ΤΟΕ – στατιστική 2 – 2024 Σεπτέμβριος Κυριακοπούλου

1. (α). [4 Μονάδες] Ένα εργοστάσιο παράγει ένα προϊόν χρησιμοποιώντας δύο μηχανές. Υποθέτουμε ότι το βάρος των παραγόμενων τεμαχίων και από τις δύο μηχανές ακολουθεί την κανονική κατανομή, όχι κατ' ανάγκη ίδια για κάθε μηχανή. Από τα προϊόντα κάθε μηχανής επιλέγουμε τυχαία από 100 τεμάχια και υπολογίζουμε αριθμητικούς μέσους και δειγματικές διακυμάνσεις για τα βάρη των τεμαχίων. Βρίσκουμε έτσι αριθμητικούς μέσους 121 kg και 117 kg και δειγματικές διακυμάνσεις 39 και 25 για τεμάχια από την πρώτη και τη δεύτερη μηχανή, αντίστοιχα.

   (i). Μπορούμε να δεχθούμε σε επίπεδο σημαντικότητας 5% ότι οι διακυμάνσεις των βαρών είναι ίσες και για τις δύο μηχανές ή υπάρχει μεγαλύτερη διασπορά βάρους στα προϊόντα της πρώτης μηχανής;

   (ii). Υπολογίστε ένα διάστημα εμπιστοσύνης για τη διαφορά των πραγματικών μέσων τιμών βάρους για τα προϊόντα των δύο μηχανών με πιθανότητα 95%.

   (iii). Υπολογίστε, με βάση τα παραπάνω δεδομένα, ένα διάστημα εμπιστοσύνης για το λόγο των διακυμάνσεων των δύο πληθυσμών με πιθανότητα 90%.

   (iv). Αν στο δείγμα από την πρώτη μηχανή βρέθηκαν 12 ελαττωματικά τεμάχια και στο δείγμα από τη δεύτερη μηχανή βρέθηκαν 8 ελαττωματικά, να ελεγχθεί, σε επίπεδο σημαντικότητας 5%, η μηδενική υπόθεση ότι η αναλογία των ελαττωματικών είναι ίδια και στους δύο πληθυσμούς έναντι της εναλλακτικής ότι στην πρώτη μηχανή είναι μεγαλύτερη από εκείνη της δεύτερης.
   Δίνονται: $f^U_{99;99;0.05} = 1.394$, $f^U_{99;99;0.025} = 1.486$, $z_{0.025} = 1.96$ και $z_{0.05} = 1.645$.

   (β). [1 Μονάδα] Τοποθετείστε σε αύξουσα σειρά (από το μικρότερο προς το μεγαλύτερο) τα εκατοστημόρια:

$$\chi^2_{27;0.51},\ t_{27;0.56},\ z_{0.56}$$

   και δικαιολογήστε την απάντησή σας, χρησιμοποιώντας επίσης και τα απαραίτητα γραφήματα.

2. (α). [2 Μονάδες] Έστω x μια παρατήρηση της συνεχούς τυχαίας μεταβλητής X. Θέλουμε να ελέγξουμε την μηδενική υπόθεση $\mathrm{H}_0 : X$ ακολουθεί κατανομή $U(0,3)$ με $f(x) = \frac{1}{3}$ έναντι της εναλλακτικής $\mathrm{H}_1 : X$ ακολουθεί κατανομή με $f(x) = \frac{2x}{9}$, $0 \le x \le 3$, έτσι ώστε αν η H_1 είναι αληθής, η πιθανότητα να δεχθούμε κατά λάθος την H_0 να είναι ίση με 0.05. [hint: πώς καλείται αυτή η πιθανότητα;]

   (i). Με βάση τα παραπάνω δεδομένα, να βρεθεί η τιμή x_0 έτσι ώστε αν $x < x_0$ να δεχθούμε την H_0.

   (ii). Να βρεθεί το επίπεδο σημαντικότητας α του στατιστικού ελέγχου.

