Ασκήσεις στη Μη Παραμετρική Στατιστική

ΕΛΕΓΧΟΙ ΒΑΣΙΣΜΕΝΟΙ ΣΤΗ ΔΙΩΝΥΜΙΚΗ ΚΑΤΑΝΟΜΗ

ΑΣΚΗΣΗ 1.

Έχει εκτιμηθεί ότι τουλάχιστον οι μισοί των αντρών που υποβάλλονται σε αφαίρεση του καρκίνου του προστάτη υποφέρουν από μια ανεπιθύμητη παρενέργεια . Σε μία προσπάθεια μείωσης των παρενεργειών μία καινούρια μέθοδος επέμβασης μελετάται. Από 19 άντρες που υποβλήθηκαν σε επέμβαση με τη καινούρια μέθοδο μόνο 3 υπέφεραν από παρενέργειες. Είναι λογικό να υποθέσουμε ότι ή νέα μέθοδος επέμβασης μειώνει τις περιπτώσεις των ανεπιθύμητων παρενεργειών; (α=0,05)
(Conover W.J. “Practical Nonparametric Statistics” 3rd edition)

Λύση

$n=19$ $\tau=3$

H_0: $p\geq 0.5$
H_1: $p<0.5$

Κάτω από την H_0 η $T=$ # ασθενών που έχουν παρενέργειες $\sim bin(n=19,p=0.5)$.

1ος τρόπος:

απορρίπτω H_0 αν $T\leq t$ τέτοιο ώστε

$\alpha=0.05\approx P(T\leq t|\ n=19,p=0.5)$

Έχω $P(T\leq \mathbf{5}|\ n=19,p=0.5)=0{,}0318$
$P(T\leq 6|\ n=19,p=0.5)=0{,}0835$

Επειδή το 0,0318 είναι πιο κοντά στο 0,05 (από το 0,0835) παίρνουμε t=5.

Επειδή $\tau=3\leq t=5$ απορρίπτω την H_0 σε επίπεδο σημαντικότητας 0,0318.

$2^{ος}$ τρόπος

Βρίσκω κρίσιμο επίπεδο $\hat{\alpha}$ (p-value) για τ=3

$\hat{\alpha}$ =P(T≤τ|n=19,p=0.5)= P(T≤3|n=19,p=0.5)=0,0022

Επειδή $\hat{\alpha}$ = 0,0022<α=0,05 απορρίπτω H_0.
(Απορρίπτω για κάθε επίπεδο σημαντικότητας μεγαλύτερο του 0,0022).

ΑΣΚΗΣΗ 2.

Μια εταιρεία κατασκευής μηχανημάτων οδοποιίας ισχυρίζεται ότι το 30% των μηχανικών που αναλαμβάνουν έργα του δημοσίου χρησιμοποιούν κάποιο από τα μηχανήματα που παράγει. Από έρευνα σε ένα τυχαίο δείγμα μηχανικών βρέθηκε ότι 5 στους 13 μηχανικούς χρησιμοποιούν μηχανήματα που παράγει η συγκεκριμένη εταιρεία.
Να ελεγχθεί σε επίπεδο σημαντικότητας α=0.05 αν ο ισχυρισμός της εταιρείας είναι βάσιμος ή όχι;
Να διατυπώσετε αναλυτικά τις υποθέσεις του ελέγχου και να ελεγχθεί η μηδενική υπόθεση με 2 τρόπους: **α)** χρησιμοποιώντας το κρίσιμο επίπεδο – p – value και **β)** συγκρίνοντας την τιμή της ελεγχοσυνάρτησης με τις κριτικές τιμές.
(Θέμα $3^{ο}$ εξεταστικής Φεβρουάριου 2010)

Λύση

n=13 τ=5 (παρατηρούμενη τιμή)

$$H_0: p=0.3$$
$$H_1: p\neq 0.3$$

Κάτω από την H_0 η T= # μηχανικών που χρησιμοποιούν το μηχανήματα της εταιρίας ~ bin(n=13,p=0.3) .

1ος τρόπος:

Απορρίπτω H_0 αν $T \le t_1$ ή $T > t_2$ για t_1 και t_2 τέτοια ώστε

$\alpha/2 = 0.025 \approx P(T \le \mathbf{t_1} | n=13, p=0.3)$ και

$\alpha/2 = 0.025 \approx P(T > \mathbf{t_2} | n=13, p=0.3) \Rightarrow 1 - 0{,}025 = 0{,}975 = P(T \le \mathbf{t_2} | n=13, p=0.3)$

Από πίνακες διωνυμικής κατανομής έχω

$\mathbf{P}(T \le \mathbf{0} | n=13, p=0.3) = 0{,}0097$

$P(T \le 1 | n=13, p=0.3) = 0{,}0637$

$\mathbf{P}(T \le 6 | n=13, p=0.3) = 0{,}9376$

$P(T \le \mathbf{7} | n=13, p=0.3) = 0{,}9818$

Επειδή το 0,0097 είναι πιο κοντά στο 0,025 (από το 0,0637) παίρνουμε $t_1=0$

Επίσης επειδή το 0,9818 είναι πιο κοντά στο 0,975 (από το 0,9376) παίρνουμε $t_2=7$

Απορρίπτω H_0 αν $T \le 0$ ή $T > 7$

Εμείς έχουμε $0 < \tau = 5 \le 7$ άρα δεν απορρίπτω H_0 για $\alpha = 0{,}0097 + (1 - 0{,}9818) = 0{,}0279$

Άρα ο ισχυρισμός είναι βάσιμος.

2ος τρόπος

Βρίσκω κρίσιμο επίπεδο $\hat{\alpha}$ (p-value)

$E\{T\} = n*p = 13*0.3 = 3.9.$

Εχω $\tau = 5 > 3{,}9$

$\frac{\hat{\alpha}}{2} = P(T \ge \tau | n=13, p=0.3) = P(T \ge 5 | n=13, p=0.3) =$

$1 - P(T \le 4 | n=13, p=0.3) = 1 - 0{,}6543 = 0{,}3457 \Rightarrow$

$\widehat{\alpha}$ = 2*0,3457=0,6914>α=0,05 άρα δεν απορρίπτω H_0.

ΑΣΚΗΣΗ 3.

Ο υπεύθυνος πωλήσεων μιας πολυεθνικής εταιρείας ειδών οικιακής χρήσης ισχυρίζεται ότι περισσότερα από το 30% των καταστημάτων που προμηθεύει η εταιρεία του έχουν σημειώσει πτώση στις πωλήσεις τους κατά το τελευταίο εξάμηνο. Με βάση τα στοιχεία έρευνας αγοράς, διαπιστώθηκε ότι σε ένα δείγμα 9 καταστημάτων υπάρχει πτώση στις πωλήσεις σε 3 από αυτά. Να εξεταστεί σε επίπεδο σημαντικότητας 0.05, το ακόλουθο ζεύγος των υποθέσεων:

H_0: p≤0.35

H_1: p>0.35

όπου p είναι το ποσοστό των καταστημάτων που προμηθεύει η εταιρεία που έχουν σημειώσει πτώση στις πωλήσεις τους κατά το τελευταίο εξάμηνο. Ποιο είναι το συμπέρασμα στο οποίο καταλήγετε;

(Θέμα $1^{ο}$ εξεταστικής Αυγούστου 2010)

Λυση

n=9 τ=35

H_0: p≤0.3

H_1: p>0.3

$1^{ος}$ τρόπος:

Απορρίπτω H_0 αν T>t για t τέτοιο ώστε

α=0.05≈P(T>**t**| n=9,p=0.35)=>

1-α=0,95= P(T≤**t**| n=9,p=0.35)

Από πίνακες διωνυμικής κατανομής βρίσκω ότι

$P(T\leq \mathbf{5}|\ n=9,p=0.35)=0,9464 \Rightarrow t=5$

Εμείς έχουμε $\tau=3<t=5$ άρα δεν απορρίπτω την H_0.

<u>2ος τρόπος</u>

Βρίσκω κρίσιμο επίπεδο $\hat{\alpha}$ (p-value)

$\hat{a}=P(T\geq\tau|n=9,p=0.35)= P(T\geq3|n=9,p=0.35)=1- P(T\leq2|n=9,p=0.35)=1-0.3373=0.6627$

¨Εχουμε $\hat{a}>\alpha$ άρα δεν απορρίπτω την H_0.

<u>ΑΣΚΗΣΗ 4.</u>

Έστω δυο προϊόντα Α και Β με τα ίδια χαρακτηριστικά. Κάποιος που επιθυμεί να ελέγξει αν το προϊόν Β προτιμάται έναντι του Α

H_0: «προϊόν Β τείνει να μην προτιμάται έναντι του Α»

H_1: «προϊόν Β τείνει να προτιμάται έναντι του Α»

επιλέγει ένα τυχαίο δείγμα 10 καταναλωτών που δοκιμάζουν τα δύο προϊόντα. Από τους καταναλωτές οι 8 προτιμούν το προϊόν Β ένας προτιμά το Α και ένας είναι αδιάφορος. Σε τι συμπέρασμα καταλήγετε;

(Conover W.J. “Practical Nonparametric Statistics” 3rd edition)

<u>Λύση</u>

Πρόκειται για προσημικο έλεγχο (sign test).

Συμβολίζω με «+» το γεγονός «προϊόν Β προτιμάται έναντι του Α»

με «-» το γεγονός «προϊόν Α προτιμάται έναντι του Β»

‘Αρα έχουμε # «+» =8

«-» =1

n= αριθμός των «+» + αριθμός των «-»=8+1=9

Οι υποθέσεις μπορούν να γραφτούν

H_0: P(+)≤1/2

H_1: P(+)>1/2

Κάτω από την H_0 η ελεγχοσυνάρτηση T=# «+»~bin(n,p=0.5)

<u>1ος τρόπος:</u>

Απορρίπτω H_0 αν T≥t για t τέτοιο ώστε

α=0.05≈P(T≥**t**| n=9,p=0.5)

=P(T≤n-**t**| n=9,p=0.5) **(επειδή η bin(n=9,p=0.5) είναι συμμετρική)**

Από πίνακες βρίσκω ότι P(T≤1| n=9,p=0.5)=0,0195

P(T≤2| n=9,p=0.5)=0,0898

Επειδή το 0,0195είναι πιο κοντά στο 0,05 (από το 0,0898) παίρνουμε n-**t** =1

Έτσι έχουμε n-**t** =1 => **t**=n-1=9-1=8

Επειδή εμείς έχουμε τ= # «+»=8≥t =8 απορρίπτω την H_0 σε επίπεδο σημαντικότητας α=0,0195, άρα μπορούμε να πούμε ότι οι καταναλωτές προτιμούν το B σε σχέση με το A προϊόν.

