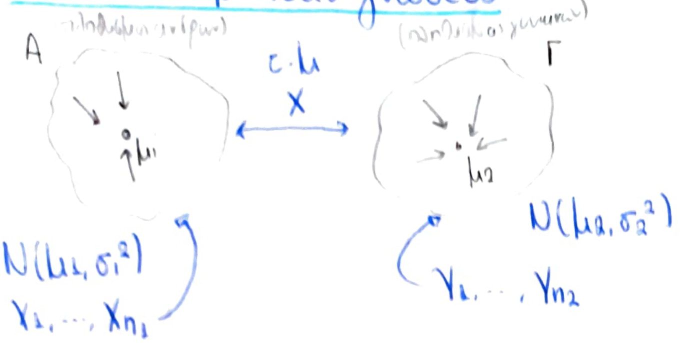
ΤΡΙΤΗ 20 ΑΠΡΙΛΙΟΥ 2021 - ΣΤΑΤΙΣΤΙΚΗ ΣΥΜΠΕΡΑΣΜΑΤΟΛΟΓΙΑ.

Διάστημα εμπιστοσύνης για την διαφορά μέσων τιμών ανεξάρτητων κανονικών πληθυσμών

(Ι) Οι διακυμάνσεις γνωστές



$\bar X$ εκτιμητής του μ_1 $\qquad$ $\bar Y$: εκτιμητής του μ_2.

Άρα $\boxed{\bar X - \bar Y}$ (σημειακός): εκτιμητής του $\mu_1-\mu_2$.

Θεωρώ τ.δ. (ανεξ.) $X_1,\dots,X_{n_1}$ από $N(\mu_1,\sigma_1^2)$ και τ.δ. $Y_1,\dots,Y_{n_2}$ από $N(\mu_2,\sigma_2^2)$ και υποθέτω ότι σ_1^2, σ_2^2 γνωστές. Ζητώ δ.ε για την διαφορά $\mu_1-\mu_2$.

Είναι: $\bar X \sim N\left(\mu_1, \frac{\sigma_1^2}{n_1}\right)$, $\bar Y \sim N\left(\mu_2, \frac{\sigma_2^2}{n_2}\right)$ - επειδή τα δείγματα είναι ανεξάρτητα:
(αν δεν είχα ανεξαρτησία θα έπρεπε να είχα και τις cov)

$$\bar X - \bar Y \sim N\left(\mu_1-\mu_2, \frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}\right)$$

Άρα: $$\boxed{\frac{\bar X - \bar Y - (\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)}$$

Θεωρώ την αναλλοίωτη ποσότητα $$Q = \frac{\bar X - \bar Y - (\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

επειδή Q αναλλοίωτη $\exists\, q_1, q_2$ $(q_1<q_2)$: $1-\alpha = P(q_1 < Q < q_2) = P\left(q_1 < \frac{\bar X - \bar Y - (\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} < q_2\right) =$

$$= P\left((\bar X - \bar Y) - q_2\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} < \mu_1-\mu_2 < (\bar X - \bar Y) - q_1\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$

Άρα, ένα δ.ε για την $\mu_1-\mu_2$ με σ.ε $100(1-\alpha)\%$ είναι:

$$\left((\bar X - \bar Y) - q_2\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}},\ (\bar X - \bar Y) - q_1\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}\right)$$

Ακολουθώντας την μεθοδολογία εύρεσης δ.ε ελαχίστου μήκους, το $100(1-\alpha)\%$ δε

→ δ.ε για την $\mu_1-\mu_2$ ελαχίστου μήκους είναι:

$$\boxed{\left((\bar{X}-\bar{Y}) - z_{a/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \ , \ (\bar{X}-\bar{Y}) + z_{a/2}\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}} \right)}$$

(II) Οι διακυμάνσεις άγνωστες και ίσες

Θεωρώ ανεξ. τ.δ. $X_1, \dots, X_{n_1}$ από $N(\mu_1, \sigma_1^2)$ και τ.δ. $Y_1, \dots, Y_{n_2}$ από $N(\mu_2, \sigma_2^2)$ και υποθέτω ότι σ_1^2, σ_2^2 άγνωστες και ίσες $\sigma_1^2 = \sigma_2^2 = \sigma^2$. Ζητώ δ.ε για την διαφορά $\mu_1-\mu_2$. Είναι: $\bar{X} \sim N(\mu_1, \frac{\sigma^2}{n_1})$, $\bar{Y} \sim N(\mu_2, \frac{\sigma^2}{n_2})$ επειδή τα δείγματα είναι ανεξάρτητα:

$$\bar{X}-\bar{Y} \sim N\left(\mu_1-\mu_2, \sigma^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)\right)$$

Άρα $\dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} \sim N(0,1)$. Δεν μπορεί να είναι αναστρεπτή, αφού περιέχει το άγνωστο σ. Άρα, αρκεί να απαλοιφεί το άγνωστο σ.

