

11^ο Φροντιστήριο – Στατιστική Ι



Κεφ. 11 Ανάλυση Παλινδρόμησης & Επαναληπτικές ασκήσεις

Άσκηση 1

Μία ερευνήτρια ενδιαφέρεται να διαπιστώσει κατά πόσο μπορούμε να προβλέψουμε την επαγγελματική ικανοποίηση από τον μισθό που παίρνει ένας εργαζόμενος. Ποια είναι η προβλεπτική ή ερμηνευτική μεταβλητή και ποια η μεταβλητή κριτήριο;

Απάντηση:

Η προβλεπτική μεταβλητή είναι ο μισθός του εργαζομένου και η μεταβλητή κριτήριο (δηλ. αυτή που θέλουμε να προβλέψουμε) είναι η επαγγελματική ικανοποίηση.

Άσκηση 3

Ποια η διαφορά ανάμεσα στην απλή και στην πολλαπλή παλινδρόμηση;

Απάντηση:

Στην ανάλυση της απλής παλινδρόμησης χρησιμοποιούμε μία προβλεπτική μεταβλητή για να προβλέψουμε τη μεταβλητή κριτήριο, ενώ στην ανάλυση πολλαπλής παλινδρόμησης χρησιμοποιούμε περισσότερες από μία προβλεπτικές μεταβλητές.

Άσκηση 5

Πως ονομάζεται το σημείο που βρίσκεται πάνω στη γραμμή παλινδρόμησης;

Απάντηση:

Ονομάζεται τιμή πρόβλεψης (predicted value)

Άσκηση 6

Πως προκύπτει το συνολικό σφάλμα εκτίμησης μιας γραμμής παλινδρόμησης;

Απάντηση:

Προκύπτει από το άθροισμα όλων των αποστάσεων d , το οποίο με τη σειρά του προκύπτει από τη διαφορά που έχουν οι πραγματικές τιμές από τις αντίστοιχες τιμές πρόβλεψης.

Άσκηση 7

Από την ανάλυση παλινδρόμησης που εφαρμόζαμε στη μεταβλητή X και στη μεταβλητή Y , διαπιστώσαμε ότι η κλίση (slope) ήταν $b = -0,68$. Μπορούμε από αυτό το νούμερο να βγάλουμε κάποιο συμπέρασμα για τη σχέση των δύο μεταβλητών που μελετάμε;

Απάντηση:

Το συμπέρασμα που μπορούμε να βγάλουμε είναι ότι υπάρχει αρνητική συσχέτιση ανάμεσα στις δύο μεταβλητές, καθώς το πρόσημο στην τιμή της κλίσης μας δείχνει και την κατεύθυνση της συσχέτισης των δύο μεταβλητών.

Άσκηση 8

Από την ανάλυση πολλαπλής παλινδρόμησης, διαπιστώσαμε ότι ο συντελεστής πολλαπλής συσχέτισης R είναι 0,55. Πόσο ποσοστό της διακύμανσης ερμηνεύουν τελικά οι προβλεπτικές μεταβλητές που συμμετείχαν στη συγκεκριμένη εξίσωση παλινδρόμησης;

Απάντηση:

Το ποσοστό της δικαύμανσης που ερμηνεύει είναι

$$0,55^2 = 0,3025$$

Άσκηση 9

Για ποιο λόγο είναι αναγκαίο να μετατρέψουμε τους συντελεστές παλινδρόμησης σε τυποποιημένους συντελεστές παλινδρόμησης;

Απάντηση:

Για να εξισώσουμε τις διακυμάνσεις των προβλεπτικών μεταβλητών που χρησιμοποιούμε στην εξίσωση παλινδρόμησης (όλες θα πρέπει να έχουν $M.O.=0$ και $t.a.=1$). Μόνο τότε μπορούμε να ερμηνεύσουμε τη σημαντικότητα του κάθε προβλεπτικού παράγοντα.

