ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΚΡΗΤΗΣ

Στατιστική II

Ενότητα 2: Δειγματοληψία

Γεώργιος Κ. Τσιώτας
Τμήμα Οικονομικών Επιστημών

Περιεχόμενα

Δειγματοληψία

Κατανομές Δειγματοληψίας

Κεντρικό Οριακό Θεώρημα

Τι είναι η Δειγματοληψία;

Στη Στατιστική εξάγουμε συμπεράσματα για τυχαία γεγονότα μέσω δειγματοληψίας. Έστω για τ.μ X λαμβάνουμε δείγμα $X_1, \ldots, X_n$.

Λόγοι χρήσης δειγματοληψίας

- Οικονομικά ασύμφωρη η χρήση όλου του πληθυσμού.
- Τεχνολογικά μη εφικτή.
- Ανάγκη άμεσης λήψης απόφασης.
- Ενδεχόμενη καταστροφή δείγματος μέσω της χρήσης του.

Είδη Δειγματοληψίας:

1. Απλή Τυχαία Δειγματοληψία: Κάθε ενδεχόμενο του πληθυσμού έχει ίση πιθανότητα επιλογής (παρ. κλήρωση αριθμών σε τυχαίο παίγνιο).
2. Συστηματική Τυχαία Δειγματοληψία: Η δειγματοληψία γίνεται κατά ομοιόμορφα (συστηματικά) διαστήματα (παρ. δειγματοληψία τουριστικών αφίξεων κατα τον μήνα Ιούνιο, μέτρηση ποιότητας ελαστικού κάθε $50km$)
3. Στρωματοποιημένη Τυχαία Δειγματοληψία: Η δειγματοληψία γίνεται σε ομοιόμορφα στρώματα του πληθυσμού (παρ. δειγματοληψία σε νέους από $18-28$ ετών).

Είδη Δειγματοληψίας:

1. Απλή Τυχαία Δειγματοληψία: Κάθε ενδεχόμενο του πληθυσμού έχει ίση πιθανότητα επιλογής (παρ. κλήρωση αριθμών σε τυχαίο παίγνιο).
2. Συστηματική Τυχαία Δειγματοληψία: Η δειγματοληψία γίνεται κατά ομοιόμορφα (συστηματικά) διαστήματα (παρ. δειγματοληψία τουριστικών αφίξεων κατα τον μήνα Ιούνιο, μέτρηση ποιότητας ελαστικού κάθε 50km)
3. Στρωματοποιημένη Τυχαία Δειγματοληψία: Η δειγματοληψία γίνεται σε ομοιόμορφα στρώματα του πληθυσμού (παρ. δειγματοληψία σε νέους από 18 — 28 ετών).

Είδη Δειγματοληψίας:

1. Απλή Τυχαία Δειγματοληψία: Κάθε ενδεχόμενο του πληθυσμού έχει ίση πιθανότητα επιλογής (παρ. κλήρωση αριθμών σε τυχαίο παίγνιο).
2. Συστηματική Τυχαία Δειγματοληψία: Η δειγματοληψία γίνεται κατά ομοιόμορφα (συστηματικά) διαστήματα (παρ. δειγματοληψία τουριστικών αφίξεων κατα τον μήνα Ιούνιο, μέτρηση ποιότητας ελαστικού κάθε $50km$)
3. Στρωματοποιημένη Τυχαία Δειγματοληψία: Η δειγματοληψία γίνεται σε ομοιόμορφα στρώματα του πληθυσμού (παρ. δειγματοληψία σε νέους από $18-28$ ετών).

Είδη Δειγματοληψίας:

1. Απλή Τυχαία Δειγματοληψία: Κάθε ενδεχόμενο του πληθυσμού έχει ίση πιθανότητα επιλογής (παρ. κλήρωση αριθμών σε τυχαίο παίγνιο).
2. Συστηματική Τυχαία Δειγματοληψία: Η δειγματοληψία γίνεται κατά ομοιόμορφα (συστηματικά) διαστήματα (παρ. δειγματοληψία τουριστικών αφίξεων κατα τον μήνα Ιούνιο, μέτρηση ποιότητας ελαστικού κάθε $50km$)
3. Στρωματοποιημένη Τυχαία Δειγματοληψία: Η δειγματοληψία γίνεται σε ομοιόμορφα στρώματα του πληθυσμού (παρ. δειγματοληψία σε νέους από $18-28$ ετών).

Μέσης Τιμής Δείγματος

Εαν ανεξάρτητες τ.μς $X_1, \ldots, X_n$ που προέρχονται από πληθυσμό με μέση τιμή μ και διακύμανση σ^2, τότε τότε η μέση τιμή της δειγματικής μέσης τιμής $\bar{X} = \frac{\sum_{i=1}^n X_i}{n}$ θα ορίζεται ως:

$$E(\bar{X}) = \frac{\sum_{i=1}^n E(x_i)}{n} = \frac{\mu + \cdots + \mu}{n} = \frac{n \cdot \mu}{n} = \mu.$$

Επίσης, η διακύμανση της δειγματικής μέσης τιμής $\bar{X}$ θα ορίζεται ως:

$$V(\bar{X}) = \frac{\sum_{i=1}^n V(X_i)}{n^2} = \frac{\sigma^2 + \cdots + \sigma^2}{n^2} = \frac{n \cdot \sigma^2}{n^2} = \frac{\sigma^2}{n}.$$

Κατανομή Μέσης Τιμής Δείγματος

Για $X_1, \ldots, X_n$ που προέρχονται από πληθυσμό με μέση τιμή μ και διακύμανση σ^2 και κατανέμονται Κανονικά (κατά *Gauss*), θα ισχύει ότι:

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1).$$

Η πιο πάνω ιδιότητα θα ισχύει ακόμη και όταν τα $X_1, \ldots, X_n$ κατανέμονται Κανονικά ασυμπτωτικά (δηλ. για $n \geq 30$).