   (iii). Εάν η τυχαία μεταβλητή X ακολουθεί κανονική κατανομή $N(2, 0.36)$, ποια είναι η πιθανότητα να βρίσκεται ανάμεσα σε 0.8 και 2.3;
   Δίνονται: $G(0.5) = 0.6915$ και $G(2) = 0.972$.

   (β)[3 Μονάδες] Ο αριθμός των ελαττωματικών τεμαχίων, που παράγονται από τρεις διαφορετικές μηχανές, που χειρίζονται τέσσερις διαφορετικοί εργάτες, δίνεται στον παρακάτω πίνακα:

	Εργάτης 1	Εργάτης 2	Εργάτης 3	Εργάτης 4	Σύνολο
Μηχανή 1	35	38	41	32	146
Μηχανή 2	31	40	38	31	140
Μηχανή 3	36	35	43	25	139
Σύνολο	102	113	122	88	425

   Να ελεγχθεί, σε επίπεδο σημαντικότητας $\alpha = 0.05$, αν υπάρχει στατιστικά σημαντική επίδραση των παραγόντων αυτών στην παραγωγή ελαττωματικών τεμαχίων.
   Δίνονται: $f_{2;6;0.05} = 5.14$ και $f_{3;6;0.05} = 4.76$.

3. (α). [2 Μονάδες] Από πληθυσμό που ακολουθεί κατανομή με συνάρτηση πυκνότητας πιθανότητας:

$$f(x) = \begin{cases} \alpha x^{\alpha-1}, & 0 < x < 1, \alpha > 0 \\ 0, & x \leq 0 \text{ ή } x \geq 1 \end{cases}$$

λαμβάνουμε τυχαίο δείγμα $\{X_1, X_2, ..., X_n\}$. Να βρεθεί η εκτιμήτρια της παραμέτρου α με τη μέθοδο της μέγιστης πιθανοφάνειας.

(β). [3 Μονάδες] Ο παρακάτω πίνακας περιέχει τις δαπάνες διαφήμισης (ως ποσοστό των συνολικών δαπανών) και τα καθαρά κέρδη (ως ποσοστό των πωλήσεων) σε ένα τυχαίο δείγμα 12 καταστημάτων αθλητικών ειδών. Βεβαιωθείτε ότι το υπόδειγμα που χρησιμοποιείτε είναι έγκυρο με τη χρήση της στατιστικής ελέγχου T και στη συνέχεια υπολογίστε τον συντελεστή προσδιορισμού και πείτε ποια είναι η ερμηνεία του με βάση το αποτέλεσμα.

διαφήμιση (X)	1.5	0.8	2.6	1.0	0.6	2.8	1.2	0.9	0.4	1.3	1.2	2.0
κέρδος (Y)	3.1	1.9	4.2	2.3	1.2	4.9	2.8	2.1	1.4	2.4	2.4	3.8

Δίνονται επίσης: $t_{8;0.025} = 2.306$ και

$$\sum_{i=1}^{12} X_i = 16.3, \quad \sum_{i=1}^{12} X_i^2 = 28.39$$

$$\sum_{i=1}^{12} Y_i = 32.5, \quad \sum_{i=1}^{12} Y_i^2 = 101.77$$

$$\sum_{i=1}^{12} X_i Y_i = 53.24$$

Λύσεις

(α). [4 Μονάδες] Ένα εργοστάσιο παράγει ένα προϊόν χρησιμοποιώντας δύο μηχανές. Υποθέτουμε ότι το βάρος των παραγόμενων τεμαχίων και από τις δύο μηχανές ακολουθεί την κανονική κατανομή, όχι κατ' ανάγκη ίδια για κάθε μηχανή. Από τα προϊόντα κάθε μηχανής επιλέγουμε τυχαία από 100 τεμάχια και υπολογίζουμε αριθμητικούς μέσους και δειγματικές διακυμάνσεις για τα βάρη των τεμαχίων. Βρίσκουμε έτσι αριθμητικούς μέσους 121 kg και 117 kg και δειγματικές διακυμάνσεις 39 και 25 για τεμάχια από την πρώτη και τη δεύτερη μηχανή, αντίστοιχα.

