F6 - Konfidensintervall, hypotesprövning, p-värde

Statistisk översiktskurs

Ulf Högnäs

Statististiska institutionen
Stockholms universitet

April 3, 2025

Contents

Konfidensintevall och den centrala gränsvärdessatsen

Konfidensintervall och den centrala gränsvärdessatsen

Idé

- När stickprovsstorleken ökar närmar sig stickprovsmedelvärdets fördelning en normalfördelning
- Stickprovsmedelvärdets varians kan skattas med stor noggrannhet m.h.a. stickprovet
- Enligt 68-95-99.7-regeln hamnar Stickprovsmedelvärdet inom två standardavvikelser från populationsparametern med sannolikhet 95%
- vi skapar ett intervall centrerat kring stickprovsmedelvärdet med samma bredd (se kludd på whiteboard)
- detta intervall kommer att fånga populationsparametern med sannolikhet 95%

Kludd

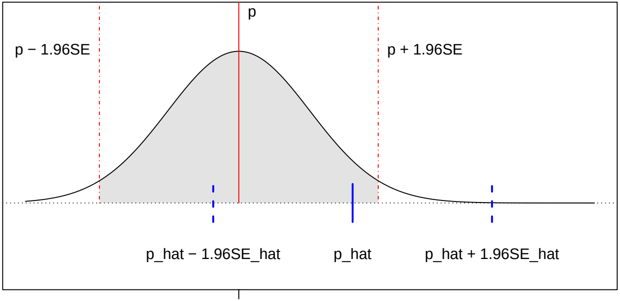


Figure: Med ungefär 95% sannolikhet hamnar stickprovsproportionen så nära den verkliga proportionen p att konfidensintervallet fångar p

Kludd

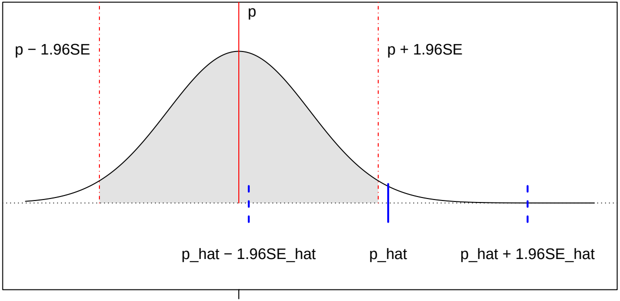


Figure: Med ungefär 5% sannolikhet hamnar stickprovsproportionen i en av stickprovsfördelningen svansar så att konfidensintervallet missar p

Standardfel och variation mellan stickprov

standardfel (SE)

Punktskattningar varierar från stickprov till stickprov, och denna variation kvantifieras med det som kallas **standardfel** (SE).

Standardfelet är detsamma som standardavvikelsen för den statistika vi undersöker. Det beskriver hur mycket ett punktskattning förväntas variera mellan olika stickprov.

Standardfelet är i sig en skattning, beräknad från stickprovsdata.
Hur vi bestämmer standardfelet beror på situationen, men oftast använder vi en formel som bygger på **centrala gränsvärdessatsen**.[1]

[1] Följande fem bilder följer boken nära, från s. 225

Felmarginal eller Margin of Error (ME)

Felmarginal

Avståndet $z^\star \times SE$ kallas för **felmarginal**.
där $z^\star$ ett värde från normalfördelningen. Det vanligaste värdet är $z^\star = 1.96$ (i 68-95-99.7-regeln är detta avrundat till 2). Detta ger en felmarginal som omfattar ungefär 95% av alla möjliga stickprovsskattningar

