Πανεπιστήμιο Πελοποννήσου

Εκτιμήσεις

Διαστήματα Εμπιστοσύνης

Έλεγχοι Υποθέσεων

Εκτιμητική

Οι κατανομές των *στατιστικών* έχουν άγνωστες παραμέτρους, οι οποίες πρέπει να εκτιμηθούν

Εκτιμητές σε σημείο (Σημειακοί εκτιμητές) μιας παραμέτρου *θ* είναι μια συνάρτηση από την οποία μπορούμε να υπολογίσουμε την αριθμητική τιμή της παραμέτρου, χρησιμοποιώντας τις μετρήσεις του δείγματος. Ο αριθμός που υπολογίζεται, λέγεται εκτιμητής σε σημείο της παραμέτρου

Εκτιμητές σε διάστημα είναι ένας τύπος από τον οποίο μπορούμε να κατασκευάσουμε ένα διάστημα στο οποίο ανήκει η παράμετρος του πληθυσμού με συγκεκριμένο βαθμό βεβαιότητας, χρησιμοποιώντας τις μετρήσεις του δείγματος

Εκτιμητές σε σημείο (Point Estimators)

Για την κατασκευή *Εκτιμητών σε Σημείο* των παραμέτρων μιας κατανομής χρησιμοποιούμε:

1. Μέθοδο των Ροπών
2. Μέθοδο Μέγιστης Πιθανοφάνειας
3. Μέθοδος Ελάχιστων Τετραγώνων

Μέθοδος των Ροπών
(Method of Moments)

Για να κατασκευάσουμε σημειακούς εκτιμητές, εξισώνουμε τις θεωρητικές ροπές που περιέχουν τις άγνωστες παραμέτρους με τις αντίστοιχες δειγματικές ροπές

Θεωρητική Ροπή k τάξεως	$\mu_k = \sum_x x^k P(X = x)$
Δειγματική Ροπή k τάξεως	$m_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k$

Μέθοδος Μέγιστης Πιθανοφάνειας (Method of Maximum Likelihood)

Ονομάζεται ***πιθανοφάνεια*** και συμβολίζεται με L(θ), η κατανομή του δείγματος, όταν θεωρείται συνάρτηση της άγνωστης παραμέτρου θ.

Αν f(x) είναι η κατανομή από όπου προέρχεται το δείγμα, τότε:

$$L(\theta)=\prod_{i=1}^{n} f(x_i)$$

Ο ***εκτιμητής μέγιστης πιθανοφάνειας*** (ΕΜΠ) της άγνωστης παραμέτρους θ βρίσκεται μεγιστοποιώντας τη συνάρτηση L(θ) ως προς το θ

Παρατήρηση

Οι ροπο-εκτιμητές και οι εκτιμητές μέγιστης πιθανοφάνειας σε άλλες κατανομές συμπίπτουν και σε άλλες δεν συμπίπτουν. Όταν δεν συμπίπτουν, διαλέγουμε τον εκτιμητή με τις καλύτερες ιδιότητες από πιθανοθεωρητικής άποψης

Τέτοιες ιδιότητες είναι

1. Αμεροληψία
2. Ελάχιστη Διακύμανση
3. Συνέπεια
4. Επάρκεια

Εκτιμητές παραμέτρων του πληθυσμού

- Μέση τιμή (μ) → $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$
- Αναλογία (p) → $\hat{p} = \frac{n_A}{n}$
- Διασπορά (σ^2) → $\hat{\sigma}^2 = s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$

$$V(\bar{x}) = s^2 / n$$

$$V(\hat{p}) = \frac{\hat{p}(1-\hat{p})}{n}$$

Πανεπιστήμιο Πελοποννήσου

Διαστήματα Εμπιστοσύνης

Διαστήματα Εμπιστοσύνης
Εκτιμητές σε διάστημα

Έστω $x_1, x_2, \ldots, x_n$ τυχαίο δείγμα και θ μία άγνωστη παράμετρος του πληθυσμού. Επιπλέον, έστω 1-α μία μεγάλη πιθανότητα. Τότε το διάστημα (Α,Δ) του οποίου τα άκρα Α (αριστερό) και Δ (δεξιό) ορίζονται από τη σχέση

$$P(A<\theta<\Delta)=1-\alpha$$

Λέγεται 100(1-α)% διάστημα εμπιστοσύνης για την παράμετρο θ και η πιθανότητα 1-α λέγεται συντελεστής εμπιστοσύνης.

