Γραπτή Εξέταση Περιόδου Ιουνίου 2008
στο Μάθημα Στατιστική

4/07/08

1. Η πιθανότητα να υπάρχει στο υπέδαφος μιας συγκεκριμένης περιοχής εκμεταλλεύσιμο κοίτασμα πετρελαίου είναι 50%. Μια εταιρεία, που πρόκειται να κάνει σχετική έρευνα στην περιοχή, προβληματίζεται αν θα εφαρμόσει, μεταξύ άλλων, και μια δαπανηρή μέθοδο της οποίας όμως το αποτέλεσμα δεν είναι πάντα σωστό. Αν πράγματι υπάρχει εκμεταλλεύσιμο κοίτασμα, η μέθοδος αυτή δίνει ευοίωνο αποτέλεσμα (δηλαδή ανιχνεύει το κοίτασμα) με πιθανότητα 80% ενώ αν δεν υπάρχει, δίνει ευοίωνο αποτέλεσμα (δηλαδή αστοχεί) με πιθανότητα 30%. **α)** Αν η μέθοδος εφαρμοσθεί στην περιοχή, ποια είναι η πιθανότητα να δώσει ευοίωνο αποτέλεσμα; **β)** Αν η μέθοδος εφαρμοσθεί στην περιοχή και δώσει ευοίωνο αποτέλεσμα, ποια είναι η πιθανότητα να υπάρχει κοίτασμα στην περιοχή; **γ)** Να συγκρίνετε την εκ των προτέρων πιθανότητα να υπάρχει κοίτασμα στην περιοχή με την αντίστοιχη εκ των υστέρων πιθανότητα που υπολογίσατε στο ερώτημα (β) και να σχολιάσετε ως προς την ανεξαρτησία τα σχετικά ενδεχόμενα.

(15 Μονάδες)

2. Σε ένα άρθρο διατυπώνεται ο ισχυρισμός ότι η διάμεσος του ετήσιου εισοδήματος των αγροτών σε μια αγροτική περιοχή είναι 22.000€. Αν δεχθούμε τον ισχυρισμό αυτό: **α)** Ποια είναι η πιθανότητα από 5 αγρότες (που επιλέξαμε τυχαία) το πολύ δύο να έχουν ετήσιο εισόδημα μικρότερο των 22.000€; **β)** Ποια είναι η πιθανότητα από 100 αγρότες (που επιλέξαμε τυχαία), οι 37 ή λιγότεροι να έχουν ετήσιο εισόδημα μικρότερο των 22.000€; **γ)** Αν από 100 τυχαία επιλεγμένους αγρότες, βρεθούν 37 να έχουν ετήσιο εισόδημα μικρότερο των 22.000€, είναι δικαιολογημένος ο ισχυρισμός του αρθρογράφου;

(20 Μονάδες)

3. Ο αριθμός των προσκλήσεων ενός κτηνιάτρου από μια κτηνοτροφική μονάδα ανά μήνα είναι διακριτή τυχαία μεταβλητή X με συνάρτηση πιθανότητας:

x	0	1	2	3	4	5
$f(x)$	0.1	0.4	0.2	c	0.1	0.05

α) Βρείτε την πιθανότητα ο κτηνίατρος σε ένα μήνα να προσκληθεί από την κτηνοτροφική μονάδα 3 φορές ακριβώς. **β)** Βρείτε τον αναμενόμενο αριθμό προσκλήσεων του κτηνιάτρου από την μονάδα ανά μήνα. **γ)** Βρείτε την πιθανότητα η τυχαία μεταβλητή X να πάρει τιμή στο διάστημα $\mu \pm 2\cdot\sigma$.

(15 Μονάδες)

4. Προκειμένου να μετρηθεί η περιεκτικότητα κάποιας ουσίας στα νερά ενός ποταμού, πάρθηκαν 81 υδάτινα δείγματα από τον ποταμό. Η μέση περιεκτικότητα της ουσίας στο δείγμα των 81 μετρήσεων ήταν 50 milligram ανά λίτρο με τυπική απόκλιση 5 milligram ανά λίτρο. Για να συγκριθεί η περιεκτικότητα της ουσίας αυτής στον ποταμό με την περιεκτικότητα της ίδιας ουσίας σε έναν παραπόταμό του, πάρθηκαν και 100 δείγματα νερού από τον παραπόταμο. Η μέση περιεκτικότητα στις 100 μετρήσεις του παραποτάμου βρέθηκε να είναι 55.3 milligram ανά λίτρο με τυπική απόκλιση 4 milligram ανά λίτρο.