****

θα μπορούσαμε να βρούμε τη κρίσιμη περιοχή του ελέγχου χωρίς να λάβουμε υπ ‘όψιν το γεγονός ότι η **bin(n=9,p=0.5) είναι συμμετρική.** Σε αυτή τη περίπτωση θα έχουμε

Απορρίπτω H_0 αν $T>t_1$ για t_1 τέτοιο ώστε

$\alpha=0.05=P(T>\mathbf{t_1}|\ n=9,p=0.5)=>$

$1-\alpha=0,95= P(T\leq \mathbf{t_1}|\ n=9,p=0.5)$

Από πίνακες βρίσκω ότι $P(T\leq 6|\ n=9,p=0.5)=0,9102$

$P(T\leq 7|\ n=9,p=0.5)=0,9805$

Επειδή το 0,9805 είναι πιο κοντά στο 0,95 (από το 0,9102) παίρνουμε $\mathbf{t_1}=7$

Επειδή εμείς έχουμε τ= # «+»=8>t_1 =7 απορρίπτω την H_0 σε επίπεδο σημαντικότητας 1-0,9805=0,0195

****

<u>2ος τρόπος:</u>

Βρίσκω κρίσιμο επίπεδο $\hat{\alpha}$ (p-value)

$\hat{a}=P(T\geq \tau|n=9,p=0.5)= P(T\geq 8|n=9,p=0.5)$

$= P(T\leq n-8|n=9,p=0.5)$ **(επειδή η bin(n=9,p=0.5) είναι συμμετρική)**

$= P(T\leq 1|n=9,p=0.5)=0.0195$

Επειδή $\hat{\alpha}<\alpha=0,05$ απορρίπτω H_0

<u>ΑΣΚΗΣΗ 5</u>

Πριν από ένα τηλεοπτικό debate μεταξύ 2 υποψήφιων προέδρων ένα δείγμα 100 ψηφοφόρων δηλώνουν ποιον προτιμούν να αναλάβει τη προεδρεία. 84 προτιμούν τον υποψήφιο D και 16 τον υποψήφιο R. Μετά το debate ξαναρωτήθηκαν τα ίδια 100 άτομα Το ¼ από αυτούς που προτιμούσαν τον D άλλαξαν γνώμη. Επίσης το ¼ από αυτούς που προτιμούσαν τον R προτιμούν τώρα τον D. Να ελεγχθεί το παρακάτω ζεύγος υποθέσεων

H_0: «ο πληθυσμός δεν άλλαξε πεποιθήσεις μετά το debate »

H_1: «άλλαξε το ποσοστό των ψηφοφόρων που προτιμά τον D »

(Conover W.J. “Practical Nonparametric Statistics” 3rd edition)

Λύση

		ΜΕΤΑ	
		D	R
ΠΡΙΝ	D	63	21
	R	4	12

Συμβολίζω με X_i=0 αν ο i προτιμά **πριν** το debate τον D

X_i=1 αν ο i προτιμά **πριν** το debate τον R

Y_i=0 αν ο i προτιμά **μετά** το debate τον D

Y_i=1 αν ο i προτιμά **μετά** το debate τον R

Επίσης συμβολίζω με «+» το γεγονός $(X_ι, Y_ι)= (0,1)$ «πριν D μετά R»

και με «-» το γεγονός $(X_ι, Y_ι)= (1,0)$ «πριν R μετά D»

\# «+»=21

\# «-»=4

n= # «+» + # «-»=25

τ= # «+»=21

H_0: P(+)=1/2

H_1: P(+)≠1/2

Κάτω από την H_0 η ελεγχοσυνάρτηση T=# «+»~bin(n=25,p=0.5) που προσεγγίζει τη N(np,npq)

<u>1^{ος} τρόπος:</u>

Απορρίπτω H_0 αν $T \leq t_1$ ή $T \geq t_2$ για t_1 και t_2 τέτοια ώστε

--*Βρίσκω t_1*

$\alpha/2 = 0.025 = P(T \leq \mathbf{t_1} \mid n=25, p=0.5) = P\left[\left(\frac{T-np}{\sqrt{npq}}\right) \leq \left(\frac{t_1-np}{\sqrt{npq}}\right) \mid n=25, p=0.5\right] =$

$P\left[Z \leq \left(\frac{t_1-np}{\sqrt{npq}}\right) \mid n=25, p=0.5\right] = P\left[Z \leq -1.96 \mid n=25, p=0.5\right]$

$=> \left[\left(\frac{t_1-np}{\sqrt{npq}}\right) = -1.96\right] =>$

$$t_1 = np - 1.96\sqrt{npq} = 25*0.5 - 1.96*\sqrt{25*0.5*0.5} = 7.6 \approx 8$$

--*Βρίσκω t_2*

$\alpha/2 = 0.025 = P(T \geq \mathbf{t_2} \mid n=25, p=0.5)$

$= P(T \leq n-\mathbf{t_2} \mid n=25, p=0.5)$ (επειδή η bin(n=25,p=0.5) είναι συμμετρική)
$= P(T \leq t_1 \mid n=25, p=0.5)$

$=> n-\mathbf{t_2} = t_1 => \mathbf{t_2} = n-t_1 = 25-8 = 17$

Επειδή $\tau=21>17=t_2$ απορρίπτω H_0

$2^{ος}$ τρόπος:

Βρίσκω κρίσιμο επίπεδο $\widehat{\alpha}$ (p-value)

Επειδή τ=21>12.5=E(T)=np, κοιτάζω τη δεξιά ουρά.

$\frac{\widehat{\alpha}}{2}$ =P(T≥τ|n=25,p=0.5)= P(T≤n-τ|n=25,p=0.5)= P(T≤25-21|n=25,p=0.5)= P(T≤4|n=25,p=0.5)=

$$P\left[\left(\frac{T-np}{\sqrt{npq}}\right) \le \left(\frac{4-np}{\sqrt{npq}}\right) \mid n=25, p=0.5\right] = P\left[Z \le \left(\frac{4-12.5}{\sqrt{6.25}}\right) \mid n=25, p=0.5\right] =$$

$P\left[Z \le -3.40\right]$=0.00036

=> $\widehat{a}$ =2*0.00036≈0.0007<α=0,05

άρα απορρίπτω H_0

ΑΣΚΗΣΗ 6

Οι παρακάτω τιμές αφορούν τις ποσότητες βροχόπτωσης της πόλης «Α» για 19 συνεχόμενα έτη.

45,25 45,83 41,77 36,26 45,37 52,25 35,37 57,16 35,37 58,32 41,05 33,72 45,73 37,90 41,72 36,07 49,83 36,24 39,90.

Να ελεγχθεί αν η βροχόπτωση τείνει να αυξάνεται ή να μειώνεται.
(Conover W.J. "Practical Nonparametric Statistics" 3rd edition)

Λύση

Πρόκειται για έλεγχο ύπαρξης τάσης Cox-Stuart .

H_0: «Δεν υπάρχει τάση»

H_1: «Υπάρχει τάση»

Πρέπει να δημιουργήσουμε ζεύγη. Έχουμε όμως 19(περιττός αριθμός) παρατηρήσεις. Για το λόγο αυτό αφαιρώ από το δείγμα μου τη μεσαία(10η) παρατήρηση (το 58,32) και έχω το νέο δείγμα έστω X={45,25 45,83 41,77 36,26 45,37 52,25 35,37 57,16 35,37 58,32 41,05 33,72 45,73 37,90 41,72 36,07 49,83 36,24 39,90}.

με 18 παρατηρήσεις. Σε αυτό το δείγμα δημιουργώ τα ζεύγη (X_i, X_{i+9}) και θέτω «+» αν $X_i < X_{i+9}$, «-» αν $X_i > X_{i+9}$ και «0» αν $X_i = X_{i+9}$.

(45,25 41,05) -

(45,83 37,72) -

(41,77 45,73) +

(36,26 37,90) +

(45,37 41,72) -

(52,25 36,07) -

(35,37 49,83) +

(57,16 36,24) -

(35,37 39,90) +

n= # «+» + # «-»=9

τ= # «+»=4

H_0: P(+)=1/2

H_1: P(+)≠1/2

Κάτω από την H_0 η ελεγχοσυνάρτηση T=# «+»~bin(n=9,p=0.5)

$1^{ος}$ τρόπος:

Απορρίπτω H_0 αν $T \leq t_1$ ή $T \geq t_2$

--*Βρίσκω* t_1

$\alpha/2 = 0.025 = P(T \leq \mathbf{t_1} | n=9, p=0.5)$

Από πίνακες βρίσκω ότι

$P(T \leq 1 | n=9, p=0.5) = 0{,}0195$

$\Rightarrow \mathbf{t_1 = 1}$

--*Βρίσκω* t_2

$\alpha/2 = 0.025 = P(T \geq \mathbf{t_2} | n=9, p=0.5) = P(T \leq n - \mathbf{t_2} | n=9, p=0.5)$

βρήκαμε πριν $\alpha/2 = P(T \leq t_1 = 1 | n=9, p=0.5) = 0{,}0195$

=>

$\mathbf{t_2} = n - t_1 = 9 - 1 = 8$

$\alpha = 2 * 0{,}0195 = 0{,}0390$

επειδή $\tau=4$ δεν απορρίπτω H_0 σε επίπεδο σημαντικότητας 0,0390.

(Απορρίπτω αν $T \leq 1$ ή $T \geq 8$)

$2^{ος}$ τρόπος:

Βρίσκω κρίσιμο επίπεδο $\widehat{\alpha}$ (p-value)

Επειδή $\tau = 4 < 4{,}5 = E(T) = np$, κοιτάζω αριστερή ουρά.

$$\frac{\widehat{\alpha}}{2} = P(T \leq \tau | n=9, p=0.5) = P(T \leq 4 | n=9, p=0.5) = 0{,}5$$

=>

$\hat{a}=1$=> Η H_0 δεν απορρίπτεται για κανένα επίπεδο σημαντικότητας

ΑΣΚΗΣΗ 7

Ο Cohran (1947) μελετά την επίδραση δύο διαφορετικών φάρμακων σε δέκα ασθενείς. Χρησιμοποιώντας μία κλίμακα μέτρησης καταγράφει τα δεδομένα τα οποία παρουσιάζονται στο παρακάτω πίνακα. Ελέγξτε αν υπάρχει θετική συσχέτιση ως προς την επίδραση των δύο φάρμακων

(Conover W.J. “Practical Nonparametric Statistics” 3rd edition)

DRUG 1	DRUG 2
0.7	1.9
-1.6	0.8
-0.2	1.1
-1.2	0.1
-0.1	-0.1
3.4	4.4
3.7	5.5
0.8	1.6
0	4.6
2	3.4

Λύση

H_0: «Δεν υπάρχει θετική συσχέτιση μεταξύ DRUG 1και DRUG 2»

H_1: «Υπάρχει θετική συσχέτιση μεταξύ DRUG 1και DRUG 2»

Ανακατατάσσω τα ζεύγη έτσι ώστε οι τιμές του DRUG 1 να διατάσσονται κατά αύξουσα σειρά.

