Föreläsning 13

Thomas Önskog

7/12 2017

Hypotesprövning, styrkefunktion

På förra föreläsningen undersökte vi tre metoder för hypotesprövning. Två typer av fel kan uppkomma vid hypotesprövning.

- **Typ I-fel**: Att förkasta H_0 trots att H_0 är sann. Sannolikheten för denna feltyp är lika med felrisken α.
- **Typ II-fel**: Att inte förkasta H_0 trots att H_0 är falsk.

Typ II-fel är svårare att kvantifiera. Hur stort detta fel är beror på det sanna värdet på parametern θ, vilket är okänt om nollhypotesen H_0 är falsk. Som en försiktighetsåtgärd bör vi därför designa ett hypotestest så att typ I-felet, det fel som vi kan mäta, är det allvarligare felet. Vi exemplifierar detta i nästa exempel. För att även kunna kvantifiera typ II-fel, så inför vi **styrkefunktionen**.

Definition. ***Styrkefunktionen*** *för ett test definieras som*

$$h(\theta) = \mathbb{P}(H_0 \textit{ förkastas} | \theta \textit{ är det sanna parametervärdet}).$$

För $H_0 : \theta = \theta_0$, så gäller $h(\theta_0) \leq \alpha$. Ett bra test har dessutom egenskapen att $h(\theta)$ är stort för alla värden på θ som uppfyller mothypotesen. Vi visar med ett exempel hur styrkefunktionen kan användas i praktiken.

Exempel. *Ett utandningsprov visar alkoholhalten $X_i = \mu + \epsilon_i$ i blodet, mätt i promille, där μ är den sanna alkoholhalten och $\epsilon_i \in N(0, \sigma)$ är mätfelet. För att avgöra om en person är rattonykter sätter vi $H_0 : \mu = 0.2$ (nykter) och $H_1 : \mu > 0.2$ (onykter). Vi väljer μ i H_0 till 0.2 eftersom detta är den högsta tillåtna alkoholhalten vid bilkörning och motsvarar ett sorts worst case scenario bland de alkoholhalter för vilka personen betecknas som nykter. Typ I-fel motsvarar här att en nykter person befinns skyldig till rattonykterhet, medan typ II-fel motsvarar att en onykter befinns oskyldig till rattonykterhet. Enligt gällande rättsprinciper, så är det värre att fälla en oskyldig än att fria en skyldig, så vi har valt våra hypoteser så att typ I-felet är det allvarligare felet.*

Om vi utför n utandningsprov, så är $\overline{X} \in N(\mu, \sigma/\sqrt{n})$. Vi antar att H_0 är sann och vill förkasta H_0 om $\overline{x}$ är stort. Det kritiska området $C = [a, \infty)$ för testvariabeln $t_{obs} = \overline{x}$

uppfyller

$$\begin{aligned}\alpha &= \mathbb{P}(H_0 \textit{ förkastas}|H_0 \textit{ är sann}) = \mathbb{P}(\overline{X} \in C|H_0 \textit{ är sann}) = \mathbb{P}(\overline{X} \geq a|H_0 \textit{ är sann})\\ &= \mathbb{P}\Big(\underbrace{\frac{\overline{X}-0.2}{\sigma/\sqrt{n}}}_{\in N(0,1) \textit{ givet } H_0} \geq \underbrace{\frac{a-0.2}{\sigma/\sqrt{n}}}_{\lambda_\alpha}|H_0 \textit{ är sann}\Big),\end{aligned}$$

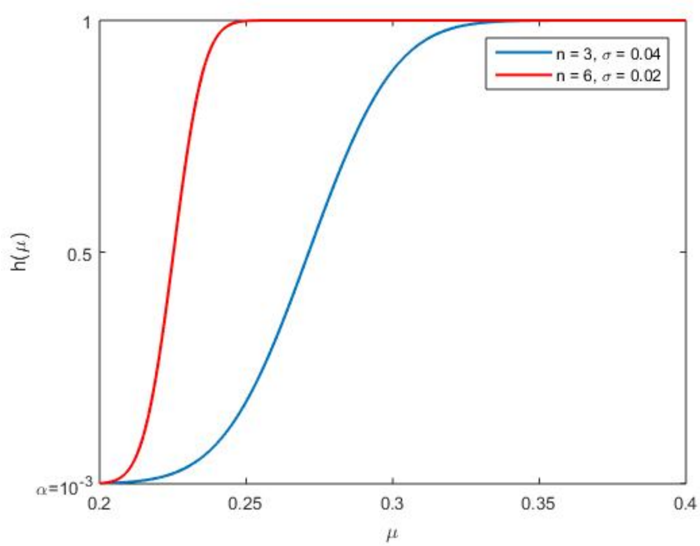
så

$$\frac{a-0.2}{\sigma/\sqrt{n}} = \lambda_\alpha \qquad \Rightarrow \qquad a = 0.2 + \lambda_\alpha \frac{\sigma}{\sqrt{n}}.$$

