Experimental Design and Statistics - AGA47a

M. Maciak (Czech University of Life Sciences, Prague)

Lab Session 6 - Summer Term 2015

1 Normalne Gaussovo rozdeleni a CLT - teoreticke priklady

- predpokladejme, ze $X \sim N(3.2, 6.2)$, $Y \sim N(-2.1, 3.5)$ a $Z \sim N(12, 7.5)$ and jsou vzajemne nezavisle. Spoctete nasledujici pravdepodobnosti:

  $X + Y \geq 0$

  $4X - 4Y + 2Z \geq 25$

  $|X + 6Y + Z| \geq 2$

- Zasadenych bylo 20 rostlin. Po trech tydnech byla zmerena velikost kazde rostliny. Predpoklada se, ze tato velikost ma normalni rozdeleni $N(\mu, 2.1^2)$ s neznamym parametrem stredni hodnoty. Spoctete 95 $\%$ konfidencny interval pro ocekavanou velikost rostliny po trech tydnech, kdyz vime, ze vyberovy prumer z 20 rostlin je $\overline{X}_{20} = 12.3\ cm$. Jak by vypadal stejny konfidencny interval, ak by byl pocet rostlin pouze 10, ale vyberovy prumer by zustal stejny?

- Aby se zabranilo sireni nakazy a infekce, byl dobytek vakcinovan. Predpokladame, ze pri jednom ockovani je pravdepodobnost nezaduci reakcie $p = 0.0005$. Tato hodnota vyjadruje, ze v prumeru jeden kus dobytka z 2000 vakcinovanych by mel vykazovat znamky nezadouci reakcie na vakcinaci. Spoctete 95 $\%$ konfidencny interval pro pocet zvirat, ktere budu mit nezadouci reakci, ak pocet kusu dobytka, ktory byli vakcinovan, je 500000.

- Pocet vajec, ktere za jeden den na drubezi farme znesou vsechny slepice dohromady ma priblizne Poissonovo rozdeleni s parametrem $\lambda = 120$ vajec za den. Spoctete 95 $\%$ interval spolehlivosti pro pocet vajec, ktere na drubezi farme "vyrobi" za jeden rok. Jaka je pravdepodobnost, ze techto vajec bude alespon 75000?

2 Studentovo t-rozdeleni v Rku

- vyzkousejte funkce `pt()`, `rt()` a `qt()`, ktere jsou v `R`ku pro Studentovo t-rozdeleni.

- nasimulujte nahodny vyber zo studentovho rozdeleni pro ruzne stupne volnosti a pomoci funkci `density()` a `plot()` porovnajte tohle rozdeleni so standardnim normalnym rozdelenim;

- pouzite kvantilove hodnoty a prislusne pravdepodobnosti Studentovho t-rozdeleni a overte, ze pro narusajici pocet stupnu volnosti konverguje tohle rozdeleni k standardnimu normalnimu $N(0, 1)$;

- pouzite funkci `qqnorm()` a vykreslite vyber zo standardniho normalniho a Studentovho t-rozdeleni. Udelejte to same pro dva vybery z $N(0, 1)$. Jaky zasadny rozdil pozorujete?

- Studentovo t-rozdeleni je v porovnani s normalnim standardnim rozdelenim rozdeleni s tezkymi chvostami. Proc?

3 χ^2 rozdeleni v Rku

- vyzkousejte funkce `pchisq()`, `rchisq()` a `qchisq()` pro pravdepodobnostne hodnoty, generovani a kvantily z chi-kvadrat rozdeleni s n stupnami volnosti;
- nasimulujte dostatecne velky nahodny vyber z χ^2 rozdeleni a vykreslite hustotu; Udelejte to same pro ruzne (postupne narustajici) hodnoty stupnu volnosti. Vyuzite prikazy textttdensity() a `plot()`;
- nagenerujte vyber z chi^2 s n stupni volnosti a nasledne tento vyber standardizujte nasledujicim zpusobem
$$\frac{X-n}{\sqrt{2n}},$$
a opet vykreslite hustotu pomoci prikazu `density()`. Jake rozdeleni hustota pripomina? Cely postup opakujte pro n dostatecne velke (velikost nahodneho vyberu by mela byt take dostatecne velka).