Ισχύουν: $\dfrac{(n_1-1)S_1^2}{\sigma_1^2} \sim X^2_{n_1-1}$ και $\dfrac{(n_2-1)S_2^2}{\sigma_2^2} \sim X^2_{n_2-1}$

Επειδή τ.δ. ανεξάρτητα, τα S_1^2, S_2^2 ανεξ. → $\dfrac{(n_1-1)S_1^2}{\sigma^2} + \dfrac{(n_2-1)S_2^2}{\sigma^2} \sim X^2_{n_1+n_2-2}$

Θεωρώ την ποσότητα:

$$Q = \frac{\dfrac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\cancel{\sigma}\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}}{\sqrt{\dfrac{\frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{\cancel{\sigma^2}}}{n_1+n_2-2}}} \xrightarrow{\text{από άποψη κατανομής}} \equiv \frac{N(0,1)}{\sqrt{\dfrac{X^2_{n_1+n_2-2}}{(n_1+n_2-2)}}} \sim t_{n_1+n_2-2}$$

αλγεβρικά

$$Q = \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}}\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$$

Θεωρούμε την ανισβερεπη:

$$Q = \frac{(\bar{X}-\bar{Y})-(\mu_1-\mu_2)}{\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_p^2}} \sim t_{n_1+n_2-2} \quad , \text{όπου } S_p^2 = \frac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$$

Ακολουθώντας την μεθοδολογία της ανισβερεπης ποσότητας το $100(1-a)\%$ δ.ε για την $\mu_1-\mu_2$ ελαχίστου μήκους είναι:

$$\left((\bar{X}-\bar{Y}) - t_{n_1+n_2-2,\frac{a}{2}}\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_p^2} \ , \ (\bar{X}-\bar{Y}) + t_{n_1+n_2-2,\frac{a}{2}}\sqrt{\left(\frac{1}{n_1}+\frac{1}{n_2}\right)S_p^2} \right)$$

◦ Διάστημα εμπιστοσύνης για το πηλίκο διακυμάνσεων ανεξάρτητων κανονικών πληθυσμών.

Θεωρώ ανεξάρτητα τ.δ. $X_1,\dots,X_{n_1}$ από $N(\mu_1,\sigma_1^2)$ και τ.δ. $Y_1,\dots,Y_{n_2}$ από $N(\mu_2,\sigma_2^2)$ και υποθέτω σ_1^2, σ_2^2 άγνωστες. Ζητώ δ.ε. για την σύγκριση των σ_1^2, σ_2^2 ή δ.ε για το πηλίκο σ_1^2/σ_2^2.

Για την κατασκευή ανισβερεπής τεντρίζουμε: $\frac{(n_1-1)S_1^2}{\sigma_1^2} \sim X^2_{n_1-1}$, $\frac{(n_2-1)S_2^2}{\sigma_2^2} \sim X^2_{n_2-1}$

$\Rightarrow \frac{S_1^2}{\sigma_1^2} \sim X^2_{n_1-1}/(n_1-1)$ <u>ανεξ</u> $\frac{S_2^2}{\sigma_2^2} \sim X^2_{n_2-1}/(n_2-1)$

Διαιρώντας την 2º με την 1º έχουμε:

$$Q = \frac{\sigma_1^2}{\sigma_2^2}\frac{S_2^2}{S_1^2} \sim F_{n_2-1,n_1-1} \quad : \text{ανισβερεπή.}$$

Αφού η Q είναι ανισβερεπη $\exists\ 0<g_1<g_2$ έτσι ώστε: $1-a = P(g_1<Q<g_2) =$
$= P\left(g_1 < \frac{\sigma_1^2}{\sigma_2^2}\frac{S_2^2}{S_1^2} < g_2\right) = P\left(g_1\frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < g_2\frac{S_1^2}{S_2^2}\right)$

Άρα, ένα δ.ε είναι: $\left(g_1\frac{S_1^2}{S_2^2} \ , \ g_2\frac{S_1^2}{S_2^2}\right)$

