

Statistika a ztroskotání Titaniku

Autor si zvolil zajímavé téma a vzhledem ke kulatému výročí této námořní katastrofy i téma aktuální. Podkladem mu byl rozsáhlý soubor základních údajů o osazenstvu lodi během její krátké plavby. Při pohledu do této databáze se nabízí otázka, na niž se autor snaží najít (statisticky podloženou) odpověď: Jaké byly šance na přežití pro jednotlivé kategorie cestujících?

Kromě zajímavě položených otázek je na práci třeba ocenit i snahu o jejich zodpovězení různými postupy. Kromě odhadu podmíněných pravděpodobností pomocí relativních četností pro jednotlivé kategorie autor zapojil také test nezávislosti v kontingenčních tabulkách a poměr determinace. Práce má pěknou úpravu, málo překlepů, řadu tabulek a vyčerpávající přílohu všech cestujících.

Nyní k nedostatkům. Asi za největší považuji, že se autor nijak nepokusil sjednotit značení a pojmy, které přejímal z několika zdrojů. Často doslova cituje celé dlouhé pasáže z učebnic. Týká se to i praktické části, kde se s každou metodou mění symbolika. Druhá hlavní výtká se týká rozvržení sil. Kapitoly 3-5 trpí nedostatkem malých četností, a proto měly posloužit spíše jako předběžná analýza. Není možné tvrdit, že prst přežití je vysoká pro kategorii, ve které je pouze jediná osoba! Nová kategorizace v kapitole 6 je mnohem rozumnější a výsledky důvěryhodnější – toto mohl autor rozvést a zdůraznit jako stěžejní část práce.

Dále některé konkrétní otázky a připomínky:

- str. 7 ... V našem případě je $\Omega = \{0,1\} = \{\text{přežil, nepřežil}\}$, $A = \{\text{přežil}\}$, $P(A)$ počítáme jako frekvenční prst pomocí relativní četnosti téhož výsledku (přežil) v mnoha pokusech (na mnoha osobách).
- str. 12 ... Vhodnější název pro podkapitolu 1.2: Testování hypotéz.
- str. 17 ... Jaký je rozdíl mezi kontingenční a korelační tabulkou?
- str. 24 ... Nešťastně zvolené pojmy: vaše elementární četnosti se správně nazývají sdružené četnosti, vaše sdružené četnosti jsou správně marginální četnosti. Tudíž místo sdružená podmínka by bylo vhodnější např. komplexní, vícehodnotová či sloučená podmínka.
- str. 25 ... Čtenáři by určitě chtěli vidět i marginální četnosti.
- str. 24n ... Vypisování všech možných sčítaných variant mi přijde zbytečně obšírné. Rovněž tak počítané prsty mohly být úsporněji vtěsnány do tabulky o menším počtu stran, s kódováním kategorií pohotově v legendě.
- str. 29, 46, 52 ... Stále mi není jasné, co jste prováděl s kapitánem.
- str. 30 ... Vzorec je zbytečně složitý: po druhé rovnosti může následovat poslední.
- str. 33 ... Lépe objasněte mezní hodnoty intervalu.
- str. 46n ... Chybí číslování tabulek, průvodní text je zde příliš skrovný.
- str. 47 ... Pozor, chybné stupně volnosti!
- str. 48 ... Dá se říct, která závislost vyšla silnější?
- str. 54 ... Elementární podmínky rozlišují všechny **tři** proměnné (znaky).
- str. 55 ... Srovnej výsledky s kapitolou 3 (na str. 58 s kapitolou 4).

- str. 59n ... Správná interpretace poměru determinace je vysvětlená část **variability** znaku, nikoli hodnoty znaku nebo hodnoty prstí.
- str. 60 ... Přestože interpretace získaných čísel je subjektivní, mně se hodnoty korelačního poměru 0,46 a 0,51 jeví jako dostatečně vysoké a s prokázanou závislostí bych byl spokojený.

Ačkoli se autor často dostával do rozporu se značením a zavedenou teorií, oceňuji jeho snahu o vlastní uchopení problému, hledání alternativních cest k odpovědím a podchycení všech možných kategorií. Se získanými závěry ohledně významných kategorií (ovlivňujících přežití) je možno souhlasit. Práci doporučuji k obhajobě, navrhuji hodnocení **velmi dobře (C)**.

V Šumperku 15. 5. 2012

RNDr. PhDr. Ivo Müller, Ph.D.
oponent