

Obsah

1 Úvod	9
2 Obecné úvahy	11
2.1 Přístupy k řešení problémů	12
2.2 Populace a výběr – základ statistické indukce	12
3 Typy sledovaných veličin	15
3.1 Co můžeme sledovat	15
3.2 Typy náhodných veličin	16
3.2.1 Alternativní veličiny	16
3.2.2 Nominální veličiny	17
3.2.3 Ordinální veličiny	17
3.2.4 Kvantitativní veličiny	18
3.2.5 Celočíselné veličiny	20
4 Základní statistické charakteristiky	21
4.1 Míry pro kvalitativní veličiny	21
4.1.1 Pravděpodobnost	22
4.1.2 Relativní četnost	22
4.2 Míry polohy	23
4.2.1 Průměr (aritmetický)	23
4.2.2 Geometrický průměr	24
4.2.3 Medián	24
4.2.4 Modus	24
4.2.5 Useknutý průměr	24
4.2.6 Kvantil	25
4.3 Míry měřítka	25
4.3.1 Rozptyl	25
4.3.2 Rozpětí	26
4.3.3 Mezikvartilové rozpětí	26
4.3.4 Variační koeficient	27
4.4 Ostatní charakteristiky	27
4.4.1 Šikmost – skewness	27
4.4.2 Špičatost – kurtosis	27
5 Modely náhodné veličiny – rozložení pravděpodobnosti	29
5.1 Nominální veličiny	29
5.2 Diskrétní (celočíselné) kvantitativní veličiny	29
5.2.1 Binomické rozložení	29
5.2.2 Multinomické rozložení	30
5.2.3 Poissonovo rozložení	30
5.2.4 Negativně binomické (Pascalovo) rozložení	30
5.2.5 Nakažlivá rozložení	30

5.3	Spojité kvantitativní veličiny	31
5.3.1	Normální (Gaussovo) rozložení	31
5.3.2	Logaritmicko-normální rozložení	31
5.3.3	Exponenciální rozložení	32
5.3.4	Weibullovo rozložení	32
5.3.5	Rovnoměrné rozložení	32
5.3.6	Logistické rozložení	32
5.4	Výběrová rozložení – rozložení testovacích statistik	32
5.4.1	χ^2 -rozložení	33
5.4.2	Studentovo t-rozložení	33
5.4.3	Fisherovo F-rozložení	33
6	Statistické odhady a testy – základní principy	35
6.1	Odhady populačních charakteristik	35
6.2	Bodové odhady	36
6.3	Intervalové odhady	36
6.3.1	Intervalové odhady populačních charakteristik – intervaly spolehlivosti	36
6.3.2	Intervalové odhady – predikční intervaly	38
6.3.3	Intervalové odhady – toleranční intervaly	38
6.4	Rozdíl interpretace intervalu spolehlivosti a tolerančního intervalu	38
6.5	Statistické testy	40
7	Ověřování typu rozložení dat – klíč k volbě modelu	43
7.1	Grafické zobrazení výběrového rozložení	43
7.2	Testy k ověření typu rozložení	44
7.2.1	χ^2 testy dobré shody	44
7.2.2	Kolmogorovův-Smirnovův test	44
7.2.3	Test normality Shapirů-Wilkův	45
7.2.4	Další možnosti	45
7.3	Význam znalosti typu rozložení	45
8	Porovnání kvantitativní veličiny jednoho výběru s pevnou hodnotou	47
8.1	Testy charakteristik	47
8.1.1	Jednovýběrový Z-test	47
8.1.2	Jednovýběrový t-test	48
8.1.3	Jednovýběrový znaménkový (mediánový) test	48
8.1.4	Jednovýběrový Wilcoxonův test	49
8.2	Intervalové odhady	49
8.2.1	Intervaly spolehlivosti	49
8.2.2	Predikční intervaly	50
8.2.3	Toleranční intervaly	50
9	Porovnání kvantitativní veličiny ve dvou různých výběrech	51
9.1	Dvě skupiny	51
9.1.1	Dvouvýběrový t-test	52
9.1.2	Porovnání dvou rozptylů	52
9.1.3	Dvouvýběrový znaménkový test (mediánový)	52
9.1.4	Dvouvýběrový Wilcoxonův test	52
9.2	Párové porovnání	52
9.2.1	Párový t-test	53
9.2.2	Párový znaménkový test	53

