Παραδείγματα στα

Διαστήματα Εμπιστοσύνης

1) Για να εκτιμήσουμε τη **μέση τιμή του λίτρου βενζίνης** στα πρατήρια μιας περιοχής, επισκεπτόμαστε 10 πρατήρια και καταγράφουμε τις τιμές:

$$X=\{1.67, 1.73, 1.71, 1.70, 1.65, 1.64, 1.62, 1.69, 1.77, 1.68\}$$

α) Να βρείτε το **δ.ε. του μέσου** με συντελεστή εμπιστοσύνης ***95%***.

β) Πόσο θα ήταν το **δ.ε. του μέσου** εάν ήταν γνωστό ότι $\sigma=0.04$;

γ) Να βρείτε το **δ.ε. για την διασπορά** της τιμής,

δ) Πόσο θα ήταν το **δ.ε. της διασποράς** εάν είναι γνωστό ότι $\mu=1.70$;

Λύση

(α) δ.ε. του μέσου με συντελεστή εμπιστοσύνης *95%*.

- Περίπτωση

$$\overline{X} - \frac{S}{\sqrt{n}} t_{n-1}(a/2) \le \mu \le \overline{X} + \frac{S}{\sqrt{n}} t_{n-1}(a/2)$$

- Από το δείγμα των n=10 τιμών βρίσκουμε τα

$$\overline{X} = \sum_{i=1}^{n} X_i = 1.686 \qquad S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = 0.0020$$

- Βρίσκουμε από τον πίνακα της Student-t την τιμή του

$$t_{n-1}(a/2) = t_9(0.025) = 2.262$$

- **Αντικαθιστώντας βρίσκουμε το δ.ε.**

$$1.646 \le \mu \le 1.6876$$

Λύση

(β) **δ.ε. του μέσου εάν ήταν γνωστό ότι *σ=0.04***

- Περίπτωση (***σ_0 = 0.04***)

$$\overline{X} - \frac{\sigma_0}{\sqrt{n}} z_{a/2} \le \mu \le \overline{X} + \frac{\sigma_0}{\sqrt{n}} z_{a/2}$$

- Από το δείγμα των n=10 τιμών βρίσκουμε το

$$\overline{X} = \sum_{i=1}^{n} X_i = 1.686$$

- Βρίσκουμε από τον πίνακα της κανονικής κατανομής την τιμή του

$$z_{a/2} = z_{0.025} = 1.96$$

- **Αντικαθιστώντας βρίσκουμε το δ.ε.**

$$1.661 \le \mu \le 1.711$$

Λύση

(γ) δ.ε. για την διασπορά της τιμής

- Περίπτωση

$$\frac{(n-1)S^2}{\chi^2_{n-1}(a/2)} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{n-1}(1-a/2)}$$

- Από το δείγμα των n=10 τιμών βρίσκουμε το

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2$$

- Βρίσκουμε από τον πίνακα της χι-τετράγωνο κατανομής τις τιμές των

$$\chi^2_{n-1}(a/2) = \chi^2_9(0.025) \qquad \chi^2_{n-1}(1-a/2) = \chi^2_9(0.975)$$

- **Αντικαθιστώντας** βρίσκουμε το δ.ε.

Λύση

(δ) **δ.ε. της διασποράς εάν είναι γνωστό ότι μ=1.70**

- Περίπτωση (**μ=1.70**)
  (γνωστό μέσο)

$$\frac{\sum_{i=1}^{n}\left(X_i-\mu\right)^2}{\chi_n^2\left(a/2\right)} \le \sigma^2 \le \frac{\sum_{i=1}^{n}\left(X_i-\mu\right)^2}{\chi_n^2\left(1-a/2\right)}$$

- Από το δείγμα των n=10 τιμών βρίσκουμε το

$$\sum_{i=1}^{n}(X_i-\mu)^2$$

- Βρίσκουμε από τον πίνακα της χι-τετράγωνο κατανομής τις τιμές των

$$\chi_{10}^2\left(0.025\right)=20.483 \qquad \chi_{10}^2\left(1-\frac{0.05}{2}\right)=\chi_{10}^2\left(0.975\right)=3.247$$

- **Αντικαθιστώντας** βρίσκουμε το δ.ε.

