

TMS136

Föreläsning 12

Hypotesprövning

- Vid hypotesprövning finns följande möjligheter

	H_0 är sann	H_0 är falsk
Inte förkasta H_0	Korrekt	Typ II-fel
Förkasta H_0	Typ I-fel	Korrekt

- Sannolikheten att göra Typ I-fel kallas signifikansnivå, betecknas α
- Sannolikheten att göra Typ II-fel, betecknas β , och vi kallar $1 - \beta$ testets styrka

Styrka hos hypotestest

- Styrkan, $1 - \beta$, är alltså sannolikheten att ta det korrekta beslutet att förkasta en falsk nollhypotes
- I praktiken är styrkan alltså sannolikheten att upptäcka en existerande avvikelse från nollhypotesen
- Vi vill naturligtvis att styrkan ska vara så stor som möjligt

Styrka hos hypotestest

- Vilka faktorer påverkar styrkan?
- Vi tänker oss ett exempel där vi tar stickprov från en normalfördelning med känd standardavvikelse $\sigma = 1$ och vill testa

$$H_0: \mu = 0$$

$$H_1: \mu > 0$$

på $\alpha = 0.05$ -nivå

Styrka hos hypotestest

- Vår testfunktion är (under nollhypotesen)

$$\frac{\bar{X}}{1/\sqrt{n}} = \sqrt{n}\bar{X}$$

- Då testfunktionen är standard normalfördelad förkastar vi nollhypotesen om vi observerar ett värde på kvoten som är större än 1.64 eller ekvivalent om vi observerar $\bar{X} > 1.64/\sqrt{n}$

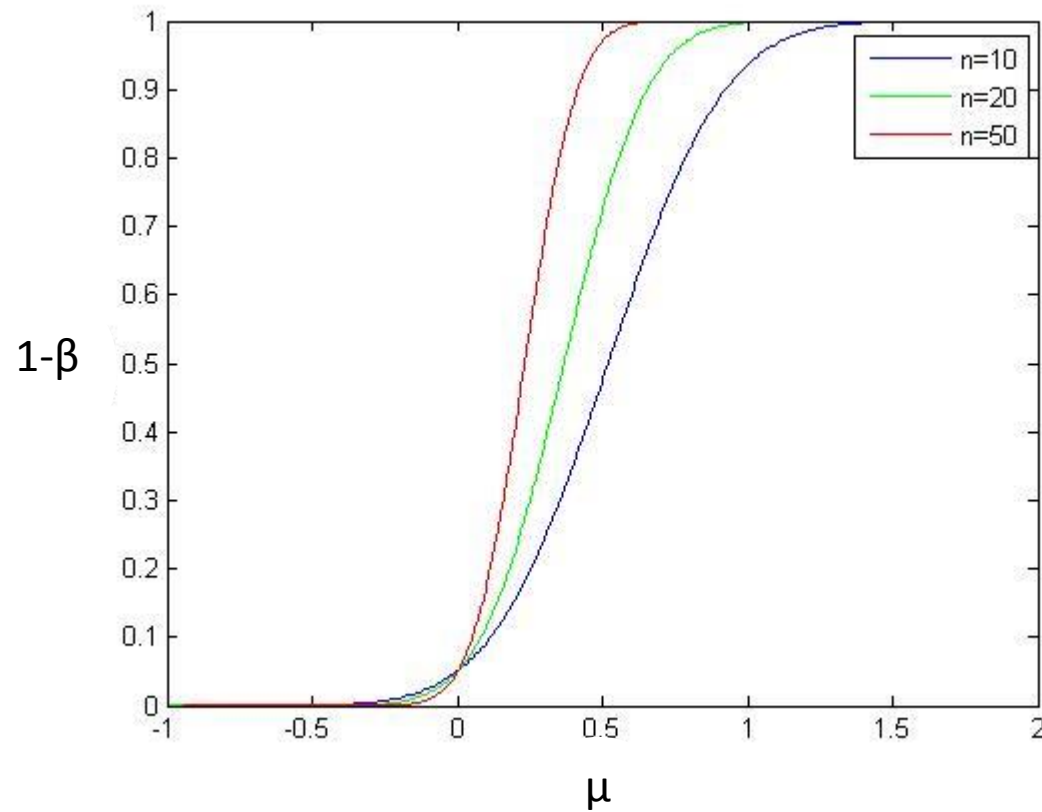
Styrka hos hypotestest

- Vi vill bestämma sannolikheten att förkasta en falsk nollhypotes (styrkan), och utgår därför från att $\mu > 0$, det vill säga mothypotesen är sann

- Vi får

$$\begin{aligned} 1 - \beta &= P(\bar{X} > 1.64/\sqrt{n}) = P(\bar{X} - \mu > 1.64/\sqrt{n} - \mu) \\ &= P(\sqrt{n}(\bar{X} - \mu) > 1.64 - \sqrt{n}\mu) \\ &= 1 - P(Z < 1.64 - \sqrt{n}\mu) \\ &= 1 - \Phi(1.64 - \sqrt{n}\mu) \end{aligned}$$