Άσκηση 10

Ποιο είναι εκείνο το φαινόμενο το οποίο, εάν υπάρχει, θα δημιουργήσει πρόβλημα εγκυρότητας στα αποτελέσματα που θα προκύψουν από την ανάλυση της πολλαπλής παλινδρόμησης;

Απάντηση:

Είναι το φαινόμενο της πολυσυγραμμικότητας. Επειδή υπάρχει υψηλή συσχέτιση ανάμεσα στις προβλεπτικές μεταβλητές, δεν μπορούμε να δεχθούμε τα αποτελέσματα της εξίσωσης παλινδρόμησης, καθώς είναι πιθανόν δύο διαφορετικά δείγματα από τον ίδιο πληθυσμό να δώσουν εντελώς διαφορετικές εξισώσεις παλινδρόμησης.

Άσκηση 11

Ένας ερευνητής θέλησε να μελετήσει τη σχέση μεταξύ της ενδοοικογενειακής βίας και τεσσάρων προβλεπτικών μεταβλητών: ετήσιο εισόδημα, έτη συμβίωσης, κοινωνικοοικονομικό επίπεδο και επιθετική συμπεριφορά. Η μεταβλητή κριτήριο (εξαρτημένη μεταβλητή) είχε δύο επίπεδα: Επεισόδιο ενδοοικογενειακής βίας (1) – Μη επεισόδιο ενδοοικογενειακής βίας (0). Τι ανάλυση θα χρησιμοποιήσει για να μελετήσει το παραπάνω ερευνητικό ερώτημα;

Απάντηση:

Θα χρησιμοποιήσει διωνυμική λογιστική παλινδρόμηση, καθώς η μεταβλητή κριτήριο είναι κατηγορική με δύο επίπεδα. Επεισόδιο Ενδοοικογενειακής βίας (1) – Μη Επεισόδιο Ενδοοικογενειακής βίας (0).

Επαναληπτικές Ασκήσεις

1η Επαναληπτική Άσκηση

Μας ενδιαφέρει να διερευνήσουμε αν οι φοιτητές/τριες ηλικίας 18-25 τη σημερινή εποχή (έτος 2025) περνούν αρκετές ώρες στις καφετέριες με φίλους. Σύμφωνα με παλαιότερες έρευνες, (τελευταία έρευνα του 2012 που καθόρισε τη συμπεριφορά του πληθυσμού) οι φοιτητές/τριες περνούσαν κατά μέσο όρο 7 ώρες/εβδομάδα στις καφετέριες με απόκλιση 2 ωρών.

1η Επαναληπτική Άσκηση

Από μία έρευνα που διεξήχθη τον Απρίλιο (2025) με δείγμα 100 φοιτητές/τριες από διαφορετικά μέρη της Ελλάδας, προέκυψε ότι ο μέσος χρόνος που περνάνε σε καφετέριες είναι 6,3. Η έρευνα είχε σκοπό να ελέγξει αν η συγκεκριμένη συνήθεια έχει μεταβληθεί στο έτος 2025 σε επίπεδο σημαντικότητας 0,05.

1η ερώτηση

Μπορούμε να αποφασίσουμε, εφόσον

α) κάνουμε έλεγχο υπόθεσης

β) λάβουμε υπόψη ότι ο μέσος όρος άλλαξε (από 7 έγινε 6,3), οπότε άλλαξε και η συμπεριφορά των φοιτητών/τριών.

γ) Είναι γνωστό ότι πλέον οι φοιτητές/τριες δεν κάθονται πολύ για καφέ, διότι έχουν άλλα ενδιαφέροντα.

2η ερώτηση

Αν αποφασίσουμε να κάνουμε έλεγχο υπόθεσης, θα κάνουμε

α) Μονόπλευρο με $H_1: \mu_0 < \mu$

β) Μονόπλευρο με $H_1: \mu_0 > \mu$

γ) Αμφίπλευρο έλεγχο με $H_1: \mu_0 \neq \mu$

δ) Δεν χρειάζεται ο έλεγχος υπόθεσης διότι ο μέσος όρος του δείγματος δείχνει ξεκάθαρα την τάση του πληθυσμού.