Nu när det kritiska området är specificerat, så kan vi bestämma styrkefunktionen för $\mu > 0.2$. *Vi får*

$$\begin{aligned}h(\mu) &= \mathbb{P}(\overline{X} \geq a|\mu \textit{ är det sanna parametervärdet})\\ &= \mathbb{P}\Big(\underbrace{\frac{\overline{X}-\mu}{\sigma/\sqrt{n}}}_{\in N(0,1)} \geq \frac{a-\mu}{\sigma/\sqrt{n}}\Big) = 1-\Phi\Big(\frac{a-\mu}{\sigma/\sqrt{n}}\Big) = 1-\Phi\Big(\frac{0.2-\mu}{\sigma/\sqrt{n}}+\lambda_\alpha\Big),\end{aligned}$$

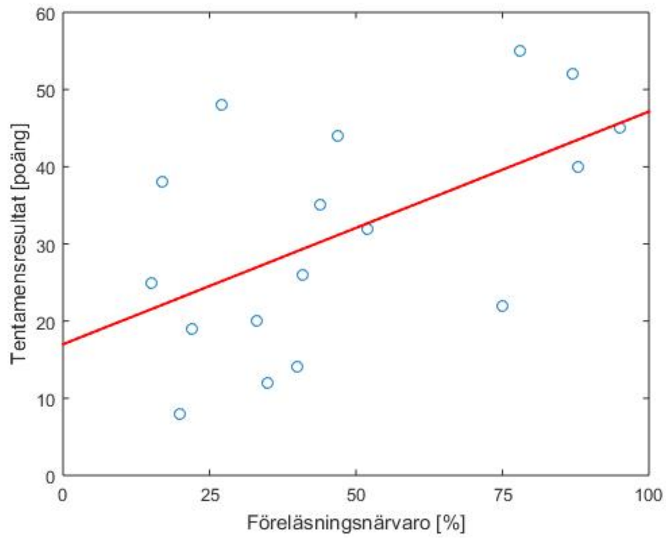
där vi i sista steget satte in $a = 0.2 + \lambda_\alpha \sigma/\sqrt{n}$. *Med* $n = 3$, $\sigma = 0.04$ *och* $\alpha = 0.001$, *så blir* $h(0.3) = 0.89$ *och* $h(0.33) = 0.99$. *Så för* $\mu \geq 0.33$ *är det mycket osannolikt att* H_0 *inte förkastas, dvs. en person med mer än* 0.33 *promille alkohol i blodet kommer med minst* 99% *sannolikhet att befinnas skyldig till rattonykterhet. Om vi anser att testet är för dåligt, så kan vi välja att göra fler utandningsprov eller att använda en mer tillförlitlig mätmetod. Med* $n = 6$, $\sigma = 0.02$ *och* $\alpha = 0.001$, *så blir* $h(0.245) = 0.99$, *så personer med mer än* 0.245 *promille alkohol i blodet kommer att befinnas skyldiga till rattonykterhet med minst* 99% *sannolikhet.*



På nästa föreläsning ska vi undersöka en viktig klass av hypotestest som alla involverar χ^2-fördelningen. Vi ägnar resten av denna föreläsning åt regressionsanalys.

Regressionsanalys

Regressionsanalys är den kanske mest använda statistiska metoden och den används för att avgöra om det finns något samband mellan två storheter. Finns det exempelvis något samband mellan studenters närvaro på föreläsningarna och deras resultat på den efterföljande tentan? **Linjär regression** är en metod för att finna sådana (linjära) samband mellan två storheter.



Vi har n par av värden

$$(x_1, y_1), (x_2, y_2), ..., (x_n, y_n),$$

där $x_1, ..., x_n$ är givna storheter (föreläsningsnärvaron i figuren) och $y_1, ..., y_n$ är observationer av oberoende s.v. $Y_1, ..., Y_n$, där $Y_i \in \mathrm{N}(\mu_i, \sigma)$ (tentamensresultatet i figuren). Observera att σ antas vara oberoende av x. Vidare antas väntevärdena μ_i vara linjärt beroende av x_i, dvs $\mu_i = \alpha + \beta x_i$. Den räta linjen $y = \alpha + \beta x$ kallas för den **teoretiska regressionslinjen** (röda linjen i figuren). Parametrarna α och β är okända, men vi kan skatta dem med hjälp av MK-metoden. Vi vill då minimera kvadratsumman

$$Q(\alpha, \beta) = \sum_{i=1}^{n} (y_i - \mu_i)^2 = \sum_{i=1}^{n} (y_i - \alpha - \beta x_i)^2,$$

