HG
April 09

Oversikt over konfidensintervall i Econ 2130

(1) La $X_1, X_2, \ldots, X_n$ være uavhengige og identisk fordelte stokastiske variable med $E(X_i) = \mu$ og $\text{var}(X_i) = \sigma^2$

Aktuelle estimatorer: $\hat{\mu} = \bar{X}$ og $\hat{\sigma}^2 = S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.

La $\alpha/2$ -kvantilen i $N(0,1)$ - fordelingen betegnes med $z_{\alpha/2}$ (slik at $P(-z_{\alpha/2} \le Z \le z_{\alpha/2}) = 1-\alpha$, der $Z \sim N(0,1)$)

La $\alpha/2$ -kvantilen i t_{n-1} - fordelingen betegnes med $t_{n-1,\alpha/2}$ (slik at $P(-t_{n-1,\alpha/2} \le T \le t_{n-1,\alpha/2}) = 1-\alpha$, der $T \sim t_{n-1}$)

Tabell 1 Konfidensintervall for μ

Situasjon	*Forutsetninger (modell)*	*n*	σ	*Standardfeil* $\sqrt{\text{var}(\hat{\mu})}$	*Estimert standardfeil*	*Pivotal*	$1-\alpha$ *KI for*	*Konfidens-grad*
1	(1) sammen med: $X_i \sim N(\mu, \sigma), \quad i = 1,2,\ldots,n$	Vilkårlig	Kjent	$\frac{\sigma}{\sqrt{n}}$	$\frac{\sigma}{\sqrt{n}}$	$\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$	$\bar{X} \pm z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$	Eksakt $1-\alpha$
2	(1) sammen med: $X_i \sim N(\mu, \sigma), \quad i = 1,2,\ldots,n$	Vilkårlig	Ukjent	$\frac{\sigma}{\sqrt{n}}$	$\frac{S}{\sqrt{n}}$	$\frac{\bar{X}-\mu}{S/\sqrt{n}} \sim t_{n-1}$	$\bar{X} \pm t_{n-1,\alpha/2}\frac{S}{\sqrt{n}}$	Eksakt $1-\alpha$
3	Bare (1) der X_i er vilkårlig fordelt	n "stor", $n \ge 30$ (til nød ≥ 20)	Ukjent	$\frac{\sigma}{\sqrt{n}}$	$\frac{S}{\sqrt{n}}$	$\frac{\bar{X}-\mu}{S/\sqrt{n}} \overset{\text{tilnærmet}}{\sim} N(0,1)$	$\bar{X} \pm z_{\alpha/2}\frac{S}{\sqrt{n}}$	Tilnærmet $1-\alpha$
4	Bare (1) der X_i er vilkårlig fordelt	n liten	Ukjent				**Ikke pensum**	

Merknad 1. Uttrykket "**pivotal**" betegner en stokastisk variabel som avhenger av ukjente parametre i modellen - en variabel som

derfor ikke er observerbar - men som har **kjent** sannsynlighetsfordeling. Pivotaler er bl.a nyttige ved konstruksjon av konfidensintervaller og tester.

Merknad 2. Siden t_{n-1}-fordelingen er tilnærmet lik $N(0,1)$ for $n \geq 30$, vil forskjellen mellom KI-ene i situasjon 2 og 3 være neglisjerbar når $n \geq 30$.

Tabell 2. Konfidensintervall for σ^2 når $X_1, X_2, \ldots, X_n$ er uavhengige og normalfordelte med $X_i \sim N(\mu, \sigma)$.

(Hvis en stokastisk variabel, *V*, er kji-kvadratfordelt (avsnitt 5.9.1) med *k* frihetsgrader, skriver vi kort: $V \sim \chi_k^2$.
p-kvantilen i denne fordelingen kaller Løvås, χ_p, som er det tallet som oppfyller $P(V > \chi_p) = p$. Noen kvantiler finnes i tabell D6.)

