Θέμα 1ο
Α. Αν σ ένα στατιστικό έλεγχο υποθέσεων, έχουμε υπολογίσει το pvalue και με βάση αυτό η μηδενική υπόθεση απορρίπτεται για επίπεδο σημαντικότητας 5%, τότε απορρίπτεται ή όχι και για επίπεδο σημαντικότητας 10%; Δικαιολογήστε την απάντησή σας
Απάντηση

Το pvalue είναι η πιθανότητα να παρατηρήσουμε μια τιμή της στατιστικής δοκιμής τουλάχιστον τόσο ακραία όσο η παρατηρούμενη, δεδομένου ότι η μηδενική υπόθεση είναι αληθής. Όταν συγκρίνουμε το pvalue με το επίπεδο σημαντικότητας α, έχουμε τα εξής:

- Εάν το pvalue είναι μικρότερο ή ίσο με α, απορρίπτουμε τη μηδενική υπόθεση H0.
- Εάν το pvalue είναι μεγαλύτερο από α, δεν απορρίπτουμε τη μηδενική υπόθεση H0.

Τώρα, αν υποθέσουμε ότι έχουμε υπολογίσει το pvalue και βάσει αυτού η μηδενική υπόθεση απορρίπτεται για επίπεδο σημαντικότητας 5% (α=0.05\), αυτό σημαίνει ότι:

p-value≤0.05

Τώρα, εξετάζουμε εάν η μηδενική υπόθεση θα απορριπτόταν και για επίπεδο σημαντικότητας 10% (α=0.10):

Αν το pvalue είναι μικρότερο ή ίσο με 0.05, τότε είναι επίσης μικρότερο ή ίσο με 0.10, καθώς:

p-value≤0.05 ⇒ p-value≤0.10

Άρα, αν απορρίπτουμε τη μηδενική υπόθεση σε επίπεδο σημαντικότητας 5%, σίγουρα θα την απορρίψουμε και σε επίπεδο σημαντικότητας 10%.

Συμπέρασμα: Ναι, η μηδενική υπόθεση **θα απορριφθεί** και για επίπεδο σημαντικότητας 10%, εφόσον έχει ήδη απορριφθεί για επίπεδο σημαντικότητας 5%.

Β. Σε 10 γραμμές το πολύ εξηγείστε ποια τεχνική δειγματοληψίας θα χρησιμοποιούσατε και γιατί, εάν θέλετε να ελέγξετε το βάρος των παιδιών σε ένα δημοτικό σχολείο
Απάντηση
Για να ελέγξουμε το βάρος των παιδιών σε ένα δημοτικό σχολείο, θα χρησιμοποιούσα τη **στρωματοποιημένη δειγματοληψία**. Σε αυτή την τεχνική, ο πληθυσμός (δημοτικό σχολείο) χωρίζεται σε στρώματα (τάξεις ή ηλικιακές ομάδες) και από κάθε στρώμα επιλέγουμε ένα τυχαίο δείγμα. Αυτή η μέθοδος εξασφαλίζει ότι κάθε τάξη ή ηλικιακή ομάδα εκπροσωπείται στο δείγμα, δίνοντας πιο ακριβή και αντιπροσωπευτικά αποτελέσματα για το βάρος των παιδιών σε όλο το σχολείο, λαμβάνοντας υπόψη την ποικιλία του πληθυσμού.
Γ.Αναφέρεται την ερμηνεία του συντελεστή εμπιστοσύνης τόσο σε ένα διάστημα εμπιστοσύνης, όσο και στον έλεγχο υποθέσεων
Απάντηση

Συντελεστής Εμπιστοσύνης σε Διάστημα Εμπιστοσύνης

Ο συντελεστής εμπιστοσύνης (συχνά εκφραζόμενος ως ποσοστό, π.χ. 95%) σε ένα διάστημα εμπιστοσύνης είναι η πιθανότητα ότι το πραγματικό παράμετρο του πληθυσμού (π.χ., η μέση τιμή ή η αναλογία) βρίσκεται εντός του υπολογισμένου διαστήματος εμπιστοσύνης. Ένα διάστημα εμπιστοσύνης 95% σημαίνει ότι, αν υπολογίσουμε διαστήματα εμπιστοσύνης από πολλά δείγματα, το 95% αυτών των διαστημάτων θα περιέχουν την πραγματική παράμετρο του πληθυσμού. Δεν σημαίνει ότι υπάρχει 95% πιθανότητα η συγκεκριμένη τιμή της παραμέτρου να βρίσκεται στο συγκεκριμένο διάστημα που υπολογίστηκε από ένα συγκεκριμένο δείγμα.

