8.1

Řešení tohoto příkladu je podobné řešením mnoha příkladů, které jsme tento semestr probírali. Doporučuji tedy vyzkoušet si řešení příkladu samostatně a následující text použít pouze pro kontrolu =).

1. Vytvoříme funkci SimTSBinom(), která pro zadané hodnoty parametrů p, N a počet simulací M vykreslí jeden graf s histogramem. Součástí histogramu bude popisek s Haldovou podmínkou. Barva histogramu se bude měnit v závislosti na hodnotě Haldovy podmínky $N \cdot p(1-p) > 9$.

   $Np(1-p) \leq 9$ => barva histogramu bude červená

   $Np(1-p) > 9$ => barva histogramu bude zelená

```
SimTSBinom <- function(N, p, M=... ){
  X <- replicate(... , rbinom(1,...)) ... vektor M=1000 náh. veličin z Bin(N,p) (M=1000x vygenerujeme 1 náh. veličinu z Bin(N,p))
  zW <- (X/N - p) / sqrt(p(1-p)/N) ... vektor M=1000 test. statistik zW pro M=1000 náh. veličin X z Bin(N,p)
  xfit <- ... posloupnost od min(zW)-2 do max(zW)+2 o délce minimálně 500
  yfit <- dnorm(...) hustota N(0,1) nad posloupností xfit
  Hp <- round(...) ... Haldova podmínka Np(1-p) zaokrouhlená na 4 des. místa
  d <- max(yfit, hist(zW, plot=F)$dens) ... najdeme nejvyšší hodnotu z výšek sloupců histogramu
                                            a hustoty N(0,1) nad posl. xfit
  if( Hp <= 9){c.hist <- 'red'; c.lines <- 'darkred'} ... nastavení červených odstínů histogramu a
                                                          křivky hustoty, když Np(1-p) ≤ 9
  if(Hp > 9){c.hist <- ... ; ...} ... nastavení zelených odstínů
                                      histogramu a křivky hustoty, když Np(1-p) > 9
  hist(zW, ... , col=c.hist, breaks= seq(floor(min(zW)), ceiling(max(zW)), by=0.5),
       ylim=c(0,d), ... ) ... histogram zW
          automatické nastavení hranic třídicích intervalů v histogramu
          floor(min(zW)): minimální hodnotu zW zaokrouhlíme dolů na nejbližší nižší celé číslo
          ceiling(max(zW)): maximální hodnotu zW zaokrouhlíme nahoru na nejbližší vyšší celé číslo
  box(bty='o') ... □
  lines(... , col=c.lines, ... ) ... křivka N(0,1); barva závisí na hodnotě vložené v proměnné c.lines
  mtext(bquote(...), ... ) ... popisek zw
  mtext(paste(...), ...) ... popisek N=..., p=..., Hp=...
}
```

2. Vytvoříme animaci ukazující změnu rozdělení test. statistiky z_W při $N \to \infty$ pro (a) $p=0.1$; (b) $p=0.5$; (c) $p=0.9$ najednou.

```
N <- c(5,10,...) ... posloupnost N ze zadání příkladu
... úvodní nastavení animace
saveLatex(for(i in 1:length(N)){
        par(...) □□□ ; 4□2 (2 nahoře, 6 dole)
        SimTSBinom(N=N[i], p=0.1) (a)
        SimTSBinom(...) (b)
        SimTSBinom(...) (c)
}
```

\# Nezapomeňte uvést komentář, zda podle vás Haldova podmínka dobře ukazuje na situace, kdy je možné rozdělení testovací statistiky z_W aproximovat normálním rozdělením $N(0,1)$ a kdy naopak ne.