D1	D 2
-1.6	0.8
-1.2	0.1
-0.2	1.1
-0.1	-0.1
0	4.6
0.7	1.9
0.8	1.6
2	3.4
3.4	4.4
3.7	5.5

Και κάνω στη σειρά D2 έλεγχο Cox-Stuart για ύπαρξη θετικής τάσης.

Δημιουργώ ζεύγη ($D2_i$, $D2_{i+5}$) και θέτω «+» αν $D2_i < D2_{i+5}$, «-» αν $D2_i > D2_{i+5}$ και «0» αν $D2_i = D2_{i+5}$.

(0,8 1,9) «+»

(0,1 1,6) «+»

(1,1 3,4) «+»

(-0,1 4,4) «+»

(4,6 5,5) «+»

H_0: P(+)≤1/2

H_1: P(+)>1/2

n= # «+» + # «-»=5

τ= # «+»=5

Κάτω από την H_0 η ελεγχοσυνάρτηση T=# «+»~bin(n=5,p=0.5)

1ος τρόπος:

Βρίσκω t τέτοιο ώστε να απορρίπτω H_0 αν $T\geq t$

α= 0.05=P(T≥**t**| n=5,p=0.5)= P(T≤n-**t**| n=5,p=0.5)

Από πίνακες διωνυμικής κατανομής έχω

P(T≤0| n=5,p=0.5)=0,0312

=>n-t=0 => t=5.

Άρα σε επίπεδο σημαντικότητας 0,0312 απορρίπτω H_0 αν T≥5. Εμείς έχουμε τ=5 άρα απορρίπτουμε H_0 σε επίπεδο σημαντικότητας 0,0312. Άρα υπάρχει θετική συσχέτιση ανάμεσα στα δύο φάρμακα.

2ος τρόπος

Βρίσκω κρίσιμο επίπεδο $\hat{\alpha}$ (p-value)

$\hat{a}$ = P(T≥5| n=5,p=0.5)= P(T≤n-5| n=5,p=0.5)= P(T≤0| n=5,p=0.5)=0,0312<α=0,05 άρα απορρίπτω H_0

ΑΣΚΗΣΗ 8

Ένας ερευνητής θέλει να ελέγξει αν το μεσημεριανό γεύμα επηρεάζει τα αντανακλαστικά σε εργαζομένους γραφείου. Για αυτόν τον λόγο συγκρίνει τον χρόνο αντίδρασης (σε κάποιο ερέθισμα) πριν το γεύμα με τον χρόνο αντίδρασης μετά το γεύμα 28 εργαζόμενων. 22 αντιδρούσαν γρηγορότερα πριν το φαγητό ενώ 2 εργαζόμενοι αντιδρούσαν το ίδιο. Ελέγξτε αν ο χρόνος αντίδρασης είναι σημαντικά μεγαλύτερος μετά το μεσημεριανό γεύμα.

H_0: Ο χρόνος αντίδρασης δεν αυξάνεται μετά το μεσημεριανό φαγητό

H_1: Ο χρόνος αντίδρασης <u>αυξάνεται</u> μετά το μεσημεριανό φαγητό
(Conover W.J. "Practical Nonparametric Statistics" 3rd edition)

<u>Λύση</u>

Έχω n'=28
Συμβολίζω με «+» αν ο χρόνος αντίδρασης αυξήθηκε μετά το μεσημεριανό γεύμα.
Συμβολίζω με «-» αν ο χρόνος αντίδρασης μειώθηκε μετά το μεσημεριανό γεύμα.
Συμβολίζω με «0» αν ο χρόνος αντίδρασης έμεινε ο ίδιος μετά το μεσημεριανό γεύμα

Έχω #«+»=22
#«-»=??
«0»=2

'Όμως n'=#«+» + #«-» + #«0» => 28=22 + #«-» + 2 =>

#«-»=4

n=#«+» + #«-»=22+4=26

Οι υποθέσεις μπορούν να γραφτούν

H_0: $P(+)\leq 1/2$
H_1: $P(+)>1/2$

Κάτω από την H_0 η ελεγχοσυνάρτηση T=# «+»~bin(n=26,p=0.5) που προσεγγίζει τη N(np,npq)=N(13,6.5)

<u>1ος τρόπος:</u>

Βρίσκω t τέτοιο ώστε να απορρίπτω H_0 αν $T\geq t$

α= 0.05=P(T≥**t**| n=26,p=0.5)= P(T≥**t**| n=26,p=0.5)=

$$P\left[\left(\frac{T-np}{\sqrt{npq}}\right)\geq\left(\frac{t\ -np}{\sqrt{npq}}\right)\mid n=26,p=0.5\right]=P\left[\left(\frac{T-np}{\sqrt{npq}}\right)\geq\left(\frac{t\ -np}{\sqrt{npq}}\right)\mid n=26,p=0.5\right]=$$

$$P\left[Z\geq\left(\frac{t\ -13}{\sqrt{6.5}}\right)\right]=\ P\left[Z\geq 1.645\right]=>$$

$$\left(\frac{t\ -13}{\sqrt{6.5}}\right)=1.645 => t=13+1.645*\sqrt{6.5}=17.193\approx 17$$

Εμείς έχουμε τ=22>17 άρα απορρίπτουμε H_0.

<u>$2^{ος}$ τρόπος:</u>

$\hat{a}$ = P(T≥22| n=26,p=0.5)=P $\left[Z\geq\left(\frac{22-13}{\sqrt{6.5}}\right)\right]$ = P $\left[Z\geq\left(\frac{22-13}{\sqrt{6.5}}\right)\right]$ = P(Z≥3.53)=

P(Z≤-3.53)=0.0002

Επειδή α=0,05> $\hat{a}$ =0,0002 απορρίπτω H_0.

<u>ΕΛΕΓΧΟΣ WILCOXON ΓΙΑ ΤΗ ΔΙΑΜΕΣΟ ΕΝΟΣ ΠΛΗΘΥΣΜΟΥ</u>

<u>ΑΣΚΗΣΗ 9</u>

Ο μέσος βαθμός ενός μαθήματος ιστορίας σε ένα πανεπιστημιακό τμήμα κατά την τελευταία δεκαπενταετία είναι 6,5 μονάδες. Την τελευταία χρονιά άλλαξε το σύστημα εισαγωγής των φοιτητών στο τμήμα και μετά την εξέταση του μαθήματος ο υπευθύνως καθηγητής επιθυμεί να ελέγξει αν διαφοροποιήθηκαν σημαντικά οι βαθμολογίες των φοιτητών σε σχέση με εκείνες των φοιτητών των προηγούμενων 15 ετών. Για το λόγο αυτό συνέλεξε ένα τυχαίο δείγμα 15 φοιτητών των οποίων οι βαθμοί είναι οι

4 4,5 1 9,5 2,5 6 7   9 9,5 8 6,5 10 10 9 6 7

Πιστεύεται ότι τα παραπάνω δεδομένα παρέχουν σημαντικές ενδείξεις ότι η βαθμολογία των νέων φοιτητών διαφέρει σημαντικά από την βαθμολογία των φοιτητών των προηγούμενων ετών; (Υποθέστε ότι η κατανομή της βαθμολογίας είναι συμμετρική)
(Ε. ΞΕΚΑΛΑΚΗ “Μη Παραμετρική Στατιστική” 2001)

Λύση

$H_0 : E(X) = 6.5$
$H_1 : E(X) \neq 6.5$

x	D=x-6,5	|D|	R(|D|)	R	R^2
4	-2,5	2,5	8	-8	64
4,5	-2	2	6	-6	36
1	-5,5	5,5	14	-14	196
9,5	3	3	10,5	10,5	110,25
2,5	-4	4	13	-13	169
6	-0,5	0,5	2,5	-2,5	6,25
7	0,5	0,5	2,5	2,5	6,25
9	2,5	2,5	8	8	64
9,5	3	3	10,5	10,5	110,25
8	1,5	1,5	5	5	25
6,5	0	0	-	-	-
10	3,5	3,5	12	12	144
9	2,5	2,5	8	8	64
6	-0,5	0,5	2,5	-2,5	6,25
7	0,5	0,5	2,5	2,5	6,25
SUM				**13**	**1007.5**

$$T = \frac{\sum R_i}{\sqrt{\sum R_i^2}} = \frac{13}{\sqrt{1007{,}5}} = 0.41$$

Απορρίπτω H_0 αν T<-1,96(=$Z_{0,025}$) ή αν T>1,96(=$Z_{0,975}$ =-$Z_{0,025}$)

Εμείς επειδή T=0,41 δεν απορρίπτουμε $H_{0.}$ Άρα τα δεδομένα δεν παρέχουν σημαντικές ενδείξεις ότι η βαθμολογία των νέων φοιτητών διαφέρει σημαντικά από την βαθμολογία των φοιτητών των προηγούμενων ετών

ΕΛΕΓΧΟΣ WILCOXON ΓΙΑ ΔΕΙΓΜΑ ΖΕΥΓΩΝ ΠΑΡΑΤΗΡΗΣΕΩΝ

ΑΣΚΗΣΗ 10

Για να ελεγχθούν τα αποτελέσματα ενός νέου εμβολίου, απαιτείται η γνώμη της μεταβολής της θερμοκρασίας του σώματος πριν και μετά τον εμβολιασμό. Οι μετρήσεις της θερμοκρασίας (σε βαθμό Κελσίου) σε 9 τυχαία επιλεγμένους ασθενείς δίνονται στον παρακάτω πίνακα.

