

## Generování z rozdělení, vícerozměrné testy

1. Na základě informací z minulé hodiny si zkuste napsat skript na generování z

- (a) alternativního rozdělení s parametrem  $p = 0.1$
- (b) binomického rozdělení s parametry  $n = 10$  a  $p = 0.1$
- (c) normálního rozdělení s parametry  $\mu = 0.5$  a  $\sigma^2 = 9$
- (d) Hotellingova rozdělení

2. Nyní si trochu oprášíme jednorozměrné testování

- Načtěte si data `ženy_diabetes.txt`
- Zaměříme se nejprve na BMI těchto žen
- vypište si základní výběrové charakteristiky pro BMI
- podívejte se na grafické zpracování údajů - histogram, boxplot, q-q graf
- určete oboustranný a jednostranné intervaly spolehlivosti pro střední hodnotu BMI
- Jak by to vypadalo s testy o střední hodnotě?

3. Teď už přejdeme k vícerozměrnému testování

Načtěte si všechna data. Uvádí informace o zdravotním stavu sta žen a mužů diabetes. U žen jsou navíc nashromážděny informace o jejich stavu po roce léčby.

- Vizualizujte si jednotlivé soubory metodami, které už znáte, tj. nechte si vypsát souhrnné statistiky, podívejte se na jednotlivé zkoumané veličiny, histogramy, boxploty a qqploty (srovnej s normálním rozdělením). Co byste o jednotlivých složkách náhodného vektoru řekli?

- zkuste vizualizovat data jako celek. Máte na základě  $2D$  vizualizací nějakou představu, zda by se například hodila shluková analýza? Co využití PCA?
- Porovnejte muže a ženy, použijte testy z tématu Vícerozměrné testy. Co byste o nich řekli?
- podívejte se na údaje o ženách po roce léčení. Co byste mohli říci o vývoji jejich stavu? Měla léčba nějaký významný efekt? pokud ano, jaký?