8.2

$X \sim Bin(N, p)$, $N = 30$, $p = 0.8$. V $N = 30$ nastalo celkem $x = 24$ úspěchů $\Rightarrow$ $\hat{p} = Pr(X=1) = \frac{24}{30} = 0.8$.

skutečná, teoretická hodnota parametru p (neznáme ji, odhadujeme ji)

odhad parametru p na základě naměřených dat (počtu úspěchů v N pokusech)

Waldův $(1-\alpha)100\%$ DIS pro parametr p:

$$(dh; hh) = \left(\frac{x}{N} - u_{1-\alpha/2}\sqrt{p(1-p)/N} \; ; \; \frac{x}{N} - u_{\alpha/2}\sqrt{p(1-p)/N} \right) = (0.657, 0.943)$$

Myšlenka výpočtu psti pokrytí Waldova $(1-\alpha)100\%$ IS je odlišná od té, kterou jsme používali při výpočtu psti pokrytí Waldova IS pro μ (resp. σ^2). Zde využíváme toho, že počet úspěchů v N pokusech je konečný. Pst pokrytí počítáme podle vzorce:

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

$p_j \in \left\{ \frac{1}{N}, \frac{2}{N}, \dots, \frac{N-2}{N}, \frac{N-1}{N} \right\}$.

Myšlenku vzorce je nejsnadnější vysvětlit názorně pomocí tabulky, která je jedním z finálních výsledků příkladu. Zaměříte-li se postupně na jednotlivé sloupce tabulky, pochopíte i jednotlivé kroky vedoucí k výsledné psti pokrytí.

x_j	p_j	dh_j	hh_j	$Pr(X=Np_j)$	$p \in IS_j$
1	1/30	-0.0309	0.0976	0	0
2	2/30	-0.0226	0.1559	0	0
⋮	⋮	⋮	⋮	⋮	⋮
13	13/30	0.2560	0.6107	0	0
14	14/30	0.2881	0.6452	0	0
15	15/30	0.3211	0.6789	0.0001	0
16	16/30	0.3548	0.7119	0.0007	0
⋮	⋮	⋮	⋮	⋮	⋮ 0 1 ⋮
26	26/30	0.7450	0.9883	0.1325	1
27	27/30	0.7926	1.0074	0.0785	1
28	28/30	0.8441	1.0226	0.0337	0
29	29/30	0.9024	1.0309	0.0093	0
x_j	x_j/N				

$x_j/N - u_{1-\alpha/2}\sqrt{p_j(1-p_j)/N}$ dolní hranice Waldova IS pro p_j

$x_j/N - u_{\alpha/2}\sqrt{p_j(1-p_j)/N}$ horní hranice Waldova IS pro p_j

$Pr(X = Np_j) = \text{dbinom}(x_j, N, p)$, Pst, že za předpokladu, že $X \sim Bin(N, p)$, bude počet úspěchů právě x_j.

$I(p \in (dh_j, hh_j))$ indikační funkce: $I(\dots) = 1$, pokud $p \in (dh_j, hh_j)$; $I(\dots) = 0$, pokud $p \notin (dh_j, hh_j)$. Náleží odhad p ($\hat{p}$) do Waldova IS?

$$Pr(\text{pokrytí}) = \sum_j Pr(X = Np_j : p \in (dh_j, hh_j)) = \text{sum}(Pr(X = Np_j) * (p \in IS_j))$$

1. Vytvoříme funkci PokrytiWSL( ), která pro zadané hodnoty N, p, α a zvolený typ intervalu (zatím pouze Waldův) vrátí (a) tabulku 29x6 uvedenou ve výsledcích ; (b) psť pokrytí. V rámci příkladu 8.3 obohatíme tuto funkci o skóre a věrohodnostní typ intervalu.

```
PokrytiWSL <- function(p, N, alpha= ... ,type='wald'){
  xj <- ...   posloupnost celých čísel od 1 do (N-1). Lze definovat např. pomocí :
  pj <- xj/N  ... vektor pravděpodobností pj (délka = N-1 = 29 (pro N=30))

  if(type=='wald'){
    dhj <- xj/N - u_{1-α/2} sqrt(pj(1-pj)/N)  ... vektor dolních hranic Waldových IS (délka = N-1)
    hhj <- xj/N - u_{α/2} sqrt(pj(1-pj)/N)    ... vektor horních hranic Waldových IS (délka = N-1)
  }
  (*)... zde potom v rámci příkladu 8.3 doplníme výpočet dolní a horní hranice skóre
         a věrohodnostního IS.

  pr.X.Npj <- dbinom(...)... Pr(X=Npj) za předp. že X~Bin(N,p) ... vektor o délce N-1
  p.in.ISj <- (p>dhj & p<hhj)... p∈ISj ... vektor o délce N-1
  tab <- data.frame(...)... tabulka výsledků se sloupci xj, pj, dhj, hhj, pr.X.Npj, p.in.ISj
  pst.pokryti <- sum(Pr(X=Npj) * (p∈ISj))... výpočet psti pokrytí (1 číslo)
  return(list(tab=tab, pst.pokryti = pst.pokryti))... seznam (list) výsledků používáme tehdy, když
                                                       chceme, aby funkce měla více výstupů. Na
                                                       každý výstup se potom odvoláváme přes $. (viz
                                                       bod 2.)
}
```

