Εισαγωγή στην Στατιστική ①
1ο online μάθημα
26/3/2020

Στο προηγούμενο μάθημα είδαμε ότι αν $\overbrace{X_1, X_2, \dots, X_n}^{\text{κεφαλαία}}$ είναι n το πλήθος ανεξάρτητες και ισόνομες τ.μ. με $X_i \sim N(0,1)$ τότε η κατανομή της τ.μ. $Y = X_1^2 + \dots + X_n^2$ λέμε ότι είναι X_n^2 (χι-τετράγωνο με n β.ε) → βαθμούς ελευθερίας

Ισχύουν $X_n^2 \underset{\text{ταυτίζεται}}{\equiv} \Gamma(\frac{n}{2}, 2)$

$$E X_n^2 = n$$

$$\operatorname{Var} X_n^2 = 2n$$

Επίσης αν $\left.\begin{array}{l} X \sim N(0,1) \\ \text{και } Y \sim X_n^2 \end{array}\right\}$ ανεξάρτητες τ.μ.

Τότε η κατανομή της τ.μ.

$$Z = \frac{X}{\sqrt{Y/n}}$$

λέμε ότι είναι η t με n βαθμούς ελευθερίας

$$E(t_n) = 0 \quad , n > 1$$

$$\operatorname{Var}(t_n) = \frac{n}{n-2} \quad , n > 2$$

Όσο $n \to \infty$ οι κατανομές t_n και $N(0,1)$ τείνουν να ταυτιστούν.

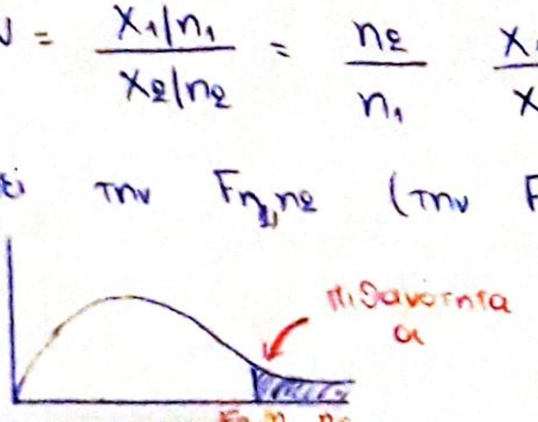
Έστω τώρα ότι

$$\left.\begin{array}{l} X_1 \sim X_{n_1}^2 \\ X_2 \sim X_{n_2}^2 \end{array}\right\} X_1, X_2 \text{ ανεξάρτητες τ.μ}$$

n_1 β.ε της X_1
n_2 β.ε της X_2

Τότε η τ.μ. $W = \dfrac{X_1/n_1}{X_2/n_2} = \dfrac{n_2}{n_1} \dfrac{X_1}{X_2}$

λέμε ότι ακολουθεί την F_{n_1,n_2} (την F κατανομή με β.ε. n_1, n_2)



(Είναι έτσι και στο χι τετράγωνο που αφ[illegible] και της παρουσίασης)

F_{a,n_1,n_2} είναι το σημείο εκείνο τ/ω

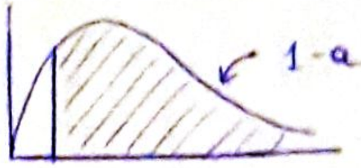
$$P[F_{n_1,n_2} > F_{a,n_1,n_2}] = a$$

Ν.δ.ο.

$$F_{1-a,n_1,n_2} = \frac{1}{F_{a,n_2,n_1}}$$

Απόδειξη: Σύμφωνα με τον ορισμό F_{1-a,n_1,n_2} είναι εκείνο το σημείο για το οποίο

$$P[F_{n_1,n_2} > F_{1-a,n_1,n_2}] = 1-a$$



F_{1-a,n_1,n_2}

Όμως $F_{n_1,n_2} = \frac{n_2}{n_1}\frac{X_1}{X_2}$ όπου $X_1 \sim X^2_{n_1}$, $X_2 \sim X^2_{n_2}$

με X_1, X_2 ανεξάρτητες τ.μ.