Διάστημα εμπιστοσύνης για τη μέση τιμή μ του πληθυσμού (διασπορά γνωστή) Δείγμα μεγάλο n≥30

$$A = \bar{x} - z_{a/2} \frac{\sigma}{\sqrt{n}}$$

$$\Delta = \bar{x} + z_{a/2} \frac{\sigma}{\sqrt{n}}$$

Όταν το δείγμα προέρχεται από κανονικό πληθυσμό, οι παραπάνω τύποι ισχύουν και για μικρό δείγμα

Διάστημα εμπιστοσύνης για τη μέση τιμή μ του πληθυσμού (διασπορά άγνωστη) Δείγμα μεγάλο n≥30

$$A = \overline{x} - z_{a/2} \frac{s}{\sqrt{n}}$$

$$\Delta = \overline{x} + z_{a/2} \frac{s}{\sqrt{n}}$$

Διάστημα εμπιστοσύνης για τη μέση τιμή μ του πληθυσμού (διασπορά άγνωστη) Δείγμα μικρό n<30

$$A = \bar{x} - t_{n-1,a/2}\frac{s}{\sqrt{n}}$$

$$\Delta = \bar{x} + t_{n-1,a/2}\frac{s}{\sqrt{n}}$$

Το δείγμα πρέπει να προέρχεται από κανονικό πληθυσμό

Διάστημα εμπιστοσύνης για τη διαφορά μέσων τιμών δύο πληθυσμών. (δείγματα ανεξάρτητα, n & m≥30)

$$A = \left(\overline{x} - \overline{y}\right) - z_{a/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$$

$$\Delta = \left(\overline{x} - \overline{y}\right) + z_{a/2}\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}$$

Όταν οι διασπορές είναι άγνωστες, χρησιμοποιούμε τις αντίστοιχες δειγματικές διασπορές

Διάστημα εμπιστοσύνης για τη διαφορά μέσων τιμών δύο πληθυσμών. (δείγματα ανεξάρτητα, n & m<30)

$$A = \left(\overline{x} - \overline{y}\right) - t_{n+m-2;a/2} s_{pooled} \sqrt{\frac{1}{n} + \frac{1}{m}}$$

$$\Delta = \left(\overline{x} - \overline{y}\right) + t_{n+m-2;a/2} s_{pooled} \sqrt{\frac{1}{n} + \frac{1}{m}}$$

$$s_{pooled}^2 = \frac{(n-1)s_1^2 + (m-1)s_2^2}{n+m-2}$$

Τα δείγματα πρέπει να προέρχονται από κανονικούς πληθυσμούς

Διάστημα εμπιστοσύνης για τη διαφορά των μέσων τιμών δύο πληθυσμών (Δείγματα εξαρτημένα – Ζευγαρωτές παρατηρήσεις)

$$A = \bar{z} - t_{n-1,a/2}\frac{s_z}{\sqrt{n}}$$

$$\Delta = \bar{z} + t_{n-1,a/2}\frac{s_z}{\sqrt{n}}$$

Το ζεύγη των παρατηρήσεων πρέπει να προέρχεται από κανονικό πληθυσμό

Όταν n≥30, τότε αντί για t χρησιμοποιούμε την κανονική κατανομή $z_{a/2}$

Διάστημα εμπιστοσύνης για την αναλογία p ενός πληθυσμού
Δείγμα μεγάλο n≥30

$$A = \hat{p} - z_{a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$\Delta = \hat{p} + z_{a/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Διάστημα εμπιστοσύνης για τη διαφορά p_1-p_2, των αναλογιών δύο πληθυσμών Δείγμα μεγάλο $n \geq 30$

$$A = \hat{p}_1 - \hat{p}_2 - z_{a/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}$$

$$\Delta = \hat{p}_1 - \hat{p}_2 + z_{a/2}\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n} + \frac{\hat{p}_2(1-\hat{p}_2)}{m}}$$