2. Pomocí funkce PokrytiWSL( ) vygenerujeme tabulku 29x6 pro (a) $N=30$, $p=0.8$; (b) $N=30$, $p=0.79$. Při tom si vyzkoušíme 'dovolání se konkrétního výsledku pomocí $.

```
tab80 <- PokrytiWSL(p=0.8, N=30, ...)$tab
round(tab80, 4)... zaokrouhlení na 4 des. místa
tab79 <- PokrytiWSL(...)$tab
round(tab79, ...)... zaokrouhlení na 4 des. místa
```

3. Pomocí funkce PokrytiWSL(...) vypočítáme psť. pokrytí pro Waldův 95% IS při (a) $N=30$, $p=0.8$; (b) $N=30$, $p=0.79$.

```
pst.pokryti80 <- PokrytiWSL(...)$pst.pokryti
pst.pokryti79 <- PokrytiWSL(...)$pst ... konkrétní výstup získáme, když jej přes $ zavoláme jeho
                                          názvem. Název však nemusíme psát celý. Vyzkoušejte si to tak,
                                          že místo $pst.pokryti napíšete pouze $pst nebo $p, apod.
tab.pst <- data.frame(...)... tabulka obsahující obě psti pokrytí.
round(tab.pst)
```

Komentář: $\alpha = 0.05 \to$ Očekávaná psť pokrytí Waldova 95% IS je tedy 0.95. Uvědomme si, že pokud máme reálnou situaci, kdy v $N=30$ pokusech sledujeme počet úspěchů, potom odhad skutečné psti nastání úspěchu p (tj. $\hat{p} = \frac{X}{N}$) je velmi hrubý, protože je založen na diskrétním počtu úspěchů x. Tento deficit se projevuje právě např. na výši psti. pokrytí. Pokud skutečná psť. úspěchu $p=0.8$ a počet úspěchů v $N=30$ pokusech je 24, tj. $\hat{p} = \frac{X}{N} = \frac{24}{30} = 0.8$, je vše v pořádku a psť. pokrytí Waldova IS je cca 0.95. Nicméně, co když je skutečná psť úspěchu $p=0.79$? Nejbližším odhadem je opět $\hat{p} = \frac{24}{30} = 0.8$, ale vidíme, že tento odhad už je nepřesný. Psť pokrytí v tomto případě klesá na cca 0.89.

8.3

V tomto příkladu se zaměříme na globální pohled na pst pokrytí Waldova 95% DIS. Zatímco v příkladu S6.2 jsme spočítali pst pokrytí Waldova 95% DIS za předpokladu, že $X \sim Bin(N, p)$, kde (a) $p = 0.80$, (b) $p = 0.79$, zde si pst pokrytí spočítáme pro každé $p \in \{0.001, 0.003, \dots, 0.997, 0.999\}$. Následně vykreslíme graf závislosti psti pokrytí na hodnotě p. Graf vykreslíme pro (i) $N = 30$; (ii) $N = 100$; (iii) $N = 500$. Nakonec do funkce PokrytiWSL( ) doprogramujeme výpočet psti pokrytí pro skóre DIS a věrohodnostní DIS a vykreslíme analogické grafy pro skóre a věrohodnostní 95% DIS.

I. Vykreslení grafu psti pokrytí Waldova 95% DIS:

Poznámka: Ze zadání víme, že máme finálně vykreslit 3 grafy pro Waldův 95% DIS, 3 grafy pro skóre 95% DIS a 3 grafy pro věrohodnostní 95% DIS. Nejelegantnější tedy bude vytvořit funkci PokrytiPlot( ), která pro zadané hodnoty N, α a typ IS vrátí příslušný graf.

