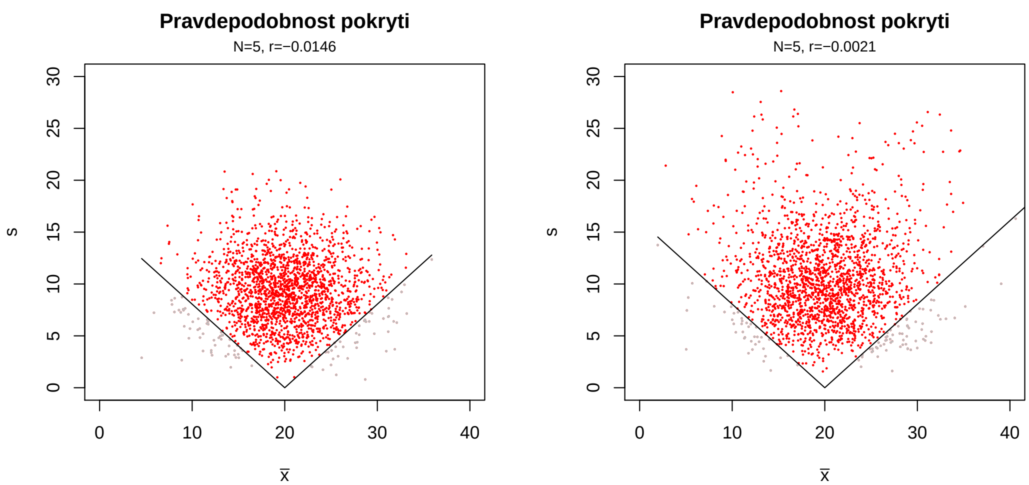
Zadání příkladů – Statistická inference II – 2017

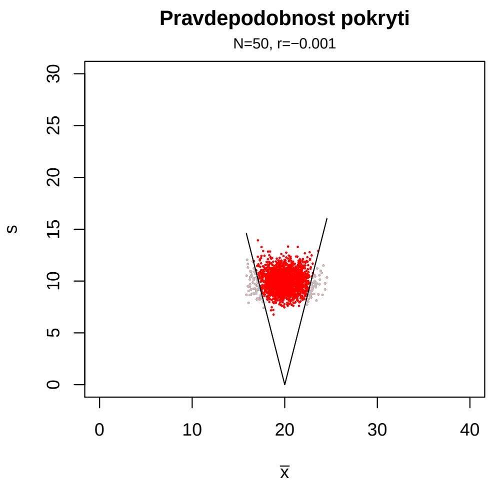
Příklad 12. nezávislost μ a σ^2; pravděpodobnost pokrytí Nechť $X \sim N(\mu, \sigma^2)$, kde $\mu = 20$ a $\sigma^2 = 100$. Pomocí simulační studie vypočítejte Pearsonův korelační koeficient $r_{\bar{X},S}$. Nakreslete šedou barvou rozptylový graf $(\bar{x}_m, s_m)$, kde $m = 1, 2, \ldots, M$, přičemž $M = 5\,000$. Černou barvou vyznačte v grafu takové body $(\bar{x}_m, s_m)$, pro které platí $t_{W,m} = \left|\frac{\bar{x}_m - \mu}{s_m}\sqrt{n}\right| < t_{n-1}(\alpha/2)$. Dále vykreslete hranice, které jsou definovány body $(\bar{x}_m, s_m)$, jež splňují vztah $t_{W,m} = t_{n-1}(\alpha/2)$. Vypočítejte pravděpodobnost pokrytí 95 % DIS pro μ jako podíl $\sum_m I(t_{W,m} < t_{n-1}(\alpha/2))/M$. Zvolte (a) $n = 5$, (b) $n = 50$ a (c) $n = 100$.

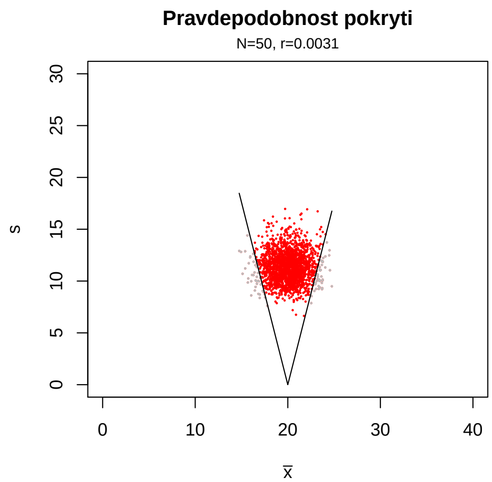
Simulaci proveďte také za předpokladu, že data pochází ze smíšeného rozdělení $X \sim [pN(\mu, \sigma_1^2) + (1-p)N(\mu, \sigma_2^2)]$, kde $p = 0.9$, $\mu = 20$, $\sigma_1^2 = 100$ a $\sigma_2^2 = 400$.

```
##                                         n
## aktualni pst.pokryti               0.9515
## nominalni pst.pokryti (spolehlivost) 0.9500
##                                        n
## aktualni pst.pokryti               0.947
## nominalni pst.pokryti (spolehlivost) 0.950
```