Διάστημα εμπιστοσύνης για τη διασπορά σ^2 ενός πληθυσμού

$$A = \frac{(n-1)s^2}{X^2_{n-1,a/2}}$$

$$\Delta = \frac{(n-1)s^2}{X^2_{n-1,1-a/2}}$$

Το δείγμα πρέπει να προέρχεται από κανονικό πληθυσμό

Διάστημα εμπιστοσύνης για το λόγο των διασπορών δύο πληθυσμών

$$\frac{s_1^2}{s_2^2}\frac{1}{F_{n-1,m-1,a/2}} < \frac{\sigma_1^2}{\sigma_1^2} < \frac{s_1^2}{s_2^2}F_{m-1,n-1,a/2}$$

Τα δείγματα πρέπει να προέρχεται από κανονικούς πληθυσμούς

Πανεπιστήμιο Πελοποννήσου

Έλεγχοι Υποθέσεων

Εισαγωγή

Σκοπός των στατιστικών ελέγχων υποθέσεων είναι η πραγματοποίηση κάποιων ελέγχων αναφορικά με τις παραμέτρους του πληθυσμού

Παράδειγμα

Για μία ασθένεια υπάρχει φάρμακο που θεραπεύει μέσα σε 10 μέρες. Ένα νέο φάρμακο ισχυρίζεται πως θεραπεύει σε συντομότερο χρονικό διάστημα. Σκοπός είναι να ελεγχθεί αν ο μέσος όρος θεραπείας του νέου φαρμάκου είναι μικρότερο από 10 μέρες.

Η υπόθεση $\mu=10$ λέγεται **μηδενική υπόθεση H_0**

Η υπόθεση $\mu<10$ λέγεται **εναλλακτική υπόθεση H_1**

Στοιχεία στατιστικού ελέγχου

Η εναλλακτική υπόθεση ελέγχεται σε σχέση με την μηδενική υπόθεση.

Για να γίνει ο έλεγχος χρησιμοποιείται ένα τυχαίο δείγμα, και με βάση αυτό υπολογίζεται ένα στατιστικό (συνάρτηση των τιμών του δείγματος)

Απορριπτική περιοχή της H_0 ονομάζεται η περιοχή που όταν το στατιστικό του ελέγχου πάρει τιμές, τότε η H_0 απορρίπτεται. Η απορριπτική περιοχή συμβολίζεται με **R.**

Στοιχεία στατιστικού ελέγχου (2)

1. Η μηδενική υπόθεση H_0
2. Η εναλλακτική υπόθεση H_1
3. Το στατιστικό του συγκεκριμένου ελέγχου
4. Η απορριπτική περιοχή R

Με βάση τα παραπάνω γίνεται ο έλεγχος και εξάγονται συμπεράσματα

Μορφές ελέγχου

- Η μηδενική υπόθεση είναι πάντα της μορφής H_0:θ=$θ_0$, όπου θ είναι η παράμετρος του πληθυσμού που ελέγχεται και $Θ_0$ είναι μία τιμή
- Η εναλλακτική υπόθεση παίρνει τις παρακάτω μορφές
  - $θ>θ_0$ (μονόπλευρος έλεγχος)
  - $Θ<θ_0$ (μονόπλευρος έλεγχος)
  - $θ\neqθ_0$ (δίπλευρος έλεγχος)

Σφάλματα

- Ονομάζεται σφάλμα τύπου I η απόρριψη της μηδενικής υπόθεσης ενώ είναι σωστή. Η πιθανότητα αυτού του σφάλματος συμβολίζεται **a.** Άρα a=P(απόρριψη της $H_0|H_0$ σωστή)
- Ονομάζεται σφάλμα τύπου II η αποδοχή της μηδενικής υπόθεσης ενώ είναι λάθος. Η πιθανότητα αυτού του σφάλματος συμβολίζεται **β.** Άρα β=P(αποδοχή της $H_0|H_0$ λάθος)

Ισχύς Ελέγχου και Παρατηρούμενη επίπεδο σημαντικότητας

- Ισχύς ενός στατιστικού ελέγχου ονομάζεται η πιθανότητα απόρριψης της H_0 όταν η H_0 είναι λάθος. Συμβολίζεται **γ=1-β**

- **Παρατηρούμενη επίπεδο σημαντικότητας** ονομάζεται η πιθανότητα να παρατηρηθεί μία τιμή του στατιστικού μεγαλύτερη από αυτή που έδωσε το δείγμα.