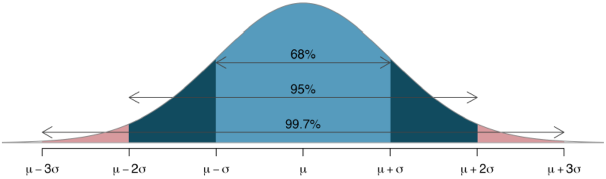


Figure: The 68-95-99.7 rule

Formel för ett 95%-igt konfidensintervall

Att konstruera ett 95%-konfidensintervall

När stickprovsfördelningen för en punktestimator (t.ex. $\hat{p}$ eller $\bar{x}$) rimligen kan antas vara normalfördelad, kommer det observerade värdet att ligga inom 1.96 standardfel från det sanna värdet ungefär 95% av gångerna. Därför kan ett 95%-konfidensintervall konstrueras som

$$\text{punktskattning} \pm 1.96 \times SE$$

Vi kan då vara 95% säkra på att detta intervall kommer att fånga det sanna värdet i populationen.

Propotioner och “the success-failure contition”

the success-failure contition

Stickprovsfördelningen för $\hat{p}$ (andel i stickprovet), baserad på ett stickprov av storlek n från en population med sann andel p, är ungefär normalfördelad när

1. Observationerna är oberoende, t.ex. från ett enkelt slumpmässigt urval.
2. Vi förväntar oss minst 10 lyckade och 10 misslyckade utfall: $np \geq 10$ och $n(1-p) \geq 10$

I så fall är $\hat{p}$ ungefär normalfördelad med

- Medelvärde p
- Standardfel $SE = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Formel för konfidensintervall för ett proportion

Konfidensintervall för en proportion, CGS

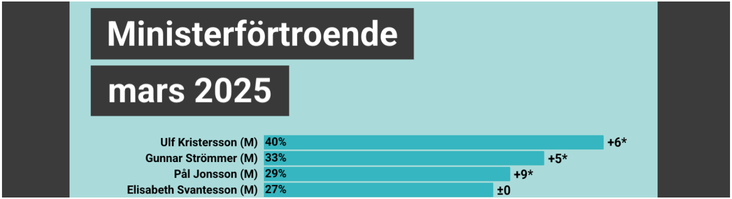
Om “the success-failure condition” är uppfyllt så kan det det 95%-iga konfidensintervallet för en proportion beräknas som

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

där $\hat{p}$ är stickprovsproportionen och n är stickprovsstorleken

Ett räkneexempel

- Novus undersökning omfattade 509 personer[2]
- "Vilket förtroende har du för följande ministrar?"
- 204 svarande gav Ulf Kristersson "ganska" eller "mycket stort"
- Beräkna ett 95%-igt konfidensintervall för andelen bland alla väljare som har ganska stort eller stort förtroende för Kristersson



[2] egentligen 1005 personer, men alla fick inte frågan om Kristersson

Lösning

1. Beräkna $\hat{p}$
$$\hat{p} = \frac{204}{509} \approx 0.4008$$

2. kontrollera "the success-failure condition"
$$n \cdot \hat{p} = 509 \cdot \frac{204}{509} = 204 \geq 10$$
$$n \cdot (1 - \hat{p}) = 509 \cdot \left(1 - \frac{204}{509}\right) = 305 \geq 10$$

OK. De förväntade antalen mycket-ganska-stort / inte mycket-ganska-stort i stickprovet är båda större än 10. Vi kan använda den centrala gränsvärdessatsen

Lösning

3. Använd formeln på bild 33

$$\hat{p} \pm 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.4008 \pm 1.96\sqrt{\frac{0.4008(1-0.4008)}{509}}$$

$$0.4008 \pm 0.04257$$

4. Beräkna den lägre och den övre gränsen, avrunda

$$(0.36, 0.44)$$

Vi kan säga med 95 procents konfidens att mellan 36% och 44% av väljarna hade mycket eller ganska stort förtroende för Kristersson, mars 2025

Statistiskt säkerställd förändring

Ministerförtroende, Mars 2025, Novus

- Mätningen visar på en ökning på $6^{\star}$ procentenheter för väljarnas förtroende för Kristerson sedan september 2024
- $\star$ är Novus markering för *statistiskt säkerställda förändringar jämfört med september 2024*
- Hur testas detta? Man måste ta hänsyn till den sammanlagda osäkerheten i de två undersökningarna
- Vi beräknar ett **konfidensintervall för skillnaden i andel** mellan de två mätningarna
- Vi skippar formel för denna typ av test och använder R istället