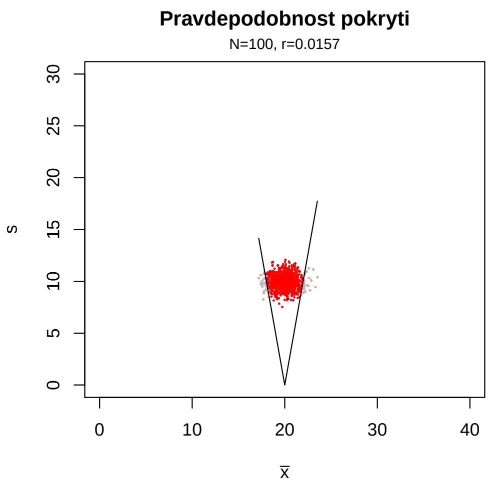


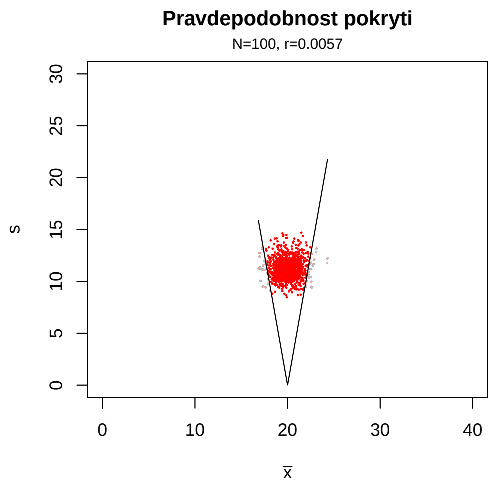

```
##                                        n
## aktualni pst.pokryti               0.946
## nominalni pst.pokryti (spolehlivost) 0.950
##                                        n
## aktualni pst.pokryti               0.958
## nominalni pst.pokryti (spolehlivost) 0.950
```






```
##                                          n
## aktualni pst.pokryti                  0.95
## nominalni pst.pokryti (spolehlivost) 0.95
##                                          n
## aktualni pst.pokryti                 0.958
## nominalni pst.pokryti (spolehlivost) 0.950
```

