

Studijní témata

1. kapitola: Základní pojmy matematické statistiky

Při aplikaci metod popisné statistiky dospíváme pomocí zjištěných dat k závěrům, které se týkají pouze výběrového souboru. Naproti tomu matematická statistika nám umožňuje na základě znalosti náhodného výběru a statistik z něj odvozených (tj. např. výběrového průměru, výběrového rozptylu, výběrového koeficientu korelace, hodnoty výběrové distribuční funkce apod.) učinit závěry o parametrech nebo tvaru rozložení, z něhož daný náhodný výběr pochází. Často se jedná o bodové či intervalové odhady parametrů a parametrických funkcí a testování hypotéz o nich.

2. kapitola: Uspořádání pokusů

Abychom mohli správně vyhodnotit výsledky pokusu, musí být pokus dobře naplánován. V závislosti na záměrech experimentátora rozeznáváme několik typů uspořádání pokusů: jednoduché pozorování (zkoumají se hodnoty náhodné veličiny pozorované za týchž podmínek), dvojné pozorování (zkoumá se rozdílnost hodnot náhodné veličiny pozorované za dvojích různých podmínek) a mnohonásobné pozorování (zkoumá se rozdílnost hodnot náhodné veličiny pozorované za $r \geq 3$ různých podmínek). Podle typu uspořádání pokusu pak volíme vhodnou statistickou metodu.

3. kapitola: Diagnostické grafy a testy normality dat

Diagnostické grafy slouží především k tomu, aby nám pomohly orientačně posoudit povahu dat a určit směr další statistické analýzy. Při zpracování dat se často předpokládá splnění určitých podmínek. V případě jednoho náhodného výběru je to především normalita (posuzujeme ji pomocí NP plotu, Q-Q plotu, histogramu) a nepřítomnost vybočujících hodnot (odhalí je krabicový diagram). U dvou či více nezávislých náhodných výběrů sledujeme kromě normality též shodu středních hodnot nebo shodu rozptylů - homoskedasticitu (porovnáváme vzhled krabicových diagramů). V případě jednoho dvourozměrného náhodného výběru často posuzujeme dvourozměrnou normalitu dat (použijeme dvourozměrný tečkový diagram s proloženou $100(1-\alpha)\%$ elipsou konstantní hustoty pravděpodobnosti).

Vzhledem k důležitosti předpokladu normality se vedle grafického posouzení doporučuje též použití některého testu normality, např. Kolmogorovova – Smirnovova testu nebo Shapirova – Wilkova testu. K závěrům těchto testů však přistupujeme s určitou opatrností. Máme-li k dispozici rozsáhlejší datový soubor (orientačně $n > 30$) a test zamítne na obvyklé hladině významnosti 0,01 nebo 0,05 hypotézu o normalitě, i když vzhled diagnostických grafů svědčí jenom o lehkém porušení normality, nedopustíme se závažné chyby, pokud použijeme statistickou metodu založenou na normalitě dat.

4. kapitola: Úlohy o jednom náhodném výběru z normálního rozložení

Mnoho náhodných veličin, s nimiž se setkáváme ve výzkumu i praxi, se řídí normálním rozložením. Za jistých předpokladů obsažených v centrální limitní větě se dá rozložení jiných náhodných veličin aproximovat normálním rozložením. Proto je zapotřebí věnovat velkou pozornost právě náhodným výběrům z normálního rozložení.

Normální rozložení je charakterizováno dvěma parametry – střední hodnotou μ a rozptylem σ^2 . Budeme tedy řešit úlohy, které se týkají těchto parametrů. Jedná se především o jednovýběrový t-test či test o rozptylu. Seznámíme se rovněž se situací, kdy máme k dispozici jeden náhodný výběr z dvourozměrného normálního rozložení a posuzujeme rozdílnost středních hodnot obou náhodných veličin. K řešení tohoto problému slouží párový t-test.

5. kapitola: Úlohy o dvou nezávislých náhodných výběrech z normálních rozložení

V tomto případě je naším úkolem porovnat střední hodnoty či rozptyly dvou normálních rozložení na základě znalosti dvou nezávislých náhodných výběrů pořízených z těchto rozložení. Zpravidla konstruuje intervaly spolehlivosti pro rozdíl středních hodnot nebo podíl rozptylů respektive hodnotíme shodu středních hodnot pomocí dvouvýběrového t-testu a shodu rozptylů pomocí F-testu.

6. kapitola: Analýza rozptylu jednoduchého třídění

Situace je charakterizována aspoň třemi nezávislými náhodnými výběry pocházejícími z normálních rozložení, která mají stejný rozptyl, ale mohou se lišit ve středních hodnotách. Zajímá nás, zda jsou všechny střední hodnoty stejné. Pokud se prokáže jejich odlišnost, zjišťujeme pomocí metod mnohonásobného porovnávání, které dvojice středních hodnot se liší na dané hladině významnosti.

7. kapitola: Pořadové testy o mediánech

Při používání t-testů či analýzy rozptylu by měl být splněn předpoklad normality dat. Pro výběry větších rozsahů ($n \geq 30$) nemá mírné porušení normality závažný dopad na výsledky. Někdy se však setkáváme s výběry malých rozsahů, které pocházejí z výrazně nenormálních rozložení. Pro práci s nimi byly vytvořeny tzv. neparametrické testy, které nevyžadují předpoklad o konkrétním typu rozložení (např. normálním), stačí např. předpokládat, že distribuční funkce rozložení, z něhož náhodný výběr pochází, je spojitá.

Tyto neparametrické testy se rovněž používají v situacích, kdy zkoumaná data nemají intervalový či poměrový charakter, ale pouze ordinální charakter.

Ve srovnání s klasickými parametrickými testy jsou však neparametrické testy slabší, tzn., že nepravdivou hypotézu zamítají s menší pravděpodobností než testy parametrické.

V této kapitole se omezíme na ty neparametrické testy, které jsou založeny na pořadí a týkají se mediánů. Nazývají se pořadové testy.

8. kapitola: Analýza závislosti dvou veličin

Při zpracování dat se velmi často setkáme s úkolem zjistit, zda dvě náhodné veličiny jsou stochasticky nezávislé. Např. nás může zajímat, zda ve sledované populaci je barva očí a barva vlasů nezávislá nebo zda počet dnů absence a věk pracovníka jsou nezávislé. Testování hypotézy o nezávislosti se provádí různými způsoby podle toho, jakého typu jsou dané náhodné veličiny – zda jsou nominální, ordinální, intervalové či poměrové.

Při zkoumání závislosti je nesmírně důležité provést logický rozbor problému. Nemá smysl se zabývat hledáním závislosti v případech, když

- z logických důvodů nemůže existovat,
- závislost je způsobena formálními vztahy mezi veličinami,
- soubor dvourozměrných dat je nehomogenní,
- závislost je způsobena společnou příčinou.

Zpravidla chceme také zjistit intenzitu případné závislosti sledovaných dvou veličin. K tomuto účelu byly zkonstruovány různé koeficienty, které nabývají hodnot od 0 do 1 (resp. od -1 do 1). Čím je takový koeficient bližší 1 (resp. -1), tím je závislost mezi danými dvěma veličinami silnější a čím je bližší 0, tím je slabší.