Εργαστήριο Μαθηματικών & Στατιστικής
Μάθημα: Στατιστική

Γραπτή Εξέταση Περιόδου Φεβρουαρίου 2012
για τα Τμήματα Ε.Τ.Τ. και Γ.Β.

16/02/2012

1ο Θέμα [20] Η ποσότητα, έστω *X*, φυτικών ινών που περιέχεται σε ψωμί ολικής άλεσης με σίκαλη είναι τυχαία μεταβλητή και, σύμφωνα με την εταιρεία παραγωγής, ακολουθεί κανονική κατανομή με μέση τιμή $\mu = 1.5gr$ και τυπική απόκλιση $\sigma = 0.5gr$ *(ανά φέτα των 25gr περίπου).* Με την υπόθεση ότι ο ισχυρισμός της εταιρείας παραγωγής είναι αληθής, ποια είναι η πιθανότητα,
α) μια τυχαία επιλεγμένη φέτα να περιέχει το πολύ 1.24*gr* φυτικών ινών
β) από τέσσερις τυχαία επιλεγμένες φέτες, ακριβώς μία να περιέχει το πολύ 1.24*gr* φυτικών ινών
γ) η **μέση** ποσότητα φυτικών ινών που περιέχεται σε 36 τυχαία επιλεγμένες φέτες να είναι το πολύ 1.24*gr*.

2ο Θέμα [40] *(συνέχεια του 1ου Θέματος)***.** Ένας φοιτητής του *Τμήματος Επιστήμης & Τεχνολογίας Τροφίμων* προκειμένου, στο πλαίσιο της πτυχιακής του εργασίας, να ελέγξει αν η μέση ποσότητα φυτικών ινών που περιέχεται ανά φέτα είναι πράγματι 1.5*gr*, πήρε σύμφωνα με ένα σχέδιο τυχαίας δειγματοληψίας 36 φέτες από την παραγωγή της συγκεκριμένης εταιρείας και μέτρησε την ποσότητα φυτικών ινών που περιέχεται σε αυτές. Οι μετρήσεις αυτές έδωσαν $\bar{x} = 1.24\,gr$.
α) Σε επίπεδο σημαντικότητας 1%, δίνουν τα δεδομένα του δείγματος που πήρε ο φοιτητής στατιστικά σημαντικές αποδείξεις ότι η μέση ποσότητα φυτικών ινών που περιέχεται ανά φέτα στο ψωμί ολικής άλεσης με σίκαλη της συγκεκριμένης εταιρείας δεν είναι 1.5*gr* αλλά μικρότερη; *(Η διασπορά* $\sigma^2 = 0.5^2\,gr^2$ *της X δεν αμφισβητείται).*
β) Με βάση το συμπέρασμά σας στο (α), μπορείτε να συμπεράνετε αν τα δεδομένα του δείγματος δίνουν, σε επίπεδο σημαντικότητας 5%, στατιστικά σημαντικές αποδείξεις ότι η μέση ποσότητα φυτικών ινών που περιέχεται ανά φέτα στο ψωμί ολικής άλεσης με σίκαλη της συγκεκριμένης εταιρείας δεν είναι 1.5*gr* αλλά μικρότερη;
γ) Το συμπέρασμα του (α) ή το συμπέρασμα του (β) ερωτήματος θα προτείνατε στον φοιτητή να αναφέρει στην πτυχιακή του εργασία; Εξηγείστε.
δ) Γνωρίζετε την πιθανότητα το συμπέρασμα στο ερώτημα (α) να είναι λάθος; Εξηγείστε.
ε) Μπορείτε να απαντήσετε στο ερώτημα (α) με βάση την απάντησή σας στο ερώτημα (1γ); Αν ναι, τι θα προτείνατε στον φοιτητή να αναφέρει στην πτυχιακή του εργασία για να υποστηρίξει με τον καλύτερο τρόπο το συμπέρασμά του;
στ) Να υπολογίσετε ένα 95% διάστημα εμπιστοσύνης για τη μέση ποσότητα φυτικών ινών που περιέχεται ανά φέτα στο ψωμί ολικής άλεσης με σίκαλη της συγκεκριμένης εταιρείας και να εξηγήσετε ποιο είναι το περιθώριο σφάλματος αυτής της εκτίμησης. Πώς αντιλαμβάνεσθε/ερμηνεύετε αυτό το διάστημα;

