

F11

Olika urvalsmetoder,
speciellt obundet
slumpmässigt urval
(OSU)

Punkt- och
intervallskattning

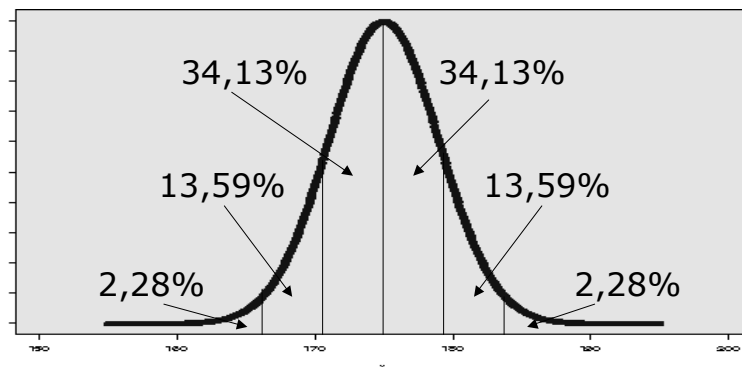
Förra gången (F10)

- Vad är viktigt att kunna?
 - Om slumpvariabeln X är normalfördelad så vet vi hur stor andel av populationen som finns inom vissa intervall
 - Ex. om längd är normalfördelad med medelvärde 175 och standardavvikelse 4 så är ungefär 95% av alla i populationen mellan 167 och 183 cm.
 - Dvs om $X \in N(175, 4)$ så finns ungefär 95% av observationerna inom ± 2 standardavvikelser från medelvärdet. Mer exakt finns 95% av observationerna inom $\pm 1,96$ standardavvikelser

Exempel på en av oändligt många normalfördelningar

- Längd: $X \in N(175,4)$

- Dvs X är normalfördelad med medelvärde 175 och standardavvikelse 4



Sannolikhetsfördelningen för X-bar

- Om $X \in N(\mu_X, \sqrt{V(X)}) \Rightarrow \bar{X} \in N\left(\mu_X, \sqrt{\frac{V(X)}{n}}\right)$
- Om X-bar är normalfördelad med ett visst medelvärde och standardavvikelse så vet vi hur sannolikhetsfördelningen ser ut och hur stor andel av alla X-bar som finns inom vissa intervall.
- Ex. om $X \in N(175,4) \Rightarrow \bar{X} \in N\left(175, \frac{4}{\sqrt{n}}\right)$
 - Om vi drar alla möjliga urval av $n=64$ så vet vi att i ungefär 95% av alla urval får vi ett X-bar som ligger mellan 174 och 176 cm.

Hur kan vi använda denna information för att bilda konfidensintervall/felmarginal?

- Om vi drar alla möjliga urval av en viss storlek (n), beräknar \bar{X} för alla urval och bildar ett intervall som är $\pm 1,96$ standardavvikelse (för \bar{X}) runt \bar{X} så vet vi att 95% av dessa intervall kommer täcka det sanna medelvärdet.
- Varför?

Konfidensintervall

- 95%-igt konfidensintervall för μ :

$$\bar{X} \pm 1,96 \sqrt{\frac{V(X)}{n}}$$

- Detta intervall bygger alltså på att sannolikhetsfördelningen för \bar{X} är normalfördelad. När är den det?

Generellt

- Vi har en okänd parameter
 - Ex μ , P (andelen) eller τ (totalvärdet)
- Vi söker ett 95%-igt konfidensintervall för t ex μ .
- Generell skattning för μ från ett slumpmässigt urval: $\hat{\mu}$
- Om vi antar att CGV gäller, dvs att $\hat{\mu}$ är approximativt normalfördelat, kan vi skriva $\hat{\mu} \pm$ felmarginalen

$$\hat{\mu} \pm 1,96 \times \sqrt{V(\hat{\mu})}$$

OSU-urval med eller utan återläggning

- Vi drar ett OSU-urval av n element från N element totalt
- Alla element i populationen har samma sannolikhet att väljas
 - Använd någon slumpalsgenerator såsom slumptabell/dator
- Det är slumpen som styr vilka element som väljs
- Vi kan skatta μ genom stickprovsmedelvärdet $\bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

OSU

- Vi kan skatta variansen med stickprovsvariansen

$$s^2 = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}$$

- *Med återläggning* kan vi då skatta variansen för skattningen (\bar{X}) som

$$\hat{V}(\bar{X}) = \frac{s^2}{n}$$

OSU utan återläggning

- *Utan återläggning* lägger vi till en ändlighetskorrektion och skattar variansen för skattningen med

$$\hat{V}(\bar{X}) = \left(1 - \frac{n}{N}\right) \frac{s^2}{n}$$

Konfidensintervall

- Vid OSU utan återläggning kan vi då beräkna ett 95%igt konfidensintervall som

$$\bar{x} \pm 1,96 \times \sqrt{\left(1 - \frac{n}{N}\right) \frac{s^2}{n}}$$

- Tolkning: Med 95% konfidens ligger det sanna medelvärdet inom gränserna

$$\left[\bar{x} - 1,96 \times \sqrt{\left(1 - \frac{n}{N}\right) \frac{s^2}{n}}; \bar{x} + 1,96 \times \sqrt{\left(1 - \frac{n}{N}\right) \frac{s^2}{n}} \right]$$

- Om vi drog ett oändligt antal urval skulle 95% av dem täcka det sanna medelvärdet μ
- Vad har vi gjort för antaganden?

Exempel

- Vi drar ett OSU urval av $n=30$ från en population på $N=300$ och beräknar den skattade medellängden och variansen
- Vi får $\bar{x} = 176,20$
- Vi får $s^2 = 20,00$
- Ett 95%-igt konfidensintervall för μ blir då:

$$176,20 \pm 1,96 \times \sqrt{\left(1 - \frac{30}{300}\right) \frac{20,00}{30}} = 176,20 \pm 1,52$$

- Dvs, den sanna medellängden ligger med 95% säkerhet mellan 174,68 och 177,72
- Vad har vi gjort för antaganden?