Föreläsning 10: Jämförelser

Matematisk statistik

David Bolin
Chalmers University of Technology
September 30, 2015

Hypotesprövning

Vi har ett stickprov med någon fördelning med okänd parameter θ.

Vi vill testa *nollhypotesen*

$$H_0 : \theta = \theta_0$$

mot *mothypotesen*

$$H_1 : \theta \neq \theta_0.$$

Om testet bekräftar mothypotesen säger vi att

- H_0 förkastas till förmån för H_1.
- θ är signifikant skild från θ_0.

I fallet då vi inte kan förkasta H_0 har vi *inte* visat att $\theta = \theta_0$. Frågan om H_0 eller H_1 är sann lämnas öppen.

Konfidensintervallmetoden

Om vi bildar ett 95% konfidensintervall I_θ för θ kan vi direkt använda detta för att göra hypotestestet.

- Vi förkastar H_0 om intervallet inte innehåller θ_0.
- Om intervallet däremot innehåller θ_0 kan vi inte förkasta H_0.

Konfidensgraden vi använder när vi beräknar konfidensintervallet är också konfidensgraden för hypotestestet vi gör.

Definition

Felrisken eller signifikansnivån definieras som

$$\alpha = \mathsf{P}(H_0 \text{ förkastas } | H_0 \text{ sann}).$$

Översatt till konfidensintervall är signifikansnivån

$$\alpha = \mathsf{P}(\theta \notin I_\theta \text{ om } \theta = \theta_0) = \mathsf{P}(\theta_0 \notin I_\theta).$$

Teststorheter

Definition

En teststorhet $T = T(X_1, \ldots, X_n)$ är en funktion av observationerna, och alltså en slumpvariabel. $T_{obs} = T(x_1, \ldots, x_n)$ är ett observerat värde av teststorheten för givna observationer.

I allmänhet innehåller teststorheten en skattning θ^* av parametern.

Givet att vi har formulerat vår teststorhet gör vi testet genom att beräkna ett p-värde.

Definition

p-värdet eller signifikanssannolikheten definieras som sannolikheten under nollhypotesen att vi får ett värde T som är åtminstone lika "extremt" som det observerade värdet T_{obs}.

p-värden

Vi vill använda en storhet T som vi känner fördelningen av under H_0, så att vi kan beräkna p-värdet.

I fallet med normalfördelning med känd varians kan vi använda

$$T = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}$$

har vi att T under H_0 är $\mathsf{N}(0,1)$-fördelad, så vi har att

$$p = \mathsf{P}(|T| \geq |T_{obs}|) = 2 \cdot \mathsf{P}(T \geq |T_{obs}|) = 2(1 - \Phi(|T_{obs}|)).$$

Vi förkastar H_0 om p är litet, och det exakta värdet på p är den minsta signifikansnivå vi kan ha för att förkasta H_0.

Om vi vill göra ett test med en fix signifikansnivå α, tex 0.05, behöver vi alltså endast beräkna p och sedan förkasta H_0 om p$< \alpha$.

Kritiskt område

Definition

Givet en signifikansnivå α definierar vi det kritiska området C_α som de värden på teststorheten T som leder till att man förkastar H_0 på nivån α.

Vi förkastar H_0 på nivån α om $p < \alpha$, vilket ekvivalent kan uttryckas som $|T| > T_{\alpha/2,H_0}$, där $T_{\alpha/2,H_0}$ är $\alpha/2$-kvantilen i T:s fördelning under H_0. Därför har vi $C_\alpha = \{T : |T| > T_{\alpha/2,H_0}\}$.

Exempel: Med test för väntevärdet hos normalfördelade data får vi:

- Om variansen är känd: T under H_0 är $\mathrm{N}(0,1)$-fördelad, förkasta H_0 på nivån α om $|T| > z_{\alpha/2}$.
- Om variansen är okänd: T under H_0 är $t(f)$-fördelad, förkasta H_0 på nivån α om $|T| > t_{\alpha/2}(f)$.

Ensidiga test

I vissa situationer kan vi vilja testa om det observerade värdet ligger för högt respektive för lågt i förhållande till referensvärdet.

$$H_0 : \theta = \theta_0$$
$$H_1 : \theta > \theta_0 \text{ (eller } \theta < \theta_0\text{)}$$

Om vi använder konfidensintervall för att göra testet ska vi nu använda ett ensidigt intervall

- om $H_1 : \theta > \theta_0$: $I_\theta = \{a, \infty\}$.
- om $H_1 : \theta < \theta_0$: $I_\theta = \{-\infty, b\}$.

Vi förkastar H_0 om intervallet inte täcker θ_0.
Använder vi oss av en teststorhet och vi beräknar då p-värdet som

- om $H_1 : \theta > \theta_0$: $p = \mathsf{P}_{H_0}(T \geq T_{obs})$.
- om $H_1 : \theta < \theta_0$: $p = \mathsf{P}_{H_0}(T \leq T_{obs})$.

och förkastar sedan H_0 om $p \leq \alpha$.

Styrka

Sannolikheten att *inte* göra ett typ 2 fel kallas för testets *styrka*, och ges av $1-\beta$, där $\beta = \mathsf{P}(\text{förkasta ej } H_0 | H_1 \text{ sann})$.