3η ερώτηση

Εφόσον επιλεγεί να γίνει έλεγχος υπόθεσης, ποιο από τα παρακάτω στατιστικά κριτήρια θα επιλέξετε;

α) $\chi^2 = \sum \frac{(\Pi - A)^2}{A}$

β) $Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

γ) $T = \bar{x} + \mu_0 + \sigma - n$

4η ερώτηση

Τελικά, άλλαξε η συμπεριφορά των φοιτητών/τριών ως προς τον χρόνο που περνάνε στις καφετέριες το 2025; Δίνονται οι αριθμοί, σε περίπτωση που τους χρειαστείτε $z_{0,95}=1,645$, $z_{0,05}=-1,645$, $z_{0,975}=1,96$ και $z_{0,025}=-1,96$.

- α) ΟΧΙ, περνάνε τον ίδιο χρόνο
- β) ΝΑΙ, άλλαξε η συμπεριφορά

ΛΥΣΗ

Ερώτηση 1 : α

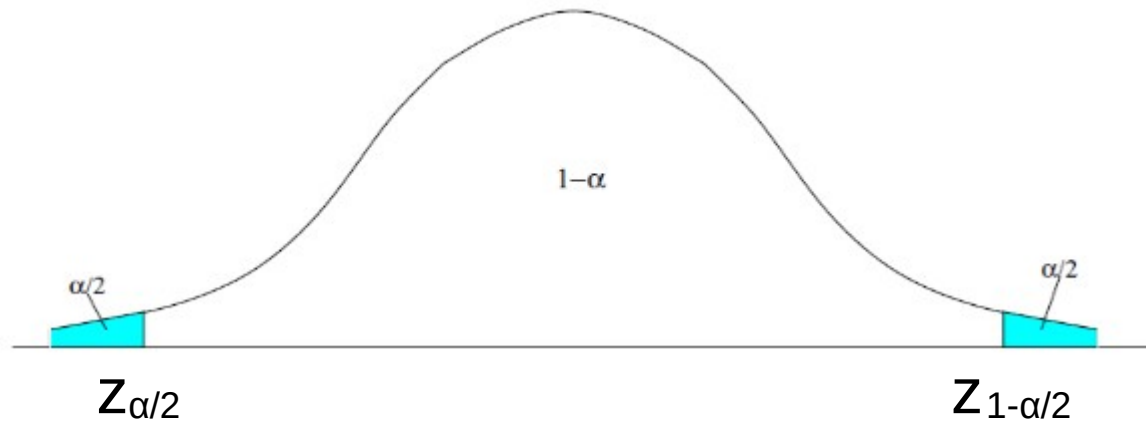
Ερώτηση 2 : γ

Ερώτηση 3 : β

Ερώτηση 4 : ?

ΛΥΣΗ

ΘΕΩΡΙΑ: Στον αμφίπλευρο έλεγχο εξετάζουμε αν η τιμή που παίρνουμε από το στατικό κριτήριο “πέφτει” ή όχι στην περιοχή απόρριψης της μηδενικής υπόθεσης (ότι δεν αλλάζει). Τα όρια της περιοχής αυτής καθορίζονται (γαλάζια περιοχή) από τους αριθμούς $Z_{\alpha/2}$ και $Z_{1-\alpha/2}$.