9.2.3	Párový Wilcoxonův test	53
10	Analýza vztahu dvou spojitých veličin	55
10.1	Společné rozložení dvou veličin	55
10.2	Kovariance – míra lineárního vztahu dvou veličin	56
10.3	Koeficient lineární korelace	57
10.4	Robustní varianty korelačních koeficientů	57
10.4.1	Spearmanův koeficient monotónní korelace	59
10.4.2	Kendallův koeficient monotónní korelace	59
10.5	Praktické ukázky různých typů závislostí	59
10.6	Lineární regresní model	60
10.6.1	Lineární regresní model normálně rozložené náhodné veličiny	61
10.6.2	Regresní modely procházející počátkem (bez interceptu) – regrese procházející počátkem	61
10.6.3	Oblasti spolehlivosti – intervalové odhady	63
10.6.4	Problémy s linearitou a normalitou – transformace modelu	64
10.6.5	Ověření předpokladu lineárního regresního modelu	65
10.6.6	Odlehlá pozorování v regresi	65
10.7	Vztah více než dvou veličin	67
10.7.1	Vícenásobná regrese	67
10.7.2	Korelace více veličin	68
10.7.3	Polynomiální regrese	68
10.8	Nelineární regrese	68
10.9	Robustní regresní metody	69
10.10	Metody vyhlazování časových řad	69
11	Porovnání kvantitativní veličiny ve více skupinách – Analýza rozptylu – ANOVA	71
11.1	Podmínky použitelnosti analýzy rozptylu	72
11.1.1	Test shody rozptylů	72
11.2	Více skupin – Analýza rozptylu jednoduchého třídění – způsob výpočtu	73
11.2.1	Kontrasty	73
11.2.2	Metody mnohonásobného srovnání	73
11.3	Neparametrické varianty analýzy rozptylu	74
11.4	Vztah mezi regresi a analýzou rozptylu	75
11.5	Analýza rozptylu dvojnásobného třídění	76
11.6	Opakované pozorování	78
11.7	Testování modelu a „podmodelu“	78
11.8	Obecnější modely analýzy rozptylu	78
12	Kvalitativní veličiny a jejich vztah	81
12.1	Odhad a testy pravděpodobnosti alternativní veličiny	81
12.1.1	Aproximace normálním rozložením	81
12.1.2	Fleissova kvadratická aproximace	82
12.1.3	Exaktní binomický test	82
12.2	Obecná kontingenční tabulka	82
12.3	Kontingenční tabulka 2×2	84
12.3.1	Míry vztahu dvou alternativních veličin	85
12.3.2	Hypotéza symetrie McNemar	86
12.3.3	Shoda dvou hodnotitelů	87

12.4	Typy studií – způsoby konstrukce kontingenčních tabulek	87
12.4.1	Průřezová studie	88
12.4.2	Kohortová studie	88
12.4.3	Studie případ-kontrola	88
12.5	Stratifikované kontingenční tabulky	88
12.6	Test trendu v kontingenční tabulce	89
12.7	Souvislost testů pro kategoriální a spojité veličiny	90
12.8	Intenzita incidence	91
12.9	Hodnocení kvality screeningových testů	92
12.10	ROC křivky	93
13	Výběr a jeho reprezentativnost	95
13.1	Rušivé faktory	96
13.2	Konstrukce výběru pro studie popisující populaci	96
13.3	Plány experimentu	97
13.3.1	Rozdělení na skupiny (do větví)	97
13.3.2	Volba kontrolní skupiny	98
13.3.3	Párové uspořádání dat	99
13.3.4	Křížový pokus	99
13.4	Stanovení rozsahu výběru	99
13.4.1	Rozsah výběru pro jednovýběrový t-test	99
13.5	Standardizace	100
13.5.1	Přímá standardizace	102
13.5.2	Nepřímá standardizace	102
13.5.3	Inverzní standardizace	102
13.5.4	Intervaly spolehlivosti pro standardizované ukazatele	102
14	Další modely pro studium závislosti veličin	103
14.1	Logistická regrese – model závislosti alternativní veličiny	103
14.1.1	Účinná dávka ED50 či LD50	105
14.2	Poissonovská regrese – model závislosti počtů na spojité či kvalitativní veličině	105
15	Analýza cenzorovaných dat	107
15.0.1	Neúplná informace – cenzorovaná data	107
15.0.2	Analýza přežití	108
15.0.3	Odhad doby do události (doby přežití)	110
15.0.4	Složitější parametrické modely pro analýzu přežití	115
15.1	Cenzorovaná data – hodnoty pod detekčním limitem	116
	Literatura	117
	Rejstřík	121