

SF1901: Sannolikhetslära och statistik

Föreläsning 7.B

Statistik: statistiska inferensproblem

Jan Grandell & Timo Koski

10.02.2012



Vi inleder med ett exempel på statistiska *inferensproblem* (inferens= slutledning).

Vi drar statistiska slutsatser med stöd av *sannolikhetsmodeller* (byggda på kunskaper från kursens första del).

Inledande om opinionsundersökning: punktskattning

Vi väljer 1000 personer ur en 'stor' population (t.ex. Sveriges befolkning). De ska svara 'JA' eller 'NEJ' till en fråga (t.ex. om medlemskap i militäralliansen NATO).

Ponera nu att

$$x = 350 \quad \text{svarat JA}$$

Vi vill veta proportionen JA-sägare i HELA populationen, dvs. punktskatta proportionen JA-sägare i HELA populationen.

En skattning av p :

$$p_{obs}^* = \frac{x}{1000} \Rightarrow p_{obs}^* = 35\%$$

Hur stor är osäkerheten i denna skattning?



Inledande diskussion: sannolikhetsmodell för skattning

$p_{obs}^* = \frac{x}{1000} = 0.35$. Hur stor är osäkerheten i denna skattning? Vi använder oss av en sannolikhetsmodell.

URNMODELL: p = andelen JA-svar i populationen. X = antalet JA-svar när vi frågar 1000 individer. N = populationens storlek, t.ex. $N = 7,3$ miljoner svenskar över 15 år (enligt SCB var Sveriges befolkning 9 201 650 invånare på nyårsafton 2008).

Vi har att

$$X \in \text{Hyp}(N, 1000, p)$$

Inledande diskussion: modell för skattning

Vi återkallar i minnet

Sats

Om X är $\text{Hyp}(N, n, p)$ -fördelad med $n/N \leq 0.1$ så är X approximativt $\text{Bin}(n, p)$ -fördelad.

Då kan vi utan stor förlust ta

$$X \in \text{Bin}(1000, p)$$

Vi uppfattar $x = 350$ som ett utfall av X . Vi kallar

$$p^* = \frac{X}{1000}$$

en *stickprovsvariabel*. $p_{\text{obs}}^* = \frac{x}{1000} = 0.35$ är ett utfall av p^*



Inledande diskussion: modell för skattning

$$X \in \text{Bin}(1000, p)$$

Vi vet att $E(X) = 1000 \cdot p$, $V(X) = 1000 \cdot p \cdot (1 - p)$. Detta ger

$$E(p^*) = E\left(\frac{X}{1000}\right) = \frac{1}{1000} E(X) = \frac{1}{1000} \cdot 1000 \cdot p = p.$$

$$V(p^*) = V\left(\frac{X}{1000}\right) = \frac{1}{1000^2} V(X) = \frac{1}{1000^2} \cdot 1000 \cdot p(1 - p) = \frac{p(1 - p)}{1000}.$$

Svårigheten är att de ovanstående uttrycken beror på den okända (parameter) p som vi vill skatta!



Inledande diskussion: medelfelet för skattning

$$E(p^*) = p, \quad V(p^*) = \frac{p(1-p)}{1000}.$$

$$D(p^*) = \sqrt{V(p^*)} = \sqrt{\frac{p(1-p)}{1000}}$$

Vår skattning av osäkerhet i skattningen är då *medelfelet*

$$d(p^*) = \sqrt{\frac{p_{obs}^*(1-p_{obs}^*)}{1000}} = \sqrt{\frac{0.35(1-0.35)}{1000}} \approx 0.0015 = 1.5\%$$

Vi har en uppsättning data

$$x_1, x_2, \dots, x_n$$

som ses som utfall av s.v.

$$X_1, X_2, \dots, X_n.$$

Dessa variabler antages vara oberoende och likafördelade och deras gemensamma fördelning beror av en okänd parameter θ , t.ex. $\text{Bin}(n, \theta)$, $N(\theta, \sigma)$, $\text{Po}(\theta)$, $N(\theta_1, \theta_2)$, osv.



- En punktskattning θ_{obs}^* av θ är en funktion $\theta^*(x_1, \dots, x_n)$

- En punktskattning θ_{obs}^* av θ är en funktion $\theta^*(x_1, \dots, x_n)$
- och motsvarande stickprovsvariabel θ^* är $\theta^*(X_1, \dots, X_n)$

- En punktskattning θ_{obs}^* av θ är en funktion $\theta^*(x_1, \dots, x_n)$
- och motsvarande stickprovsvariabel θ^* är $\theta^*(X_1, \dots, X_n)$
- Exemplet ovan:

$$p_{obs}^* = \frac{350}{1000} = 0.35$$

är en punktskattning.

$$p^* = \frac{X}{1000}, \quad X \in \text{Bin}(1000, p)$$

är motsvarande stickprovsvariabel.

Vad menas med en bra skattning?

Vi ger tre kriterier:

Definition

- 1) En punktskattning θ_{obs}^* av θ är väntevärdesriktig om $E(\theta^*(X_1, \dots, X_n)) = \theta$.

Vad menas med en bra skattning?

Vi ger tre kriterier:

Definition

- 1) En punktskattning θ_{obs}^* av θ är väntevärdesriktig om $E(\theta^*(X_1, \dots, X_n)) = \theta$.
- 2) En punktskattning θ_{obs}^* av θ är konsistent om $P(|\theta^*(X_1, \dots, X_n) - \theta| > \varepsilon) \rightarrow 0$ då $n \rightarrow \infty$.

Vad menas med en bra skattning?

Vi ger tre kriterier:

Definition

- 1) En punktskattning θ_{obs}^* av θ är väntevärdesriktig om $E(\theta^*(X_1, \dots, X_n)) = \theta$.
- 2) En punktskattning θ_{obs}^* av θ är konsistent om $P(|\theta^*(X_1, \dots, X_n) - \theta| > \varepsilon) \rightarrow 0$ då $n \rightarrow \infty$.
- 3) Om θ_{obs}^* och θ_{obs}^{**} är väntevärdesriktiga skattningar av θ så säger man att θ_{obs}^* är effektivare än θ_{obs}^{**} om $V(\theta^*(X_1, \dots, X_n)) < V(\theta^{**}(X_1, \dots, X_n))$.

$$X \in \text{Bin}(1000, p)$$

$$p^* = \frac{X}{1000}, \quad E(p^*) = p.$$

Således är p_{obs}^* väntevärdesriktig.

Sats

Stickprovsmedelvärdet $\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$ som skattning av väntevärdet μ är

- 1) Väntevärdesriktig;
- 2) Konsistent;
- 3) Ej nödvändigtvis effektiv, dvs. den effektivaste möjliga skattningen.

Bevis.

- 1) $E(\bar{X}) = \mu$.
- 2) $V(\bar{X}) = \sigma^2 / n$ och stora talens lag gäller.
- 3) Det finns motexempel (den intresserade hänvisas till Blom, Englund et. al.)