(i). Μπορούμε να δεχθούμε σε επίπεδο σημαντικότητας 5% ότι οι διακυμάνσεις των βαρών είναι ίσες και για τις δύο μηχανές ή υπάρχει μεγαλύτερη διασπορά βάρους στα προϊόντα της πρώτης μηχανής;

$$n_1 = 100, \quad \bar{x}_1 = 121, \quad s_1^2 = 39$$
$$n_2 = 100, \quad \bar{x}_2 = 117, \quad s_2^2 = 25$$

i) F ελεγχος λόγου διακυμάνσεων (για ισότητα διακυμάνσεων) με α=0,05

$$H_0: \sigma_1^2 = \sigma_2^2$$
$$H_1: \sigma_1^2 > \sigma_2^2$$

$$F = \frac{s_1^2}{s_2^2} = \frac{39}{25} = 1{,}56 \ \boxed{>} \ F^U_{99,99,0.05} = 1{,}394$$

Απορρίπτεται η Ho συνεπώς στατιστικά είναι μεγαλύτερη η διασπορά βάρους στα προιόντα της πρώτης μηχανής

(ii). Υπολογίστε ένα διάστημα εμπιστοσύνης για τη διαφορά των πραγματικών μέσων τιμών βάρους για τα προϊόντα των δύο μηχανών με πιθανότητα 95%. $\to \alpha = 0{,}05$

$$\left.\begin{matrix} n_1 = 100 \geq 30 \\ n_2 = 100 \geq 30 \end{matrix}\right\} \text{μεγαλο δέιγμα}$$

95% Δ. Ε.
$$\bar{x}_1 - \bar{x}_2 \pm z_{\alpha/2} \cdot \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$
$$121 - 117 \pm z_{0.025} \cdot \sqrt{\frac{39}{100} + \frac{25}{100}}$$
$$4 \pm 1{,}96 \cdot 0{,}8$$
$$4 \pm 1{,}568$$
$$(4 - 1{,}568 \ , \ 4 + 1{,}568)$$
$$(2{,}432 \ , \ 5{,}568)$$

(iii). Υπολογίστε, με βάση τα παραπάνω δεδομένα, ένα διάστημα εμπιστοσύνης για το λόγο των διακυμάσεων των δύο πληθυσμών με πιθανότητα 90%. $\to \alpha = 0{,}1$

90% Δ. Ε.
$$\left(\frac{s_1^2}{s_2^2} \cdot F^L_{99,99,0.05} \ , \ \frac{s_1^2}{s_2^2} F^U_{99,99,0.05}\right)$$
$$\left(\frac{39}{25} \cdot \frac{1}{1{,}394} \ , \ \frac{39}{25} \cdot 1{,}394\right)$$
$$(1{,}12 \ , \ 2{,}17)$$

(iv). Αν στο δείγμα από την πρώτη μηχανή βρέθηκαν 12 ελαττωματικά τεμάχια και στο δείγμα από τη δεύτερη μηχανή βρέθηκαν 8 ελαττωματικά, να ελεγχθεί, σε επίπεδο σημαντικότητας 5%, η μηδενική υπόθεση ότι η αναλογία των ελαττωματικών είναι ίδια και στους δύο πληθυσμούς έναντι της εναλλακτικής ότι στην πρώτη μηχανή είναι μεγαλύτερη από εκείνη της δεύτερης.
Δίνονται: $f^U_{99;99;0.05} = 1.394$, $f^U_{99;99;0.025} = 1.486$, $z_{0.025} = 1.96$ και $z_{0.05} = 1.645$.