Jämfört med Typ 1 fel är Typ 2 fel är lite svårare att hantera eftersom de beror på mothypotesen men ett sätt att visualisera dessa är att beräkna testets *styrkefunktion*

Definition

Vi vill testa $H_0 : \theta = \theta_0$ och har bestämt en teststorhet T och en signifikansnivå α. Testets styrkefunktion $\pi(\theta)$ definieras som

$$\pi(\theta) = \mathsf{P}(T \in C_\alpha | \theta) = \mathsf{P}(H_0 \text{ förkastas} | \text{ parametervärdet är } \theta)$$

Testets styrkefunktion ger alltså sannolikheten för att testet ger ett signifikant resultat om det sanna värdet är θ, vilket alltså är testets styrka vid värdet θ.

Sats

Antag att vi har n observationer av en normalfördelning $\mathsf{N}(\mu, \sigma^2)$ med känd varians och att vi vill testa $H_0 : \mu = \mu_0$. Testets styrka ges då av

- Om $H_1 : \mu > \mu_0$:

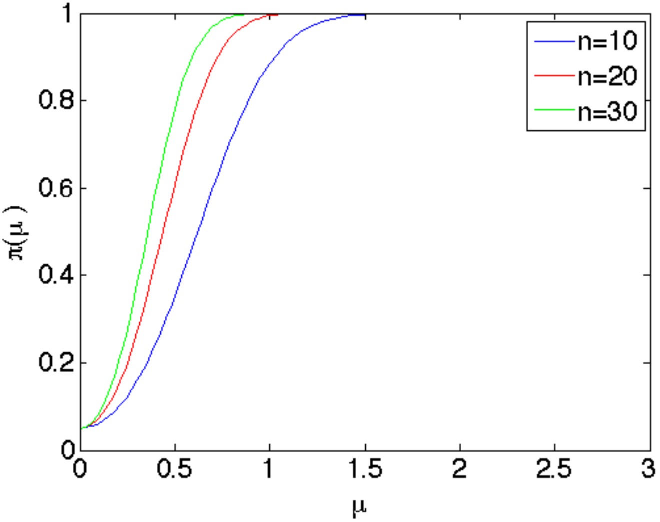
$$\pi(\mu) = 1 - \Phi\left(z_\alpha - \frac{(\mu - \mu_0)\sqrt{n}}{\sigma}\right)$$

- Om $H_1 : \mu < \mu_0$:

$$\pi(\mu) = \Phi\left(-z_\alpha - \frac{(\mu - \mu_0)\sqrt{n}}{\sigma}\right)$$

- Om $H_1 : \mu \neq \mu_0$:

$$\pi(\mu) = 1 + \Phi\left(-z_{\alpha/2} - \frac{(\mu - \mu_0)\sqrt{n}}{\sigma}\right) - \Phi\left(z_{\alpha/2} - \frac{(\mu - \mu_0)\sqrt{n}}{\sigma}\right)$$



Styrkefunktion för test av $H_0 : \mu = 0$ mot $H_1 : \mu \neq 0$ för $\mathsf{N}(\mu, 1)$-fördelningen på nivå $\alpha = 0.05$.

Hur många observationer krävs?

Det vi kan se från styrkefunktionerna i de olika fallen är att testet har lättare att upptäcka en avvikelse $\mu - \mu_0$ om

- Risken för typ 1 fel, α, är stor.
- Antalet observationer, n, är stort.
- Variansen σ^2 är liten.

Om vi bestämmer oss för en viss avvikelse $\mu - \mu_0$ vi vill att testet ska upptäcka samt fixerar signifikansnivån α kan vi använda styrkefunktionen för att beräkna hur många observationer vi bör göra för att få en given styrka $1 - \beta$ på testet.

Sats

Antag att vi har normalfördelade observationer $\mathrm{N}(\mu, \sigma^2)$ och önskar testa $H_0 : \mu = \mu_0$. Om vi vill att testet ska ge utslag för avvikelsen $\mu - \mu_0$ med sannolikhet $1 - \beta$ ska antal observationer väljas som:

- Känd varians:

$$\text{Ensidigt test: } n = \frac{\sigma^2}{(\mu - \mu_0)^2}(z_\alpha + z_\beta)^2$$

$$\text{Tvåsidigt test: } n = \frac{\sigma^2}{(\mu - \mu_0)^2}(z_{\alpha/2} + z_\beta)^2$$

- Om variansen skattas med s^2:

$$\text{Ensidigt test: } n \approx \frac{\sigma^2}{(\mu - \mu_0)^2}(z_\alpha + z_\beta)^2 + z_\alpha^2/2$$

$$\text{Tvåsidigt test: } n \approx \frac{\sigma^2}{(\mu - \mu_0)^2}(z_{\alpha/2} + z_\beta)^2 + z_{\alpha/2}^2/2$$

Jämförelser

En vanlig situation är att man vill göra jämförelser av olika stickprov. Exempel på när detta kan vara av intresse är:

- Vi vill jämföra luftkvalitén i två städer.
- Vi vill se hur olika lufttryck i en typ av bildäck påverkar bromssträckan.
- Vi vill jämföra utbytet av en viss kemisk reaktion vid olika temperaturer.

Vi kommer idag studera två typfall av jämförelser

- Oberoende stickprov
- Stickprov i par

Täthetsfunktionen för $F(f_1, f_2)$-fördelningen