11. Porovnání empirického a teoretického rozložení

11.1. Motivace

11.2. Popis Kolmogorovova – Smirnovova testu a jeho Lilieforsovy varianty

11.3. Příklad: Jsou dány hodnoty 10, 12, 8, 9, 16. Pomocí Lilieforsovy varianty K- S testu ověřte na hladině významnosti 0,05, zda tato data pocházejí z normálního rozložení.
Řešení: Odhadem střední hodnoty je výběrový průměr m = 11, odhadem rozptylu je výběrový rozptyl $s^2 = 10$. Uspořádaný náhodný výběr je (8, 9, 10, 12, 16). Vypočteme hodnoty výběrové distribuční funkce:

$$x < 8 : F_5(x) = 0, 8 \le x < 9 : F_5(x) = \frac{1}{5} = 0{,}2\,, 9 \le x < 10 : F_5(x) = \frac{2}{5} = 0{,}4,$$

$$10 \le x < 12 : F_5(x) = \frac{3}{5} = 0{,}6, 12 \le x < 16 : F_5(x) = \frac{4}{5} = 0{,}8, x \ge 16 : F_5(x) = 1$$

Hodnoty teoretické distribuční funkce $\Phi_T(x)$ v bodech 8, 9, 10, 12, 16:

$$\Phi_T(8) = \Phi\left(\frac{8-11}{\sqrt{10}}\right) = \Phi(-0{,}95) = 1 - \Phi(0{,}95) = 1 - 0{,}82894 = 0{,}17106$$

$$\Phi_T(9) = \Phi\left(\frac{9-11}{\sqrt{10}}\right) = \Phi(-0{,}63) = 1 - \Phi(0{,}63) = 1 - 0{,}73565 = 0{,}26435$$

$$\Phi_T(10) = \Phi\left(\frac{10-11}{\sqrt{10}}\right) = \Phi(-0{,}32) = 1 - \Phi(0{,}32) = 1 - 0{,}62552 = 0{,}37448$$

$$\Phi_T(12) = \Phi\left(\frac{12-11}{\sqrt{10}}\right) = \Phi(0{,}32) = 0{,}62552$$

$$\Phi_T(16) = \Phi\left(\frac{16-11}{\sqrt{10}}\right) = \Phi(1{,}58) = 0{,}94295$$

(Φ je distribuční funkce rozložení N(0,1).)

Rozdíly mezi výběrovou distribuční funkcí $F_5(x)$ a teoretickou distribuční funkcí $\Phi_T(x)$:
$d_1 = 0{,}2 - 0{,}17106 = 0{,}02894$; $d_2 = 0{,}4 - 0{,}26435 = 0{,}13565$; $d_3 = 0{,}6 - 0{,}37448 = 0{,}22552$;
$d_4 = 0{,}8 - 0{,}62552 = 0{,}17448$; $d_5 = 1 - 0{,}94295 = 0{,}05705$.
Testová statistika: $D_5 = 0{,}22552$, modifikovaná kritická hodnota pro n = 5, $\alpha = 0{,}05$ je 0,343.
Protože 0,22552 < 0,343, hypotézu o normalitě nezamítáme na hladině významnosti 0,05.

11.4. Popis Shapirova – Wilkova testu

11.5. Poznámka: Další testy normality

11.6. Popis testu dobré shody v diskrétním a spojitém případě

11.7. Příklad: Byl zjišťován počet poruch určitého zařízení za 100 hodin provozu ve 150 disjunktních 100 h intervalech. Výsledky měření:

Počet poruch za 100 hodin provozu	0	1	2	3	4 a víc
Absolutní četnost	52	48	36	10	4

Na asymptotické hladině významnosti 0,05 testujte hypotézu, že náhodný výběr $X_1, ..., X_{150}$ pochází z rozložení Po(1,2).

