

Εργασία 2

(1)

Δίνεται αρχείο στην R, διαθέσιμο στις εργασίες - ηλεκτρ. τάση με τα δεδομένα ύψους φοιτητών της Στατιστικής Ι, όπως δήλωσαν οι φοιτητές σε φωνηματολόγιο το χειμ. εξάμηνο του 2016.

(i) ανοίξτε το αρχείο με την R (αφού πρώτα κάνετε εγκατάσταση του λογισμικού), θα παρατηρήσετε ότι είναι διαθέσιμα τα

δεδομένα ύψους και φύλου 83 φοιτητών της Στατιστικής Ι.

Μπορείτε να χρησιμοποιήσετε οποιοδήποτε λογισμικό θέλετε για να αναλύσετε τα δεδομένα.

• Υπολογίστε τον αριθμητικό μέσο όρο \bar{x} και τη βεγματική διασπορά s^2 που αντιστοιχεί στο συγκεκριμένο αυτό δείγμα.

Αναφέρετε τα αποτελέσματά σας.

(ii) Υποθέστε ότι έχετε ένα τ.δ. από $N(\mu, \sigma^2)$ με άγνωστη μέση τιμή και διασπορά και εκτιμήστε το μ και το σ^2 , με κάποια από τις μεθόδους που μάθαμε. Ιδανικά χρησιμοποιήστε τουλάχιστον 2 μεθόδους.

(iii) Κάντε ένα ιστόγραμμα των δεδομένων και στη συνέχεια σχεδιάστε τη σ.π.π. της $N(\hat{\mu}, \hat{\sigma}^2)$ που εκτιμήσατε (με μία μέθοδο αρκεί). Τι παρατηρείτε?

(iv) Χωρίστε το δείγμα σε 2 ομάδες, σε άντρες A και γυναίκες B, και επαναλάβετε τα βήματα (i), (ii) και (iii) ξεχωριστά για τα 2 δείγματα. Εδώ υποθέτουμε ότι $N(\mu_a, \sigma_a^2)$ και $N(\mu_g, \sigma_g^2)$ είναι 2 διαφορετικές κανον. κατανομές που αντιστοιχούν σε άντρες και γυναίκες αντίστοιχα. Τι παρατηρείτε?

Σημαντικό ρόλο στην ανάλυση αυτή παίζει το γεγονός ότι η πληροφορία αν είναι άντρας ή γυναίκα είναι διαθέσιμη.

(v) Υποθέστε τώρα ότι $\sigma_a^2 = \sigma_\gamma^2 \equiv \sigma^2$, και

$$X_i \sim \mathcal{N}(\mu_a, \sigma^2), \quad 1 \leq i \leq n_a, \quad Y_i \sim \mathcal{N}(\mu_\beta, \sigma^2), \quad 1 \leq i \leq n_\beta$$

(για άντρες) (για γυναίκες).

Εκτιμήστε τα μ_a, μ_β και σ^2 , συνδυάζοντας τα 2 δείγματα για την εκτίμηση του σ^2 που είναι κοινή παράμετρος.

Συμπεριλάβετε οπωσδήποτε τη μέθοδο μέγιστης πιθανοφάνειας.

(vi) Υποθέστε τώρα ότι $(C_i, Z_i)_{1 \leq i \leq n}$ είναι το

τ.δ. που εκφράζει το φύλο C_i και το ύψος Z_i , του i -ατόμου,

όπου $P(C_i=1) = p$, όπου $1 \rightarrow$ γυναίκα και

και $P(C_i=0) = 1-p$, όπου $0 \rightarrow$ άντρα, και

$$[Z_i | C_i=0] \sim \mathcal{N}(\mu_a, \sigma_a^2) \quad (\text{δηλ. παίρνει το ρόλο της } X_i \text{ πριν}).$$

$$\text{και } [Z_i | C_i=1] \sim \mathcal{N}(\mu_\beta, \sigma_\beta^2) \quad (\text{δηλ. παίρνει το ρόλο της } Y_i \text{ πριν}).$$

Το p θεωρείται άγνωστη παράμετρος που πρέπει να εκτιμηθεί.

Ορίστε μια κατάλληλη συνάρτηση πιθανοφάνειας του τ.δ., και

εκτιμήστε τα $p, \mu_a, \sigma_a^2, \mu_\beta, \sigma_\beta^2$. Τι παρατηρείται σε σχέση με το (iv) ?

(vii) Υποθέστε τώρα ότι η πληροφορία αν είναι άντρας ή γυναίκα δεν είναι διαθέσιμη. Γνωρίζεται όμως την τιμή του p , των σ_a^2 και σ_β^2 , και είναι αυτά που εκτιμήσατε στο (vi).

Προτείνετε μεθόδους εκτίμησης των μ_a και μ_β με μόνη γνώση τα αποτελέσματα Z_1, Z_2, \dots, Z_n .

Αν μπορούσε δώστε εκτιμήσεις των μ_a και μ_β , είτε αναλυτικά, είτε με τη βοήθεια λογισμικού.