Δίνονται ομαδοποιημένες οι ημερήσιες καταναλώσεις ηλεκτρικής ενέργειας (σε 100-άδες κιλοβατώρες) μιας χημικής βιομηχανίας για 40 ημέρες:

ενέργεια	9-13	13-17	17-21	21-25	25-29
ημέρες	4	10	14	8	4

Εάν μ είναι η μέση τιμή της ημερήσιας κατανάλωσης ηλεκτρικής ενέργειας, σε επίπεδο σημαντικότητας α=10%, δεχόμαστε την υπόθεση $H_0 : \mu = 18$ έναντι της $H_1 : \mu \neq 18$; ΝΑΙ ☐ ή ΟΧΙ ☐

Οι παρατηρήσεις του δείγματος, μεγέθους $n = 40$, δίνονται ομαδοποιημένες κατά συνέπεια ο δειγματικός μέσος υπολογίζεται από τον τύπο:

$$\overline{x} = \frac{\sum_{i=1}^{\kappa} x_i f_i}{n} \overset{\text{εδώ}}{=\!=} \frac{11 \times 4 + 15 \times 10 + 19 \times 14 + 27 \times 4}{40} = \frac{752}{40} = \underline{\underline{18.8 \text{ 100-δες } kWh}},$$

και η δειγματική διασπορά από τον τύπο:

$$s^2 = \frac{1}{n-1} \sum_{i=1}^{\kappa} (x_i - \overline{x})^2 f_i \overset{\text{εδώ}}{=\!=} \frac{(11 - 18.8)^2 \times 4 + \cdots + (27 - 18.8)^2 \times 4}{39}$$

$$= 20.472 = (4.52)^2, \quad \text{οπότε η τυπική απόκλιση είναι:} \quad s = \underline{\underline{4.52 \text{ 100-δες } kWh}}$$

Ο έλεγχος είναι αμφίπλευρος και η εναντία υπόθεση είναι $H_1 : \mu \neq 18$

Τρόπος 1ος Κατασκευάζουμε διάστημα εμπιστοσύνης (δ.ε.) συντελεστού εμπιστοσύνης (σ.ε.) $1-\alpha$, όπου α το επίπεδο σημαντικότητας (ε.σ.) του ελέγχου. Εδώ $1-\alpha = 1-0.1 = 0.9 (= 90\%)$.

Απορρίπτουμε την H_0 όταν το 18 δεν περιέχεται στο διάστημα αυτό.

Το δ.ε. για την μέση τιμή οποιασδήποτε κανονικής με άγνωστη διασπορά και μεγάλο μέγεθος δείγματος n=40, σ.ε. $1-\alpha$, δίδεται από τον τύπο

$$\left[\overline{X} - z_{\alpha/2}\frac{S}{\sqrt{n}},\ \ \overline{X} + z_{\alpha/2}\frac{S}{\sqrt{n}}\right] \overset{\text{εδώ}}{=\!=} \left[\overline{X} - z_{0.05}\frac{S}{\sqrt{40}},\ \ \overline{X} + z_{0.05}\frac{S}{\sqrt{40}}\right]$$

εφόσον, $1-\alpha = 0.9 \Rightarrow \frac{\alpha}{2} = 0.05$ και $z_{0.05}$ είναι τέτοιο ώστε:

$$P(Z > z_{0.05}) = 0.05 \Rightarrow \Phi(z_{0.05}) = P(Z \leqslant z_{0.05}) = 0.95, \Phi(1.645) = 0.95 \Rightarrow z_{0.05} = 1.645$$

$$\left[18.8 - 1.645\frac{4.52}{\sqrt{40}},\ \ 18.8 + 1.645\frac{4.52}{\sqrt{40}}\right] = [18.8 - 1.645 \times 0.715,\ \ 18.8 + 1.645 \times 0.715] =$$
$$[18.8 - 1.17,\ \ 18.8 + 1.17] = \underline{\underline{[17.62,\ \ 19.97]}}$$

είναι το δ.ε. σ.ε. 90%, για το οποίο παρατηρούμε ότι:

Κάτω Άκρο < 18 < Άνω Άκρο

άρα αποδεχόμαστε την H_0 σε ε.σ. 10%.

