

F9, Konfidensintervall (en population)

Christian Tallberg

Statistiska institutionen

Stockholms universitet

Med känd standardavvikelse σ

Exempel:

Låt $X_1, \dots, X_n$ vara ett stickprov ($n = 16$) av oberoende normalfördelade variabler med okänt medelvärde μ och känd standardavvikelse $\sigma = 12$. Punktskattningen är $\bar{x} = 62$. Ett 95 %-igt k.i. för μ ges av

$$\begin{aligned}\bar{x} \pm 1.96 \frac{\sigma}{\text{sqrt}(n)} &= 62 \pm 1.96 \frac{12}{4} \\ &= 62 \pm 5.88.\end{aligned}$$

Tolkning: Med konfidensgraden 95 % täcker intervallet (56.12; 67.88) populationsparametern μ .

Alternativa konfidensgrader:

konfg.	$z_{\alpha/2}$ -värde
90 %	1.64
95 %	1.96
99 %	2.58
99.9 %	3.29

där $z_{\alpha/2}$ är ett värde (sannolikhetstal) från $N(0; 1)$ som bestämmer konfidensgraden och α är signifikansnivån (risknivån). Notera att $\alpha = 1 - \text{konfidensgraden}$.

Generellt skrivs ett k.i. med konf.grad $(1 - \alpha) 100$ % som

$$\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

vilket betyder att intervallets längd är

$$2 \cdot z_{\alpha/2} \frac{\sigma}{\sqrt{n}}.$$

OBS!:

Om skattningen $\hat{\theta}$ är normalfördelad (approximativt normalfördelad) gäller att ett k.i. med konfidensgrad $(1 - \alpha) 100 \%$ för populationsparametern θ ges av $\hat{\theta} \pm \text{sannolikhetstalet (t ex } z_{\alpha/2}) \cdot \text{standardavvikelsen för skattningen}$ (eller skattade standardavvikelsen för skattningen).

Intervallets längd bestäms av

- standardavvikelsen σ
- konfidensgraden, dvs värdet på $z_{\alpha/2}$
- stickprovets storlek n

Forts. exemplet: Ett 90 % igt k.i. för μ ges av

$$\begin{aligned}\bar{x} \pm 1.64 \frac{\sigma}{\sqrt{n}} &= 62 \pm 1.64 \frac{12}{4} \\ &= 62 \pm 4.92 \\ &\quad (57.08; 66.92)\end{aligned}$$

Ett 99 % igt k.i. för μ ges av

$$\begin{aligned}\bar{x} \pm 2.58 \frac{\sigma}{\sqrt{n}} &= 62 \pm 2.58 \frac{12}{4} \\ &= 62 \pm 7.74 \\ &\quad (54.26; 69.74)\end{aligned}$$

Stora stickprov

Om fördelningen för X är okänd, utnyttjar vi att fördelningen för $\bar{X}$ är approximativt normalfördelad om n är stort ($n \geq 30$). Alltså

$$\bar{X} \sim approx N \left(\mu; \frac{\sigma}{\sqrt{n}} \right) \text{ enligt CGS.}$$

Om populationsvariansen σ^2 också är okänd skattas den med stickprovsvariansen s^2 . Ett $(1 - \alpha) 100$ %igt k.i. för μ ges då av

$$\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$$

Exempel: I ett stickprov av storlek $n = 64$ är medelinkomsten 15 tusen kr och standardavvikelsen 9 tusen kr. Ett 95 %-igt k.i. för medelinkomsten i populationen ges av

$$\begin{aligned}\bar{x} \pm 1.96 \frac{s}{\sqrt{n}} &= 15 \pm 1.96 \frac{9}{8} \\ &= 15 \pm 2.20 \\ &\quad (12.8; 17.2)\end{aligned}$$

Tolkning: Med 95 % tillförlitlighet ligger pop.medelinkomsten i intervallet.

Normalfördelade variabler med okänd varians

Sedan tidigare vet vi att om $X_1, \dots, X_n$ är ett stickprov av oberoende normalfördelade variabler med okänt medelvärde μ och känd standardavvikelse σ så gäller att

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

Om vi har samma förutsättningar men σ^2 är okänd och skattas i stickprovet med s^2 , då gäller att

$$\frac{\bar{X} - \mu}{\frac{s}{\sqrt{n}}} \sim t(n - 1),$$

dvs t-fördelad med $(n - 1)$ frihetsgrader. Ett $(1 - \alpha) 100$ %-igt k.i. för μ ges då av

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

Exempel:

Ett telefonbolag vill skatta den genomsnittliga längden på långdistanssamtal under helger. Låt X = "samtalslängd". Låt $X_1, \dots, X_n$ vara ett stickprov ($n = 20$) av oberoende normalfördelade variabler med okänt medelvärde μ och okänd standardavvikelse σ . I stickprovet är $\bar{x} = 14.5$ minuter och $s = 5.6$ minuter. Ett 95 %-igt k.i. för μ ges av

$$\begin{aligned}\bar{x} \pm t_{0.025} (20 - 1) \frac{s}{\sqrt{n}} &= 14.5 \pm 2.09 \frac{5.6}{\sqrt{20}} \\ &= 14.5 \pm 2.6 \\ &\quad (11.9; 17.1)\end{aligned}$$

Tolkning: Med konfidensgraden 95 % täcker intervallet populationsparametern μ .

Sammanfattning: k.i. för μ

1. Undersökningsvar är normförd

σ är känd: $\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

σ är okänd: $\bar{x} \pm t_{\alpha/2}(n-1) \frac{s}{\sqrt{n}}$

2. Förd undersökningsvar okänd, stort stickprov

σ är känd: $\bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

σ är okänd: $\bar{x} \pm z_{\alpha/2} \frac{s}{\sqrt{n}}$

K.i för pop.andelar p

Om $X_1, \dots, X_n$ är ett stickprov av oberoende variabler från

$$X \sim \text{Bin}(n; p)$$

så gäller att

$$X \sim \text{approx } N(np; np(1-p))$$

om $npq > 5$. Då gäller även att

$$\hat{P} = \frac{X}{n} \sim \text{approx } N\left(p; \frac{p(1-p)}{n}\right).$$

Dessutom är då

$$\frac{\hat{P} - p}{\text{sqrt}\left(\frac{p(1-p)}{n}\right)} \sim \text{approx } N(0; 1).$$

Ett $(1 - \alpha)$ 100 %-igt k.i. för p ges då av

$$\hat{p} \pm z_{\alpha/2} \text{sqrt}\left(\frac{\hat{p}(1-\hat{p})}{n}\right),$$

där $\hat{p}$ (litet) är den observerade skattningen i stickprovet.

Exempel:

I en presidentvalsundersökning svarade 777 av $n = 1500$ personer att de skulle rösta på George W.? Uppskatta andelen som skulle rösta på George W. i hela väljarkåren.

Andelen anhängare till George W. i stickprovet är

$$\hat{p} = \frac{777}{1500} = 0.518.$$

Då $n\hat{p}\hat{q} = 375 > 5$ är

$$\hat{P} \sim approx N \left(p; \frac{p(1-p)}{n} \right).$$

Ett 95 %-igt k.i. för p ges då av

$$\begin{aligned} & \hat{p} \pm z_{0.025} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \\ &= 0.518 \pm 1.96 \sqrt{\frac{0.518 \cdot 0.482}{1500}} \\ &= 0.518 \pm 0.025. \end{aligned}$$

Med 95 % tillförlitlighet ligger andelen anhängare till George W. i intervallet (0.49; 0.54).

Korrektionsfaktor för ändlig pop.

Vid slumpmässigt urval utan återläggning ur en ändlig population då stickprovet är stort i förhållande till populationen (tumregel $n/N > 0.1$), ges variansen för en skattning under vissa förutsättningar t. ex av

$$V(\bar{X}) = \frac{\sigma^2}{n} \frac{N-n}{N-1},$$

där faktorn $\frac{N-n}{N-1}$ kallas korrektionsfaktorn. Ett $(1 - \alpha)$ 100 %-igt k.i. för μ ges då av

$$\bar{x} \pm z_{\alpha/2} \sqrt{\frac{\sigma^2}{n} \frac{N-n}{N-1}}.$$

Bestämning av stickprovsstorlek

Vi vill dra ett stickprov av oberoende variabler $X_1, \dots, X_n$ från en normalfördelning med okänt medelvärde μ och känd varians $\sigma^2 = 16$. Hur stort stickprov krävs om vi vill att ett 99 %-igt k.i. för μ skall ha en precision så att längden på k.i. inte överstiger fyra? Ett 99 %-igt k.i. för μ ges av

$$\begin{aligned} \bar{x} \pm z_{0.005} \frac{\sigma}{\sqrt{n}} \\ = \bar{x} \pm 2.58 \frac{4}{\sqrt{n}} \end{aligned}$$

Det betyder att halva k.i. längd skall vara

$$\begin{aligned} 2.58 \frac{4}{\sqrt{n}} &\leq 2 \\ \frac{2.58^2 \cdot 4^2}{2^2} &\leq n \\ 26.63 &\leq n \end{aligned}$$

Alltså det krävs att $n \geq 27$.