(α) Σε επίπεδο σημαντικότητας 1%, αποδεικνύουν τα δεδομένα αυτά ότι η μέση συγκέντρωση της ουσίας στον παραπόταμο είναι αυξημένη σε σχέση με τον ίδιο τον ποταμό;

(β) Σε επίπεδο σημαντικότητας 5%, αποδεικνύουν τα δεδομένα αυτά ότι η μέση συγκέντρωση της ουσίας στον παραπόταμο είναι αυξημένη περισσότερο από 3 milligram ανά λίτρο, σε σχέση με τον ίδιο τον ποταμό;

(γ) Δώστε 98% διάστημα εμπιστοσύνης για τη μέση συγκέντρωση της ουσίας στον κύριο ποταμό.
(δ) Δώστε 95% διάστημα εμπιστοσύνης για τη διαφορά στην συγκέντρωση μεταξύ παραποτάμου και ποταμού.

(20 Μονάδες)

5. Ένα φάρμακο, χρήσιμο σε ασθενείς που πάσχουν από υψηλή πίεση δίνεται πειραματικά σε 200 υπερτασικά άτομα με τα παρακάτω αποτελέσματα

ΑΠΟΤΕΛΕΣΜΑΤΑ	*Πλήθος ασθενών*
Α. Βαθμιαία ελάττωση της πίεσης	110
Β. Μέτρια ελάττωση της πίεσης	60
Γ. Μικρή ελάττωση της πίεσης	20
Δ. Μικρή αύξηση της πίεσης	10

Το φάρμακο αυτό συγκρίνεται με κάποιο άλλο που ήδη κυκλοφορεί και έχει την παρακάτω αποτελεσματικότητα για τις τέσσερες κατηγορίες Α:50%, Β:30%, Γ:10%, Δ:10%. Μπορούμε να ισχυριστούμε ότι τα δύο φάρμακα έχουν την ίδια αποτελεσματικότητα (α = 0.05);

(15 Μονάδες)

6. Κάποιο χημικό πείραμα έλαβε χώρα με 4 διαφορετικούς καταλύτες και σε 3 διαφορετικές θερμοκρασίες. Ελέγξτε σε επίπεδο σημαντικότητας α=0.05 εάν το αποτέλεσμα του πειράματος αλλάζει: **α)** διαφοροποιώντας τον καταλύτη (Α), **β)** διαφοροποιώντας τη θερμοκρασία (Β).

	Καταλύτης (Α)			
Θερμοκρασία (Β)	A_1	A_2	A_3	A_4
B_1	53	59	58	50
B_2	57	65	62	60
B_3	52	62	54	52

(Δίνονται: SSA=132, SSB=96, SST=252). **(15 Μονάδες)**

Δίνονται:

- Τιμές Φ(z) της συνάτρησης κατανομής της τυποποιημένης Κανονικής Κατανομής στο σημείο z.

Φ(0.5)=0.6915, Φ(0.75)=0.7734, Φ(1)=0.8413, Φ(1.33)=0.9083 Φ(1.5)=0.9332, Φ(1.59)=0.9441, Φ(2)=0.9772, Φ(2.5)=0.9938, Φ(2.6)=0.9953.

- Κριτικές τιμές Z_α της τυποποιημένης κανονικής κατανομής σε επίπεδο σημαντικότητας α

$Z_{0.005}=2.57$ $Z_{0.01}=2.33$ $Z_{0.02}=2.05$ $Z_{0.025}=1.96$ $Z_{0.05}=1.64$ $Z_{0.10}=1.28$

- Κριτικές τιμές X_κ^2 (α) της X^2 κατανομής με k βαθμούς ελευθερίας και για επίπεδο σημαντικότητας α.