Řešení:
Pravděpodobnost, že náhodná veličina s rozložením Po(λ), kde $\lambda = 1,2$ bude nabývat hodnot $p_0, ..., p_4$ a víc je $p_j = \frac{\lambda^j}{j!} e^{-\lambda} = \frac{1,2^j}{j!} e^{-1,2}$, $j = 0,1,2,3$, $p_4 = 1 - (p_0 + p_1 + p_2 + p_3)$.
Výpočty potřebné pro stanovení testové statistiky K uspořádáme do tabulky.

j	n_j	p_j	np_j	$\frac{(n_j - np_j)^2}{np_j}$
0	52	0,301	150.0,301=45,15	1,039
1	48	0,361	150.0,361=54,15	0,698
2	36	0,217	150.0,217=32,55	0,366
3	10	0,087	150.0,087=13,05	0,713
4	4	0,034	150.0,034=5,1	0,237

K = 1,039 + 0,698 + 0,713 + 0,237 = 3,053, r = 5, $\chi^2_{0,95}(4) = 9,488$. Protože 3,053 < 9,488, nulovou hypotézu nezamítáme na asymptotické hladině významnosti 0,05.

11.8. Příklad (test dobré shody pro spojité rozložení): Byl pořízen náhodný výběr rozsahu n = 100. Jeho číselné realizace byly roztříděny do 5 ekvidistantních třídicích intervalů o délce 0,04, přičemž dolní mez prvního třídicího intervalu je 3,92. Absolutní četnosti jednotlivých třídicích intervalů jsou: 11, 20, 44, 19, 6.
Výběrový průměr se realizoval hodnotou m = 4,02 a výběrová směrodatná odchylka hodnotou s = 0,04.
Na asymptotické hladině významnosti 0,05 testujte hypotézu, že náhodný výběr pochází z normálního rozložení.
Řešení:
Výpočty potřebné pro stanovení testové statistiky K uspořádáme do tabulky.
Přitom symbolem Φ značíme distribuční funkci rozložení $N(\mu,\sigma^2)$, kde $\mu = 4,02$ a $\sigma = 0,04$.

$\langle u_j, u_{j+1} \rangle$	n_j	$p_j=\Phi(u_{j+1})- \Phi(u_j)$	np_j	$(n_j - np_j)^2$	$\frac{(n_j - np_j)^2}{np_j}$
$\langle 3,92, 3,96 \rangle$	11	0,060598	6,0598	24,4060	4,0276
$\langle 3,96, 4,00 \rangle$	20	0,241730	24,1730	17,4142	0,7204
$\langle 4,00, 4,04 \rangle$	44	0,382925	38,2925	32,5756	0,8507
$\langle 4,04, 4,08 \rangle$	19	0,241730	24,1730	26,7608	1,1070
$\langle 4,08, 4,12 \rangle$	6	0,060598	6,0598	0,0036	0,0006

K = 4,0276 + 0,7204 + 0,8507 + 1,1070 + 0,0006 = 6,7063
Kritický obor: $W = \langle \chi^2_{1-\alpha}(r-1-p), \infty) = \langle \chi^2_{0,95}(5-1-2), \infty) = \langle 5,9915, \infty)$
Protože testová statistika se realizuje v kritickém oboru, hypotézu o normalitě zamítáme na asymptotické hladině významnosti 0,05.

11.9. Poznámka: Použití testu dobré shody na data, jejichž rozložení je určeno intuitivně nebo na základě zkušenosti

11.10. Příklad: Ve svých pokusech pozoroval J.G. Mendel 10 rostlin hrachu a na každé z nich počet žlutých a zelených semen. Výsledky pokusu:

č.rostliny	1	2	3	4	5	6	7	8	9	10
počet žlutých	25	32	14	70	24	20	32	44	50	44
počet zelených	11	7	5	27	13	6	13	9	14	18
celkem	36	39	19	97	37	26	45	53	64	62

