

# Předmluva

Text je určen především studentům matematické statistiky či ekonometrie na mat-fyzy Univerzity Karlovy v Praze. Doufám však, že bude užitečný i dalším zájemcům, kteří budou ochotni překonat použitý formalismus. U čtenáře předpokládám znalost základů lineární algebry, přesto nejdůležitější potřebná tvrzení shrnuji v appendixu. Předpokládám samozřejmě také znalost základů matematické statistiky.

Nijak nechci skrývat dojem, který na mě udělalo programové prostředí R, proto jsem jej učinil svým hlavním fámulem. Doufám, že jsem tím neodradil příliš zájemců o regresní model, *erko opravdu stojí za to!* Není to jen další program pro statistické výpočty, je to opravdu celé prostředí pro statistické výpočty. Prostředí schopné se přizpůsobit, vhodné pro interaktivní práci, prostředí volně šířitelné. Impozantní (až nepřehledný) je počet dostupných knihoven. V létě roku 2007 jsem jich napočítal zhruba 1080. Až tento text vyjde, uvedené číslo jistě bude od skutečnosti hodně daleko. Stačí se podívat např. na <http://cran.at.r-project.org/>. Tam se najde nejen samotný program, ale i řada pomůcek, které zájemci usnadní vniknout do tajů programu.

Tento text má několik cílů. Není to učebnice prostředí R, není to kuchařka s návodem, jak aplikovat teorii lineárních modelů. Kladl jsem si za cíl vyložit principy, na nichž lineární modely spočívají a současně ukázat, jak lze regresní modely i modely analýzy rozptylu aplikovat při řešení statistických úloh. Chtěl jsme tedy naznačit nejen, jak se něco počítá, ale také proč to má fungovat. Navíc jsem se snažil popsat dva modely, které se aproximují pomocí lineárního modelu, totiž model logistické regrese a model nelineární regrese. Pokud snad někdo hledá jen použití erkového prostředí k aplikaci lineárních modelů, nechť sáhne například po knížce Faraway (2005), pokud snad čtenáře zajímají také vnitřnosti tohoto prostředí, může se poučit například v knížce Fox (2002). Mnohem širší záběr statistických postupů obsahuje v mnoha vydáních postupně rozšiřovaná kniha Venables, Ripley (1997).

Ještě několik technických detailů. Konce důkazů jsou označeny klasickým čtverečkem. Podobně je na konci každého příkladu umístěn kroužek. Číslované poznámky jsou psány menším písmem. Ostatně, číslování vzorců, vět, definic, tabulek, obrázků i poznámek v sobě nese informaci o kapitole. Vektory jsou zásadně sloupcové a jsou psány tučně. K označení matic konstant jsou použita velká písmena psaná tučným bezpatkovým písmem, podobně vektory konstant jsou označeny malými písmeny psanými bezpatkovým písmem. Výjimkou je náhodná složka lineárního modelu. Jde o náhodnou veličinu, takže by měla být psána velkým písmenem. Když jsem nechtěl použít někdy užívané řecké  $\epsilon_i$ , hrozila při použití velkého  $E_i$

záměna s operátorem střední hodnoty. Proto jsem se rozhodl pro u nás častější malé  $e_i$ . Na podobný problém jsem narazil u indexů výběrových korelačních koeficientů. Nakonec jsem zpravidla použil odpovídající malá písmena, výjimkou je zavedení koeficientu determinace, kde byl kladem důraz na souvislost mezi  $\mathbf{Y}$  a  $\mathbf{Y}$ . Výstupy z programu R jsou z typografických důvodů poněkud upravovány, zejména jsou zrušeny prázdné řádky. Ještě poznámka k indexu na konci knihy. Jde vlastně o tři indexy v jednom, položky se liší použitým písmem. V první části (matematická kurzíva) je uvedeno několik symbolů označujících součty čtverců apd., druhá část (neproporciální písmo) obsahuje jména použitých knihoven, erkových funkcí a jejich parametrů, teprve třetí část je klasickým rejstříkem pojmů.

Data použitá v příkladech včetně příkazových souborů všech výpočtů vystavím na své internetové stránce <http://www.karlin.mff.cuni.cz/~zvara/>.

Děkuji oběma recenzentům za shovívavé posudky, za pečlivé přečtení rukopisu, za mnohá užitečná doporučení a za upozornění na spoustu nedopatření. Ostatně, řadu překlepů pomohly odstranit generace studentů, kteří pracovní verze tohoto textu používali jako učební pomůcku. Také jim děkuji. Za všechny nedostatky, které v textu zůstaly, jsem odpovědný já sám.

5. května 2008, KZv.

## Ke druhému vydání

V tomto vydání byla opravena řada tiskových chyb a podobných nedopatření. Za jejich vyhledání děkuji především prof. Jiřímu Andělovi, ing. Jiřímu Morávkovi, ale také řadě studentů, kteří aspoň některé části knížky opravdu přečetli. Programové prostředí R se od prvního vydání knížky mohutně rozrostlo, ale v tomto ohledu jsem žádné úpravy neudělal. Věřím, že si čtenář dokáže vyhledat modernější balíčky sám. Základní balíčky, s nimiž jsem před lety pracoval, jistě fungují dál. Data použitá v příkladech včetně příkazových souborů všech výpočtů jsou k dispozici na přiloženém CD.

1. října 2019, KZv.