

Τρίτη 26/02/2013

## 3ο ΜΑΘΗΜΑ "ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ"

- ⊙ Από τις μετρήσεις και αναλύσεις των βιοικείων ενός δείγματος κάποιου πληθυσμού, εξαχονται συμπεράσματα που αφορούν ολόκληρο τον πληθυσμό αυτών. Είναι ένα από τα κύρια θέματα της Δεξιότητας  
Επαγγελματικής Στατιστικής (1)  
(Inferential Statistics).
  - ⊙ Η Επαγγελματική Στατιστική ασχολείται με
    - Εκτίμηση των παραμέτρων του πληθυσμού
    - Έλεγχος Στατιστικών Υποθέσεων
  - ⊙ Η εκτίμηση των παραμέτρων του πληθυσμού μπορεί να είναι
    - Σημειακή
    - Εκτίμηση σε διάστημα (διάστημα εμπιστοσύνης)
  - ⊙ Η εκτίμηση σε διάστημα γίνεται αφού προηγουμένως οριστεί μια μικρή πιθανότητα  $\alpha$
- (1) Αλεξάνδρου Αν. κ.ά. "ΣΤΑΤΙΣΤΙΚΗ", ΣΥΓΧΡΟΝΗ ΕΚΔΟΣΤΙΚΗ, ΑΘΗΝΑ 1994, ISBN: 960-7344-33-2

η οποία είναι η πιθανότητα (λάθος) η παράμετρος  $\theta$  του πληθυσμού (που εκτιμούμε από το δείγμα) να βρίσκεται εκτός του διαστήματος  $(\theta_1, \theta_2)$  ενώ το οποίο εκτιμούμε ότι θα βρίσκεται η τιμή της παραμέτρου  $\theta$ .

Η πιθανότητα (μεγάλη)  $1-\alpha$  (επιτυχίας) που ορίζεται από την σχέση

$$P(\theta_1 < \theta < \theta_2) = 1-\alpha$$

λέγεται συντελεστής εμπιστοσύνης και το  $\alpha$  στάθμη σημαντικότητας (significance) και το διάστημα  $(\theta_1, \theta_2)$  λέγεται

$100(1-\alpha)\%$  διάστημα εμπιστοσύνης για την εκτιμώμενη παραμετρο  $\theta$

Στην πράξη συνήθως χρησιμοποιούμε ένα

95% ή 99% διάστημα εμπιστοσύνης

με  $\alpha=0,05$  ή  $0,01$  αντίστοιχα.

Τρίτη 25/02/2013

3ο ΜΑΘΗΜΑ "ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ"

⊙ ΕΚΤΙΜΗΤΕΣ (Η ΕΚΤΙΜΗΤΡΙΕΣ) ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ ΤΟΥ ΠΛΗΘΥΣΜΟΥ.

Λέγονται εκείνες οι συναρτήσεις των τιμών του δείγματος που οδηγούν στην καλύτερη εκτίμηση μιας παραμέτρου του πληθυσμού.

⊙ Ιδιότητες των εκτιμητών.

1. Αμερόληπτος (Unbiased).

- Εάν οριστεί ένα τυχαίο μέγεθος δείγματος  $n_1$
- Δημιουργηθούν ΟΛΑ τα δυνατά δείγματα μεγέθους  $n_1$
- $n_1 < n$  : μέγεθος πληθυσμού.
- Εάν η μέση τιμή της παραμέτρου  $\theta_i$ ,  $i=1, \dots, n_1$  που προκύπτει από ΟΛΑ τα δείγματα μεγέθους  $n_1$  ΣΥΜΠΤΗΤΕΙ με την μέση τιμή του πληθυσμού. (Μπορεί ΝΑΙ)

2. Αποτελεσματικός (efficient)

- Εάν μεταξύ όλων των ΑΜΕΡΟΛΗΠΤΩΝ εκτιμητών της παραμέτρου  $\theta$ , υπάρχει κάποιος του οποίου η κατανομή έχει την μικρότερη διασπορά από όλους.

### 3. Συνεπής (Consistent)

- Εάν αυξανόμενου του μεγέθους του δείγματος η τιμή του εκτιμητή τείνει προς την πραγματική τιμή της παραμέτρου του πληθυσμού που εκτιμά. (mean ναι, median όχι)

### 4. Επαρκής (adequate)

- Εάν για να δημιουργηθεί χρησιμοποιεί όλες τις μετρήσεις του δείγματος. (mean ναι, median όχι)

### 5. Στιβαρός (Robust)

- Εάν δεν επηρεάζεται από ακραίες τιμές (outliers). (mean όχι, median ναι).

Τρίτη 25/02/2013

### 3ο ΜΑΘΗΜΑ "ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ"

#### ⊙ ΕΛΕΓΧΟΣ ΣΤΑΤΙΣΤΙΚΗΣ ΥΠΟΘΕΣΗΣ

- Με κάποιο βκεπτικό θεωρούμε ότι μία παράμετρος  $\theta$  τού πληθυσμού έχει συγκεκριμένη τιμή  $\theta_0$
- Ο έλεγχος της υπόθεσης ότι  $\theta = \theta_0$  θα βασιστεί στις μετρήσεις του διαδεξιμου δείγματος

⊙ ΣΤΑΤΙΣΤΙΚΗ ΥΠΟΘΕΣΗ είναι μια παραδοχή για την κατανομή πιθανότητας <sup>(ή περισοτήτων)</sup> ενός πληθυσμού

Οι στατιστικές υποθέσεις γίνονται συνήδως για τις τιμές των παραμέτρων ενός πληθυσμού, όπως π.χ.

