SF1901: Sannolikhetslära och statistik
Föreläsning 12.
Mer hypotesprövning. χ^2-test

Jan Grandell & Timo Koski

25.02.2016

KTH Matematik

Innehåll

- Hypotesprövning och konfidensmetoden; exempel på tvåsidga test

Innehåll

- Hypotesprövning och konfidensmetoden; exempel på tvåsidga test
- Hypotesprövning: allmänt

KTH Matematik

- Hypotesprövning och konfidensmetoden; exempel på tvåsidga test
- Hypotesprövning: allmänt
  - Nollhypotes,mothypotes, Teststorhet (testvariabel), Kritiskt område, Signifikansnivå.

INNEHÅLL

- Hypotesprövning och konfidensmetoden; exempel på tvåsidga test
- Hypotesprövning: allmänt
  - Nollhypotes,mothypotes, Teststorhet (testvariabel), Kritiskt område, Signifikansnivå.
  - signifikant*, *signifikant***, *signifikant****

INNEHÅLL

- Hypotesprövning och konfidensmetoden; exempel på tvåsidga test
- Hypotesprövning: allmänt
  - Nollhypotes,mothypotes, Teststorhet (testvariabel), Kritiskt område, Signifikansnivå.
  - signifikant*, *signifikant***, *signifikant****
  - p -värden

- Hypotesprövning och konfidensmetoden; exempel på tvåsidga test
- Hypotesprövning: allmänt
  - Nollhypotes,mothypotes, Teststorhet (testvariabel), Kritiskt område, Signifikansnivå.
  - signifikant*, *signifikant***, *signifikant****
  - p -värden
  - Styrkefunktion

Innehåll

- Hypotesprövning och konfidensmetoden; exempel på tvåsidga test
- Hypotesprövning: allmänt
  - Nollhypotes,mothypotes, Teststorhet (testvariabel), Kritiskt område, Signifikansnivå.
  - signifikant*, *signifikant***, *signifikant****
  - p -värden
  - Styrkefunktion
  - ensidiga test

KTH Matematik

Innehåll

- Hypotesprövning och konfidensmetoden; exempel på tvåsidga test
- Hypotesprövning: allmänt
  - Nollhypotes,mothypotes, Teststorhet (testvariabel), Kritiskt område, Signifikansnivå.
  - signifikant*, *signifikant***, *signifikant****
  - p -värden
  - Styrkefunktion
  - ensidiga test
- χ^2 -test (inledning)

Hypotesprövning: från påstående till procedur

- Vi har ett påstående om t.ex. ett populationsmedelvärde.

Exempel

Vid rökning omvandlas nikotin till cotinin, en metabolite av nikotin, som kan mätas. Det påstås att den genomsnittliga nivån av cotinin hos alla rökare är 200.0 ng/mL.

- Vi kollar nivån av cotinin hos n rökare, och får cotininvärdena $x_1, \ldots, x_n$.
- Vi använder en statistisk procedur för att checka om påståendet om medelnivån 200.0 kan anses vara förenlig med dessa mätdata.
- Procedur: inför en statistisk modell, teststorhet, kritiskt område, beslutsregel, signifikansnivå och använd 'rare event principle'

KTH Matematik

Statistisk Hypotesprövning: rare event rule

If, under a given assumption, the probability of an observed event is very small, we conclude that the assumption is likely not correct.

$X_1, X_2, \ldots, X_n$ är oberoende och $N(\mu, \sigma)$-fördelade, där μ och σ är okända.

$$H_0: \; \mu = \mu_0$$

mot

$$H_1: \; \mu \neq \mu_0$$

Tvåsidiga test

$X_1, X_2, \ldots, X_n$ är oberoende och $N(\mu, \sigma)$-fördelade, där μ och σ är okända.

$$H_0: \ \mu = \mu_0$$

mot

$$H_1: \ \mu \neq \mu_0$$

KTH Matematik

Här verkar det rimligt att utgå från

$$t(X_1, \ldots, X_n) = \frac{\overline{X} - \mu_0}{S/\sqrt{n}},$$

som under H_0 är $t(n-1)$-fördelad, och att förkasta H_0 om $|t(x_1, \ldots, x_n)|$ är för stor.
Vi får då

$$\alpha = P(|t(X_1, \ldots, X_n)| > c \text{ om } H_0 \text{ sann}),$$

vilket ger $c = t_{\alpha/2}(n-1)$.

