

Testování hypotéz II.

5. 1. 2012

Úvodní nastavení. Budeme pokračovat v práci s daty `zaci.dat`. Spusťte si R Commander (`Packages`→`Load Package`→`Rcmdr`), změňte si svůj pracovní adresář (`File`→`Change working directory`) a načtěte si do R data `zaci.dat` (`Data`→`Import Data`→`from text file, clipboard or URL`). Nezapomeňte opět převést kategoriální veličiny na faktory pomocí `Data`→`Manage variables in active data set` →`Convert numeric variables to factors`.

Veškeré statistické testy provádějte na hladině významnosti 5%.

1. Společná práce

1. Zajímá nás, zda v osmé třídě dochází u žáků ke zlepšení známek oproti konci sedmé třídy. Podíváme se proto na rozdíl mezi průměrnou známkou na konci 7. třídy a průměrnou známkou v pololetí osmé třídy.

(a) Spočítejte si tuto novou veličinu.

`Data`→`Manage variables in active data set` →`Compute new variable`. Zde do políčka `New variable name` napište `rozdil` (jméno nové veličiny), do `Expression` napište `ZN8-ZN7K` (známka z osmé třídy minus známka z konce sedmé).

Podívejte se na základní popisné statistiky této veličiny. Na základě nich posuďte, zda dochází v osmé třídě spíše ke zlepšení nebo zhoršení.

(b) Posuďte normalitu rozdělení této veličiny pomocí vhodných obrázků a pomocí Shapiro-Wilkova testu.

(c) Otestujte, zda u žáků dochází ke zmíněnému zlepšení známek.

- Jak vyjádříme domněnku o zlepšení pomocí výše definované veličiny `rozdil`? Formulujte nulovou a alternativní hypotézu, kterou chcete testovat.
- Proved'te jednovýběrový t-test a podívejte se na výsledek. Je tento test pro danou situaci vhodný či nikoliv?
- Proved'te test pomocí jednovýběrového Wilcoxonova testu.

R Commander bohužel nenabízí „klikací“ variantu jednovýběrového Wilcoxonova testu, a proto je nutné jej zadat ručně pomocí příkazu `wilcox.test`. V okénku `Script Window` tedy přepište poslední řádek (t-test) tak, že `t.test` nahradíte `wilcox.test`. Poté klikněte na tlačítko `Submit`.

- Připomeňte si, na čem je založen Wilxonův test. Jaká je konkrétní hodnota testové statistiky?
- Porovnejte výsledek Wilcoxonova testu s t-testem. Interpretujte závěr.

2. Zajímáme se o porovnání IQ chlapců a dívek.

- (a) Pomocí popisných statistik si udělejte představu o tom, zda se typické IQ chlapců a dívek liší.

[Statistics](#) → [Summaries](#) → [Numerical summaries](#)

Vykreslete si dále vhodný ilustrativní obrázek.

- (b) Ověřte, zda lze předpokládat, že IQ chlapců a dívek jsou nezávislé výběry z normálního rozdělení se shodným rozptylem.

- Číselné porovnání směrodatných odchylek (např.): [Statistics](#) → [Summaries](#) → [Table of statistics](#). Zde vyberte IQ v závislosti na POHLAVI a zvolte [Standard deviation](#).
- [Test normality](#): Poslední řádek ve [Script Window](#) přepište tak, že namísto `sd, na.rm=TRUE` napíšete `shapiro.test`. Poté dejte [Submit](#).
- Test shody rozptylů provedeme pomocí F-testu. Připomeňte si jeho předpoklady, nulovou a alternativní hypotézu.

[Statistics](#) → [Variances](#) → [Two Variance F-test](#).

- (c) Otestujte, zda je typické IQ chlapců vyšší než typické IQ dívek.

- Uveďte uvažovaný model.
- Formulujte nulovou a alternativní hypotézu.
- Proveďte vhodný test, tj. dvouvýběrový t-test. Připomeňte si předpoklady tohoto testu. Jsou splněny?

Dvouvýběrový t-test provedeme pomocí [Statistics](#) → [Means](#) → [Independent samples t-test](#). Zde vhodně doplňte jednotlivá políčka (především správná volba alternativní hypotézy a předpokladu shody rozptylů).

- Jaká je p-hodnota testu? Zamítáme nebo nezamítáme nulovou hypotézu? Interpretujte.
- Jaká je hodnota testové statistiky? Jaké má testová statistika dvouvýběrového t-testu rozdělení? Vzpomněli byste si na vzoreček pro testovou statistiku?

3. Zjistěte, zda mají chlapci na konci sedmé třídy lepší známky než dívky.

- (a) Nejdříve se podívejte na popisné statistiky a vhodný obrázek.

- (b) Posuďte, jak je to s normalitou průměrné známky z konce sedmé třídy pro dívky a chlapce zvlášť.

- (c) Odpovězte na zkoumanou otázku pomocí vhodného testu.

- Je vhodné provést porovnání pomocí dvouvýběrového t-testu? Zdůvodněte odpověď, případně test proveďte (je-li vhodný).
- Proveďte porovnání pomocí dvouvýběrového Wilcoxonova testu. Formulujte nulovou a alternativní hypotézu. Připomeňte si, na čem je založen Wilcoxonův dvouvýběrový test.

[Statistics](#) → [Nonparametric tests](#) → [Two-Sample Wilcoxon test](#)

- Jaká je p-hodnota testu? Jaká je odpověď na naši otázku?

2. Samostatná práce

Stáhněte si z internetu soubor `hypermarket.dat`. Data obsahují informace o tržbách náhodně vybraných hypermarketů jednoho velkého obchodního řetězce. Pro každý hypermarket známe jeho lokalitu (Čechy vs. Morava) a tržby před, během a po reklamní kampani (tržby jsou v tisících Kč za den).

7. Načtěte si data a prohlédněte si základní popisné statistiky. Co lze na základě číselných údajů říci o průměrných tržbách před, během a po kampani?
8. Pomocí vhodného obrázku se podívejte, jak spolu souvisí tržby před a po kampani.
9. Ověřte (testem), zda byl typický zisk během kampaně vyšší než 1 500 000 Kč denně.
 - (a) Uveďte uvažovaný model, zformulujte nulovou a alternativní hypotézu.
 - (b) Proveďte vhodný test a ověřte jeho předpoklady. Interpretujte svůj závěr.
10. Odhadněte tržbu před kampaní pro Čechy a Moravu zvlášť. Porovnejte také graficky.
11. Vedení obchodního řetězce tvrdí, že hypermarkety v Čechách mají vyšší tržby než hypermarkety na Moravě. Ověřte tuto domněnku.
 - (a) Formulujte nulovou a alternativní hypotézu.
 - (b) Vyberte vhodný test a ověřte jeho předpoklady.
 - (c) Proveďte test a interpretujte závěr.
12. Management přirozeně zajímá, zda byla reklamní kampaň úspěšná, tj. zda vedla ke zvýšení tržeb.
 - (a) Formulujte nulovou a alternativní hypotézu.
 - (b) Vyberte vhodný test a ověřte jeho předpoklady.

Pozor, tržba před a po kampani v konkrétním obchodě jsou závislé veličiny! Nelze tedy použít dvouvýběrový t-test, vhodný je párový t-test (Paired t-test).
 - (c) Proveďte test a interpretujte závěr.
13. Neměla kampaň náhodou úplně opačný účinek než by vedení požadovalo? Otestujte!