2) Ζητείται να εκτιμήσουμε το μέσο χρόνο ζωής των κατοίκων μιας περιοχής. Για τον σκοπό αυτό πήραμε ένα τυχαίο δείγμα ***n=50*** ατόμων και βρήκαμε ότι $\boldsymbol{\Sigma_i X_i = 3959{,}6}$, $\boldsymbol{\Sigma_i X^2_i = 77265}$.

Να βρεθεί για το χρόνο ζωής

α) το **δ.ε. για τον μέσο** με συντελεστή εμπιστοσύνης **95%** και

β) το **δ.ε. για την τυπική απόκλιση** με συντελεστή εμπιστοσύνης **95%**

Λύση

(α) δ.ε. για τον μέσο με συντελεστή εμπιστοσύνης 95%

- Περίπτωση $$\bar{X} - \frac{S}{\sqrt{n}} t_{n-1}(a/2) \leq \mu \leq \bar{X} + \frac{S}{\sqrt{n}} t_{n-1}(a/2)$$

- Ισχύει ότι $$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 + n\bar{X}^2 - 2\bar{X}\sum_{i=1}^{n} X_i\right) =$$
$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right)$$

- Βρίσκουμε από τον πίνακα της t-Student την τιμή του

$$t_{n-1}(a/2) = t_{49}(0.025) \approx 2.021$$

- Αντικαθιστώντας βρίσκουμε το **διάστημα εμπιστοσύνης του μέσου**

Λύση

(β) **δ.ε. για την τυπική απόκλιση με συντελεστή εμπιστοσύνης 95%**

- Περίπτωση

$$\frac{(n-1)S^2}{\chi^2_{n-1}(a/2)} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{n-1}(1-a/2)}$$

- Υπολογίζουμε το ***S***^{2}

$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 + n\bar{X}^2 - 2\bar{X}\sum_{i=1}^{n} X_i\right) =$$

$$= \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right) = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - \frac{\left(\sum_{i=1}^{n} X_i\right)^2}{n}\right)$$

Λύση

(β) **δ.ε. για την τυπική απόκλιση με συντελεστή εμπιστοσύνης 95%**

- Περίπτωση

$$\frac{(n-1)S^2}{\chi^2_{n-1}(a/2)} \le \sigma^2 \le \frac{(n-1)S^2}{\chi^2_{n-1}(1-a/2)}$$

- Υπολογίζουμε το ***S*²**

- Βρίσκουμε από τον **πίνακα της χι-τετράγωνο** την τιμή των

$$\chi^2_{49}(0.025) \approx 71.420 \qquad \chi^2_{49}\left(1-\frac{0.05}{2}\right) = \chi^2_{49}(0.975) \approx 32.357$$

- **Αντικαθιστώντας** βρίσκουμε το **διάστημα εμπιστοσύνης της διασποράς**

Παραδείγματα στον

Έλεγχο Υποθέσεων

Παραδείγματα σε «**Ελέγχους Υποθέσεων**»

- Το 2015 η μέση ηλικία των διευθυντικών στελεχών μεγάλων επιχειρήσεων στις ΗΠΑ ήταν 48 ετών. Σε ένα τυχαίο δείγμα μεγέθους 25 βρέθηκε ότι η μέση τιμή τέτοιων στελεχών ήταν 46 και η τυπική απόκλιση 5. Να βρεθεί σε στάθμη σημαντικότητας 1%, αν η μέση ηλικία στις μεγάλες εταιρίες άλλαξε από το 2015.

