Fjármálatölfræði

Haust 2016

3. vika: Upprifjun í líkindafræði, kafli 9-10 í Walpole

Mat á stikum, öryggisbil og tilgátupróf

Athugið að þessar glærur koma ekki í stað fyrirlestra. Meira efni er skrifað upp á töfluna í fyrirlestrum.

Tölfræðileg ályktunarfræði skiptist í **matsfræði eða mat á stikum** (kafli 9) og **tilgátuprófun** (kafli 10).

Gerum ráð fyrir að við höfum sanna líkindadreifingu $f(\underline{x}, \underline{\theta})$, þar sem $\underline{\theta} = (\theta_1, \ldots, \theta_n)$ geymir sönnu stikana, t.d. $\underline{\theta} = (\mu, \sigma)$.

Punktmat fyrir stikann θ er eitthvað gildi $\widehat{\theta}$ sem reiknihendingin $\widehat{\Theta}$ tekur. $\widehat{\Theta}$ er þá kallaður **metill** fyrir θ.

Við þurfum einhvern mælikvarða á það hversu góður metill $\widehat{\Theta}$ er fyrir θ:

Skilgreining. *Við segjum að reiknihendingin $\widehat{\Theta}$ sé* ***meðalgildis-réttur*** *metill fyrir θ ef*

$$E\left(\widehat{\Theta}\right) = \theta.$$

Skilgreining. *Látum $\widehat{\Theta}_1$ og $\widehat{\Theta}_2$ vera tvo metla fyrir θ. Ef*

$$\sigma^2_{\widehat{\Theta}_1} < \sigma^2_{\widehat{\Theta}_2},$$

þá segjum við að $\widehat{\Theta}_1$ sé ***skilvirkari [e. more efficient]*** *en $\widehat{\Theta}_2$. Sá metill fyrir θ sem hefur minnstu dreifnina kallast* ***skilvirkasti metillinn*** *fyrir θ.*

Punktmatið er ekki nákvæmt enda er það háð hendingu og við fáum mismunandi gildi í hvert skipti. En við getum metið líkurnar á því hvort það sé nálægt sanna gildinu.

Við skilgreinum $(1-\alpha)100\%$ **öryggisbil** fyrir θ sem bilið $[\widehat{\theta}_L, \widehat{\theta}_U]$ þannig að það séu $(1-\alpha)$ líkur á því að θ liggi á því bili, þ.e.

$$P(\widehat{\theta}_L < \theta < \widehat{\theta}_U) = 1-\alpha.$$

Hér kallast $1-\alpha$ **öryggisstuðull** og er tala á milli 0 og 1 (yfirleitt 95% eða 99%).

Tvíhliða $100(1-\alpha)\%$ öryggisbil fyrir μ þegar σ er gefið:

Ef $\bar{x}$ er meðalgildi úrtaks af stærð n frá þýði með dreifni σ^2, þá er $100(1-\alpha)\%$ öryggisbil fyrir μ gefið sem

$$\bar{x} - z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x} + z_{\alpha/2}\frac{\sigma}{\sqrt{n}},$$

þar sem $P(Z \leq -z_{\alpha/2}) = \alpha/2$ (Z er staðlaða normaldreifingin).

Einhliða $100(1-\alpha)\%$ öryggisbil fyrir μ þegar σ er gefið:

Með efri mörkum:

$$\left(-\infty, \bar{x} + z_{\alpha}\frac{\sigma}{\sqrt{n}}\right)$$

og með neðri mörkum:

$$\left(\bar{x} - z_{\alpha}\frac{\sigma}{\sqrt{n}}, \infty\right)$$

Skoðum nú tilfellið þegar σ er óþekkt og X_i eru normalhendingar. Rifjum upp að þá gildir að

$$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$$

hefur t-dreifingu með $n-1$ frígráðu. Höfum þá $100(1-\alpha)\%$ einhliða öryggisbilið

$$\bar{x} - t_{\alpha/2}\frac{s}{\sqrt{n}} < \mu < \bar{x} + t_{\alpha/2}\frac{s}{\sqrt{n}},$$

og tvíhliða $100(1-\alpha)\%$ öryggisbilin eru

$$\mu < \bar{x} + t_{\alpha}\frac{s}{\sqrt{n}}$$

og

$$\mu > \bar{x} - t_{\alpha}\frac{s}{\sqrt{n}}$$

Mat á mismun tveggja væntigilda fyrir tvö úrtök:

Ef $\bar{x_1}$ og $\bar{x_2}$ eru meðalgildi úr óháðum úrtökum af stærð n_1 og n_2 úr þýðum með dreifni σ_1^2 og σ_2^2, í sömu röð, þá er $100(1-\alpha)\%$ öryggisbil fyrir $\mu_1 - \mu_2$ gefið sem

$$(\bar{x_1} - \bar{x_2}) - z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}} < \mu_1 - \mu_2 < (\bar{x_1} - \bar{x_2}) + z_{\alpha/2}\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}.$$

Dæmi. Unnið var að rannsókn þar sem tvær bílvelar, A og B, voru bornar saman. Úr þessum rannsóknum kom í ljós að meðaleyðsla bílvelar A var 8.9L/100km og eyðsla bílvelar B var 10.5L/100km. Fimmtíu mælingar voru gerðar á bílvel A og 74 mælingar voru gerðar á bílvel B. Finnið 95% öryggisbil fyrir $\mu_B - \mu_A$. G.r.f. Að staðalfrávik þýðanna eru 6 fyrir vél A og 8 fyrir Vél B.