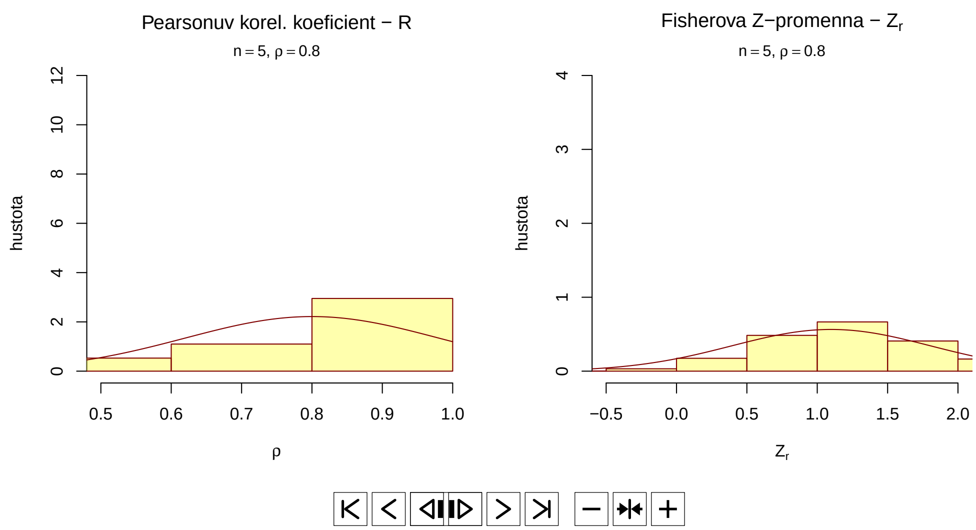




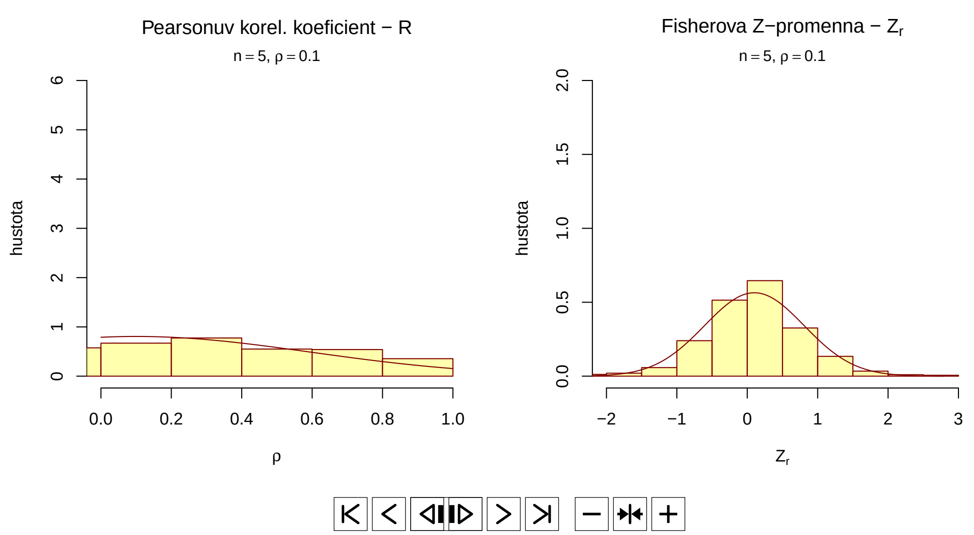
Příklad 13. konvergence ρ a ξ k normálnímu rozdělení

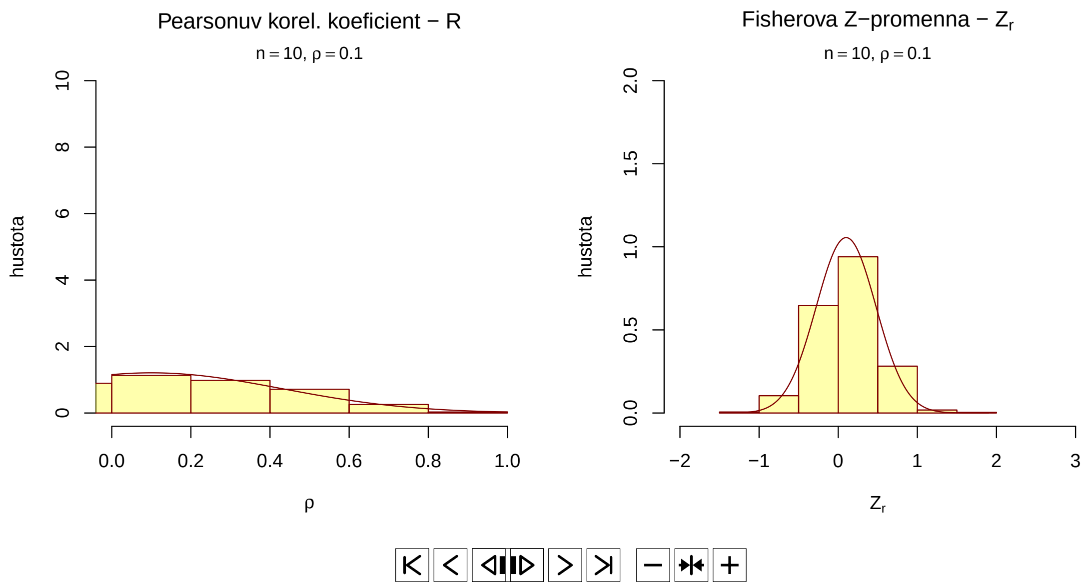
1. Proveďte simulaci pseudonáhodných čísel z $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, kde $\mu_1 = 0$, $\mu_2 = 0$, $\sigma_1 = 1$, $\sigma_2 = 1$, $M = 10\,000$, $rho = 0.8$.