Kritiska området är alltså $|t(x_1,\dots,x_n)| > t_{\alpha/2}(n-1)$

$$\Leftrightarrow \begin{cases} t(x_1,\dots,x_n) > t_{\alpha/2}(n-1) & \text{om } t(x_1,\dots,x_n) > 0 \text{ (A)} \\ -t(x_1,\dots,x_n) > t_{\alpha/2}(n-1) & \text{om } t(x_1,\dots,x_n) < 0 \text{ (B)} \end{cases}$$

Fall (A):

$$t(x_1,\dots,x_n) > t_{\alpha/2}(n-1) \Leftrightarrow \frac{\bar{x}-\mu_0}{s/\sqrt{n}} > t_{\alpha/2}(n-1)$$

$$\Leftrightarrow \bar{x}-\mu_0 > t_{\alpha/2}(n-1)s/\sqrt{n} \Leftrightarrow \mu_0 < \bar{x} - t_{\alpha/2}(n-1)s/\sqrt{n}$$

P.s.s. fås i fallet (B) att

$$\mu_0 > \bar{x} + t_{\alpha/2}(n-1)s/\sqrt{n}$$

KTH Matematik

Testet kan alltså utföras så att man bestämmer ett konfidensintervall och ser efter om det *hypotetiska värdet* μ_0 ligger utanför eller innanför detta; i det förra fallet förkastas H_0, inte i det senare.

- Bilda ett konfidensintervall $I_\mu = \left(\bar{x} - t_{\alpha/2}(f)s/\sqrt{n}, \bar{x} + t_{\alpha/2}(f)s/\sqrt{n}\right)$, $f = n - 1$ och förkasta H_0 om

$$I_\mu \not\ni \mu_0.$$

Hypotesprövning: sammanfattning av exemplet

$X_1, X_2, \ldots, X_n$ är oberoende och $N(\mu, \sigma)$-fördelade, μ och σ okända.

$$H_0: \ \mu = \mu_0$$

mot

$$H_1: \ \mu \neq \mu_0.$$

- Bilda ett konfidensintervall $I_\mu = \left(\bar{x} - t_{\alpha/2}(f)d, \bar{x} + t_{\alpha/2}(f)d\right)$ och förkasta H_0 om
$$I_\mu \not\ni \mu_0.$$
- Detta verkar rimligt. I_μ ger ju de "troliga" värdena på μ, och om μ_0 inte hör dit, så bör ju H_0 förkastas.

Ex. Stickprov i par

$$H_0 : \ \Delta = 0$$

mot

$$H_1 : \ \Delta \neq 0.$$

$$\underline{I_\Delta = \bar{z} \pm \lambda_{\alpha/2}\sigma/\sqrt{n}}$$

om σ känd och

$$\underline{I_\Delta = \bar{z} \pm t_{\alpha/2}(n-1)s/\sqrt{n}}$$

om σ okänd. Förkasta H_0 om

$$I_\Delta \not\ni 0.$$

- En forskare mäter en storhet θ som på grund av ett mätfel är $N(0, \sigma)$, och får värdet x, dvs. x är således en observation från $N(\theta, \sigma)$.
- Forskaren vill pröva en viss hypotes

$$H_0 : \theta = 2.0.$$

  Signifikansnivån skall vara 0.05.
- Som testvariabel väljs det erhållna mätvärdet x $(= t(x))$.
- Kritiska området blir (se föreläsning 10)

$$\begin{array}{lll} \text{Om} & |x - 2.0| \geq 1.96 \cdot \sigma & \text{förkasta } H_0 \\ \text{Om} & |x - 2.0| < 1.96 \cdot \sigma & \text{förkasta ej } H_0. \end{array}$$

KTH Matematik

Men $(x - 1.96\sigma, x + 1.96\sigma)$ är ett tvåsidigt 95 % konfidensintervall för θ. Testet kan alltså utföras så att man bestämmer konfidensintervallet och ser efter om det *hypotetiska värdet* θ_0 ligger utanför eller innanför detta; i det förra fallet förkastas H_0, inte i det senare.

