

Στατιστικά Προγράμματα, 29-11-04

Homework #1

Έγιναν μετρήσεις σε ένα δείγμα 400 ενηλίκων ανδρών και γυναικών καταγράφοντας στις εξής μεταβλητές:

- | | |
|------------------------------------|--|
| 1. THIGH: περιφέρεια μηρού (σε cm) | 4. WAIST: περιφέρεια μέσης (σε cm) |
| 2. WEIGHT: βάρος (σε kgr) | 5. GENDER: φύλο (1 - male, 0 - female) |
| 3. HEIGHT: ύψος (σε cm) | |

Τα δεδομένα τα οποία βρίσκονται σε αρχείο dat (οι μεταβλητές στις στήλες δίνονται με την παραπάνω σειρά) στην ιστοσελίδα του μαθήματος

<http://www.unipi.gr/faculty/mbouts/statprog/index.htm>

προέρχονται από ένα μεγαλύτερο σύνολο δεδομένων που ελήφθη από διάφορα πανεπιστήμια και περιοχές της Καλιφόρνια (βλ. Heinz, Peterson, Johnson and Kerk (2003) «Exploring Relationships in Body Dimensions», *Journal of Statistics Education* Vol. 11, No 2)

1) Να δοθούν: η δειγματική μέση τιμή, η δειγματική τυπική απόκλιση, η μεγαλύτερη και η μικρότερη τιμή των μεταβλητών 1-4 (συνολικά και ανά γυναίκες – άνδρες). Να δοθεί δ.ε. 95% για τους (πληθυσμιακούς) μέσους των μεταβλητών 1-4 (ανά γυναίκες – άνδρες).

2) Σε ποιες μεταβλητές (από τις 1-4) υπάρχει στατιστικά σημαντική διαφορά στις διασπορές και σε ποιες στους μέσους μεταξύ ανδρών – γυναικών (δώστε τα p-value για τα αμφίπλευρα τεστ). Η μέση περιφέρεια του μηρού είναι μεγαλύτερη στις γυναίκες; (δώστε το p-value για το μονόπλευρο τεστ)

3) Να ελέγξετε αν οι μεταβλητές 2-4 προέρχονται από την κανονική κατανομή (ανά φύλο): να δοθούν τα αντίστοιχα χα ιστογράμματα (χρησιμοποιώντας 10 κλάσεις), Q-Q plots και p-values από το K-S test.

4) Να ελέγξετε αν η μεταβλητή THIGH ακολουθεί την κατανομή Γάμμα (ανά φύλο). Να δοθούν τα αντίστοιχα ιστογράμματα (χρησιμοποιώντας 10 κλάσεις), Q-Q plots και p-values από το K-S test.

5) Ένα εργοστάσιο ενδυμάτων επιθυμεί να κατασκευάσει γυναικείες φόρμες σε 3 διαφορετικά μεγέθη (S, M, L). Σε αυτά τα 3 μεγέθη θα πρέπει να εντάσσεται το 95% των ενηλίκων γυναικών (εξαιρείται το 2.5% των γυναικών με πολύ μεγάλη και το 2.5% των γυναικών με πολύ μικρή μέση). Το μέγεθος S θα αφορά μεγέθη μέσης (Waist) από a έως $a+h$, το M από $a+h$ έως $a+2h$, και το L από $a+2h$ έως $a+3h$ (σε cm). Το εργοστάσιο αναθέτει σε εσάς την μελέτη για την εκτίμηση των a , h . Αν το εργοστάσιο κατασκευάσει 1000 φόρμες, πόσες θα πρέπει να είναι S, πόσες M και πόσες L; Για να προσδιοριστεί και το μήκος που θα έχουν οι φόρμες, να εκτιμήσετε το μέσο ύψος των γυναικών σε κάθε μία από τις τρεις κατηγορίες S, M, L. Να απαντήσετε με βάση τα παραπάνω δεδομένα (χωρίς να θεωρήσετε κάποια συγκεκριμένη κατανομή για την μεταβλητή Waist).

6) Σύμφωνα με κάποιες επιδημιολογικές μελέτες όταν αυξάνεται ο δείκτης BMI (Body Mass Index) ο οποίος ισούται με το πηλίκο του βάρους (σε kg) δια του τετραγώνου του ύψους ενός ατόμου (σε m), αυξάνει παράλληλα και η ένταση θνησιμότητας του ατόμου. Συγκεκριμένα, όταν ο δείκτης BMI είναι μεγαλύτερος του 30 τότε αυξάνεται η ένταση θνησιμότητας (από όλες τις αιτίες αλλά κυρίως από καρδιαγγειακά νοσήματα) κατά 50 -100% σε σχέση με άτομα που έχουν BMI στο διάστημα 20-25. Να κατασκευάσετε μια νέα μεταβλητή που να εκφράζει το BMI κάθε ατόμου, και να εκτιμήσετε τη μέση της τιμή (ανά φύλο). Να κατασκευάσετε ένα πίνακα που να δείχνει το πλήθος των ατόμων του δείγματος που έχουν BMI στα διαστήματα $A = (0, 20]$, $B = (20, 25]$, $C = (25, 30]$, $D = (30, \infty)$ ανά φύλο. Ο πίνακας να περιλαμβάνει και τα αντίστοιχα ποσοστά των ατόμων που εμπίπτουν σε κάθε μία από τις $4 \times 2 = 8$ κατηγορίες.

7) Από προηγούμενη μελέτη έχει βρεθεί ότι ο μέσος BMI (σε άνδρες και γυναίκες από την ίδια περιοχή) είναι ίσος με 23.75. Επαληθεύεται αυτό από τα παραπάνω δεδομένα; (δώστε το p-value για το αμφίπλευρο και τα μονόπλευρα τεστ).

8) Σε μία άλλη περιοχή έχει βρεθεί ότι 5%, 50%, 38% και 7% των ανδρών ανήκουν στα διαστήματα A, B, C, D αντίστοιχα. Να ελέγξετε αν ισχύει το ίδιο και στην περιοχή από την οποία προήλθαν τα παραπάνω δεδομένα (δώστε το p-value κατάλληλου ελέγχου).

9) Η κατανομή του δείκτη BMI στους άνδρες είναι διαφορετική από την κατανομή του δείκτη BMI στις γυναίκες; (δώστε το p-value). Η κατανομή του BMI στους άνδρες είναι ίση με την κατανομή του BMI στις γυναίκες μετατοπισμένη κατά 2,5 μονάδες πάνω; (δώστε τα p-value τριών διαφορετικών παραμετρικών ελέγχων);

(Όλοι οι έλεγχοι να γίνουν σε ε.σ. $\alpha = 5\%$)