Στατιστική ΙΙ

Γ. Ψαρράκος

Πανεπιστήμιο Πειραιώς

Οικονομικό Τμήμα

Β΄ Εξάμηνο

28 Φεβρουαρίου 2013

Ύλη

ΣΤΑΤΙΣΤΙΚΗ ΙΙ: Τυχαία δειγματοληψία. Διακριτές και Συνεχείς κατανομές. Χαρακτηριστικές παράμετροι κατανομών. Νόμος των μεγάλων αριθμών και κεντρικό οριακό θεώρημα. Κατανομές που απορρέουν από την κανονική. Σημειακοί εκτιμητές. Αμεροληψία. Συνέπεια. Επάρκεια. Μέθοδοι εκτίμησης. Διαστημική εκτίμηση πληθυσμιακών παραμέτρων. Διαστήματα εμπιστοσύνης μέσου, διασποράς, ποσοστού, διαφορών και αθροισμάτων. Έλεγχος υποθέσεων μέσου, διασποράς, ποσοστού, διαφορών και αθροισμάτων.

Ορισμός Στατιστικής

Στατιστική είναι η επιστήμη που ασχολείται με τον σχεδιασμό πειραμάτων, τη συλλογή και ανάλυση αριθμητικών δεδομένων και την εξαγωγή συμπερασμάτων. Η Στατιστική ορίζεται επίσης ως η επιστήμη της λήψης αποφάσεων όταν υπάρχει αβεβαιότητα.

Ο σχεδιασμός των πειραμάτων και η στατιστική ανάλυση γίνεται σε ένα τυχαίο δείγμα και η εξαγωγή συμπερασμάτων - λήψη αποφάσεων αφορά τον πληθυσμό.

Το τυχαίο δείγμα μεγέθους n (από ένα πληθυσμό) συμβολίζεται με $X_1, X_2, \ldots, X_n$ όπου X_i, $i - 1, 2, \ldots, n$ είναι ανεξάρτητες και ισόνομες τ.μ.

Τυχαίες Μεταβλητές

Η τυχαία μεταβλητή (τ.μ.) X είναι μια συνάρτηση

$$X : \Omega \to R \ \text{ με } \ X(\omega) = x,$$

όπου Ω ο δειγματικός χώρος και R το σύνολο των πραγματικών αριθμών (μπορεί να είναι και ένα υποσυνολό του).
Κάθε τ.μ. έχει μια κατανομή πιθανότητας.

Χαρακτηριστικά τ.μ. και κατανομών

Η μέση τιμή της X είναι

$$\mu = E(X) = \begin{cases} \sum_x x P(X = x), & \text{αν } X \text{ διακριτή τ.μ.,} \\ \int_{-\infty}^{+\infty} x\, f(x)\, dx, & \text{αν } X \text{ συνεχής τ.μ.} \end{cases}$$

- $P(X = x)$: συνάρτηση πιθανοτητας (σπ)
- $f(x) = P(x \leq X < x + dx)$: η συνάρτηση πυκνότητας πιθανότητας (σππ)

Γενικά η μέση τιμή της τ.μ. $h(X)$, όπου h μια συνάρτηση είναι

$$E(h(X)) = \begin{cases} \sum_x h(x)\, P(X = x), & \text{αν } X \text{ διακριτή τ.μ.,} \\ \int_{-\infty}^{+\infty} h(x)\, f(x)\, dx, & \text{αν } X \text{ συνεχής τ.μ.} \end{cases}$$

Η διακύμανση ή διασπορά της τ.μ. X είναι

$$\sigma^2 = Var(X) = E[(X-\mu)^2] = E(X^2) - \mu^2, \quad \text{όπου } \mu = E(X)$$

και η τυπική απόκλιση $\sigma = \sqrt{Var(X)}$.

