Διάλεξη 9η

10/05/2019

Στατιστική Συμπερασματολογία

1

▸ Έστω τ.δ. $X_1,\dots,X_n$ από $N(\mu,\sigma^2)$. Ενδιαφερόμαστε για το δ.ε. για την σ^2, όταν:

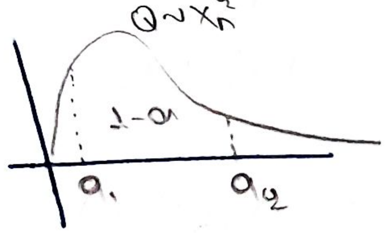
i) μ: γνωστό:

$$X_i \sim N(\mu,\sigma^2) \Rightarrow \frac{X_i-\mu}{\sigma}\sim N(0,1) \Rightarrow \left(\frac{X_i-\mu}{\sigma}\right)^2 \sim N(0,1)^2 \equiv \chi_1^2.$$

$$\xRightarrow[\text{ανεξ.}]{X_i} \sum_{i=1}^n \frac{(X_i-\mu)^2}{\sigma^2} \sim \chi^2_{\sum_{i=1}^n 1} \Rightarrow \frac{\sum (X_i-\mu)^2}{\sigma^2}\sim \chi_n^2,$$ η χ_n^2 είναι ανεξ. της σ^2.

Θεωρώ την αντιστρέψιμη ποσότητα: $Q=\frac{\sum_{i=1}^n (X_i-\mu)^2}{\sigma^2}\sim\chi_n^2$

Αφού η Q: αντιστρέψιμη: $\exists\, q_1,q_2$ $(0<q_1<q_2<\infty)$ ανεξ. της σ^2 (αφού η κατανομή της Q ανεξ.: σ^2) τέτοια ώστε:



$$1-\alpha = P(q_1<Q<q_2) : (*)$$

$$= P\left(q_1 < \frac{\sum(X_i-\mu)^2}{\sigma^2} < q_2\right)$$

$$\Rightarrow 1-\alpha = P\left(\frac{\sum(X_i-\mu)^2}{q_2}<\sigma^2<\frac{\sum(X_i-\mu)^2}{q_1}\right)$$

$\Downarrow$

Ένα $100(1-\alpha)\%$ δ.ε. για την σ^2 είναι: $\left(\frac{\sum(X_i-\mu)^2}{q_2}, \frac{\sum(X_i-\mu)^2}{q_1}\right)$

Προσοχή: Εύρεση δ.ε. ελαχ. μήκους.

$$l = \sum(X_i-\mu)^2\left(\frac{1}{q_1}-\frac{1}{q_2}\right)$$

Ζητώ q_1,q_2 που ελαχιστοποιούν το l ή ζητώ q_1,q_2 που ελαχισ. το $l^* = \frac{1}{q_1}-\frac{1}{q_2}$, ως προς q_1,q_2 υπό την προϋπόθεση (*)

$$F_{\chi_n^2}(q_2) - F_{\chi_n^2}(q_1) = 1-\alpha : (**)$$

$$\frac{d\ell^*}{dq_1} = -\frac{1}{q_1^2} - \frac{dq_2}{dq_1}\left(\frac{d}{dq_2}\,\frac{1}{q_2}\right)$$

$$= -\frac{1}{q_1^2} + \frac{1}{q_1^2}\,\frac{dq_2}{dq_1}$$

Από (**) $\dfrac{d}{dq_1} F_{X_n^2}(q_2) - \dfrac{d}{dq_1} F_{X_n^2}(q_1) = 0$

$$\Rightarrow \frac{dq_2}{dq_1}\,\frac{dF_{X_n^2}(q_2)}{dq_2} - f_{X_n^2}(q_1) = 0$$

$$\Rightarrow \frac{dq_2}{dq_1} f_{X_n^2}(q_2) - f_{X_n^2}(q_1) = 0 \Rightarrow \frac{dq_2}{dq_1} = \frac{f_{X_n^2}(q_1)}{f_{X_n^2}(q_2)}$$

Επομένως $\dfrac{d\ell^*}{dq_1} = -\dfrac{1}{q_1^2} + \dfrac{1}{q_2^2}\,\dfrac{f_{X_n^2}(q_1)}{f_{X_n^2}(q_2)}$

