3ο online μάθημα ①

Εισαγωγή στην Στατιστική

2/4/2020

Άσκηση 1η: Έστω $X_1, X_2, \ldots, X_9$ τ.δ. από πληθυσμό με κανονική κατανομή $N(18, 4)$. Υπολογίστε την $P(\bar{x} < 20)$

Λύση: Γνωρίζουμε ότι αν $X_1, \ldots, X_9$ τ.δ. από $N(18,4)$ τότε

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

Πώς μπορώ να το θυμάμαι;

$\bar{X}$ είναι γραμμικός συνδυασμός

Γραμμικός συνδυασμός κανονικών $\rightsquigarrow$ κανονική

Επίσης εύκολα $E\bar{X} = E\left(\frac{1}{n}\sum X_i\right) = \mu$

$Var\,\bar{X} = \sigma^2/n$ $\quad$ $Var\,\bar{X} = Var\left(\frac{1}{n}\sum X_i\right) = \frac{1}{n}\sum Var\,X_i$

Είναι λοιπόν: $\quad$ $X_i \sim N(18,4)$

$$\bar{X} \sim N\left(\mu = 18, \frac{\sigma^2}{n} = \frac{4}{9}\right)$$

και (τυπικός μετασχηματισμός)

$$Z = \frac{\bar{X} - 18}{\sqrt{\frac{4}{9}}} = \frac{(\bar{X} - 18)\cdot 3}{2} \sim N(0,1)$$

Ζητώ $\quad P(\bar{x} < 20) = P\left(\frac{\bar{x}-18}{2}\cdot 3 < \frac{20-18}{2}\cdot 3\right)$

$\Downarrow$

$$P(\bar{x} < 20) = P\left(Z \le \frac{2}{2/3}\right) = P(Z < 3) \overset{\text{πίνακες}}{\underset{N(0,1)}{=\!=\!=}} 0.9987$$

Άσκηση 2η: Έστω $X_1, X_2, \dots, X_n$, $Y_1, \dots, Y_m$ δύο ανεξάρτητα τυχαία δείγματα από κανονικούς πληθυσμούς με ίσες διακυμάνσεις. Να βρεθεί η κατανομή της τ.μ. $\frac{S_1^2}{S_2^2}$

Λύση: $X_1, \dots, X_n$ τ.δ. $N(\mu_1, \sigma^2)$ ίσες διακυμάνσεις

$Y_1, \dots, Y_m$ τ.δ. $N(\mu_2, \sigma^2)$

Γνωρίζω ότι:

$$\left.\begin{array}{l} \frac{(n-1)S_1^2}{\sigma^2} \sim \chi^2_{n-1} \\ \\ \frac{(m-1)S_2^2}{\sigma^2} \sim \chi^2_{m-1} \end{array}\right\} \text{ανεξάρτητες}$$

Τότε από ορισμό της F κατανομή

X, Y ανεξ.

$X \sim \chi^2_{n_1}$ $\quad F = \frac{X/n_1}{Y/n_2} \sim F_{n_1, n_2}$

$Y \sim \chi^2_{n_2}$

$$\frac{\frac{(n-1)S_1^2}{\sigma^2} \Big/ (n-1)}{\frac{(m-1)S_2^2}{\sigma^2} \Big/ (m-1)} \sim F_{n-1, m-1}$$

ή $\frac{S_1^2}{S_2^2} \sim F_{n-1, m-1}$

Άσκηση 3η: Αν $X_1, X_2, \dots, X_n$ τ.δ. ($N(0,1)$) από την τυπική κανονική κατανομή να βρεθεί η κατανομή της $U = n\bar{X}^2$

