

## Statistika

Tento předmět navazuje na předměty

Pravděpodobnost

Statistika

Časové řady

Bayesovské metody

Průzkumová mnohorozměrná statistika

Vybrané metody strojového učení

## SZZDS

### 1. Bodové a intervalové odhady a princip testování statistických hypotéz.

Náhodný výběr, parametrický prostor (příklady), bodové odhady střední hodnoty a rozptylu, vlastnosti bodových odhadů, metody konstrukce bodových odhadů; definice a interpretace intervalového odhadu, příklad konstrukce intervalového odhadu; postup pro testování hypotéz, chyba 1. a 2. druhu, hladina testu, kritický obor, p-hodnota.

### 2. Statistická analýza vztahů dvou veličin (kvantitativních, kvalitativních).

Dva kvalitativní znaky: testy nezávislosti (asymptotický a exaktní) v kontingenčních tabulkách, poměr šancí; použití McNemarova testu a testu homogenity

Kvalitativní a kvantitativní znak: dvouvýběrové testy o shodě středních hodnot a o shodě rozptylů za předpokladu normality, dvouvýběrový Wilcoxonův test, jednofaktorová ANOVA (předpoklady, testovaná hypotéza, myšlenka konstrukce testové statistiky), myšlenka mnohonásobného porovnávání, Kruskalův-Wallisův test, použití Leveneova testu

### 3. Bayesova věta a její použití, apriorní a aposteriorní rozdělení, Monte Carlo metody.

Rozdíl mezi inferencí, predikcí a rozhodováním. Odvození Bayesovy věty a její aplikace. Testování hypotéz versus bayesovský přístup. Srovnávání více modelů. Exaktní metody, Laplaceova aproximace, variační metody pro získání posterioru. Podstata a smysl Monte Carlo metod, příklady samplingu. Baysovský přístup k regresi.

### 4. Klasifikace, klasické metody a metody pro vysoce-dimenzionální data.

Úloha klasifikace. Lineární diskriminační analýza, Fisherova diskriminační analýza, logistická regrese, jejich vzájemný vztah. Metoda k nejbližším sousedům. Metoda částečných nejmenších čtverců a její použití pro diskriminační analýzu, výpočetní algoritmy. Vyhodnocení kvality klasifikace, role křížové validace při výběru modelu.

### 5. Regresní analýza a její aplikace, konstrukce a verifikace modelů, korelace.

Korelační koeficient – definice, obor hodnot, interpretace, odhad, podmínky využití k testování nezávislosti dvou náhodných veličin, Spearmanův korelační koeficient – definice, interpretace, srovnání s Pearsonovým k.k.; model regresní přímky a jeho předpoklady, odhad parametrů (ilustrovat graficky), vektorový zápis modelu regresní přímky, vlastnosti odhadů; residua a jejich využití, residuální součet čtverců, koeficient determinace a jeho interpretace. Vztah mezi regresí a korelací.

**6. Průzkumová statistická analýza, shlukování, redukce dimenze.**

Předzpracování dat (standardizace, role měřítka), robustní statistika a detekce odlehlých hodnot. Hierarchické shlukování, metoda k průměrům, modelové shlukování, hodnocení kvality shlukování. Singulární rozklad datové matice, metoda hlavních komponent (včetně její robustní verze) a faktorová analýza, vztahy mezi nimi.

**7. Modelování časových řad, trend a sezónnost, náhodná složka, srovnání s Box-Jenkinsovou metodologií.**

Princip dekompozičního přístupu (aditivní versus multiplikatívni), různé funkce pro modelování trendu, adaptivní modelování trendu, model konstantní sezónnosti a proporcionální sezónnosti (+odhady parametrů v těchto modelech), předpoklady o náhodné složce v dekompozičním přístupu; (slabá) stacionarita časové řady, srovnání dekompozičního přístupu s Box-Jenkinsovým přístupem.

**8. Supervised machine learning, neuronové sítě, support vector machines, decision trees.**

Supervised machine learning – vysvětlení problematiky, typy řešených úloh. Neuronové sítě – biologická motivace, konstrukce, algoritmy pro trénování, predikce, aplikace. Support Vector Machine – motivace, princip, aplikace. Decision trees – motivace, princip, aplikace, modifikace (např. random forests). Srovnání metod mezi sebou a srovnání s alternativními přístupy k řešení daných problémů.