

Design Experimentu a Statistika - AGA46E

Homework 3 - 14.04.2015

Deadline: 28.04.2015 (paper copy)

Teoretická část

□ Problem 1

Uvažujme náhodný výběr X_1, \dots, X_{30} z rozdělení $N(\mu, \sigma^2)$, kde oba parametry jsou neznáme. Příslušné výběrové odhady pro μ a σ^2 jsou $\bar{X}_{30} = 0.25$ a $s_{30}^2 = 1.1$. Na základě 95% intervalu spolehlivosti pro neznámý parameter μ a σ^2 rozhodnete, jestli je lze předpokládat, že náhodný výběr byl generován z standardního normálního rozdělení $N(0, 1)$.
[2 body]

□ Problem 2

Množství pšenice, kterou sezbíráme na poli o rozloze 1 akr, má přibližně normální rozdělení $N(\mu, 400)$ v metrických tunech. Na poli o rozloze 10 akru bylo na každém akru sezbíráno postupně 17.5, 49, 39, 60, 53, 64, 42.5, 68, 70, a 34 metrických tun pšenice (předpokládáme, že množství na každém akru je nezávislé od množství na ostatních akrech); Najděte 95% interval spolehlivosti pro očekávané množství pšenice na jednom akru (neznámý parameter střední hodnoty $\mu \in \mathbb{R}$). [2 body]

□ Problem 3

Použijte obyčejnou minci a 30 krát ji vyhodte a pokazde zaznamenajte, jestli padla hlava nebo lic. Použijte centralní limitní větu a ověřte, že mince je spravedlivá (hlava s pravděpodobností 0.5 a stejně tak lic s pravděpodobností 0.5). Navod: pomocí CLV najdete buď konfidenční interval s danou spolehlivostí, nebo testujte příslušnou nulovou hypotézu, $H_0 : p = 0.5$, kde p je pravděpodobnost, že padne hlava.
[2 points]

□ Bonus Problem

Průměrná výška mužů a žen v populaci se liší. Abychom tenhle rozdíl statisticky odhadli, změřili jsme 20 mužů a 20 žen a zaznamenali jsme jejich výšku. Předpokládáme, že výška mužů má normální rozdělení $N(\mu_1, \sigma^2 = 64)$, a výška žen v populaci se řídí rozdělením $N(\mu_2, \sigma^2 = 64)$. Výběrové průměry jsou $\bar{X}_{20\text{males}} = 181$ a $\bar{Y}_{20\text{females}} = 168.5$ (měřeno v centimetrech). Najděte 95% interval spolehlivosti při rozdílu mezi průměrnou výškou mužů a průměrnou výškou žen. Otestujte nulovou hypotézu H_0 , že tenhle rozdíl je 10 centimetrů oproti alternativě, že tomu tak není.

Prakticka cast v programu R

1. Stahnete a v Rku nactete data `passengerData2.csv`. Predpokladejte, ze promene *TotalTime* a *WaitingTime* maji normalne rozdeleni s neznamym parametrem stredni hodnoty a neznamym parametrem rozptylu. Spocetete 95% interval spoehlivosti pro stredni hodnotu celkoveho casu (*TotalTime*) a stredni hodnotu doby cekani (*WaitingTime*) a to samostatne pro pasazeri - muze a pro pasazeri - zeny.
2. Pomoci analyzy rozptylu (prikaz `anova()` v R-ku) testujte nulovou hypotezu, ze cekaci doby na letisti (promenna *WaitingTime*) je stejnna pro vsechny typy letu (promenna *FlightType*). Udelejte vhodny graf a interpretujte. vysledky.
3. Pomoci linearniho regresniho modelu overte, jestli existuje linearni zavislost mezi dobou cekani - *WaitingTime* a celkovym trvanim letu - *FlightTime*. Udelejte vhodny graf a vysledky interpretujte.