Λύση: $X_1, X_2, \dots, X_n$ τ.δ. $N(0,1)$ $\quad \mu = 0$, $\sigma = 1$

Τότε γνωρίζω ότι αν $X_i \sim N(\mu, \sigma^2)$

$$\bar{X} \sim N(\mu, \sigma^2/n)$$

ή ισοδύναμα εδώ

$$\bar{X} \sim N(0, 1/n)$$

Γνωρίζω ότι $(N(0,1))^2 \to \chi^2_1$

Δεν ξέρω τι κάνει $(N(0, 1/n))^2$

Επομένως, $\frac{\bar{X} - 0}{\sqrt{1/n}} = \bar{X}\sqrt{n} \sim N(0,1)$

Τότε $n\bar{X}^2 \sim \chi_1^2 \equiv (N(0,1))^2$

Άσκηση 4η: Έστω X, Y δύο ανεξάρτητες τ.μ. με κατανομές $N(\mu, \sigma^2)$ και X^2_n αντίστοιχα. Να βρεθεί η κατανομή της τ.μ.

$$\frac{n(X-\mu)^2}{\sigma^2 Y}$$

Λύση: Είναι $X \sim N(\mu, \sigma^2)$ άρα (τυπικός μετασχηματισμός)

$$Z = \frac{X-\mu}{\sigma} \sim N(0,1)$$

και

$$\frac{(X-\mu)^2}{\sigma^2} \sim X^2_1 \quad (\text{ορισμός } X^2)$$

Είναι τότε

$$\frac{\frac{(X-\mu)^2}{\sigma^2} \Big/ 1}{Y/n} \underset{\text{ορισμός}}{\overset{X,Y \text{ ανεξ.}}{\sim}} F_{1,n}.$$

Άσκηση 5η: Ένα συνηθισμένο ζάρι ρίχνεται 120 φορές. Να βρεθεί η πιθανότητα το άθροισμα των 120 ενδείξεων να είναι μεγαλύτερο ή ίσο από 400.

Λύση: Έστω X_i η τ.μ. που παριστάνει την ένδειξη στην i-οστή ρίψη ζαριού. Δυνατές τιμές της $X_i = 1, 2, 3, 4, 5, 6$ $i = 1, \ldots, 120$

με $P(X_i = x) = \frac{1}{6}$, $x = 1, 2, 3, 4, 5, 6$.

Ζητείται η εύρεση πιθανότητας

$$P\left(\sum_{i=1}^{120} X_i \geq 400\right)$$

άθροισμα ανεξάρτητων και ισόνομων τ.μ.

Μπορούμε να ~~xxx~~ προσεγγίσουμε την πιθανότητα με το Κ.Ο.Θ.

Ξέρουμε ότι X_i ανεξάρτητες και ισόνομες τ.μ. με $EX_i = \mu < +\infty$

$Var X_i = \sigma^2 < +\infty$ τότε

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

Εδώ είναι

$$EX_i = \sum_{i=1}^{6} x_i P(X = x_i) \Rightarrow$$

$$EX_i = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \dots + \frac{1}{6} \cdot 6 \Rightarrow$$

$$EX_i = \frac{1}{6}(1+2+3+4+5+6) = \frac{21}{6}$$

$$Var X_i = EX_i^2 - (EX_i)^2$$

$$EX_i^2 = \frac{1}{6}(1^2 + 2^2 + 3^2 + 4^2 + 5^2 + 6^2) = \frac{91}{6}$$

Άρα $Var X_i = \frac{91}{6} - \left(\frac{21}{6}\right)^2 = \frac{91 \cdot 6 - 21^2}{6^2} \Rightarrow$

$$Var X_i = \frac{546 - 441}{36} = \frac{105}{36}$$

Άρα

$$P\left(\sum_{i=1}^{120} X_i \geq 400\right) \underset{\text{ΔΙΟΡΘΩΣΗ ΣΥΝΕΧΕΙΑΣ}}{\overset{\text{προσοχή}}{=\!=}} P\left(\sum X_i \geq 399.5\right)$$

$$= P\left(\frac{\sum X_i - 120 \cdot \frac{21}{6}}{\sqrt{120 \cdot \frac{105}{36}}} \geq \frac{399.5 - 120 \cdot \frac{21}{6}}{\sqrt{120 \cdot \frac{105}{36}}}\right)$$

Επομένως

$$P\left(\sum X_i \geq 400\right) \approx P\left(Z \geq \frac{399.5 - 420}{\sqrt{350}}\right)$$

όπου Z ακολουθεί τυπική κατανομή

Η συνέχεια αφήνεται ...