Ασθενείς	**Θερμοκρασία**	
	Πριν	**Μετά**
1	37.0	38.0
2	37.0	37.2
3	36.4	37.3
4	36.7	38.6
5	37.0	37.8
6	36.9	36.9
7	37.0	36.9
8	37.0	39.6
9	37.0	37.5

Να θεωρήσετε ότι η κατανομή της διαφοράς των θερμοκρασιών του σώματος (d = θερμοκρασία μετά – θερμοκρασία πριν) είναι περίπου συμμετρική. Αποτελούν τα δεδομένα ένδειξη αύξησης στη θερμοκρασία του σώματος μετά τον εμβολιασμό (H_0:$d_{0.5}\leq 0$ vs H_1:$d_{0.5}>0$); Να διεξάγετε τον παρακάτω έλεγχο σε επίπεδο σημαντικότητας α=5%.
(Θέμα 3ο εξεταστικής Αυγούστου 2010)

Λύση

$$H_0: d_{0.5} \leq 0$$
$$H_1: d_{0.5} > 0$$

Π	**Μ**	**D=M-Π**	**|D|**	**R(|D|)**	**R**
37.0	38.0	1	1	6	6
37.0	37.2	0,2	0.2	2	2
36.4	37.3	0,9	0,9	5	5
36.7	38.6	1,9	1,9	7	7
37.0	37.8	0,8	0,8	4	4
36.9	36.9	0	0	το αγνοούμε	
37.0	36.9	-0,1	0.1	1	-1
37.0	39.6	2,6	2.6	8	8
37.0	37.5	0,5	0.5	3	3

n'=9 n=8 Επειδή δεν υπάρχει ταύτιση τιμών στο |D| η ελεγχοσυνάρτηση είναι η

T^+=άθροισμα θετικών R_i

Απορρίπτω H_0 αν $T^+ > w_{1-\alpha}$

$w_\alpha = w_{0,05} = 6$ (για n=8)

$$w_{1-\alpha} = \frac{n(n+1)}{2} - w_\alpha \Rightarrow w_{0.95} = \frac{8(9)}{2} - 6 = 30$$

Στο δείγμα μας έχουμε άθροισμα θετικών R_ι =35>30 (=$w_{0.95}$) άρα απορρίπτω H_0.

ΕΛΕΓΧΟΣ ΤΩΝ MANN-WHITNEY

ΑΣΚΗΣΗ 11

Το τμήμα στεγαστικών δανείων μιας τράπεζας ενδιαφέρεται να εκτιμήσει αν υπάρχει διαφορά στο μέσο ύψος των δανείων που χορηγήθηκαν στην διάρκεια του περασμένου Δεκεμβρίου σε δύο κεντρικά καταστήματα της. Για το σκοπό αυτό επέλεξε τυχαίο δείγμα 7 δανείων από το σύνολο των στεγαστικών δανείων που εγκρίθηκαν τον περασμένο Δεκέμβριο σε κάθε ένα από τα δύο καταστήματα και κατέγραψε το ποσό των δανείων σε χιλιάδες ευρώ:

Κατάστημα Α	127	210	203	108	178	100	85
Κατάστημα Β	115	110	95	83	117	280	145

Να ελεγχθεί σε επίπεδο σημαντικότητας α=0.05 αν υπάρχει διαφορά στο μέσο ύψος των δανείων που χορηγήθηκαν στην διάρκεια του περασμένου Δεκεμβρίου στα δύο κεντρικά καταστήματα της τράπεζας.

(Θέμα 2^{o} εξεταστικής Φεβρουάριου 2010)

Λύση

$$H_0: E(A)=E(B)$$

$$H_1: E(A)\neq E(B)$$

n=m=7

N=n+m=14

A	**B**	**R(A)**	**R(B)**
127	115	9	7
210	110	13	6
203	95	12	3
108	83	5	1
178	117	11	8
100	280	4	14
85	145	2	10
		56	**49**

Δεν έχουμε ισοπαλίες. Άρα η ελεγχοσυνάρτηση είναι η Τ= $\sum R(A_i)$ =56

και απορρίπτω αν Τ<$w_{0.025}$ ή Τ>$w_{0.975}$

$w_{0.025}$=37

$w_{0.975}$=n(N+1)-$w_{0.25}$=7(14+1)-37=68

αφού Τ=56 >37 και <68 δεν απορρίπτω H_0.

<u>ΜΕΤΡΑ ΣΥΣΧΕΤΙΣΗΣ ΤΑΞΗΣ ΜΕΓΕΘΟΥΣ</u>

<u>ΑΣΚΗΣΗ 12</u>

Ερευνητής ενδιαφέρεται να μελετήσει τον βαθμό ευφυΐας των παιδιών των οικογενειών που έχουν 2 αγόρια. Για το σκοπό αυτό, επιλέχτηκε τυχαίο δείγμα 5 οικογενειών και καταγράφηκε ο βαθμός σε ένα τεστ IQ των δύο παιδιών. Τα αποτελέσματα καταγράφονται στον παρακάτω πίνακα:

	Βαθμός IQ 1^{ου} παιδιού	**Βαθμός IQ 2^{ου} παιδιού**
1	110	115
2	108	103
3	120	112
4	105	120
5	95	88

α) να υπολογίσετε τον συντελεστή συσχέτισης του Kendall

(Θέμα 4^{ο} εξεταστικής Φεβρουάριου 2010)

<u>Λύση</u>

Διατάσσουμε τα ζεύγη κατά αύξουσα τάξη μεγέθους των τιμών IQ του 1^{ου} παιδιού.

IQ_1	**IQ_2**	**εναρμονοσμένα ζεύγη**	**μη εναρμονισμένα**
95	88	4	0
105	120	0	3
108	103	2	0
110	115	0	1
120	112	0	0
		N_C=6	**N_d=4**

$$\tau=\frac{N_c - N_d}{\frac{n(n-1)}{2}} = \frac{6-4}{10} = 0.2$$

<u>ΑΣΚΗΣΗ 13</u>

Είναι εύλογο να υποθέσει κανείς ότι οι βαθμοί των δύο τεστ στο παραπάνω παράδειγμα είναι θετικά συσχετισμένοι, σε επίπεδο σημαντικότητας 0.05;

(Θέμα 4^{o} εξεταστικής Φεβρουάριου 2010)

<u>Λύση</u>

H_0: IQ_1 και IQ_2 είναι ασυσχέτιστα

H_1: IQ_1 και IQ_2 είναι θετικά συσχετισμένα

Η ελεγχοσυνάρτηση είναι η $T=N_c-N_d$ απορρίπτω H_0 αν $T>w_{0.95}=6$

Αφού $N_c-N_d=6-4=2<6$ δεν απορρίπτεται η μηδενική υπόθεση ότι δεν υπάρχει θετική συσχέτιση

<u>ΑΣΚΗΣΗ 14</u>

Ερευνητής ενδιαφέρεται να μελετήσει τον βαθμό ευφυΐας των παιδιών των οικογενειών που έχουν 2 αγόρια. Για το σκοπό αυτό, επιλέχτηκε τυχαίο δείγμα 5 οικογενειών και καταγράφηκε ο βαθμός σε ένα τεστ IQ των δύο παιδιών. Τα αποτελέσματα καταγράφονται στον παρακάτω πίνακα:

	Βαθμός IQ 1^{ου} παιδιού	**Βαθμός IQ 2^{ου} παιδιού**
1	110	115
2	108	103
3	120	112
4	105	120
5	95	88

Α) να υπολογίσετε τον συντελεστή συσχέτισης του Spearman

Β) Διεξάγετε μονόπλευρο έλεγχο για θετική συσχέτιση

<u>Λύση</u>

Α)

X(=IQ$_1$)	**Y(=IQ$_2$)**	**R(X)**	**R(Y)**	**R(X)- R(Y)**	**(R(X)- R(Y))^2**
110	115	4	4	0	0
108	103	3	2	1	1
120	112	5	3	2	4
105	120	2	5	-3	9
95	88	1	1	0	0
					SUM=16

Δεν έχουμε ισοπαλίες άρα

$$\rho = 1 - \frac{6\sum_{i-1}^{5}\left[R(X_i) - R(Y_i)\right]^2}{n(n^2-1)} = 1 - \frac{6*14}{5(25-1)} = 1 - 0{,}7 = 0{,}3$$

B)

H_0: X και Y είναι ασυσχέτιστα

H_1: X και Y είναι θετικά συσχετισμένα

Απορρίπτω H_0 αν $\rho > w_{0.95}$

Από πίνακες (για Spearman ελεγχοσυνάρτηση) έχω $w_{0.95}=0{,}8$

Αφού $\rho=0{,}3<0{,}8$ δεν απορρίπτω H_0.

<u>ΑΣΚΗΣΗ 15</u>

Πέντε απόφοιτοι MBA μελετώνται ώστε να ερευνήσουμε τη συσχέτιση μεταξύ του GMAT-score και το βαθμό του μεταπτυχιακού διπλώματος. Τα scores και οι βαθμοί δίνονται στο παρακάτω πίνακα.

Φοιτητης	Βαθμος MBA	GMAT-score
1	3,5	610
2	3,8	580
3	3,9	640
4	4	610
5	4	710

A) Υπολογίστε το συντελεστή συσχέτισης του **Spearman**
B) Ελέγξτε την ύπαρξη θετικής συσχέτισης μεταξύ GMAT και MBA βαθμούς.

Λύση

n=5 και έχουμε και ισοπαλίες.

Φοιτητης	**X**(=Βαθμος MBA)	**Y**(=GMAT-score)	R(X)	R(Y)	**R(X) R(Y)**	**R(X)^2**	**R(Y)^2**
1	3,5	610	1	2.5	2.5	1	6,25
2	3,8	580	2	1	2	4	1
3	3,9	640	3	4	12	9	16
4	4	610	4.5	2.5	11.25	20,25	6,25
5	4	710	4.5	5	22.5	20,25	25
					50,25	**54,5**	**54,5**

$n\left(\frac{n+1}{2}\right)^2 = 5\frac{6^2}{4}$ =45

ρ= $\frac{\sum R(X_i)R(Y_i) - n\left(\frac{n+1}{2}\right)^2}{\sqrt{\left[\sum R(X_i)^2 - n\left(\frac{n+1}{2}\right)^2\right]\left[\sum R(Y_i)^2 - n\left(\frac{n+1}{2}\right)^2\right]}} = \frac{50.25-45}{\sqrt{(54.5-45)(54.5-45)}} =$

=0,5526

B)

H_0: X και Y είναι ασυσχέτιστα

H_1: X και Y είναι θετικά συσχετισμένα

Απορρίπτω H_0 αν $\rho > w_{0.95}$

Από πίνακες (για Spearman ελεγχοσυνάρτηση) έχω $w_{0.95}$=0,8

Αφού ρ=0,5526<0,8 δεν απορρίπτω H_0.