3ο Θέμα [20] Ένας ερευνητής μέτρησε τη συγκέντρωση γλυκόζης (σε *mg/dLi*) στο αριστερό και το δεξί μάτι 35, τυχαία επιλεγμένων, υγιών σκυλιών συγκεκριμένης ράτσας. Ας συμβολίσουμε με *X* και *Y* τη συγκέντρωση γλυκόζης στο αριστερό και το δεξί μάτι, αντίστοιχα, υγιών σκύλων της συγκεκριμένης ράτσας. Για τις 35 μετρήσεις, $x_1, x_2, ..., x_{35}$, στο αριστερό μάτι και τις 35 μετρήσεις, $y_1, y_2, ..., y_{35}$, στο δεξί, ο ερευνητής υπολόγισε τη μέση τιμή και την τυπική απόκλισή τους, $\bar{x} = 84.6$ *mg/dLi*, $s_x = 11.64$ *mg/dLi*, $\bar{y} = 84.83$ *mg/dLi*, $s_y = 11.72$ *mg/dLi*, αντίστοιχα. Υπολόγισε επίσης την τυπική απόκλιση, s_d, των διαφορών, $d_i = x_i - y_i$, $i = 1,2,...,35$, μεταξύ αριστερού και δεξιού ματιού και βρήκε ότι, $s_d = 2.16$ *mg/dLi*. Χρησιμοποιείστε <u>όσα και όποια</u> από αυτά τα στατιστικά κρίνετε, για να απαντήσετε στο εξής ερώτημα: σε επίπεδο σημαντικότητας 5%, υποστηρίζουν οι μετρήσεις που έκανε ο ερευνητής ότι υπάρχει στατιστικά σημαντική διαφορά στη μέση συγκέντρωση γλυκόζης μεταξύ των δύο ματιών υγιών σκύλων της συγκεκριμένης ράτσας; Χρειάσθηκε να κάνετε κάποιες παραδοχές;

4^{ο} Θέμα [20] Γνωρίζουμε (από πρόσφατη έρευνα) ότι στο 1% των αλλαντικών που διατίθενται στην ελληνική αγορά, η ποσότητα ενός συγκεκριμένου συντηρητικού που περιέχεται σε αυτά υπερβαίνει το ανώτατο επιτρεπτό όριο. Η μέθοδος που εφαρμόζει ο ΕΦΕΤ όταν πραγματοποιεί σχετικούς ελέγχους δεν είναι απολύτως ακριβής. Συγκεκριμένα: δίνει θετικό αποτέλεσμα στο 98% των δειγμάτων που πράγματι υπερβαίνουν το όριο αλλά και στο 5% των δειγμάτων που δεν υπερβαίνουν το όριο.
α) Ποια είναι η πιθανότητα σε ένα τυχαία επιλεγμένο δείγμα αλλαντικού ο έλεγχος του ΕΦΕΤ να δώσει θετικό αποτέλεσμα;
β) Αν σε ένα τυχαία επιλεγμένο δείγμα αλλαντικού ο έλεγχος του ΕΦΕΤ έδωσε θετικό αποτέλεσμα, ποια είναι η πιθανότητα η ποσότητα συντηρητικού που περιέχεται σε αυτό πράγματι να υπερβαίνει το όριο;

5^{ο} Θέμα. [20] Ένας οργανισμός ποιοτικού ελέγχου τροφίμων κατατάσσει τα μήλα, ως προς διάφορα ποιοτικά χαρακτηριστικά, σε τέσσερις κατηγορίες, Α, Β, Γ και Δ, αντίστοιχα. Τα μήλα που παράγονται στο οροπέδιο της Τεγέας, θεωρείται (από εμπειρία πολλών ετών) ότι κατατάσσονται σε αυτές τις τέσσερις κατηγορίες σε ποσοστό, 0.20, 0.40, 0.30 και 0.10, αντίστοιχα. Εξακόσια μήλα που επελέγησαν, σύμφωνα με ένα σχέδιο τυχαίας δειγματοληψίας, από το κτήμα ενός συγκεκριμένου παραγωγού του οροπεδίου της Τεγέας, κατετάγησαν στις τέσσερις κατηγορίες ως εξής:

	Κατηγορία			
	Α	Β	Γ	Δ
Παρατηρηθείσα συχνότητα	150	290	140	20

Σε επίπεδο σημαντικότητας 1%, δίνουν αυτά τα δεδομένα στατιστικά σημαντικές αποδείξεις, ότι τα ποσοστά των μήλων παραγωγής του συγκεκριμένου παραγωγού στις τέσσερις κατηγορίες, διαφέρουν από τα αντίστοιχα που θεωρείται ότι ισχύουν για την παραγωγή σε όλο το οροπέδιο;

Πρέπει να απαντήσετε στα θέματα 1, 2 και σε <u>δύο</u> από τα 3, 4, 5 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 $X \sim N(1.5,\ 0.5^2)$

α) $P(X \le 1.24) = P(\frac{X-1.5}{0.5} \le \frac{1.24-1.5}{0.5}) = P(Z \le -0.52) = 1-\Phi(0.52) = 0.3015$