Αποτέλεσμα :

$$P(Z \geq -1.09) = 0.5 + 0.3621 = 0.8621$$

Άσκηση 6η: Το βάρος των ατόμων ενός πληθυσμού ακολουθεί $N(81, 10^2)$. Από τον πληθυσμό διαλέγουμε τυχαία 6 άτομα. Να βρεθεί η πιθανότητα το συνολικό βάρος να υπερβαίνει τα 500 κιλά.

n

Γι' αυτό δεν χρησιμοποιούμε Κ.Ο.Θ.

Λύση: Έστω X_i η τ.μ. που παριστάνει το βάρος του i-οστού ατόμου

Τότε $X_i \sim N(81, 10^2)$ $i=1,\dots,6$

Είναι γνωστό τότε ότι

$$Y=\sum_{i=1}^{6} X_i \sim N\left(\sum_{i=1}^{6} 81, \sum_{i=1}^{6} 10^2\right)$$

$$\equiv N(n\mu, \sigma^2)$$

δηλαδή $Y \sim N(6\cdot 81, 6\cdot 10^2)$

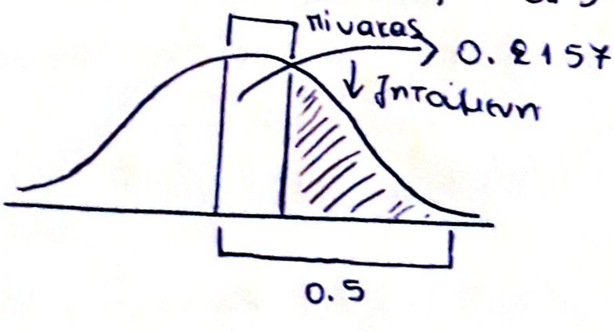
Τότε $$P(Y \geq 500) = P\left(\frac{Y-486}{\sqrt{6\cdot 10^2}} \geq \frac{500-486}{\sqrt{6\cdot 10^2}}\right)$$

Τότε $$Z = \frac{Y-486}{\sqrt{6\cdot 10^2}} \sim N(0,1)$$

Και $$P(Y \geq 500) = P\left(Z \geq \frac{500-486}{\sqrt{6\cdot 10^2}}\right)$$

Ερώτηση προς συζήτηση: γιατί όχι Κ.Ο.Θ???

Αποτέλεσμα: $P(Z \geq 0.57) = 0.5 - 0.2157$



Άσκηση $7^{η}$: Αν οι τ.μ. $X_1, \dots, X_9$ ακολουθούν $N(\mu, \sigma^2)$ να βρεθεί η πιθανότητα

$$P\left(\frac{\sigma}{2} \leq \bar{X} - E(\bar{X}) \leq \sigma\right)$$

Λύση: Όταν $X_1, \dots, X_9 \sim N(\mu, \sigma^2)$ τότε

$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

$$\frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0,1) \quad \text{όπου } n = 9$$

Επομένως $\frac{\bar{X} - \mu}{\sigma/\sqrt{9}} = 3 \frac{\bar{X} - \mu}{\sigma} \sim N(0,1)$

Επίσης ξέρω ότι $E\bar{X} = \mu$

Επομένως

$$P\left(\frac{\sigma}{2} \leq \bar{X} - \mu \leq \sigma\right) \xrightarrow[3 \cdot \frac{\bar{X}-\mu}{\sigma}]{\text{σχηματίζω}} P\left(3 \frac{\sigma}{2\sigma} \leq 3 \cdot \frac{\bar{X}-\mu}{\sigma} \leq \frac{3\sigma}{\sigma}\right)$$