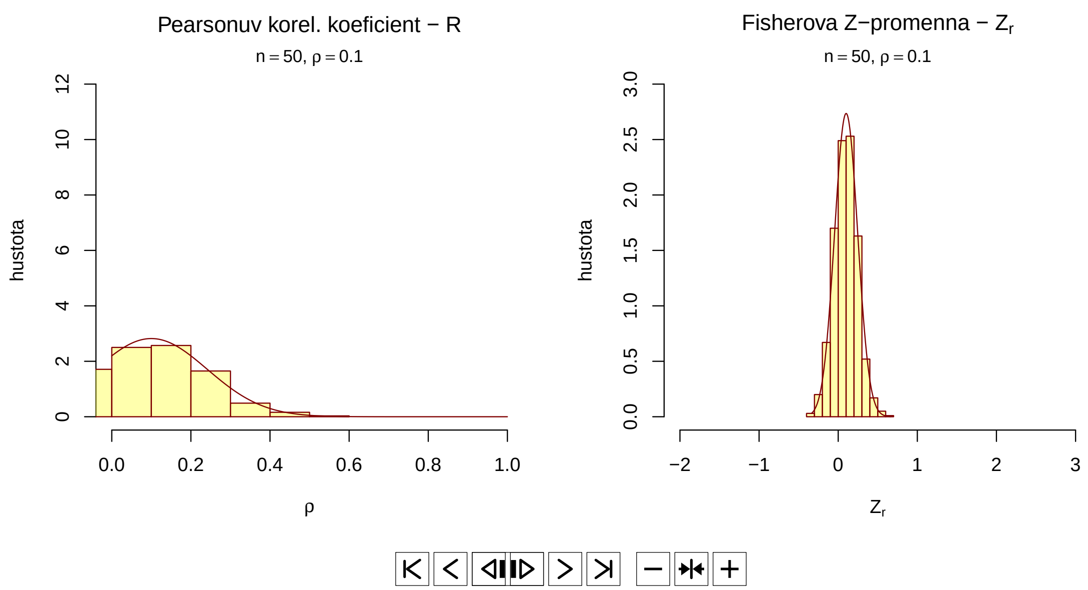
   Pro každé $m = 1, 2, \ldots, M$, vypočítejte realizaci výběrového korelačního koeficientu r_m a Fisherovy Z-proměnné $z_{R,m}$. Zobrazte histogramy simulovaných r_m a $z_{R,m}$ a superponujte je teoretickými hustotami příslušných normálních rozdělení. Vytvořte animaci zobrazující rozdělení výběrového korelačního koeficientu R a Fisherovy Z transformace pro různé rozsahy náhodného výběru $n \in \{5, 10, \ldots, 70\}$.



2. Vytvořte animaci zobrazující rozdělení výběrového korelačního koeficientu $R \in \{0.1, 0.2, \ldots, 0.9\}$ a Fisherovy Z transformace pro tyto různé korelační koeficienty. Animace vytvořte pro (a) $n = 5$, (b) $n = 10$, (c) $n = 100$.




Pearsonuv korel. koeficient − R
n = 10, ρ = 0.1
hustota
ρ
Fisherova Z−promenna − Zr
n = 10, ρ = 0.1
hustota
Zr


Pearsonuv korel. koeficient − R
n = 50, ρ = 0.1
hustota
ρ
Fisherova Z−promenna − Zr
n = 50, ρ = 0.1
hustota
Zr

Příklad 14. Test o korelačním koeficientu ρ: Mějme data one-sample-correlation-skull-mf.txt a proměnné největší výška mozkovny skull.pH a morfologická výška tváře face.H (obě v mm) starověké egyptské mužské populace, o kterých předpokládáme, že mají dvourozměrné normální rozdělení $N_2(\boldsymbol{\mu}, \boldsymbol{\Sigma})$.

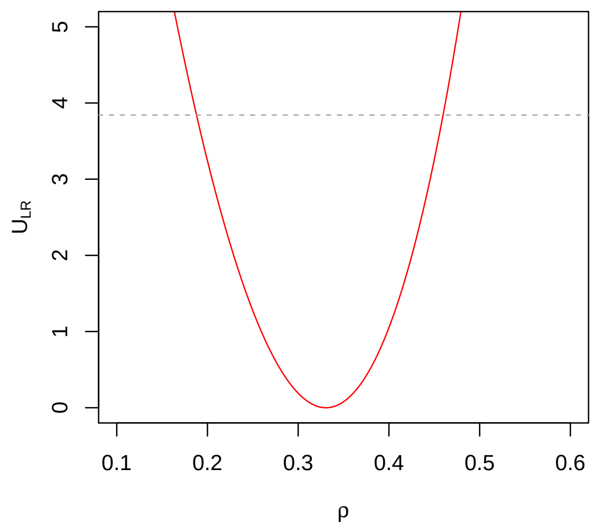
1. Otestujte hypotézu o tom, že korelační koeficient největší výšky mozkovny a morfologické výšky tváře je rovný 0.251.

2. Vypočítejte $100 \times (1-\alpha)\%$ empirický DIS pro korelační koeficient, kde koeficient spolehlivosti $1-\alpha = 0.95$.

