Matematisk statistik V
Föreläsningsanteckningar
Tommy Norberg
4 maj 2004

F11: Ch 8 Statistisk testteori

Nyckelord: noll och alternativhypotes, typ I och typ II-fel, signifikansnivå och styrka, teststatistika, P-värde

Dagens föreläsning ska handla om hur man med statistiska metoder "bevisar" påståenden. Man gör det med statistiska (hypotes-) test och p g a att vi analyserar slumpmässiga fenomen är vi tvingade att acceptera att det faktiskt kan hända att vi felaktigt tror oss om att ha bevisat något. Vi ska lära oss hur man gör ett statistiskt test, så att risken att man felaktigt tror något viktigt är liten. Denna risk brukar kallas för testets nivå och betecknas α.

I en statistisk testsituation kan man också råka utför att man p g a slumpen inte lyckas att bevisa något som faktiskt är sant. Detta är en annan typ av fel och risken att det inträffar ska vi beteckna β.

Tyvärr går det inte utan att öka stickprovets storlek att samtidigt skydda sig mot dessa båda typer av fel.

Antag att vi har 2000 längdskidåkare vid en tävling. Under tävlingen genomförs en dopingkontroll och 850 prövas positivt för doping. Proportionen dopade var alltså

$$\hat{p} = 0.425$$

Kan man p.g.a detta resultat påstå att färre än 50% av alla längdskidåkare dopar sig?

Lösning: Om $p = P(\text{en godtycklig åkare dopar sig}) = 0.5$, så följer att antalet dopade åkare i tävlingen är binomial$(2000, 0.5) \approx \mathrm{N}(1000, 22.36)$. Sannolikheten att 850 eller färre åkare dopar sig är under detta antagande är lika med

$$\sum_{x=0}^{850} \binom{2000}{k} \left(\frac{1}{2}\right)^k \left(1 - \frac{1}{2}\right)^{2000-k}$$

$$\approx P\left(Z \leq \frac{850-1000}{22.36}\right) = P\left(Z \leq -6.71\right) \approx 0$$

Att så få som 850 dopar sig är alltså väldigt osannolikt om det vore så att $p = 0.5$. Slutsatsen vi drar ifrån denna beräkning är att antagandet är felaktigt. Det är alltså inte så att hela 50% av skidåkarkollektivet dopar sig.

Antag att Pontus Foxtrot vill slå vad med dig om att han med tärningen han håller i handen kommer att kasta minst 25 sexor på 100 försök. Pontus vill satsa 10 USD och undrar lite småtyket om du vågar satsa lika mycket på att han misslyckas?

Lösning: Du vet att antalet sexor x är binomialfördelat med parametrar $n = 100$ och $p = 1/6$, och räknar snabbt ut att

$$\mu = np = 100/6 \approx 16.7$$

samt att

$$\sigma = \sqrt{np(1-p)} = \sqrt{500/36} \approx 3.73$$

Så $x \geq 25$ innebär faktiskt att Pontus räknar med att slå fler är $\mu + 2\sigma$ och du vet att centrala gränsvärdessatsen säger att denna sannolikhet är så liten som ≈ 0.025. Pontus chans att vinna vadet är alltså inte större än ca $1/40$.

Det är lätt att frestas att sätta emot om det inte vore så att Pontus är känd för att gärna luras. Så slutsatsen av kalkylen ovan är att Pontus antagligen har en osymmetrisk tärning.

Men du sätter emot ändå, för 10 USD är ju inte så himla mycket och det kunde kanske vara kul att se hur det går.

Hur det gick? Jo, Pontus slår 28 sexor på sina 100 kast och vinner dina 10 dollar.

Fuskade Pontus?

Om han gjorde det, hur ska vi då bevisa det utan att dissekrera tärningen. Vi kan gott anta att den ser helt normal ut.

Här följer ett allmänt tillvägagångssätt: Vår utgångspunkt H_0 är att tärningen är symmetrisk, vilket bl a innebär att $P(\text{sexa}) = 1/6$, och att antalet sexor x är binomialfördelad med parametrar $n = 100$ och $p = 1/6$. Centrala gränsvärdessatsen talar om för oss att x är approximativt normalfördelad med väntevärde $\mu \approx 16.7$ och standardavvikelse $\sigma \approx 3.72$.

Sannolikheten för ett resultat som är minst lika extremt som det Pontus fick, är

$$P(X \geq 28) = P(X > 27.5)$$
$$= P\left(\frac{X - 16.67}{3.73} > \frac{27.5 - 16.67}{3.73}\right)$$
$$= P(Z > 2.91) = 1 - 0.9982 \approx 0.002$$

(Obs Z betecknar en variabel som är standardiserat normalfördelad.)

Slutsatsen av denna kalkyl är: antingen är tärningen OK och något väldans osannolikt har inträffat; eller så är tärningen osymmetrisk. Den som är införstådd med Pontus allmänna karaktär föredrar nog att tro det sistnämnda och anse att kalkylen faktiskt bevisar att Pontus fuskade.

Men helt säkra kan vi inte vara. Risken att vi felaktigt anklagar Pontus för falskspel är så låg som $\approx 0.002 = 0.2\%$. Den risken att ha fel kan man nog acceptera.

