Analýza dat v neurologii

XX. Statistické testy pro četnosti kategorií – test dobré shody

L. Dušek, T. Pavlík, J. Koptíková
Institut biostatistiky a analýz
Masarykova univerzita, Brno

doc. RNDr. Ladislav Dušek, Dr.
Institut biostatistiky a analýz
Masarykova univerzita, Brno
e-mail: dusek@cba.muni.cz

Je velmi potěšující, že nám jubilejní 20. díl seriálu vyšel právě na test dobré shody. Tento často používaný test patří bez nadsázky mezi několik málo statistických testů, které se staly přímo součástí „know-how" biologů i lékařů. Každý si jistě vzpomene na cvičení z obecné genetiky, kde jsme pomocí tohoto testu ověřovali, zda výskyt homozygotních a heterozygotních jedinců v dceřiných populacích odpovídá Mendelovým zákonům. V tomto díle se pokoušíme přinést praktické ukázky i jiných, taktéž velmi rozšířených, aplikací tohoto testu.

Ačkoli si tedy mnozí spojujeme test dobré shody s ověřováním Mendelových zákonů, sám Mendel jej v době svých pokusů v ruce neměl. Test dobré shody (χ^2 test, chí-kvadrát test) byl publikován až v roce 1900 významným anglickým statistikem Karlem Pearsonem (1857–1936). Proto v anglickém jazyce test najdeme pod heslem Pearson's chi-square (χ^2) test. S osobností Karla Pearsona se v našem seriálu rozhodně nesetkáváme naposledy, minimálně se o něm zmíníme při výkladu o korelačních koeficientech.

Test dobré shody je určen pro srovnání pozorovaných četností jevu (jevů) s četnostmi očekávanými. Máme tedy celkem *n* jedinců (subjektů), u kterých sledujeme výskyt jednoho jevu (u každého jedince získáváme binární kód: jev nastal/nenastal, 1/0) nebo výskyt více jevů (např. krevní skupina A/B/AB/0). Součet pozorovaných četností jednotlivých jevů v experimentu musí samozřejmě dát dohromady celkovou hodnotu *n*. Zároveň předpokládáme, že pozorování jednotlivých jedinců jsou vzájemně zcela nezávislá. Očekávané četnosti dopočítáme podle nějakého předem daného klíče (např. předpokládaný poměr sledovaných kategorií dle Mendelových zákonů). Následně ověřujeme nulovou hypotézu, že se očekávané a pozorované četnosti neliší. Formálně můžeme výpočet statistiky testu dobré shody zapsat takto:

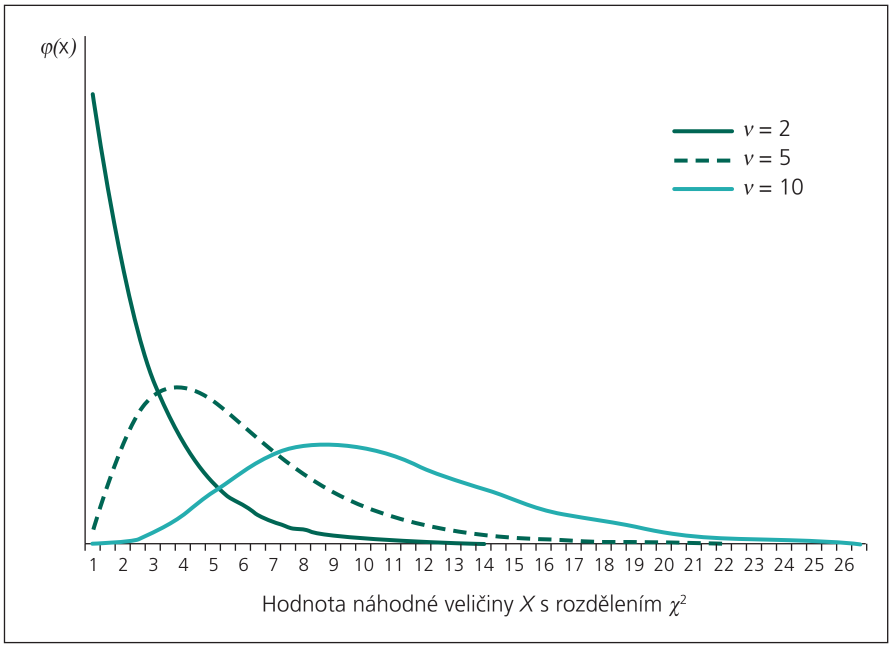
$$\chi^2_v = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \ldots\ldots + \frac{(O_k - E_k)^2}{E_k}$$