Použijte

(a) Waldovu testovací statistiku Z_W;

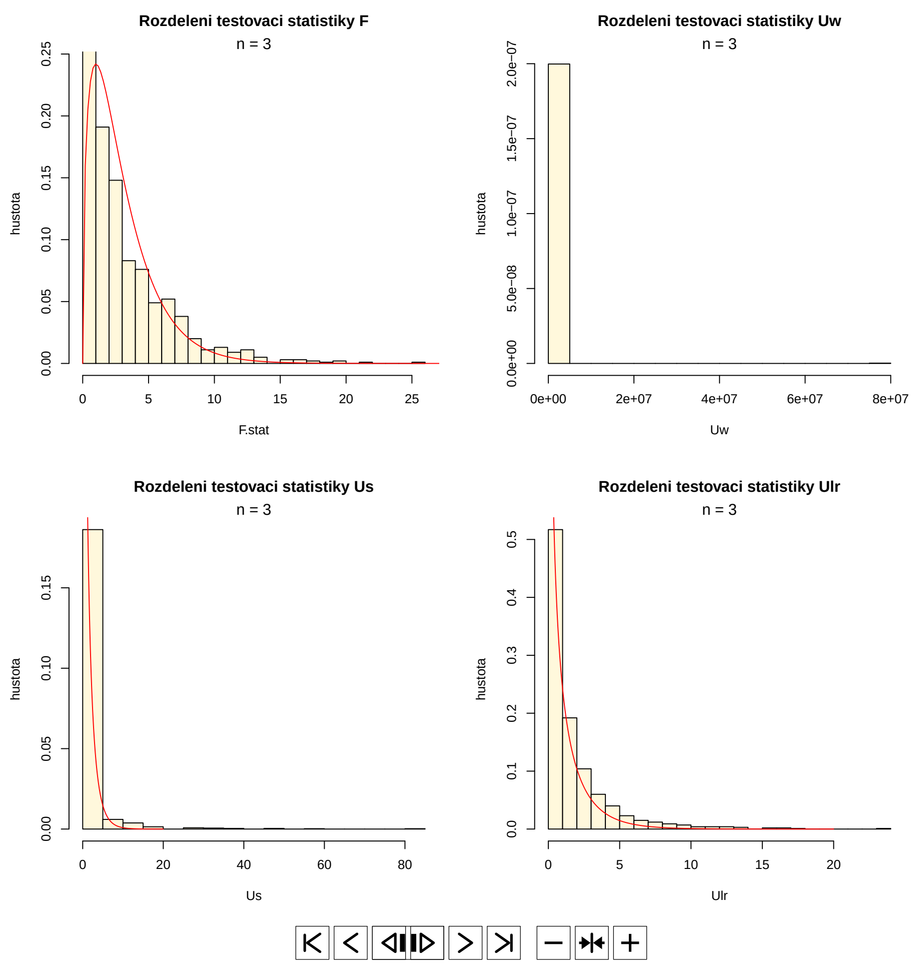
(b) testovací statistiku poměrem věrohodnosti U_{LR}.



	test.statistika	W	kritický obor	p-hodnota
Waldův přístup	1.104799	$(-\infty, -1.959964\rangle \cup \langle 1.959964; \infty)$	$(0.1868617, 0.4605561)$	0.2692467
Věrohodnostní přístup	1.241757	$\langle 3.841459; \infty)$	$(0.1880626, 0.4590998)$	0.2651328

Příklad 15. Rozdělení testovacích statistik F, U_W, U_S a U_{LR} jednovýběrového testu o rozptylu σ^2
Pomocí simulační studie ověřte, že

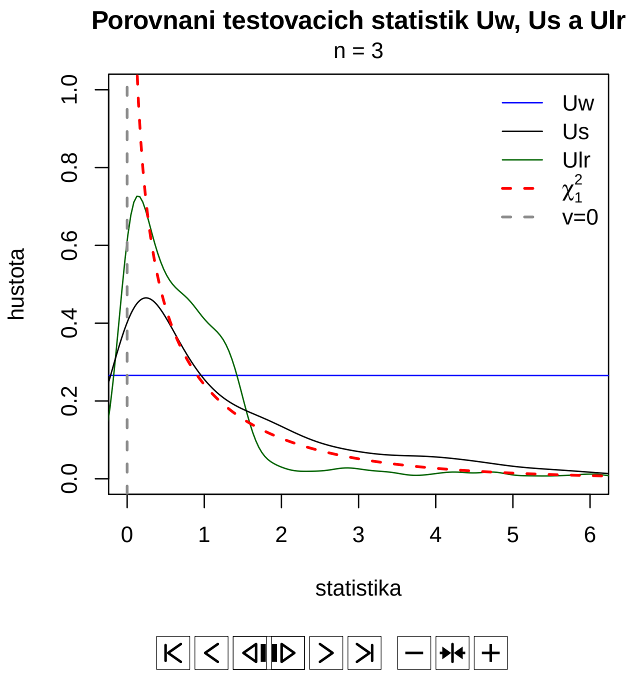
- statistika $F = \dfrac{nS_n^2}{\sigma_0^2} \sim \chi_n^2$
- statistika $U_w = \dfrac{n}{2}\left(\dfrac{n}{F_{obs}} - 1\right)^2 \sim \chi^2(1)$
- statistika $U_S = \dfrac{n}{2}\left(\dfrac{F_{obs}}{n} - 1\right)^2 \sim \chi^2(1)$
- statistika $U_{LR} = F_{obs} - n\left(1 + \ln\left(\dfrac{F_{obs}}{n}\right)\right) \sim \chi^2(1)$
- Vygenerujte $M = 1000$ pseudonáhodných výběrů o rozsahu $n = 5$. Pro každý náhodný výběr vypočítejte hodnoty realizací testovacích statistik F, U_W, U_S a U_{LR}. Vytvořte histogramy těchto testovacích statistik a superponujte je křivkami příslušného rozdělení.
- Vytvořte animaci, pomocí které bude zřejmé, jak se s rostoucím n rozdělení testovacích statistik F, U_W, U_S a U_{LR} asymptoticky blíží k příslušnému rozdělení. Animaci spravte pro měnící se $n \in \{3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000\}$