<u>ΑΣΚΗΣΗ 16</u>

Πέντε απόφοιτοι MBA μελετώνται ώστε να ερευνήσουμε τη συσχέτιση μεταξύ του GMAT-score και το βαθμό του μεταπτυχιακού διπλώματος. Τα scores και οι βαθμοί δίνονται στο παρακάτω πίνακα.

Φοιτητης	Βαθμος MBA	GMAT-score
1	3,5	610
2	3,8	580
3	3,9	640
4	4	610
5	4	710

Α) Υπολογίστε το συντελεστή συσχέτισης του **Kendal**

Β) Ελέγξτε την ύπαρξη θετικής συσχέτισης μεταξύ GMAT και MBA βαθμούς.

<u>Λύση</u>

Α) n=5 και έχουμε και ισοπαλίες.

Φοιτητης	**X**(=Βαθμος MBA)	**Y**(=GMAT-score)	**εναρμονοσμένα ζεύγη**	**μη εναρμονισμένα**	**Ισοπαλιες**
1	3,5	610	2	1	1
2	3,8	580	3	0	0
3	3,9	640	1	1	0
4	4	610	0	0	1
5	4	710	0	0	0
			N_C=6	**N_d=2**	

$$\tau = \frac{N_c - N_d}{\frac{n(n-1)}{2}} = \frac{6-2}{10} = 0.4$$

B) H_0: X και Y είναι ασυσχέτιστα

H_1: X και Y είναι θετικά συσχετισμένα

Απορρίπτω H_0 αν $T=N_c-N_d > w_{0.95}=6$

Αφού $N_c-N_d=6-2=4<6$ δεν απορρίπτεται η μηδενική υπόθεση ότι δεν υπάρχει θετική συσχέτιση.

X^2 ΕΛΕΓΧΟΙ ΚΑΛΗΣ ΠΡΟΣΑΡΜΟΓΗΣ

ΑΣΚΗΣΗ 17

.Σε ένα πείραμα γενετικής δύο διαφορετικές ποικιλίες ενός καθορισμένου είδους διασταυρώνονται. Οι απόγονοι αυτής της διασταύρωσης έχουν ένα ορισμένο χαρακτηριστικό σε μία από τις τρεις γενοτυπικές κατηγορίες Α,Β και Γ. Σύμφωνα με ένα παραδεκτό πιθανοθεωρητικό πρότυπο οι πιθανότητες για να εμφανιστεί το παραπάνω χαρακτηριστικό σε μία από τις κατηγορίες Α,Β και Γ είναι , αντίστοιχα, 1/2, 1/3 και 1/6. Να ελεγχθεί η εγκυρότητα του παραπάνω πιθανοθεωρητικού προτύπου σε επίπεδο στατιστικής σημαντικότητας $\alpha=0.10$, εάν από 60 απογόνους, 36,18 και 6 έχουν το χαρακτηριστικό της κατηγορίας Α,Β και Γ αντίστοιχα.

(Θέμα εξεταστικής Φεβρουαρίου 2010)

ΛΥΣΗ

H_0: Ισχύει το πιθανοθεωρητικό πρότυπο
H_1: Δεν ισχύει το πιθανοθεωρητικό πρότυπο

Κλάση	$O_ι$	p_i^0
1	36	1/2
2	18	1/3
3	6	1/6

$n'=36+18+6=60$

$$T = \sum_{i=1}^{3} \frac{(O_i - E_i)^2}{E_i} \sim X^2_{3-1}$$

$E_i = n' p_i^0 \Rightarrow \quad E_1=60*1/2=30$

$E_2=60*1/3=20$

$E_3=60*1/6=10$

$$T = \sum_{i=1}^{3} \frac{(O_i - E_i)^2}{E_i} = \frac{(36-30)^2}{30} + \frac{(18-20)^2}{20} + \frac{(6-10)^2}{10} = \frac{36}{30} + \frac{4}{20} + \frac{16}{10} = 3$$

Απορρίπτω H_0 αν $T=3 > X^2_{2,0.9}=4.605$ που δεν ισχύει. Άρα δεν απορρίπτω H_0 σε α=0,10 και το πιθανοθεωρητικο πρότυπο είναι έγκυρο.

<u>ΑΣΚΗΣΗ 18</u>

.Σε βιομηχανική παραγωγή ηλεκτρονικών εξαρτημάτων ορισμένου τύπου, μηχανικός ποιοτικού ελέγχου λαμβάνει ημερησίως δείγματα μεγέθους 10 και στη συνέχεια ελέγχει για ελαττωματικά εξαρτήματα. Αν σε 200 διαδοχικές ημέρες ο μηχανικός έλαβε 112 δείγματα που περιέχουν 0 ελαττωματικά, 76 δείγματα που περιέχουν 1 ελαττωματικό και τέλος 12 δείγματα που περιέχουν 2 ελαττωματικά εξαρτήματα, να ελεγχθεί σε επίπεδο σημαντικότητας α=0.05 η υπόθεση ότι τα δείγματα αυτά προέρχονται από πληθυσμό με κατανομή τη διωνυμική.
(Θέμα εξεταστικής Αυγούστου 2010)

ΛΥΣΗ

H_0: Η κατανομή των ελαττωματικών εξαρτημάτων είναι Bin(n=10,p)

H_1: Η κατανομή των ελαττωματικών εξαρτημάτων δεν είναι Bin(n=10,p)

Για την εκτίμηση της παραμέτρου p έχουμε:

$\hat{p} = \dfrac{0\cdot 112 + 0.1\cdot 76 + 0.2\cdot 12}{112+76+12} = \dfrac{10}{200}$ =0,05

Άρα

Πλήθος ελαττωματικών εξαρτημάτων	**O_i**	**p_i^0 = (Πιθανότητες διωνυμικής με n=10 και p=0,05)**	**E_i = n’ p_i^0**
0	112	0,5987	119,74
1	76	0,3152	63,04
2	12	0,0746	14,92
3 και άνω	0	0,0115	2,3
Σύνολο	**n’= 200**	**1**	**200**

Η στατιστική συνάρτηση ελέγχου είναι η

$T = \sum_{i=1}^{4} \dfrac{(O_i - E_i)^2}{E_i} = \dfrac{(112-119.74)^2}{119.74} + \ldots + \dfrac{(0-2.3)^2}{2.3}$ =6,036.

Για α=0.05 απορρίπτω H_0 αν $T > \chi^2_{4-1-1,0.95} = \chi^2_{2,0.95}$ =5,991 (έχουμε 4 κλάσεις και εκτιμήσει μία παράμετρο).

Εμείς έχουμε *T=6,036* και συνεπώς η μηδενική υπόθεση απορρίπτεται.

ΑΣΚΗΣΗ 19

Μία γεννήτρια τυχαίων αριθμών παράγει μονοψήφιους αριθμούς. Προκειμένου να ελεγχθεί, με την βοήθεια του χ^2 ελέγχου, αν όντως η γεννήτρια αυτή παράγει τυχαίους αριθμούς, καταγράφηκε ο αριθμός των εμφανίσεων κάθε ψηφίου σε 100 μονοψήφιους αριθμούς που παρήγαγε η γεννήτρια.

(Θέμα εξεταστικής)

ι) Με βάση τις εμφανίσεις των ψηφίων κατασκευάζεται ο ακόλουθος πίνακας συχνοτήτων από τον οποίο λείπουν στοιχεία τα οποία πρέπει να συμπληρώσετε:

$p_ι^0=1/10$ για $i=1,..,10$

$E_ι=n'* p_ι^0=100*1/10=10$ για $i=1,..,10$

Ψηφίο	Αριθμός εμφανίσεων	Αναμενόμενος αριθμός εμφανίσεων
0	6	10
1	6	10
2	14	10
3	11	10
4	12	10
5	8	10
6	8	10
7	6	10
8	14	10
9	15	10

ιι) Ποιο συμπέρασμα θα ήταν εύλογο για την γεννήτρια, σε επίπεδο σημαντικότητας α=0.05, αν η τιμή της ελεγχοσυνάρτησης του ελέγχου χ^2 είναι 11.8;

α) $\chi^2_{10,0.95}$=18.3070. Άρα, η γεννήτρια παράγει τυχαίους αριθμούς.

β) $\chi^2_{9,0.95}$=16.9190. Άρα, η γεννήτρια παράγει τυχαίους αριθμούς.

γ) $\chi^2_{9,0.05}$=3.3251. Άρα, η γεννήτρια δεν παράγει τυχαίους αριθμούς.

δ) $\chi^2_{10,0.05}$=3.9403. Άρα, η γεννήτρια δεν παράγει τυχαίους αριθμούς.

ΑΣΚΗΣΗ 20

Προκειμένου να ελεγχθεί, με την βοήθεια του χ^2 ελέγχου, αν ένα ζάρι είναι 'δίκαιο', καταγράφηκε ο αριθμός των εμφανίσεων κάθε σημείου σε 300 ρίψεις του ζαριού και τα αποτελέσματα παρουσιάζονται στον πίνακα που ακολουθεί.

α) Με βάση τις εμφανίσεις κάθε σημείου και τα στοιχεία που σας δίνονται να συμπληρώσετε τα στοιχεία που λείπουν από τον παραπάνω πίνακα.

p_i^0=1/6 για $i=1,..,6$

E_i= n'* p_i^0=300*1/6=50 για $i=1,..,6$

Σημείο	*Αριθμός εμφανίσεων*	*Αναμενόμενος αριθμός εμφανίσεων*
1	*43*	*50*
2	*57*	*50*
3	*61*	*50*
4	*54*	*50*
5	*38*	*50*
6	*47*	*50*

β) Σε επίπεδο σημαντικότητας α=0.05, σε ποιο συμπέρασμα οδηγήστε (είναι ‘δίκαιο’ το ζάρι ή όχι);

(Θέμα εξεταστικής)

Λύση

H0:Δίκαιο ζάρι
H1:Όχι δίκαιο ζάρι

$T=\sum_{i=1}^{10}\frac{(O_i-E_i)^2}{E_i}=\frac{(43-50)^2}{50}+\frac{(57-10)^2}{50}+\frac{(61-50)^2}{50}+\frac{(54-50)^2}{50}+\frac{(38-50)^2}{50}+\frac{(37-50)^2}{50}$=7.76.