Kde:
k je počet jevů (kategorií), které sledujeme
$v = k - 1$ jsou stupně volnosti pro určení kritické hodnoty testu
E_i je očekávaný počet výskytů jevu i
O_i je pozorovaný počet výskytů jevu i

Je zřejmé, že člen $(O_i - E_i)^2 / E_i$ je počítán pro každý sledovaný jev i. Výše uvedený vztah má tedy tolik sčítanců, kolik je sledovaných jevů, a pro každý z nich přímo srovnáváme pozorované a očekávané četnosti. Ve výpočtu můžeme mít minimálně dva sčítance (kdy sledujeme pouze, zda nějaká vlastnost nastala, nebo nenastala); maximální počet kategorií není omezen a je označován *k*.

Pevně věříme, že čtenáři ocení výpočet testu pro jeho jednoduchost. Skutečně málokterý test nabízí tak transparentní vysvětlení principu výpočtu jako právě χ^2 test. Pozorované a očekávané četnosti jsou pro každý jednotlivý jev jasně odečítány a rozdíl je započítán jako druhá mocnina $(O_i - E_i)^2$. Čím větší je rozdíl mezi pozorovanými a očekávanými četnostmi, tím větší je testová statistika, a tedy i pravděpodobnost, že se O a E mezi sebou liší.

Rozdíl pozorovaných a očekávaných četností měříme testovou statistikou, která má chí-kvadrát (χ^2) rozdělení. Základním parametrem tohoto rozdělení je již výše uvedené v, tedy tzv. počet stupňů volnosti ($v = k - 1$, tedy o jednu méně než počet jevů). Testová statistika χ^2 může tudíž mít nejnižší možný počet stupňů volnosti 1 (pokud sledujeme pouze nastání/nenastání jednoho jevu). Vliv stupňů volnosti na tvar teoretického chí-kvadrát



Obr. 1. Pearsonovo χ^2 rozdělení při různých stupních volnosti (v).

1. 10 000 lidí házelo mincí s výsledkem 4 000 dopadů na lícovou a 6 000 na rubovou stranu mince.
2. Nulová hypotéza: mince dopadá na rubovou i lícovou stranu se stejnou četností (shoda pozorovaných četností s očekávaným poměrem obou jevů 1 : 1).
3. Předpoklady: u ideální mince je pravděpodobnost dopadu na každou její stranu stejná, jednotlivé hody mincí jsou vzájemně zcela nezávislé.
4. Očekávané četnosti: je vypočten očekávaný počet dopadů na lícovou (5 000) a rubovou stranu (5 000).
5. Výpočet testu: pro ověření platnosti nulové hypotézy je využit test dobré shody $$\chi^2_v = \frac{(E_{lic} - O_{lic})^2}{E_{lic}} + \frac{(E_{rub} - O_{rub})^2}{E_{rub}} = \frac{(5000-6000)^2}{5000} + \frac{(5000-4000)^2}{5000} = 400$$
6. Počet stupňů volnosti je odvozen od počtu sledovaných jevů (rub a líc): $v = 2 - 1 = 1$.
7. Kritická hodnota testové statistiky χ^2 pro $\alpha = 0{,}05$ je $\chi^2_{0,95}(v - 1) = 3{,}84$.
8. Reálná hodnota $\chi^2 = 400$ je větší než nalezená kritická hodnota $\chi^2_{0,95}(1) = 3{,}84$, což vede k zamítnutí nulové hypotézy.
9. **Závěr:** na hladině významnosti $\alpha = 0{,}05$ byl prokázán statisticky významný rozdíl v počtu dopadů mince na rubovou a lícovou stranu oproti předpokladu jejich stejné četnosti.