Låt oss betrakta den allmänna situationen, dvs.:
Vi har en uppsättning data $x_1, x_2, \ldots, x_n$ som ses som utfall av s.v.

$$X_1, X_2, \ldots, X_n.$$

Dessa variabler antages vara oberoende och likafördelade och deras gemensamma fördelning beror av en okänd parameter θ.

En hypotes är en mängd av θ-värden. Formaliserat betyder detta att vi vill testa en *nollhypotes*

$$H_0: \ \theta \in H_0$$

mot ett *alternativ* (eller en mothypotes)

$$H_1: \ \theta \in H_1.$$

Att testa H_0 är detsamma som att avgöra om våra data är "förenliga" med H_0.

- Om H_0 ej är sann vill vi *förkasta* H_0 till förmån för H_1.

Att testa H_0 är detsamma som att avgöra om våra data är "förenliga" med H_0.

- Om H_0 ej är sann vill vi *förkasta* H_0 till förmån för H_1.
- Vi bildar därför en *teststorhet* $t = t(x_1, \ldots, x_n)$

Att testa H_0 är detsamma som att avgöra om våra data är "förenliga" med H_0.

- Om H_0 ej är sann vill vi *förkasta* H_0 till förmån för H_1.
- Vi bildar därför en *teststorhet* $t = t(x_1, \ldots, x_n)$
- och ett *kritiskt område* C.

- Test: Förkasta H_0 om $t \in C$.

Hypotesprövning: signifikansnivå

- Test: Förkasta H_0 om $t \in C$.
- t bestäms av situationen och C av *signifikansnivån* (eller felrisken) α:

$$\text{signifikansnivån} = \alpha \geq P(H_0 \text{ förkastas om } H_0 \text{ sann})$$

$$= P(T(X_1, \ldots, X_n) \in C \text{ om } H_0 \text{ sann}).$$

KTH Matematik

Man arbetar med flera signifikansnivå er samtidigt, t.ex. med nivåerna 0.05, 0.01 och 0.001. Kodbeteckningarna *signifikant**, *signifikant*** respektive *signifikant**** är för att markera att ett resultat är signifikant på en nivå (men inte för ett lägre α-värde bland dessa tre). Med denna symbolik betyder alltså signifikant* att $\alpha = 0.05$ ger signifikans men inte $\alpha = 0.01$, alltså inte heller $\alpha = 0.001$.

HYPOTESPRÖVNING: p-VÄRDET

Lå t oss betrakta ett test sådant att hypotesen H_0 förkastas om testvariabeln $t = t_{\text{obs}} = t(x)$ är "stor".

Vi förutsätter också att stickprovsvariabeln $t(X)$ har en given känd sannolikhetsfördelning under H_0. *p-värdet* eller *observerade signifikansnivån* är sannolikheten

$$p = P\big(t(X) \geq t(x)\big),$$

beräknad under förutsättningen att H_0 är sann. Om p-värdet är "tillräckligt litet" förkastas H_0; vi tror alltså inte på H_0 om testresultatet är "osannolikt" då H_0 är sann.

KTH Matematik

- Med risknivån garderar vi oss således mot felet att förkasta H_0 då H_0 är sann.

- Med risknivån garderar vi oss således mot felet att förkasta H_0 då H_0 är sann.
- Vi bör välja H_0 så att detta är det allvarligaste felet.