Příklad 16. Pokračování příkladu 15

- Vygenerujte $M = 1000$ pseudonáhodných výběrů o rozsahu $n = 5$. Pro každý náhodný výběr stanovte hodnoty realizací testovacích statistik F, U_W, U_S a U_{LR}. Pomocí funkce `density()` vypočítejte jádrové odhady rozdělení testovacích statistik U_W, U_S a U_{LR}. Tyto jádrové odhady zobrazte do jednoho grafu a barevně je odlište. Do grafu dokreslete křivku hustoty χ^2 rozdělení o jednom stupni volnosti a vertikální referenční čáru v hodnotě 0.

- Vytvořte animaci, pomocí které bude zřejmé, jak se s rostoucím n rozdělení testovacích statistik U_W, U_S a U_{LR} asymptoticky blíží k $\chi^2(1)$ rozdělení. Animaci spravte pro měnící se $n \in \{3, 4, 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 60, 70, 80, 90, 100, 200, 300, 400, 500, 600, 700, 800, 900, 1000\}$



Příklad 17. Pravděpodobnost pokrytí – 1. část

1. Nechť $X \sim Bin(N,p)$, kde $N = 30$ a $p = 0.8$ a pravděpodobnost úspěchu $\widehat{p} = \frac{24}{30} = 0.8$, kde $x = 24$ a $N = 30$. Waldův 95 % empirický DIS pro p je rovný $(d,h) = (0.657, 0.943)$. Vypočítejte pravděpodobnost pokrytí tohoto intervalu.

   Poznámka: Pravděpodobnost pokrytí Waldova 95 % DIS pro p vypočítáme následovně:

   $$\Pr(pokryti) = \sum_j \Pr(X = Np_j : p \in \text{ Waldovu 95 \% DIS pro } p_j),$$

   kde $p_j \in \mathcal{M}_J = \{\frac{1}{30}, \frac{2}{30}, \ldots, 1 - \frac{1}{30}\}$, t.j. jde o součet takových funkčních hodnot pravděpodobnostní funkce v bodech Np_j, kde $p \in$ Waldovu 95 % DIS pro p_j. Výsledky uspořádejte do tabulky, jejíž sloupce budou x_j, p_j, d_j (dolní hranice Waldova 95 % DIS pro p_j), h_j (horní hranice Waldova 95 % DIS pro p_j), Pr(pokrytí) a *pokrytí* (indikace toho, zda p náleží nebo nenáleží do Waldova 95 % DIS pro p_j).

2. Celý postup opakujte tentokrát pro parametr $p = 0.79$. Pozorujte, jak se změnila pravděpodobnost pokrytí.

1	xj	pj	dj	hj	P(pokryti)	pokryti
2	0	0	0	0	0	0
3	1	0.0333	-0.0309	0.0976	0	0
4	2	0.0667	-0.0226	0.1559	0	0
5	3	0.1	-0.0074	0.2074	0	0
6	4	0.1333	0.0117	0.255	0	0
7	5	0.1667	0.0333	0.3	0	0
8	6	0.2	0.0569	0.3431	0	0
9	7	0.2333	0.082	0.3847	0	0
10	8	0.2667	0.1084	0.4249	0	0
11	9	0.3	0.136	0.464	0	0
12	10	0.3333	0.1646	0.502	0	0
13	11	0.3667	0.1942	0.5391	0	0
14	12	0.4	0.2247	0.5753	0	0
15	13	0.4333	0.256	0.6107	0	0
16	14	0.4667	0.2881	0.6452	0	0
17	15	0.5	0.3211	0.6789	2e-04	0
18	16	0.5333	0.3548	0.7119	7e-04	0
19	17	0.5667	0.3893	0.744	0.0022	0
20	18	0.6	0.4247	0.7753	0.0064	0
21	19	0.6333	0.4609	0.8058	0.0161	1
22	20	0.6667	0.498	0.8354	0.0355	1
23	21	0.7	0.536	0.864	0.0676	1
24	22	0.7333	0.5751	0.8916	0.1106	1
25	23	0.7667	0.6153	0.918	0.1538	1
26	24	0.8	0.6569	0.9431	0.1795	1
27	25	0.8333	0.7	0.9667	0.1723	1
28	26	0.8667	0.745	0.9883	0.1325	1
29	27	0.9	0.7926	1.0074	0.0785	1
30	28	0.9333	0.8441	1.0226	0.0337	0
31	29	0.9667	0.9024	1.0309	0.0093	0
32	30	1	1	1	0.0012	0

