

Föreläsning 5

Kapitel 6, sid 153-185

Inferens om en population

Agenda

- Statistisk inferens om populationsmedelvärde
 - Statistisk inferens om populationsandel
 - Punktskattning
 - Konfidensintervall
 - Hypotesprövning
-

Statistisk inferens om populationsmedelvärde

Punktskattning

- Att använda en stickprovsstatistika som en uppskattning av motsvarande parameter
- Stickprovsstatistikor är slumpvariabler och antar olika värden för varje stickprov
- Hur ska vi hantera osäkerheten?

Krav för konfidensintervall för medelvärde

1. Stickprovet är draget som ett OSU
 - Garanterar oberoende mellan observationerna

2. Samplingfördelningen för stickprovsmedelvärdet kan betraktas som normalfördelad
 - Antingen genom centrala gränsvärdessatsen, $n \geq 30$
 - Populationen kan betraktas som normalfördelad

- Om kraven är uppfyllda kan vi skatta osäkerheten genom ett så kallat konfidensintervall

Punktskattning och intervallskattning

- Om kraven är uppfyllda kan vi bilda ett konfidensintervall för populationsmedelvärdet: vi lägger ett osäkerhetsintervall kring punktskattningen vilket tillåter oss att med en viss säkerhet säga att den okända populationsparametern täcks av intervallet.

Dubbelsidigt konfidensintervall för populationsmedelvärde när σ är okänd

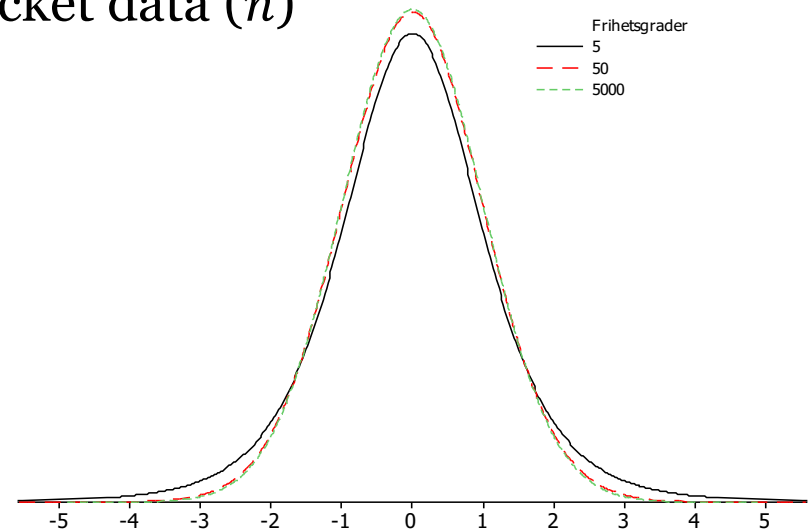
- Givet de två antaganden bildas ett dubbelsidigt konfidensintervall för μ enligt:

$$\bar{x} \pm t_{n-1; 1-\frac{\alpha}{2}} \cdot \frac{s}{\sqrt{n}}$$

- Värdet på t hämtas från t-fördelningen, där $n - 1$ är frihetsgrader och α är signifikansnivån

t-fördelningen

- Fördelningen liknar normalfördelningen men beror på frihetsgrader
- Används om stickprovet är litet och om σ är okänd
- Frihetsgrader bestäms av hur mycket data (n) man har
- Fördelningen konvergerar mot normalfördelningen
- Approximativt normal då $n \geq 30$



Exempel

- Ett slumpmässigt urval om 40 studenter vid Linköpings universitet ger medelåldern 21.2 år och standardavvikelsen 4.4 år.

Bestäm ett intervall som med 95 procents säkerhet täcker den sanna medelåldern bland studerande vid Linköpings universitet.

Enkelsidiga konfidensintervall (KI)

- Nedåt begränsat KI: $\mu > \bar{x} - t_{n-1;1-\alpha} \cdot \frac{s}{\sqrt{n}}$
- Uppåt begränsat KI: $\mu < \bar{x} + t_{n-1;1-\alpha} \cdot \frac{s}{\sqrt{n}}$

Exempel fråga

- Styrelsen i en bostadsrättsförening får in klagomål på att golvvärmen i badrummen är för låg. Man drar ett OSU om 30 badrum bland de omkring 400 badrum som finns i föreningens fastigheter och mäter golvvärmen där. Medeltemperaturen beräknas till 21 grader och standardavvikelsen till 1.6 grader.

Energimyndigheten rekommenderar att golvvärmen ska ligga på minst 20 grader för att man ska undkomma problem med fuktskador. Föreligger risk för fuktskador i föreningens badrum?

Hypotesprövning för populationsmedelvärde när σ är okänd

Förutsätts åter att

1. stickprovet är draget som ett OSU
 2. samplingfördelningen för stickprovsstatistikan kan betraktas som normalfördelad
- En **hypotesprövning** testar om ett påstående (en hypotes) är förenlig med ett observationerna i ett stickprov
 - En hypotesprövning kan utföras i 4 steg

Exempel

- I ett OSU omfattande 40 personer bland medlemmarna i ett politiskt parti i en region är medelåldern 42.3 år och standardavvikelsen 7.1 år.

Testa på 5 procents signifikansnivå om medelåldern bland medlemmarna i partiet överstiger 45 år.