- Med risknivån garderar vi oss således mot felet att förkasta H_0 då H_0 är sann.
- Vi bör välja H_0 så att detta är det allvarligaste felet.
- Det andra möjliga felet är att ej förkasta H_0 då H_0 är falsk. Vi bildar *styrkefunktionen*

$$h(\theta) = P(H_0 \text{ förkastas}) \quad \text{om } \theta \text{ är det sanna värdet.}$$

KTH Matematik

$$h(\theta) = P(H_0 \text{ förkastas}) \quad \text{om } \theta \text{ är det sanna värdet.}$$

- För $\theta \in H_0$ gäller således att $h(\theta) \leq \alpha$.
- Ett test är "bra" om $h(\theta)$ är stor då $\theta \in H_1$.

För att få något bestämt ta $\sigma = 0.04$ Antag forskaren vill pröva hypotesen $H_0 : \theta = 2.0$. Låt den alternativa hypotesen vara

$$H_1 : \theta \leq 1.9 \text{ eller } \theta \geq 2.1.$$

Forskaren är angelägen om att verkligen förkasta H_0 om θ skulle avvika så mycket från 2.0 som H_1 anger. Vi undersöker hur stor sannolikheten är att forskaren gör detta.

Vi tergår til exemplet ovan: Testet lyder ju nu: Förkasta H_0 om

$$|x - 2.0| > 1.96 \cdot 0.04.$$

Alltså blir

$$h(\theta) = P(|X - 2.0| > 1.96 \cdot 0.04) \quad \text{om } X \in N(\theta, 0.04).$$

Den komplementära sannolikheten att X ligger mellan gränserna $2.0 \pm 1.96 \cdot 0.04$:

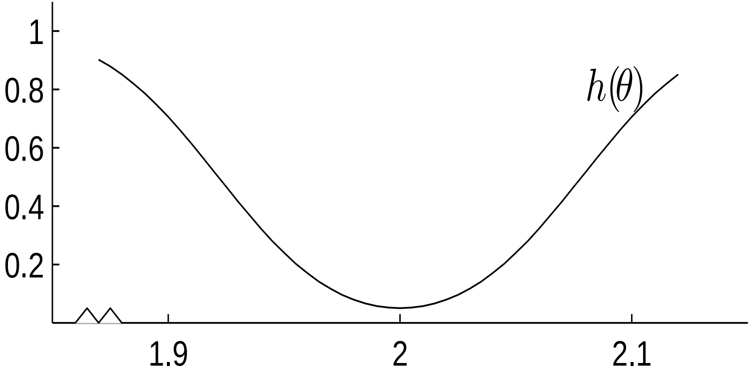
$$1 - h(\theta) = \Phi\left(\frac{2.0 + 1.96 \cdot 0.04 - \theta}{0.04}\right) - \Phi\left(\frac{2.0 - 1.96 \cdot 0.04 - \theta}{0.04}\right)$$

varav

$$h(\theta) = 1 - \Phi\left(1.96 + \frac{2.0 - \theta}{0.04}\right) + \Phi\left(-1.96 + \frac{2.0 - \theta}{0.04}\right).$$

Styrkefunktion (tvåsidigt test).

Denna styrkefunktion har har minimum (= 0.05, vilket är den fastställda signifikansnivån) för $\theta = 2.0$ och att den är ungefär 0.70 eller däröver för θ-värden inom H_1. Styrkan är inte särskilt stor, det finns två utvägar att förbättra den: att öka mätprecisionen dvs. minska σ eller utföra experiment med flera observationer.



Om H_0 omfattar flera θ-värden kommer sannolikheten $P(H_0$ förkastas$)$, om H_0 är sann, att bli beroende av vilket θ-värde inom H_0 som är det rätta, d.v.s. vår gamla beskrivning av α fungerar inte längre. Därför sätter vi i stället

$$\alpha = \max_{\theta \in H_0} h(\theta).$$

Vi låter denna högsta felrisk vara den på förhand bestämda signifikansnivån. Om vi tar t.ex. $\alpha = 0.01$, är risken att vi förkastar H_0, om H_0 är sann, högst lika med 0.01. Problemet är alltså att komma underfund med för vilket $\theta \in H_0$ som felrisken är störst, och det brukar vara lätt.

