

Induktivní statistika (statistická inference, statistické usuzování)

Cílem stat. usuzování je odvodit na základě dat týkajících se výběrového souboru závěry o celé populaci. Zobecnění z náhodného výběru na populaci s jistým stupněm jistoty (spolehlivosti).

Dvě základní metody – odhady a testování hypotéz.

Parametry – číselné hodnoty, jež platí pro celou populaci, teoretické, často nejsou známy, označují se řeckými písmeny.

Statistiky – odhady parametrů, získáváme je z dat výběrového souboru, označují se latinkou.

ODHADY

Bodové odhady x Intervalové odhady

Intervaly spolehlivosti – kratší intervaly jsou přesnější než dlouhé.

Hladina spolehlivosti je pravděpodobnost, s jakou se odhadovaný populační parametr (statistika) ocitne v intervalu spolehlivosti při opakovaném provádění výběru. Nejběžněji používané hladiny jsou 95% nebo 99% (konsensus).

Alfa je pravděpodobnost, s níž výzkumník toleruje, že interval spolehlivosti nepokryje daný parametr. Pak $(1-\text{alfa}) \cdot 100\% = \text{hladina spolehlivosti}$.

TESTOVÁNÍ HYPOTÉZ

Nulová hypotéza – předpoklad, že vybrané charakteristiky zkoumaných souborů dat se od sebe v zásadě neliší. Tuto hypotézu testujeme (statistickými procedurami), snaha o její zamítnutí, tudíž můžeme přijmout její negaci (tzn. zkoumané soubory dat se od sebe liší).

Určení hladiny významnosti (alfa) – pravděpodobnost, že se zamítne nulová hypotéza, přestože platí. Tuto hladinu si určuje výzkumník (resp. vědecká komunita).

Hodnota významnosti p – odpovídá na otázku: „Jestliže nulová hypotéza platí, jaká je pravděpodobnost, že získáme právě vypočítanou hodnotu nebo ještě neobvyklejší hodnotu testovací statistiky?“ Hodnota p tedy kvantifikuje pravděpodobnost realizace hodnoty testovací statistiky, pokud nulová hypotéza platí. Jestliže je nízká, je to důkaz toho, že nulová hypotéza neplatí. **Jestliže p hodnota je menší než hladina alfa nebo se jí rovná, data přinášejí doklad pro zamítnutí nulové hypotézy.** Pokud je p hodnota větší než alfa, nulová hypotéza se přijímá (respektive vystavuje dalšímu ověřování).

Párový t-test

Dle odpovědi – porovnání zda se liší proměnná x dle proměnné pohlaví.

Dle proměnných – srovnání výsledků před návštěvou knihovny a po návštěvě knihovny.

Korelace

Zjišťuje míru vztahu mezi dvěma proměnnými.

Nabývá rozmezí od -1 do +1, příklady, grafy.

Použití vhodného korelačního koeficientu závisí na typu proměnných, s nimiž pracujeme.

Součinný koeficient korelace r pro intervalové proměnné. Spearmanův koeficient R , pokud jsou obě proměnné pořadové.