Άρα έχουμε:

$$P\left[\frac{n_2}{n_1}\frac{X_1}{X_2} > F_{1-a,n_1,n_2}\right] = 1-a$$

$$\Rightarrow P\left[\frac{n_1}{n_2}\frac{X_2}{X_1} \leq \frac{1}{F_{1-a,n_1,n_2}}\right] = 1-a$$

(Πώς; Σκεφτόμαστε ότι όταν $k>\lambda$ με $k,\lambda>0$ τότε $\frac{1}{k} < \frac{1}{\lambda}$)

Επομένως, $P\left[\frac{n_1}{n_2}\frac{X_2}{X_1} \leq \frac{1}{F_{1-a,n_1,n_2}}\right] = 1-a$

$$\Rightarrow 1 - P\left[\frac{n_1}{n_2}\frac{X_2}{X_1} \geq \frac{1}{F_{1-a,n_1,n_2}}\right] = 1-a$$

συνεχείς τ.μ.

διότι $P(A') = 1-P(A)$

Άρα $P\left[\frac{n_1}{n_2}\frac{X_2}{X_1} \geq \frac{1}{F_{1-a,n_1,n_2}}\right] = a$

$X_2 \sim X^2_{n_2}$, $X_1 \sim X^2_{n_1}$, $\frac{X_2/n_2}{X_1/n_1} \sim F_{n_2,n_1}$

$$\Rightarrow P\left[F_{n_2,n_1} \geq \frac{1}{\underset{\uparrow\ F_{a,n_2,n_1}}{F_{1-a,n_1,n_2}}}\right] = a$$

Γιατί;

$X_2 \sim X^2_{n_2}$, $X_1 \sim X^2_{n_1}$ } ανεξάρτητες τμ άρα $\frac{X_2/n_2}{X_1/n_1} \sim F_{n_2,n_1}$

$\Downarrow$

$\frac{n_1}{n_2}\frac{X_2}{X_1} \sim F_{n_2,n_1}$

Από την τελευταία σχέση και ενθυμούμενοι ότι

$$P[F_{n_2,n_1} \geq F_{a,n_2,n_1}] = a$$

Έχουμε ότι:

$$F_{a,n_2,n_1} = \frac{1}{F_{1-a,n_1,n_2}}$$

Πρόταση: Ισχύει ότι $t^2_n = F_{1,n}$ και $t^2_{a,n} = F_{a,1,n}$

Απόδειξη: Θυμόμαστε ότι εξ' ορισμού αν $X \sim N(0,1)$, $Y \sim X^2_n$ } X, Y ανεξάρτητες τ.μ.

τότε $Z = \frac{X}{\sqrt{Y/n}} \sim t_n$

Επομένως $t_n^2 = Z^2 = \frac{X^2}{Y/n}$

Όμως $X \sim N(0,1)$ άρα $X^2 \sim X^2_1$, $Y \sim X^2_n$ } ανεξάρτητες αφού X, Y είναι ανεξάρτητες

$\Downarrow$

$$Z^2 = \frac{X^2/1}{Y/n} \sim F_{1,n}$$

Επιπρόσθετα $t_{a,n}$ είναι το σημείο εκείνο τ.ω

$$P[t_n \geq t_{a,n}] = a \Rightarrow P[t^2_n \geq t^2_{a,n}] = a$$

$$\xRightarrow{t^2_n = F_{1,n}} P[F_{1,n} \geq t^2_{a,n}] = a$$

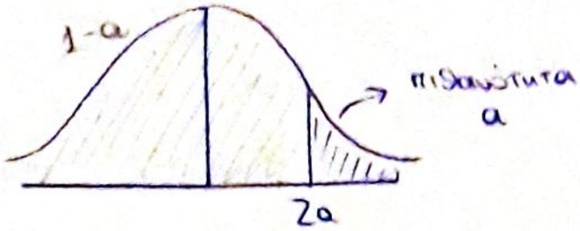
Όμως εξ' ορισμού $P[F_{1,n} \geq F_{a,1,n}] = a$