dvs. summan av kvadraterna av avstånden ϵ_i i lodrät led mellan observationerna och den teoretiska regressionslinjen. Derivering av Q med avseende på α respektive β ger

$$\frac{\partial Q}{\partial \alpha} = -2 \sum_{i=1}^{n} (y_i - \alpha - \beta x_i) = 0 \qquad \text{och} \qquad \frac{\partial Q}{\partial \beta} = -2 \sum_{i=1}^{n} x_i (y_i - \alpha - \beta x_i) = 0.$$

$$\underset{\text{algebra}}{\Rightarrow} \qquad \beta^*_{\text{obs}} = \frac{\sum_{i=1}^{n} (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^{n} (x_i - \overline{x})^2} \qquad \text{och} \qquad \alpha^*_{\text{obs}} = \overline{y} - \beta^*_{\text{obs}} \overline{x}.$$

Notera att β^*_{obs} är kvoten av stickprovskovariansen mellan $x_1, ..., x_n$ och $y_1, ..., y_n$ och stickprovsvariansen för $x_1, ..., x_n$. Med hjälp av punktskattningarna α^*_{obs} och β^*_{obs}, så får vi en skattad regressionslinje $y = \alpha^*_{\text{obs}} + \beta^*_{\text{obs}} x$. Vi undersöker om punktskattningen av β är väntevärdesriktig och effektiv. Vi kan skriva β^*_{obs} som en linjärkombination av observationerna, dvs.

$$\beta^*_{\text{obs}} = \sum_{i=1}^{n} c_i y_i, \qquad \text{där } c_i = \frac{x_i - \overline{x}}{\sum\limits_{i=1}^{n} (x_i - \overline{x})^2},$$

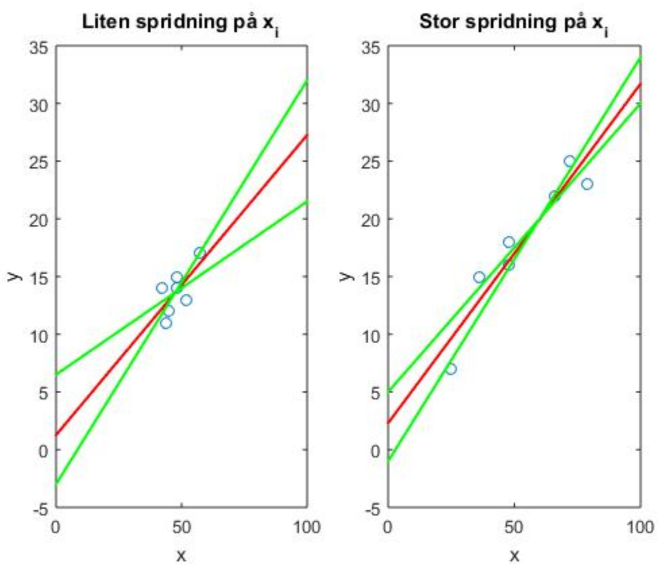
så

$$E(\beta^*) = E\Big(\sum_{i=1}^{n} c_i Y_i\Big) = \sum_{i=1}^{n} c_i \mu_i = \sum_{i=1}^{n} c_i(\alpha + \beta x_i) = \Big[\sum_{i=1}^{n} c_i = 0, \sum_{i=1}^{n} x_i c_i = 1\Big] = \beta,$$

så skattningen är väntevärdesriktig. Vidare gäller

$$V(\beta^*) = V\Big(\sum_{i=1}^{n} c_i Y_i\Big) = \sum_{i=1}^{n} c_i^2 \sigma^2 = \frac{\sigma^2}{\sum\limits_{i=1}^{n} (x_i - \overline{x})^2}.$$

Punktskattningen β^*_{obs} blir alltså mer effektiv om σ är litet och om spridningen hos x_i-värdena är stor. Det sistnämnda faktumet illustreras i figuren nedan.



Eftersom β^*_{obs} är en linjärkombination av ett normalfördelat stickprov, så kan vi enkelt bestämma konfidensintervall för β. I fallet då σ är känt får vi att ett konfidensintervall för β med konfidensgrad p ges av

$$I_\beta = (\beta^*_{\text{obs}} - \lambda_{p/2} D, \beta^*_{\text{obs}} + \lambda_{p/2} D), \qquad \text{där } D = \frac{\sigma}{\sqrt{\sum_{i=1}^{n} (x_i - \overline{x})^2}}.$$

Se boken för fallet med okänt σ. Vi kan använda detta konfidensintervall för att utföra hypotestest med $H_0 : \beta = 0$ och $H_1 : \beta \neq 0$. Om I_β inte innehåller origo, så förkastar vi nollhypotesen och drar därmed slutsatsen att det finns ett signifikant (linjärt) samband mellan storheterna x och y.