Příklad 1. Aplikace testu dobré shody při hodu mincí.

1. Celkem bylo zkoumáno 250 semen určitého druhu rostliny a roztříděno do následujících kategorií: žluté/hladké; žluté/vrásčité; zelené/hladké; zelené/vrásčité. Předpokládaný poměr výskytu těchto kategorií v populaci je 9 : 3 : 3 : 1.

	žluté/hladké	žluté/vrásčité	zelené/hladké	zelené/vrásčité	n
$f_{poz.}$	152	39	53	6	250
$f_{oček.}$	140,625	46,875	46,875	15,625	

2. Nulová hypotéza: pozorovaný výskyt jednotlivých kategorií se shoduje s očekávaným poměrem *(očekávaný poměr je v tomto a obdobných případech určen podle externího pravidla, zde odvozeného z předpokladů o vlivu kombinací alel genů na vlastnosti semen rostliny)*
3. Pro výpočet je využito testu dobré shody

$$\chi^2 = \frac{11.3750^2}{140.6250} + \frac{7.8750^2}{46.8750} + \frac{6.1250^2}{46.8750} + \frac{9.6250^2}{15.6250} = 8.972$$

$v = 4 - 1 = 3$; kritická hodnota $\chi^2_{0,95}(3) = 7{,}15 \rightarrow$ zamítáme nulovou hypotézu

4. V této fázi si ale jistě položíme otázku, která z kategorií vlastností semen nejvíce přispívá k výslednému χ^2 (výsledná hodnota χ^2 je součtem příspěvku jednotlivých kategorií): srovnání jednotlivých členů rovnice ukazuje, že největší odchylku od očekávaného počtu vykazuje varianta zelené/vrásčité. Na základě tohoto zjištění otestujeme dílčí hypotézu, zda se ostatní varianty vyskytují dle předpokládaného poměru 9 : 3 : 3.

	žluté/ hladké	žluté/ vrásčité	zelené/ hladké	n
$f_{poz.}$	152	39	53	244
$f_{oček.}$	146,4	48,8	48,8	

5. Pro analýzu této dílčí hypotézy je opět využito testu dobré shody, nyní s $v = 2$ stupni volnosti

$$\chi^2 = \frac{5.600^2}{146.40} + \frac{9.800^2}{48.80} + \frac{4.200^2}{48.80} = 2.544$$

$v = 3 - 1 = 2$; kritická hodnota $\chi^2_{0,95}(2) = 5{,}99 \rightarrow$ nulovou hypotézu nezamítáme

6. Na základě výsledku dílčího testu (bod 4 a 5) slučujeme první tři kategorie semen a testujeme hypotézu, zda se sloučené kategorie vs kategorie zelené/vrásčité vyskytují dle předpokládaného poměru, který nyní činí 15 : 1. Pro analýzu této dílčí hypotézy je opět využito testu dobré shody.

	ostatní	zelené/vrásčité	n
$f_{poz.}$	244	6	250
$f_{oček.}$	234,375	15,625	

$$\chi^2 = \frac{9.625^2}{15.625} + \frac{9.625^2}{234.375} = 6.324$$

$v = 2 - 1 = 1$; kritická hodnota $\chi^2_{0,95}(1) = 3{,}84 \rightarrow$ nulovou hypotézu zamítáme

7. **Závěr:** Na hladině významnosti $\alpha = 0{,}05$ byl prokázán statisticky významný rozdíl mezi pozorovanými četnostmi typů semen oproti očekávanému předpokladu. Zdrojem rozdílu mezi pozorovanými a očekávanými četnostmi je varianta semen zelené/vrásčité, která se oproti předpokladu vyskytla méně často. Dílčí hodnoty testové statistiky χ^2 v bodech 6 a 5 se v součtu rovnají celkové hodnotě χ^2 vypočítané v bodě 3.

Příklad 2. Aplikace testu dobré shody pro situaci s více než dvěma jevy.