Z genetických modelů vyplývá, že pravděpodobnost výskytu žlutého semene by měla být 0,75 a zeleného 0,25. Na asymptotické hladině významnosti 0,05 testujte hypotézu, že výsledky Mendelových pokusů se shodují s modelem.
Řešení:
Výpočty potřebné pro stanovení testové statistiky K uspořádáme do tabulky.

j	n_j	p_j	np_j	$\frac{(n_j - np_j)^2}{np_j}$
1	25	0,75	36.0,75=27	0,148148
2	32	0,75	39.0,75=29,25	0,258547
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
10	44	0,75	62.0,75=46,5	0,134409

$K = 0{,}148148 + 0{,}258547 + ... + 0{,}134409 = 1{,}797495$, $r = 10$, $\chi^2_{0,95}(9) = 16{,}9$.
Protože $1{,}797495 < 16{,}9$, nulovou hypotézu nezamítáme na asymptotické hladině významnosti 0,05.

11.11. Poznámka: Jednoduchý test Poissonova rozložení

11.12. Příklad: Studujeme rozložení počtu pacientů, kteří během 75 dnů přijdou na pohotovost. Osmihodinovou pracovní dobu rozdělíme do půlhodinových intervalů a v každém intervalu zjistíme počet příchozích pacientů:

Počet pacientů	Pozorovaná četnost
0	79
1	188
2	282
3	275
4	196
5	114
6	45
7	10
8	7
9	3
10	1

Na asymptotické hladině významnosti 0,05 testujte hypotézu, že daný náhodný výběr pochází z Poissonova rozložení. Použijte jednoduchý test Poissonova rozložení.

Řešení:

Nejprve musíme vypočítat realizaci výběrového průměru a výběrového rozptylu:

$$m = \frac{1}{1200}(0 \cdot 79 + 1 \cdot 188 + \ldots + 10 \cdot 1) = 2{,}80\overline{3}$$

$$s^2 = \frac{1}{1199}\left[79 \cdot (0 - 2{,}80\overline{3})^2 + 188 \cdot (1 - 2{,}80\overline{3})^2 + \ldots + 1 \cdot (10 - 2{,}80\overline{3})^2\right] = 2{,}708579$$

$$K = \frac{(n-1)S^2}{M} = \frac{1199 \cdot 2{,}708579}{2{,}80\overline{3}} = 1158{,}579,$$

Kritický obor: $W = \left\langle 0, \chi^2_{\alpha/2}(n-1)\right\rangle \cup \left\langle \chi^2_{1-\alpha/2}(n-1), \infty\right) = \left\langle 0; 1104{,}93\right\rangle \cup \left\langle 1296{,}86; \infty\right),$

H_0 nezamítáme na asymptotické hladině významnosti 0,05.

11.13. Poznámka: Jednoduchý test exponenciálního rozložení

11.14. Příklad: Byla zkoumána doba životnosti 45 součástek (v hodinách). Zjistili jsme, že průměrná doba životnosti činila m = 99,93 h a rozptyl $s^2 = 7328{,}91\ h^2$. Na asymptotické hladině významnosti 0,05 testujte hypotézu, že daný náhodný výběr pochází z exponenciálního rozložení.

Řešení:

Testová statistika: $K = \frac{(n-1)S^2}{M^2} = \frac{44 \cdot 7328{,}91}{99{,}93^2} = 32{,}2924$

Kritický obor:

$$W = \left\langle 0, \chi^2_{\alpha/2}(n-1)\right\rangle \cup \left\langle \chi^2_{1-\alpha/2}(n-1), \infty\right) = \left\langle 0, \chi^2_{0{,}025}(44)\right\rangle \cup \left\langle \chi^2_{0{,}975}(44), \infty\right) =$$

$$= \left\langle 0{,}27{,}575\right\rangle \cup \left\langle 64{,}202, \infty\right)$$

Protože se testová statistika nerealizuje v kritickém oboru, hypotézu o exponenciálním rozložení nezamítáme na asymptotické hladině významnosti 0,05.