Τρόπος 2^{ος} Απορρίπτουμε την $H_0: \mu = \mu_0$ έναντι της $H_1: \mu \neq \mu_0$ σε ε.σ. α, όταν ισχύει:

$$|T| = \left| \frac{\overline{X} - \mu_0}{\frac{S}{\sqrt{n}}} \right| > z_{\alpha/2}.$$

Εδώ $\mu_0 = 18$, $\alpha = 0.1$ και αντιστοίχως έχουμε

$$|T| = \left| \frac{18.8 - 18}{\frac{4.52}{\sqrt{40}}} \right| = 1.118 < 1.645 = z_{0.05},$$

άρα αποδεχόμαστε την H_0 σε ε.σ. 10%.

Να σημειωθεί ότι το p-value=2*(1-CDF.NORMAL(1.118))=0.2635.

Τα δεδομένα στην περίπτωση που μελετάμε προέρχονται από άγνωστη κατανομή με μέση τιμή μ για την οποία δε γνωρίζουμε τη διασπορά. Συνεπώς, εφόσον έχουμε αρκετές παρατηρήσεις $n = 40$, χρησιμοποιήσουμε το κεντρικό οριακό θεώρημα σύμφωνα με το οποίο ισχύει ότι:

$$\frac{\overline{X} - \mu}{\frac{S}{\sqrt{n}}} \overset{\text{προσεγ.}}{\sim} N(0, 1).$$

Εάν γνωρίζουμε ότι η μέση τιμή της ημερήσιας κατανάλωσης ηλεκτρικής ενέργειας είναι $\mu = 18$ και η διασπορά της είναι $\sigma^2 = 20$. Να υπολογιστεί η πιθανότητα για 80 ημέρες η μέση ημερήσια κατανάλωση να κυμανθεί από 17.5 έως 19 100-άδες κιλοβατώρες ______.

Συμβολίζουμε με την τ.μ. X_i: την κατανάλωση ηλεκτρικής ενέργειας (σε 100-άδες kWh) την i ημέρα $i = 1, \ldots, 80$. Τότε, σύμφωνα με τα δεδομένα του προβλήματος, $\mu := E(X_i) = 18$, $\sigma^2 := Var(X_i) = 20$.
Η τ.μ. $\overline{X} = \frac{1}{80}\sum_{i=1}^{80} X_i$ είναι ο μέσος όρος των 80 μετρήσεων. Για αυτήν ισχύει ότι:

$$E\left(\overline{X}\right) = \frac{1}{80}\left(\sum_{i=1}^{80} X_i\right) = \frac{1}{80}\sum_{i=1}^{80} E\left(X_i\right) = \frac{1}{80}80 \cdot \mu = \mu = \underline{\underline{18(\text{ 100-άδες kWh })}},$$

$$Var\left(\overline{X}\right) = \frac{1}{80^2}Var\left(\sum_{i=1}^{80} X_i\right) \overset{\text{οι τ.μ. } X_i}{\underset{\text{είναι ανεξ.}}{=\!=\!=}} \frac{1}{80^2}\sum_{i=1}^{80} Var\left(X_i\right) = \frac{\sigma^2}{80} = \frac{1}{4} = \underline{\underline{0.5^2}}.$$

Επειδή έχουμε $n = 80$ παρατηρήσεις ισχύει το κεντρικό οριακό θεώρημα και συνεπώς

$$\overline{X} \overset{\text{προσεγ}}{\sim} N\left(18, 0.5^2\right) \Rightarrow Z = \frac{\overline{X} - E(\overline{X})}{\sqrt{Var(\overline{X})}} = \frac{\overline{X} - 18}{0.5} \overset{\text{προσεγ}}{\sim} N\left(0, 1\right).$$