Κατανομή Διαφοράς Μέσων Τιμών Δείγματος

Εαν ανεξάρτητες τ.μς $X_1, \ldots, X_{n_A}$ που προέρχονται από πληθυσμό A με μέση τιμή μ_A και διακύμανση σ_A^2, και δείγμα $X_1, \ldots, X_{n_B}$ προερχόμενο από πληθυσμό B με μέση τιμή μ_B και διακύμανση σ_B^2 τότε για Κανονικά κατανεμημένες τ.μς ή έχοντας $n_A, n_B \geq 30$ θα ισχύει:

$$\frac{\bar{X}_A - \bar{X}_B - (\mu_A - \mu_B)}{\sqrt{\frac{\sigma_A^2}{n_A} + \frac{\sigma_B^2}{n_B}}} \sim N(0, 1).$$

Για σ_A^2, σ_B^2 άγνωστες αλλά ίσες ($\sigma_A^2 = \sigma_B^2$)

$$\frac{\bar{X}_A - \bar{X}_B - (\mu_A - \mu_B)}{\sqrt{S^2\left(\frac{1}{n_A} + \frac{1}{n_B}\right)}} \sim t_{n_A+n_B-2},$$

κατανέμεται σύμφωνα με την $t - Student$ κατανομή με $n_A + n_B - 2$ βαθμούς ελευθερίας, όπου:

$$S^2 = \frac{S_A^2(n_A - 1) + S_B^2(n_B - 1)}{n_A + n_B - 2}.$$

Το Κεντρικό Οριακό Θεώρημα (Κ.Ο.Θ.)

Εαν έχουμε ανεξάρτητες τ.μς $X_1, \ldots, X_n$ που προέρχονται από πληθυσμό με μέση τιμή μ και διακύμανση σ^2 τότε θα ισχύει ότι:

$$\frac{\sum_{i=1}^{n} X_i - \sum_{i=1}^{n} \mu}{\sqrt{\sum_{i=1}^{n} \sigma^2}} \sim N(0, 1),$$

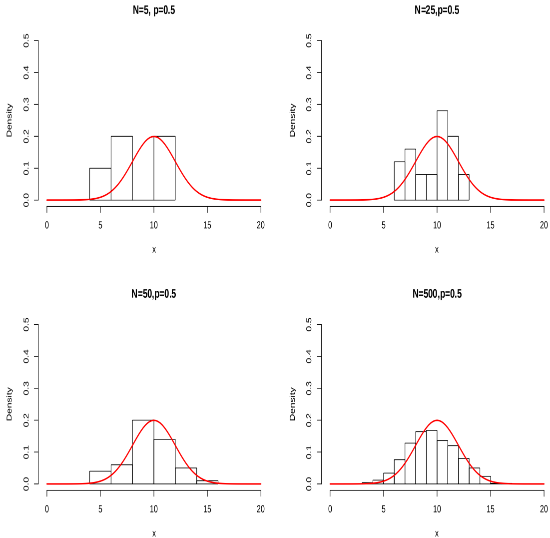
ασχέτως τρόπου κατανομής των $X_1, \ldots, X_n$ για $n \to \infty$ (στην πράξη θα ισχύει για $n \geq 30$).

Κ.Ο.Θ. στη διωνυμική κατανομή

Για ανεξάρτητες *Bernoulli* τ.μς $X_1, \ldots, X_n$ με p την αναλογία στο δείγμα, θα ισχύει

$$\frac{\sum_{i=1}^{n} X_i - n \cdot p}{\sqrt{n \cdot p \cdot (1-p)}} \sim N(0,1).$$

Κ.Ο.Θ. στη διωνυμική κατανομή (συν.)

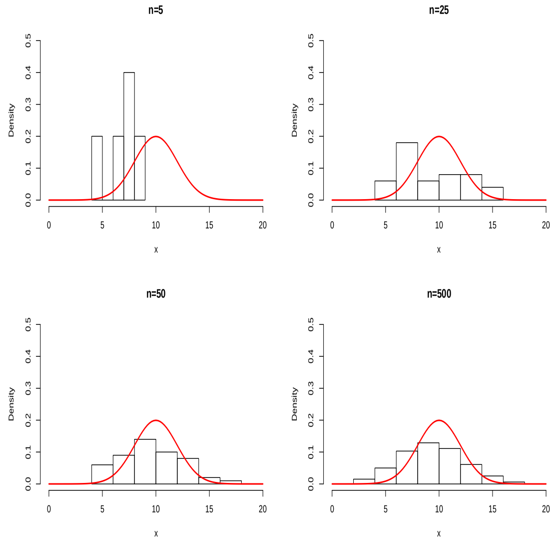


Κ.Ο.Θ. στη *Poisson* κατανομή

Για ανεξάρτητες *Poisson* τ.μ.ς $X_1, \ldots, X_n$, θα ισχύει

$$\frac{\sum_{i=1}^{n} X_i - n \cdot \lambda}{\sqrt{n \cdot \lambda}} \sim N(0, 1).$$

Κ.Ο.Θ. στη *Poisson* κατανομή (συν.)



Τέλος Ενότητας