Λύση

$$\overline{X} = 46 \quad , \quad S = 5 \quad , \quad n = 25$$

$\alpha = 0.01$

$H_0 : \mu = \mu_0 = 48$

$H_1 : \mu \neq \mu_0$

$$\overline{X} \geq \mu_0 + \frac{S}{\sqrt{n}} t_{n-1}(a/2) \qquad \overline{X} \leq \mu_0 - \frac{S}{\sqrt{n}} t_{n-1}(a/2)$$

$$t_{24}(0.005) = 2.797$$

48 + 2.5 = 50.5 ή 45.5

~~46 ≥ 50.5~~ ~~46 ≤ 45.5~~

Επομένως η *H_0* ΔΕΝ απορρίπτεται

- Ο μέσος χρόνος μεταξύ δύο επισκευών των φωτοτυπικών μηχανημάτων που χρησιμοποιούνται στα σχολεία μιας πόλης πέρυσι ήταν 135 μέρες χρήσης. Αν φέτος σε ένα δείγμα 25 τέτοιων μηχανημάτων η μέση τιμή του χρόνου X μεταξύ δύο επισκευών βρέθηκε να είναι 131 ημέρες με τυπική απόκλιση 11 ημέρες, να εξεταστεί σε στάθμη σημαντικότητας 5% αν αυτός ο χρόνος μειώθηκε σε σχέση με πέρυσι.

Λύση

$$\overline{X} = 131 \quad , \quad S = 11 \quad , \quad n = 25$$

$H_0 : \mu = \mu_0 = 135$

$H_1 : \mu < \mu_0$

$$\overline{X} \leq \mu_0 - \frac{S}{\sqrt{n}} t_{n-1}(a)$$

t_{24} (0.05) = 1,71

135-(11/5) * 1,71 = 131,23

131<131,23

επομένως **ΑΠΟΡΡΙΠΤΕΤΑΙ η *H_0*** και άρα μπορούμε να ισχυριστούμε ότι ο χρόνος μειώθηκε

- Να ελεγχθεί σε στάθμη σημαντικότητας 5% αν το ποσοστό των ψηφοφόρων ενός κόμματος είναι 40%, αν σε ένα δείγμα 150 ατόμων βρέθηκαν 68 ψηφοφόροι του κόμματος. Να βρεθεί επίσης η ***P-value*** του ελέγχου.

Λύση

$$\overline{X} = \frac{68}{150} = 0.453 \;,\; a = 0.05$$

$H_0 : \rho = \boldsymbol{\rho_0} \mathbf{=0.4}$

$\mathbf{H_1 : \rho \neq \rho_0}$

$$\overline{X} \geq \rho_0 + \frac{\sqrt{\rho_0(1-\rho_0)}}{\sqrt{n}} z_{a/2}$$

$$\overline{X} \leq \rho_0 - \frac{\sqrt{\rho_0(1-\rho_0)}}{\sqrt{n}} z_{a/2}$$

$\boldsymbol{z_{a/2} = z_{0.025} = 1{,}96}$

$0.4 \pm 0.078 =$ **0.4784** **ή** **0.3216**

Έτσι **0.3216 < 0.453 < 0.4784** **ΔΕΝ απορρίπτεται η H_0** - ποσοστό 40%

$$P-value = 2\left(1-\Phi\left(\frac{\bar{x}-\mu_0}{\sigma_0/\sqrt{n}}\right)\right) = 2\left(1-\Phi\left(\frac{\bar{x}-\rho_0}{\sqrt{\rho_0(1-\rho_0)}/\sqrt{n}}\right)\right) = 2(1-\Phi(1.333)) = 2(1-0.9082) = 0.1836$$

Παρατηρούμε ότι ***a=0.05 < P-value*** *(άρα δεν απορρίπτουμε την H_0)*

- Γνωρίζουμε ότι το μήνα Νοέμβριο η μέση τιμή πώλησης ενός προϊόντος σε μια περιοχή είναι 1000 ευρώ. Η μέση τιμή πώλησης του ίδιου προϊόντος σε ένα τυχαίο δείγμα n=10 καταστημάτων το μήνα Δεκέμβριο βρέθηκε ίση με 1030 ευρώ. Μπορούμε σε επίπεδο σημαντικότητας 5% να πούμε ότι η μέση τιμή πώλησης αυξήθηκε κατά το μήνα Δεκέμβριο (γνωρίζουμε ότι η τυπική απόκλιση του προϊόντος είναι 50). Να βρεθεί στη συνέχεια το ***P-value***.