Άρα $t^2_{a,n} = F_{a,1,n}$

Υπενθύμιση - Σύνοψη

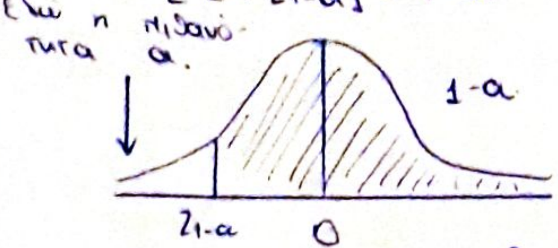
z_a είναι εκείνο το σημείο για το οποίο $P[Z \geq z_a] = a$

με $Z \sim N(0,1)$

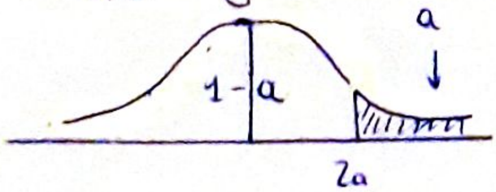


Τι ισχύει για το z_{1-a};

$$P[Z \geq z_{1-a}] = 1-a$$



Λόγω συμμετρίας



Επομένως z_{1-a} το συμμετρικό του z_a, δηλαδή $z_{1-a} = -z_a$

$z_{0,95} = -z_{0,05}$

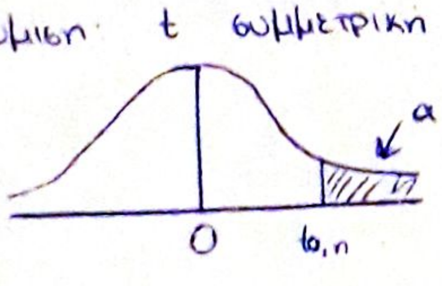
$z_{0,99} = -z_{0,01}$

$t_{a,n}$ είναι εκείνο το σημείο για το οποίο

$$P[t \geq t_{a,n}] = a$$

με $t \sim t_n$

Υπενθύμιση: t συμμετρική γύρω από το 0.



Τι ισχύει για το $t_{1-a,n}$;

Με παρόμοιο σκεπτικό προκύπτει ότι $t_{1-a,n} = -t_{a,n}$ ← Να το δω σπίτι.

Υπενθύμιση: $F_{1-a,n_1,n_2} = \dfrac{1}{F_{a,n_2,n_1}}$

Θεώρημα 1ο:

Έστω $X_1, X_2, \dots, X_n$ τυχαίο δείγμα (ανεξάρτητες και ισόνομες τ.μ.) από έναν πληθυσμό με μέση τιμή μ και διακύμανση σ^2.

σημαίνει ότι

η κάθε μία έχει ίδια κατανομή με την άλλη

$EX = \mu$

$Var\, X = \sigma^2$

Ισχύει ότι:

(i) $E\bar{X} = \mu$

(ii) $Var\,\bar{X} = \sigma^2/n$

(iii) $ES^2 = \sigma^2$

$E(aX) = aEX \quad E(b) = b$

$E(aX+b) = aEX + b \quad E(X+Y) = EX + EY$

Απόδειξη:

(i) $E\bar{X} \xlongequal{\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i} E\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \xRightarrow[\text{μέσης τιμής}]{\text{ιδιότητες}} E\bar{X} = \frac{1}{n}\sum_{i=1}^{n} EX_i$

$\Longrightarrow E\bar{X} = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu. \qquad \frac{1}{n}\cdot n\mu = \mu$

(ii) $Var\,\bar{X} = Var\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \Longrightarrow Var\,\bar{X} = \frac{1}{n^2}\sum_{i=1}^{n} Var\, X_i$

ανεξάρτητες

(η προηγούμενη σχέση ισχύει από ιδιότητες διακύμανσης και λόγω της ανεξαρτησίας των τ.μ.)

