12. cvičení z PSt — 25.5.2021

Testování hypotéz

1. (Všimněte si podobnosti a rozdílu oproti příkladu z minule.) Máme jedno měření $X \sim N(\mu, 1)$. Chceme ověřit hypotézu H_0: $\mu = 5$ s hladinou významnosti $\alpha = 5$ %.

(a) Jaký zvolíme kritický obor – množinu měření, ve které hypotézu zamítneme?

(b) Místo jednoho měření jich provedeme n (pochopitelně nezávislých). Jaký bude kritický obor pro $\bar{X}_n$?

(c) Pokud je ve skutečnosti $\mu = 4$ a máme $n = 10$ měření, jaká je pravděpodobnost, že hypotézu nezamítneme?

(d) Nechť X má stále střední hodnotu μ a rozptyl 1, ale není už nutně normální. Co se změní?

2. Podle slibu výrobce bude jeho stroj dělat chyby nejvýše ve 3 % případů. Z 600 pokusů došlo k chybě v 28 případech. Posuďte slib výrobce (coby nulovou hypotézu) na hladině významnosti 5 %.

(a) Počet chyb modelujte přesně, tj. pomocí binomického rozdělení.

(b) Počet chyb modelujte přibližně pomocí normálního rozdělení (s vhodným μ, σ^2).

Test dobré shody

3. Otestujte náhodnost kostky. Můžete házet ručně, máte-li kostku po ruce, nebo využít vhodný generátor (na webu, např. `https://www.random.org/dice/?num=60` – výsledek jde kopírovat a vložit jako posloupnost čísel – nebo softwarový). Případně přímo v R (např. sample(6,100, replace=T)). Použijte Pearsonův χ^2 test dobré shody.

4. Pojďme zkontrolovat, zda počty emailů jsou opravdu dobře modelovány Poissonovským rozdělením. Můžete použít svoje data, nebo použít moje, za loňský listopad. (Nejsou to všechny emaily, jenom „ty důležité“ podle klasifikace Gmailu, ale to by na statistických vlastnostech nemělo nic měnit.)

$$moje\_emaily11 = c(0, 6, 14, 8, 8, 9, 3, 3, 12, 12, 15, 7, 15, 2, 5, 13, 5, 17, 15, 11, 9, 2, 16, 8, 9, 11, 6, 2, 2, 9)$$

Rozmyslete, jak použít test dobré shody, zjistěte velikost λ a test proveďte. V R můžete použít buď vzorec z přednášky a dosadit do distribuční funkce ($pchisq(T, df)$), nebo použít příkaz *chisq.test*. Ideálně zkuste obojí.

Pro případné zvýšení přesnosti lze použít příkaz options(digits = 22).

Zamyslete se nad výsledkem.

Lineární regrese

5. Pro hodnoty $x = 1, 5, 9, 10, 13, 16, 20, 30$ nám vyšlo $y = 6.1982, 12.9892, 23.8005, 23.8891, 30.0391, 35.7535, 49.0685, 63.1825$. Proveďte lineární regresi. Jak dobrá je? Nakreslete i obrázek, kde vynesete body a proloženou přímku.

K procvičení

6. Bezpečnostní kamera pozoruje hlídanou oblast a naměří signál $X = W$, pokud není přítomen lupič (hypotéza H_0) nebo $X = W + \vartheta$, pokud lupič přítomen je (hypotéza H_1). Víme, že $\vartheta > 0$, ale přesnou hodnotu neznáme (závisí na velikosti lupiče).

Předpokládáme, že $W \sim N(0, 0.5)$.

(a) Naměříme jednu hodnotu $X = 0.96$. Máme nulovou hypotézu zamítnout na hladině 5 %?

(b) Naměříme pět nezávislých hodnot $0.96, -0.34, 0.85, 0.51, -0.24$. Máme nulovou hypotézu zamítnout na hladině 5 %?

(c) Opakujte část (b) pomocí t-distribuce (bez předpokladu znalosti rozptylu).

7. Náklon šikmé věže v Pise je měřen vzdáleností pevného bodu ve věži od jeho „správné“ polohy. V letech 1975 až 1987 tato poloha rostla následujícím způsobem: 2.9642, 2.9644, 2.9656, 2.9667, 2.9673, 2.9688, 2.9696, 2.9698, 2.9713, 2.9717, 2.9725, 2.9742, 2.9757. Proveďte lineární regresi, znázorněte i graficky.