ΒΑΣΙΚΟ ΠΑΡΑΔΕΙΓΜΑ 1:

Το όριο αντοχής ενός τύπου καλωδίων είναι τυχαία μεταβλητή X με μέση τιμή $\mu = 1500$ kgr και τυπική απόκλιση $\sigma = 175$ kgr. Το εργοστάσιο που τα κατασκευάζει, ισχυρίζεται ότι βελτίωσε τα υλικά που χρησιμοποιεί και πλέον το όριο αντοχής των καλωδίων έχει αυξηθεί. Αν επιλέξουμε τ.δ. καλωδίων πλήθους $n = 50$ με $\bar{x} = 1550$ kgr η πρώτη εκτίμηση, με επίπεδο σημαντικότητας $\alpha = 0{,}05$. Να γίνει κατάλληλος στατιστικός έλεγχος για τη μέση τιμή μ.

ΛΥΣΗ

1º βήμα:

Ορίζουμε 2 υποθέσεις:

$H_0: \mu \leq 1500$ kgr και $H_a: \mu > 1500$ kgr

Δηλ. αυτή που αμφισβητείται από τον ισχυρισμό μας (ή αυτή η οποία δεν έχει καμία επίδραση στην εξαρτημένη μεταβλητή για τον πληθυσμό)

Βελτίωση των υλικών άρα και η μέση τιμή αντοχής καλωδίων αυξάνεται αλλαγή στον πληθυσμό

2º βήμα

Ορίζουμε στατιστική συνάρτηση:

$$Z = \frac{\bar{x} - 1500}{175/\sqrt{50}} = \frac{1550 - 1500}{175/\sqrt{50}} = 2.02$$

3º βήμα

Ορίζουμε κρίσιμη περιοχή (περιοχή Απόρριψης της H_0):

$$Z \geq Z_\alpha = Z_{0,05} = 1{,}645 \Rightarrow C = \{Z : Z \geq Z_{0,05}\} = [1{,}645, +\infty)$$

4º βήμα

Λόγω ότι $Z = 2{,}02 \in [1{,}645, +\infty) = C$ τότε η H_0 θα απορριφθεί σε επίπεδο σημαντικότητας $5\% = \alpha$

Άρα, προφανώς θα έχουμε αύξηση της αντοχής των καλωδίων

ΒΑΣΙΚΟ ΠΑΡΑΔΕΙΓΜΑ 2:

153 σεισμικές δονήσεις που μετρήθηκαν μετά το 1997 έδωσαν σε κλίμακα Richter: $\bar{x} = 3.2$. Πριν το 1997 το μέγεθος των σεισμών είχε μέση τιμή $\mu = 2,9$ με $\sigma^2 = 0,08$. Μπορούμε να πούμε ότι το μέγεθος των σεισμών είχε αυξηθεί απο το 1997 σε επίπεδο σημαντικότητας $\alpha = 0,02$;

ΛΥΣΗ

1º Βήμα

Ορίζουμε 2 υποθέσεις

$H_0: \mu \leq 2.9$ και $H_a: \mu > 2.9$

2º Βήμα

Ορίζουμε στατιστική συνάρτηση

$$Z = \frac{\bar{x} - 2,9}{\sqrt{0,08}/\sqrt{153}} = \frac{3,2 - 2,9}{\sqrt{0,08}/\sqrt{153}} = 13,12$$

3º Βήμα

Ορίζουμε κρίσιμη περιοχή (ή περιοχή απόρριψης της H_0):

$Z \geq Z_a = Z_{0,02} = 2,055 \Rightarrow C = \{Z : Z \geq Z_a\} = [2,055, +\infty)$

4º Βήμα

Λόγω ότι $Z = 13,12 \in C = [2,055, +\infty)$ H_0: θα απορριφθεί συνεπώς οι σεισμοί έχουν αυξηθεί απο το 1997

ΠΧ1) Ενας υγειονομικός σταθμός θέλει να ελέγξει εαν ο μέσος αριθμός βακτηριδίων ανά μονάδα όγκου θαλασσινού νερού σε μια παραλία υπερβαίνει το επίπεδο ασφαλείας των 200. Δώδεκα δείγματα νερού συλλέγονται και βρίσκονται οι ακόλουθοι αριθμοί βακτ. ριδίων ανά μονάδα όγκου:

170, 175, 190, 198, 215, 185
184, 207, 210, 193, 196, 180

(Να υποθέσετε ότι οι μετρήσεις αποτελούν ένα τδ με κανονική κατανομή και επίπεδο σημαντικότητας $\alpha = 1\%$)