Ο συντελεστής λοξότητας είναι

$$\frac{E(X-\mu)^3}{\sigma^3}$$

και ο συντελεστής κύρτωσης είναι

$$\frac{E(X-\mu)^4}{\sigma^4}.$$

Δύο βασικές ιδιότητες της μέσης τιμής και της διασποράς

Αν a, b σταθεροί αριθμοί, τότε ισχύουν

$$E(aX + b) = aE(x) + b$$

και

$$Var(aX + b) = a^2\, Var(X).$$

Από την τελευταία σχέση προκύπτει άμεσα ότι

$$\sigma_{aX+b} = \sqrt{Var(aX + b)} = |a| \sqrt{Var(X)} = |a|\, \sigma_X.$$

Διακριτές κατανομές

Έστω ότι η τ.μ. X είναι διακριτή με τιμές $\{x_1, x_2, x_3, \ldots\}$ και σπ

$$P(X = x), \quad x = x_1, x_2, x_3, \ldots$$

Η συνάρτηση $P(X = x)$ ικανοποιεί τις ιδιότητες:

- $P(X = x) \geq 0$.
- $\sum_x P(X = x) = P(X = x_1) + P(X = x_2) + \ldots = 1$.

Ακόμα κάποιες επιπλέον ιδιότητες είναι

$$P(X \leq x_3) = P(X = x_1) + P(X = x_2) + P(X = x_3),$$

$$\begin{aligned} P(X > x_2) &= P(X = x_3) + P(X = x_4) + P(X = x_5) + \ldots \\ &= 1 - P(X = x_1) - P(X = x_2) = 1 - P(X \leq x_2), \end{aligned}$$

$$P(x_2 \leq X \leq x_5) = P(X = x_2) + P(X = x_3) + P(X = x_4) + P(X = x_5)$$

και

$$P(x_2 \leq X < x_5) = P(X = x_2) + P(X = x_3) + P(X = x_4).$$

Οι κυριότερες διακριτές κατανομές είναι η Διωνυμική, η Υπεργεωμετρική, η Γεωμετρική και η κατανομή Poisson.

Παραδείγματα

- Έστω X: το πλήθος των περιπτώσεων κατά τις οποίες θα έρθει 'γράμματα' από την ταυτόχρονη ρίψη τριών νομισμάτων. Να βρεθεί η κατανομή πιθανότητας της τ.μ. X.
- Ρίχνουμε δύο ζάρια. Αν X: το άθροισμα των αριθμών των δύο πλευρών, να κατασκευάσετε τη συνάρτηση πιθανότητας.

Διωνυμική κατανομή

Ένα πείραμα τύχης λέγετε διωνυμικό όταν τα μόνα δυνατά αποτελέσματα είναι δύο, επιτυχία (E) ή αποτυχία (A) και ισχύει ότι:

- Η πιθανότητα επιτυχίας (p) ή αποτυχίας $(q = 1 - p)$ παραμένη σταθερή σε όλες τις δοκιμές του πειράματος.
- Οι δοκιμές είναι στατιστικώς ανεξαρτητες, δηλ. το αποτέλεσμα μιας ορισμένης δοκιμής δεν επηρεάζει το αποτέλεσμα οποιασδήποτε άλλης δοκιμής.

Η σπ της διωνυμικής κατανομής, συμβ. $X \sim B(n, p)$, είναι

$$P(X = x) = \binom{n}{x} p^x \, q^{n-x}, \quad x = 0, 1, 2, \ldots, n,$$

όπου

$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \quad \text{συνδιασμός των } n \text{ ανά } x.$$

Η διωνυμική τ.μ. X: ο αριθμός των επιτυχιών σε n δοκιμές, μπορεί να εκφραστεί ως

$$X = I_1 + I_2 + \ldots + I_n,$$

όπου για $i = 1, 2, \ldots, n$ είναι

$$I_i = \begin{cases} 1, & \text{το αποτέλεσμα της } i \text{ δοκιμής είναι επιτυχία,} \\ 0, & \text{το αποτέλεσμα της } i \text{ δοκιμής είναι αποτυχία.} \end{cases}$$

Τονίζουμε ότι οι τ.μ. $I_1, I_2, \ldots I_n$ είναι μεταξύ τους ανεξάρτητες και ισόνομες.

