

Föreläsning 9: Hypotesprövning

Matematisk statistik

David Bolin
Chalmers University of Technology
Oktober 1, 2018



Konfidsensintervall för μ i normalfördelningen

Låt $X_1, \dots, X_n$ vara oberoende $N(\mu, \sigma^2)$.

Känd varians σ^2

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

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

Okänd varians σ^2

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

är ett konfidsensintervall för μ med konfidsensgrad $1 - \alpha$. Här är s^2 stickprovsvariansen och $t_{\alpha/2}(n-1)$ är $(1 - \alpha/2)$ -kvantilen i $t(n-1)$ -fördelningen.

Konfidsensintervall

Konfidsensintervall

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 $100(1 - \alpha)\%$ konfidsensintervall för θ med konfidsensgraden $1 - \alpha$ 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 $X_1, \dots, X_n$.
- Tolkning: Låt $\mathbf{x}_1 = (x_{11}, \dots, x_{n1})$, $\mathbf{x}_2 = (x_{12}, \dots, x_{n2})$, ... vara upprepade mätningar av $X_1, \dots, X_n$. Om vi bildar konfidsensintervallet för θ baserat på varje $\mathbf{x}_i$ så kommer $100(1 - \alpha)\%$ av dessa intervall täcka det sanna värdet θ_0 .

Centrala gränsvärdessatsen

- Kom ihåg att CGS säger att $\bar{X}$ är approximativt $N(\mu, \sigma^2/n)$ -fördelad om n är stor.
- Om vi har ett stickprov med känd varians σ^2 kommer

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

vara ett konfidsensintervall för väntevärdet μ med konfidsensgrad approximativt $1 - \alpha$.

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

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

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 $\chi_{\alpha/2}^2(n-1)$ ($1 - \alpha/2$)-kvantilen i $\chi^2(n-1)$ fördelningen.

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

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

Viktigt: Till skillnad från konfidsensintervall för väntevärdet så är konfidsensintervall för variansen mycket känsliga för avvikelser från normalfördelningen.

Ensidiga konfidsensintervall

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

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 $100(1 - \alpha)\%$ konfidsensintervall för θ med konfidsensgraden $1 - \alpha$ är ett intervall $[a(X_1, \dots, X_n), \infty]$ sådant att $P(a \leq \theta_0) = 1 - \alpha$.
- Ett övre begränsat konfidsensintervall 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 fall vi tittat på fås ett ensidigt konfidsensintervall genom att endast beräkna en av gränserna samt att ändra $\alpha/2$ till α i kvantilen som används.

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.