Vzorek (i)	Počet jedinců ve vzorku (n_i)	Počet leváků ve vzorku (r_i)	Relativní četnost leváků ve vzorku (p_i)
1	14	11	0,79
2	16	12	0,75
3	20	5	0,25
4	28	14	0,50
5	17	4	0,24
6	22	5	0,23

1. Z populace bylo provedeno 6 nezávislých výběrů (S = 6) a v každém z nich určen podíl leváků a praváků.

2. Nulová hypotéza: zastoupení (relativní četnost) leváků je ve všech šesti výběrech (pozorováních) shodné.

3. Pro výpočet je využita obdoba testu dobré shody (tzv. test homogenity binomických rozdělení)

$$\chi^2_{\nu = S-1} = \frac{\left(\sum n_i \; p_i - \bar{p} \sum n_i\right)}{\bar{p}\left(1-\bar{p}\right)} \qquad \text{kde} \quad \bar{p} = \frac{\sum p_i}{S}$$

4. Po dosazení do vztahu získáme

$\chi^2 = 30{,}2 \qquad \nu = 6 - 1 = 5$

5. Kritická hodnota testové statistiky χ^2 pro a = 0,05 je $\chi^2_{0,95}(5) = 11{,}07$.

6. Zjištěná hodnota $\chi^2 = 30{,}2$ je větší než nalezená kritická hodnota $\chi^2_{0,95}(5) = 11{,}07$; což vede k zamítnutí nulové hypotézy.

7. **Závěr:** Na hladině významnosti $\alpha = 0{,}05$ byla zamítnuta shoda zastoupení leváků v jednotlivých výběrech. Výběry není možné na základě tohoto závěru sloučit

Příklad 3. Test homogenity binomických rozdělení.

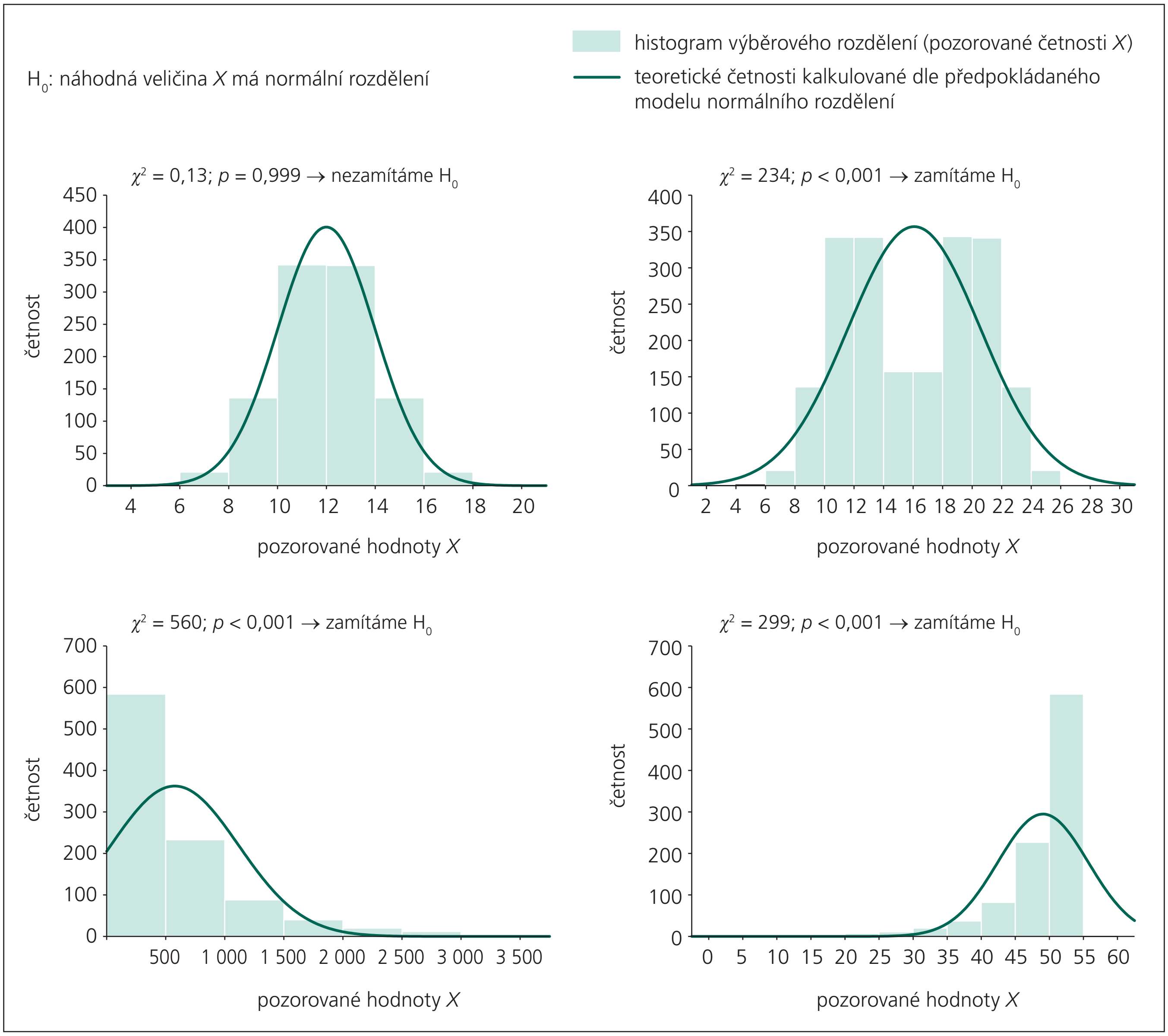
rozdělení ukazuje obr. 1. Všimněme si, že již sama hodnota χ^2 je druhou mocninou χ, a tedy nabývá pouze kladných hodnot.

