Válasz a 2006-os előadások után feltett néhány kérdésre

32. Egy valószínűségi változó relatív szórása (a szórásnégyzet négyzetgyöke osztva a várható értékkel) 10 %. Milyen pontossággal van értelme megadni 850 mérés átlagát?

A minta középértéke: $T(x) = \dfrac{\sum_{i=1}^{n} x_i}{n}$

Jelölés: $\bar{x}$ – eloszlása általában nem ismert!

ha $x_i \sim N(\mu, \sigma^2)$, akkor $\bar{x} \sim N\left(\mu, \dfrac{\sigma^2}{n}\right)$

$M(\bar{x}) = M(x) = \mu$ μ ***torzítatlan becslése***

$D^2(\bar{x}) = \dfrac{\sigma^2}{n}$ ezért a ***minta*** középértéke

N.B. n növelésével csak $\dfrac{1}{\sqrt{n}}$-szeresére csökken a szórás!

$$\frac{\sqrt{\sigma^2(x)}}{M(x)} = \frac{1}{10} \quad , \quad \sqrt{D^2(\bar{x})} = \sqrt{\frac{\sigma^2}{850}} = \frac{\sqrt{\sigma^2}}{29{,}15} \qquad \frac{\sqrt{\sigma^2}}{29{,}15} \cdot \frac{1}{10} = \frac{\sqrt{\sigma^2}}{291{,}5}$$

Ebből az következik, hogy míg egyetlen x érték 10 % relatív pontossággal mérhető, addig a 850 mérés átlaga a várható érték kb. 1/300-ad részének megfelelő szórású. azaz kb. 3.3 % relatív pontosságú. Az x mérésekor tehát legfeljebb 1 számjegy a pontos, a második már csak közelítő, addig 850 mérés átlagával a pontosság kb. 30-szoros lesz, ezért a második számjegy pontos, a harmadik pedig jó közelítéssel pontos. Így 3 számjegyre mindenképpen érdemes megadni az eredményt, de 4-nél többre semmiképpen nem. **A jó válasz tehát 4 számjegy.**

33. Írja fel F(T1, T2, ..., Tr) függvény tapasztalati szórásnégyzetét a T1, T2, ..., Tr becslőfüggvények tapasztalati szórásnégyzetének ismeretében

Függvények várható értékének és szórásának becslése:

① $\vartheta_i^* = t_i$ a T_i statisztikák realizációja

② ϕ $(\theta_1, \theta_2, ...\theta_r)$ becslése $\phi(t_1, t_2, ...t_r)$

③ D^2 (T_i) becslése S^2 (T_i)

C (T_i, T_j) becslése $\hat{C}$ (T_i, T_j)

④ ***$D^2(\phi)$* becslése:**

$$\boxed{S^2[\phi(T_1, T_2 \ldots T_r)] = \sum_{i=1}^{r} \left(\frac{\partial \phi}{\partial \vartheta_i}\right)_{t_i}^2 \cdot S^2(T_i) + 2\sum_{j=1}^{r} \sum_{i<j} \left(\frac{\partial \phi}{\partial \vartheta_i}\right)_{t_i} \left(\frac{\partial \phi}{\partial \vartheta_j}\right)_{t_j} \hat{C}(T_i, T_j)}$$

35. Mit nevezünk szignifikancia-valószínűségnek?

H_0-t *megvédjük* vagy *elvetjük* α szignifikanciaszinten (α: 0,1; 0,05; 0,01)

A döntés alapja az

$$\alpha = P(T \geq c \,|\, H_0) \quad \text{vagy} \quad \alpha = P(|T| \geq c \,|\, H_0)$$

Elvetjük H_0-t, ha $\quad$ $\boldsymbol{t \geq c}$ $\quad$ **vagy** $\quad$ $\boldsymbol{|t| \geq 0}$

c : kritikus érték

Honnan tudjuk 1. t értékét?

2. Az α valószínűséget?

Válasz: Ismerjük (vagy azt hisszük, hogy ismerjük!!) T eloszlását,

és abból kiszámíthatjuk fix α-hoz a c-t.

Ezt a c-t hasonlítjuk a mintából számított t realizációhoz.

Ha $t \cong$ c vagy $|t| \cong 0$, érdemes új mintát (több adatot) vizsgálni.

Ha $t <<$ c vagy $|t| << 0$, lehet kisebb az α.

