρχουν διαφορές μεταξύ των σκευασμάτων των τριών εταιρειών ως προς το μέσο αριθμό σωματιδίων που περιέχουν *(ανά λίτρο)*; Εξηγείστε τι έλεγχο, γιατί και με ποιες παραδοχές πρέπει να κάνετε.

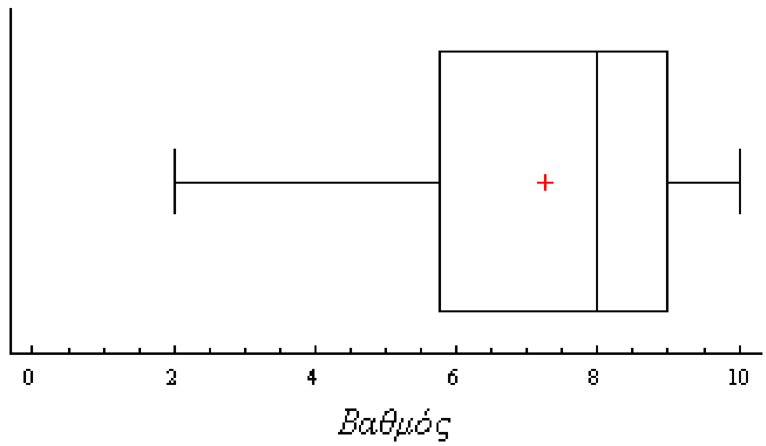
(Δίνονται: SSTot=260400, SSE=146754, $F_{15,2;0.05} = 19.4$, $F_{2,15;0.05} = 3.68$, $F_{2,17;0.05} = 3.59$, $F_{15,17;0.05} = 2.31$*).*

5. [20] Σε μια αγροτική περιοχή έγινε μια μελέτη για να ελεγχθεί η αποτελεσματικότητα ενός νέου εμβολίου, το οποίο χορηγείται σε δύο δόσεις, για προστασία από τη γρίπη. Στους 1000 κατοίκους της περιοχής δόθηκε η δυνατότητα να κάνουν το εμβόλιο δωρεάν (και εθελοντικά). Για κάθε κάτοικο, η ερευνητική ομάδα κατέγραψε πόσες δόσεις του εμβολίου έκανε (καμία, μία ή δύο) και αν αρρώστησε ή όχι από τη γρίπη. Τα δεδομένα που προέκυψαν δίνονται στον πίνακα που ακολουθεί.

	Δόσεις		
	Καμία	Μια	Δύο
Αρρώστησαν	24	9	13
Δεν Αρρώστησαν	289	100	565

Σε επίπεδο σημαντικότητας 5%, δίνουν αυτά τα δεδομένα σημαντικές αποδείξεις ότι ο αριθμός των περιστατικών γρίπης εξαρτάται από τον αριθμό των δόσεων;

6. [20] Στο σχήμα που ακολουθεί δίνεται το *θηκόγραμμα* της κατανομής των βαθμών όσων συμμετείχατε στη 2η πρόοδο του μαθήματος κατά το τρέχον ακαδημαϊκό έτος.



Δίνεται επίσης, η μέση τιμή των βαθμών, $\bar{x} = 7.28$, και η τυπική απόκλισή τους, $s = 2.01$. **α)** Με βάση αυτές τις πληροφορίες, τι συμπεραίνετε για τη θέση και τη μορφή της κατανομής των βαθμών; *(Γράψτε μια μικρή παράγραφο)*. **β)** Η *z-τιμή* του βαθμού ενός συμφοιτητή σας είναι ίση με 1.5. Πώς αντιλαμβάνεσθε (ερμηνεύετε) αυτή την τιμή; **γ)** Ο βαθμός ενός συμφοιτητή σας είναι 5.5. Τι μπορείτε να πείτε για τη θέση αυτού του βαθμού στην κατανομή των βαθμών; **δ)** Το ποσοστό των φοιτητών με βαθμό μεγαλύτερο του μέσου, είναι μεγαλύτερο ή μικρότερο του 50%;

Πρέπει να απαντήσετε στα θέματα 1, 2, 3 και σε δύο από τα 4, 5, 6 που εσείς θα επιλέξετε.
Για το άριστα (10) απαιτούνται 100 μόρια και για τη βάση (5) απαιτούνται 50 μόρια.