Ett test till

Ett mynt ger krona med den okända sannolikheten p. Man vill pröva $H_0 : p = 1/2$ mot $H_1 : p \neq 1/2$. Myntet kastas $n = 30$ gånger, varvid krona kommer upp $x = 5$ gånger.
Fixera signifikansnivån till 0.01. Vi skall förkasta H_0 om $x \leq a$ eller $x \geq b$. a och b väljs så att, om H_0 sann, $P(X \leq a) \lesssim 0.005$, $P(X \geq b) \lesssim 0.005$.

KTH Matematik

Ett test till

En tabell över binomialfördelningen ger $a = 7$, $b = 30 - 7 = 23$. ($P(X \leq a) = P(X \geq b) = 0.0026$ signifikansnivån är 0.0052) I det angivna fallet föreligger alltså signifikans på den valda nivån, ty $x \leq 7$.

Låt oss som övning alternativt använda P-värdet (jämför ovan).

$$P/2 = \sum_{i=0}^{5} \binom{30}{i} \left(\frac{1}{2}\right)^{i} \left(\frac{1}{2}\right)^{30-i}$$

varav $P = 0.00032$. Resultatet är så ledes signifikant***, och vi förkastar alltså hypotesen att myntet är välgjort.

$X_1, X_2, \ldots, X_n$ är oberoende och $N(\mu, \sigma)$-fördelade, där μ och σ är okända.

$$H_0: \ \mu = \mu_0$$

mot

$$H_1: \ \mu > \mu_0 \quad \text{eller } \mu < \mu_0.$$

Låt oss anta att stort värde på μ är en önskad egenskap. Det kan vara naturligt att vi gör en åtgärd, t.ex. köper någon ny utrustning, som bör öka värdet på μ.
Det är naturligt att vi endast vill köpa denna nya utrustning om vi är någolunda säkra på att den verkligen ger ett högre värde på μ än μ_0

Det är då naturligt att testa

$$H_0: \ \mu = \mu_0$$

mot

$$H_1: \ \mu > \mu_0.$$

Ensidiga test

Testet blir då att vi förkastar H_0 om $t(x_1, \ldots, x_n)$ är för stor, eller mera precist om

$$t > t_\alpha(n-1) \quad \text{eller om } \bar{x} > \mu_0 + t_\alpha(n-1)s/\sqrt{n}.$$

Tolkningen är att vi kräver, för att förkasta H_0, att $\bar{x}$ är tillräckligt mycket större än μ_0 för att det inte ska vara troligt att skillnaden är slumpmässig.

KTH VETENSKAP OCH KONST

KTH Matematik

Ensidiga test

Givetvis skulle vi mycket väl kunna vilja påstå att $\mu = \mu_0$, och då skulle vi ju vilja testa $H_0: \ \mu \neq \mu_0$ mot $H_1: \ \mu = \mu_0$.

Detta går inte, eftersom inga observationer i världen skulle kunna få oss att förkasta detta H_0.

Den som gör ett test, "vill" därför ofta att H_0 ska förkastas. Det är nog detta som gör att begreppet signifikant misstolkas.

KTH Matematik

χ^2-testet är ett så kallat **"goodness of fit"**[1]-test

Den enklaste situationen:

Ett försök kan utfalla på r olika sätt: $A_1, A_2, \ldots, A_r$. Låt $x_1, x_2, \ldots, x_r$ vara antalet gånger som alternativen $A_1, A_2, \ldots, A_r$ förekommer i n försök.