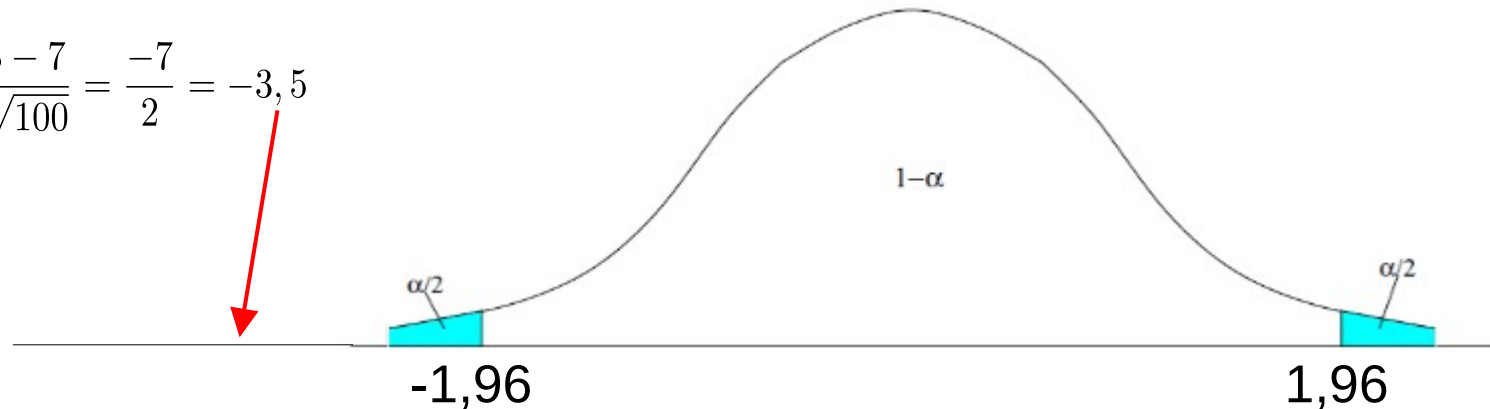


ΛΥΣΗ

4: Στην άσκηση εφόσον το $\alpha=0,05$, η περιοχή απόρριψης οριοθετείται από τους αριθμούς $z_{0,025}=-1,96$ και $z_{0,975}=1,96$.

Υπολογίζουμε την τιμή που δίνει το κριτήριο και βλέπουμε που “πέφτει”:

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{6,3 - 7}{2/\sqrt{100}} = \frac{-7}{2} = -3,5$$



ΛΥΣΗ

4: Εφόσον $-3,5 < -1,96$, απορρίπτουμε τη μηδενική υπόθεση, δηλαδή αποφασίζουμε ότι με επίπεδο σημαντικότητας 0,05, οι ώρες τις οποίες περνάνε οι φοιτητές/τριες στις καφετέριες έχουν μεταβληθεί.
Άρα, η σωστή απάντηση είναι το β)

2η Επαναληπτική Άσκηση

Μας ενδιαφέρει να διερευνήσουμε αν οι φοιτητές/τριες ηλικίας 18-25 τη σημερινή εποχή (έτος 2025) περνούν αρκετές ώρες στις καφετέριες με φίλους. Σύμφωνα με παλαιότερες έρευνες, (τελευταία έρευνα του 2012 που καθόρισε τη συμπεριφορά του πληθυσμού) οι φοιτητές/τριες περνούσαν κατά μέσο όρο 7 ώρες/εβδομάδα στις καφετέριες με απόκλιση 2 ωρών.

2η Επαναληπτική Άσκηση

Από μία έρευνα που διεξήχθη τον Απρίλιο (2025) με δείγμα 100 φοιτητές/τριες από διαφορετικά μέρη της Ελλάδας, προέκυψε ότι ο μέσος χρόνος που περνάνε σε καφετέριες είναι 7,2. Η έρευνα είχε σκοπό να ελέγξει αν η συγκεκριμένη συνήθεια των φοιτητών/τριών έχει αυξηθεί το έτος 2025 σε επίπεδο σημαντικότητας 0,05.

ΛΥΣΗ

Η λέξη **αυξηθεί** υποδηλώνει μονόπλευρο έλεγχο.