Styrka hos hypotestest



Styrka hos hypotestest

- Vi ser alltså att styrkan ökar ”ju mer falsk” nollhypotesen är, alltså då μ växer
- Ju större avvikelser från nollhypotesen är desto lättare är den att upptäcka
- Vi ser även att större stickprov ger större styrka

Styrka hos hypotestest

- Vi jobbar så här



Styrka hos hypotestest

- Alternativt kan man tänka sig att man först bestämmer vad man vill upptäcka, exempelvis vilken skillnad mellan före och efter en medicinsk behandling som anses kliniskt relevant
- Om vi specificerar den relevanta skillnaden, nivån α och styrkan $1 - \beta$ kan man beräkna hur stort stickprov som behövs för att med den valda styrkan påvisa den kliniskt relevanta skillnaden

Styrka hos hypotestest

- Vi tänker oss parvisa stickprov från normalfördelningar

$$\left((X_1, Y_1), \dots, (X_n, Y_n) \right)$$

från före och efter behandling av n individer

- För att analysera för skillnad bildar vi differenserna

$$(D_1 = Y_1 - X_1, \dots, D_n = Y_n - X_n)$$

Styrka hos hypotestest

- Våra hypoteser är med $\mu_d = \mu_y - \mu_x$

$$H_0: \mu_d = 0$$

$$H_1: \mu_d > 0$$

- Vår testfunktion är alltså den (under nollhypotesen) standard normalfördelade kvoten

$$\frac{\bar{D}}{\sigma_d/\sqrt{n}} = \frac{\frac{1}{n} \sum_{i=1}^n D_i}{\sigma_d/\sqrt{n}}$$

Styrka hos hypotestest

- Det är vanligt att man i medicinska sammanhang sätter den kliniskt relevanta skillnaden till en halv standardavvikelse
- Vi antar att $\alpha = 0.01$, $1 - \beta = 0.95$ och vill bestämma vilken stickprovsstorlek som behövs för att upptäcka den kliniskt relevanta skillnaden dvs $\mu_d = \sigma_d/2$

Styrka hos hypotestest

- Vi förkastar alltså nollhypotesen om vi observerar att testfunktionens värde är större än 2.33
(Appendix
- Styrkan ska vara 0.95 för $\mu_d = \sigma_d/2$ så

$$0.95 = P(\bar{D} > 2.33\sigma_d/\sqrt{n})$$

$$= P\left(\left(\bar{D} - \frac{\sigma_d}{2}\right) / \sigma_d/\sqrt{n} > \left(2.33\sigma_d/\sqrt{n} - \frac{\sigma_d}{2}\right) / \sigma_d/\sqrt{n}\right)$$

$$= P(Z > 2.33 - \sqrt{n}/2) = 1 - \Phi(2.33 - \sqrt{n}/2)$$

Styrka hos hypotestest

- Så enligt tabell 1 ska vi ha

$$2.33 - \frac{\sqrt{n}}{2} = -1.64$$

- Löser vi ekvationen får vi $n \approx 64$
- Vi behöver alltså ett stickprov av storlek 64 för att med den föreskrivna sannolikheten 0.95 upptäcka den kliniskt relevanta skillnaden en halv standardavvikelse.

Hypotestest för varians

- Antag att $X_1, \dots, X_n$ är ett stickprov från en normalfördelning med okänd varians σ^2
- Vi vill kanske göra hypotestestet

$$H_0: \sigma^2 = \sigma_0^2$$

$$H_1: \sigma^2 > \sigma_0^2$$

Hypotestest för varians

- För att utföra testet ska vi använda kvoten

$$\frac{(n-1)S^2}{\sigma_0^2}$$

där

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Hypotestest för varians

- Det gäller att

$$\frac{(n - 1)S^2}{\sigma_0^2}$$

är χ^2_{n-1} -fördelad

- Så för att testa hypoteserna ska det observerade värdet på kvoten (testfunktionen) jämföras med lämplig chi-2-kvantil

Hypotestest för varians (exempel)

- Vi tar ett stickprov av storlek 10 från en normalfördelning och observerar stickprovsvariansen 2.26

- Vi vill testa

$$H_0: \sigma^2 = 1.5$$

$$H_1: \sigma^2 > 1.5$$

på 0.10-nivå

Hypotestest för varians (exempel)

- Det observerade värdet på kvoten blir

$$\frac{(n-1)s^2}{\sigma_0^2} = \frac{9 \cdot 2.26}{1.5} \approx 13.56$$

- Den aktuella chi-2-kvantilen är (appendix)

$$\chi^2_{0.10,9} = 14.68$$

så vi kan inte förkasta nollhypotesen

Hypotestest för varians (exempel)

- Vi kan även testa tvåsidigt

$$H_0: \sigma^2 = 1.5$$

$$H_1: \sigma^2 \neq 1.5$$

- Om vi fortfarande har $\alpha = 0.10$ blir förkastelseområdet värden på kvoten mindre än $\chi^2_{0.95,9} = 3.33$ och värden på kvoten större än $\chi^2_{0.05,9} = 16.92$ och vi kan inte heller här förkasta nollhypotesen i exemplet

Goodness of fit

- *Goodness of fit* är ett test som kan användas för att undersöka om ett stickprovdata kommer från en viss fördelning
- Hur själva proceduren vid användandet ser ut beror på om man väljer att skatta parametrar från data eller gör antaganden kring parametrarna
- Nollhypotesen är alltid att data kommer från en viss fördelning och man jämför observerade och (under nollhypotesen) förväntade värden

Goodness of fit?