4 Konfidencne intervaly a testy hypotez

- Velikost kukurice v centimetrech ma priblizne normalne rozdeleni s parametrem stredni hodnoty $\mu \in \mathbb{R}$, ktery je neznamy a rozptylem $\sigma^2 = 25$ centimetru. Nahodny vyber o velikosti 16 kusu byl nazbierany a parameter μ byl odhadnout jako $\overline{X}_{15} = 25.5\ cm$. Jaky je 95 % konfidencny interval pro neznamy parameter μ?
- Uvazujme predchozi pripad, avsak parametr rozptylu je stejne neznamy,plati pouze $\sigma^2 > 0$. Z nahodneho vyberu jsme tento rozptyl odhadli pomoci vyberoveho rozptylu a prislusna hodnosta je $s_{16}^2 = 25$. Jaky je 95% konfidencny interval pro neznamy parametr μ v teto situaci?
- Testujte nulovou hypotezu, ze skutecna hodnota parametru μ je rovna hodnote 25 centimetru, vzhledem k alternativni hypoteze, ze tomu tak neni.
- Predpokladame, ze hmotnost jedneho kusu hoveziho dobytka ma priblizne normalne rozdeleni $N(\mu, 3600)$ v kilogramech. Bylo zmerenych 100 kusu dobytka a neznamy parameter stredni hodnoty, parametr μ, byl odhadnout vyberovym prumerem jako $\overline{X}_{100} = 635\ kgs$. Pro hladinu spolehlivosti $\alpha = 0.05$ testujte nulovou hypotezu, ze skutecna hodnota parametru (stredni ocekavana hmostnost) je rovna hodnote 650 kgs. Spoctete 95% konfidencny interval pro μ a porovnajte interval s vysledkem statistickeho testu.
- V predchozim prikladu udelejte stejny test hypotezy, pro stejnou hladinu spolehlivosti, avsak parameter $\sigma^2 > 0$ je neznamy, pouze jsme ho odhadli pomoci vyberoveho rozptylu jako $s_{100}^2 = 3600$.
- Cas, ktery stravite cekanim na autobus kazdy den ma exponencialne rozdeleni s paramtrem $\lambda = 6$ (v minutach). Predpokladejme, ze behem semestru musite jit do skoly (a teda i cakat na autobus) 110 krat. Jaka je pravdepodobnost, ze dohromady nestravite cekanim na autobus vice, nez 12 hodin?
- Uvazujte predchozi pripad, akurat je parameter $\lambda > 0$ neznamy. Na zaklade pozorovani za 60 dnu jsme ale zjistili, ze jsme cekanim stravili dohromady380 minut. Odhadnite parameter λ na zaklade teto informace. Sestrojte 90% konfidencny interval pro $\lambda > 0$?
- Uvazujme jednou minci, o ktere nevime, jestli je spravodliva, nebo ne. Proto ji otestujeme a 1000 ji vyhodime a zaznamename vysledek. Nasledne odhadneme parameter $p \in (0,1)$, co je pravdepodobnost, ze sledujeme hlavu. Z celkoveho poctu 1000 hodu jsme pozorovali 580 hlav. Jaky je prislusny 95% konfidencny interval pro neznamy parameter $p \in (0,1)$?
- V predchozim pripade testujte nulovou hypotezu, ze mince je spravodliva, teda ze plati, ze $p = 0.5$ ($H_0 := 0.5$ vs. $H_1 : p \neq 0.5$).

