

Bohumír Procházka

Biostatistika pro lékaře – principy základních metod a jejich interpretace s využitím statistického systému R

Kniha je určena lékařům, popisuje základní statistické postupy a ukazuje interpretaci získaných výsledků. V knize jsou zmíněny i chyby, kterých se může lékař při analýze dat dopustit.

Lze ji doporučit především těm, kdo se zajímají o „evidence-based medicine“. Text knihy je možno číst i tak, že se čtenář soustředí na způsob interpretace výsledků a předpoklady pro použití jednotlivých statistických metod, aniž by se vyčerpával studiem, jak programovat analytický balík R. Je užitečná nejen pro lékaře, ale i pro další profese, jak k seznámení se statistickými metodami, tak i jako návod k nalezení společného jazyka mezi lékařem a analytikem. Jejich neporozumění v praxi přináší mnoho interpretačních problémů, jak v oblasti volby adekvátních metod, tak i v správné interpretaci.

Poslední kategorií čtenářů, pro které bude text velmi užitečný, jsou i samotní statistici a biostatistici, protože poukazuje na rozdíly v jazyce, kterým se vyjadřují lékaři a statistici. Což samozřejmě platí také obráceně, tj. i pro lékaře. Zde je třeba podtrhnout autorovy letité zkušenosti s výukou biostatistiky na lékařských fakultách.

Autor pokryl prakticky celé spektrum problémů, na které lékař může, při čtení lékařské literatury prezentující data, narazit a to jazykem, který je srozumitelný výše definovaným cílovým skupinám.

Používaný program R je volně šiřitelný poskytovatelem R Foundation for Statistical Computing, Vienna, Austria, <http://www.R-project.org/>. Kniha seznamuje s prací v tomto programu a zbavuje strachu z provádění výpočtů. Tento program je používán nejen analytiky, ale i lékaři. Kniha se pokouší zaplnit prostor v české literatuře v oblasti popisu práce s programem R. Zdaleka však není pouhým návodem k používání programu R, ale zaměřuje se především na interpretaci a je tedy přínosná i pro uživatele jiných statistických systémů.

Recenzenti: prof. RNDr. Jan Hendl, CSc., MUDr. Zdeněk Šmerhovský, Ph.D.

Z obsahu knihy:

Předkládaná monografie navazuje na dříve publikovanou knihu „Stručná biostatistika pro lékaře“ stejného autora. Tato kniha je rozšířena jednak o popis práce s volně šiřitelným programem R, ale hlavně o příklady chyb při použití statistických metod. V textu celé knihy jsou průběžně čarou a symbolem R označeny bloky obsahující příkazy tohoto programu a jeho výstupy. Podrobnější základy s programem R jsou v první příloze. Kniha obsahuje i CD s cvičnými daty a skripty pro výpočty v jednotlivých kapitolách.

Kniha se skládá z jednotlivých kapitol:

První kapitola uvádí problematiku statistického uvažování a jeho použití v jiných oblastech vědy. Krátce zmiňuje i historický rámec vzniku matematické statistiky a teorie pravděpodobnosti.

V druhé kapitole jsou obecné úvahy o rozdílném přístupu biologie a matematiky k řešení problémů reálného světa. Je zde především popsán způsob induktivního uvažování, který je zásadní pro použití a zobecnění získaných zkušeností. Jsou zde popsány i základní úvahy o vztahu populace a výběru.

Třetí kapitola se zaměřuje na popis různých typů sledovaných veličin a ukazuje jejich příklady.

Čtvrtá kapitola popisuje různé typy statistických charakteristik míry polohy jako je například aritmetický průměr nebo medián a míry polohy jako je směrodatná odchylka a střední chyba průměru a zmiňuje jejich rozdílnost.

Pátá kapitola popisuje matematické modely pro rozložení různých typů veličin.

Šestá kapitola se zabývá popisem vlastností odhadů statistických charakteristik. Podrobně popisuje principy intervalových odhadů a jejich interpretaci. Popisuje rozdíly mezi intervalem spolehlivosti a tolerančním nebo predikčním intervalem. Poukazuje na jejich chybnou konstrukci a interpretaci. Dále je zde popsán obecný princip konstrukce testů včetně popisu P-hodnoty, která je používána pro hodnocení statistické významnosti prováděných testů.

Sedmá kapitola popisuje jak ověřit typ rozložení analyzovaných veličin, což je důležitý předpoklad pro provádění statistických testů.

Osmá kapitola se zabývá testováním rozdílu mezi sledovaným výběrem a pevnou hodnotou a s tím související konstrukcí různých intervalových odhadů.