Συντελεστής Εμπιστοσύνης στον Έλεγχο Υποθέσεων

Στον έλεγχο υποθέσεων, ο συντελεστής εμπιστοσύνης συμπληρώνει το επίπεδο σημαντικότητας (α). Αν το επίπεδο σημαντικότητας είναι 5% (α = 0.05), ο συντελεστής εμπιστοσύνης είναι 95%. Αυτό υποδηλώνει ότι υπάρχει 95% πιθανότητα να μην απορρίψουμε τη μηδενική υπόθεση όταν αυτή είναι αληθής (δηλαδή, 95% πιθανότητα να μην κάνουμε σφάλμα τύπου I). Στην πράξη, αν το p-value είναι μικρότερο από το επίπεδο σημαντικότητας (π.χ. 0.05), τότε απορρίπτουμε τη μηδενική υπόθεση.

Δ. Τα τελευταία 15 χρόνια καταγράφηκαν 20 θαλάσσιοι πνιγμοί σε ένα πολυσύχναστο beach bar της Χαλκιδικής. Ο χρόνος που μεσολαβεί μεταξύ διαδοχικών πνιγμών σε αυτό το beach bar μπορεί να περιγραφεί από την …
(i) εκθετική κατανομή.
(ii) γεωμετρική κατανομή.
(iii) κανονική κατανομή.
(iv) κατανομή Poisson **.**
<u>Απάντηση</u>
(i) **εκθετική κατανομή**.
Η εκθετική κατανομή είναι κατάλληλη για να περιγράψει τον χρόνο που μεσολαβεί μεταξύ διαδοχικών ανεξάρτητων γεγονότων που συμβαίνουν με σταθερό μέσο ρυθμό.

Ε. Δώστε τους ακόλουθους ορισμούς: α) ανεξάρτητα γεγονότα β) συμπληρωματικά γεγονότα, κάνοντας σχόλιο για το πλήθος των δοκιμών που αφορούν (ή που σχετίζονται με αυτά)
<u>Απάντηση</u>

α) Ανεξάρτητα Γεγονότα

Ορισμός: Δύο γεγονότα AAA και BBB είναι ανεξάρτητα αν η εμφάνιση του ενός δεν επηρεάζει την πιθανότητα εμφάνισης του άλλου. Δηλαδή, η πιθανότητα του AAA και του BBB να συμβούν μαζί ισούται με το γινόμενο των πιθανοτήτων τους:

P(A∩B)=P(A)⋅P(B)P(A \cap B) = P(A) \cdot P(B)P(A∩B)=P(A)⋅P(B)

Σχόλιο για το πλήθος των δοκιμών: Η έννοια της ανεξαρτησίας είναι θεμελιώδης και δεν εξαρτάται από το πλήθος των δοκιμών. Η ανεξαρτησία μπορεί να ισχύει σε μεμονωμένες δοκιμές ή σε πολλαπλές δοκιμές ενός πειράματος, και αναφέρεται στη σχέση μεταξύ των γεγονότων σε οποιοδήποτε αριθμό δοκιμών.

β) Συμπληρωματικά Γεγονότα

Ορισμός: Δύο γεγονότα AAA και A′A'A′ (όπου A′A'A′ είναι το συμπλήρωμα του AAA) είναι συμπληρωματικά αν η εμφάνιση του ενός αποκλείει την εμφάνιση του άλλου και το άθροισμα των πιθανοτήτων τους είναι 1. Δηλαδή:

$P(A)+P(A')=1P(A) + P(A') = 1P(A)+P(A')=1$

Σχόλιο για το πλήθος των δοκιμών: Τα συμπληρωματικά γεγονότα αναφέρονται σε ένα συγκεκριμένο σύνολο δοκιμών ή σε ένα συγκεκριμένο πείραμα. Σε μία μόνο δοκιμή, ένα γεγονός είτε συμβαίνει είτε όχι, καλύπτοντας έτσι όλες τις πιθανότητες. Για παράδειγμα, σε μια ρίψη ενός νομίσματος, το γεγονός "κεφάλι" και το γεγονός "γράμματα" είναι συμπληρωματικά, καθώς το νόμισμα θα δείξει είτε κεφάλι είτε γράμματα σε κάθε δοκιμή.