και $\dfrac{d\ell^*}{dq_1} = 0 \Rightarrow q_1^2 f_{X_n^2}(q_1) = q_2^2 f_{X_n^2}(q_2)$: (1)

Αλλά $f_{X_n^2}(x) = \dfrac{1}{\Gamma(n/2)\,2^{n/2}}\, x^{n/2-1} \cdot e^{-x/2}$: (2)

Από τις (1) κ' (2) έχω: $q_1^{n/2+1} e^{-q_1/2} = q_2^{n/2+1} e^{-q_2/2}$ (3)

Έχω ότι $\int_{q_1}^{q_2} f_{X_n^2}(x)dx = 1-\alpha.$ m

$$\int_{q_1}^{q_2} \frac{1}{\Gamma(n/2)2^{n/2}} x^{\frac{n}{2}-1} e^{-x/2}\, dx = 1-\alpha. \quad :(4)$$

Τα q_1, q_2 που ελαχ το ℓ^* προκύπτουν από την λύση των (3) κ' (4)

Το σύστημα δεν λύνεται αναλυτικά αλλά αριθμητικά

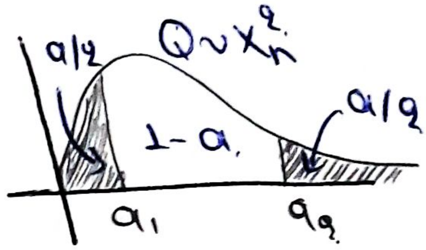
Διαστήματα ίσων ουρών:

(όταν δεν μπορεί να βρεθεί αναλυτικά το ελαχίστου μήκους)

Δ.Ε ίσων ουρών είναι εκείνο που αποκόπτει ίσες ουρές από την κατανομή της Q.

Δηλαδή είναι εκείνο που προκύπτει για q_1, q_2 τ.ω

$$P(Q>q_2)=\frac{a}{2} \quad \text{και} \quad P(Q<q_1)=\frac{a}{2}.$$



Από ορισμό αντίστροφων εκατοστημορίων της χ^2_n:

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

Από άσκηση προκύπτει $q_2 = \chi^2_{n,a/2}$:

$$P(Q<q_1)=a/2 \Rightarrow 1-P(Q>q_1)=a/2 \Rightarrow P(Q>q_1)=1-a/2$$
$$\Rightarrow q_1 = \chi^2_{n,1-\frac{a}{2}}$$

Το $100(1-a)\%$ δ.ε ίσων ουρών για την σ^2, όταν μ: γνωστό είναι:

$$\left(\frac{\Sigma(x_i-\mu)^2}{\chi^2_{n,a/2}}, \frac{\Sigma(x_i-\mu)^2}{\chi^2_{n,1-a/2}}\right)$$

Διάστημα εμπ. για την σ: Καταλήγουμε ότι:

$$1-a = P\left(\frac{\Sigma(x_i-\mu)^2}{\chi^2_{n,a/2}} < \sigma^2 < \frac{\Sigma(x_i-\mu)^2}{\chi^2_{n,1-a/2}}\right)$$

$$= P\left(\sqrt{\frac{\Sigma(x_i-\mu)^2}{\chi^2_{n,a/2}}} < \sigma < \sqrt{\frac{\Sigma(x_i-\mu)^2}{\chi^2_{n,1-a/2}}}\right) : \text{το δ.ε για } \sigma$$

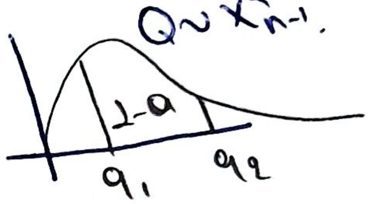
▸ Έστω τ.δ. $X_1, \dots, X_n$ από $N(\mu, \sigma^2)$: δ.ε. για την σ^2 όταν:

(II) μ: άγνωστο:

Θεωρώ την $Q = \frac{(n-1)S^2}{\sigma^2} \sim X^2_{n-1}$.

Επειδή η Q είναι συνάρτηση του τ.δ. της άγνωστης παραμέτρου που με ενδιαφέρει και η κατανομή της είναι ανεξ. της σ^2.

