

Στατιστική II

η προβλεπτική τιμή

$\hat{Y}_i$ η πρόβλεψη της τιμής του Y το X_i

Σφάλμα $e = Y_i - \hat{Y}_i$

Σημειώσεις - Λογμένο θέματα
e-students.gr
☎ 698 93 94 741

Απλή Γραμμική Παλινδρόμηση

Η Απλή Γραμμική Παλινδρόμηση (Simple Linear Regression) χρησιμοποιείται να προβλέψουμε και να προβλέψουμε τη σχέση ανάμεσα σε δύο μεταβλητές:

- Ανεξάρτητη μεταβλητή (x): προβλεπτική / εξαρτημένη
- Εξαρτημένη μεταβλητή (y): τιμή που θέλουμε να προβλέψουμε

Η υπόθεση είναι ότι η σχέση τους μπορεί να προσεγγιστεί από μία ευθεία:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

β_0 : σταθερός όρος

- β_0 «κατεβάζει» ή «ανεβάζει» όλη την συνθήκη στον άξονα y

β_1 : συντελεστής κλίση

- $\beta_1 > 0$: θετική σχέση (όσο αυξάνει το x , αυξάνει και το y)
- $\beta_1 < 0$: αρνητική σχέση (όσο αυξάνει το x , μειώνεται το y)
- $\beta_1 = 0$: δεν υπάρχει γραμμική σχέση

ε_i : σφάλμα (τυχαία απόκλιση)

Σημειώσεις - Λογμένο θέματα
e-studente.gr
☎ 698 93 94 741

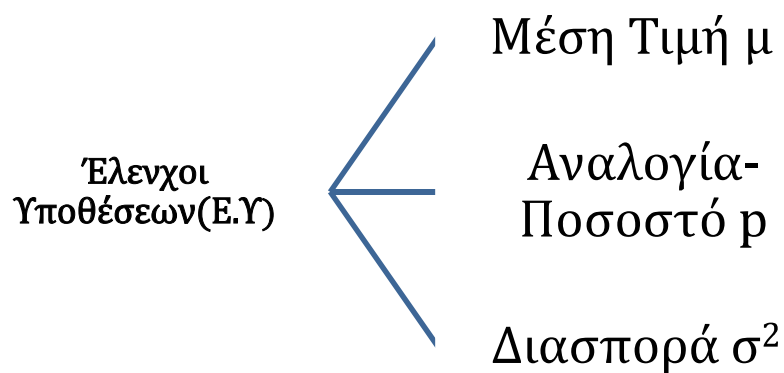
Έλεγχοι Υποθέσεων

Διαδικασία για να αποφασίσουμε αν τα δεδομένα μας συμφωνούν με έναν ισχυρισμό για τον πληθυσμό

Μηδενική υπόθεση H_0 : συνήθως ισχυρισμός ισότητας (π.χ. $\mu = \mu_0$)

Εναλλακτική υπόθεση H_1 : αυτό που θέλουμε να ελέγξουμε

1. Διπλής κατεύθυνσης (π.χ $\mu \neq \mu_0$)
2. Μονής κατεύθυνσης (π.χ $\mu > \mu_0$)
3. Μονής κατεύθυνσης (π.χ $\mu < \mu_0$)



Έλεγχος υπόθεσης για μέση τιμή μ

• Ένας πληθυσμός

A. Περίπτωση A: σ γνωστή

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

$\bar{X}$: η δειγματική μέση τιμή (ο μέσος όρος από το δείγμα μας)

μ_0 : η τιμή της μέσης τιμής που δηλώνει η μηδενική υπόθεση H_0

σ : η τυπική απόκλιση του πληθυσμού (γνωστή).

n : το μέγεθος του δείγματος (πόσες παρατηρήσεις έχουμε)

$\sigma / \sqrt{n}$: το τυπικό σφάλμα του μέσου όρου

$H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$
Αν $ Z > z_{\alpha/2}$ απορρίπτουμε την H_0

$H_0: \mu = \mu_0$
$H_1: \mu > \mu_0$
Αν $Z > z_{\alpha}$ απορρίπτουμε την H_0

$H_0: \mu = \mu_0$
$H_1: \mu < \mu_0$
Αν $Z < -z_{\alpha}$ απορρίπτουμε την H_0

B. Περίπτωση B: σ άγνωστη

$$T = \frac{\bar{X} - \mu_0}{s / \sqrt{n}}$$

$H_0: \mu = \mu_0$
$H_1: \mu \neq \mu_0$
Αν $ T > t_{\alpha/2, n-1}$ απορρίπτουμε την H_0

$H_0: \mu = \mu_0$
$H_1: \mu > \mu_0$
Αν $T > t_{\alpha, n-1}$ απορρίπτουμε την H_0

$H_0: \mu = \mu_0$
$H_1: \mu < \mu_0$
Αν $T < -t_{\alpha, n-1}$ απορρίπτουμε την H_0

• Δύο Πληθυσμοί – Έλεγχος διαφοράς μέσων ($\mu_1 - \mu_2$)