$X_1^2(0.05) = 3.8$ $X_2^2(0.05) = 6.0$ $X_3^2(0.05) = 7.8$ $X_4^2(0.05) = 9.5$ $X_5^2(0.05) = 11.1$

- Κριτικές τιμές $t_k(\alpha)$ της t κατανομής με k βαθμούς ελευθερίας και για επίπεδο σημαντικότητας α.

$t_4(0.05) = 2.13$ $t_4(0.025) = 2.78$ $t_4(0.01) = 3.75$ $t_4(0.005) = 4.60$
$t_7(0.05) = 1.89$ $t_7(0.025) = 2.36$ $t_7(0.01) = 3.00$ $t_7(0.005) = 3.50$

- Κριτικές τιμές $F_{\mu,\nu}(\alpha)$ της F κατανομής με μ και ν βαθμούς ελευθερίας για επίπεδο σημαντικότητας α.

$F_{2,6}(0.05) = 5.14$ $F_{3,6}(0.05) = 4.76$ $F_{2,11}(0.05) = 3.98$ $F_{3,11}(0.05) = 3.59$

<u>Ενδεικτικές απαντήσεις</u>

1ο Θέμα

Ορίζουμε τα ενδεχόμενα, Ε: το αποτέλεσμα της μελέτης είναι ευοίωνο και Ο: υπάρχει εκμεταλλεύσιμο κοίτασμα στην περιοχή. Δίνονται οι πιθανότητες:

$P(O) = 0.5$, $P(E/O) = 0.8$, $P(E/O') = 0.3$.

α) Ζητάμε την πιθανότητα

$P(E) = P(E/O)P(O) + P(E/O')P(O') = 0.8 \cdot 0.5 + 0.3 \cdot 0.5 = 0.55$

β) Ζητάμε την εκ των υστέρων πιθανότητα

$$P(O/E) = \frac{P(E/O)P(O)}{P(E)} = \frac{0.8 \cdot 0.5}{0.55} = 0.73$$

γ) Τα ενδεχόμενα O και E δεν είναι ανεξάρτητα αφού $P(O/E) = 0.73 \neq P(O) = 0.5$. Με όρους του προβλήματος, αυτό σημαίνει ότι αν το αποτέλεσμα της μεθόδου είναι ευοίωνο η πιθανότητα να υπάρχει εκμεταλλεύσιμο κοίτασμα στην περιοχή επηρεάζεται και μάλιστα αυξάνει από 50% σε 73%.

2ο Θέμα

Σύμφωνα με τον ισχυρισμό του αρθρογράφου, το διάμεσο ετήσιο εισόδημα των αγροτών είναι 22.000€ και επομένως *P(εισόδημα μικρότερο των 22.000€)* = 0.5.

α) Έστω X ο αριθμός των αγροτών (από τους 5 που επελέγησαν τυχαία) που έχουν εισόδημα μικρότερο των 22.000€. Προφανώς $X \sim B(5,\ 0.5)$ και επομένως:

$$P(X \le 2) = P(X=0) + P(X=1) + P(X=2) = \binom{5}{0} \cdot 0.5^0 \cdot 0.5^5 + \binom{5}{1} \cdot 0.5^1 \cdot 0.5^4 + \binom{5}{2} \cdot 0.5^2 \cdot 0.5^3 = 0.5$$

β) Έστω Y ο αριθμός των αγροτών (από τους 100 που επελέγησαν τυχαία) που έχουν εισόδημα μικρότερο των 22.000€. Προφανώς $Y \sim B(100,\ 0.5)$. Επειδή $n = 100$ (μεγάλο) με $\nu \cdot p = 100 \cdot 0.5 = 50 \ge 5$, η Y προσεγγίζεται ικανοποιητικά από την κανονική κατανομή με $\mu = 100 \cdot 0.5 = 50$ και $\sigma^2 = 100 \cdot 0.5 \cdot (1 - 0.5) = 25$, δηλαδή, $Y \sim N(50,\ 5^2)$ και επομένως:

$$P(Y \le 37) = P\left(Z < \frac{37-50}{5}\right) = P(Z < -2.6) = \Phi(-2.6) = 1 - \Phi(2.6) = 1 - 0.9953 = 0.0047.$$