β) Έστω Y ο αριθμός φετών από τις τέσσερις που επελέγησαν, κάθε μία από τις οποίες περιέχει ποσότητα φυτικών ινών το πολύ 1.24gr. Προφανώς, $Y \sim B(4,\ 0.3015)$. Ζητάμε την πιθανότητα, $P(Y=1) = \binom{4}{1} \cdot 0.3015^1 \cdot 0.6985^3 = \ldots$

γ) Έστω X_i η ποσότητα φυτικών ινών στη φέτα i $(i = 1,2,\ldots,36)$. Δίνεται ότι, $X_i \sim N(1.5,\ 0.5^2)$, $i = 1,2,\ldots,36$. Με την υπόθεση ότι οι X_i είναι ανεξάρτητες μεταξύ τους, $\overline{X} = \frac{\sum_{i=1}^{36} X_i}{36} \sim N(1.5,\ \frac{0.5^2}{36})$ ή $\overline{X} \sim N(1.5,\ 0.0833^2)$ άρα,

$P(\overline{X} \le 1.24) = P(\frac{\overline{X}-1.5}{0.0833} \le \frac{1.24-1.5}{0.0833}) = P(Z \le -3.12) = 1-\Phi(3.12) = 0.0009$.

2ο Θέμα

α) Έστω μ η ελεγχόμενη (άγνωστη) πληθυσμιακή μέση ποσότητα φυτικών ινών που περιέχεται στο ψωμί ολικής άλεσης με σίκαλη (ανά φέτα) που παράγει η συγκεκριμένη εταιρεία. Θα κάνουμε τον έλεγχο της μηδενικής υπόθεσης $H_0 : \mu = 1.5gr$ έναντι της εναλλακτικής $H_1 : \mu < 1.5gr$ σε επίπεδο σημαντικότητας $\alpha = 0.01$. Επειδή η διασπορά του πληθυσμού, σ^2, είναι γνωστή, η απορριπτική περιοχή του ελέγχου είναι: $z = \frac{\overline{x}-1.5}{\sigma/\sqrt{n}} < -z_{0.01}$ ή $z = \frac{\overline{x}-1.5}{\sigma/\sqrt{n}} < -2.33$

και επειδή, $z = \frac{1.24-1.5}{0.5/\sqrt{36}} = -3.12 < -2.33$, η μηδενική υπόθεση, σε επίπεδο σημαντικότητας 1%, απορρίπτεται και επομένως σε επίπεδο σημαντικότητας 1%, τα δεδομένα του δείγματος δίνουν στατιστικά σημαντικές αποδείξεις ότι η μέση ποσότητα φυτικών ινών που περιέχεται στο ψωμί ολικής άλεσης με σίκαλη (ανά φέτα) που παράγει η συγκεκριμένη εταιρεία δεν είναι $1.5gr$ αλλά μικρότερη.

β) Εφόσον η μηδενική υπόθεση σε επίπεδο σημαντικότητας 1% απορρίπτεται, κατά μείζονα λόγο απορρίπτεται και σε οποιοδήποτε μεγαλύτερο (αφού σε μεγαλύτερο επίπεδο σημαντικότητας απαιτούνται λιγότερο σημαντικές αποδείξεις εναντίον της μηδενικής). **γ)** Προφανώς του ερωτήματος (α) γιατί υποδηλώνει ισχυρότερες αποδείξεις εναντίον της μηδενικής. **δ)** Ναι. Από τον ορισμό του επιπέδου σημαντικότητας, είναι το πολύ 1%. **ε)** Στο ερώτημα (1γ) προφανώς υπολογίσαμε την *P-Value* του ελέγχου και επομένως μπορούμε να απαντήσουμε στο ερώτημα (α) συγκρίνοντας το επίπεδο σημαντικότητας με την *P-Value*. Στον φοιτητή θα προτείναμε να αναφέρει την *P-Value* γιατί έτσι θα καταδείξει επακριβώς πόσο ισχυρές αποδείξεις εναντίον της μηδενικής προκύπτουν από το συγκεκριμένο δείγμα. Είναι μάλιστα πολύ ισχυρές αφού η *P-Value* είναι πολύ μικρή (0.0009). **στ)** Το ζητούμενο 95% διάστημα εμπιστοσύνης είναι $\overline{x} \pm z_{0.025}\frac{\sigma}{\sqrt{n}}$ ή $1.24 \pm 1.96\frac{0.5}{\sqrt{36}}$ ή

1.24 ± 0.16. Το περιθώριο σφάλματος, με πιθανότητα 0.95, είναι 0.16. Το διάστημα αυτό περιέχει την άγνωστη μέση τιμή με πιθανότητα 0.95.