Κατασκευάζω σε δ.ε ίσων ουρών για q_1, q_2 τέτοια ώστε:

$$P(Q \geq q_2) = \frac{a}{2} \leadsto q_2 = F_{n_2-1, n_1-1, \frac{a}{2}}$$

και

$$P(Q \leq q_1) = \frac{a}{2} \leadsto q_1 = F_{n_2-1, n_1-1, 1-\frac{a}{2}}$$

- **Κατασκευή Αναστρεπτής Ποσότητας**

ΠΡΟΤΑΣΗ 1: (α) Έστω τ.μ. X με συνεχή Α.σ.κ. F_θ, $\theta \in \Theta \subseteq \mathbb{R}$. Τότε η τ.μ. $Y = F_\theta(x) \sim U(0,1)$

(β) Η τ.μ. $Z = -2\log F_\theta(x) \sim \chi^2_2$

(γ) Η τ.μ. $W = -2\log(1 - F_\theta(x)) \sim \chi^2_2$

ΠΡΟΤΑΣΗ 2: Έστω τ.δ. $X_1, \dots, X_n$ από συνεχή κατανομή με ασκ. F_θ, $\theta \in \Theta \subseteq \mathbb{R}$. Τότε

(α) $Q = -2\sum_{i=1}^{n} \log F_\theta(x_i) \sim \chi^2_{2n}$

(β) $Q = -2\sum_{i=1}^{n} \log(1 - F_\theta(x_i)) \sim \chi^2_{2n}$

και επομένως είναι αναστρεπτές.

Απόδειξη: $\rightarrow$ Αλλαγή μεταβλητών με μέθοδο μετασχηματισμού.

Παράδειγμα: Έστω τ.δ. $X_1, \dots, X_n$ από κατανομή $Βήτα(\theta, 1)$, $\theta > 0$ και σ.π.π. $f(x,\theta) = \theta x^{\theta - 1}$, $0 < x < 1$, $\theta > 0$. Να βρεθεί δ.ε ίσων ουρών για την θ με σ.ε $100(1-a)\%$.

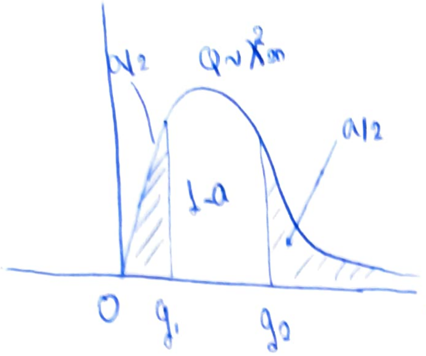
Λύση

$$F_\theta(x) \overset{\text{ορ.}}{=} \int_{-\infty}^{x} f(t,\theta)\,dt = \begin{cases} 0, & x \leq 0 \\ \int_0^x \theta t^{\theta-1}\,dt, & 0 < x < 1 \\ \int_{-\infty}^{x} f(t,\theta)\,dt, & x \geq 1 \end{cases} = \begin{cases} 0, & x \leq 0 \\ x^\theta, & 0 < x < 1 \\ 1, & x \geq 1. \end{cases}$$

Θεωρώ την $Q = -2\sum_{i=1}^{n} \log F_\theta(x_i)$ με κατανομή χ^2_{2n}

$$\Rightarrow Q = -2\sum_{i=1}^{n} \log x_i^\theta \Rightarrow Q = -2\theta \sum_{i=1}^{n} \log x_i \sim \chi^2_{2n}$$

Επειδή Q αναστρεπτή ∃ g_1, g_2 $(0<g_1<g_2<\infty)$.



$$1-a = P(g_1<Q<g_2) = P\left(g_1 < -2\theta\sum_{i=1}^{n}\log x_i < g_2\right) \qquad \frac{0<x_i<1}{\log x_i<0}$$

$$= P\left(-\frac{g_1}{2\sum\log x_i} < \theta < -\frac{g_2}{2\sum \log x_i}\right)$$

Άρα, ένα δ.ε. είναι: $\left(-\frac{g_1}{2\sum_{i=1}^{n}\log x_i}, -\frac{g_2}{2\sum_{i=1}^{n}\log x_i}\right)$

δ.ε ίσων ουρών: Επιλέγω g_1, g_2:

$$P(Q\ge g_2) = \frac{a}{2} \Rightarrow g_2 = \chi^2_{2n,a/2}$$

και

$$P(Q\le g_1) = \frac{a}{2} \Rightarrow g_1 = \chi^2_{2n,1-a/2}$$

Άρα το δ.ε ίσων ουρών είναι: $\left(-\frac{\chi^2_{2n,1-a/2}}{2\sum\log x_i}, -\frac{\chi^2_{2n,a/2}}{2\sum\log x_i}\right)$

Παρατήρηση: Το δ.ε εκφράζεται μέσω του επαρκούς που είναι $\prod_{i=1}^{n} x_i$

Παράδειγμα: Έστω τ.δ. $X_1,\dots,X_n$ από $U(0,\theta)$, $\theta>0$. Να βρεθεί δ.ε ίσων ουρών για την θ.