Μέση τιμή και διασπορά της διωνυμικής

Υπολογίζουμε πρώτα τη μέση τιμή και τη διασπορά της I_i, $i = 1, 2, \ldots, n$. Είναι

$$E(I_i) = \sum_{x \in \{0,1\}} x\, P(I_i = x) = 0\, P(I_i = 0) + 1\, P(I_i = 1) = 0\, q + 1\, p = p,$$

$$E(I_i^2) = \sum_{x \in \{0,1\}} x^2\, P(I_i = x) = 0^2\, P(I_i = 0) + 1^2\, P(I_i = 1) = p$$

$$\sigma_{I_i}^2 = Var(I_i) = [E(I_i^2)] - [E(I_i)]^2 = p - p^2 = p(1-p) = pq.$$

Χρησιμοποιώντας ότι $X = I_1 + I_2 + \ldots + I_n$ με τα I_i ανεξάρτητα και ισόνομα μεταξύ τους, προκύπτει ότι

$$E(X) = n\, E(I_i) = n\, p$$

και

$$\sigma_X^2 = Var(X) = n\, Var(I_i) = np(1-p) = npq.$$

Υπεργεωμετρική κατανομή

Έστω ότι ο πληθυσμός αποτελείται από N αντικείμενα m από τα οποία έχουν την ιδιότητα E (επιτυχία) και $N-m$ (αποτυχία).

Περιγραφή πειράματος: Έστω ότι έχουμε ένα δοχείο με N (ο πληθυσμός) σφαιρίδια, από τα οποία τα m είναι λευκά (επιτυχία) και $N-m$ είναι μαύρα (αποτυχία). Έστω ότι εξάγουμε n από τα N σφαιρίδια **χωρίς επανάθεση** και

$$X: \quad \text{ο αριθμός των λευκών } (E) \text{ σφαιριδίων.}$$

Τότε η X ακολουθεί την Υπεργεωμετρική κατανομή, συμβ. $X \sim Hg(N, n, m)$ με σπ

$$P(X=x) = \frac{\binom{N-m}{n-x}\binom{m}{x}}{\binom{N}{n}}, \quad x = 0, 1, 2, \ldots, \min\{m, n\}.$$

Η μέση τιμή και η διασπορά είναι

$$\mu = n\frac{m}{N} = np \quad \text{και} \quad \sigma^2 = n\frac{m}{N}\frac{N-m}{N}\frac{N-n}{N-1} = npq\frac{N-n}{N-1},$$

όπου $p = m/N$ (θυμίζουμε $q = 1-p$).

Διωνυμική προσέγγιση της υπεργεωμετρικής. Όταν το N είναι μεγάλο και το n μικρό η $Hg(N, n, m)$ προσεγγίζεται από την $B(n, p = m/N)$

$$\frac{\binom{N-m}{n-x}\binom{m}{x}}{\binom{N}{n}} \approx \binom{n}{x} p^x (1-p)^{n-x}.$$

Γεωμετρική κατανομή

Σε ένα διωνυμικό πείραμα έστω η τ.μ.

X : ο αριθμός των δοκιμών μέχρι την πρώτη επιτυχία.

Τότε η X ακολουθεί τη γεωμετρική κατανομή, συμβ.
$X \sim Geo(p)$ με σπ

$$P(X = x) = p\, q^{x-1}, \quad x = 1, 2, 3, \ldots$$

Η μέση τιμή και η διασπορά είναι

$$\mu = \frac{1}{p} \quad \text{και} \quad \sigma^2 = \frac{q}{p^2}.$$

Κατανομή Poisson

Η κατανομή Poisson εφαρμόζεται όταν έχουμε μια σειρά τυχαίων ενδεχομένων τα οποία γίνονται στη μονάδα του χρόνου π.χ. το πλήθος αφίξεων σε ένα αεροδρόμιο, το πλήθος των τηλεφωνημάτων που φθάνουν σε ένα κεντρικό σταθμό, το πλήθος των αποζημιώσεων που φθάνουν σε μια ασφαλιστική εταιρεία κτλ.
Η σπ της Poisson με παράμετρο $\lambda > 0$, συμβ.$X \sim Po(\lambda)$, είναι