A. Περίπτωση A: γνωστές διασπορές

$$Z = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

$\bar{X}_1$: η δειγματική μέση τιμή από τον πρώτο πληθυσμό

$\bar{X}_2$: η δειγματική μέση τιμή από τον δεύτερο πληθυσμό

d_0 : η υποθετούμενη διαφορά μέσων κάτω από το H_0 . Συνήθως $d_0=0$ (δηλαδή ελέγχουμε αν οι μέσοι είναι ίσοι). Μπορεί να είναι και άλλη τιμή, αν θέλουμε να ελέγξουμε διαφορά ίση με κάποια σταθερά (π.χ. 5 μονάδες)

σ_1^2 : η διασπορά του πρώτου πληθυσμού (γνωστή)

σ_2^2 : η διασπορά του δεύτερου πληθυσμού (γνωστή)

n_1 : το μέγεθος δείγματος από τον πρώτο πληθυσμό

n_2 : το μέγεθος δείγματος από τον δεύτερο πληθυσμό

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 \neq \mu_2$

Αν $|Z| > z_{\alpha/2}$
απορρίπτουμε
την H_0

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 > \mu_2$

Αν $Z > z_\alpha$
απορρίπτουμε
την H_0

$H_0: \mu_1 = \mu_2$

$H_1: \mu_1 < \mu_2$

Αν $Z < -z_\alpha$
απορρίπτουμε
την H_0

Β. Περίπτωση Β: άγνωστες αλλά ίσες διασπορές

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}$$

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Αν $|T| > t_{\alpha/2, n_1+n_2-2}$
απορρίπτουμε την
 H_0

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 > \mu_2$$

Αν $T > t_{\alpha, n_1+n_2-2}$
απορρίπτουμε
την H_0

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 < \mu_2$$

Αν $T < -t_{\alpha, n_1+n_2-2}$
απορρίπτουμε
την H_0

Γ. Περίπτωση Γ: άγνωστες αλλά άνισες διασπορές

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - d_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 \neq \mu_2$$

Αν $|T| > t_{\alpha/2, df}$
απορρίπτουμε
την H_0

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 > \mu_2$$

Αν $T > t_{\alpha, df}$
απορρίπτουμε
την H_0

$$H_0: \mu_1 = \mu_2$$

$$H_1: \mu_1 < \mu_2$$

Αν $T < -t_{\alpha, df}$
απορρίπτουμε
την H_0

$$df = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

Άσκηση

Μια εταιρεία ισχυρίζεται ότι η μέση διάρκεια ζωής των λαμπτήρων της είναι 1000 ώρες. Ένα δείγμα 64 λαμπτήρων έδωσε μέση διάρκεια 980 ώρες, με γνωστή διασπορά $\sigma=80$ ώρες. Να ελεγχθεί στο 5% αν η μέση διάρκεια ζωής διαφέρει από 1000 ώρες.

Λύση

$$\bar{X}=980, \mu_0=1000, \sigma=80, n=64$$

$$\alpha=0,05 \rightarrow \alpha/2=0,025$$

$$z_{0,025}=1,96$$

$$H_0:\mu=1000$$

$$H_1:\mu \neq 1000$$

$$Z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

$$\Rightarrow Z = \frac{980 - 1000}{\frac{80}{\sqrt{64}}}$$

$$\Rightarrow Z = \frac{-20}{\frac{80}{8}}$$

$$\Rightarrow Z = \frac{-20}{10}$$

$$\Rightarrow Z = -2$$

Υπολογίσαμε $Z = -2,00$

Επειδή $|-2,00| > 1,96 \rightarrow$ **Απορρίπτουμε την H_0**

Στο 5% επίπεδο σημαντικότητας, τα δεδομένα δείχνουν ότι η μέση διάρκεια ζωής διαφέρει από τις 1000 ώρες.

Άσκηση

Ένας κατασκευαστής ισχυρίζεται ότι ο μέσος όγκος γέμισης ενός σακουλιού είναι τουλάχιστον 500 ml. Δείγμα 9 σακουλιών έχει μέσο 492 ml και δειγματική τυπική απόκλιση $s = 6$ ml. Ελέγξτε στο $\alpha=0.05$ εάν ο πραγματικός μέσος είναι μικρότερος από 500.

Λύση

$$\bar{X}=492, \mu_0=500, s=6, n=9$$

$$\alpha=0.05, df=8 \rightarrow t_{0.05,8} \approx 1.86$$

$$H_0: \mu=500$$

$$H_1: \mu < 500$$

$$T = \frac{\bar{X} - \mu_0}{s / \sqrt{n}}$$

$$\Rightarrow T = \frac{492 - 500}{\frac{6}{\sqrt{9}}}$$

$$\Rightarrow T = \frac{-8}{\frac{6}{3}}$$

$$\Rightarrow T = \frac{-8}{2}$$

$$\Rightarrow T = -4$$

Υπολογίσαμε $T = -4$

Επειδή $-4 < -1.86 \rightarrow$ **Απορρίπτουμε την H_0**

Σε επίπεδο 5% υπάρχουν ισχυρά στοιχεία ότι ο μέσος όγκος είναι μικρότερος από 500 ml