

## Εφαρμοσμένη Στατιστική – Εξετάσεις Ιουνίου 2018 – Ομάδα θεμάτων 2.

Διάρκεια εξέτασης: 2 ώρες.

### Θέμα 1. (2 μονάδες)

Αποδείξτε ότι στο μοντέλο ανάλυσης διακύμανσης κατά ένα παράγοντα, η μέση τιμή του αθροίσματος των τετραγώνων των αποκλίσεων από το συνολικό μέσο του δείγματος είναι μεγαλύτερη από τη διακύμανση του μοντέλου, όταν η μηδενική υπόθεση πως οι μέσοι όροι ανά επίπεδο είναι ίσοι δεν αληθεύει. Θα πρέπει μαζί με την απόδειξη να δώσετε το μοντέλο που υποθέτετε, την μηδενική υπόθεση που χρησιμοποιείτε και γενικά να είναι ξεκάθαρο πως προκύπτουν οι πράξεις σας).

### Θέμα 2. (3 μονάδες)

1. Σε ένα τυχαίο δείγμα 18 μαθητών, ο μέσος όρος των βαθμών τους σε ένα τεστ είναι 510 με δειγματική τυπική απόκλιση 140 βαθμών.
  - a) Κατασκευάστε ένα 98% διάστημα εμπιστοσύνης για τη μέση τιμή. Ερμηνεύστε το δ/μα.
  - b) Πιστεύετε ότι η πραγματική μέση τιμή είναι 500. Ελέγξτε την υπόθεση ότι η δειγματική μέση τιμή δεν είναι σημαντικά διαφορετική από την πραγματική τιμή. υποθέστε  $\alpha = 0.01$ , διατυπώστε τα συμπεράσματά σας και την ερμηνεία του ελέγχου σας.
  - c) Πως σχετίζονται τα αποτελέσματα του δ/το εμπιστοσύνης και του ελέγχου υποθέσεων;
2. Με βάση ένα τ.δ.  $X_1, X_2, \dots, X_n \sim N(\mu, 8^2)$  θέλουμε να ελέγξουμε σε ε.σ.  $\alpha = 5\%$  την  $H_0: \mu = 80$  έναντι της  $H_1: \mu = 76$ . Να δειχθεί ότι η κρίσιμη περιοχή είναι της μορφής  $K: \bar{X} < c_\alpha$ . Να βρεθεί το ελάχιστο  $n$  και το  $c_\alpha$  έτσι ώστε  $P[\text{σφάλμα τύπου I}] \leq 0.05$ ,  $P[\text{σφάλμα τύπου II}] \leq 0.10$ .

### Θέμα 3. (3 μονάδες)

Μία μεγάλη αλυσίδα super market συγκέντρωσε στοιχεία για την παραγωγικότητα και το κόστος των εγκαταστάσεων της τα οποία περιέχονται στο αρχείο "GroceryRetailer.csv". Κάθε γραμμή περιέχει τις μετρήσεις μίας εβδομάδας και οι μεταβλητές που περιέχονται είναι: CasesShipped = αριθμός από συσκευασίες που έφυγαν από τις εγκαταστάσεις σε μία εβδομάδα, Costs = κόστος εργασιών των εγκαταστάσεων ανά εβδομάδα, LaborHours = ώρες εργασίας των υπαλλήλων στις εγκαταστάσεις ανά εβδομάδα και Holiday = 0 ή 1 ανάλογα με το αν η εβδομάδα είχε κάποια γιορτή ή όχι.

Η αλυσίδα super market θέλει να μελετήσει την επίδραση του κόστους, των κιβωτίων που έφυγαν από τις εγκαταστάσεις και των αργιών στις ώρες εργασίας των υπαλλήλων των εγκαταστάσεων της.

- a) Ποια στατιστική μέθοδο θα χρησιμοποιήσετε, ποια η εξαρτημένη και ποιες οι ανεξάρτητες μεταβλητές; Εξετάστε τη γραμμικότητα των ανεξάρτητων μεταβλητών σε σχέση με την εξαρτημένη.
- b) Προσαρμόστε το μοντέλο παλινδρόμησης, δώστε την εξίσωση παλινδρόμησης και ερμηνεύστε τους συντελεστές των ανεξάρτητων μεταβλητών.
- c) σχολιάστε το συγκεκριμένο μοντέλο - είναι κατάλληλο για να βγάλουμε συμπεράσματα;
- d) Υπολογίστε τα υπόλοιπα του μοντέλου και κάντε τη γρ. παράσταση τους έναντι της εξαρτημένης και των ανεξάρτητων μεταβλητών - τι πληροφορία μας δίνουν; Πώς θα ελέγξετε την κανονικότητα των υπολοίπων;