$$P(X = x) = \frac{e^{-\lambda}\,\lambda^x}{x!}, \quad x = 0, 1, 2, \ldots$$

Η μέση τιμή και η διασπορά είναι

$$\mu = \lambda \ \text{ και } \ \sigma^2 = \lambda.$$

- $B(n, p) \approx Po(\lambda = np)$ για μικρά p και μεγάλα n.

Στοχαστική διαδικασία Poisson

Η κατανομή Poisson έχει συχνή εφαρμογή σε διαδικασίες που εξελίσσονται με το χρόνο, δηλ. έχουμε $X \sim Po(\lambda\, t)$ (αντί $X \sim Po(\lambda)$), όπου $t > 0$ ο χρόνος.
Έστω X_t ο αριθμός των ενδεχομένων (αφίξεων ή εκπομπών) μέχρι τη χρονική στιγμή $t > 0$. Η $X(t) \sim Po(\lambda t)$, με $\lambda > 0$ όταν

$$P(X(t) = x) = \frac{e^{-\lambda\, t}\,(\lambda\, t)^x}{x!}, \quad x = 0, 1, 2, \ldots$$

- P(να συμβεί ένα ενδεχόμενο σε μικρό χρονικό διάστημα)=ανάλογη του μήκους του διαστήματος.
- P(να συμβούν δύο ή περισσότερα ενδεχόμενα σε μικρό χρονικό διάστημα) = αμελητέα.

Συνεχείς κατανομές

Μέχρι τώρα ασχοληθήκαμε με διακριτές τ.μ., θεωρήσαμε δηλαδή ότι η τ.μ. X λαμβάνει τιμές σε διακριτό σύνολο, πεπερασμένο (βλέπε Διωνυμική, Υπεργεωμετρική) ή άπειρο (βλέπε Γεωμετρική, Poisson). Στη συνέχεια θα μελετήσουμε συνεχείς τ.μ. οι οποίες ορίζονται στο διάστημα $(-\infty, \infty)$ (ή σε υποσυνολό του), με σππ $f(x) = P(x \leq X < x + dx)$ και

$$P(a \leq X \leq b) = \int_a^b f(x)\, dx \geq 0,$$

δηλαδή $P(a \leq X \leq b)$ είναι το εμβαδόν του χωρίου που σχηματίζεται από τη γραφική παράσταση της $f(x)$ στο διάστημα $[a, b]$.

- $$P(-\infty < X < \infty) = \int_{-\infty}^{\infty} f(x)\,dx = 1$$
- Λόγω συνέχειας ισχύει ότι

$$P(X = a) = P(a \leq X \leq a) = \int_{a}^{a} f(x)\,dx = 0.$$

Επομένως,

$$P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b).$$

- Αθροιστική συνάρτηση κατανομής (ασκ), $F(x) = P(X \leq x)$

$$F(x) = P(X \leq x) = \int_{-\infty}^{x} f(t)\,dt.$$

Ισχύει ότι $F'(x) = f(x)$, $F(x)$ αύξουσα ως προς x,
$F(-\infty) = 0$ και $F(\infty) = 1$.