Θέμα 2ο

Στο πλαίσιο μιας έρευνας σχετικά με τις συνήθειες των φοιτητών και διαπιστώθηκε ότι οι ώρες ανάπαυσής τους ημερησίως ακολουθούν κανονική κατανομή με μέση τιμή 7,389 ώρες και τυπική απόκλιση 1,67 ώρες.

α) Ποια είναι η πιθανότητα οι ώρες ανάπαυσης ενός φοιτητή να μην ξεπερνούν τις 4,5 ώρες;

<u>Απάντηση</u>

$Z=(X-\mu)/\sigma$

Για $X=4.5$:

$Z=(4.5-7.389)/1.67=-2.889/1.67\approx-1.73$

Τώρα, βρίσκουμε την πιθανότητα $P(Z\leq-1.73)$ χρησιμοποιώντας πίνακες της τυπικής κανονικής κατανομής:

$P(Z\leq-1.73)\approx0.0418$

Άρα, η πιθανότητα οι ώρες ανάπαυσης ενός φοιτητή να μην ξεπερνούν τις 4,5 ώρες είναι περίπου 0.0418 ή 4.18%.

β) Βρείτε το πλήθος των ωρών πάνω από τις οποίες αναπαύεται μόνο το 5% των φοιτητών.

<u>Απάντηση</u>

Αναζητούμε την τιμή X για την οποία $P(X>x0)=0.05$. Αυτό ισοδυναμεί με $P(X\leq x0)=0.95$ γιατί $P(X>x0)=1-P(X\leq x0)=1-0,95=0,05$

Χρησιμοποιούμε τον πίνακα της τυπικής κανονικής κατανομής για να βρούμε το αντίστοιχο Z-score για $P(Z\leq z0)=0.95$:

$Z0.95\approx1.645$

Τώρα, μετατρέπουμε το Z-score στην αρχική κατανομή:

$x0=\mu+Z0.95\cdot\sigma$

$x0=7.389+1.645\cdot1.67$

$x0 \approx 10.136$

Άρα, μόνο το 5% των φοιτητών αναπαύεται πάνω από 10.136 ώρες.

γ) Ποια είναι η πιθανότητα η δειγματική μέση τιμή των ωρών ανάπαυσης ενός τυχαίου δείγματος 9 φοιτητών να είναι μεταξύ 6 και 8 ωρών;
<u>Απάντηση</u>

Η δειγματική μέση τιμή X ενός δείγματος μεγέθους n=9 φοιτητών ακολουθεί κανονική κατανομή με μέση τιμή μ=7.389 και τυπική απόκλιση $\sigma X^{-}=\sigma/sqr(n)$

$\sigma X^{-}=1.67/sqr(9)=1.67/3\approx 0.5567$

Χρησιμοποιούμε τον τυπικό μετασχηματισμό Z:

Για $X^{-}=6$

$Z=6-7.3890.5567=-1.3890.5567\approx -2.496Z$

Για $X^{-}=8$:

$Z=8-7.3890.5567=0.6110.5567\approx 1.098Z$

Βρίσκουμε τις πιθανότητες από τους πίνακες της τυπικής κανονικής κατανομής ή έναν υπολογιστή:

$P(Z\leq -2.496)\approx 0.0063$

$P(Z\leq 1.098)\approx 0.8643$

Η πιθανότητα η δειγματική μέση τιμή να είναι μεταξύ 6 και 8 ωρών είναι:

$P(-2.496\leq Z\leq 1.098)=P(Z\leq 1.098)-P(Z\leq -2.496)$

$=0.8643-0.0063=0.8580$

Άρα, η πιθανότητα η δειγματική μέση τιμή των ωρών ανάπαυσης ενός τυχαίου δείγματος 9 φοιτητών να είναι μεταξύ 6 και 8 ωρών είναι περίπο**υ 0.8580 ή 85.80%.**

Θέμα 3ο
Μια μηχανή συσκευάζει καλαμπόκι σε τσουβάλια των 25 Kgr και έχει ρυθμιστεί έτσι ώστε οι ποσότητες καλαμποκιού ανά τσουβάλι να έχουν τυπική απόκλιση 1,5 Kgr. Ο υπεύθυνος παραγωγής γνωρίζει ότι οι ποσότητες καλαμποκιού ανά τσουβάλι ακολουθούν κανονική κατανομή. Επειδή υποψιάζεται ότι η μηχανή έχει απορυθμιστεί τελευταία και θέλει να ελέγξει αν όντως η τυπική απόκλιση των ποσοτήτων καλαμποκιού ανά τσουβάλι είναι 1,5 Kgr επέλεξε τις προάλλες από την παραγωγή 30 τσουβάλια και κατέγραψε τα βάρη τους: υπολόγισε δε ότι το μέσο βάρος τους ήταν 24,8 Kgr και η τυπική τους απόκλιση 1,6 Kgr.
α) Σε επίπεδο σημαντικότητας 5%, διαπιστώνετε ότι είναι βάσιμες οι υποψίες του υπεύθυνου παραγωγής;
<u>Απάντηση</u>