$\chi^2_{n-1,0.95}=\chi^2_{5,0.95}$=11.075 (n=6 κλάσεις)

T=7,76<11.075= $\chi^2_{5,0.95}$ Άρα δεν απορρίπτω H0 άρα το ζάρι είναι δίκαιο.

ΑΣΚΗΣΗ 21

Το τυχαίο δείγμα 20 παρατηρήσεων μπορεί να αποτελεί δείγμα πάνω σε μία μεταβλητή Χ της οποίας η κατανομή είναι κανονική με μέση τιμή 50 κ διακύμανση 144:

37 41 42 44 44 44 46 46 48 49 49 53 56 57 60 61 62 66 67 80

(Ε. ΞΕΚΑΛΑΚΗ “Μη Παραμετρική Στατιστική” 2001)

ΛΥΣΗ

n'=20

H_0: F(x)=N(50,144)

H_1: F(x)≠N(50,144)

Θα δημιουργήσω 4 ισοπίθανες κλάσεις

Αν Z~N(0,1)=> X=μ+σZ

$x_a=\mu+\sigma z_a$ είναι το a ποσοστιαίο σημείο της $N(\mu,\sigma^2)$

Έτσι

$x_{0,25}=50+12(-0,67=z_{0,25})=41,96$ είναι το 0,25 ποσοστιαίο της $N(50,12^2)$

$x_{0,5}=50+12(0)=50$

$x_{0,75}=50+12(0.67)=58.04$

	<41.96	41.96-50	50-58.04	58.04<
E_i	5	5	5	5
O_i	2	9	3	6

$p_\iota^0=0.25$ για ι=1,..4

$E_\iota=n'\, p_\iota^0=20*0,25=5$ για ι=1,..4

$T=\sum_{i=1}^{4}\frac{(O_i-E_i)^2}{E_i}=\frac{9}{5}+\frac{16}{5}+\frac{4}{5}+\frac{1}{5}=6$ το οποίο είναι μικρότερο από

$\chi^2_{n-1,0.95}=\chi^2_{3,0.95}=7,8147$

Άρα δεν απορρίπτω H_0 σε α=0,05.

<u>ΑΣΚΗΣΗ 22</u>

Το τυχαίο δείγμα 20 παρατηρήσεων μπορεί να αποτελεί δείγμα πάνω σε μία μεταβλητή Χ της οποίας η κατανομή είναι κανονική ;

37 41 42 44 44 44 46 46 48 49 49 53 56 57 60 61 62 66 67 80

ΛΥΣΗ

n’=20

H_0: F(x)=Normal
H_1: F(x)≠Normal

$\bar{x} = \frac{\Sigma x_i}{n'}$ =52.6 $s^2 = \frac{\Sigma x_i^2 - n\bar{x}^2}{n'-1} = 117.3053$ =>$s=\sqrt{117.3053}$=10.83

$x_{0.25}=\bar{x}+sz_{0.25}$ =52.6+10.83*-0.67=45.344
$x_{0.5}=\bar{x}+sz_{0.5}$ =52.6+10.83*0=52.60
$x_{0.75}=\bar{x}+sz_{0.75}$ =52.6+10.83*0.67=59.8561

	<45.344	*45.344-52.6*	*52.6-59.8561*	*59.8561<*
E_i	*5*	*5*	*5*	*5*
O_i	*6*	*5*	*3*	*6*

p_ι^0=0.25 για ι=1,..4
$E_\iota=n'p_\iota^0$=20*0,25=5 για ι=1,..4

$T=\sum_{i=1}^{4}\frac{(O_i-E_i)^2}{E_i}=\frac{1}{5}+\frac{0}{5}+\frac{4}{5}+\frac{1}{5}$=1.2 το οποίο είναι μικρότερο από

$\chi^2_{n-k-1,0.95}$(όπου k=2 ο αριθμός των παραμέτρων (μ και σ) που έχουμε εκτιμήσει)

$= \chi^2_{4-2-1,0.95}=\chi^2_{1,0.95}$=3,8425

Άρα δεν απορρίπτω H_0 σε α=0,05

<u>ΑΣΚΗΣΗ 23</u>

Ο διευθυντής ενός υποκαταστήματος μιας τράπεζας θέλει να ελέγξει την υπόθεση ότι οι πελάτες της τράπεζας φθάνουν στο κατάστημα με τυχαίο τρόπο. Για το σκοπό αυτό κατέγραψε τους χρόνους που μεσολαβούν μεταξύ διαδοχικών αφίξεων πελατών κατά τη διάρκεια ενός πρωινού. Οι χρόνοι δίνονται παρακάτω

3,6 6,2 12,7 14,2 38 3,8 10,8 6,1 10,1 22,1 4,2 4,6 1,4 3,3 8,2

Όπως είναι γνωστό μία σειρά γεγονότων είναι τυχαία αν οι χρόνοι που μεσολαβούν μεταξύ διαδοχικών γεγονότων ακολουθούν την εκθετική κατανομή

$$f(x) = \lambda e^{-\lambda x}, \lambda>0, x>0$$

Τι συμπερασματολογία θα μπορούσατε να κάνετε με βάση τα στοιχεία? (Δημιουργείστε 3 ισοπίθανες κλάσεις).

(Ε. ΞΕΚΑΛΑΚΗ “Μη Παραμετρική Στατιστική” 2001)

ΛΥΣΗ

n’=15

Αν $f(x) = \lambda e^{-\lambda x}$ => $F(x) = \int_0^X f(x)dx = 1 - e^{-\lambda x}$

H_0: F(x)= $1 - e^{-\lambda x}$

H_1: F(x)≠ $1 - e^{-\lambda x}$

Ο μέσος της εκθετικής κατανομής (με παράμετρο λ) είναι λ^{-1}. Επομένως η εκτιμήτρια της παραμέτρου λ είναι

$$\widehat{\lambda} = \overline{X}^{-1} = \left(\frac{\Sigma X_i}{n'}\right)^{-1} = 9{,}9533^{-1} = 0{,}1005 \approx 0{,}1$$

Θα δημιουργήσω τρεις ισοπίθανες (p_ι^0=1/3) κλάσεις έτσι ώστε Eι= n' p_ι^0 =15*1/3
5 για ι=1,2,3.
Πρέπει να βρω τα 1/3 και 2/3 ποσοστιαία σημεία της exp(0.1).

****

Έστω $p=F(x_p)= 1-e^{-\lambda x_p}$ => $x_p=F^{-1}(p|\lambda)$=> ($p=1-e^{-\lambda x_p}$ =>$-\lambda x_p=\ln(1-p)$)

$$x_p=-\frac{\ln(1-p)}{\lambda}$$

δηλαδή το p ποσοστιαίο σημείο της exp(λ) ισούται με x_p

$$x_{1/3} = -\frac{\ln(1-1/3)}{0.1} = 4.0547 \Rightarrow P(X\leq 4.0547|\lambda=0{,}1)=1/3$$

$$x_{2/3} = -\frac{\ln(1-2/3)}{0.1} = 10.9861 \Rightarrow P(X\leq 10{,}9861|\lambda=0{,}1)=2/3$$

	<4,0547	4,0547-10,9861	10,9861<
E_i	5	5	5
O_i	4	7	4

p_ι^0=1/3 για ι=1,2,3

$E_\iota=n'p_\iota^0$=15*1/3=5 για ι=1,2,3

$$T=\sum_{i=1}^{3}\frac{(O_i-E_i)^2}{E_i}=\frac{1}{5}+\frac{4}{5}+\frac{1}{5}=1{,}2$$

Απορρίπτω την H_0 αν $T > \chi^2_{n-k-1,0.95} = \chi^2_{3-1-1,0.95} = \chi^2_{1,0.95}$ =3,84

Επειδή 1,2<3,84 δεν απορρίπτω την H_0.

ΕΛΕΓΧΟΙ KOLMOGOROV

ΑΣΚΗΣΗ 24

Έστω ότι από μία κατανομή F(x) παίρνουμε το εξής τυχαίο δείγμα

547 559 553 542 557 560.

Να ελεγχθεί η υπόθεση ότι το δείγμα μας προέρχεται από κανονική κατανομή μέσο 550 και διακύμανση 10^2.

Λυση

$$H_0: F(x)=N(550,100)(=F_0(x))$$

$$H_1: F(x)\neq N(550,100)$$

Κάτω από την H_0 η μεταβλητή $Z=\dfrac{X-550}{10} \sim N(0,1)$

ι	x_i	z_i	$\Phi(z_i)=F_0(x_i)$	$S(x_i)$	$F_0(x_i)-S(x_i)$	$F_0(x_i)-S(x_{i-1})$
1	542	-0.8	0.2129	0.1666	0.0452	0.2119
2	547	-0.3	0.3821	0.3333	0.0488	0.2154
3	553	0.3	0.6179	0.5000	0.1179	**0.2846**
4	557	0.7	0.7580	0.6667	0.0914	0.2580
5	559	0.9	0.8159	0.8323	-0.0174	0.1493
6	660	1	0.8413	1.0000	-0.1587	0.0080

Απορρόπτω H_0 αν $T>T_{0.95}$(για αμφίπλευρο έλεγχο)=0.519

Η ελεγχοσυνάρτηση είναι η

T=$\sup\limits_{x}\left(|\,F_0(x)-S(x)\,|\right)$=0,2846

Άρα εφόσον 0,2846<0,519=$T_{0,95}$ δεν απορρίπτω H_0.

(p-value)

$\hat{\alpha}$ =P(T≥0.2846|H_0)=1- P(T<0.2846|H_0)> 1- P(T<0.410|H_0)=1-0.8=0.20

<u>ΑΣΚΗΣΗ 25</u>

Έστω ότι από μία κατανομή F(x) παίρνουμε το εξής τυχαίο δείγμα

547 559 553 542 557 560.

Να ελεγχθεί η υπόθεση

H_0:F(x)≥N(550,100)(=F_0(x)) για κάθε x

H_1:F(x)<N(550,100) (=F_0(x)) για τουλάχιστον μία τιμή του x

<u>Λύση</u>

Κάτω από την H_0 η μεταβλητή Z=$\dfrac{X-550}{10}$~N(0,1)

ι	x_i	z_i	$\Phi(z_i)=F_0(x_i)$	$S(x_i)$	$F_0(x_i)$- $S(x_i)$	$F_0(x_i)$- $S(x_{i-1})$
1	542	-0.8	0.2129	0.1666	0.0452	0.2119
2	547	-0.3	0.3821	0.3333	0.0488	0.2154
3	553	0.3	0.6179	0.5000	0.1179	**0.2846**
4	557	0.7	0.7580	0.6667	0.0914	0.2580
5	559	0.9	0.8159	0.8323	-0.0174	0.1493
6	660	1	0.8413	1.0000	-0.1587	0.0080

Απορρίπτω H_0 αν T>$T_{0.95}$(για μονόπλευρο έλεγχο)=0.4680

Η ελεγχοσυνάρτηση είναι η

T=$\sup_x \left(F_0(x) - S(x)\right)$=0,2846

Άρα εφόσον 0,2846<0,4680=$T_{0,95}$ δεν απορρίπτω H_0.