Η Q είναι αντιστρέψιμη.



Άρα $\exists\, q_1, q_2$ $(0 < q_1 < q_2 < \infty)$ έτσι ώστε

$$1-a = P(q_1 < Q < q_2) = P\left(q_1 < \frac{(n-1)S^2}{\sigma^2} < q_2\right)$$
$$= P\left(\frac{(n-1)S^2}{q_2} < \sigma^2 < \frac{(n-1)S^2}{q_1}\right)$$

Ένα $100(1-a)\%$ δ.ε. της σ^2 είναι το $\left(\frac{(n-1)S^2}{q_2}, \frac{(n-1)S^2}{q_1}\right)$.

Ακολουθώντας την περίπτωση μ: γνωστό, δεν μπορεί να βρεθεί αναλυτικά δ.ε. ελαχίστου μήκους.

Δ.Ε. ίσων ουρών: Επιλέγω q_1, q_2 έτσι ώστε:

$$P(X^2_{n-1} > q_2) = \frac{a}{2} \quad \text{κ'} \quad P(X^2_{n-1} < q_1) = \frac{a}{2}.$$

$$q_2 = X^2_{n-1, a/2} \qquad q_1 = X^2_{n-1, 1-a/2}.$$

Έτσι το 100(1-α)% δ.ε. ίσων ουρών για σ^2 είναι:

$$\left(\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}, \frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}\right)$$

και το 100(1-α)% δ.ε. ίσων ουρών για την σ

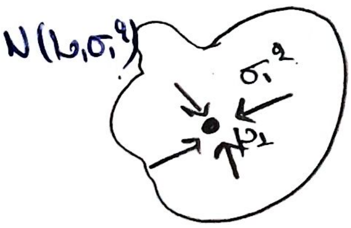
είναι: $$\left(\sqrt{\frac{(n-1)s^2}{\chi^2_{n-1,\alpha/2}}}, \sqrt{\frac{(n-1)s^2}{\chi^2_{n-1,1-\alpha/2}}}\right)$$

Δύο κανονικοί πληθυσμοί (ανεξάρτητοι)
δ.ε. για την διαφορά των μέσων τιμών:

(I) Διακυμάνσεις γνωστές:

Έστω 2 ανεξάρτητοι κανονικοί πληθ. $N(\mu_1,\sigma_1^2)$, $N(\mu_2,\sigma_2^2)$
Έστω 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 τα τ.δ. είναι ανεξάρτητα.

Πληθ. 1: φοιτητές



Έστω ενδιαφέρομαι ένα κοινό χαρακτηριστικό τους μία κοινή τ.μ.

Πληθ. 2: φοιτήτριες

$N(\mu_2,\sigma_2^2)$

Με ενδιαφέρει να έχω άποψη για την διαφορά $\mu_1-\mu_2$ γιατί αυτή με πληροφορεί για την συμπεριφορά των 2 πληθυσμών ως προς το κοινό τους χαρακτηριστικό.

Εκτιμητής της $\mu_1-\mu_2$: $\bar{X}-\bar{Y}$

$$\bar{X} = \frac{1}{n_1}\sum X_i, \qquad \bar{Y} = \frac{1}{n_2}\sum Y_i$$

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

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

$$\Rightarrow Z = \frac{\bar{X}-\bar{Y}-(\mu_1-\mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}} \sim N(0,1).$$

Η Z αντιστρεπτή. Άρα $\exists\, q_1, q_2$ $(-\infty < q_1 < q_2 < +\infty)$ έτσι ώστε:

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

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

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

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

Το δ.ε ελάχ. μήκους προκύπτει

για $q_2 = z_{\alpha/2}$, $q_1 = -z_{\alpha/2}$

II) Διακυμάνσεις άγνωστες αλλά ίσες:

Έστω 2 ανεξ. πληθυσμοί κανονικοί, $N(\mu_1,\sigma_1^2)$ και $N(\mu_2,\sigma_2^2)$
$\sigma_1^2=\sigma_2^2=\sigma^2$

Έστω τ.δ. $X_1,\dots,X_n$ από $N(\mu_1,\sigma^2)$ και τ.δ. $Y_1,\dots,Y_n$ από $N(\mu_2,\sigma^2)$ και είναι ανεξάρτητα.