Ομοιόμορφη κατανομή

Αν $X \sim U(a, b)$, τότε η σππ δίνεται από

$$f(x) = \begin{cases} \frac{1}{b-a}, & \text{αν } x \in [a, b], \\ 0, & \text{αν } x \notin [a, b] \end{cases}$$

Η ασκ είναι

$$F(x) = \int_a^x f(t)\, dt = \begin{cases} 0, & \text{αν } x < a, \\ \frac{x-a}{b-a}, & \text{αν } x \in [a, b] \\ 1, & \text{αν } x > b \end{cases}$$

Η μέση τιμή, η δεύτερη ροπή και η διασπορά είναι, αντίστοιχα

$$\mu = E(X) = \int_a^b x\, f(x)\, dx = \frac{b+a}{2},$$

$$E(X^2) = \int_a^b x^2\, f(x)\, dx = \frac{b^2+ab+a^2}{3}$$

και

$$\sigma^2 = Var(X) = E(X^2) - [E(X)]^2 = \frac{(b-a)^2}{12}.$$

Εκθετική κατανομή

Αν $X \sim Exp(\lambda)$, με $\lambda > 0$, τότε η σππ δίνεται από

$$f(x) = \lambda\, e^{-\lambda x}, \quad x \geq 0.$$

Η ασκ είναι

$$F(x) = \int_0^x f(t)\, dt = 1 - e^{-\lambda x}, \quad x \geq 0.$$

Η μέση τιμή, και η διασπορά είναι

$$\mu = E(X) = \int_0^\infty x\, f(x)\, dx = \frac{1}{\lambda}$$

και

$$\sigma^2 = Var(X) = E(X^2) - [E(X)]^2 = \frac{1}{\lambda^2}.$$

Κανονική κατανομή

Η σππ της κανονικής κατανομής με μέση τιμή $-\infty < \mu < \infty$ και διασπορά $0 < \sigma^2 < \infty$, είναι ($\pi \approx 3.14$ και $e \approx 2.71$)

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}, \quad -\infty < x < \infty.$$

Μερικές φορές γράφουμε $f(x; \mu, \sigma)$, θεωρώντας ότι το x είναι η μεταβλητή, ενώ τα μ και σ παράμετροι. Στην ειδική περίπτωση όπου $\mu = 0$ και $\sigma = 1$ λαμβάνουμε την τυποποιήμενη κανονική με σππ

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, \quad -\infty < x < \infty.$$

$$X \sim N(\mu, \sigma^2) \quad \Leftrightarrow \quad Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

Η ασκ της $X \sim N(\mu, \sigma^2)$ είναι

$$F(x) = P(X \leq x) = \frac{1}{\sigma\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{(t-\mu)^2}{2\sigma^2}}\, dt.$$

Η ασκ της $Z \sim N(0,1)$ είναι

$$\Phi(z) = P(Z \leq z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-\frac{t^2}{2}}\, dt.$$

Για την ασκ της τυποποιημένη κανονική ισχύουν οι ακόλουθες σχέσεις

$$\Phi(-z) = 1 - \Phi(z), \quad -\infty < z < \infty$$

και

$$\Phi(z) = \frac{1}{2} + \frac{1}{\sqrt{2\pi}} \int_{0}^{z} e^{-\frac{t^2}{2}}\, dt, \quad x \geq 0.$$

Η κανονική κατανομή (γνωστή και ως Γκαουσιανή κατανομή) χρησιμοποιείται ως προσέγγιση για να περιγραφούν τ.μ., οι οποίες τείνουν να συγκεντρώνονται γύρω από μια μέση τιμή. Η κανονική κατανομή είναι η πιο σημαντική κατανομή της στατιστικής μεθοδολογίας για τους εξής βασικούς λόγους:

- Την κανονική κατανομή ακολουθούν με μεγάλη προσέγγιση τα περισσότερα συνεχή φαινόμενα, π.χ. πολλά πληθυσμιακά χαρακτηριστικά, όπως το ύψος, το βάρος η βαθμολογία σε διαγώνισμα, κ.λ.π.
- Σύμφωνα με το κεντρικό οριακό θεώρημα (το άθροισμα ενός ικανοποιητικά μεγάλου αριθμού ανεξάρτητων και ισόνομων τυχαίων μεταβλητών προσεγγίζεται από την κανονική κατανομή) αποτελεί τη βάση της στατιστικής συμπερασματολογίας.
- Τυχαία σφάλματα που εμφανίζονται σε διάφορες μετρήσεις έχουν κανονική κατανομή. Γι΄ αυτό το λόγο η κανονική κατανομή αναφέρεται πολλές φορές και ως κατανομή σφαλμάτων.

