

Föreläsning 9: Hypotesprövning

Matematisk statistik

David Bolin
Chalmers University of Technology
September 28, 2015



Konfidensintervall

Konfidensintervall

Låt X_1, \dots, X_n vara slumpvariabler med en fördelning som har θ som en parameter med θ_0 som sant okänt värde. Ett $100(1 - \alpha)\%$ konfidensintervall för θ med konfidensgraden $1 - \alpha$ är ett intervall $[a(X_1, \dots, X_n), b(X_1, \dots, X_n)]$ sådant att

$$P(a \leq \theta_0 \leq b) = 1 - \alpha.$$

- (a, b) är ett slumpmässigt intervall, eftersom a och b är slumpvariabler som beror av X_1, \dots, X_n .
- Konfidensintervallet ska alltså tolkas som att om vi gör upprepade mätningar av X_1, \dots, X_n och bildar konfidensintervallet för alla dessa mätningar så kommer $100(1 - \alpha)\%$ av dessa intervall täcka det sanna värdet θ_0 .

Konfidsensintervall för μ i normalfördelningen

Känd varians

Låt X_1, \dots, X_n vara oberoende $N(\mu, \sigma^2)$ där σ^2 är känd. Intervallet

$$I_\mu = (\bar{X} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{X} + z_{\alpha/2}\sigma/\sqrt{n})$$

är ett konfidsensintervall för μ med konfidsensgrad $1 - \alpha$.

Okänd varians

Låt X_1, \dots, X_n vara oberoende $N(\mu, \sigma^2)$ där σ^2 är okänd. Intervallet

$$I_\mu = (\bar{X} - t_{\alpha/2}(n-1)S/\sqrt{n}, \bar{X} + t_{\alpha/2}(n-1)S/\sqrt{n})$$

är ett konfidsensintervall för μ med konfidsensgrad $1 - \alpha$. Här är $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ och $t_{\alpha/2}(n-1)$ är $\alpha/2$ -kvantilen i t-fördelningen med $n - 1$ frihetsgrader.

Centrala gränsvärdessatsen (CGS)

Kom ihåg att CGS säger att \bar{X} är approximativt $N(\mu, \sigma^2/n)$ -fördelad om n är stort.

- Om vi har ett stickprov med känd varians σ^2 kommer

$$I_\mu = (\bar{X} - z_{\alpha/2}\sigma/\sqrt{n}, \bar{X} + z_{\alpha/2}\sigma/\sqrt{n})$$

vara ett approximativt konfidensintervall för väntevärdet μ .

- Om vi inte känner variansen kan vi skatta σ med S , men det är då viktigt att n är stort och att fördelningen för X_i inte är allt för tungsvansad om skattningen ska bli någorlunda bra.
- Eftersom n är stor kommer i detta fall $t_{\alpha/2}(n-1) \approx z_{\alpha/2}$, så vi använder

$$I_\mu = (\bar{X} - z_{\alpha/2}S/\sqrt{n}, \bar{X} + z_{\alpha/2}S/\sqrt{n})$$

som ett approximativt konfidensintervall i fallet då σ är okänd.

Konfidsensintervall för σ^2 i normalfördelningenKonfidsensintervall för σ^2

Om X_1, \dots, X_n är oberoende $N(\mu, \sigma^2)$ fås ett konfidsensintervall med konfidsensgrad $1 - \alpha$ för σ^2 som

$$I_{\sigma^2} = \left(\frac{(n-1)S^2}{\chi_{\alpha/2}^2(n-1)}, \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2(n-1)} \right).$$

Här är $\chi_{\alpha/2}^2(n-1)$ $\alpha/2$ -kvantilen i χ^2 fördelningen med $n-1$ frihetsgrader.

Om Z_i är oberoende $N(0, 1)$, så gäller att

$$\sum_{i=1}^n Z_i^2$$

är $\chi^2(n)$ -fördelad.

Konfidsensintervall för σ^2 i normalfördelningenKonfidsensintervall för σ

Om X_1, \dots, X_n är oberoende $N(\mu, \sigma^2)$ fås ett konfidsensintervall med konfidsensgrad $1 - \alpha$ för σ som

$$I_\sigma = \left(\sqrt{\frac{(n-1)S^2}{\chi_{\alpha/2}^2(n-1)}}, \sqrt{\frac{(n-1)S^2}{\chi_{1-\alpha/2}^2(n-1)}} \right).$$

Här är $\chi_{\alpha/2}^2(n-1)$ $\alpha/2$ -kvantilen i χ^2 fördelningen med $n-1$ frihetsgrader.

En viktig anmärkning här är att till skillnad från konfidsensintervall för väntevärdet så är konfidsensintervall för variansen mycket känsliga för avvikelser från normalfördelningen.

Ensidiga konfidensintervall

Ibland är man endast intresserad av en trolig övre eller undre gräns för parametern. Vi kan då bilda ett ensidigt konfidensintervall.

Låt X_1, \dots, X_n vara slumpvariabler med en fördelning som har θ som en parameter med θ_0 som sant okänt värde.

- Ett undre begränsat ensidigt $100(1 - \alpha)\%$ konfidensintervall för θ med konfidensgraden $1 - \alpha$ är ett intervall $[a(X_1, \dots, X_n), \infty]$ sådant att

$$P(a \leq \theta_0) = 1 - \alpha.$$

- Ett övre begränsat konfidensintervall ges på samma sätt av ett intervall $[-\infty, b(X_1, \dots, X_n)]$ sådant att

$$P(\theta_0 \leq b) = 1 - \alpha.$$

I de fallen vi tittat på få ensidiga konfidensintervall genom att endast beräkna en av gränserna samt att vi byter $\alpha/2$ -kvantilen till en α -kvantil.

Hypotesprövning

Ett viktigt problem inom statistiken är att kunna testa om en teori eller en hypotes är sann. Exempel på sådana frågeställningar kan vara

- Ger ett nytt läkemedel någon effekt?
- Dör rökare tidigare än icke rökare?
- Har mätinstrumentet ett systematiskt fel?

Svaret som den statistiska analysen ger kan vara

- ① att hypotesen styrks eller bekräftas av försöket, och man får då också ett mått på hur säker denna slutsats är.
- ② att datan inte ger något stöd för hypotesen.