Έτσι

$$\begin{aligned} P\left(17.5 \leqslant \overline{X} \leqslant 19\right) &= P\left(\frac{17.5 - 18}{0.5} \leqslant \frac{\overline{X} - 18}{0.5} \leqslant \frac{19 - 18}{0.5}\right) \\ &= P\left(-1 \leqslant Z \leqslant 2\right) \\ &\simeq \Phi(2) - \Phi(-1) = \Phi(2) - (1 - \Phi(1)) \\ &= 0.977 - 1 + 0.841 = \underline{\underline{0.818}} \end{aligned}$$

Αυτό ισχύει, επειδή η σ.π. της $N(0, 1)$ είναι συμμετρική ως προς το 0, δηλαδή,

$$P(Z < -1) = P(Z > 1) \Rightarrow P(Z < -1) = 1 - P(Z \leqslant 1) \Rightarrow \Phi(-1) = 1 - \Phi(1).$$

Σε ένα εργαστήριο συμμετέχουν 9 φοιτητές και 6 φοιτήτριες και οι βαθμοί τους στην πρώτη άσκηση, σε κλίμακα 0-20, ακολουθούν αντίστοιχα $N(\mu_A, \sigma^2)$ και $N(\mu_K, \sigma^2)$ με σ^2 άγνωστο. Υπολογίστηκε για τους φοιτητές ότι $\overline{X_A} = 14.1$, $S_A^2 = 7.61$ και για τις φοιτήτριες $\overline{X_K} = 16.9$, $S_K^2 = 6.7$. Σε επίπεδο σημαντικότητας α=5%, δεχόμαστε την $H_0 : \mu_A \overset{(<)}{=} \mu_K$ έναντι της $H_1 : \mu_A > \mu_K$; ΝΑΙ ☐ ή ΟΧΙ ☐

Επιλέχθηκαν οι 70 μαθητές της πρώτης δημοτικού από μια σχολική μονάδα, για να μελετηθεί η σχέση ανάμεσα στο φύλο των παιδιών και τη χρήση των χεριών τους. Τα αποτελέσματα που πήραμε δίδονται στον πίνακα που ακολουθεί:

	Φύλο	
"Καλό" χέρι	Αγόρια	Κορίτσια
αριστερό	6	12
δεξί	28	24

1. Σε επίπεδο σημαντικότητας (ε.σ.) α=10%, εξαρτάται η χρήση των χεριών από το φύλο; ΝΑΙ ☐ ή ΟΧΙ ☐
2. Είναι ίδια η απόφασή σας σε ε.σ. α=5%; ΝΑΙ ☐ ή ΟΧΙ ☐ (να σχολιάσετε). Ποια είναι η σχέση του p-value των δεδομένων με τα προηγούμενα ε.σ.; ____ $\leqslant$ 5% $\leqslant$ ____ $\leqslant$ 10% $\leqslant$ ____

Ο έλεγχος που ζητείται είναι ένας έλεγχος ανεξαρτησίας:

H_0 : P(“Καλό” χέρι $\cap$ Φύλο) $=$ P(“Καλό” χέρι)P(Φύλο),

H_1 : διαφορετικά,

με “Καλό” χέρι $\in$ {αριστερό, δεξί} και Φύλο $\in$ {Αγόρια, Κορίτσια}

	Φύλο		
"Καλό" χέρι	Αγόρια	Κορίτσια	
αριστερό	6 (8.7)	12 (9.3)	$\pi_{1\bullet} = 18$
δεξί	28 (25.3)	24 (26.7)	$\pi_{2\bullet} = 52$
	$34 = \pi_{\bullet 1}$	$36 = \pi_{\bullet 2}$	$n = 70$

Για τον έλεγχο, χρησιμοποιείται η στατιστική συνάρτηση (σ.σ.)