Ενδεικτικές Απαντήσεις

1ο Θέμα

Δίνεται ότι $X \sim N(40, 1^2)$

α) $P(X > 42) = P(\frac{X-40}{1} > \frac{42-40}{1}) = P(Z > 2) = 1 - P(Z \le 2) = 1 - \Phi(2) = 0.0228$

β) Έστω Y ο αριθμός φιαλών από τις τέσσερις που επελέγησαν, κάθε μία από τις οποίες περιέχει ποσότητα συντηρητικού μεγαλύτερη από 42mgr. Προφανώς, $Y \sim B(4, 0.0228)$.

$$P(Y \le 1) = P(Y = 0) + P(Y = 1) = \binom{4}{0} \cdot 0.0228^0 \cdot 0.9772^4 + \binom{4}{1} \cdot 0.0228^1 \cdot 0.9772^3 = 0.9969$$

γ) Έστω X_i η ποσότητα συντηρητικού στη φιάλη i $(i = 1,2,3,4)$. Δίνεται ότι, $X_i \sim N(40, 1^2)$, $i = 1,2,3,4$. Με την υπόθεση ότι οι X_i είναι ανεξάρτητες μεταξύ τους, $\overline{X} = \frac{\sum_{i=1}^{4} X_i}{4} \sim N(40, \frac{1^2}{4})$ άρα,

$$P(\overline{X} > 42) = P(\frac{\overline{X}-40}{1/2} > \frac{42-40}{1/2}) = P(Z > 4) = 1 - P(Z \le 4) = 1 - \Phi(4) \approx 0 .$$

2ο Θέμα

Ι. α) Είναι εξαρτημένα γιατί αν ήταν ανεξάρτητα έπρεπε $P(A/B) = P(A)$ ενώ αυτό δε συμβαίνει. β) Δε μπορεί να είναι $P(B/A) = P(B)$ γιατί τότε τα *A, B,* έπρεπε να είναι ανεξάρτητα όμως είναι εξαρτημένα. **ΙΙ.** Δες σχετικές σημειώσεις.

3ο Θέμα

α) Με την υπόθεση ότι το δείγμα προέρχεται από κανονικό πληθυσμό, ένα 95% διάστημα εμπιστοσύνης για τη μέση ποσότητα κορεσμένων λιπαρών που περιέχεται στην κατσικίσια γραβιέρα που παράγει η συγκεκριμένη γεωργοκτηνοτροφική μονάδα είναι $\overline{x} \pm t_{n-1;0.025} \frac{s}{\sqrt{n}}$ ή $7.65 \pm t_{8;0.025} \frac{1}{\sqrt{9}}$ ή $7.65 \pm 2.306 \cdot \frac{1}{3}$ ή 7.65 ± 0.7687 ή $[6.88,\ 8.42]$. Αυτό σημαίνει ότι: Το διάστημα $[6.88,\ 8.42]$ έχει 95% πιθανότητα να περιέχει την άγνωστη (πληθυσμιακή) μέση ποσότητα κορεσμένων λιπαρών που περιέχεται στην κατσικίσια γραβιέρα που παράγει η συγκεκριμένη γεωργοκτηνοτροφική μονάδα. **β)** Η ακρίβεια της εκτίμησης θα μειωθεί (το διάστημα θα γίνει πιο πλατύ). **γ)** Έστω μ η άγνωστη πληθυσμιακή μέση ποσότητα κορεσμένων λιπαρών που περιέχεται στην κατσικίσια γραβιέρα που παράγει η συγκεκριμένη γεωργοκτηνοτροφική μονάδα. Θα κάνουμε τον έλεγχο της μηδενικής υπόθεσης $H_0 : \mu = 7.5 gr / μερίδα$ έναντι της εναλλακτικής $H_1 : \mu > 7.5 gr / μερίδα$ σε επίπεδο σημαντικότητας $\alpha = 0.05$. Με την υπόθεση ότι το δείγμα προέρχεται από κανονικό πληθυσμό, η απορριπτική περιοχή του ελέγχου είναι: $t = \frac{\overline{x} - 7.5}{s/\sqrt{n}} > t_{n-1;0.05}$ ή

$$t = \frac{\overline{x} - 7.5}{s/\sqrt{n}} > t_{8;0.05} \text{ ή } t = \frac{\overline{x} - 7.5}{s/\sqrt{n}} > 1.860$$