Υπάρχει λόγος ανησυχίας;

ΛΥΣΗ

Έστω μ ο μέσος όρος βακτηριδίων ανά μονάδα όγκου και σ^2 η διακύμανση. Το 200 θεωρείται ως το ανώτατο επιτρεπτό όριο. Άρα, προκύπτουν:

Μηδενική υπόθεση (H_0): $\mu \geq 200$
Εναλλακτική υπόθεση (H_a): $\mu < 200$

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \quad \left(\sim t_{n-1} \text{ όταν } \mu = \mu_0\right) \quad (1)$$

Η t έχει 11 β.ε. και κριτ. περιοχή:

$$C = (-\infty, -t_{0{,}01,11}) = (-\infty, -2{,}718)$$

$$\bar{X} = \frac{170+175+190+198+215+185+184+207+2010+193+196+180}{12} =$$

$$\Rightarrow \boxed{\bar{X} = 191{,}9}$$

$$S^2 = \frac{1}{11}\sum_{i=1}^{12}(X_i - \bar{X})^2 \Rightarrow \boxed{S = 14{,}03}$$

$$(1): \quad t = \frac{191{,}9 - 200}{14{,}03/\sqrt{12}} = -2{,}197 \notin C = (-\infty, -2{,}718)$$

Άρα, η H_0 δεν απορρίπτεται στο επίπεδο του 1%.

Ένα 95% διάστημα εμπιστοσύνης του μ, είναι:

$$\bar{X} \pm t_{0{,}025,11}\frac{S}{\sqrt{n}} = 191{,}9 \pm 2{,}201 \cdot 4{,}05 = 191{,}9 \pm 8{,}91.$$

ΠΧ2) 25 παιδιά εξετάζονται με ένα βασικό τεστ ευφυΐας. Κατόπιν τα παιδιά αυτά παρακολουθούν ένα ειδικό μάθημα όπου ο σκοπός του είναι η αύξηση του δείκτη ευφυΐας. Στο τέλος του μαθήματος εξετάζονται για $2^{η}$ φορά. Η διαφορά των βαθμών μεταξύ $2^{ης}$ και $1^{ης}$ καταγράφεται για κάθε παιδί. Έστω ότι η μ.τ. των διαφορών τους είναι $\bar{X}=6$ και η διακύμανση $S^2=64$ (σε κλίμακα 0-100). Έχει η εκπαίδευση αυξήσει τον βαθμό ευφυΐας; (Θεωρήστε $a=5\%$).

<u>ΛΥΣΗ</u>

Έστω μ η μ.τ των διαφορών.

Μηδενική Υπόθεση H_0: $\underline{\mu=0}$ (να παραμένει στάσιμη)

Εναλλακτική Υπόθεση H_a: $\underline{\mu>0}$ (να υπάρχει θετική αύξηση)

Θα χρησιμοποιήσουμε δεξιόπλευρο t, τεστ

$$t_{0{,}05,\,24}=1{,}711$$

$$t=\frac{\bar{X}-\mu_0}{s/\sqrt{n}} \quad \text{και} \quad C=(t_{a,n-1},+\infty)=(1{,}711,+\infty)$$

(προσοχή: αυτό επίσης είναι ότι θέλουμε να "μην" το $\frac{a}{2}$ και στο τέλος να έχουμε τελικά λάθος 2,5%).

$$t=\frac{6-0}{8/5}=3{,}75>1{,}711=t_{a,n-1}$$

άρα, $t\in C=(1{,}711,+\infty)$

Έτσι απορρίπτουμε την H_0 και καταλήγουμε στο ότι η εκπαίδευση αυξάνει το δείκτη ευφυΐας των παιδιών (στη $2^{η}$ φορά) με ποσοστό σφάλματος 5%

ΑΣΚΗΣΕΙΣ

1) Δώδεκα φυτά επιλέγονται τυχαίως και από κάθε τέτοιο φυτό επιλέγεται ένα φύλλο στην τύχη. Για κάθε φύλλο προσδιορίζεται το περιεχόμενο του σε ασκορβικό οξύ με τα ακόλουθα αποτελέσματα:

9.35	8.65	11.68	12.77	8.81	9.52
8.68	9.82	10.28	10.39	10.76	10.55

i) Να εκτιμηθεί το μέσο περιεχόμενο ασκορβικού οξέως και να βρεθεί το τυπικό σφάλμα του εκτιμητή

ii) Να βρεθεί ένα 95% διάστημα εμπιστοσύνης του μέσου περιεχομένου.