- Antag att vi tillfrågar 300 personer vilken av chokladbitarna Coco, Daim eller Japp de helst äter
- Vi tror att preferenserna skiljer sig åt och vi vill testa detta.
- Nollhypotesen blir att en tillfrågad lika gärna äter vilken som helst av bitarna, alltså att sannolikheten är $1/3$ för val av var och en av chokladbitarna

Coco, Daim eller Japp?

- De förväntade antalen under nollhypotesen blir att 100 väljer Coco, 100 väljer Daim och 100 väljer Japp
- Vi observerar att 73 helst äter Coco, 123 helst äter Daim och 104 helst äter Japp

Coco, Daim eller Japp?

- För vårt test skapar vi summan

The diagram illustrates the chi-square test formula with three terms. Each term consists of a squared difference between an observed count and an expected count, divided by the expected count. Blue arrows point from the text labels to the corresponding parts of the formula.

$$\frac{(73 - 100)^2}{100} + \frac{(123 - 100)^2}{100} + \frac{(104 - 100)^2}{100}$$

Observerade antal

Förväntade antal under nollhypotesen

= 12.74

Coco, Daim eller Japp?

- Summans värde ska utvärderas mot en χ^2 -kvantil med $3-1=2$ frihetsgrader (appendix)
- För signifikansnivån $\alpha = 0.05$ hittar vi $\chi^2_{2,0.05} = 10.60$ så vårt observerade 12.74 ligger i förkastelseregionen.
- Slutsatsen blir att folk hellre väljer en viss chokladbit över övriga

Mer generellt

- I goodness of fit-testet har vi

H_0 : data kommer från F

H_1 : data kommer inte från F

- Teststatistikan är

$$\sum_{i=1}^k \frac{(O_i - E_i)^2}{E_i}$$

där O_i och E_i är observerade respektive förväntade (under nollhypotesen) värden

Mer generellt

- Man kan visa att under nollhypotesen så är testfunktionen χ^2_{k-p-1} -fördelad där p är antal parametrar man skattar i F
- Vi tänker oss ett exempel där vi vill undersöka om antalet fel på en viss tillverkad produkt kan antas följa en Poisson-fördelning

Exempel

- Vi observerar 60 produkter och konstaterar att 32 av dessa inte har något fel, 15 har 1 fel, 9 har 2 fel och 4 har 3 fel
- Om vi ska undersöka att dessa antal verkar följa en Poisson-fördelning vill vi skatta λ i Poisson-fördelningen
- Vi vet att λ är väntevärdet i Poisson-fördelningen så vi skattar lambda med $\frac{(32 \cdot 0 + 15 \cdot 1 + 9 \cdot 2 + 4 \cdot 3)}{60} = 0.75$

Exempel

- Så om X är antalet fel har vi

$$P(X = 0) = e^{-0.75} \frac{0.75^0}{0!} \approx 0.472$$

$$P(X = 1) = e^{-0.75} \frac{0.75^1}{1!} \approx 0.354$$

$$P(X = 2) = e^{-0.75} \frac{0.75^2}{2!} \approx 0.133$$

$$P(X \geq 3) \approx 1 - (0.472 + 0.354 + 0.133) \approx 0.041$$

Vi approximerar alltså sannolikheten för minst tre fel eftersom vi har så lite information (få förekomster i stickprovet) om denna

Exempel

- Vi får nu att de förväntade felfrekvenserna under nollhypotesen blir

$$E_1 = 60 \cdot 0.472 \approx 28.32$$

$$E_2 = 60 \cdot 0.354 \approx 21.24$$

$$E_3 = 60 \cdot 0.133 \approx 7.98$$

$$E_4 = 60 \cdot 0.041 \approx 2.46$$

Eftersom det blir så få (färre än tre är tumregeln) i den sista kategorin kombinerar vi den med den näst sista så vi sätter alltså $E_3 = 7.98 + 2.46 = 10.44$

Exempel

- För testfunktionen observerar vi alltså värdet

$$\frac{(32 - 28.32)^2}{28.32} + \frac{(15 - 21.24)^2}{21.24} + \frac{(13 - 10.44)^2}{10.44} \approx 2.94$$

- Eftersom vi skattat en parameter (λ) är teststatistikan $\chi^2_{3-1-1} = \chi^2_1$ -fördelad
- Om vi väljer att testa på 0.05-nivå ser vi i appendix att $\chi^2_{0.05,1} = 3.84$ så vi kan inte förkasta nollhypotesen
- Vi har alltså inget skäl att tro antalet fel inte skulle kunna vara Poisson-fördelat