Επομένως

$$P\left(\frac{\sigma}{2} \leq \bar{X} - E\bar{X} \leq \sigma\right) = P\left(\frac{3}{2} \leq Z \leq 3\right)$$

με $Z \sim N(0,1)$ και η τελευταία υπολογίζεται από πίνακες της τυπικής κατανομής (να συνεχιστεί)

Αποτέλεσμα: $0.4987 - 0.4332 = 0.0655$

Άσκηση $8^{η}$: Αν $\bar{X}$ είναι η μέση τιμή τυχαίου δείγματος μεγέθους n από την κανονική $N(\mu, 100)$ να βρεθεί το n έτσι ώστε

$$P(\mu - 5 < \bar{X} < \mu + 5) = 0.954$$

Λύση: $\bar{X} \sim N(\mu, \sigma^2/n)$

Επομένως $\bar{X} \sim N(\mu, 100/n)$

Έτσι είναι $P(\mu - 5 < \bar{x} < \mu + 5) = 0.954 \Rightarrow$

$$P\left(\frac{\mu-5-\mu}{\sqrt{\frac{100}{n}}} < \frac{\bar{x}-\mu}{\sqrt{\frac{100}{n}}} < \frac{\mu+5-\mu}{\sqrt{\frac{100}{n}}}\right) = 0.954$$

ή ισοδύναμα

$$P\left(-\sqrt{n}\,\frac{5}{10} < Z < \sqrt{n}\,\frac{5}{10}\right) = 0.954$$

ή

$$2 \cdot P(0 < Z < 0.5\sqrt{n}) = 0.954$$

ή $P(0 < Z < 0.5\sqrt{n}) = \dfrac{0.954}{2} \Rightarrow$

$0.5\sqrt{n} = 2 \Rightarrow \sqrt{n} = 4 \Rightarrow n = 16$

Άσκηση 9η: Δίνεται τ.δ. $X_1, X_2, \dots, X_{10}$ από πληθυσμό $N(0,1)$ και $\bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i$. Έστω X_{11} μια άλλη ανεξάρτητη παρατήρηση από τον ίδιο πληθυσμό και

$$Y_1 = \sum_{i=1}^{10}(X_i - \bar{X})^2 + X_{11}^2$$

$$Y_2 = \frac{\sqrt{10}\,X_{11}}{\sqrt{\sum_{i=1}^{10} X_i^2}}$$

$$Y_3 = \frac{9(10\bar{X}^2 + X_{11}^2)}{2\sum_{i=1}^{10}(X_i - \bar{X})^2}$$

Να βρεθούν οι σταθερές c_1, c_2, c_3 έτσι ώστε

(i) $P(Y_1 \leq c_1) = 0.01$

(ii) $P(Y_2 \geq -c_2) = 0.9$

(iii) $P(Y_3 \geq c_3) = 0.95$

Λύση: Υπενθύμιση αποτελεσμάτων δειγματοληψίας από κανονικούς πληθυσμούς

$$\bar{X} \sim N(\mu, \sigma^2/n) \qquad \frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

$\bar{x}, S^2$ ανεξάρτητα

Έχουμε ότι $X_1, X_2, \ldots, X_{10}$ τ.δ. $N(0,1)$

αυτό σημαίνει ότι

$$\bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i \sim N(0, 1/10)$$

και ότι

$$\frac{(n-1)S^2}{\sigma^2} = \frac{9 \cdot \frac{1}{9}\sum_{i=1}^{10}(X_i - \bar{X})^2}{1} \sim \chi^2_9$$

Άρα προκύπτει ότι

$$\bar{X} \sim N\left(0, \frac{1}{10}\right) \quad (1)$$

$$\sum_{i=1}^{10}(X_i - \bar{X})^2 \sim \chi^2_9$$

είναι $N(0,1)$ το τετράγωνο -> άρα χ^2_1

Είναι τότε χ^2_9

$$Y_1 = \sum_{i=1}^{10}(X_i - \bar{X})^2 + X_{11}^2 \equiv \chi^2_{10}$$

έχω δείξει ότι ακολουθεί χ^2_9

χ^2_1

αφού $\chi^2_9 + \chi^2_1$

με X_{11} ανεξάρτητες από $X_1, \ldots, X_{10}$.