ΛΥΣΗ

i) $\bar{X}$ ο εκτιμητής του μ

S ο εκτιμητής του σ

$$\bar{X} = \frac{1}{12} \cdot \sum_{i=1}^{12} x_i = \dots = 10,156$$

$$S^2 = \frac{1}{11} \cdot \sum_{i=1}^{12} (x_i - \bar{x})^2 = \dots = 0,367$$

ii) $P(L < \mu < U) = 1 - a$ (= διάστημα εμπιστοσύνης)

$$1 - a = 0,95 \Rightarrow a = 0,05$$

Τα όρια εμπιστοσύνης είναι:

$$L = \bar{X} - Z_{a/2} \frac{\sigma}{\sqrt{n}} = 10,156 - 1,96 \cdot \frac{0,367}{\sqrt{12}} = 9,948$$

$$U = \bar{X} + Z_{a/2} \frac{\sigma}{\sqrt{n}} = 10,156 + 1,96 \frac{0,367}{\sqrt{12}} = 10,364$$

2) Η μέση ετήσια παραγωγή γάλακτος μιας συγκεκριμένης φυλής αγελάδων είναι 4000 kgr/αγελάδα. Ένας ερευνητής θέλει να ελέγξει αν στις κτηνοτροφικές μονάδες της Μακεδονίας και της Θράκης οι αγελάδες της συγκεκριμένης φυλής έχουν μέση ετήσια απόδοση που αναφέρεται στη βιβλιογραφία. Για το σκοπό αυτό και με βάση ένα σχέδιο τυχαίας δειγματοληψίας, επέλεξε ένα δείγμα 40 αγελάδων της συγκεκριμένης φυλής από μονάδες της Μακεδονίας και της Θράκης και κατέγραψε κάθε μέρα, επί ένα έτος, την παραγωγή γάλακτος κάθε μιας αγελάδας του δείγματος. Η μέση ετήσια παραγωγή των 40 αγελάδων, βρέθηκε 3910 kgr με τυπική απόκλιση 250 kgr (επίπεδο σημαντικότητας $\alpha=0{,}05$ εάν βρέθηκε ότι η μέση ετήσια απόδοση των αγελάδων στη Μακεδονία-Θράκη διαφέρει από τη μέση ετήσια απόδοση που αναφέρεται στη βιβλιογραφία). Να ελεγχθεί η αρχική υπόθεση έναντι της εναλλακτικής.

<u>ΛΥΣΗ</u>

Έστω X: ετήσια παραγωγή γάλακτος σε kgr, μιας αγελάδας της συγκεκριμένης φυλής στη Μακεδονία & στη Θράκη και με $X_1, X_2, \dots, X_{40}$ οι ετήσιες αποδόσεις από 40 αγελάδες του πληθυσμού. Στο συγκεκριμένο δείγμα που πήρε ο ερευνητής οι τιμές του δείγματος $x_1, x_2, \dots, x_{40}$ έδωσαν:

$\bar{x} = 3910$ kgr με $S = 250$ kgr.

Όταν n→∞ (ν≥30) τότε $t_\infty \equiv N(0,1)$

Υποθέτουμε:

- Μηδ. Υπόθεση (H_0): $\mu = \overbrace{4000}^{\mu_0}$ kg (αυτή που αμφισβητείται)
- Εναλ. Υπόθεση (H_a): $\mu \neq 4000$ kg (ελέγχω αν θα έχει αύξ. ή μείωση)

Ως σ.σ. χρησιμοποιούμε την $\dfrac{\bar{x}-\mu_0}{S/\sqrt{n}} = Z \sim N(0,1)$

διότι η διασπορά του πληθυσμού μας είναι άγνωστη και το πλήθος αρκετά μεγάλο ($n=40>30$)

Επίπεδο σημαντικότητας $\alpha = 0{,}05$ άρα η περιοχή απόρριψης (ή κρίσιμη περιοχή) είναι η:

$$C = (-\infty, -z_{\alpha/2}] \cup [z_{\alpha/2}, +\infty) = (-\infty, -z_{0{,}025}] \cup [z_{0{,}025}, +\infty) =$$
$$= (-\infty, -1{,}96] \cup [1{,}96, +\infty)$$

Ενώ, $z = \frac{(\bar{x}-\mu_0)}{s/\sqrt{n}} = -2,28 \in (-\infty, -1,96)$

άρα $z \in C$ και έτσι H_0 απορρίπτεται με $\alpha = 5\%$ Επίπεδο σημαντικότητας (ή ποσοστό σφάλματος 5%)