- ευνοϊκή πιθανότητα γεγονότος
- μέση τιμή
- αναδοχία επιτυχιών στο δείγμα
- διαφορές μέσων τιμών
- διαφορές αναδοχιών
- διασπορές
- λόγους διασπορών

Οι παράμετροι του πληθυσμού σε συνδυασμό με κάποιες παραμέτρους του δείγματος (που υπολογίζονται εύκολα) είναι μέρος κάποιων συναρτήσεων πιθανότητας που ακολουθούν γνωστή κατανομή.

Ανάλογα ποιά παράμετρο του πληθυσμού θέλουμε να ελέγξουμε, επιλέγουμε την γνωστή κατανομή πιθανότητας που ακολουθεί η συνάρτηση που περιέχει τις παραμέτρους του δείγματος και την παράμετρο του πληθυσμού. (π.χ.  $T = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \rightsquigarrow N(0,1)$ )

Για μια προκαθορισμένη <sup>επίπεδο</sup> (επίπεδο) σημαντικότητας που ΕΜΕΙΣ επιλέγουμε  $1-\alpha$ , μπορούμε να βρούμε τα άκρα του διαστήματος που αντιστοιχεί ε'αυτό το <sup>της παραπάνω ποσ. πιθαν.</sup> επίπεδο σημαντικότητας  $(1-\alpha)$  που εμείς ορίσαμε.

Τα άκρα του διαστήματος θα είναι τιμές της Τ.Μ. που ακολουθεί τη συνάρτηση κατανομής (που περιέχει την παράμετρο του πληθυσμού και παραμέτρους του δείγματος) π.χ.  $t_{\frac{\alpha}{2}}$ ,  $t_{\frac{\alpha}{2}}$ . Τέλος βρίσκουμε αν η προτεινόμενη τιμή περιέχεται στο διάστημα αυτό.

ΤΡΙΤΗ 25/02/2013

### 3ο ΜΑΘΗΜΑ "ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ"

#### ΒΗΜΑΤΑ ΕΦΑΡΜΟΓΗΣ ΕΝΟΣ ΣΤΑΤΙΣΤΙΚΟΥ ΕΛΕΓΧΟΥ

1. Ορισμός της μηδενικής  $H_0: \mu = \mu_0$  και της εναλλακτικής υποθέσεως  $H_1: \mu \neq \mu_0$  (διπλευρός έλεγχος) (εάν π.χ.  $\theta = \mu$ ) ή  $H_1: \mu > \mu_0$  ή μόνο  $\mu < \mu_0$  (μονόπλευρος)  
- Δηλ. Θέλουμε να δούμε εάν το δείγμα προέρχεται από τον πληθυσμό που εμείς ορίσαμε την παράμετρο.
2. Ορίζεται το <sup>συμπληρωματικό</sup> επίπεδο (ή η στάθμη) σημαντικότητας  $\alpha$  που είναι η πιθανότητα απόρριψης της  $H_0$  (συνήθως 5% ή 1%)
3. Προσδιορίζεται η κατανομή πιθανότητας (ανάλογα με τα στοιχεία του δείγματος) <sup>του στατιστικού π.χ.  $t, (\bar{x}, s, n, \mu_0)$</sup>  στην οποία θα οριστεί και η περιοχή αποδοχής (ή απόρριψης) της  $H_0$   
(Μπορεί να είναι η κανονική, ή η Student-t ή άλλη κατανομή)
4. Από τον τύπο της κατανομής πιθανότητας (π.χ.  $t_0 = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$ ) προσδιορίζεται η τιμή του στατιστικού το ποσό συνδέεται με την υποθετική τιμή της παραμέτρου  $\mu_0$  και τις παραμέτρους του δείγματος  $\bar{x}, s, n$
5. Από τους πίνακες των κατανομών βρίσκουμε ανάμεσα σε ποιές τιμές π.χ.  $t_1$  και  $t_2$  βρίσκεται η τιμή  $t_0$  και σε ποιές πιθανότητες αντιστοιχούν π.χ.  $p_1$  και  $p_2$

6. Εάν η πιθανότητα που δίνει το λογισμικό (π.χ. SP.  
περιέχεται ανάμεσα από τις δύο  $p_1$  και  $p_2$   
δεχόμαστε την  $H_0$



Τρίτη 26/02/2013

3ο ΜΑΘΗΜΑ "ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ"

- 3,00
- 3,10
- 3,00
- 2,50
- 3,20
- 2,20
- 3,20
- 2,60
- 3,50
- 3,30
- 4,20
- 3,10
- 3,20
- 2,80
- 3,70

**One-Sample Statistics**

	N	Mean	Std. Deviation	Std. Error Mean
VAR00001	15	3,1067	,48766	,12591

**One-Sample Test**

Test Value = 3.01

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
VAR00001	* ,768	14	,455	,09667	-,1734	,3667

**One-Sample Statistics**

	N	Mean	Std. Deviation	Std. Error Mean
VAR00001	15	3,1067	,48766	,12591

**One-Sample Test**

Test Value = 3.08

	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
VAR00001	,212	14	,835	,02667	-,2434	,2967

$\sqrt{15} = 3,8729833$   
 $\approx 3,873$   
 $\mu_0 = 3,01$   
 $\bar{y} = 3,1067$   
 $S = 0,48766$

\*  $t = \frac{\bar{y} - \mu_0}{S} \cdot \sqrt{n} = \frac{3,1067 - 3,01}{0,48766} \sqrt{15} = \frac{0,0967}{0,48766} \cdot 3,873 = 0,1983 \cdot 3,873 \approx 0,768$