Απ.

Η ασ.κ $F(x,\theta) = \begin{cases} 0, & x\le 0 \\ \frac{x}{\theta}, & 0<x<\theta \\ 1, & x\ge\theta \end{cases}$

Άρα, θεωρώ $Q = -2\sum_{i=1}^{n}\log F(x_i,\theta) = -2\sum\log\frac{x_i}{\theta} \Rightarrow Q = -2\sum\log x_i + 2n\log\theta \sim \chi^2_{2n}$.

επομένως, είναι αναστρεπτή.

επειδή Q αναστρεπτή ∃ g_1, g_2 $(0<g_1<g_2<+\infty)$ τέτοια ώστε:

$$1-a = P(g_1<Q<g_2) = P\left(g_1 < -2\sum\log x_i + 2n\log\theta < g_2\right) =$$

$$= P\left(e^{\frac{q_1}{2n}} \left(\prod_{i=1}^{n} x_i\right)^{1/n} < \theta < e^{\frac{q_2}{2n}} \left(\prod_{i=1}^{n} x_i\right)^{1/n} \right)$$

Το δ.ε ίσων ουρών $q_1 = \chi^2_{2n,1-a/2}$, $q_2 = \chi^2_{2n,a/2}$

Άρα το ίσων ουρών:

$$\left(e^{\chi^2_{2n,1-a/2}} \left(\prod_{i=1}^{n} x_i\right)^{1/n} , \; e^{\chi^2_{2n,a/2}} \left(\prod_{i=1}^{n} x_i\right)^{1/n} \right)$$

Παρατήρηση: Επαρκές για την θ είναι το $X_{(n)}$

Παράδειγμα: Έστω τ.δ. $X_1, \dots, X_n$ από $U(0,\theta)$, $\theta>0$ και έστω $X_{(n)} = \max\{X_1, \dots, X_n\}$

(α) Νδο η $X_{(n)}/\theta$ είναι αντιστρεπτή.

(β) Να βρεθεί $100(1-a)\%$ δ.ε ίσων ουρών για θ.

(γ) Να βρεθεί $100(1-a)\%$ δ.ε ελαχίστου μήκους για την θ.

Απ

(α) Έστω $Y = X_{(n)}$ και $Q = \frac{X_{(n)}}{\theta} = \frac{Y}{\theta}$

Επειδή $Y = X_{(n)} \xrightarrow{1^{\text{ο}}\ \text{μαθ}} f_Y(y) = n[F_X(y)]^{n-1} f_X(y)$, όπου $F_X(y) = \frac{y}{\theta}$, $0<y<\theta$ και

$f_X(y) = \frac{1}{\theta}$, $0<y<\theta$.

Επομένως, έχουμε ότι: $f_Y(y) = \frac{ny^{n-1}}{\theta^n}$, $0<y<\theta$

Άρα, με μέθοδο μετασχηματισμού, η κατανομή της $Q = h(Y) = \frac{Y}{\theta}$ είναι:

$$f_Q(q) = \begin{cases} nq^{n-1} , & 0<q<1 \\ 0 , & \text{αλλού.} \end{cases}$$

Άρα $Q = \frac{X_{(n)}}{\theta}$ είναι αντιστρεπτή ποσότητα.

(β) Αφού Q αντιστρεπτή ∃ g_1, g_2 ($0<g_1<g_2<1$) τέτοια ώστε:

$$1-\alpha = P(g_1<Q<g_2) = P\left(g_1 < \frac{X_{(n)}}{\theta} < g_2\right) = P\left(\frac{X_{(n)}}{g_2} < \theta < \frac{X_{(n)}}{g_1}\right)$$

Άρα ένα $100(1-\alpha)\%$ δ.ε για το θ είναι $\left(\frac{X_{(n)}}{g_2}, \frac{X_{(n)}}{g_1}\right)$

