

Katedra aplikované matematiky si Vám v rámci [Statistické laboratoře](#) dovoluje nabídnout **sérii workshopů** na téma

# Statistická analýza dat se softwarem R

**Lektorky:**

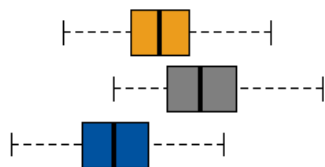
**Ing. Martina Litschmannová, Ph.D.**

**Mgr. Adéla Vrtková**

**Workshopy budou probíhat v LS 2016/2017,  
vždy v pátek od 9:00h do 10:30h, učebna EB 413 (FEI)**

Termín	Téma
17. února	Pravděpodobnost je...
24. února	Jak modelovat výsledky náhodných pokusů?
3. března	Vybraná rozdělení náhodné veličiny
10. března	Máme data – a co dál? (část 1)
17. března	Máme data – a co dál? (část 2)
24. března	Máme data – a co dál? (část 3)
31. března	Neurčitost v datech – bodové a intervalové odhady
7. dubna	Testování statistických hypotéz (princip, vybrané jednovýběrové testy)
21. dubna	Testování statistických hypotéz (vybrané dvouvýběrové testy)
28. dubna	Vícevýběrové testy par. hypotéz – jednoduchá analýza rozptylu
5. května	Úvod do analýzy kategoriálních dat – dvourozměrné kontingenční tabulky
12. května	Úvod do korelační a regresní analýzy

Kapacita workshopů je omezená. Počítáme s 20 místy pro zájemce o celou sérii workshopů a 10 místy pro zájemce o vybraný workshop. Přihlásit se na celou sérii, popř. na vybraný workshop, se můžete na <http://k470.vsb.cz/statlab/>.



## **Anotace jednotlivých workshopů:**

### **1. Pravděpodobnost je...**

V tomto workshopu si připomeneme, co je to pravděpodobnost a ukážeme si, jak řešit základní pravděpodobnostní úlohy. Procvičíme si vše na konkrétních příkladech tak, aby účastníci kurzu neměli následně problém s porozuměním „matematickému zápisu“.

### **2. Jak modelovat výsledky náhodných pokusů?**

Víte, co je to náhodná veličina, pravděpodobnostní funkce, distribuční funkce, hustota pravděpodobnosti, ...? Přístupnou formou zavedeme základní pojmy a následně vše procvičíme na praktických příkladech. Pro řešení příkladů začneme používat software R – zatím ve formě „inteligentní“ kalkulačky.

### **3. Vybraná rozdělení náhodné veličiny**

V rámci tohoto tématu se naučíme modelovat vybrané typy náhodných veličin. Ukážeme si, jak pracovat s binomickým, Poissonovým, rovnoměrným, exponenciálním a normálním rozdělením, včetně pravidla  $3\sigma$ . A to vše s využitím Rka.

### **4. Máme data – a co dál? (část 1)**

Statistika nuda je... Nebo není? Seznámíme se se základními pojmy z popisné statistiky a následně se zaměříme na vizualizaci a charakteristiky kategoriálních (nominálních a ordinálních) dat. Ukážeme si jak v Rku načíst datový soubor, jak data připravit pro následnou analýzu a jak analyzovat jednorozměrná kategoriální data.

### **5. Máme data – a co dál? (část 2)**

V návaznosti na předchozí lekci budeme pokračovat s popisnou statistikou numerických (poměrových) dat. V teoretické části lekce se zaměříme zejména na správnou interpretaci měř polohy a heterogenity dat a způsoby vizualizace. Následně vše procvičíme na praktických příkladech v Rku.

### **6. Máme data – a co dál? (část 3)**

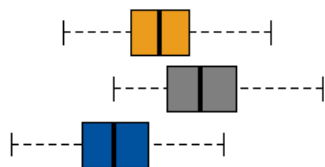
V této lekci dokončíme, co jsme nestihli a následně se budeme věnovat dalšímu procvičení popisné statistiky (explorační analýzy) a prohloubení praktických dovedností v Rku.

### **7. Neurčitost v datech – bodové a intervalové odhady**

Pomocí popisné statistiky zjišťujeme informace o našich datech (výběrových souborech). Statistika nám však umožňuje více. Na základě výběrového souboru dokážeme odhadnout (s jistou spolehlivostí), jak se chová příslušná populace. Ukázat jak na to je cílem tohoto workshopu.

### **8. Testování statistických hypotéz (princip, vybrané jednovýběrové testy)**

Pomocí intervalových odhadů odhadujeme na základě náhodných výběrů populační charakteristiky. Jiným způsobem, jak zjistit na základě náhodného výběru „něco“ o populaci, je testování hypotéz. Vysvětlíme si princip testování hypotéz, seznámíme se s pojmem p-hodnota a s typy chyb, které se při testování hypotéz mohou vyskytnout. Na vybraných jednovýběrových testech si ukážeme, jak se testování provádí v praxi (samozřejmě s využitím Rka).



## 9. Testování statistických hypotéz (vybrané dvouvýběrové testy)

Tento workshop přímo navazuje na předcházející. Seznámíme se s pojmy nezávislá a párová data a ukážeme si, jak testovat shodu středních hodnot (resp. mediánů) a shodu rozptylů u dvouvýběrových dat.

## 10. Vícevýběrové testy par. hypotéz – jednoduchá analýza rozptylu

Workshop je pokračováním v tématu Testování hypotéz. Zaměříme se na testy shody úrovně (stř. hodnot, mediánů) a ověření homogenity vícevýběrových dat, tj. ANOVu, Kruskalův – Wallisův test, Bartlettův a Leveneův test.

## 11. Úvod do analýzy kategoriálních dat – dvourozměrné kontingenční tabulky

Toto téma je věnováno základní analýze závislosti dvou kategoriálních proměnných. Ukážeme si, jak s pomocí Rka provést základní explorační analýzu, seznámíme se s vhodnými měrami asociace, ukážeme si, jak sledovanou závislost vizualizovat a jak testovat její statistickou významnost. Pokud nám to čas dovolí, ukážeme si i speciální přístupy k analýze dat zapsaných v asociačních tabulkách.

## 12. Úvod do korelační a regresní analýzy

Závěrečný workshop je věnován základní analýze závislosti numerických proměnných. Ukážeme si, jak měřit sílu závislosti mezi numerickými proměnnými a podrobněji se pak budeme zabývat základy lineární regrese.