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

$Var(aX + bY) \xlongequal[\text{ξάρτητα}]{X,Y \text{ ανε-}} a^2\, Var\, X + b^2\, Var\, Y.$

Επομένως,

$$Var\,\bar{X} = \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 = \sigma^2/n$$

(iii) $ES^2 = E\left[\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2\right]$

Ας αναπτύξουμε τον όρο $\sum_{i=1}^{n}(X_i - \bar{X})^2$.

Είναι:

$$\sum_{i=1}^{n}(X_i - \bar{X})^2 = \sum_{i=1}^{n} X_i^2 - n\bar{X}^2$$

Είναι τότε από ιδιότητες αναμενόμενης τιμής

$$ES^2 = \frac{1}{n-1} E\left[\sum_{i=1}^{n} X_i^2 - n\bar{X}^2\right] = \frac{1}{n-1}\left[\sum_{i=1}^{n} EX_i^2 - nE\bar{X}^2\right]$$

Είναι όμως:

$$Var X_i = EX_i^2 - (EX_i)^2 \Rightarrow EX_i^2 = Var X_i + (EX_i)^2$$

Επομένως

$$EX_i^2 = \sigma^2 + \mu^2$$

Επίσης με παρόμοιο σκεπτικό

$$E\bar{X}^2 = Var\bar{X} + (E\bar{X})^2 \xRightarrow[(ii)]{\text{βλέπε (i) και}}$$

$$E\bar{X}^2 = \sigma^2/n + \mu^2$$

Συνοψίζοντας έχουμε:

$$ES^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n}\overbrace{(\sigma^2+\mu^2)}^{EX_i^2} - n\overbrace{\left(\frac{\sigma^2}{n}+\mu^2\right)}^{E\bar{X}^2}\right] \Rightarrow$$

$$ES^2 = \frac{1}{n-1}\left[n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2\right] \Rightarrow ES^2 = \frac{1}{n-1}(n-1)\sigma^2 \Rightarrow$$

$$ES^2 = \sigma^2$$

Ποια είναι η αξία των αποτελεσμάτων (i) και (iii); $E\bar{X} = \mu$ $ES^2 = \sigma^2$

Η μέση τιμή της δειγματικής κατανομής του $\bar{X}$ και S^2 αντίστοιχα ισούται με την πληθυσμιακή μέση τιμή μ και πληθυσμιακή διακύμανση σ^2 αντίστοιχα. $X_1, X_2, \dots, X_n$ $\bar{X}$ S^2

Αν μια δειγματοληψία επαναληφθεί πολλές φορές και καταγράφοντας τα $\bar{X}$ και S^2 τότε ο μ όρος του r προσεγγίζει τα μ, σ^2 αντίστοιχα.

$m_{X_1}(t)$ $m_{X_2}(t)$ $m_{X_n}(t)$

Πόρισμα*: Έστω $X_1, \ldots, X_n$ τ.δ. από έναν πληθυσμό με μέση τιμή μ_1 και διακύμανση σ_1^2. Έστω $Y_1, \ldots, Y_n$ τ.δ. από έναν πληθυσμό με μέση τιμή μ_2 και διακύμανση σ_2^2.

Υποθέτουμε ότι είναι τα δύο τ.δ. ανεξάρτητα μεταξύ τους.

Τότε ισχύουν τα ακόλουθα:

(i) $E(\bar{X} - \bar{Y}) = \mu_1 - \mu_2$

(ii) $Var(\bar{X} - \bar{Y}) = \dfrac{\sigma_1^2}{n} + \dfrac{\sigma_2^2}{m}$ όπου $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\bar{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$

Απόδειξη: Θυμηθείτε ότι

$E(aX + bY) = aEX + bEY$ και ότι αν X, Y ανεξάρτητες τότε

$Var(aX + bY) = a^2 Var X + b^2 Var Y$

*

ΔΕΙΓΜΑΤΟΛΗΨΙΑ ΑΠΟ ΚΑΝΟΝΙΚΟΥΣ ΠΛΗΘΥΣΜΟΥΣ.