Κεντρικό Οριακό Θεώρημα (ΚΟΘ)

Έστω $X_1, X_2, \ldots, X_n$ ανεξάρτητες μεταξύ τ.μ. που ακολουθούν την ίδια κατανομή (δηλ. είναι ισόνομες) με

$$E(X_i) = \mu \text{ και } Var(X_i) = \sigma^2, \text{ για} i = 1, 2, \ldots, n.$$

Αποδεικνύεται ότι αν

$$X_1, X_2, \ldots, X_n \sim N(\mu, \sigma^2) \text{ τότε } \sum_{i=1}^{n} X_i = X_1 + X_2 + \ldots + X_n \sim N(n\mu, n\sigma^2).$$

Αυτό σημαίνει ότι

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

Στην περίπτωση όπου οι $X_1, X_2, \ldots, X_n$ δεν ακολουθούν την κανονική κατανομή ισχύει το αποτέλεσμα της τελευταίας σχέσης για μεγάλα n ;

Θεώρημα

Αν οι τμ $X_1, X_2, \ldots, X_n$ *είναι ανεξάρτητες και ισόνομες με* $E(X_i) = \mu$ *και* $Var(X_i) = \sigma^2$ *(πεπερασμένα),* $i = 1, 2, \ldots, n$*, τότε η τμ*

$$Z = \frac{\sum_{i=1}^{n} X_i - n\mu}{\sqrt{n}\,\sigma} \tag{1}$$

ακολουθεί ασυμπτωτικά (για μεγάλο n*, συνήθως* $>$ *του 30) την* $N(0, 1)$.

Από την (1) βλέπουμε ότι

$$\sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2) \tag{2}$$

και

$$\text{Δειγματικός μέσος} = \overline{X} = \frac{\sum_{i=1}^{n} X_i}{n} \sim N(\mu, \frac{\sigma^2}{n}). \tag{3}$$

Κανονικής Προσέγγιση της Διωνυμικής Κατανομής

Θυμίζουμε ότι η διωνυμική τ.μ. $X \sim B(n, p)$: ο αριθμός των επιτυχιών σε n δοκιμές, μπορεί να εκφραστεί ως

$$X = I_1 + I_2 + \ldots + I_n,$$

όπου I_i ανεξάρτητες και ισόνομες δίτιμες Bernoulli, για $i = 1, 2, \ldots, n$,

$$I_i = \begin{cases} 1, & \text{το αποτέλεσμα της } i \text{ δοκιμής είναι επιτυχία,} \\ 0, & \text{το αποτέλεσμα της } i \text{ δοκιμής είναι αποτυχία.} \end{cases}$$

Εφαρμόζοντας το ΚΟΘ για το άθροισμα $X = I_1 + I_2 + \ldots + I_n$, με $\mu = E(I_i) = p$ και $Var(I_i) = pq$, όπου $q = 1 - p$, είναι

$$X \sim N(nE(I_i), nVar(I_i)), \text{ δηλαδή } X \sim N(np, npq).$$

Συμβολικά είναι $B(n, p) \approx N(np, npq)$ και η προσέγγιση αυτή είναι ικανοποιητική όταν npq μεγάλο (> 10 ή καλύτερα > 30).

Νόμος των μεγάλων αριθμών

Θεώρημα

Έστω $X_1, X_2, \ldots$ μια ακολουθία ανεξάρτητων και ισόνομων τ.μ. με $E(X_i) = \mu$. Τότε,

$$\frac{X_1 + X_2 + \ldots + X_n}{n} \rightarrow \mu, \quad \text{όταν } n \rightarrow \infty.$$

Αν εκτελέσουμε ένα (τυχαίο) πείραμα n φορές και X_i είναι η τ.μ. που εκφράζει το αποτέλεσμα του i - πειράματος, τότε σύμφωνα με το νόμο των μεγάλων αριθμών ο μέσος όρος των X_i συγκλίνει στη μέση τιμή $E(X_i) = \mu$ για $n \rightarrow \infty$.