Tabulka 1: Contingency table of absolute frequencies – Womac

	p=0.8	p=0.79
pst.pokryti	0.9463	0.8876

Tabulka 2: Contingency table of absolute frequencies – Womac

Příklad 18. Pravděpodobnost pokrytí – 2. část Nechť $X_i \sim Bin(N, p_i)$. Vypočítejte pravděpodobnosti pokrytí

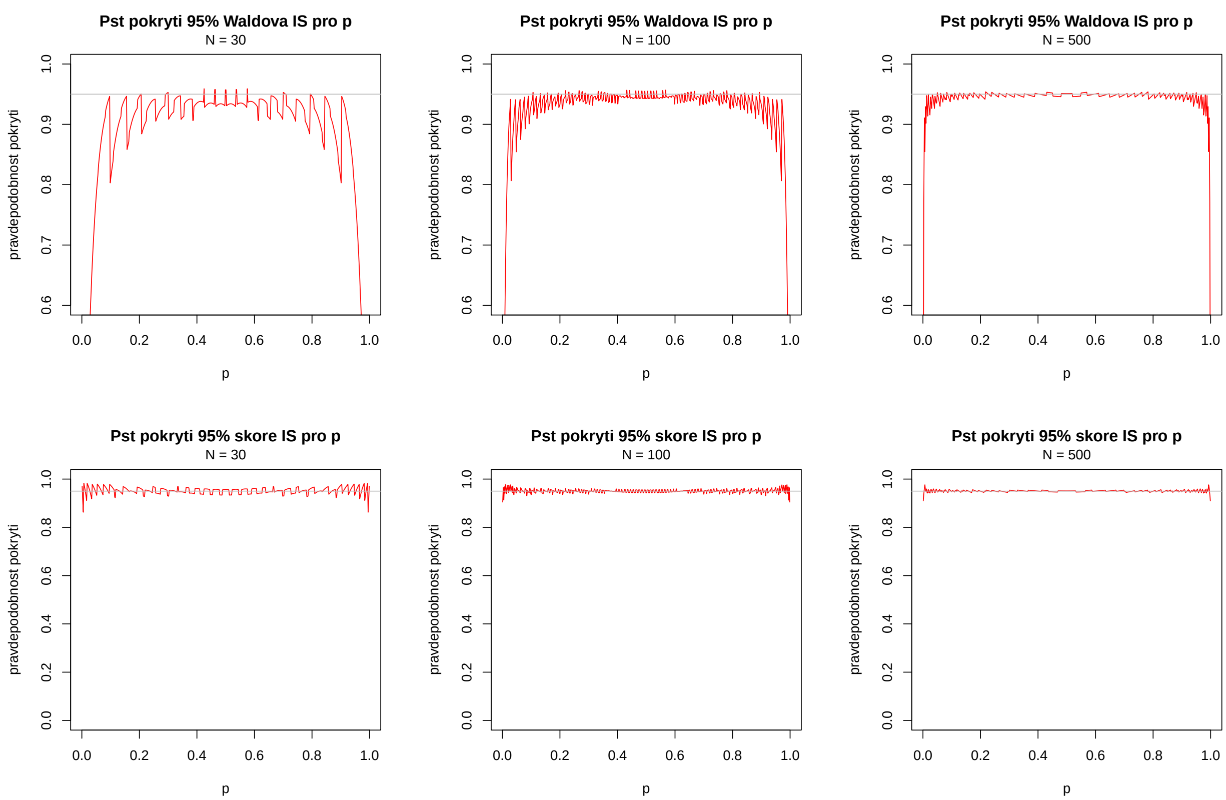
1. Waldova 95 % DIS,
2. skóre 95 % DIS,
3. věrohodnostního 95 % DIS