Έλεγχος υπόθεσης για τη μέση τιμή μ του πληθυσμού (διασπορά γνωστή)

Δείγμα μεγάλο n≥30 δεδομενα από οποιαδήποτε κατανομή
Δείγμα μικρόο n<30 δεδομενα από κανονική κατανομή

$H_0: \mu=\mu_0$ $H_1: \mu>\mu_0$	$H_0: \mu=\mu_0$ $H_1: \mu<\mu_0$	$H_0: \mu=\mu_0$ $H_1: \mu\neq\mu_0$
$R=\{z>z_a\}$	$R=\{z<-z_a\}$	$R=\{\{z>z_{a/2}\}$ ή $\{z<-z_{a/2}\}\}$

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}}$$

Έλεγχος υπόθεσης για τη μέση τιμή μ του πληθυσμού (διασπορά άγνωστη)

Δείγμα μεγάλο $n \geq 30$ δεδομένα από οποιαδήποτε κατανομή (προσεγγιστικά)
Δείγμα μικρό $n < 30$ δεδομένα από κανονική κατανομή

$H_0: \mu = \mu_0$ $H_1: \mu > \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu < \mu_0$	$H_0: \mu = \mu_0$ $H_1: \mu \neq \mu_0$
$R = \{t > t_{n-1,a}\}$	$R = \{t < -t_{n-1,a}\}$	$R = \{\{t > t_{n-1,a/2}\}$ ή $\{t < -t_{n-1,a/2}\}\}$

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Έλεγχος υπόθεσης για τη διαφορά των μέσων όρων δύο πληθυσμών (δείγματα ανεξάρτητα, διασπορές γνωστές ή άγνωστες) Δείγματα μεγάλα m & n ≥ 30

$H_0: \mu_1 - \mu_2 = \delta$ $H_1: \mu_1 - \mu_2 > \delta$	$H_0: \mu_1 - \mu_2 = \delta$ $H_1: \mu_1 - \mu_2 < \delta$	$H_0: \mu_1 - \mu_2 = \delta$ $H_1: \mu_1 - \mu_2 \neq \delta$
$R = \{z > z_a\}$	$R = \{z < -z_a\}$	$R = \{\{z > z_{a/2}\}$ ή $\{z < -z_{a/2}\}\}$

$$z = \frac{\bar{x} - \bar{y} - \delta}{\sqrt{\frac{\sigma_1^2}{n} + \frac{\sigma_2^2}{m}}}$$

ή

$$z = \frac{\bar{x} - \bar{y} - \delta}{\sqrt{\frac{s_1^2}{n} + \frac{s_2^2}{m}}}$$

Έλεγχος υπόθεσης για τη διαφορά των μέσων όρων δύο πληθυσμών (δείγματα ανεξάρτητα, διασπορές ίσες). Δείγματα μεγάλα m & n <30

$H_0: \mu_1-\mu_2=\delta$ $H_1: \mu_1-\mu_2>\delta$	$H_0: \mu_1-\mu_2=\delta$ $H_1: \mu_1-\mu_2<\delta$	$H_0: \mu_1-\mu_2=\delta$ $H_1: \mu_1-\mu_2\neq\delta$
$R=\{t>t_{n+m-2,a}\}$	$R=\{t<-t_{n+m-2,a}\}$	$R=\{\{t>t_{n+m-2,a/2}\}$ ή $\{t<-t_{n+m-2,a/2}\}\}$

$$t=\frac{\bar{x}-\bar{y}-\delta}{\sqrt{\frac{(n-1)s_1^2+(m-1)s_2^2}{n+m-2}}\sqrt{\frac{1}{n}+\frac{1}{m}}}$$