---

p-value

$\hat{\alpha}$ =P(T≥0.2846|H_0)=1- P(T<0.2846|H_0)> 1- P(T<0.410|H_0)=1-0.9=0.10

<u>**ΑΣΚΗΣΗ 26**</u>

Έστω ότι από μία κατανομή F(x) παίρνουμε το εξής τυχαίο δείγμα

547 559 553 542 557 560.

Να ελεγχθεί η υπόθεση

H_0:F(x)≤N(550,100)(=F_0(x)) για κάθε x

H_1:F(x)>N(550,100) (=F_0(x)) για τουλάχιστον μία τιμή του x

<u>Λυση</u>

Κάτω από την H_0 η μεταβλητή Z=$\frac{X-550}{10}$~N(0,1)

ι	x_i	z_i	$\Phi(z_i)=F_0(x_i)$	$S(x_i)$	$S(x_i)-F_0(x_i)$
1	542	-0.8	0.2129	0.1666	-0.0452
2	547	-0.3	0.3821	0.3333	-0.0488
3	553	0.3	0.6179	0.5000	-0.1179
4	557	0.7	0.7580	0.6667	-0.0914
5	559	0.9	0.8159	0.8323	0.0174
6	660	1	0.8413	1.0000	0.1587

Απορρίπτω H_0 αν $T>T_{0.95}$(για μονόπλευρο έλεγχο)=0.4680

Η ελεγχοσυνάρτηση είναι η

T=$\sup\limits_{x}\left(S(x)-F_0(x)\right)$=0,1587

Άρα εφόσον 0,1587<0,4680=$T_{0,95}$ δεν απορρίπτω H_0.

ΕΛΕΓΧΟΙ LILLEFORS

ΑΣΚΗΣΗ 27

Έστω ότι από μία κατανομή F(x) παίρνουμε το εξής τυχαίο δείγμα

547 559 553 542 557 560.

Να ελεγχθεί η υπόθεση ότι το δείγμα μας προέρχεται από κανονική κατανομή

Λυση

H_0:F(x)=Normal

H_1:F(x)≠Normal

$\bar{x}=\dfrac{\sum x_i}{n}$ =553 s=$\sqrt{\dfrac{\sum\left(x_i-\bar{x}\right)^2}{n-1}}$ =7.183

$z_i=\dfrac{x_i-553}{7.189}$

ι	x_i	z_i	$\Phi(z_i)=F_0(x_i)$	$S(x_i)$	$F_0(x_i)- S(x_i)$	$F_0(x_i)- S(x_{i-1})$
1	542	-1.531	0.0628	0.1666	-0.1038	0.0628
2	547	-0.835	0.2018	0.3333	-0.1316	0.0351
3	553	0	0.500	0.5000	0	0.1667
4	557	0.556	0.711	0.6667	0.0445	0.2112
5	559	0.835	0.798	0.8323	-0.0351	0.1316
6	660	974	0.835	1.0000	0.1649	0.0018

$T_1 = \sup_x\left(|\,F_0(x) - S(x)\,|\right) = 0{,}2112$

Απορρίπτω H_0 αν $T_1 > T_{0.95}$(πίνακες για κανονικότητα LILLEFORS)=0.319

=> Δεν απορρίπτω H_0 αφού *0,2112<0,319*

<u>ΑΣΚΗΣΗ 28</u>

Έστω ότι από μία κατανομή F(x) παίρνουμε το εξής τυχαίο δείγμα

1.15 1.34 4.12 4.82 8.39 16.84

Να ελεγχθεί η υπόθεση ότι το δείγμα μας προέρχεται από μία εκθετική κατανομή.

<u>Λύση</u>

$$H_0: F(x)=1-e^{-\lambda x} \text{ για } x>0$$
$$H_1: F(x)\neq 1-e^{-\lambda x}$$

$\bar{x} = \frac{\sum x_i}{n}$ =6.11

ι	x_i	$z_i = x_i/\bar{x}$	$F_0(z_i)=1-e^{-zi}$	$S(z_i)$	$F_0(z_i)- S(z_i)$	$F_0(z_i)- S(z_{i-1})$
1	1.15	0.188	0.1716	0.1666	0.0049	0.1716
2	1.34	0.219	0.1969	0.3333	-0.136	0.0303
3	4.12	0.674	0.4905	0.5000	-0.009	0.157
4	4.82	0.788	0.545	0.6667	-0.1210	0.045
5	8.39	1.373	0.746	0.8323	-0.086	0.08
6	16.84	2.756	0.936	1.0000	-0.06	0.1031

$T_1 = \sup_x \left(| F_0(z) - S(z) | \right) = 0,1716$

Απορρίπτω H_0 αν $T_1 > T_{0.95}$(πίνακες για εκθετικότητα LILLEFORS)=0.4085

=> Δεν απορρίπτω H_0 αφού *0,1716<0,4085*

<u>ΕΛΕΓΧΟΙ ΤΥΧΑΙΟΤΗΤΑΣ</u>

<u>ΑΣΚΗΣΗ 29</u>

Παραδειγμα 4.4.1

(Ε. ΞΕΚΑΛΑΚΗ “Μη Παραμετρική Στατιστική” 2001)

<u>ΑΣΚΗΣΗ 30</u>

Ένας μετεωρολόγος διεξάγει ένα πείραμα για να προσδιορίσει αν τα επίπεδα υγρασίας, που κατεγράφησαν στις 1200 το μεσημέρι για 20 διαδοχικές ημέρες τον Ιούλιο του 2000, κατανέμονται με τυχαίο τρόπο ως προς το εάν αυτά βρίσκονται πάνω η κάτω από το μέσο επίπεδο υγρασίας που καταγράφη κατά τη διάρκεια του Ιουλίου των ετών 1995-1999. Συμβολίζοντας με «+» τα επίπεδα υγρασίας που ήταν υπεράνω του μέσου επιπέδου υγρασίας του Ιουλίου και με «-» αυτά που ήταν κάτω από το μεσο επίπεδο, οι 20 συγκρίσεις έδωσαν τα εξής αποτελέσματα:

+ + + - - - + + - - + - + - + - - - + +

(Ε. ΞΕΚΑΛΑΚΗ “Μη Παραμετρική Στατιστική” 2001)

<u>Λύση</u>

H_0: Τα δύο είδη συμβόλων εμφανίζονται με τυχαία σειρά

H_1: Τα δύο είδη συμβόλων <u>δεν</u> εμφανίζονται με τυχαία σειρά

+ + +	- - -	+ +	- -	+	-	+	-	+	- - -	+ +
1	2	3	4	5	6	7	8	9	10	11

n_1= # “+”=10

n_2= # “-”=10

T= # αριθμός ροών

Απορρίπτω H_0 αν $T\leq w_{0.025,}n_1,n_2=6$ ή $T\geq w_{0.025,}n_1,n_2=16$

Εμείς έχουμε 11 ροές άρα δεν απορρίπτω H_0

<u>ΠΙΝΑΚΕΣ ΣΥΝΑΦΕΙΑΣ</u>

<u>ΑΣΚΗΣΗ 31</u>

Ένα τυχαίο δείγμα μαθητών από ιδιωτικά σχολεία και ένα τυχαίο δείγμα μαθητών από δημόσια σχολεία δίνει πανελλήνιες εξετάσεις με τα ακόλουθα αποτελέσματα

	0-275	276-350	351-425	426-500	
ΙΔΙΩΤΙΚΟ	6	14	17	9	$46=n_1$
ΔΗΜΟΣΙΟ	30	32	17	3	$82=n_2$
	$36(=C_1)$	$46(=C_2)$	$34(=C_3)$	$12(=C_4)$	128

Ελέγξτε τη μηδενική υπόθεση ότι η κατανομή της βαθμολογίας είναι η ίδια για μαθητές ιδιωτικών και δημόσιων σχολείων.

Λύση

H_0: Οι πιθανότητες σε κάθε στήλη είναι ίσες μεταξύ τους

ή ($p_{1j}=p_{2j},$ για όλα τα j)

H_1: Τουλάχιστον σε μία στήλη οι πιθανότητες δεν είναι ίσες μεταξύ τους

($p_{1j}=p_{2j},$ για κάποια j)

$$E_{ij}=\frac{n_i C_j}{N} \Rightarrow E_{11}=\frac{n_1 C_1}{N}=\frac{46*36}{128}=12.9$$

$$E_{12}=\frac{n_1 C_2}{N}=\frac{46*46}{128}=16.5$$

$$E_{13}=..........$$

…………..

$E_{ij}=$

12.9	16.5	12.2	4.3
23.1	29.5	21.8	7.7

$$T = \sum_{i=1}^{2}\sum_{j=1}^{4}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} = \frac{(O_{11}-E_{11})^2}{E_{11}} + \frac{(O_{12}-E_{12})^2}{E_{12}} + ... + \frac{(O_{21}-E_{21})^2}{E_{21}} + ... =$$

$$\frac{(6-12.9)^2}{12.9} + \frac{(14-16.5)^2}{16.5} + ... + \frac{(30-23.1)^2}{23.1} + ... =$$

3.69+0.38+1.89+5.14+2.06+0.21+1.06+2.87=17.3

Απορρίπτω H_0 αν $T \geq X^2_{0.95,(r-1)(c-1)} = X^2_{0.95,3} = 7,815$

(r-1)(c-1)= (2-1)(4-1)=3

Άρα εφόσον 17,3>7,815 απορρίπτω H_0 σε α=0,05.

<u>ΑΣΚΗΣΗ 32</u>

Ένα τυχαίο δείγμα φοιτητών ενός συγκεκριμένου πανεπιστημίου των ΗΠΑ ταξινομείται ανάλογα με τη σχολή στην οποία φοιτούν καθώς και ανάλογα με το αν αποφοίτησαν από λύκειο της πολιτείας στην οποία βρίσκεται το πανεπιστήμιο ή όχι.