Αλλά, τί είδους;

α) Μονόπλευρο με $H_1: \mu_0 < \mu$

β) Μονόπλευρο με $H_1: \mu_0 > \mu$

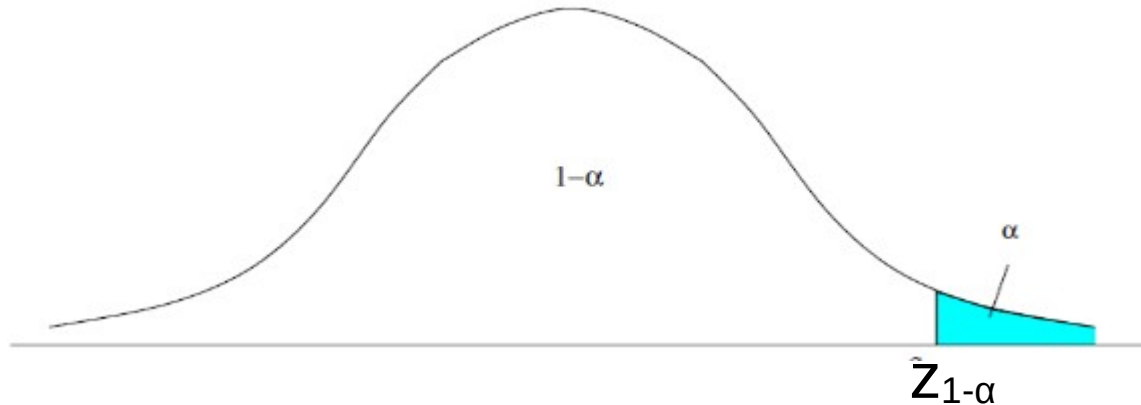
ΛΥΣΗ

Εφόσον μιλάμε για αύξηση, σημαίνει ότι ο πληθυσμός σήμερα θα παρουσιάζει μεγαλύτερο μέσο όρο ωρών από αυτόν που λέει η θεωρία, δηλαδή από τον μ_0 . Οπότε επιλέγουμε

α) Μονόπλευρο με $H_1: \mu_0 < \mu$

ΘΕΩΡΙΑ

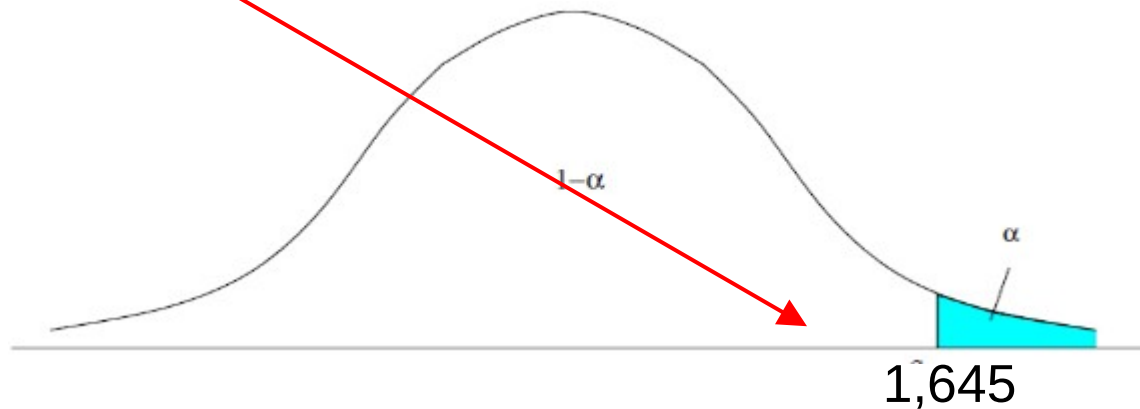
Στον συγκεκριμένο μονόπλευρο έλεγχο εξετάζουμε αν η τιμή που παίρνουμε από το στατικό κριτήριο “πέφτει” ή όχι στην περιοχή απόρριψης της μηδενικής υπόθεσης (ότι δεν αλλάζει ή μειώνεται). Το όριο της περιοχής αυτής καθορίζεται (γαλάζια περιοχή) από τον αριθμό $z_{1-\alpha}$.