Je zřejmé, že nejnižší možná hodnota testové statistiky χ^2 je 0, což je případ absolutní shody očekávaných a pozorovaných četností u všech sledovaných jevů. S rostoucími rozdíly mezi O_i a E_i roste také hodnota χ^2 a ve chvíli, kdy přesáhne kritickou mez $\chi^2_{1-\alpha}$ (pro ν stupňů volnosti), zamítáme hypotézu shody četností na hladině významnosti α.

Test dobré shody je využitelný pro náhodné veličiny nominální, ordinální i diskrétní; tedy pro všechny typy dat, kde umíme spočítat četnost kategorií. Avšak i spojitá data (třeba výšku postavy) můžeme převést na kategorie, pokud by bylo třeba sledovat výskyt těchto kategorií proti nějakým předpokládaným hodnotám. Tento postup je často využíván v klinické praxi, kdy nepracujeme s primárními hodnotami laboratorních markerů, ale s jejich kategoriemi fyziologickými, rizikovými a patologickými. Jinou modifikací je aplikace testu na výskyt jednoho jevu v časových intervalech. Jednoduše sledujeme, kolikrát nastal určitý jev ve vybraných časových intervalech u kohorty *n* jedinců (např. počty unikátních pacientů s nějakou komplikací po operaci nebo počty unikátních pacientů s opakovanou hospitalizací po propuštění), a srovnáváme tyto počty s očekávanými četnostmi (dány buď teoreticky, nebo literaturou apod.). Součet pozorovaných a očekávaných četností jednotlivých jevů musí být samozřejmě i v tomto případě shodný a musí se rovnat počtu sledovaných pacientů *n*. Jednotlivá pozorování pacientů a jejich chování musí být vzájemně nezávislá. Očekávané a pozorované četnosti srovnáváme pro jednotlivé časové intervaly a vztah pro výpočet χ^2 tedy bude mít *k* sčítanců, kde *k* je rovno počtu časových intervalů.

