Roosevelt

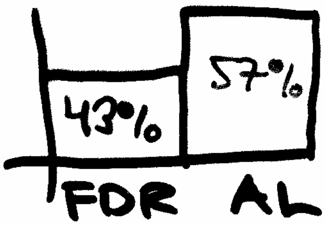
Landon

Am. presidentval, 1936

Opinionsundersökning av Literary Digest:

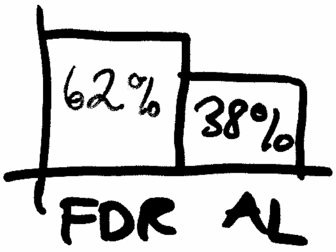
<u>2,4 miljoner</u> personer frågades, slumpmässigt valda från telefonkatalogen, bilregistret osv.

Prognos:



Problem: Systematisk exkludering av folk utan bil, telefon, etc

Resultat:



Fördläsning 10

Statistisk inferens (=slutledning)

Vill ställa en fråga som kan ge olika utfall.
→ statistisk undersökning

Steg-för steg

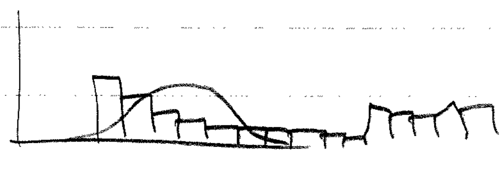
1. Formulera problemet. Vad ska mätas?
   ex. Vad är fördelningen av svenska mäns längd? (se till att alla är barfota)

2. Försöksplanering. Hur mäter vi?
   ex. Välj 1000 män slumpmässigt? Telefonkatalog, gata på stan?

3. Datainsamling: mäter längder → observationer → stickprov
   alla måste ha likadan linjal & avrunda p.s.s.

4. Bearbetning av data
   ex. modell: längd ~ $N(\mu, \sigma)$
   kompensera, korrigera

5. Presentation.
   ex. skattning av μ, σ. Hur bra är dom skattningarna? Var normalfördelning rimligt antagande?



Statistisk inferens - hundbegrepp

- Stickprov: observationer från en population med någon fördelning (i regel oberoende observationer). Används för att dra slutsatser om hela populationen.

- (punkt)skattning: Skatta värdet på en parameter; tex μ i en $N(\mu, 2)$-fördelning.

- konfidensintervall: ett intervall $[a,b]$ inom vilket en parameters verkliga värde ligger med en viss sannolikhet.

- hypotesprövning/Statistiskt test: ansätt en hypotes och avgör om data stöder denna tillräckligt väl.

Skattning av väntevärde

Anta vi har n oberoende stickprov $x_1, \ldots, x_n$ från en fördelning med okänt väntevärde $\mu = E[X_i]$

Vi skattar väntevärdet μ med en skattare, nämligen stickprovsmedelvärdet

$$\hat{\mu} = \bar{x} = \frac{x_1 + \ldots + x_n}{n}$$

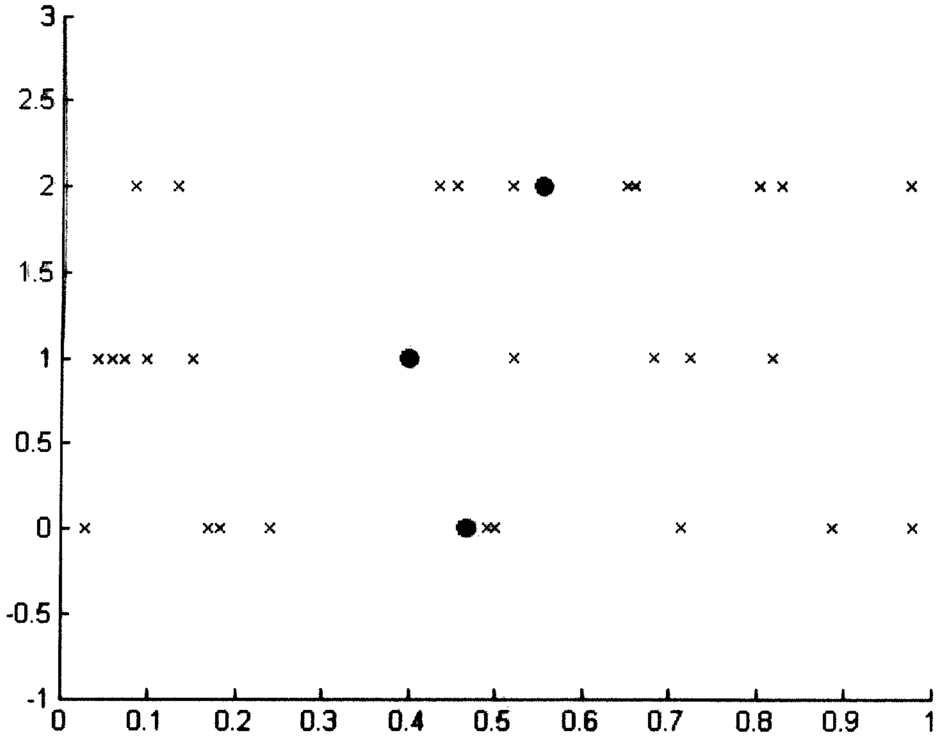
Hur "bra" är det att göra så? Vad betyder "bra"?

Betrakta nu ist. s.v. med denna fördelning; uttryck deras medelvärde:

$$\bar{X} = \frac{X_1 + \ldots + X_n}{n}$$

$X_1, \ldots, X_n$ är s.v. $\Rightarrow$ $\bar{X}$ är s.v.