Konfidensintervall för skillnad i andel med R

Mätning	Förtroende	Tillfrågade
sep-2024	182	509
mar-2025	204	536

Table: Med "förtroende" menas de antal som svarade "mycket högt" eller "ganska högt"

- Punktskattningen för skillnaden i andel är $\frac{204}{536} - \frac{182}{509} \approx 0.0612$
- Vi använder

```
prop.test(c(204, 182), c(536, 509))
```

Där `c(204, 182)` är en vektor med de två täljarana
och `c(536, 509)` är en vektor med de två nämnarna

En statistiskt säkerställd förändring

```
> prop.test(c(204, 182), c(509, 536))

	2-sample test for equality of proportions with continuity correction

data:  c(204, 182) out of c(509, 536)
X-squared = 3.9437, df = 1, p-value = 0.04705
alternative hypothesis: two.sided
95 percent confidence interval:
 0.0008402383 0.1216269933
sample estimates:
   prop 1    prop 2
0.4007859 0.3395522
```

- Vi kan säga med 95% konfidens att skillnaden i andel, andel mars minus andel september, ligger i intervallet $(0.00084, 0.12)$, alltså mellan 0.084% och 12%
- Eftersom noll inte finns med i intervallet så säger vi att förändringen är **statistiskt säkerställd**

Hypotestest och den centrala gränsvärdessatsen

Från föreläsning 5

Noll- och alternativa hypoteser

Nollhypotesen H_0 (eng. the null hypothesis) det skeptiska perspektivet; hypotesen att ingen skillnad från ett visst skeptiskt antagande existerar i verkligheten

Den alternativa hypotesen H_A (eng. the alternative hypothesis) ett alternativt påstående som försöket eller undersökningen eventuellt kan hitta stöd för

Hypoteser, exempel

Exempel

Vi vill testa om Ulf Kristerssons förtroende är över 40%. Vi väljer hypoteserna

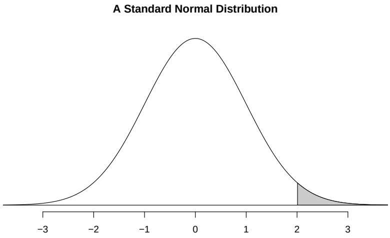
$$\begin{aligned} H_0: &\quad p = 0.4 \\ H_A: &\quad p > 0.4 \end{aligned}$$

där p är andelen av väljarna som har ganska stort eller mycket stort förtroende för Kristersson.

I en ny Novusundersökning svarar 220 av 500 tillfrågade "ganska stort" eller "mycket stort". Vad blir hypotestestets resultat?

Z-score

- Istället för att simulera med the randomization test procedure kommer vi att utnyttja central limit theorem
- När stickprovet är tillräckligt stort är ju stickprovsmedelvärdet approximativt normalfördelat
- För att få ett mått på hur ovanligt ett resultat kan man beräkna det som kallas **Z score**



Z-score

Definition

Z-score mäter avståndet till medelvärdet, räknat i standardavvikelser. För en observation x som följer en fördelning med medelvärde μ och standardavvikelse σ beräknas Z-score som

$$Z = \frac{x - \mu}{\sigma}$$

Om x är normalfördelad med medelvärde μ och standardavvikelse σ, så är Z också normalfördelad med medelvärde 0 och standardavvikelse 1.

Vad blir Z score om $x = \mu$?

Z-score och hypotestest

Z-score vid hypotestest

Vid hypotesprövning beräknas Z-score för en punktskattning som

$$Z = \frac{\text{punktskattning} - \text{nollvärde}}{\text{SE}}$$

där **SE** (standard error) är motsvarigheten till standardavvikelsen för punktskattningen, och **nollvärdet** kommer från påståendet i nollhypotesen.