Λύση

$$\overline{X} = 1030 \quad , \quad \sigma_0 = 50 \quad , \quad n = 10$$

$H_0 : \mu = \mu_0 = 1000$

$H_1 : \mu > \mu_0$

$$\overline{X} \geq \mu_0 + \frac{\sigma_0}{\sqrt{n}} z_a$$

$a=0.05$, $z_a = 1,64$

1025.93

επειδή ***1030 > 1025,93*** **ΑΠΟΡΡΙΠΤΕΤΑΙ** η H_0

Παρατηρούμε ότι ***a=0.05 > P-value***
(άρα απορρίπτουμε την H_0)

$$P-value = \Phi\left(\frac{\mu_0 - \bar{x}}{\sigma_0 / \sqrt{n}}\right) = \Phi(-1.8974) = 1 - \Phi(1.8974) = 1 - 0.977 = 0.023$$

- Σε προβλήματα ελέγχου ποιότητας εκτός από τη διατήρηση ενός σταθερού μέσου μας ενδιαφέρει και η διατήρηση της διασποράς σε χαμηλά επίπεδα (διότι αυξάνεται ο κίνδυνος απόρριψης του προϊόντος). Από τη παραγωγή τυχαίο δείγμα ***n=16*** έδωσε δειγματική απόκλιση ***5.5***. Αν η μέγιστη επιτρεπτή τυπική απόκλιση είναι ***4*** να εξεταστεί εάν η παραπάνω υπέρβαση είναι στατιστικά σημαντική ή όχι (***α=0.05***).

Λύση

$H_0 : \sigma^2 = \sigma^2_0 = 4^2$

$H_1 : \sigma^2 > \sigma^2_0$

$$S = 5.5 \quad , \quad n = 16$$

$$S^2 \geq \frac{\sigma_0^2}{n-1} x_{n-1}^2(a)$$

$\chi^2_{15}(0.05) = 24.996$

$16/15 * 24.996 = 26.662$

Καθώς ισχύει ότι $S^2 = 5.5^2 = 30.25 > 26.662$ η H_0 **απορρίπτεται** και επομένως μπορούμε να υποθέσουμε ότι η απόκλιση μεγάλωσε **σημαντικά**.

- Σε ένα τυχαίο δείγμα 400 τηλεθεατών οι 100 δήλωσαν ότι παρακολουθούν μία ορισμένη τηλεοπτική σειρά. Μπορούμε να δεχτούμε σε επίπεδο εμπιστοσύνης 5%, ότι εκείνοι που παρακολουθούν τη σειρά είναι λιγότεροι το 30% του συνόλου των τηλεθεατών; Να βρεθεί στη συνέχεια η P-τιμή του δείγματος?

Λύση

H0: $\rho = \rho_0 = 0.3$

H1: $\rho < \rho_0$

Περίπτωση:

$$\bar{X} \le \rho_0 - \frac{\sqrt{\rho_0(1-\rho_0)}}{\sqrt{n}} z_a$$

mean(X)=100/400=0.25

α=0.05 και z_α = 1.64

$$\rho_0 - \frac{\sqrt{\rho_0(1-\rho_0)}}{\sqrt{n}} z_a = 0.3 - \sqrt{\frac{0.3*0.7}{400}}1.64 = 0.2624$$

$$\bar{X} = 0.25 \le 0.2624$$

Άρα απορρίπτεται η H0 και επομένως μειώθηκε η τηλεθέαση από 30%

$$P-value = P(Y < c(x)) = \Phi\left(\frac{\bar{X}-\mu_0}{\sigma_0/\sqrt{n}}\right) = \Phi\left(\frac{\bar{X}-\rho_0}{\sqrt{\rho_0(1-\rho_0)}/\sqrt{n}}\right) = \Phi(-2.1822) = 1-\Phi(2.1822) = 1-0.9854 = 0.0146$$

Έτσι με β.ε. 5% απορρίπτεται η H0 καθώς α=0.05>0.014 , αλλά

Με β.ε. 1% ΔΕΝ απορρίπτεται η H0 καθώς α=0.01 < 0.014