Έλεγχος υπόθεσης για τη διαφορά των μέσων όρων δύο πληθυσμών (δείγματα εξαρτημένα – ζευγαρωτές παρατηρήσεις)

$H_0: \mu_1 - \mu_2 = \delta$ $H_1: \mu_1 - \mu_2 > \delta$	$H_0: \mu_1 - \mu_2 = \delta$ $H_1: \mu_1 - \mu_2 < \delta$	$H_0: \mu_1 - \mu_2 = \delta$ $H_1: \mu_1 - \mu_2 \neq \delta$
$R = \{t > t_{n+m-2,a}\}$	$R = \{t < -t_{n+m-2,a}\}$	$R = \{\{t > t_{n+m-2,a/2}\}$ ή $\{t < -t_{n+m-2,a/2}\}\}$

$$t = \frac{\bar{z} - \delta}{\sqrt{s_z^2 / n}}$$

Προϋπόθεση: η κατανομή των διαφορών είναι περίπου κανονική

Τύποι για ζευγαρωτές παρατηρήσεις

$$\bar{z} = \frac{1}{n}\sum_{i=1}^{n}\left(x_i - y_i\right) = \frac{1}{n}\sum_{i=1}^{n} z_i$$

$$s_z^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(z_i - \bar{z}\right)^2$$

Έλεγχος υπόθεσης για την αναλογία ενός πληθυσμού

H_0: $p=p_0$ H_1: $p>p_0$	H_0: $p=p_0$ H_1: $p<p_0$	H_0: $p=p_0$ H_1: $p\neq p_0$
$R=\{z>z_a\}$	$R=\{z<-z_a\}$	$R=\{\{z>z_{a/2}\}$ ή $\{z<-z_{a/2}\}\}$

$$z=\frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

Έλεγχος υπόθεσης για τη διαφορά p_1-p_2 των αναλογιών δύο πληθυσμών

H_0: $p_1-p_2=\delta$ H_1: $p_1-p_2>\delta$	H_0: $p_1-p_2=\delta$ H_1: $p_1-p_2<\delta$	H_0: $p_1-p_2=\delta$ H_1: $p_1-p_2\neq\delta$
$R=\{z>z_a\}$	$R=\{z<-z_a\}$	$R=\{\{z>z_{a/2}\}$ ή $\{z<-z_{a/2}\}\}$

$$z=\frac{\hat{p}_1-\hat{p}_2-\delta}{\sqrt{\frac{\hat{p}_1(1-\hat{p}_1)}{n}+\frac{\hat{p}_2(1-\hat{p}_2)}{m}}}$$

Έλεγχος υπόθεσης για τη διασπορά ενός πληθυσμού

$H_0: \sigma^2=\sigma^2_0$ $H_1: \sigma^2>\sigma^2_0$	$H_0: \sigma^2=\sigma^2_0$ $H_1: \sigma^2<\sigma^2_0$	$H_0: \sigma^2=\sigma^2_0$ $H_1: \sigma^2\neq\sigma^2_0$
$R=\{X^2>x^2_{n-1,a}\}$	$R=\{X^2<x^2_{n-1,1-a}\}$	$R=\{\{X^2>x^2_{n-1,a/2}\}$ ή $\{X^2<x^2_{n-1,1-a/2}\}\}$

$$X^2 = \frac{(n-1)s^2}{\sigma_0^2}$$

Προϋπόθεση: δείγμα από κανονικό πληθυσμό

Έλεγχος υπόθεσης για το λόγο των διασπορών δύο πληθυσμών

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$$

$$H_0 : \frac{\sigma_1^2}{\sigma_2^2} > 1 \qquad H_0 : \frac{\sigma_1^2}{\sigma_2^2} < 1 \qquad H_0 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$

$$R = \{F > F_{n_1, n_2; a}\} \quad R = \{F > F_{n_1, n_2; a}\} \quad R = \{F > F_{n_1, n_2; a/2}\}$$

$$F = \begin{cases} \dfrac{s_1^2}{s_2^2}, \text{ όταν } s_1^2 > s_2^2 \\ \dfrac{s_1^2}{s_2^2}, \text{ όταν } s_1^2 < s_2^2 \end{cases}$$