3ο Θέμα

α) Πρόκειται για σύγκριση κατά ζεύγη. Επομένως αν $D = X - Y$, θα κάνουμε, σε επίπεδο σημαντικότητας 5%, τον έλεγχο της μηδενικής υπόθεσης,

$H_1 : \mu_D = 0$

έναντι της εναλλακτικής,

$H_1 : \mu_D \neq 0$

Επειδή το δείγμα των διαφορών είναι μεγάλο (>30) η απορριπτική περιοχή είναι:

$z = \dfrac{|\bar{d}|}{s_d / \sqrt{n}} > z_{0.025} = 1.96$ και επειδή, $d = \bar{x} - \bar{y} = -0.23$ και $s_d = 2.16$, έχουμε,

$$z = \frac{|\bar{d}|}{s_d / \sqrt{n}} = \frac{0.23}{2.16 / \sqrt{35}} = 0.63$$

Επομένως η μηδενική υπόθεση, σε επίπεδο σημαντικότητας 5%, δεν απορρίπτεται δηλαδή, σε επίπεδο σημαντικότητας 5%, τα ευρήματα στα δείγματα δε δίνουν στατιστικά σημαντικές αποδείξεις ότι υπάρχει διαφορά στη μέση συγκέντρωση γλυκόζης μεταξύ των δύο ματιών υγιών σκύλων της συγκεκριμένης ράτσας. Δεν χρειάσθηκε να κάνουμε κάποια παραδοχή γιατί τα δείγματα είναι μεγάλα.

4ο Θέμα

Θεωρούμε τα ενδεχόμενα:

Α: Το αλλαντικό περιέχει ποσότητα συντηρητικού που υπερβαίνει το ανώτατο επιτρεπτό όριο

Θ: Το αποτέλεσμα του ελέγχου είναι θετικό.

Δίνονται οι πιθανότητες: $P(A) = 0.01$, $P(\Theta / A) = 0.98$ και $P(\Theta / A') = 0.05$

α) Ζητείται η πιθανότητα $P(\Theta)$. Από το θεώρημα ολικής πιθανότητας έχουμε:

$$P(\Theta) = P(\Theta / A) \cdot P(A) + P(\Theta / A') \cdot P(A') = 0.98 \cdot 0.01 + 0.05 \cdot (1 - 0.01) = 0.0593$$

β) Ζητείται η πιθανότητα $P(A / \Theta)$. Από τον τύπο του Bayes έχουμε:

$$P(A / \Theta) = \frac{P(\Theta / A) P(A)}{P(\Theta)} = \frac{0.98 \cdot 0.01}{0.0593} = 0.1653.$$

5ο Θέμα

α) Έστω p_1, p_2, p_3, p_4, η πιθανότητα, αντίστοιχα, ένα μήλο παραγωγής του συγκεκριμένου παραγωγού να κατατάσσεται στην κατηγορία Α, Β, Γ, Δ. Σε επίπεδο σημαντικότητας $\alpha = 0.01$, θα ελέγξουμε τη μηδενική υπόθεση,

$H_0 : p_1 = 0.20$ και $p_2 = 0.40$ και $p_3 = 0.30$ και $p_4 = 0.10$

έναντι της εναλλακτικής,

$H_1 : p_1 \neq 0.20$ ή $p_2 \neq 0.40$ ή $p_3 \neq 0.30$ ή $p_4 \neq 0.10$.

Επειδή $600 \cdot p_{i0} \geq 5$ για κάθε $i = 1,2,3,4$, θα κάνουμε X^2 έλεγχο καλής προσαρμογής. Με την υπόθεση ότι η H_0 είναι αληθής, θα υπολογίσουμε για κάθε κατηγορία κατάταξης την αναμενόμενη (θεωρητική) συχνότητα, $E_i = 600 \cdot p_{i0}$, $i = 1,2,3,4$.

	A	B	Γ	Δ	Σύνολο
O_i	150	290	140	20	600
p_i	0.20	0.40	0.30	0.10	1.00
$E_i = 600 \cdot p_{i0}$	120	240	180	60	600

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

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

$$X^2 = \frac{(150-120)^2}{120} + \frac{(290-240)^2}{240} + \frac{(140-180)^2}{180} + \frac{(20-60)^2}{60} = 53.48 > 11.34\,,$$

η μηδενική υπόθεση, σε επίπεδο σημαντικότητας 1% απορρίπτεται και επομένως τα συγκεκριμένα δεδομένα, σε επίπεδο σημαντικότητας 1%, δίνουν στατιστικά σημαντικές αποδείξεις ότι τα ποσοστά κατάταξης των μήλων της παραγωγής του συγκεκριμένου παραγωγού διαφέρουν από τα ποσοστά κατάταξης που θεωρείται ότι ισχύουν για τα μήλα που παράγονται σε όλο το οροπέδιο.