γ) Με την υπόθεση ότι το διάμεσο ετήσιο εισόδημα είναι 22.000€, στο ερώτημα (β), βρέθηκε ότι το ενδεχόμενο να εμφανισθούν 37 ή λιγότεροι αγρότες με εισόδημα μικρότερο των 22.000€ είναι σπάνιο να συμβεί. Εντούτοις βρέθηκαν 37, δηλαδή, το σπάνιο ενδεχόμενο συνέβη! Αυτό μας οδηγεί στο να αμφιβάλουμε για τον ισχυρισμό του αρθρογράφου!

3ο Θέμα

α) Πρέπει $f(0) + f(1) + f(2) + f(3) + f(4) + f(5) = 1 \Rightarrow f(3) = 1 - 0.85 = 0.15$. Άρα $P(X = 3) = f(3) = 0.1$.

β) $\mu = E(X) = 0 \cdot 0.1 + 1 \cdot 0.4 + 2 \cdot 0.2 + 3 \cdot 0.15 + 4 \cdot 0.1 + 5 \cdot 0.05 = 1.9$.

γ) $E(X^2) = 0 \cdot 0.1 + 1 \cdot 0.4 + 4 \cdot 0.2 + 9 \cdot 0.15 + 16 \cdot 0.1 + 25 \cdot 0.05 = 5.4$.

Άρα $\sigma^2 = E(X^2) - [E(X)]^2 = 5.4 - 1.9^2 = 1.79 \Rightarrow \sigma = 1.34$ και επομένως το διάστημα $\mu \pm 2 \cdot \sigma$ είναι το διάστημα $(-0.78,\ 4.58)$. Ζητείται η πιθανότητα $P(-0.78 < X < 4.58) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) = 0.95$

4ο Θέμα

α) Αν μ_1 ο άγνωστος πληθυσμιακός μέσος της περιεκτικότητας της ουσίας στον ποταμό και μ_2 ο άγνωστος πληθυσμιακός μέσος της περιεκτικότητας της ουσίας στον παραπόταμο, κάνουμε τον έλεγχο:

$H_0 : \mu_2 = \mu_1$

$H_1 : \mu_2 > \mu_1$

Δίνεται ότι: $n_1 = 81,\ \bar{x}_1 = 50,\ s_1 = 5,\ n_2 = 100,\ \bar{x}_2 = 55.3,\ s_2 = 4.$ Επειδή οι πληθυσμιακές διασπορές $\sigma_1^2,\ \sigma_2^2$ είναι άγνωστες και $n_1, n_2 \geq 30$, η απορριπτική περιοχή ορίζεται από την ανισότητα:

$$t = \frac{\bar{x}_2 - \bar{x}_1}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} > z_\alpha .$$

Με αντικατάσταση βρίσκουμε $t \approx 7.68$ και επειδή $z_{0.01} = 2.33$ η μηδενική υπόθεση απορρίπτεται και επομένως, σε επίπεδο σημαντικότητας 1%, μπορούμε να ισχυρισθούμε ότι η μέση συγκέντρωση της ουσίας στον παραπόταμο είναι αυξημένη σε σχέση με τη μέση συγκέντρωση της ουσίας στον ποταμό.

β) Κάνουμε τον έλεγχο:

$H_0 : \mu_2 - \mu_1 = 3$

$H_1 : \mu_2 - \mu_1 > 3$

Δίνεται ότι: $n_1 = 81,\ \bar{x}_1 = 50,\ s_1 = 5,\ n_2 = 100,\ \bar{x}_2 = 55.3,\ s_2 = 4.$ Επειδή οι πληθυσμιακές διασπορές $\sigma_1^2,\ \sigma_2^2$ είναι άγνωστες και $n_1, n_2 \geq 30$, η απορριπτική περιοχή ορίζεται από την ανισότητα:

$$t = \frac{\bar{x}_2 - \bar{x}_1 - 3}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} > z_\alpha .$$

Με αντικατάσταση βρίσκουμε $t \approx 3.3$ και επειδή $z_{0.05} = 1.64$ η μηδενική υπόθεση απορρίπτεται και επομένως σε επίπεδο σημαντικότητας 5% μπορούμε να ισχυρισθούμε ότι η μέση συγκέντρωση της ουσίας στον παραπόταμο είναι αυξημένη σε σχέση με τη μέση συγκέντρωση της ουσίας στον ποταμό περισσότερο από 3 milligram ανά λίτρο.