V našem výkladu samozřejmě nemůže chybět ukázka nejjednodušší aplikace testu, kterou zde nabízí klasický hod mincí v příkladu 1. Příklad 2 dále dokumentuje velmi sympatickou vlastnost testové statistiky χ^2, a sice její aditivitu. Každý jednotlivý zlomek z výše uvedeného vztahu pro test dobré shody je vlastně dílčí komponentou výsledné hodnoty testové statistiky χ^2. A jelikož tyto komponenty sčítáme, lze jednoduše zjistit, jakým podílem se jednotlivé zlomky (kategorie, jevy) promítají do celkové hodnoty testové statistiky. Tuto vlastnost oceníme především při srovnání očekávaných a pozorovaných četností u více kategorií. Ne všechny kategorie se v reálném výskytu liší stejně od očekávaných četností. Dílčími testy a rozborem hodnoty χ^2 statistiky (tedy jednotlivých zlomků, které ji v součtu tvoří) můžeme dokonce zjistit, že i při celkovém zamítnutí shody četností se některé kategorie s očekávanými četnostmi prokazatelně shodují a jiné naopak prokazatelně ne. Takové zjištění může mít velký význam. Příklad 2 takto dokumentuje sledování výskytu čtyř typů semen, kdy celkový test dobré shody zamítl shodu pozorovaných a očekávaných četností. Dílčí testy ale odhalily, že za statistickou významnost celkového testu je odpovědná pouze jedna ze čtyř kategorií semen, a pokud je ze sledování vyloučena, pak zbývající tři typy semen se vyskytují přesně ve shodě s očekávanými poměry.

Jinou modifikací testu dobré shody je tzv. test homogenity (shody) binomických



Obr. 2. Hodnocení normality výběrového rozdělení náhodné veličiny X pomocí testu dobré shody – příklady různých výsledků.

rozdělení. Za poněkud složitým názvem se skrývá velmi užitečný postup, který umožňuje ověřit, zda se více vzájemně nezávislých průzkumů (experimentů) liší v relativní četnosti určitého znaku (jevu). Příkladem může být výskyt jevu *A* v *S* definovaných kohortách pacientů, které získáme rešerší literatury. Původní autoři ale pracovali s různě velkými soubory, a výstupy tedy nelze bez exaktního ověření jednoduše spojit. Test uvedený v příkladu 3 slouží k ověření následujících hypotéz:

- je výskyt jevu *A*, resp. jeho relativní četnost, shodná (homogenní) ve všech *S* dostupných pozorováních, a můžeme tedy pozorování sloučit a odhadnout tak celkovou relativní četnost výskytu jevu?
- pokud relativní četnost jevu *A* není v dostupných pozorováních shodná, jak se od sebe jednotlivá pozorování liší? Vytvářejí vnitřně homogenní shluky? Jsou mezi nimi viditelně odlehlá pozorování, jejichž vyloučením bychom získali homogenní skupinu?

Test dobré shody je velmi často využíván pro ověření předpokladu, že rozdělení pravděpodobnosti náhodné veličiny *X* je určitého typu, například že odpovídá modelu normálního rozdělení, Poissonova rozdělení apod. Zde se test dobré shody stává metodou matematické statistiky, která ověřuje, zda má náhodná veličina určité, předem dané rozdělení. Vlastní výpočet se nijak zásadně neliší od již uvedeného postupu. Pozorované hodnoty náhodné veličiny *X* rozdělíme do kategorií, ve kterých zjistíme pozorované četnosti. Očekávané četnosti spočítáme z modelu teoretického rozdělení, které na datech ověřujeme. Jestliže je toto modelové rozdělení dáno předem včetně všech svých parametrů, je počet stupňů volnosti pro χ^2 test i nadále $k - 1$. Je-li ale některý parametr modelového rozdělení neznámý, snižuje se počet stupňů volnosti o jednotku za každý neznámý parametr (neznámé parametry musí být nejprve z naměřených dat odhadnuty a následně využity k výpočtům očekávaných četností podle daného modelu). Test dobré shody je tedy jedním z testů normality rozdělení nebo obecně jedním z tzv. testů goodness-of-fit. Do této rodiny patří i ostatní testy o rozdělení sledované veličiny, které jsme probrali v díle XVIII tohoto seriálu (např. Shapiro-Wilkův test, Kolmogorov-Smirnovův test). Vybrané příklady aplikace testu dobré shody jako testu normality přibližuje obr. 2.

Literatura

Plackett RL. Karl Pearson and the Chi-Squared Test. International Statistical Review 1983; 51(1): 59–72.