Στην ενότητα αυτή θα δοθούν κάποια αποτελέσματα που αφορούν την δειγματοληψία από κανονικούς πληθυσμούς.

Έστω $X_1, X_2, \dots, X_n$ ανεξάρτητες τ.μ. με $X_i \sim N(\mu_i, \sigma_i^2)$

Το ερώτημα είναι ποια η κατανομή της τ.μ.

$$Y=\sum_{i=1}^{n} X_i$$

Μια πολύ εύχρηστη (αλλά όχι πάντοτε αποτελεσματική) μέθοδος για την εύρεση της κατανομής αθροίσματος τ.μ. είναι η μέθοδος της ροπογεννήτριας.

Υπενθύμιση: Αν X τ.μ. τότε η ροπογεννήτρια της

$$m_X(t)=E[e^{tX}] \quad \forall t\in(-\delta,\delta)$$

Η μέθοδος βασίζεται στο μονοσήμαντο των ροπογεννητριών που λέει ότι η ροπογεννήτρια μιας κατανομής χαρακτηρίζει μονοσήμαντα μια κατανομή.

Είναι η ροπογεννήτρια της τ.μ. $Y=\sum_{i=1}^{n} X_i$:

$$m_Y(t)=E(e^{tY}) \xRightarrow{Y=\sum_{i=1}^{n} X_i} m_Y(t)=E(e^{tX_1}e^{tX_2}\cdots e^{tX_n})$$

Ισχύει ότι (θα το δούμε στο μάθημα 531)

$$E(e^{tX_1})\,E(e^{tX_2})\cdots E(e^{tX_n})$$

$$E[h(X_1)\dots h(X_n)] \xrightarrow[\text{ανεξ.}]{X_1\dots X_n} E[h(X_1)]\dots E[h(X_n)]$$

$$m_{X_1}(t) \qquad m_{X_2}(t) \qquad m_{X_n}(t)$$

Επομένως

$$m_Y(t) = m_{X_1}(t) \dots m_{X_n}(t)$$

Από τυπολόγιο ξέρουμε ότι $m_{X_i}(t) = e^{\mu_i t + \frac{\sigma_i^2 t^2}{2}}$ όταν

$$X_i \sim N(\mu_i, \sigma_i^2)$$

Επομένως

$$m_Y(t) = e^{\mu_1 t + \frac{\sigma_1^2 t^2}{2}} \dots e^{\mu_n t + \frac{\sigma_n^2 t}{2}}$$

$$\Rightarrow m_Y(t) = e^{t \sum_{i=1}^{n} \mu_i + \frac{t^2}{2} \sum_{i=1}^{n} \sigma_i^2}$$

Που ταυτίζεται με τη ροπογεννήτρια της

$$N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

Αποδείξαμε ότι αν $X_i \sim N(\mu_i, \sigma_i^2)$

τότε

$$\sum_{i=1}^{n} X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} \sigma_i^2\right)$$

Με παρόμοιο τρόπο

$$Y_1 = \sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

Καθώς εύκολα - όμοια

$$m_{Y_1}(t) = m_{a_1 X_1}(t) \dots m_{a_n X_n}(t)$$

$$\Downarrow$$

$$m_{Y_1}(t) = m_{X_1}(a_1 t) \dots m_{X_n}(a_n t)$$

$$= e^{\mu_1 a_1 t + \frac{\sigma_1^2}{2} a_1^2 t^2} \dots e^{\mu_n a_n + \frac{\sigma_n^2}{2} a_n^2 t^2}$$

$$\Rightarrow m_{Y_1}(t) = e^{t \sum_{i=1}^{n} a_i \mu_i + \frac{t^2}{2} \sum_{i=1}^{n} a_i^2 \sigma_i^2}$$

$$\sum a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

Θεώρημα:

Έστω $X_1, X_2, \ldots, X_n$ τ.δ. από $N(\mu, \sigma^2)$. Τότε

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

Το αποτέλεσμα αυτό προκύπτει ως ειδική περίπτωση του ότι

$$\sum_{i=1}^{n} a_i X_i \sim N\left(\sum_{i=1}^{n} a_i \mu_i, \sum_{i=1}^{n} a_i^2 \sigma_i^2\right)$$

για $a_i = \frac{1}{n}$, $\mu_i = \mu$, $\sigma_i^2 = \sigma^2$ (Όλα αυτά όταν έχω κανονική κατανομή)

(ii) $$\frac{(n-1)S^2}{\sigma^2} \sim X^2_{n-1}$$

ή ισοδύναμα $$\sum_{i=1}^{n} \frac{(X_i - \bar{x})^2}{\sigma^2} \sim X^2_{n-1}$$

(Χωρίς απόδειξη)

(iii) $\bar{X}, S^2$ ανεξάρτητες σ.σ. (Χωρίς απόδειξη)

σ.σ. → στατιστικές συναρτήσεις

Θεώρημα:

Έστω $X_1, X_2, \ldots, X_n$ τ.δ. $N(\mu, \sigma^2)$. Τότε

(i) αν σ^2 γνωστή ισχύει ότι:

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

(ii) αν σ^2 άγνωστη ισχύει ότι

$$\frac{\bar{X} - \mu}{S / \sqrt{n}} \sim t_{n-1}.$$

Απόδειξη:

(i) $\bar{X} \sim N(\mu, \sigma^2/n)$ από προηγούμενο θεώρημα.

Από ιδιότητες κανονικής κατανομής (τυπικός μετασχηματισμός) $X \sim N(\mu, \sigma^2)$

$$\frac{\bar{X}-\mu}{\sqrt{\sigma^2/n}} \sim N(0,1) \quad , \text{ δηλαδή } \quad \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1) \quad \frac{X-\mu}{\sigma} \sim N(0,1)$$

(ii) Από την (i) έχουμε ότι $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$ $\quad ES^2 = \sigma^2$

Όμως η διακύμανση σ^2 είναι άγνωστη. Γνωρίζουμε ότι:

$$\frac{(n-1)S^2}{\sigma^2} \sim X^2_{n-1}$$

Επιπλέον, $\bar{X}$, S^2 ανεξάρτητες.

Τότε ενθυμούμενοι τον ορισμό της t είναι

$$\frac{\overbrace{\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}}^{N(0,1)}}{\sqrt{\underbrace{\frac{(n-1)S^2}{\sigma^2}}_{X^2}\Big/\underbrace{(n-1)}_{\text{β.ε.}}}} \sim t_{n-1}$$

που αποδεικνύει το ζητούμενο.

Θεώρημα:

Έστω $X_1, X_2, \ldots, X_n$ τ.δ. $N(\mu_1, \sigma^2)$

$Y_1, Y_2, \ldots, Y_n$ τ.δ. $N(\mu_2, \sigma^2)$

τα οποία είναι ανεξάρτητα.

Τότε:

$$\frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{S_p\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim t_{n+m-2} \qquad \bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2, \frac{\sigma^2}{n}+\frac{\sigma^2}{m}\right)$$

όπου

$$\underbrace{S_p}_{\text{pooled (μικτό δείγμα)}}{}^2 = \frac{(n-1)S_x^2+(m-1)S_Y^2}{n+m-2}$$

με $S^2_X = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$

$S^2_Y = \frac{1}{m-1} \sum_{i=1}^{m} (Y_i - \bar{Y})^2$

Απόδειξη:

Είναι $\bar{X} \sim N(\mu_1, \sigma^2/n)$

$\bar{Y} \sim N(\mu_2, \sigma^2/m)$ (Προηγούμενο θεώρημα)

Λόγω ανεξαρτησίας (δες πρόταση $\sum_{i=1}^{n} a_i X_i$ για κανονικούς)

$$\bar{X} - \bar{Y} \sim N\left(\mu_1 - \mu_2, \frac{\sigma^2}{n} + \frac{\sigma^2}{m}\right)$$

Από ιδιότητες κανονικής

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma^2}{n} + \frac{\sigma^2}{m}}} \sim N(0,1)$$

ή

$$\frac{(\bar{X} - \bar{Y}) - (\mu_1 - \mu_2)}{\sigma\sqrt{\frac{1}{n} + \frac{1}{m}}} \sim N(0,1)$$

Το πρόβλημα λοιπόν είναι όταν το σ^2 είναι άγνωστη παράμετρος. Αν δεν είναι τότε μένουμε στο προηγούμενο αποτέλεσμα.