γ) Επειδή η πληθυσμιακή διασπορά σ_1^2 είναι άγνωστη και $n_1 = 81 \geq 30$ το ζητούμενο 98% διάστημα εμπιστοσύνης είναι $\bar{x}_1 \pm \frac{s_1}{\sqrt{n_1}} z_{\alpha/2}$ ή $50 \pm \frac{5}{\sqrt{81}} z_{0.01}$, με $z_{0.01} = 2.33$ ή 50 ± 1.29

δ) Επειδή οι πληθυσμιακές διασπορές $\sigma_1^2,\ \sigma_2^2$ είναι άγνωστες και $n_1, n_2 \geq 30$ το ζητούμενο 95% διάστημα εμπιστοσύνης είναι $\bar{x}_2 - \bar{x}_1 \pm z_{\alpha/2}\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$ ή $5.3 \pm z_{0.025}\sqrt{\frac{25}{81} + \frac{16}{100}}$ με $z_{0.025} = 1.96$ ή 5.3 ± 1.35

5ο Θέμα

Πρόκειται για n **πολυωνυμικές δοκιμές** με k δυνατά αποτελέσματα. Ειδικότερα, πρόκειται για $n = 200$ ανεξάρτητες δοκιμές με $k = 4$ δυνατά αποτελέσματα. Κάθε αποτέλεσμα ταξινομείται σε ακριβώς μία από τις τέσσερις κατηγορίες **Α, Β, Γ, Δ,** με πιθανότητα p_1, p_2, p_3, p_4 αντίστοιχα. Θα κάνουμε X^2 έλεγχο **καλής προσαρμογής**. Ειδικότερα, ελέγχουμε τη μηδενική υπόθεση

$H_0 : p_1 = 0.5,\ p_2 = 0.3,\ p_3 = 0.1,\ p_4 = 0.1$ έναντι της εναλλακτικής

$H_1 :\ p_1 \neq 0.5 \text{ ή } p_2 \neq 0.3 \text{ ή } p_3 \neq 0.1 \text{ ή } p_4 \neq 0.1$.

Η απορριπτική περιοχή ορίζεται από την ανισότητα

$X^2 = \sum_{i=1}^{4} \frac{(\pi_i - \theta_i)^2}{\theta_i} > X^2_{3;0.05} = 7.8$, όπου π_i οι παρατηρούμενες συχνότητες και θ_i οι θεωρητικές. Δημιουργούμε τον πίνακα,

	Κατηγορίες				**Σύνολα**
	Α	**Β**	**Γ**	**Δ**	
π_i	110	60	20	10	**200**
$\theta_i = n \cdot p_i = 200 \cdot p_i$	100	60	20	20	**200**

και υπολογίζουμε το στατιστικό

$$X^2 = \sum_{i=1}^{4} \frac{(\pi_i - \theta_i)^2}{\theta_i} = \frac{(110-100)^2}{100} + \frac{(60-60)^2}{60} + \frac{(20-20)^2}{20} + \frac{(10-20)^2}{20} = 6.$$

Επειδή $\theta_i \geq 5, i = 1,2,3,4$ και $X^2 = 6 < X^2_{3;0.05} = 7.8$, τα πειραματικά δεδομένα, σε επίπεδο σημαντικότητας 5%, δεν υποστηρίζουν απόρριψη της μηδενικής υπόθεσης και επομένως μπορούμε να ισχυρισθούμε ότι τα δύο φάρμακα έχουν την ίδια αποτελεσματικότητα.