Αντίστροφα εκατοστιαία σημεία της τυπικής κανονικής

Έστω $Z \sim N(0,1)$. Τα αντίστροφα εκατοστιαία σημεία της Z, συμβ. z_a, ορίζονται ως

$$P(Z > z_a) = a.$$

Ισοδύναμα έχουμε

$$\Phi(z_a) = 1 - a,$$

Για $a = 5\%$, $a = 2.5\%$ και $a = 0.5\%$ έχουμε αντίστοιχα

$$z_{0.05} = 1.645, \quad z_{0.025} = 1.96, \quad z_{0.005} = 2.58.$$

Ισχύει ότι $z_{1-a} = -z_a$.

Κατανομές που απορρέουν από την κανονική

Έστω $X_1, X_2, \ldots X_n$ ένα τυχαίο δείγμα (ανεξάρτητες και ισόνομες) από ένα κανονικό πληθυσμό με μέση τιμή μ και διασπορά σ^2. Τότε γνωρίζουμε ότι το άθροισμα $\sum_{i=1}^{n} X_i$ ακολουθεί πάλι την κανονική κατανομή με μέση τιμή $n\mu$ και διασπορά $n\sigma^2$.

$$X_1, X_2, \ldots X_n \sim N(\mu, \sigma^2) \Rightarrow \sum_{i=1}^{n} X_i \sim N(n\mu, n\sigma^2).$$

Όμοια ισχύει για την τυπική κανονική κατανομή ότι

$$Z_1, Z_2, \ldots Z_n \sim N(0, 1) \Rightarrow \sum_{i=1}^{n} Z_i \sim N(0, n).$$

Υπάρχουν άλλες κατανομές που απορρέουν από την κανονική;

Κατανομή χι-τετράγωνο με n βαθμούς ελευθερίας (β.ε.)

Έστω $Z_1, Z_2, \ldots, Z_n$ ανεξάρτητες και ισόνομες τ.μ. Τότε

$$Z_1, Z_2, \ldots, Z_n \sim N(0,1) \Rightarrow Y = \sum_{i=1}^{n} Z_i^2 \sim \chi^2_n. \quad (\chi^2_n = \sum_{1}^{n} N(0,1)^2.)$$

Η Y είναι συνεχής τ.μ. ορισμένη στο $[0, \infty)$ με σππ

$$f_{\chi^2_n}(x) = \frac{x^{\frac{n}{2}-1}\, e^{-\frac{x}{2}}}{2^{\frac{n}{2}}\, \Gamma(\frac{n}{2})}, \quad x > 0.$$

Η τ.μ. Y αποτελεί μία ειδική περίπτωση της κατανομής $\Gamma(a, b)$ με σππ (για $a = n/2$ και $b = 1/2$)

$$f_{\Gamma(a,b)}(x) = \frac{b^a\, x^{a-1}\, e^{-bx}}{\Gamma(a)}, \quad x > 0, \quad a, b > 0,$$

όπου $\Gamma(a) = \int_0^\infty x^{a-1}\, e^{-x}\, dx$ η συνάρτηση Γ.

Η χ^2_n κατανομή έχει λοξότητα από δεξιά και η πυκνοτητά της εξαρτάται μόνο από τους από τους βαθμούς ελευθερίας n. Όσο το n μεγαλώνει η λόξότητα ελαττώνεται και για μεγάλα n προσεγγίζεται από την κανονική. Ακόμα

$$E(\chi^2_n) = n \ \text{ (η μέση τιμή της } \Gamma(a,b) \text{ είναι } a/b),$$

$$Var(\chi^2_n) = 2n \ \text{ (η διασπορά της } \Gamma(a,b) \text{ είναι } a/b^2)$$

Οι συνήθεις στατιστικοί πίνακες της χ^2_n ορίζονται με τη σχέση

$$P(\chi^2_n > \chi^2_{a,n}) = a.$$

Για παράδειγμα $\chi^2_{0.01,5} = 15.086$, $\chi^2_{0.05,20} = 31.41$. Στο πίνακα η μεγαλύτερη τιμή του n είναι 30. Για τιμές $n > 30$, προσεγγίζουμε τις τιμές της χ^2_n χρησιμοποιώντας το ΚΟΘ.