	ΤΕΧΝΟΛΟΓ.	ΝΟΜΙΚΗ	ΟΙΚΟΝΟΜΙΚΗ	ΑΛΛΟ	ΣΥΝΟΛΟ
ΕΝΤΟΣ ΠΟΛΙΤΕΙΑΣ	16	14	13	13	56(=R_1)
ΕΚΤΟΣ ΠΟΛΙΤΕΙΑΣ	14	6	10	8	38(=R_2)
ΣΥΝΟΛΟ	30(=C_1)	20(=C_2)	23(=C_3)	2123(=C_4)	94(=N)

Ελέγξτε σε επίπεδο σημαντικότητας 0,05 την υπόθεση ότι το τι σπουδάζει κάποιος φοιτητής στο συγκεκριμένο πανεπιστήμιο δεν εξαρτάται από το αν ο φοιτητής πήγε σε λύκειο της συγκεκριμένης πολιτείας.
(Conover W.J. “Practical Nonparametric Statistics” 3rd edition)

ΛΥΣΗ

H_0:p_{ij}=$p_i p_j$ για κάθε i και j

H_1:p_{ij}≠$p_i p_j$ για κάποια i , j

$$E_{ij} = \frac{R_i C_j}{N} \Rightarrow E_{11} = \frac{R_1 C_1}{N} = \frac{56*30}{94} = 17.872$$

$$E_{12} = \frac{R_1 C_2}{N} = \frac{56*20}{94} = 11.915$$

$$E_{13} = ...$$

17,872	11,915	13,702	12,511
12,128	8,0851	9,2979	8,4894

…………..

$E_{ij} =$

$$T = \sum_{i=1}^{2}\sum_{j=1}^{4}\frac{(O_{ij}-E_{ij})^2}{E_{ij}} = \frac{(O_{11}-E_{11})^2}{E_{11}} + \frac{(O_{12}-E_{12})^2}{E_{12}} + ... + \frac{(O_{21}-E_{21})^2}{E_{21}} + ... =$$

$$\frac{(16-17.872)^2}{17.872} + \frac{(14-11.915)^2}{11.915} + ... + \frac{(14-12.128)^2}{12.128} + ... =$$

0.1961+0.3649+0.0360+0.0191+0.2891+0.5377+0.0530+0.0282=1.5241

Απορρίπτω H_0 αν $T > X^2_{0.95,(r-1)(c-1)}$

$X^2_{0.95,(r-1)(c-1)} = X^2_{0.95,(2-1)(4-1)} = X^2_{0.95,(3)}$ =*7,815*

Εμείς έχουμε T=1,5241<7,815 άρα δεν απορρίπτω H_0

<u>ΑΣΚΗΣΗ 33</u>

Παραδειγμα 8.1.3
(Ε. ΞΕΚΑΛΑΚΗ "Μη Παραμετρική Στατιστική" 2001)

<u>ΜΗ ΠΑΡΑΜΕΤΡΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ</u>

<u>ΑΣΚΗΣΗ 34</u>

Στον παρακάτω πίνακα καταγράφονται το GMAT-score και ο βαθμός του μεταπτυχιακού διπλώματος 12 αποφοίτων MBA.

GMAT(X)	MBA-Grade(Y)
710	4
610	4
640	3.9
580	3.8
545	3.7
560	3.6
610	3.5
530	3.5
560	3.5
540	3.3
570	3.2
560	3.2

α. Εκτιμείστε την ευθεία παλινδρόμησης ελαχίστων τετραγώνων μεταξύ Gmat και βαθμού μεταπτυχιακού (χρησιμοποιείστε σαν εξαρτημένη μεταβλητή το βαθμό του μεταπτυχιακού.).
β. Μια έρευνα αναφέρει ότι μία αύξηση 40 πόντων στο GMAT-score έχει ως αποτέλεσμα μία αύξηση τουλάχιστον 0.4 στο βαθμό του πτυχίου. Ελέγξτε αν τα

δεδομένα από το δείγμα μας των 12 αποφοίτων συνηγορούν υπέρ αυτής της υπόθεσης.
(Conover W.J. “Practical Nonparametric Statistics” 3^{rd} edition)

ΛΥΣΗ

Α) Συμβολίζω με $Y_ι$ το βαθμό του πτυχίου και με $X_ι$ το GMAT-score.
Από τα δεδομένα έχουμε

$$\sum_{i=1}^{12} x_i = 7015 \quad \sum_{i=1}^{12} x_i^2 = 4129525 \quad \bar{x} = 584.58$$

$$\sum_{i=1}^{12} y_i = 43.2 \quad \sum_{i=1}^{12} x_i y_i = 25360.5 \quad \bar{y} = 3.6$$

$$b = \frac{\sum_{i=1}^{12}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{12}(x_i - \bar{x})^2} = \frac{n\sum_{i=1}^{12} x_i y_i - \sum_{i=1}^{12}(x_i)\sum_{i=1}^{12}(y_i)}{n\sum_{i=1}^{12} x_i^2 - \left(\sum_{i=1}^{12} x_i\right)^2} = \frac{12*25360-43.2*7015}{12*4129525-7015^2}$$

$= 0.003714$

a= $\bar{y}$ -b $\bar{x}$ $=$3.6-0.003714*584.58=1.428

$$\hat{y} = 1.428+0.003714x$$

Β) Επειδή η κλίση μιας συνάρτησης *f(x)* ορίζεται ως *d(f(x))/dx* η έρευνα υποδηλώνει ότι η κλίση στη παλινδρόμηση είναι τουλάχιστον 0.4/40=0.01. Άρα πρέπει να ελέγξουμε

$$H_0: β \geq 0.01$$
$$H_1: β < 0.01$$

Για να πραγματοποιήσουμε τον παραπάνω έλεγχο υπολογίζουμε τον Spearman συντελεστή συσχέτισης μεταξύ X_i και $U_i=Y_i-0.01X_i$

X_i	U_i $(=Y_i-0.01X_i)$	$R(X_i)$	$R(U_i)$
710	-3.1	12	1
610	-2.1	9.5	7
640	-2.5	11	3.5
580	-2.0	8	9.5
545	-1.75	3	12
560	-2.0	5	9.5
610	-2.6	9.5	2
530	-1.8	1	11
560	-2.1	5	7
540	-2.1	2	7
570	-2.5	7	3.5
560	-2.4	5	5

$$\sum_{i=1}^{12} R(X_i)R(U_i) = 405 \quad \sum_{i=1}^{12} R(X_i)^2 = 647.5 \quad \sum_{i=1}^{12} R(U_i)^2 = 647$$

$$\rho = \frac{\sum_{i=1}^{12} R(X_i)R(U_i) - n\left(\frac{n+1}{2}\right)^2}{\left[\sum_{i=1}^{12} R(X_i)^2 - n\left(\frac{n+1}{2}\right)^2\right]^{1/2}\left[\sum_{i=1}^{12} R(U_i)^2 - n\left(\frac{n+1}{2}\right)^2\right]^{1/2}} = -0.7273$$

Απορρίπτω την H_0 αν $\rho=-0{,}7273<w_{0.05}= -0.4965$ πράγμα που ισχύει.

Άρα για α=0,05 απορρίπτω την H_0. Άρα μπορούμε να πούμε ότι αποτελέσματα που προκύπτουν από το δείγμα μας δε συμπίπτουν με τα αποτελέσματα της έρευνας.

ΜΗ ΠΑΡΑΜΕΤΡΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

Μέθοδος του THEIL

ΑΣΚΗΣΗ 35

Ερευνητής ενδιαφέρεται να μελετήσει τον βαθμό ευφυΐας των παιδιών των οικογενειών με 2 αγόρια . Για το σκοπό αυτό επιλέχτηκε τυχαίο δείγμα 5 οικογενειών και καταγράφηκε ο βαθμός σε ένα τεστ IQ των δύο παιδιών. Τα αποτελέσματα καταγράφηκαν στον παρακάτω πίνακα.

	IQ $1^{ου}$ παιδιού	IQ $2^{ου}$ παιδιού
1	110	115
2	108	103
3	120	112
4	105	120
5	95	88

Να εκτιμήσετε την εξίσωση παλινδρόμησης του THEIL.(Να χρησιμοποιήσετε σαν εξαρτημένη μεταβλητή τον βαθμό IQ του $2^{ου}$ παιδιού.).
(Θέμα εξεταστικής Αυγούστου 2010)

ΛΥΣΗ

Υπολογίζω σχετικές κλίσεις

$S_{ij} = \frac{Y_j - Y_i}{X_j - X_i}, i < j =>$

$$S_{12} = \frac{103-115}{108-110} = \frac{-12}{-2} = 6,$$

$$S_{13} = \frac{112-115}{120-110} = \frac{-3}{10} = -0.3$$

$$S_{14} = \frac{120-115}{105-110} = \frac{5}{-5} = -1$$

$$S_{15} = \frac{88-115}{95-110} = \frac{-27}{-15} = 1.8$$

$$S_{23} = \frac{112-103}{120-108} = \frac{9}{12} = 0.75$$

$$S_{24} = \frac{120-103}{105-108} = \frac{17}{-3} = -5.67$$

$$S_{25} = ... = 1.15$$

$$S_{34} = ... = -0.53$$

$$S_{35} = ... = 0.96$$

$$S_{45} = ... = 3.2$$

Διατάσσω τα S_{ij} κατά αύξουσα σειρά

-5.67 -1 -0.53 -0.3 0.75 0.96 1.15 1.8 3.2 6

και έχω $\widetilde{\beta} = median(S_{ij}) = \dfrac{0.75+0.96}{2} = 0{,}855$

$$\widetilde{\alpha} = median(y_i - \widetilde{\beta}x_i)$$

$$y_1 - \widetilde{\beta}x_1 = 115 - 0.855 * 110 = 20.95$$

$$y_2 - \widetilde{\beta}x_2 = 103 - 0.855 * 108 = 10.66$$

$$y_3 - \widetilde{\beta}x_3 = ... = 9.4$$

$$y_4 - \widetilde{\beta}x_4 = ... = 30.225$$

$$y_5 - \widetilde{\beta}x_5 = ... = 6.775$$

Διατάσσω τα $(y_i - \widetilde{\beta}x_i)$ κατά αύξουσα σειρά: 6.775 9.4 10.66 20.95 30.225

$$\widetilde{\alpha} = median(y_i - \widetilde{\beta}x_i) = 10.66$$

$$\widetilde{y} = 10.66 + 0.855x$$