Kiszámítható közvetlenül az $\alpha = P(T \geq c \,|\, H_0)$ **szignifikancia-valószínűség** is!

36. Mi az egyoldali és mi a kétoldali nullhipotézis, és melyiket mikor tartjuk meg, illetve vetjük el?

Példák: $\quad H_0: \; \vartheta = \vartheta_0 \quad \begin{cases} H_1: & \vartheta \neq \vartheta_0 & |\text{ kétoldali} \\ H_1': & \vartheta < \vartheta_0 & |\text{ egyoldali} \\ H_1'': & \vartheta > \vartheta_0 & |\text{ egyoldali} \end{cases}$

Lehet: H_0: $\theta \leq \theta_0$ vagy $\theta \geq \theta_0$ is.

H_0-t megvédjük* vagy *elvetjük α szignifikanciaszinten (α: 0,1; 0,05; 0,01)

A döntés alapja az

$$\alpha = P(T \geq c \,|\, H_0) \quad \text{vagy} \quad \alpha = P(|T| \geq c \,|\, H_0)$$

Elvetjük H_0-t, ha $\quad$ $\boldsymbol{t \geq c}$ $\quad$ **vagy** $\quad$ $\boldsymbol{|t| \geq 0}$

c : kritikus érték

39. Hogyan adjunk meg egy minta (méréssorozat) alapján becsült eredményt?

***n* mérésből átlag számítása esetén:**

HOGYAN ADJUNK MEG EGY BECSÜLT EREDMÉNYT?

① n darab mérés $\rightarrow \quad \bar{x}, s^2(x)$

$\bar{x}$ a μ paraméter becsült értéke, $s^2(x)$ a σ^2 paraméter becsült értéke

$s(\bar{x}) = \dfrac{s(x)}{\sqrt{n}}$ a becsült $\bar{x}$ szórása μ körül. (Innen marad benne az $\dfrac{1}{\sqrt{n}}$.)

Mi az $s(\bar{x})$ információtartalma?

Legyen $T = \dfrac{\bar{x}-\mu}{s/\sqrt{n}}$; $T \sim t_{n-1}$

Ekkor: $P\left[t_{n-1}(\alpha/2) \le T \le t_{n-1}(1-\alpha/2)\right] = 1-\alpha$

ÍGY adjuk meg: Az $1-\alpha$ megbízhatósági szintű konfidencia-intervallum:

$$\bar{x} \pm t_{n-1}\left(1-\frac{\alpha}{2}\right)\cdot\frac{s}{\sqrt{n}}$$

$$\boxed{\bar{x} \pm s\cdot\frac{t_{n-1}\left(1-\frac{\alpha}{2}\right)}{\sqrt{n}}}$$

Függvény paramétereinek becslése esetén:

HOGYAN ADJUK MEG A BECSÜLT PARAMÉTEREKET??

n db mérés $\longrightarrow$ r paraméterre: ϑ_i, $s^2(\vartheta_i)$

Mi az $s^2(\vartheta_i)$ információtartalma??

Legyen: $T_i = \dfrac{\hat{\vartheta}_i - \vartheta_i}{S(\vartheta_i)}$; $T \sim t_{n-r}$ (A T_i $n-r$ szabadsági fokú Student eloszlású.)

újfent: $P\left[t_{n-r}(\alpha/2) \le T \le t_{n-r}(1-\alpha/2)\right] = 1-\alpha$

ÍGY adjuk meg: Ennek alapján az $1-\alpha$ megbízhatósági szintű konfidencia-intervallum:

$$\boxed{\hat{\vartheta}_i \pm t_{n-r}(1-\alpha/2)\cdot S(\vartheta_i)}$$

40. Mi a feltételi egyenlete egy Y = f(X, p) függvény p paramétervektor-elemei legkisebb négyzetes becslésének az {xi , yi} , i = 1,......., n rendezett minta alapján?

A leggyakrabban használt becslő módszer: legkisebb négyzetes

$$Q = \sum_{i=1}^{n} w_i\left(Y_i - f(x_i)\right)^2 \quad \textbf{legyen \textit{minimális}}$$

w_i súlyok számítása – hogy az $f(x)$ paramétereire MVU-becslést kapjunk.

feltétel: $M\left(\hat{\vartheta}\right) = \vartheta$

és $D^2\left(\hat{\vartheta}\right)$ minimális