

Kognitiv psykologi

Moment 1: Statistik, 3 poäng

VT 2007

Lärare: Maria Karlsson

1

Samplingfördelningar (Kapitel 4 i Howell)

2

Parametrar och statistikor

- En *parameter* är en konstant som "karakteriserar" en population eller en modell.
- Exempel:
 - Populationsmedelvärdet
 - Parametern p i binomialfördelningen

3

Vi har hittills berört problem av följande typ:

Antag att proportionen moderater i en population är 0.20. Vad är sannolikheten att få två moderater om vi slumpmässigt väljer tio individer ur populationen?

4

När vi kommer över till statistikteorin handlar det i stället om att **utifrån ett stickprov dra slutsatser om en population.**

5

- **Estimation:**

- Vi vill med hjälp av ett stickprov skatta proportionen moderater i populationen. Hur gör vi detta på "bästa" sätt och hur säker är vår uppskattning?

- **Hypotesprövning:**

- Vår hypotes är att proportionen moderater i populationen är minst 0.25. Hur skall vi med hjälp av ett stickprov avgöra om hypotesen är falsk eller ej?

6

- Vid såväl estimation som hypotesprövning använder vi observationerna i stickprovet för att räkna fram numeriska värden av olika slag.
- Exempel:
 - Vi använder kanske stickprovsproportionen för att uppskatta populationsproportionen.

7

- Numeriska värden som på detta sätt beräknas med hjälp av observationerna i ett stickprov kallas för **statistikor**.
- Observera att en statistika är en slumpvariabel eftersom dess värde bestäms av ett slumpmässigt försök och varierar från stickprov till stickprov.
- Exempel på statistikor:
 - Stickprovsmedelvärdet
 - stickprovsstandardavvikelsen

8

- Statistikor som används för estimation kallas för **estimatorer**.
- Observationer på estimatorer kallas för **estimat**.
- Statistikor som används vid hypotesprövning brukar kallas för **teststatistikor**.

9

Samplingfördelningar

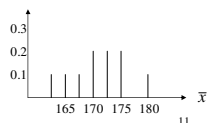
- En **samplingfördelning** är en sannolikhetsfördelning för en statistika (t ex ett medelvärde), dvs fördelningen av värden på statistikan vid upprepad (*oändligt många gånger*/"*alla möjliga stickprov*") stickprovsdragning (eller experiment) från en population.
- Utseendet beror på populationen.
- "Räknas ut" utifrån sannolikhetslära.

10

Exempel: samplingfördelning

- Population: 5 st personer, $X =$ längd
Pelle, 165cm, Svea, 160, Bertil, 180, Kerstin, 170 cm, Lars, 180
- Stickprov: 2 personer

PS	$(165+160)/2 = 162.5$
PB	$(165+180)/2 = 172.5$
PK	167.5
PL	172.5
SB	170.0
SK	165.0
SL	170.0
BK	175.0
BL	180.0
KL	175.0



11

Centrala gränsvärdessatsen

- Oavsett formen på den populationsfördelning ett slumpmässigt stickprov hämtas från, förutsatt att fördelningen har ändligt väntevärde och varians, så går fördelningen för stickprovsmedelvärdet mot en normalfördelning då stickprovsstorleken ökar.

12

- Centrala gränsvärdesatsen är även tillämpbar på summor av oberoende slumpvariabler.
 - Exempel: En binomialfördelad variabel, som består av en summa av oberoende bernoullifördelade variabler, tenderar att bli normalfördelad då n ökar.

13

- Hur stort stickprov som behövs för att normalfördelningen skall kunna användas som en approximativ modell beror på hur populationsfördelningen ser ut.
- Tumregel: I de flesta fall är normalfördelningen en tillräckligt god approximation redan vid stickprovsstorleken 30.

14

Samplingfördelningen för stickprovsmedelvärdet

- Om vi samplar från en normalfördelning är stickprovsmedelvärdet alltid normalfördelat oavsett stickprovsstorlek.
- Om vi samplar från någon annan fördelning är stickprovsmedelvärdet approximativt normalfördelat om stickprovet är "stort".

15

Stickprovsmedelvärdets väntevärde och varians

- Låt n vara stickprovsstorleken och låt μ och σ^2 vara medelvärdet och variansen i den population vi tar stickprovet ifrån då blir
- Väntevärde för stickprovsmedelvärdet:

$$E(\bar{X}) = \mu$$
- Varians för stickprovsmedelvärdet:

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

16

Exempel: Skilsmässobarn

- Total Behavior Problem poäng på Achenbach's formulär är normalfördelat i hela populationen av "normala barn" med ett populationsmedelvärdet $\mu = 50$ poäng och en populationsstandardavvikelse på $\sigma = 10$ poäng.

17

Exempel: Skilsmässobarn (forts.)

- Tar ett stickprov om 5 st skilsmässobarn slumpmässigt. Stickprovsmedelvärdet blir 56. Är detta slumpen eller är skilsmässobarn speciella?
- Om skilsmässobarn tillhörde den normala barngruppen – vad skulle då kunna vänta sig för stickprovsmedelvärde om man tar ett stickprov om 5 st barn slumpmässigt?

18

Exempel: Skilsmässobarn (forts.)

- Stickprovsmedelvärdet är normalfördelat med väntevärde 50 och standardavvikelse $10/\sqrt{5}=4.47$.
 - Stickprovsmedelvärden mellan säg 48 och 52 skulle vara mycket vanliga. Stickprovsmedelvärden på 56 poäng skulle också inträffa ganska ofta, men ett stickprovsmedelvärde på 70 skulle inträffa mycket sällan.

19

Fördelningen för stickprovproportionen

- Exempel:
 - Tag ett stickprov av n st individer och sätt man=1 och kvinna=0.
 - Vi har då observationer på n st bernoullifördelade variabler, X_1, \dots, X_n .
 - Om vi antar oberoende mellan variablerna är summan av dem, dvs. antalet män i stickprovet, binomialfördelad.

20

- Bilda andelen män i stickprovet, dvs. stickprovproportionen:

$$\hat{p} = \frac{\sum_{i=1}^n X_i}{n}$$

21

- Som vi ser i exemplet är stickprovproportionen inget annat än ett stickprovsmedelvärde där variablerna i summan i täljaren är bernoullifördelade.
- Om stickprovet är stort kan vi, med stöd av centrala gränsvärdesatsen, hävda att stickprovproportionen är approximativt normalfördelad.
- Tumregel: $np \geq 5$ och $n(1-p) \geq 5$

22

Väntevärde och varians

- Vi såg tidigare att väntevärdet och variansen för stickprovsmedelvärdet var lika med populationens medelvärde och populationens varians dividerad med stickprovstorleken.
- När populationen är bernoullifördelad är populationens medelvärde p och populationens varians $p(1-p)$.

23

- Väntevärdet för stickprovproportionen:

$$E(\hat{p}) = p$$

- Variansen för stickprovproportionen:

$$V(\hat{p}) = \frac{p(1-p)}{n}$$

24