$\Rightarrow$ $\bar{X}$ har en fördelning!



$$X \sim U([0,1])$$

$$\bar{X} = \frac{X_1 + \dots + X_n}{n} \sim ?$$

Kan beräkna väntevärde och varians för $\bar{X}$:

$$E[\bar{X}] = E\left[\frac{X_1 + \dots + X_n}{n}\right] = \frac{1}{n} E\left[\sum_{i=1}^{n} X_i\right] = \frac{1}{n} \sum_{i=1}^{n} \underbrace{E[X_i]}_{=\mu} =$$

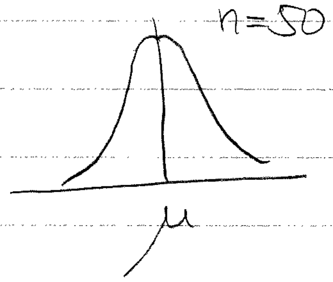
$$= \frac{1}{n} \sum_{i=1}^{n} \mu = \frac{1}{\cancel{n}} \cancel{n}\mu = \mu .$$

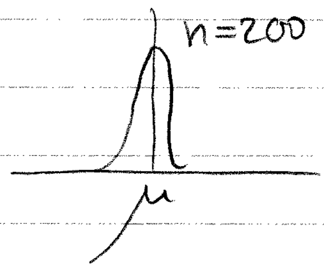
$\Rightarrow$ Så $E[\hat{\mu}] = \mu$; i 'medeltal' skattar vi precis rätt!

$$V(\bar{X}) = V\left(\frac{X_1 + \dots + X_n}{n}\right) = \frac{1}{n^2} V(X_1 + \dots + X_n) = (\text{oberoende}) =$$

$$= \frac{1}{n^2} \sum_{i=1}^{n} V(X_i) = \frac{1}{n^2} n V(X_i) = \frac{\sigma^2}{n}$$

$\Rightarrow$ Skattaren blir mer precis (effektivare) när stickprovets storlek ökar.

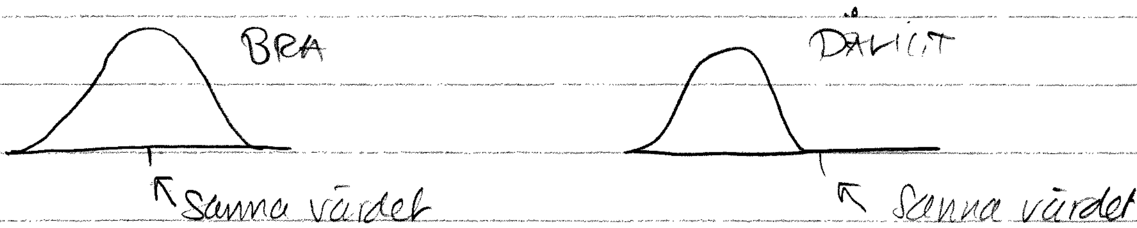




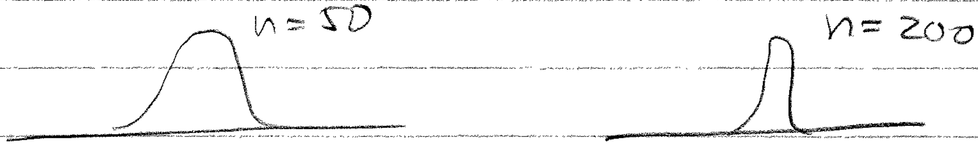
Leder oss in på...

Vilka egenskaper vill vi ha hos en skattare?

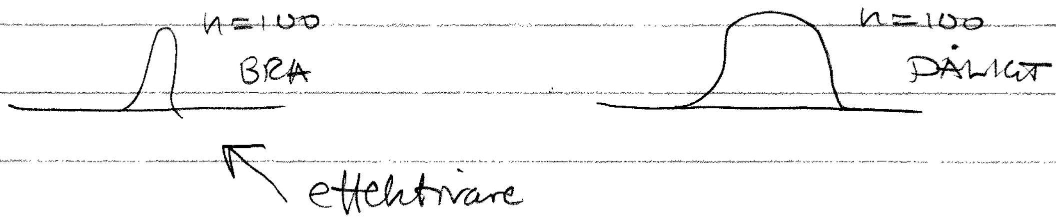
Väntevärdesriktighet: "i medeltal rätt"



Konsistens: "ju större stickprov desto bättre skattning"



Effektivitet: liten varians



$\bar{X}$ är väntevärdesriktig. Kan göra den effektivare genom att skala ner den, men då ej väntevärdesriktig

Skattning av varians

Varians är "medelvärde av kvadrerad avvikelse från medelvärdet", verkar logiskt att skatta $\sigma^2 = Var(X)$ som

$$\hat{\sigma}^2 = S_*^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2$$

men S_*^2 är inte väntevärdesriktig utan har systematiskt fel (underestimering)

$$E[S_*^2] = \frac{n-1}{n}\sigma^2 \neq \sigma^2 \quad (<\sigma^2)$$

$\Rightarrow$ Använd istället $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i-\bar{x})^2$

$E[S^2] = \sigma^2 \Rightarrow$ väntevärdesriktig

(i gengäld är S^2 mindre effektiv än S_*^2; $Var(S_*^2) = \left(\frac{n-1}{n}\right)^2 Var(S^2) < Var(S^2)$)