Κατανομή τι με n βαθμούς ελευθερίας

$$Z \sim N(0,1), Y \sim \chi^2_n \text{ (ανεξ. μεταξύ τους) } \Rightarrow W = \frac{Z}{\sqrt{Y/n}} \sim t_n.$$

Συμβολικά η κατανομή t_n, γνωστή ως κατανομή Student, είναι

$$t_n = \frac{N(0,1)}{\sqrt{\chi^2_n/n}}.$$

Η W είναι συνεχής τ.μ. ορισμένη στο $(-\infty, \infty)$ με σππ

$$f_{t_n}(t) = \frac{\Gamma(\frac{n+1}{2})}{\Gamma(\frac{n}{2})\sqrt{\pi n}} \left(1 + \frac{t^2}{n}\right)^{-\frac{n+1}{2}}, \quad -\infty < t < \infty.$$

Ακόμα είναι συμμετρική περί τον άξονα $x = 0$ (όπως η $N(0,1)$) με

$$E(t_n) = 0 \text{ και } Var(t_n) = \frac{n}{n-2} > 1, \quad n > 2.$$

Επιπλέον, $t_\infty \equiv N(0,1)$.

Οι συνήθεις στατιστικοί πίνακες της t_n ορίζονται με τη σχέση

$$P(t_n > t_{a,n}) = a.$$

Λόγω συμμετρίας έχουμε

$$P(t_n \leq -t_{a,n}) = P(t_n > t_{a,n}) = a$$

$$t_{1-a,n} = -t_{a,n}$$

Για παράδειγμα $t_{0.01,15} = 2.602,\ t_{0.05,\infty} = 1.645.$

Κατανομή F με n_1, n_2 βαθμούς ελευθερίας

$$X_1 \sim \chi^2_{n_1}, X_2 \sim \chi^2_{n_2} \text{ (ανεξ. μεταξύ τους) } \Rightarrow \frac{\chi^2_{n_1}/n_1}{X^2_{n_2}/n_2} \sim F_{n_1,n_2}.$$

Συμβολικά

$$F_{n_1,n_2} = \frac{\chi^2_{n_1}/n_1}{\chi^2_{n_2}/n_2}.$$

Είναι συνεχής τ.μ. ορισμένη στο $[0, \infty)$ με

$$E(F_{n_1,n_2}) = \frac{n_2}{n_2 - 2}, \quad n_2 > 2$$

$$Var(F_{n_1,n_2}) = \frac{2n_2^2(n_1 + n_2 - 2)}{n_1(n_2 - 2)^2(n_2 - 4)}, \quad n_2 > 4.$$

Είναι ασύμμετρη και $t_n^2 = F_{1,n}$.

Τα αντίστροφα εκατοστιαία σημεία της F είναι

$$P(F_{n_1,n_2} > F_{a,n_1,n_2}) = a$$

Τ. Παπαιωάννου και Λούκας Σ.Β. Εισαγωγή στη Στατιστική, 2002.

Χ.Κ. Αγιακλόγλου και Θ.Ε. Μπένος. Εισαγωγή στην Οικονομετρική Ανάλυση, 2003.

Γ. Ηλιόπουλος. Βασικές Μέθοδοι Εκτίμησης Παραμέτρων, 2006.

Χ.Κ. Φράγκος. Στατιστική Επιχειρήσεων, 1998.

Δ.Α. Ιωαννίδης. Στατιστικές Μέθοδοι, 1999.

Μ. Μπούτσικας. Στατιστική ΙΙΙ, Σημειώσεις 2003.