Devátá kapitola popisuje testování rozdílu mezi dvěma skupinami měření. Je zde popsán test jak pro veličinu normálně rozloženou tak i test bez tohoto předpokladu. Je zmíněn i nejjednodušší problém opakovaného měření - párové testy.

V desáté kapitole je nejprve popsán kauzální i nekauzální vztah dvou spojitých veličin. Je popsán lineární regresní model a korelační koeficient, je zde zmínka jak o měření nejen lineárního, ale i monotónního vztahu. Zmíněny jsou i složitější modely vztahu více veličin a nelineárního vztahu. V kapitole jsou i zmíněny problémy s kterými se v těchto modelech setkáváme.

Jedenáctá kapitola se zabývá zobecněním porovnávání dvou skupin na porovnání více skupin, případně vytváření složitějších modelů obsahujících více veličin a to i spojitých. Je zmíněno i rozšíření modelu o interakce (společný vliv více veličin). Popisuje i vícenásobné srovnávání. Další

problematikou je opakované měření a modely s náhodnými efekty (kdy hodnoty některých vysvětlujících veličin jsou náhodné veličiny).

Dvanáctá kapitola se zabývá problematikou často používanou v oblasti epidemiologie a to analýzou vztahu dvou kvalitativních veličin. Je zde popsána kontingenční tabulka a jak testovat nezávislost. Jsou popsány i míry závislosti v kontingenční tabulce relativní riziko a odds ratio. Zmíněný je i test pro opakované sledování kvalitativní veličiny a test trendu v kontingenční tabulce. Jsou zde popsány typy studií a jejich způsob hodnocení. Krátce je zmíněna i podobnost testů pro kvantitativní a kvalitativní veličiny. Je zmíněn i problém rušivého faktoru, hodnocení screeningového testu a ROC křivky.

Třináctá kapitola popisuje vliv rušivých faktorů na studii a eliminaci tohoto vlivu způsobem sběru dat (náhodný, systematický, kvótní a vícestupňový výběr), plánem studie, volbou kontrolní skupiny nebo uspořádáním studie (např. párové uspořádání nebo křížový pokus). Jsou zde zmíněny i vážené odhady a různé způsoby standardizace nemocnosti a konstrukce jejich intervalů spolehlivosti. V neposlední řadě se zde autor věnuje i problematice stanovení rozsahu výběru.

Čtrnáctá kapitola se zabývá obecným lineárním modelem, tedy modelem, kde vysvětlovaná proměnná má jiné než normální rozložení. Nejprve je zde popsána logistická regrese – model závislosti binomické veličiny na více charakteristikách jak kategoriálních tak i kvantitativních. Tento model je často používán pro vyloučení vlivu rušivého faktoru v kontingenční tabulce, nebo na modelování závislosti binomické veličiny na kvantitativní (např. závislost pravděpodobnosti onemocnění na věku). Je zmíněna i problematika výpočtu ED₅ či LD₅₀ (padesátiprocentní účinné, či smrtelné dávky), nebo poissonovská regrese.

Patnáctá kapitola se zabývá analýzou cenzorovaných dat - analýzou přežití, tedy daty s neúplnou informací. Jsou zde zmíněny tabulky přežití, ale i další metody jako je Kaplan-Mayerův odhad nebo Coxův semiparametrický model proporcionálního hazardu. Dalším řešením je plně parametrický model, například exponenciálního umírání nebo model kombinující dvě populace s různou rychlostí umírání. Je zde předvedena i analýza dat s hodnotami pod detekčním limitem pro veličinu s normálním či logaritmicke-normálním rozložením. V závěru této kapitoly autor popisuje netradiční použití metod pro analýzu cenzorovaných dat v epidemiologii pro nalezení epidemického prahu, či pro odhad počtu úmrtí zvýšeného výskytem epidemie nebo pro nalezení týdnů s nadměrným výskytem (s vyslovením podezření na epidemii) .

První příloha obsahuje základy práce s programem R.

Druhá příloha krátce vyjmenovává různé programy pro statistickou analýzu dat.

Třetí příloha ukazuje různé příklady využití grafů pro presentaci výsledků především z pohledu, jakých se můžeme dopustit chyb a jak mohou být nevhodně použité grafy matoucí.

Čtvrtá kapitola popisuje na reálných příkladech z praxe chyby, nesprávného použití statistických metod a chyby v interpretaci, kterých se můžeme dopustit.

Poslední příloha odkazuje na příložené CD s cvičnými daty a skripty programu R pro provádění výpočtů z jednotlivých kapitol.