Συμπέρασμα: Σε βαθμό σφάλματος 5% το δείγμα δίνει στατιστικά σημαντικές αποδείξεις ότι η μέση ετήσια απόδοση των αγελάδων της συγκεκριμένης φυλής Μακεδ-Θράκης διαφέρει από τη μέση ετήσια απόδοση που αναφέρεται στη βιβλιογραφία

3) Από έναν πληθυσμό που ακολουθεί κανονική κατανομή, πήραμε ένα δείγμα μεγέθους $n=9$ με $\bar{x}=60$ μονάδες και $s=12$ μονάδες. Σε επίπεδο σημαντικότητας 5%, να γίνει έλεγχος της μηδενικής υπόθεσης $H_0: \mu=65$ έναντι της εναλλακτικής υπόθεσης $H_a: \mu \neq 65$.

Λύση

Μηδενική Υπόθεση $H_0: \mu = 65$

Εναλλακτική Υπόθεση $H_a: \mu \neq 65$

Προφανώς, κατάλληλο είναι το t τεστ (Κανονικός πληθυσμός με άγνωστη διασπορά)

με κρίσιμη περιοχή $C = (-\infty, -t_{0,05,8}] \cup [t_{0,05,8}, +\infty) \Rightarrow$

$\Rightarrow C = (-\infty, -2,306] \cup [2,306, +\infty)$.

$$t = \frac{\bar{x}-\mu_0}{s/\sqrt{n}} = \frac{60-65}{12/3} = -\frac{5}{4} = -1,25 \notin C$$

Επομένως, σε επίπεδο σημαντικότητας $\alpha = 0,05$ δεν απορρίπτουμε την H_0

Ερώτηση: Απορρίπτοντας την $H_0: \mu=65$ αποδείξαμε ότι είναι αληθής;

Απάντηση: Όχι! Δεν αποδείξαμε ότι $\mu=65$, αλλά απλώς αποτύχαμε να απορρίψουμε την H_0. Δεν είπαμε ότι τη δεχόμαστε όμως

4) Από έναν πληθυσμό με άγνωση κατανομή και άγνωση διασπορά πήραμε ένα τ.δ. μεγέθους $n=36$. Από παλαιότερες έρευνες είναι γνωστό ότι η μέση τιμή του πληθυσμού είναι $\mu=83$, όμως υπάρχουν υπόνοιες ότι έχει αλλάξει. Το δείγμα που πήραμε έδωσε:
$\bar{x}=86{,}2$ και $s=10$

α. Να γίνει σε επίπεδο σημαντικότητας $\alpha=5\%$, κατάλληλος στατιστικός έλεγχος για τη μέση τιμή του πληθυσμού

β. Αν αλλαγή της μέσης τιμής σημαίνει μόνο αύξηση, αλλάζει κάτι στον έλεγχο που πρέπει να κάναμε;

<u>ΛΥΣΗ</u>

α. Θα κάνουμε έλεγχο τις

$H_0: \mu=83$ και $H_a: \mu\neq 83$

Λόγω ότι $n>30$: Όταν $n\to\infty$ ($\nu\geq 30$) τότε $t_\infty \equiv N(0,1)$

$$|z| = \frac{|(\bar{x}-\mu_0)|}{s/\sqrt{n}} \geq z_{\alpha/2} = z_{0{,}025} \Rightarrow z\geq 1{,}96 \text{ ή } z\leq -1{,}96$$

Και η τιμή της σσ ελέγχου, είναι:

$$z = \frac{86{,}2-83}{10/\sqrt{36}} = 1{,}92 \notin C = (-\infty,-1{,}96]\cup[1{,}96,+\infty) \text{ και άρα}$$

η Ho <u>δεν</u> απορρίπτεται. (Δηλ. η μ δεν έχει πραγματι αλλάξει)

β. Προφανώς τώρα θα έχουμε την εναλλακτική

$H\alpha: \mu>83$ (δεξιόπλευρος έλεγχος)

$z\geq z_{0{,}05} \Rightarrow C=(1{,}645,+\infty)$

$z_{0{,}05}=1{,}645 < 1{,}92 = z$

Άρα, εδώ η Ho απορρίπτεται σε ποσοστό $\alpha=5\%$

Άρα, πλέον είναι στατιστικά σημαντικές αποδείξεις ότι η μέση τιμή έχει αυξηθεί