Hypotestest med sannolikhetsmodell

Hypotestest för en andel.

1. Formulera hypoteser och välj gräns för statistisk signifikans.
2. Vi kontrollerar the success-failure contition.
   Om detta krav är uppfyllt kommer $\hat{p}$, som beräknas med hjälp av stickprovet, att vara approximativt normalfördelat med medelvärde p och standardavvikelse
   $$\sqrt{\frac{p(1-p)}{n}}$$
   Standardfelet approximeras med standardfelet
   $$SE = \sqrt{\frac{p_0(1-p_0)}{n}}$$
   Där p_0 är andelen som antas vara sann i nollhypotesen.

Metod

3. Beräkna $\hat{p}$ och $SE = \sqrt{\frac{p_0(1-p_0)}{n}}$.
4. Beräkna
$$Z = \frac{\text{punktskattning} - \text{nollvärde}}{\text{SE}}$$
I vårt fall
$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$
5. Använd R för att ta reda på hur ovanligt det observerade värdet $\hat{p}$ är om nollhypotesen är sann. Detta är vårt p-värde
6. Jämför p-värdet med gräns för statistisk signifikans och dra en slutsats

Lösning

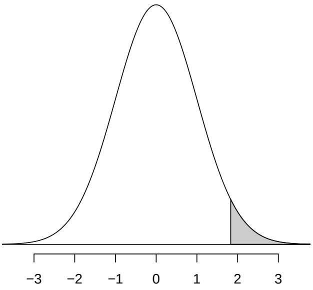
1. Hypoteserna är angivna i frågan. Vi väljer 5% som gräns.
2. Vi kontrollerar the success-failure condition (whiteboard)
3. Beräkna $\hat{p}$
$$\hat{p} = \frac{220}{500} = 0.44$$
4. Beräkna vår teststatistika z_{obs}
$$z_{obs} = \frac{0.44 - 0.4}{\sqrt{\frac{0.4(1-0.4)}{500}}} \approx 1.8257$$

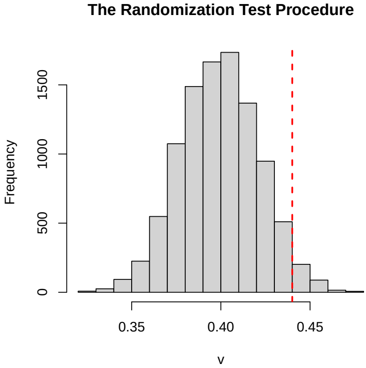
Lösning

5. Använd R

```
# Calculate and save z_obs
z_obs <- (.44-.4)sqrt(.4*(1-.4)/500)
# Look at the value
z_obs
# load the openintro package
library(openintro)
# Draw a picture
normTail(0, 1, U = z_obs)
# Find the probability
pnorm(z_obs, lower.tail = FALSE)
```

Jämförelse med the Randomization Test Procedure





Lösning, fortsättning

6. Vi hittade vårt p-värde

```
> pnorm(z_obs, lower.tail = FALSE)
[1] 0.03394458
```

Det är mindre än vår gräns på 5%. Vi förkastar nollhypotesen. Vi har hittat statistisk signifikanta bevis för att över 40% av väljarna har förtroende för stadsministern.

Hypoteser, en-sidiga och två-sidiga test

Exempel

- **Sex discrimination** Det finns en skillnad i sannolikheten för befordran och kvinnor missgynnas

$$\begin{aligned} H_0: &\quad p_M - p_K = 0 \\ H_A: &\quad p_M - p_K > 0 \end{aligned}$$

- **Sex discrimination** Det finns en skillnad i sannolikheten för befordran. Variabeln `sex` påverkar sannolikheten i <u>någon</u> riktning

$$\begin{aligned} H_0: &\quad p_M - p_K = 0 \\ H_A: &\quad p_M - p_K \neq 0 \end{aligned}$$

Om vi har tid

Vi använder `prop.test()` och `normTail()` för att prata om p-värden, en-sidiga och två-sidiga test.