Επομένως $Y_1 \sim \chi^2_{10}$

Επίσης $X_1, \dots, X_{10} \sim N(0,1)$

άρα $X_1^2 + \dots + X_{10}^2 = \sum_{i=1}^{10} X_i^2 \sim \chi^2_{10}$

Ενώ $X_{11} \sim N(0,1)$

Επομένως $\dfrac{\overset{\to N(0,1)}{X_{11}}}{\underbrace{\sqrt{\sum_{i=1}^{10} X_i^2/10}}_{\chi^2_{10}} \to \text{α.ε.}} \sim t_{10}$

δηλαδή $Y_2 \sim t_{10}$

Είναι $X_{11}^2 \sim \chi^2_1$

$\bar{X} \sim N(0, \frac{1}{10}) \Rightarrow$

$\dfrac{\bar{X}}{\sqrt{\frac{1}{10}}} \sim N(0,1) \Rightarrow$

$\sqrt{10}\,\bar{X} \sim N(0,1) \Rightarrow$

$10\bar{X}^2 \sim \chi^2_1$

Είναι $\bar{X}, X_{11}$ ανεξάρτητα $\left(\bar{X} = \frac{1}{10}\sum_{i=1}^{10} X_i\right)$

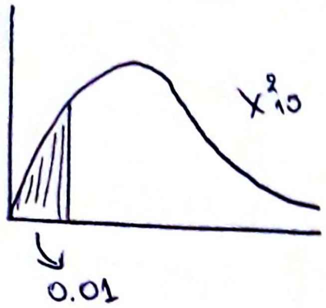
οπότε $10\bar{X}^2 + X_{11}^2 \sim \chi^2_2$

Άρα $\dfrac{10\bar{X}^2 + X_{11}^2}{\sum_{i=1}^{10}(X_i - \bar{X})^2/9} \overset{\text{οριακώς}}{\sim} F_{2,9}$ ← Θα πρέπει να επιχειρηματολογήσω ότι αυτό είναι ανεξάρτητο

Δηλαδή $Y_3 \sim F_{2,9}$

(i) $P(Y_1 \le c_1) = 0.01 \Rightarrow$

$P(\chi^2_{10} \le c_1) = 0.01$

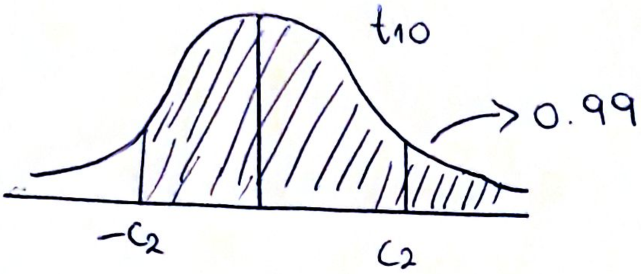


Άρα $P[X^2_{10} \geq c_1] = 0.99$

$\Rightarrow c_1 = X^2_{0.99,10} = 2.558$

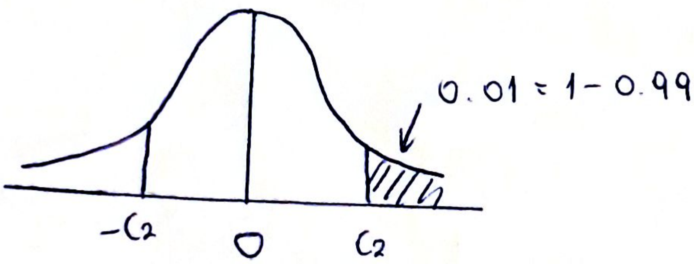
(ii) $P(Y_2 \geq -c_2) = 0.99$

όπου $Y_2 \sim t_{10}$



Άρα $-c_2$ πρέπει να είναι αρνητικός (c_2 θετικός, αλλιώς η πιθανότητα δεν μπορεί να είναι 0.99)