...en ef σ_1, σ_2 óþekkt en jöfn, þ.e. $\sigma_1 = \sigma_2$:

$$(\bar{x}_1 - \bar{x}_2) - t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}} < \mu_1 - \mu_2 < (\bar{x}_1 - \bar{x}_2) + t_{\alpha/2} s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}.$$

þar sem

$$s_p^2 = \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2}.$$

Sjá einnig tilfelli þegar $\sigma_1 \neq \sigma_2$ og bæði óþekkt.

Öryggisbil fyrir p hlutfall 'success' í tvíliðudreifingu:

$$\widehat{p} - z_{\alpha/2}\sqrt{\frac{\widehat{p}\widehat{q}}{n}} < p < \widehat{p} + z_{\alpha/2}\sqrt{\frac{\widehat{p}\widehat{q}}{n}}$$

þar sem $\widehat{p}$ er punktmat frá hendingunni $\widehat{P} = X/n$ og X táknar fjölda 'success' í úrtakinu. (sjá útleiðslu á töflu).

Dæmi. Slembiúrtak sem innihélt $n = 500$ fjölskyldur úr Grafarvogi, af þessum fjölskyldum voru $x = 230$ sem horfðu á lokaþátt Biggest Loser. Finnið 95% öryggisbil fyrir hlutfall fjölskylda sem horfðu á Biggest Loser.

Fyrir mat á dreifni, þá rifjum við upp að ef $X_i \sim N(\mu, \sigma)$ fyrir öll i, þá fylgir

$$\frac{(n-1)S^2}{\sigma^2}$$

χ^2 dreifingu með $n-1$ frígráðum. Þetta ákvarðar $100(1-\alpha)\%$ öryggisbilið fyrir dreifnina σ^2:

$$\frac{(n-1)s^2}{\chi^2_{\alpha/2}} < \sigma^2 < \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$$

Fyrir samanburð á dreifni úr tveimur þýðum, þá notum við F dreifinguna og höfum öryggisbilið

$$\frac{s_1^2}{s_2^2}\frac{1}{f_{\alpha/2}(v_1, v_2)} < \frac{\sigma_1^2}{\sigma_2^2} < \frac{s_1^2}{s_2^2} f_{\alpha/2}(v_1, v_2)$$

Skoðum nú hálíknamatið á töflunni.

Kafli 10 – Tilgátupróf

- Stundum er vandamálið sem við stöndum frammi fyrir ekki það að við viljum meta stika þýðis heldur viljum við draga ályktanir um eitthvað kerfi útfrá gögnum
- Dæmi: Auka reykingar líkur á krabbameini
- Þurfum við setja fram tilgátu og draga ályktinar útfrá gögnum þ.e eru gögnin í samræmi við tilgátuna okkar?

SKILGREINING. *Núlltilgáta*, táknuð með H_0, er staðhæfing um stika þýðis sem „eðlilegt“ er að halda fram í upphafi. *Gagntilgáta*, táknuð með H_1, staðhæfir öfugt við H_0. Við *höfnum* H_0 ef gögnin benda sterklega til þess að H_0 sé röng. Við *höfnum ekki* H_0 ef gögnin benda ekki nægjanlega sterkt til þess að H_0 sé röng.

Dæmi. Meðallíftími ákveðinnar tegundar af rafhlöðum er 325 klukkustundir. Hönn-unin á rafhlöðunum er endurbætt og spurningin er hvort meðallíftíminn hafi lengst. Núlltilgátan og gagntilgátan eru hér

$$H_0 : \mu \leq 325 \qquad \text{á móti} \qquad H_1 : \mu > 325.$$

Ritum einnig

$$H_0 : \mu = 325 \qquad \text{á móti} \qquad H_1 : \mu > 325.$$

- ATH: Við göngum útfrá H0 nema vísbending um að hún sé röng sé staðfest með gögnunum. Það að hafna ekki H0 þýðir ekki að við samykkjum hana heldur aðeins að gögnin eru ekki nægilega sterk til að hafna henni.
- Saklaus uns sekt er sönnuð: hér er H0 óbreytt ástand – það að sakborningur sé ekki dæmdur sekur þýðir EKKI að hann sé saklaus heldur séu gögnin ekki nægilega sterk til að „sanna“ sekt hans
- H1 er venjulega spurningin sem við viljum fá svar við í tilgátuprófi. Því er framsetning hennar mikilvæg.

Tilgátupróf fyrir meðalgildi í normaldreifingu, dreifnin þekkt

Dæmi. Þykkt 50 glerja í gleraugu eru mæld. Meðaltal úrtaksins er $\bar{x} = 3{,}05$ mm og staðalfrávikið er þekkt, $\sigma = 0{,}34$ mm. Glerin eiga að hafa meðalþykkt $\mu = 3{,}20$ mm. Gefa gögnin til kynna að $\mu = 3{,}20$ mm eða gefa þau til kynna að $\mu \neq 3{,}20$ mm. Notum $\alpha = 0{,}05$.