Για να ελέγξουμε αν η τυπική απόκλιση των ποσοτήτων καλαμποκιού ανά τσουβάλι είναι 1,5 Kgr, χρησιμοποιούμε τον έλεγχο χ^2 για την διακύμανση. Ορίζουμε τις υποθέσεις μας ως εξής:

- Μηδενική υπόθεση (H0): $\sigma=1.5$
- Εναλλακτική υπόθεση (H1): $\sigma\neq1.5$

Χρησιμοποιούμε τον εξής τύπο για τον έλεγχο χ^2:

$\chi^2=((n-1)s^2)/\sigma_0^2$

όπου:

- $n=30$ (μέγεθος δείγματος)
- $s=1.6$ Kgr (τυπική απόκλιση δείγματος)
- $\sigma_0=1.5$ Kgr (υποτιθέμενη τυπική απόκλιση)

Υπολογίζουμε το χ^2 ως εξής:

$\chi2=[(30-1)\cdot(1.6)^2]/(1.5)^2=(29*2.56)/2.25\approx74.24/2.25\approx32.993$

Η κατανομή χ^2 με 29 βαθμούς ελευθερίας (n-1) έχει κριτικές τιμές που μπορούμε να βρούμε από τους πίνακες χ^2. Για επίπεδο σημαντικότητας 5% ($\alpha=0.05$) σε διπλής όψης έλεγχο:

- Κατώτερη κριτική τιμή (2.5%): $\chi^2(0.025,29)\approx16.047$
- Ανώτερη κριτική τιμή (2.5%): $\chi^2(0.975,292)\approx45.722$

Εφόσον η υπολογιζόμενη τιμή χ^2 (32.993) βρίσκεται μεταξύ των κριτικών τιμών (16.047 και 45.722), δεν απορρίπτουμε τη μηδενική υπόθεση H0.

Συμπέρασμα: Σε επίπεδο σημαντικότητας 5%, δεν υπάρχουν αρκετά στοιχεία για να υποστηριχθεί ότι η τυπική απόκλιση έχει αλλάξει από 1.5 Kgr.

β) Να βρεθεί και να ερμηνευθεί το 98% διάστημα εμπιστοσύνης της μεταβλητότητας των ποσοτήτων καλαμποκιού που συσκευάζονται ανά τσουβάλι
Απάντηση

Για να βρούμε το 98% διάστημα εμπιστοσύνης για τη διακύμανση, χρησιμοποιούμε τους πίνακες χ^2 για να βρούμε τις κριτικές τιμές χ^2 για $\alpha=0.02$ (το οποίο δίνει $\alpha/2=0.01$ για τις ακραίες τιμές).

Οι κριτικές τιμές για χ^2 με 29 βαθμούς ελευθερίας είναι:

- $\chi^2(0.01,292)\approx13.121$
- $\chi^2(0.99,292)\approx50.892$

Το διάστημα εμπιστοσύνης για τη διακύμανση (σ^2) δίνεται από τον τύπο:

$[((n-1)s^2)/\chi^2(0.99,n-1),((n-1)s^2)/\chi^2(0.01,n-1)]$

Υπολογίζουμε τα όρια του διαστήματος:

$((29 \cdot (1.6)^2)/50.892,(29 \cdot (1.6)^2)/13.121)$

$(74.24/50.892,74.24/13.121)$

$(1.459,5.656)$

Το διάστημα εμπιστοσύνης για την τυπική απόκλιση (σ^2) είναι το τετράγωνο ρίζας του διαστήματος για τη διακύμανση:

(sqr 1.459, sqr5.656)

$(1.209,2.377)$

Συμπέρασμα: Το 98% διάστημα εμπιστοσύνης για την τυπική απόκλιση της ποσότητας καλαμποκιού ανά τσουβάλι είναι από 1.209 Kgr έως 2.377 Kgr. Αυτό σημαίνει ότι είμαστε 98% βέβαιοι ότι η πραγματική τυπική απόκλιση βρίσκεται εντός αυτού του διαστήματος.