$$X^2 = \sum_{i=1}^{\kappa_1} \sum_{j=1}^{\kappa_2} \frac{(\pi_{ij} - \theta_{ij})^2}{\theta_{ij}} \left(= \sum_{i=1}^{\kappa_1} \sum_{j=1}^{\kappa_2} \frac{\pi_{ij}^2}{\theta_{ij}} - n \right),$$

όπου π_{ij} οι παρατηρούμενες, $\theta_{ij} = \frac{\pi_{i\bullet} \times \pi_{\bullet j}}{n}$ οι αντίστοιχες θεωρητικές τιμές υπό την H_0, και κ_1, κ_2 το πλήθος των γραμμών και των στηλών αντιστοίχως του πίνακα συνάφειας (εδώ $\kappa_1 = 2$, $\kappa_2 = 2$).

Όταν ισχύει η H_0, τότε: $X^2 \overset{\text{ασυμπτ.}}{\sim} \mathcal{X}^2_{(\kappa_1 - 1)\times(\kappa_2 - 1)}$.

Έτσι απορρίπτουμε την H_0, σε ε.σ. α, όταν $X^2 \geqslant \mathcal{X}^2_{(\kappa_1 - 1)\times(\kappa_2 - 1)}(\alpha)$.

Από τα δεδομένα παίρνουμε, $X^2 = \frac{(6-8.7)^2}{8.7} + \frac{(12-9.3)^2}{9.3} + \frac{(28-25.3)^2}{25.3} + \frac{(24-26.7)^2}{26.7} = 0.8605 + 0.8126 + 0.2978 + 0.2813 = 2.2523$

οπότε $X^2 = 2.2523 < 2.71 = \mathcal{X}_1^2(0.1)$,
άρα σε ε.σ. 10% αποδεχόμαστε την H_0, δηλ. στους μαθητές αυτούς η χρήση του χεριού είναι ανεξάρτητη του φύλου τους.

Αφού η H_0 είναι δεκτή σε ε.σ. 10% θα είναι δεκτή και σε για κάθε ε.σ. μικρότερο του 10%, συνεπώς είναι δεκτή και στο 5%.

Άρα για το p-value του ελέγχου ισχύει: $\underline{\underline{10\% < \text{p-value}}}$, μιας και αυτό είναι το μικρότερο ε.σ. για το οποίο απορρίπτουμε την H_0.

(p-value: $P(\mathcal{X}_1^2 > 2.2523) = 0.1334$, δηλ. αποδέχομαι την H_0 για ε.σ. $\alpha < 0.1334$,
και απορρίπτω την H_0 για ε.σ. $\alpha > 0.1334$.
Συνεπώς, $0.05 < 0.1 < 0.1334$)

	Φύλο	
“Καλό” χέρι	Αγόρια	Κορίτσια
αριστερό	6	12
δεξί	28	24

3. Εάν συμβολίσουμε με p_A το ποσοστό των αριστερόχειρων παιδιών, να δώστε ένα διάστημα εμπιστοσύνης, συντελεστού εμπιστοσύνης 95%, για το p_A: [_____, _____].
   Ποια κατανομή και ποιο θεώρημα χρησιμοποιήσατε για την κατασκευή του διαστήματος αυτού και γιατί;

Δώστε ένα παράδειγμα ζεύγους ανεξάρτητου (X) και εξαρτημένου χαρακτηριστικού (Y) για την μελέτη των οποίων μπορούμε να χρησιμοποιήσουμε την θεωρία της ευθείας παλινδρόμησης.
Εάν έχω $n = 15$ παρατηρήσεις με $x_{(1)} = -18$, $x_{(15)} = 19$, $\widehat{\beta_0} = 3.47$ και $\widehat{\beta_1} = -0.09$ να γίνει η γραφική παράσταση της ευθείας παλινδρόμησης και να δοθούν οι αναμενόμενες τιμές του εξαρτημένου χαρακτηριστικού για $x_0 = 15$ ______ και $x_0 = 25$ ______ αντιστοίχως.