και επειδή, $t = \dfrac{7.65 - 7.5}{1/\sqrt{9}} = 0.45$, η μηδενική υπόθεση, σε επίπεδο σημαντικότητας 5%, δεν απορρίπτεται και επομένως σε επίπεδο σημαντικότητας 5%, τα δεδομένα του δείγματος δε δίνουν στατιστικά σημαντικές αποδείξεις ότι η μέση ποσότητα κορεσμένων λιπαρών που περιέχεται στην κατσικίσια γραβιέρα που παράγει η συγκεκριμένη γεωργοκτηνοτροφική μονάδα υπερβαίνει τα 7.5*gr/μερίδα*.
δ) Εφόσον η μηδενική υπόθεση σε επίπεδο σημαντικότητας 5% δεν απορρίπτεται, κατά μείζονα λόγο δεν απορρίπτεται και σε οποιοδήποτε μικρότερο (αφού σε μικρότερο επίπεδο σημαντικότητας απαιτούνται πιο σημαντικές αποδείξεις εναντίον της μηδενικής). **ε)** Ζητείται αν μπορούμε να υπολογίσουμε την πιθανότητα σφάλματος τύπου ΙΙ. Επακριβώς δε μπορούμε. Μόνο ως συνάρτηση της πραγματικής τιμής $\mu_1 > 7.5$ *gr/μερίδα* μπορούμε (που όμως μας είναι άγνωστη). **στ)** Όχι γιατί τότε, σε επίπεδο σημαντικότητας 5% η μηδενική υπόθεση έπρεπε να απορρίπτεται.

4ο Θέμα

Έστω $\mu_i, i = 1,2,3$ ο μέσος αριθμός σωματιδίων (ανά λίτρο) που περιέχεται στο σκεύασμα της εταιρείας $A_i, i = 1,2,3$. Πρόκειται για έλεγχο των μέσων τριών πληθυσμών με βάση τρία τυχαία και ανεξάρτητα δείγματα, ένα από κάθε πληθυσμό (εντελώς τυχαιοποιημένο σχέδιο). Με την υπόθεση ότι τα δείγματα προέρχονται από κανονικούς πληθυσμούς με ίσες διασπορές θα κάνουμε έλεγχο ANOVA με έναν παράγοντα (εταιρεία παραγωγής), έστω *A*. Δηλαδή, σε επίπεδο σημαντικότητας 5%, θα ελέγξουμε τη μηδενική υπόθεση, $H_0 : \mu_1 = \mu_2 = \mu_3$ έναντι της εναλλακτικής, $H_1 : \mu_i \neq \mu_j$ για τουλάχιστον ένα ζευγάρι $i,\ j\ \mu\varepsilon\ i, j = 1,2,3$.

Ο πίνακας ANOVA είναι:

Πηγή μεταβλητότητας	Β.Ε.	Άθροισμα τετραγώνων SS	Μέσο άθροισμα τετραγώνων MS	Κριτήριο F
Παράγοντας Α	2	$SSA = 113646$	$MSA = \dfrac{SSA}{2} = 56823$	$F_A = \dfrac{MSA}{MSE} = 5.81$
Σφάλμα	15	$SSE = 146754$	$MSE = \dfrac{SSE}{15} = 9784$	
Ολική	17	$SSTot = 260400$		

Η απορριπτική περιοχή του ελέγχου είναι, $F_A > F_{2,15;0.05}$ ή $F_A > 3.68$ και επειδή $F_A = 5.81 > 3.68$, η μηδενική υπόθεση, σε επίπεδο σημαντικότητας 5%, απορρίπτεται και επομένως, σε επίπεδο σημαντικότητας 5%, τα ευρήματα στα δείγματα δίνουν στατιστικά σημαντικές αποδείξεις ότι υπάρχουν διαφορές μεταξύ των σκευασμάτων των τριών εταιρειών ως προς το μέσο αριθμό σωματιδίων που περιέχουν *(ανά λίτρο)*.