5 Pokryti konfidencnych intervalu - spolehlivost

- Predpokledejme, ze mame nekolik ruznych vyberu z rozdeleni $N(\mu, \sigma^2)$:

$$
\begin{aligned}
\mathcal{S}_1 &= \{12.17, 9.98, 7.54, 15.93, 8.81, 8.56, 8.00, 7.63, 5.84, 8.92\} \\
\mathcal{S}_2 &= \{9.17, 8.27, 9.51, 8.08, 12.24, 9.84, 10.47, 9.64, 13.91, 9.67\} \\
\mathcal{S}_3 &= \{7.30, 12.57, 10.03, 10.40, 10.94, 6.02, 13.07, 9.45, 4.52, 6.92\} \\
\mathcal{S}_4 &= \{12.61, 9.84, 12.40, 17.67, 12.19, 12.18, 12.31, 8.85, 9.68, 13.30\} \\
\mathcal{S}_5 &= \{6.97, 13.32, 7.63, 15.09, 12.39, 6.64, 5.79, 7.34, 13.88, 6.37\} \\
\mathcal{S}_6 &= \{11.63, 9.62, 12.30, 11.59, 9.61, 5.37, 15.54, 6.44, 7.63, 11.88\} \\
\mathcal{S}_7 &= \{9.03, 8.52, 12.00, 15.43, 12.96, 6.53, 9.03, 13.82, 11.70, 9.31\} \\
\mathcal{S}_8 &= \{6.18, 10.42, 11.01, 12.19, 0.82, 6.73, 5.11, 7.85, 8.51, 5.05\} \\
\mathcal{S}_9 &= \{6.30, 5.20, 16.40, 12.46, 10.20, 12.67, 10.67, 14.05, 15.74, 15.82\} \\
\mathcal{S}_{10} &= \{11.75, 12.59, 8.63, 9.69, 11.52, 12.25, 11.86, 9.16, 7.94, 4.69\} \\
\mathcal{S}_{11} &= \{12.89, 8.81, 14.36, 10.38, 15.81, 13.68, 9.49, 8.82, 8.95, 14.20\} \\
\mathcal{S}_{12} &= \{8.36, 5.31, 6.88, 11.66, 9.90, 5.43, 11.41, 8.62, 8.92, 11.59\} \\
\mathcal{S}_{13} &= \{6.51, 8.59, 12.16, 5.40, 12.22, 11.81, 4.78, 12.26, 8.61, 11.85\} \\
\mathcal{S}_{14} &= \{8.58, 11.75, 6.45, 11.58, 13.22, 12.51, 5.89, 13.26, 16.42, 10.24\} \\
\mathcal{S}_{15} &= \{9.66, 13.48, 10.23, 14.39, 10.67, 13.12, 12.67, 14.81, 10.79, 9.33\} \\
\mathcal{S}_{16} &= \{8.37, 12.44, 8.60, 9.04, 13.01, 9.99, 13.27, 4.63, 8.24, 15.81\} \\
\mathcal{S}_{17} &= \{8.90, 9.93, 12.37, 14.28, 10.11, 8.49, 13.08, 14.70, 13.91, 17.32\} \\
\mathcal{S}_{18} &= \{7.69, 9.54, 12.59, 14.02, 5.68, 11.58, 10.06, 10.28, 7.82, 3.64\} \\
\mathcal{S}_{19} &= \{6.62, 12.49, 10.64, 11.83, 8.90, 9.64, 6.81, 9.23, 12.05, 9.07\} \\
\mathcal{S}_{20} &= \{8.09, 6.68, 9.19, 11.63, 10.19, 12.48, 6.82, 7.29, 12.22, 11.71\}
\end{aligned}
$$

  Vytvorte 90% konfidencny interval pro parameter $\mu \in \mathbb{R}$ ak $\sigma^2 > 0$ ije neznamy. Udelejte to same, ale s dodatecnym predpokladem, ze $\sigma^2 = 10$. Oba intervaly porovnejte.

- Udelejte to same pro kazdy nahodny vyber S_i, kde $i = 1, \ldots, 20$ a srovnejte mezi sebou. Jaky efekt pozorujete?

- Spoctete prislusny 90% konfidencny interval pro neznamy parameter $\sigma^2 > 0$?