ΛΥΣΗ

Υπολογίζουμε την τιμή που δίνει το κριτήριο και βλέπουμε που “πέφτει” σε σχέση με το $z_{0,975}=1,645$:

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{7,2 - 7}{2/\sqrt{100}} = \frac{2}{2} = 1$$



ΛΥΣΗ

Αφού $1 < 1,645$, μπορούμε σε επίπεδο σημαντικότητας 0,05 να πούμε ότι τελικά οι φοιτητές/τριες δεν έχουν αυξήσει τις ώρες που περνάνε στις καφετέριες.

3η Επαναληπτική Άσκηση

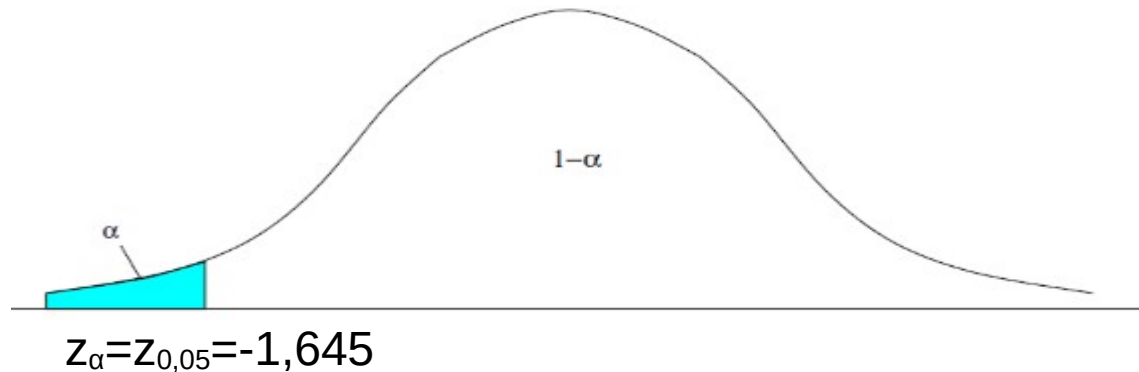
Μας ενδιαφέρει να διερευνήσουμε αν οι φοιτητές/τριες ηλικίας 18-25 τη σημερινή εποχή (έτος 2025) περνούν αρκετές ώρες στις καφετέριες με φίλους. Σύμφωνα με παλαιότερες έρευνες, (τελευταία έρευνα του 2012 που καθόρισε τη συμπεριφορά του πληθυσμού) οι φοιτητές/τριες περνούσαν κατά μέσο όρο 7 ώρες/εβδομάδα στις καφετέριες με απόκλιση 2 ωρών.

3η Επαναληπτική Άσκηση

Από μία έρευνα που διεξήχθη τον Απρίλιο (2025) με δείγμα 100 φοιτητές/τριες από διαφορετικά μέρη της Ελλάδας, προέκυψε ότι ο μέσος χρόνος που περνάνε σε καφετέριες είναι 6,8. Η έρευνα είχε σκοπό να ελέγξει αν η συγκεκριμένη συνήθεια των φοιτητών/τριών έχει ελαττωθεί το έτος 2025 σε επίπεδο σημαντικότητας 0,05.

ΘΕΩΡΙΑ

Η λέξη **έχει ελαττωθεί** παραπέμπει σε μονόπλευρο έλεγχο με $H_1: \mu_0 > \mu$. Υπολογίζουμε και εδώ το κριτήριο και ελέγχουμε αν “πέφτει” στην περιοχή απόρριψης της μηδενικής υπόθεσης, η οποία οριοθετείται από τον αριθμό z_α .



ΛΥΣΗ

Υπολογίζουμε τη τιμή του κριτηρίου

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{6,8 - 7}{2/\sqrt{100}} = \frac{-2}{2} = -1$$

Αφού $-1,645 < -1$, η τιμή “πέφτει” μέσα στην περιοχή που δεχόμαστε τη μηδενική υπόθεση και συνεπώς αποφασίζουμε ότι η συνήθεια των φοιτητών/τριών δεν μειώθηκε.

ΤΕΛΟΣ