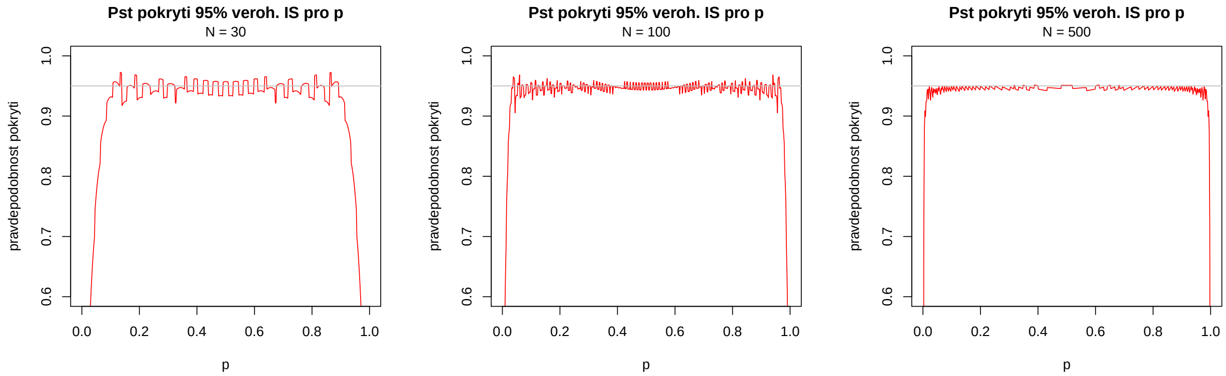
pro každé p_i, kde p_i náležící množině $\mathcal{M}_I = \langle \frac{1}{N}, 1 - \frac{1}{N} \rangle$ jsou ekvidistantně vzdálené hodnoty ležící mezi $\frac{1}{N}$ a $1 - \frac{1}{N}$ a jejich počet je $M = 1\,000$. Nakreslete obrázek, kde na ose x budou p_i a na ose y bude pravděpodobnost pokrytí Pr_i(pokrytí). Zvolte (a) $N = 30$, (b) $N = 100$ a (c) $N = 1000$.

Poznámka: Pravděpodobnosti pokrytí 95 % DIS pro p_i vypočítáme následovně

$$\text{Pr}_i(\text{pokrytí}) = \sum_j \text{Pr}(X = Np_j : p_i \in 95\,\%\ \text{DIS pro}\ p_j),$$

kde $p_j \in \mathcal{M}_J = \{\frac{1}{N}, \frac{2}{N}, \ldots, 1 - \frac{1}{N}\}$, t.j. jde o součet takových funkčních hodnot pravděpodobnostní funkce v bodech Np_j, kde $p_i \in 95\,\%$ DIS pro p_j. Pro ty DIS, které mají pro $p = 0$ a $p = 1$ nenulovou šířku, můžeme použít $\mathcal{M}_I = \langle \frac{0}{N}, \frac{N}{N} \rangle$




Pst pokryti 95% veroh. IS pro p
N = 30
pravdepodobnost pokryti
p
Pst pokryti 95% veroh. IS pro p
N = 100
pravdepodobnost pokryti
p
Pst pokryti 95% veroh. IS pro p
N = 500
pravdepodobnost pokryti
p

Příklad 19. Monotónní poměr věrohodnosti Testujeme $H_0 : \theta \in \Theta_0$ oproti $H_1 : \theta \in \Theta_1$ testem $\mathcal{T}$ na (nominální) hladině významnosti α. Nechť pro $\theta_1 \in \Theta$ a $\theta_2 \in \Theta$ platí $\theta_1 < \theta_2$. Pokud poměr věrohodností $L(\theta_2|\mathbf{x})/L(\theta_1|\mathbf{x})$ je rostoucí funkcí $T(\mathbf{X})$, potom $\mathcal{T}$ je rovnoměrně nejsilnější test $\mathcal{T} = \mathcal{T}_{UMP}$.

Příklad 20. Monotónní poměr věrohodnosti; binomické rozdělení Nechť $X \sim Bin(N,p)$, kde $\theta = p$. Funkce věrohodnosti je definovaná jako

$$L(\theta|x) = p^x(1-p)^{N-x}, \tag{1}$$

kde $x = \sum_{i=1}^N x_i$.

1. Ukažte, že pro nějaké dvě hodnoty p_1, p_2, $0 < p_1 < p_2 < 1$ je monotonní poměr věrohodnosti $L(p_2|\mathbf{x})/L(p_1|\mathbf{x})$ rostoucí funkcí statistiky $T(\mathbf{X}) = X = \sum_{i=1}^N X_i$. Zvolte $N = 20$, $p_1 = 0.4$, $p_2 = 0.5$.

2. vytvořte animaci zobrazující, že monotonní poměr věrohodnosti $L(p_2,\mathbf{x})/L(p_1|\mathbf{x})$ je rostoucí funkcí statistiky $T(\mathbf{X}) = X = \sum_{i=1}^N X_i$ pro pevně zvolené $N = 20$, $p_2 = 0.5$, $p_1 \in \{0.49, 0.48, \ldots, 0.02, 0.01\}$.