Sannolikheten 0.002 kallas för *P-värdet* eller den *observerade signifikansnivån.*

Man brukar formulera den här typen av problem m.h.a. två hypoteser kallade *noll- och alternativhypotesen.* Se rutan längst ned på sidan 326.

The **null hypothesis**, denoted by H_0, is the claim about one or more population or process characteristics that is initially assumed to be true. The **alternative hypothesis**, denoted by H_a, is the claim that is contradictory to H_0.

The null hypothesis will be rejected in favor of the alternative hypothesis only if sample evidence suggests that H_0 is false. If the sample does not strongly contradict H_0, we will continue to believe in the truth of the null hypothesis.

The two possible conclusions from a hypothesis-testing analysis are then *reject* H_0 or *fail to reject* H_0.

Observera att alternativhypotesen är det man vill påvisa med den statistiska utredningen. Nollhypotesen representerar normalläget och är ofta, men inte alltid, komplementet till alternativhypotesen. Man sägar att man *testar* H_0 *mot* H_a.

I skidåkarexemplet är alternativhypotesen $H_a : p > 0.5$ och nollhypotesen $H_0 : p = 1/2$.

I exemplet med Pontus Foxtrot är alternativhypotesen $H_a : p \neq 1/6$ eller möjligen $H_a : p > 1/6$ och nollhypotesen är $H_0 : p = 1/6$.

Efter det att noll- och alternativhypotesen är formulerade räknar man ut, under antagandet att nollhypotesen är sann, hur starkt data tyder på alternativet. Man antingen *förkastar nollhypotesen* och accepterar då samtidigt alternativhypotesen, eller så *misslyckas man att förkasta* beroende på resultatet av beräkningen.

Definition på sida 328.

A **type I error** is the error of rejecting H_0 when H_0 is actually true.

A **type II error** consists of not rejecting H_0 when H_0 is false.

Definition på sida 329:

Maximala sannolikheten att begå ett typ-I-fel skrives α och kallas för testets *signifikansnivå*. Så att om man t.ex. väljer att testa på "nivån" $\alpha = 0.01$, så ska risken att felaktigt förkasta en sann nollhypotes vara ≤ 0.01.

Sannolikheten att göra ett typ-II-fel skrives β.

$1 - \beta$ brukar man kalla testets *styrka*.

Rutan om P-värde på sidan 331:

P-värdet eller den observerade signifikansnivån är sannolikheten för ett minst lika extremt värde som det vi erhöll (beräknad under antagandet att H_0 är sann). Ju lägre P-värdet är, desto mer osannolikare är nollhypotesen.

Ofta bestämmer man sig på förhand för att testa på en viss nivå α. Man bestämmer sig alltså för hur stor risk man är beredd att acceptera för att göra ett typ I-fel (d v s att felaktigt förkasta en sann nollhypotes). Om det erhållna P-värdet $\leq \alpha$ förkastas H_0. Då räknar vi med att alternativet H_a är sannt och vet att risken för att detta ska vara fel är $\leq \alpha$.

Exempel 8.4 på sidorna 331-332. (Se OH)

Rutan på sidan 338 (”The one-sample t-test”). (se OH)

Förutsättning: n st oberoende $\mathrm{N}(\mu, \sigma)$-observationer $x_1, \ldots, x_n$

Beräkna $\bar{x},\ s$ och sedan $t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

Teststatistika: $T = \dfrac{\bar{X} - \mu_0}{S/\sqrt{n}} \sim t(n-1)$

Test av $H_0 : \mu \leq \mu_0$ mot $H_a : \mu > \mu_0$
förkasta H_0 då $P(T > t) \leq \alpha$

Test av $H_0 : \mu \geq \mu_0$ mot $H_a : \mu < \mu_0$
förkasta H_0 då $P(T < t) \leq \alpha$

Test av $H_0 : \mu = \mu_0$ mot $H_a : \mu \neq \mu_0$
förkasta H_0 då $P(T > |t|) \leq \alpha/2$

Exempel 8.5 på sida 339. (Se OH)

Rutan på sidan 341: ”The two-sample t-test”

Förutsättning: Ett stickprov bestående av n_1 st oberoende $\mathrm{N}(\mu_1, \sigma_1)$-observationer och ett stickprov bestående av n_2 st oberoende $\mathrm{N}(\mu_2, \sigma_2)$-observationer. De två stickproven ska vara oberoende.

Data: $n_1,\ \bar{x}_1,\ s_1$ samt $n_2,\ \bar{x}_2,\ s_2$

Teststatistika: $T = \dfrac{\bar{X}_1 - \bar{X}_2}{\sqrt{S_1^2/n_1 + S_2^2/n_2}} \sim t(\mathrm{df})$ där

$$\mathrm{df} \approx \frac{\left(s_1^2/n_1 + s_2^2/n_2\right)^2}{\dfrac{s_1^2/n_1}{n_1 - 1} + \dfrac{s_2^2/n_2}{n_2 - 1}}$$

Test av $H_0 : \mu_1 \leq \mu_2$ mot $H_a : \mu_1 > \mu_2$
förkasta H_0 då $P(T > t) \leq \alpha$

Test av $H_0 : \mu_1 = \mu_2$ mot $H_a : \mu_1 \neq \mu_2$
förkasta H_0 då $P(T > |t|) \leq \alpha/2$

Exempel 8.6 på sidorna 341-344. (Se OH)