5ο Θέμα

Έστω p_1 η πιθανότητα του ενδεχομένου A_1: *ένας κάτοικος αρρώστησε*, p_2 η πιθανότητα του ενδεχομένου A_2: *ένας κάτοικος δεν αρρώστησε*, q_0 η πιθανότητα του ενδεχομένου D_0:*ένας κάτοικος έκανε μηδέν δόσεις*, q_1 η πιθανότητα του ενδεχομένου D_1:*ένας κάτοικος έκανε μια μόνο δόση* και q_2 η πιθανότητα του ενδεχομένου D_2: *ένας κάτοικος έκανε δύο δόσεις*. Έστω επίσης, $p_{ij} = P(A_i \cap D_j),\ i = 1,2\ \kappa\alpha\iota\ j = 0,1,2$.

Σε επίπεδο σημαντικότητας $\alpha = 0.05$, θα ελέγξουμε τη μηδενική υπόθεση, $H_0 : p_{ij} = p_i \cdot q_j, \ \ \forall i, j$, έναντι της εναλλακτικής,

$H_1 : p_{ij} \neq p_i \cdot q_j$, για τουλάχιστον ένα ζευγάρι $i, \ j \ \mu\varepsilon \ \ i = 1,2 \ \kappa\alpha\iota \ j = 0,1,2$.

Ή αλλιώς, θα ελέγξουμε τη μηδενική υπόθεση,

H_0 : *Ο αριθμός των περιστατικών γρίπης <u>δεν εξαρτάται</u> (δεν επηρεάζεται) από τον αριθμό των δόσεων εμβολίου,*

έναντι της εναλλακτικής,

H_1 : *Ο αριθμός των περιστατικών γρίπης <u>εξαρτάται</u> από τον αριθμό των δόσεων εμβολίου.*

Θα κάνουμε X^2 έλεγχο ανεξαρτησίας. Με την υπόθεση ότι η H_0 είναι αληθής, θα εκτιμήσουμε για κάθε συνδυασμό (αρρώστησε ή όχι-αριθμός εμβολίων) την αναμενόμενη (θεωρητική) συχνότητα ($\hat{E}_{ij} = n \cdot \hat{p}_i \hat{q}_j$). Ο Πίνακας με τις παρατηρηθείσες συχνότητες O_{ij} και τις αναμενόμενες $\hat{E}_{ij}$ είναι:

	Καμία	Μια	Δύο	Σύνολο
Αρρώστησαν	24 (14.40)	9 (5.01)	13 (26.59)	46
Δεν Αρρώστησαν	289 (298.6)	100 (103.99)	565 (551.41)	954
Σύνολο	313	109	578	1000

Επειδή $\hat{E}_{ij} = n \cdot \hat{p}_i \hat{q}_j \geq 5, \forall i, j$, η απορριπτική περιοχή του ελέγχου είναι: $X^2 = \sum_{\forall i,j} \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} > \chi^2_{2;0.05}$ ή $X^2 = \sum_{\forall i,j} \frac{(O_{ij} - \hat{E}_{ij})^2}{\hat{E}_{ij}} > 5.991$ και επειδή,

$$X^2 = \frac{(24 - 14.40)^2}{14.40} + \ldots + \frac{(565 - 551.41)^2}{551.41} = 17.33 > 5.991,$$

η μηδενική υπόθεση, σε επίπεδο σημαντικότητας 5% απορρίπτεται και επομένως τα πειραματικά δεδομένα, σε επίπεδο σημαντικότητας 5%, δίνουν στατιστικά σημαντικές αποδείξεις ότι ο αριθμός των περιστατικών γρίπης εξαρτάται από τον αριθμό των δόσεων εμβολίου.

6^o Θέμα

α) $Q_1 = 5.75, \ Q_2 = \delta = 8, \ Q_3 = 9$, ελάχιστος βαθμός 2, μέγιστος βαθμός 10, δεν έχει ακραίες τιμές. Παρουσιάζει αρνητική ασυμμετρία.

β) Ο βαθμός του συμφοιτητή μας βρίσκεται 1.5 τυπικές αποκλίσεις πάνω από τη μέση τιμή των βαθμών.

γ) Η *z-τιμή* αυτού του βαθμού είναι $\frac{5.5 - 7.28}{2.01} = -0.89$ άρα βρίσκεται 0.89 τυπικές αποκλίσεις κάτω από τη μέση τιμή των βαθμών. Επίσης, ανήκει στο 25% των χαμηλότερων βαθμών.

δ) Επειδή η μέση τιμή είναι μικρότερη από τη διάμεσο, το ποσοστό αυτό είναι μεγαλύτερο του 50%.