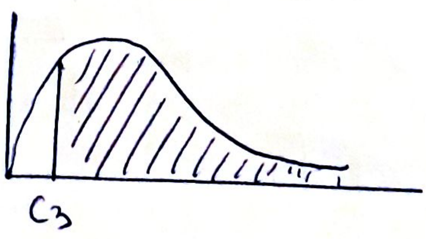
Από σχήμα



Άρα $P(Y_2 \geq c_2) = 0.01 \Rightarrow c_2 = t_{0.01,10} = 2.764$

(iii) $P(Y_3 \geq c_3) = 0.95$

όπου $Y_3 \sim F_{2,9}$



Άρα από ορισμό

$c_3 = F_{0.95,2,9}$

Το σημείο αυτό δεν υπάρχει στους πίνακες. Πως θα το βρω;

Θα δείξω ότι

$$F_{0.95,2,9} = \frac{1}{F_{0.05,9,2}} = \frac{1}{99.4} = 0.01$$ (Δείτε απόδειξη στη θεωρία)

Άσκηση 10η: Έστω $X_1, \ldots, X_n$ τ.δ. από $N(0,1)$. Έστω επίσης

$$\bar{X}_k = \frac{1}{k} \sum_{i=1}^{k} X_i$$

$$\bar{X}_{n-k} = \frac{1}{n-k} \sum_{k+1}^{n} X_i$$

Να βρεθούν οι κατανομές των:

(i) $\frac{\bar{X}_k + \bar{X}_{n-k}}{2}$

(ii) $k \bar{X}_k^2 + (n-k) \bar{X}_{n-k}$

(iii) $\underset{\chi_1^2}{X_1^2} / \underset{\chi_1^2}{X_2^2} \to$ εύκολα $F_{1,1}$ διότι $\frac{X_1^2 / 1}{X_2^2 / 1}$

Λύση: Από γνωστή πρόταση

$$\bar{X}_k \sim N\left(\mu = 0, \ \frac{\sigma^2}{k} = 1/k\right)$$

και $\bar{X}_{n-k} \sim N\left(\mu = 0, \ \frac{\sigma^2}{n-k} = \frac{1}{n-k}\right)$ } ανεξάρτητες

με $\bar{X}_k, \bar{X}_{n-k}$ ανεξάρτητες

Είναι $\overset{N(0, 1/k)}{\bar{X}_k} + \overset{N(0, \frac{1}{n-k})}{\bar{X}_{n-k}} \sim N\left(0, \ \frac{1}{k} + \frac{1}{n-k}\right)$

ή $\bar{X}_k + \bar{X}_{n-k} \sim N\left(0, \ \frac{n-k+k}{k(n-k)}\right)$

ή $\frac{\bar{X}_k + \bar{X}_{n-k}}{k} \sim N\left(0, \ \frac{n}{4k(n-k)}\right)$

(ii) $\bar{X}_k \sim N\left(0, \frac{1}{k}\right) \Rightarrow$

$$\frac{\bar{X}_k}{\sqrt{\frac{1}{k}}} \sim N(0,1) \Rightarrow k \bar{X}_k^2 \sim \chi_1^2$$

Με όμοιο τρόπο $(n-k) \bar{X}_{n-k}^2 \sim \chi_1^2$

Ανεξάρτητες άρα $k \bar{X}_k^2 + (n-k) \bar{X}_{n-k}^2 \sim \chi_2^2$