Modell	*n*	*Estimator*	*Pivotal*	*Nedre konfidensgrense*	*Øvre konfidensgrense*	*Konfidensgrad*
$X_1, X_2, \ldots, X_n$ er uavhengige og identisk fordelte med $X_i \sim N(\mu, \sigma)$	Vilkårlig	$\hat{\sigma}^2 = S^2$	$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$ (Regel 5.22)	$\frac{(n-1)S^2}{\chi_{\alpha/2}}$	$\frac{(n-1)S^2}{\chi_{1-\alpha/2}}$	Eksakt $1-\alpha$

Merknad 3. Har vi funnet et KI for populasjonsvariansen, σ^2, kan vi lett finne et for standardavviket, σ, også. Hvis $[A, B]$ er et $1-\alpha$ KI for σ^2 (slik at $P(A \leq \sigma^2 \leq B) = 1-\alpha$), så vil et $1-\alpha$ KI for σ rett og slett være gitt ved $[\sqrt{A}, \sqrt{B}]$. Dette skyldes at begivenhetene $(A \leq \sigma^2 \leq B)$ og $(\sqrt{A} \leq \sigma \leq \sqrt{B})$ er logisk ekvivalente (og derfor like sannsynlige) siden funksjonen $y = \sqrt{x}$ er en voksende funksjon av *x*.

Øvelse. Vis at konfidensintervallet for σ^2 følger av pivotal-utsagnet i tabellen (dvs regel 5.22). **Hint:** Sett $V = (n-1)S^2/\sigma^2$. Forklar hvorfor $P(\chi_{1-\alpha/2} \leq V \leq \chi_{\alpha/2}) = 1-\alpha$, sett inn for *V*,…. osv.)

Tabell 3 Tilnærmet konfidensintervall basert på regel 5.20 (normaltilnærming for binomisk, hypergeometrisk og poisson fordeling)

Modell	*Estimator*	*Standardfeil*	*Estimert standardfeil*	*Betingelse for normaltil-nærmelse*	*Pivotal*	*Konfidensintervall*
$X \sim \text{bin}(n, p)$	$\hat{p} = \frac{X}{n}$	$\sqrt{\frac{p(1-p)}{n}}$	$\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$	$\text{var}(X) \geq 5$ $(np(1-p) \geq 5$	$\frac{\hat{p}-p}{\sqrt{\hat{p}(1-\hat{p})/n}} \overset{\text{tilnærmet}}{\sim} N(0,1)$	$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$
$X \sim$ hypergeom. (n, M, N) $(p = M/N)$	$\hat{p} = \frac{X}{n}$	$\sqrt{\frac{p(1-p)}{n} \cdot \frac{N-n}{N-1}}$	$\sqrt{\frac{\hat{p}(1-\hat{p})}{n} \cdot \frac{N-n}{N-1}}$	$\text{var}(X) \geq 5$	$\frac{\hat{p}-p}{\sqrt{\frac{\hat{p}(1-\hat{p})}{n} \cdot \frac{N-n}{N-1}}} \overset{\text{tilnærmet}}{\sim} N(0,1)$	$\hat{p} \pm z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n} \cdot \frac{N-n}{N-1}}$
$X \sim \text{pois}(t\lambda)$	$\hat{\lambda} = \frac{X}{t}$	$\sqrt{\frac{\lambda}{t}}$	$\sqrt{\frac{\hat{\lambda}}{t}}$	$\text{var}(X) \geq 5$ $(t\lambda \geq 5)$	$\frac{\hat{\lambda}-\lambda}{\sqrt{\hat{\lambda}/t}} \overset{\text{tilnærmet}}{\sim} N(0,1)$	$\hat{\lambda} \pm z_{\alpha/2}\sqrt{\frac{\hat{\lambda}}{t}}$

Merknad 4 Merk at de tre KI-ene i tabell 3 samt KI-ene i situasjon 1 og 3 i tabell 1 alle har den generelle formen angitt i regel 6.7 der SE står for standardfeil eller estimert standardfeil dersom $\text{SE}(\hat{\theta})$ avhenger av ukjente parametre. Unntak fra regel 6.7 er gitt i tabell 2 og situasjon 2 under tabell 1. Argumentasjonen fra pivotal-utsagnet til konfidensintervallet er gitt i avsnitt 6.3.1. rett etter regel 6.7.

Merknad 5 I mange KI (jfr tabell 1 og 3) bruker vi altså den estimerte versjonen av standardfeilen (i tilfelle standardfeilen er ukjent) når vi utleder et KI. Det er ikke på noen måte opplagt at vi har lov til dette. Det er rimelig å tenke seg at en slik fremgangsmåte ville kunne ødelegge tilnærmelsen til N(0, 1), noe som ville gjøre konfidensgraden tvilsom. Det at vi faktisk ”har lov til” å erstatte SE med en estimert versjon uten å berøre konfidensgraden vesentlig, bygger på teoremer i sannsynlighetsregning som det blir for teknisk å ta opp i dette kurset. Noe teori om dette (som er mye brukt i økonometri) blir tatt opp i STAT 2.