[1] **En källa översätter detta som "anpassningsgrad".**

Låt $p_1, p_2, \ldots, p_r$ vara givna sannolikheter, dvs $\sum_{i=1}^{r} p_i = 1$. Vi vill testa

$$H_0: \; P(A_i) = p_i \text{ för } i = 1, \ldots, r$$

mot

$$H_1: \; \text{ej alla } P(A_i) = p_i.$$

För att göra detta bildar vi

$$Q_{\text{obs}} = \sum_{i=1}^{r} \frac{(x_i - np_i)^2}{np_i}.$$

Man kan visa att Q är approximativt $\chi^2(r-1)$-fördelad under H_0.

För att göra resultatet troligt, betraktar vi $r = 2$. Då gäller, med $X = X_1$ och $p = p_1$ att

$$Q = \frac{(X_1 - np_1)^2}{np_1} + \frac{(X_2 - np_2)^2}{np_2} = \frac{(X - np)^2}{np} + \frac{(n - X - n(1-p))^2}{n(1-p)}$$

$$= \frac{(X - np)^2}{np} + \frac{(X - np))^2}{n(1-p)} = \frac{(X - np)^2}{np(1-p)} .$$

Eftersom X är $\mathrm{Bin}(n, p)$ så gäller att $\frac{X-np}{\sqrt{np(1-p)}}$ är appr. $N(0, 1)$. Således följer att $\frac{(X-np)^2}{np(1-p)}$ är appr. $\chi^2(1)$.

Vi gör nu följande test:

Förkasta H_0 om

$$Q_{\text{obs}} > \chi^2_\alpha(r-1).$$

Om n är stort, har detta test approximativt signifikansnivån α.

Det går inte att generellt ange hur stort n skall vara för att approximationen skall vara tillfredsställande. Som tumregel brukar man använda: Tillse att insatta värden på np_j är minst lika med 5.

Vid 96 kast med en tärning erhölls följande antal ettor, tvåor, etc: 15, 7, 9, 20, 26, 19. Man önskar pröva om tärningen kan antas vara symmetrisk. Hypotesen H_0 blir alltså

$$H_0 : P(A_1) = 1/6, \dots, P(A_6) = 1/6.$$

χ^2-metoden kan användas, ty alla $np_j = 96 \cdot 1/6 = 16$ och är alltså tillräckligt stora.

χ^2-test: Kast med en tärning

Man får

$$Q_{\text{obs}} = \frac{(15-16)^2}{16} + \frac{(7-16)^2}{16} + \frac{(9-16)^2}{16} + \frac{(20-16)^2}{16} + \frac{(26-16)^2}{16} + \frac{(19-16)^2}{16} = 16.0.$$

Antalet frihetsgrader är $f = r - 1 = 6 - 1 = 5$. Eftersom Q_{obs} något överstiger $\chi^2_{0.01}(5) = 15.1$ kan man påstå att tärningen inte är symmetrisk: Resultatet är signifikant**. Med dator erhålls $P = P(Q > 16.0) \approx 0.0068$ då Q är $\chi^2(5)$.

KTH Matematik

χ^2-TEST MED SKATTNING AV PARAMETRAR

Ofta vill vi låta sannolikheterna $p_1, p_2, \ldots, p_r$ bero av en okänd parameter $\theta = (\theta_1, \ldots, \theta_s)$, och testa hypotesen

$$H_0 : \; P(A_i) = p_i(\theta), \; \text{för } i = 1, \ldots, r,$$

och för något värde på θ.

χ^2-TEST MED SKATTNING AV PARAMETRAR

Skattar vi θ med ML-metoden, och bildar

$$Q_{\text{obs}} = \sum_{i=1}^{r} \frac{(x_i - np_i(\theta^*_{\text{obs}}))^2}{np_i(\theta^*_{\text{obs}})},$$

så är Q approximativt $\chi^2(r-s-1)$-fördelad under H_0.
Detta resultat kallas ibland för *stora χ^2-satsen*.

χ^2-TEST MED SKATTNING AV PARAMETRAR: ANTALET FRIHETSGRADER

Grundregeln är att antalet frihetsgrader fås av

$$\text{antalet fria kvadratsummor} - \text{antalet skattade parametrar.}$$