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{12}{100} = 0,12$$

$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{8}{100} = 0,08$$

$$\hat{p} = \frac{12+8}{100+100} = 0,1$$

$$H_0: p_1 - p_2 = 0$$

$$H_1: p_1 - p_2 > 0$$

$$z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0,12 - 0,08}{\sqrt{0,1 \cdot 0,9 \left(\frac{1}{100}+\frac{1}{100}\right)}} = 0,94$$

$$z = 0,94 \boxed{<} z_{a} = z_{0.05} = 1,645$$

Δεν απορρίπτεται η Ho συνεπώς το ποσοστό ελλατωματικών της μηχανής 1 είναι στατιστικά μεγαλύτερο απο της μηχανής 2

(β). [1 Μονάδα] Τοποθετείστε σε αύξουσα σειρά (από το μικρότερο προς το μεγαλύτερο) τα εκατοστημόρια:

$$\chi^2_{27;0.51},\ t_{27;0.56},\ z_{0.56}$$

και δικαιολογήστε την απάντησή σας, χρησιμοποιώντας επίσης και τα απαραίτητα γραφήματα.

$$\underset{<0}{t_{27,0.56}} < \underset{<0}{z_{0.56}} < \underset{>0}{\chi^2_{27,0.51}}$$

2. (α). [2 Μονάδες] Έστω x μία παρατήρηση της συνεχούς τυχαίας μεταβλητής X. Θέλουμε να ελέγξουμε την μηδενική υπόθεση H_0 : X ακολουθεί κατανομή $U(0,3)$ με $f(x) = \frac{1}{3}$ έναντι της εναλλακτικής H_1 : X ακολουθεί κατανομή με $f(x) = \frac{2x}{9}$, $0 \le x \le 3$, έτσι ώστε αν η H_1 είναι αληθής, η πιθανότητα να δεχθούμε κατά λάθος την H_0 να είναι ίση με 0.05. [hint: πώς καλείται αυτή η πιθανότητα;]

(i). Με βάση τα παραπάνω δεδομένα, να βρεθεί η τιμή x_0 έτσι ώστε αν $x < x_0$ να δεχθούμε την H_0.

(ii). Να βρεθεί το επίπεδο σημαντικότητας α του στατιστικού ελέγχου.

(iii). Εάν η τυχαία μεταβλητή X ακολουθεί κανονική κατανομή $N(2, 0.36)$, ποια είναι η πιθανότητα να βρίσκεται ανάμεσα σε 0.8 και 2.3;
Δίνονται: $G(0.5) = 0.6915$ και $G(2) = 0.972$.

$$P(H_0 \mid H_1) = 0{,}05$$

$$P\left(x < x_0 \mid f(x) = \frac{2x}{9}\right) = 0{,}05$$

$$\int_0^{x_0} \frac{2x}{9}\,dx = 0{,}05 \rightarrow \left.\frac{x^2}{2}\right]_0^{x_0} = 0{,}225$$

$$\rightarrow \frac{x_0^2}{2} = 0{,}225 \rightarrow x_0 = \pm\sqrt{0{,}45}$$

το αρνητικό απορρίπτεται καθώς είναι εκτος πεδίου ορισμού άρα

$$x_0 = 0{,}67$$

ii) επίπεδο σημαντικότητας α

$$P(H_1 \mid H_0) = a$$

$$P\left(x > 0{,}67 \,/\, f(x) = \frac{1}{3}\right) = \int_{0{,}67}^{3} \frac{1}{3}\,dx =$$

$$= \left.\frac{1}{3}x\right]_{0{,}67}^{3} = \frac{1}{3}(3 - 0{,}67) = 0{,}777$$

άρα $a = 0{,}777$

iii) $P(0,8 < X < 2,3) = P\left(\frac{0,8-2}{\sqrt{0,36}} < Z < \frac{2,3-2}{\sqrt{0,36}}\right) =$

$= P(-2 < Z < 0,5) = \Phi(0,5) - \Phi(-2) =$

$= \Phi(0,5) - (1 - \Phi(2)) = \Phi(0,5) + \Phi(2) - 1$

$= 0,6915 + 0,972 - 1 = 0,6635$

(β)[3 Μονάδες] Ο αριθμός των ελαττωματικών τεμαχίων, που παράγονται από τρεις διαφορετικές μηχανές, που χειρίζονται τέσσερις διαφορετικοί εργάτες, δίνεται στον παρακάτω πίνακα:

	Εργάτης 1	Εργάτης 2	Εργάτης 3	Εργάτης 4	Σύνολο
Μηχανή 1	35	38	41	32	146
Μηχανή 2	31	40	38	31	140
Μηχανή 3	36	35	43	25	139
Σύνολο	102	113	122	88	425

Να ελεγχθεί, σε επίπεδο σημαντικότητας $\alpha = 0.05$, αν υπάρχει στατιστικά σημαντική επίδραση των παραγόντων αυτών στην παραγωγή ελαττωματικών τεμαχίων.
Δίνονται: $f_{2;6;0.05} = 5.14$ και $f_{3;6;0.05} = 4.76$.

Ανάλυση Διακύμανσης κατα 2 παράγοντες $\quad k = 3, r = 4$

$H_0^A: \mu_1 = \mu_2 = \mu_3$

H_1^A: τουλαχιστον ένα μi διαφέρει i=1,2,3

$H_0^B: \mu_1 = \mu_2 = \mu_3 = \mu_4$

H_1^B: τουλάχιστον ένα μj διαφέρει j=1,2,3,4

$$MO\Delta = \sum\sum x_{ij}^2 - \frac{T^2}{kr} = (35^2+38^2+\ldots+25^2) - \frac{425^2}{3\cdot 4}$$

$$= 15.335 - 15.052{,}08 = 282{,}91$$

$$MM\Delta_A = \frac{\sum T_{i.}^2}{r} - \frac{T^2}{kr} = \frac{146^2+140^2+139^2}{4} - 15.052{,}08 = 7{,}17$$

$$MM\Delta_B = \frac{\sum T_{.j}^2}{k} - \frac{T^2}{kr} = \frac{102^2+113^2+122^2+88^2}{3} - 15.052{,}08 = 214{,}92$$

$$ME\Delta = 282{,}91 - 7{,}17 - 214{,}92 = 60{,}83$$

$$F^A = \frac{MM\Delta_A/(k-1)}{ME\Delta/(k-1)(r-1)} = \frac{7{,}17/(3-1)}{60{,}83/(3-1)(4-1)} = 0{,}35 < F_{2,6,0.05} = 5{,}14$$

άρα δεν απορρίπτεται η H_0^A συνεπώς δεν υπάρχει διαφορά μεταξύ των μηχανών

$$F^B = \frac{MM\Delta_B/(r-1)}{ME\Delta/(k-1)(r-1)} = \frac{214{,}92/(4-1)}{60{,}83/(3-1)(4-1)} = 7{,}06 > F_{3,6,0.05} = 4{,}76$$

άρα απορρίπτεται η H_0^B συνεπώς υπάρχει διαφορά μεταξύ των εργατών

3. (α). [2 Μονάδες] Από πληθυσμό που ακολουθεί κατανομή με συνάρτηση πυκνότητας πιθανότητας:

$$f(x) = \begin{cases} \alpha x^{\alpha-1}, & 0 < x < 1, \alpha > 0 \\ 0, & x \le 0 \text{ ή } x \ge 1 \end{cases}$$

λαμβάνουμε τυχαίο δείγμα $\{X_1, X_2, ..., X_n\}$. Να βρεθεί η εκτιμήτρια της παραμέτρου α με τη μέθοδο της μέγιστης πιθανοφάνειας.