Το δ.ε ίσων ουρών προκύπτει για g_1, g_2:

$$P(Q \ge g_2) = \frac{\alpha}{2} \Rightarrow \int_{g_2}^{1} f_Q(q)\,dq = \frac{\alpha}{2} \Rightarrow \int_{g_2}^{1} n q^{n-1}\,dq = \frac{\alpha}{2} \Rightarrow q^n\Big|_{g_2}^{1} = \frac{\alpha}{2} \Rightarrow$$

$$\Rightarrow g_2 = \sqrt[n]{1-\alpha/2}$$

$$P(Q \le g_1) = \frac{\alpha}{2} \Rightarrow \int_{0}^{g_1} f_Q(q)\,dq = \frac{\alpha}{2} \Rightarrow \int_{0}^{g_1} n q^{n-1}\,dq = \frac{\alpha}{2} \Rightarrow q^n\Big|_{0}^{g_1} = \frac{\alpha}{2} \Rightarrow$$

$$\Rightarrow g_1 = \sqrt[n]{\alpha/2}$$

Άρα το δ.ε ίσων ουρών είναι: $\left(\frac{X_{(n)}}{\sqrt[n]{1-\alpha/2}}, \frac{X_{(n)}}{\sqrt[n]{\alpha/2}}\right)$ (Τα άκρα του εξαρτώνται από το επαρκές)

(γ) Ζητώ το ελάχιστο μήκος, δηλαδή, ζητώ g_1, g_2 ($0<g_1<g_2<1$) που ελαχιστοποιούν το $\ell = X_{(n)}\left(\frac{1}{g_1} - \frac{1}{g_2}\right)$, ή, το $\ell^* = \frac{1}{g_1} - \frac{1}{g_2}$ υπό την συνθήκη $1-\alpha = P(g_1<Q<g_2)$

Ο περιορισμός: $1-\alpha = P(g_1<Q<g_2) = \int_{g_1}^{g_2} f_Q(q)\,dq = \int_{g_1}^{g_2} n q^{n-1}\,dq \Rightarrow \dots \Rightarrow g_2^n - g_1^n = 1-\alpha$

Ζητώ g_1, g_2 ($0<g_1<g_2<1$) που ελαχιστοποιούν το μήκος $\ell^* = \frac{1}{g_1} - \frac{1}{g_2}$, υπό τον περιορισμό: $g_2^n - g_1^n = 1-\alpha$ (Π)

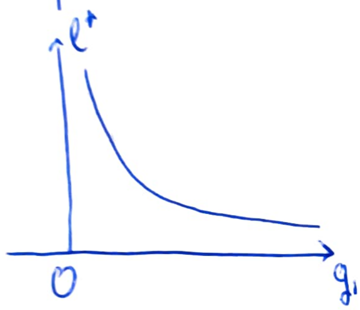
$$\frac{dl^*}{dg_1} = -\frac{1}{g_1^2} - \frac{dg_2}{dg_1}\frac{d}{dg_2}\left(\frac{1}{g_2}\right) = -\frac{1}{g_1^2} + \frac{1}{g_2^2}\frac{dg_2}{dg_1} \quad (1)$$

Από (Π) $\frac{d}{dg_1}\left(g_2^n - g_1^n\right) = \frac{d}{dg_1}(1-a) \Rightarrow \frac{dg_2}{dg_1}\left(\frac{d}{dg_2}(g_2^n)\right) - n g_1^{n-1} = 0$

$$\Rightarrow n g_2^{n-1}\frac{dg_2}{dg_1} - n g_1^{n-1} = 0 \Rightarrow \frac{dg_2}{dg_1} = \frac{g_1^{n-1}}{g_2^{n-1}} \quad (2)$$

Από (1), (2) $\frac{dl^*}{dg_1} = -\frac{1}{g_1^2} + \frac{1}{g_2^2}\left(\frac{g_1}{g_2}\right)^{n-1} < 0$ επειδή $g_1 < g_2$

Άρα, η l^* είναι φθίνουσα ως προς g_1.



άρα η l^* ελαχιστοποιείται στην μεγαλύτερη τιμή που μπορεί να πάρει το g_1.

Από την (Π): $g_2^n - g_1^n = 1-a \Rightarrow g_1^n = g_2^n - 1 + a \xRightarrow{g_2<1}$

$\Rightarrow g_1^n \le 1-1+a = a \Rightarrow \boxed{g_1 \le \sqrt[n]{a}}$

Άρα, η l^* ελαχιστοποιείται για $g_1 = \sqrt[n]{a}$.

Για αυτό το g_1 από την (Π): $g_2^n = 1-a+g_1^n = 1-a+\left(\sqrt[n]{a}\right)^n \Rightarrow \boxed{g_2 = 1}$

Άρα, το Σ.Ε ελαχίστου μήκους είναι το $\left(X(n), \frac{X(n)}{\sqrt[n]{a}}\right)$

το ελ. μήκος βασίζεται στο επαρκής/ΑΣΕΔ ...