Όπως προηγουμένως:

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

Έστω $Y_1=\frac{(n_1-1)S_1^2}{\sigma^2}\sim X^2_{n_1-1}$ και $Y_2=\frac{(n_2-1)S_2^2}{\sigma^2}\sim X^2_{n_2-1}$

και $S_1^2=\frac{1}{n_1-1}\Sigma(X_i-\bar{X})^2$ και $S_2^2=\frac{1}{n_2-1}\Sigma(Y_i-\bar{Y})^2$

Έστω $Y=Y_1+Y_2=\frac{(n_1-1)S_1^2}{\sigma_1^2}+\frac{(n_2-1)S_2^2}{\sigma_2^2}$ ανεξ. $X^2_{n_1-1}+X^2_{n_1-1}\equiv X^2_{n_1+n_2-2}$

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

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

Άρα $$Q=\frac{Z}{\sqrt{Y/(n_1+n_2-2)}}=\frac{N(0,1)}{\sqrt{\frac{X^2_{n_1+n_2-2}}{(n_1+n_2-2)}}}\overset{\text{ανεξ}}{\sim} t_{n_1+n_2-2}$$

Έτσι Q: αντιστρέψιμη.

Ακολουθώντας την ίδια διαδικασία με το δ.ε. για την μ ενός $N(\mu, \sigma^2)$ όταν σ^2: άγνωστο, το $100(1-a)\%$ δ.ε για $\mu_1 - \mu_2$ (λαχ. μήκους είναι:

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

όπου $$S_p^2 = \frac{(n_1-1) S_1^2 + (n_2-1) S_2^2}{n_1+n_2-2}$$

Διαστήματα εμπιστοσύνης για σ_1^2/σ_2^2:

$$\frac{(n_1-1) S_1^2}{\sigma_1^2} \sim X^2_{n_1-1} \, , \quad \frac{(n_2-1) S_2^2}{\sigma_2^2} \sim X^2_{n_2-1}$$

Επειδή τα τ.δ. είναι ανεξ. το S_1^2, ανεξ το S_2^2.

Θεωρώ $$F = \frac{\dfrac{(n_2-1) S_2^2}{\sigma_2^2} \Big/ (n_2-1)}{\dfrac{(n_1-1) S_1^2}{\sigma_1^2} \Big/ (n_1-1)} \underset{\text{ανεξ.}}{\sim} \frac{X^2_{n_2-1} / (n_2-1)}{X^2_{n_1-1} / (n_1-1)}$$

$\Rightarrow F \sim F_{n_2-1, n_1-1}$.

$$F = \frac{\sigma_1^2}{\sigma_2^2} \cdot \frac{S_2^2}{S_1^2}$$

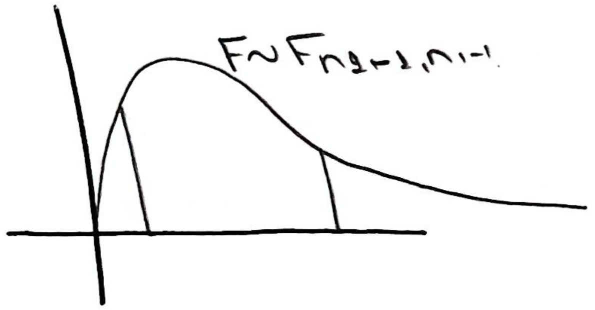
Συνθέτοντας αν $F=\frac{\sigma_1^2}{\sigma_2^2}\,\frac{S_2^2}{S_1^2} \sim F_{n_2-1,n_1-1}$ αντίστρεπτα.

Άρα $\exists\, q_1, q_2\ (0<q_1<q_2<\infty)$ τέτοια ώστε:

$$1-a = P(q_1 < F < q_2) = P\left(q_1 < \frac{\sigma_1^2}{\sigma_2^2}\,\frac{S_2^2}{S_1^2} < q_2\right)$$

$$= P\left(q_1 \frac{S_1^2}{S_2^2} < \frac{\sigma_1^2}{\sigma_2^2} < q_2 \frac{S_1^2}{S_2^2}\right)$$

Ένα $100(1-a)\%$ δ.ε. για $\frac{\sigma_1^2}{\sigma_2^2}$ είναι $\left(q_1\frac{S_1^2}{S_2^2},\ q_2\frac{S_1^2}{S_2^2}\right)$



Δεν υπάρχει αναλυτικό δ.ε. ελάχιστου μήκους.
Το δ.ε. ίσων ουρών προκύπτει για:

$q_1 = F_{n_2-1,\,n_1-1,\,1-\frac{a}{2}}$

$q_2 = F_{n_2-1,\,n_1-1,\,a/2}$

Κατασκευή αντιστρεπτής ποσότητας:

SOS!!!