πιθανοφάνεια

$$L = a \cdot x_1^{a-1} \cdot a \cdot x_2^{a-1} \cdot \ldots \cdot a \cdot x_n^{a-1} = a^n \cdot \prod x_i^{a-1}$$

λογαριθμική πιθανοφάνεια

$$\ln L = \ln a^n + \ln \prod x_i^{a-1} = n \ln a + (a-1) \sum \ln x_i$$

1.o.c. $\dfrac{\partial \ln L}{\partial a} = 0 \rightarrow \dfrac{n}{a} + \sum \ln x_i = 0 \rightarrow \boxed{\hat{a} = \dfrac{-n}{\sum \ln x_i}}$

S.o.c. $\dfrac{\partial^2 \ln L}{\partial a^2} = -\dfrac{n}{a^2} < 0$ αρα μέγιστο

(β). [3 Μονάδες] Ο παρακάτω πίνακας περιέχει τις δαπάνες διαφήμισης (ως ποσοστό των συνολικών δαπανών) και τα καθαρά κέρδη (ως ποσοστό των πωλήσεων) σε ένα τυχαίο δείγμα 12 καταστημάτων αθλητικών ειδών. Βεβαιωθείτε ότι το υπόδειγμα που χρησιμοποιείτε είναι έγκυρο με τη χρήση της στατιστικής ελέγχου T και στη συνέχεια υπολογίστε τον συντελεστή προσδιορισμού και πείτε ποιά είναι η ερμηνεία του με βάση το αποτέλεσμα.

διαφήμιση (X)	1.5	0.8	2.6	1.0	0.6	2.8	1.2	0.9	0.4	1.3	1.2	2.0
κέρδος (Y)	3.1	1.9	4.2	2.3	1.2	4.9	2.8	2.1	1.4	2.4	2.4	3.8

Δίνονται επίσης: $t_{8;0.025} = 2.306$ και

$$\sum_{i=1}^{12} X_i = 16.3, \quad \sum_{i=1}^{12} X_i^2 = 28.39$$

$$\sum_{i=1}^{12} Y_i = 32.5, \quad \sum_{i=1}^{12} Y_i^2 = 101.77$$

$$\sum_{i=1}^{12} X_i Y_i = 53.24$$

$$b_1 = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sum x_i^2 - n\bar{x}^2} = \frac{53,24 - 12\cdot\left(\frac{16,3}{12}\right)\cdot\left(\frac{32,5}{12}\right)}{28,39 - 12\cdot\left(\frac{16,3}{12}\right)^2}$$

$$= \frac{9,09}{6,25} = 1,45$$

$$b_0 = \bar{y} - b_1\bar{x} = \frac{32,5}{12} - 1,45\left(\frac{16,3}{12}\right)$$

$$b_0 = 0,74$$

αρα $\hat{y}_i = 0,74 + 1,45\, x_i$

t έλεγχος σημαντικότητας του συντελεστή κλίσης β1

$H_0: b_1 = 0$ (ΟΧΙ)

$H_1: b_1 \neq 0$ (ΝΑΙ)

$$t = \frac{\hat{b}_1 - 0}{s_{b_1}} \quad (1)$$

$$s_{b_1} = \sqrt{\frac{s_\varepsilon^2}{\sum x_i^2 - n\bar{x}^2}} \quad (2)$$

$$s_\varepsilon^2 = \frac{SSE}{n-2} \quad (3)$$

$$SSE = SST - SSR \quad (4)$$

$$SSR = 1,45^2 \cdot 6,25 = 13,14$$

$$SST = \sum y_i^2 - n\bar{y}^2 = 101,77 - 12\cdot\left(\frac{32,5}{12}\right)^2 = 13,75$$

④ $SSE = 13,75 - 13,14 = 0,61$

③ $S_E^2 = \frac{0,61}{10} = 0,061$

② $S_{\hat{\beta}_1} = \sqrt{\frac{0,061}{6,25}} = 0,099$

① $t = \frac{1,45}{0,099} = 14,68$

$|t| = |14,68| = 14.68 \boxed{>} t_{10,\,0.025} = 2,306$

Απορρίπτεται η Ηο συνεπώς το υπόδειγμα είναι εγκυρο.

$$R^2 = \frac{SSR}{SST} = \frac{13,14}{13,75} = 0,956$$

Ερμηνεία του συντελεστή προσδιορισμού:
Το 95,6% της μεταβλητότητας των καθαρών κερδών ερμηνέυονται απο την ευθεία παλινδρόμησης.