6ο Θέμα

Πρόκειται για πρόβλημα ανάλυσης διασποράς για δύο παράγοντες (Α: ο καταλύτης και Β: η θερμοκρασία). Υποθέτουμε ότι κάθε αποτέλεσμα του πειράματος είναι μία παρατήρηση, δηλαδή ένα δείγμα μεγέθους 1, από έναν πληθυσμό που ακολουθεί κανονική κατανομή. Υποθέτουμε δηλαδή ότι για κάθε συνδυασμό καταλύτη-θερμοκρασίας, ο αντίστοιχος πληθυσμός αποτελεσμάτων με αυτόν τον καταλύτη και αυτή τη θερμοκρασία ακολουθεί κανονική κατανομή. Υποθέτουμε επίσης ότι όλοι αυτοί οι πληθυσμοί έχουν ίδια διασπορά και ότι τα δείγματα από κελί σε κελί είναι ανεξάρτητα. Με την προϋπόθεση ότι ικανοποιούνται οι παραπάνω υποθέσεις, θα κάνουμε τους ελέγχους:

i) $H_{0a} : \alpha_1 = \alpha_2 = \alpha_3 = \alpha_4$ (ο παράγοντας Α – ο καταλύτης - δεν επιδρά στο αποτέλεσμα του πειράματος, δηλαδή, το αποτέλεσμα είναι ίδιο για τους τέσσερις καταλύτες)

$H_{1a} :$ τα $\alpha_i, i = 1,2,3,4$ δεν είναι όλα ίσα (ο παράγοντας Α επιδρά στο αποτέλεσμα του πειράματος, δηλαδή, το αποτέλεσμα δεν είναι ίδιο για τους τέσσερις καταλύτες)

ii) $H_{0\beta} : \beta_1 = \beta_2 = \beta_3$ (ο παράγοντας Β –η θερμοκρασία- δεν επιδρά στο αποτέλεσμα του πειράματος, δηλαδή, το αποτέλεσμα είναι ίδιο για τις τρεις θερμοκρασίες)

$H_{1\beta} :$ τα $\beta_i, i = 1,2,3$ δεν είναι όλα ίσα (ο παράγοντας Β επιδρά στο αποτέλεσμα του πειράματος, δηλαδή, το αποτέλεσμα δεν είναι ίδιο για τις τρεις θερμοκρασίες)

Ο πίνακας ανάλυσης διασποράς είναι ο παρακάτω:

Πηγή Μεταβολής	Άθροισμα Τετραγώνων	Βαθμοί Ελευθερίας	Μέσα Τετράγωνα	F
Παράγοντας Α (Καταλύτης)	*SSA=132*	*k-1=3*	*MSA=132/3=44*	$F_A = \frac{MSA}{MSE} = 11$
Παράγοντας Β (Θερμοκρασία)	*SSB=96*	*λ-1=2*	*MSB=96/2=48*	$F_B = \frac{MSB}{MSE} = 12$
Τυχαία σφάλματα	*SSE=24*	*(k-1)(λ-1)=6*	*MSE=24/6=4*	
Σύνολο	*SST=252*	*kλ-1=11*		

Το *SSE* υπολογίσθηκε από τη σχέση *SSA+SSB+ SSE=SST*.

Από τις τιμές που υπολογίσθηκαν στον πίνακα, έχουμε:

i) Η μηδενική υπόθεση H_{0a} σε επίπεδο σημαντικότητας 5% απορρίπτεται διότι $F_A = 11 > F_{3,6,0.05} = 4.76$ και επομένως με βάση τα πειραματικά δεδομένα, σε επίπεδο σημαντικότητας 5%, ο καταλύτης επιδρά στο αποτέλεσμα του πειράματος, δηλαδή τα πειραματικά δεδομένα δείχνουν ότι το αποτέλεσμα του πειράματος δεν είναι ίδιο για τους τέσσερις καταλύτες.

ii) Η μηδενική υπόθεση $H_{0\beta}$ σε επίπεδο σημαντικότητας 5% απορρίπτεται διότι $F_B = 12 > F_{2,6,0.05} = 5.14$ και επομένως με βάση τα πειραματικά δεδομένα, σε επίπεδο σημαντικότητας 5%, η θερμοκρασία επιδρά στο αποτέλεσμα του πειράματος, δηλαδή, τα πειραματικά δεδομένα δείχνουν ότι το αποτέλεσμα δεν είναι ίδιο για τις τρεις θερμοκρασίες.