Η σ^2 είναι η διακύμανση ______ Άρα λογικό είναι να πρέπει να αξιοποιηθεί η διαθέσιμη πληροφορία και από τα δύο δείγματα για αυτήν δηλαδή τόσο από το $X_1 \dots X_n$ όσο και από το $Y_1 \dots Y_m$. Είναι από το προηγούμενο θεώρημα:

$$\left.\begin{aligned} \frac{(n-1) S^2_X}{\sigma^2} &\sim \chi^2_{n-1} \\ \frac{(m-1) S^2_Y}{\sigma^2} &\sim \chi^2_{m-1} \end{aligned}\right\} \text{ανεξάρτητες}$$

από το προηγούμενο θεώρημα

από το προηγούμενο θεώρημα

Από τον ορισμό της X^2 κατανομής εύκολα οδηγούμαστε στο ότι:

$$\frac{(n-1)S_x^2}{\sigma^2} + \frac{(m-1)S_y^2}{\sigma^2} \sim X^2_{(n-1)+(m-1)}$$

Άρα $$\frac{(n-1)S_x^2 + (m-1)S_y^2}{\sigma^2} \sim X^2_{n+m-2}$$

Επομένως

$$\frac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}} \sim N(0,1)$$

$$\frac{(n-1)S_x^2 + (m-1)S_y^2}{\sigma^2} \sim X^2_{n+m-2}$$

με ανεξάρτητα $\bar{X}, \bar{Y}, S_x^2, S_y^2$ (Βλέπε Πρόταση (iii)) και από τον ορισμό t κατανομής προκύπτει ότι:

$$\frac{\dfrac{(\bar{x}-\bar{y})-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n}+\frac{1}{m}}}}{\sqrt{\dfrac{(n-1)S_x^2+(m-1)S_y^2}{\sigma^2} \Big/ (n+m-2)}} \sim t_{n+m-2}$$

που αποδεικνύει το ζητούμενο.

Θεώρημα:

Έστω $X_1, X_2, \dots, X_n$ τ.δ. $N(\mu_1, \sigma_1^2)$ και $Y_1, Y_2, \dots, Y_n$ τ.δ. $N(\mu_2, \sigma_2^2)$, ανεξάρτητα μεταξύ τους.

Τότε ισχύει ότι:

$$\frac{S_x^2 / \sigma_1^2}{S_y^2 / \sigma_2^2} \sim F_{n-1,m-1}$$

όπου $S_X^2 = \frac{1}{n-1} \sum_{i=1}^{n} (X_i - \bar{X})^2$

$S_Y^2 = \frac{1}{m-1} \sum_{i=1}^{m} (Y_i - \bar{Y})^2$

Απόδειξη:

Είναι γνωστό ότι

$$\left.\begin{array}{l} \frac{(n-1) S_X^2}{\sigma_1^2} \sim \chi^2_{n-1} \\ \\ \frac{(m-1) S_Y^2}{\sigma_2^2} \sim \chi^2_{m-1} \end{array}\right\}$$

ανεξαρτησία λόγω ανεξαρτησία δειγμάτων.

Τότε εξ ορισμού της F κατανομής

$$\frac{\frac{(n-1) S_X^2}{\sigma_1^2} \Big/ (n-1)}{\frac{(m-1) S_Y^2}{\sigma_2^2} \Big/ (m-1)} \sim F_{n-1,\, m-1}$$

που αποδεικνύει το ζητούμενο