

---

## CVIČENÍ 30.4.2013

---

POPIS DAT. Data `zaci.dat` obsahují informace o náhodně vybraných žácích osmé třídy základní školy. Pro každého máme k dispozici jeho IQ, pohlaví (kódování 0 dívky a 1 chlapci) a dále pak průměrnou známku na vysvědčení v pololetí sedmé třídy, na konci sedmé třídy a v pololetí osmé třídy.

ZAK	identifikační číslo žáka,
POHLAVI	pohlaví žáka (0 - žena, 1 - muž)
IQ	hodnota IQ,
ZN7	průměrná známka na vysvědčení v pololetí sedmé třídy,
ZN7K	průměrná známka na vysvědčení na konci sedmé třídy,
ZN7	průměrná známka na vysvědčení v pololetí osmé třídy.

1. Načtěte si do R Commanderu soubor data `zaci.dat` a prohlédněte si základní popisné statistiky. Podívejte se, které veličiny jsou kategoriální, a případně změňte jejich formát na `factor`.
2. Pomocí vhodných obrázků a popisných statistik si udělejte představu
  - (a) zda lze rozdělení IQ považovat za normální,
  - (b) zda a jak se liší IQ chlapců a dívek,
  - (c) zda a jak se liší známky chlapců a dívek,
  - (d) zda a jak se liší IQ žáků v jednotlivých školách,
3. Pomocí vhodného obrázku prozkoumejte, jak spolu souvisí známka z konce 7. třídy a z pololetí 8. třídy. Příslušný vztah popište také číselně pomocí korelačního koeficientu.
4. Na Wikipedii (<http://cs.wikipedia.org/wiki/Intelligence>) se podívejte na tabulku charakterizující inteligenci dle výše IQ. Zjistěte, jaké je procentuální zastoupení dětí s IQ v kategoriích do 90 (podprůměr), 91-110 (průměr), nad 101 (nadprůměr). Máme v souboru i nějakého génia (nad 140)?
5. Optimističtí pracovníci Ministerstva školství se domnívají, že je střední IQ žáků osmé třídy určitě vyšší než 105. Ověřte tuto jejich domněnku na hladině 5%.
  - (a) Zformulujte nulovou a alternativní hypotézu.
  - (b) Ověřte předpoklady jednovýběrového t-testu (graficky i Shapirovým-Wilkovým testem).
  - (c) Proveďte test a interpretujte výsledek.  
Statistics→Means→Single-sample t-test
6. Některé příručky uvádějí, že je typické IQ dětí takto starých je 106. Ověřte tuto domněnku na hladině 5%.
7. Pomocí stejné procedury si nechte vypsát oboustranný intervalový odhad středního IQ na hladině 90 %, 95% a 99%.
  - (a) Jak se mění interval, když zvyšujeme spolehlivost?

- (b) Jak souvisí tyto intervaly spolehlivosti s výsledkem posledního testu?
8. Nechte si vypsát také levostranný a pravostranný interval spolehlivosti na hladině 95%. Který z nich by bylo možné použít pro test domněnky ministerstva (viz výše)?
9. Zkonstruujte intervalový odhad střední průměrné známky na hladině 95% pro jednotlivé známky, které máte k dispozici.
- (a) Posuďte vždy předpoklad normality.
- (b) Výsledky porovnejte - pro kterou ze známek je interval nejširší a pro kterou nejúžší? Čím je to způsobeno?
10. **Párový t-test.** Zajímá nás, zda v osmé třídě dochází ke zlepšení prospěchu v porovnání se sedmou třídou. Podíváme se proto na rozdíl mezi průměrnou známkou na konci 7. třídy a průměrnou známkou v pololetí osmé třídy.
- (a) Spočítejte si novou veličinu `rozdil` pomocí `Data→Manage variables in active data set →Compute new variable`.
- (b) Podívejte se na základní popisné statistiky této veličiny. Co z nich soudíte o porovnání známek v sedmé a osmé třídě?
- (c) Posuďte, zda lze rozdělení této veličiny považovat za normální (pomocí obrázků a a Shapirova-Wilkova testu).
- (d) Otestujte, zda v osmé třídě dochází u žáků ke zlepšení známek oproti konci sedmé třídy.
- Připomeňte si model předpokládaný pro použití párového t-testu. Jsou tyto předpoklady splněny?
  - Formulujte nulovou a alternativní hypotézu, kterou chcete testovat.
  - Proveďte test.  
(Lze provést buď jednovýběrový t-test na veličinu `rozdil` nebo párový test pomocí nabídky `Statistics→Means →Paired t-test`.)
  - Vyzkoušejte, že oba uvedené způsoby dávají naprosto totožné výsledky.
  - Interpretujte výsledek.
- (e) Jaký je intervalový odhad rozdílu mezi známkami? Jaká je zde souvislost s výsledkem testu?
11. Otestujte, zda se liší průměrná známka v pololetí sedmé a osmé třídy. Nezapomeňte posoudit předpoklady testu, podívat se na intervalový odhad rozdílu a správně interpretovat výsledek.
12. Předchozí porovnání proveďte zvlášť pro chlapce a zvlášť pro dívky. Liší se nějak závěry oproti výsledku v předchozím bodě?

**SAMOSTATNÁ PRÁCE** Uvažujte data `Davis.txt`, která obsahují údaje o mužích a ženách, kteří pravidelně navštěvují náhodně vybranou posilovnu. U každé osoby známe jeho/její pohlaví, skutečnou váhu a výšku a váhu a výšku, kterou osoba veřejně uvádí.

Pohlavi	pohlaví osoby,
Vaha	skutečná (naměřená) hmotnost v kg,
Vyska	skutečný (naměřená) výška v cm,
RVaha	hmotnost v kg, kterou osoba veřejně uvádí,
RVyska	výška v cm, kterou osoba veřejně uvádí.

1. Načtete si do R Commanderu soubor data a prohlédněte si základní popisné statistiky.
2. Bude nás zajímat, zda lidé v posilovně uvádějí jinou hmotnost, než ve skutečnosti mají.
  - (a) Spočtete novou proměnnou, která bude udávat tento rozdíl.
  - (b) Prohlédněte si popisné statistiky a obrázky této proměnné.
  - (c) Posuďte, zda lze rozdělení této proměnné považovat za normální.
  - (d) Otestujte, zda lze na základě daných dat prokázat, že lidé v posilovně uvádějí jinou hmotnost, než skutečně mají.
3. Otestujte, zda muži uvádějí statisticky významně vyšší hmotnost než skutečně mají.
4. Zjistěte, zda ženy uvádějí statisticky významně nižší hmotnost než skutečně mají.
5. Otestujte, zda lidé uvádějí statisticky významně větší výšku než skutečně mají.
  - (a) Nejprve proveďte srovnání pro všechny osoby, bez rozlišení pohlaví.
  - (b) Poté proveďte test pro muže a ženy zvlášť.