① **Πρόταση:** (α) Έστω τ.μ. X με α.σ.κ. F_θ, $\theta \in \Theta$

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. $W = -2\log(1 - F_\theta(X)) \sim \chi^2_2$

② **Πρόταση:** Έστω τ.δ. $X_1, \dots, X_n$ από συνεχή α.σ.κ. $F_\theta(x)$. 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 Q και Q^* είναι αντιστρεπτές.

Αποδ

α) Λόγω του (β) της πρότασης ①:

$-2\log F_\theta(X_i) \sim \chi^2_2, \ \forall i = 1, \dots, n$

$$Q = -2\sum_{i=1}^{n} \log F_\theta(X_i) = -2\log F_\theta(X_1) - \dots - 2\log F_\theta(X_n) \sim$$

$$\sim \underbrace{\chi^2_2 + \chi^2_2 + \dots + \chi^2_2}_{\text{ανεξ.}}$$

$$\simeq \chi^2_{2n}$$

Παραδ: Έστω τ.δ. $X_1,\dots,X_n$ από πληθυσμό με $f(x,\theta)=\theta x^{\theta-1}$, $0<x<1$, $\theta>0$. (Αυτή είναι η Βήτα$(\theta,1)$) Να κατασκευάσετε δ.ε. ίσων ουρών για την θ.

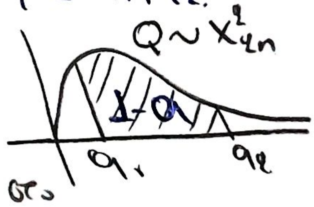
Λύση

Θα βρω την α.σ.κ. $F(x,\theta) \overset{\text{ορ}}{=} P(X\le x)=\int_{-\infty}^{x} f(t,\theta)\,dt$

$$= \begin{cases} 0, & x<0 \\ \int_0^x \theta t^{\theta-1}\,dt = x^\theta, & 0<x<1 \\ 1, & x>1\end{cases}$$

Άρα $F(x,\theta)=x^\theta$, $0<x<1$. Θεωρώ την αντίστροφη ποσότητα $Q=-2\sum_{i=1}^{n} \log F(X_i,\theta)\sim \chi^2_{2n} = -2\sum_{i=1}^{n} \log X_i^\theta$

$\Rightarrow Q=-2\theta\sum_{i=1}^{n} \log X_i \sim \chi^2_{2n}$. Αφού Q αντίστροφη, $\exists\, q_1,q_2$ $(0<q_1<q_2<\infty)$ έτσι ώστε: $1-a=P(q_1<Q<q_2)$



$\Rightarrow 1-a = P(q_1<-2\theta\sum\log X_i<q_2)$ (*)

(*) X_i παίρνει τιμές στο $(0,1)$ οπότε το $\log X_i<0$. Έτσι $(-2)<0$ και $\log X_i<0$ είναι >0

$=P\left(\frac{q_1}{-2\sum\log X_i}<\theta<\frac{q_2}{-2\sum\log X_i}\right)$

Άρα ένα δ.ε. για την θ με β.ε. $(1-a)$ ή $100(1-a)\%$ είναι το $\left(-\frac{q_1}{2\sum\log X_i}, -\frac{q_2}{2\sum\log X_i}\right)$.

Δεν μπορώ να βρω δ.ε. ελαχ. μήκους αναλυτικά.

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: $f(\underline{x},\theta) = \Pi \theta x_i^{\theta-1} = \theta^n (\Pi x_i)^{\theta-1} \Rightarrow \Pi x_i$ επαρκές

$2\Sigma \log x_i = 2 \log \Pi x_i$.

Άρα το δ.ε. βασίζεται στο επαρκές.