Hypotesprövning

- Steg 1: Formulera hypoteser och välj signifikansnivå

$$H_0: \mu = \mu_0$$

$$H_a: \mu > \mu_0$$

$$H_a: \mu < \mu_0$$

$$H_a: \mu \neq \mu_0$$

- Valet av mothypotes, H_a , bestäms av frågeställningen
- α = signifikansnivån = risken att förkasta H_0 trots att H_0 är sann
 - Vanliga värden är 1%, 5%, 10%

Hypotesprövning

- Steg 2: Bestäm testvariabeln

$$t_{test} = \frac{\bar{x} - \mu_0}{\frac{s}{\sqrt{n}}}$$

Hypotesprövning

- Steg 3: Beräkna det kritiska området

- Om $H_a: <$, till vänster om

$$t_{krit} = -t_{n-1;1-\alpha}$$

- Om $H_a: >$, till höger om

$$t_{krit} = t_{n-1;1-\alpha}$$

- Om $H_a: \neq$, på båda sidor om

$$t_{krit} = \pm t_{n-1;1-\frac{\alpha}{2}}$$

- Visualisera detta med ett diagram!

Hypotesprövning

- Steg 4: Dra slutsatser och tolka
 - Besluta att förkasta/inte förkasta H_0
 - Besvara frågeställningen i ord

Om σ är känd

- Väckigt ovanligt att σ är känd
- I detta fallet byts t-fördelningen ut mot normalfördelningen (ex. byts $t_{n-1;1-\alpha}$ ut mot $z_{1-\alpha}$) i formlerna. s byts ut mot σ .

Inferens om populationsandel

Konfidensintervall för populationsandel

- Givet att:
 1. Stickprovet är draget som ett OSU
 2. $np(1 - p) > 5$bildas ett dubbelsidigt KI för π enligt:

$$p \pm z_{1-\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}}$$

- Där värdet på z hämtas ur normalfördelningstabellen

Enkelsidiga intervall

- Nedåt begränsat KI:

$$\pi > p - z_{1-\alpha} \sqrt{\frac{p(1-p)}{n}}$$

- Uppåt begränsat KI:

$$\pi < p + z_{1-\alpha} \sqrt{\frac{p(1-p)}{n}}$$

Exempel

- I en hälsoenkät tillfrågades 100 slumpmässigt utvalda anställda vid ett stort företag om huruvida man regelbundet motionerar eller ej.

Svar erhöles från 84 anställda och av dessa svarade 65 ja.

Bestäm ett 95-procentigt konfidensintervall för andelen av de anställda vid det stora företaget som regelbundet motionerar.

Hypotesprövning

- Förutsätter att kraven för andelar är uppfyllda
- Steg 1:

$$H_0: \pi = \pi_0$$

$$H_a: \pi > \pi_0$$

$$H_a: \pi < \pi_0$$

$$H_a: \pi \neq \pi_0$$

Hypotesprövning

- Steg 2:

$$z_{test} = \frac{p - \pi_0}{\sqrt{\frac{\pi_0(1 - \pi_0)}{n}}}$$

- Steg 3:

$$z_{krit} = -z_{1-\alpha}; z_{1-\alpha}; z_{1-\frac{\alpha}{2}}$$

– Beror på H_a

Hypotesprövning

- Steg 4:
 - Dra slutsats och tolka

Exempel

- I en hälsoenkät tillfrågades 100 slumpmässigt utvalda anställda vid ett stort företag om huruvida man regelbundet motionerar eller ej.

Svar erhöles från 84 anställda och av dessa svarade 65 ja.

Undersök om det på 5 procents signifikansnivå finns belägg för påståendet att andelen regelbundna motionärer bland de anställda vid företaget understiger 85 procent.

Ytterligare om inferens

p-värdesmetoden

- Som ett alternativt steg 3 och 4 i en hypotesprövning
- p-värdet är sannolikheten att testvariabeln ska anta ett värde som det vi observerat eller ännu längre ifrån μ_0 sett i den riktning som mothypotesen pekar
- Om p-värdet är litet är H_0 osannolik
- Beslutsregel: om p-värdet $<$ signifikansnivån α så förkastas H_0
 - Vid dubbelsidig mothypotes beräknas p-värdet multiplicerat med 2

Exempel

- I en hälsoenkät tillfrågades 100 slumpmässigt utvalda anställda vid ett stort företag om huruvida man regelbundet motionerar eller ej.

Svar erhöles från 84 anställda och av dessa svarade 65 ja.

Undersök om det på 5 procents signifikansnivå finns belägg för påståendet att andelen regelbundna motionärer bland de anställda vid företaget understiger 85 procent genom att beräkna testets p-värde.

Relation mellan hypotesprövning och KI

- Om μ_0 el. π_0 täcks av intervallet kan H_0 inte förkastas
- Vid $H_a: <$ beräknas ett uppåt begränsat KI
- Vid $H_a: >$ beräknas ett nedåt begränsat KI
- Vid $H_a: \neq$ beräknas ett dubbelsidigt KI
- <http://rpsychologist.com/d3/CI/>

Feltyper och styrka

- Typ I-fel: att förkasta H_0 fastän den är sann
- Typ II-fel: att inte förkasta H_0 fastän den är falsk

- α är sannolikheten (risken) för typ I-fel

Feltyper och styrka

| Beslut baserat på stickprov | Sanning om populationen | | |
|-----------------------------|-------------------------|----------------|------------|
| | | H_0 sann | H_a sann |
| Förkasta H_0 | Typ I-fel | Korrekt beslut | |
| Inte förkasta H_0 | Korrekt beslut | Typ II-fel | |

- Det råder ett motsatsförhållande mellan risken för typ I-fel och risken för typ II-fel
- Inom samhällsvetenskaperna brukar man ange $\alpha = 0.05; 0.01; 0.10$ som ger en bra avvägning mellan typerna av fel