1. Vytvoříme fci:

```
PokrytiPlot <- function(N, alpha= ... , type= 'wald') {
    pst <- ... posl. od 0.001 do 0.999 po kroku 0.002
    pokryti <- NULL ... příprava prázdného vektoru
    for(i in 1:length(pst)) {
        pokryti[i] <- PokrytiWSL(p=pst[i], N=N, alpha=alpha, type=type)
    } ... cyklus ... pro každou hodnotu z vektoru pst spočítá pst pokrytí.

    if(type == 'wald') {text <- bquote(paste(...))} ... popisek 'Waldův XX% IS, N=...'
(*) if(   -||-   'score') {text <- ..............} ... popisek 'Skóre XX% IS, N=...'
    if(   -||-   'likely') {text <- ..............} ... popisek 'Věrohodnostní XX% IS, N=...'

    par(...) ... 4□2 (2 nahoře, 5 dole)
    plot(pst, pokryti, ylim=c(0.6, 1), ...) ... vykreslení grafu psti pokrytí
    mtext(...) ... popisek 'p'
    mtext(text, ...) ... popisek zvolený v závislosti na typu IS (viz (*))
    abline(...) ... vodorovná šedá čára v hodnotě koeficientu spolehlivosti (1-α)
}
```

α ve fci PokrytiWSL( ) bude taková, jakou zadáme ve fci PokrytiPlot( ).

Typ IS ve fci PokrytiWSL( ) bude stejný jako si zvolíme ve fci PokrytiPlot( ).

XX … automaticky se mění podle α: .((1-alpha)*100)

N=... automaticky se mění podle N: .(N)

2. Fci PokrytiPlot( ) použijeme pro vykreslení grafu psti pokrytí Waldova 95% DIS pro (i) $N = 30$; (ii) $N = 100$; (iii) $N = 500$:

```
par(...) ... okno rozdělíme na 9 grafů: 3x3
N <- c(30, 100, 500)
for (i in 1:length(N)) { PokrytiPlot(N= N[i], type='wald') }
```

Tip: Analogicky si ověřte vykreslete pst pokrytí Waldova 99% DIS. Zkontrolujte si, zda se vám správně změnil popisek pod osou x i poloha horizontální referenční čáry.

Zopakujte si definici konzervativního a liberálního IS (viz cv. 04). V komentářích uveďte zda je Waldův DIS pro parametr p konzervativní, liberální, nebo ani jedno.

II. Doplnění vzorců na výpočet pstí pokrytí skóre a věrohodnostního $(1-\alpha)100\%$ DIS.

V této části se vrátíme k funkci PokrytiWSL( ), kde na vyznačené místo (*) doplníme výpočet vektoru dolních hranic a vektoru horních hranic (1) skóre DIS pro p_j, $j=1,\dots,N-1$; (2) věrohodnostních DIS pro p_j, $j=1,\dots,N-1$. V obou případech pojmenujeme vektor dolních hranic dhj a vektor horních hranic hhj. Tím zajistíme hladkou návaznost následujících příkazů (pr.X.Npj <- …, (…), return(list(…)) bez ohledu na zvolený typ IS.

1. Doplnění výpočtu hranic skóre $(1-\alpha)100\%$ DIS ve fci PokrytiWSL( ):

Hranice $(1-\alpha)100\%$ skóre DIS vyjádříme explicitně pomocí následujícího vzorce (odvození finálního tvaru viz teorie, slidy k přednáškám nebo kniha Aplikovaná štatistická inferencia I (AŠI I) (Kalina, Králik, Hupková))

$$IS = (dh;\, hh) = \left(\hat{p}\left(\frac{N}{N+u^2}\right) + \frac{1}{2}\frac{u^2}{N+u^2} - u\sqrt{\frac{1}{N+u^2}\left[\hat{p}(1-\hat{p})\frac{N}{N+u^2} + \frac{1}{4}\frac{u^2}{N+u^2}\right]}\;;\right.$$
$$\left. \hat{p}\left(\frac{N}{N+u^2}\right) + \frac{1}{2}\frac{u^2}{N+u^2} + u\sqrt{\frac{1}{N+u^2}\left[\hat{p}(1-\hat{p})\frac{N}{N+u^2} + \frac{1}{4}\frac{u^2}{N+u^2}\right]} \right),$$

kde $u = u_{1-\alpha/2}$ = qnorm(1-alpha/2).

```
if(type == 'score'){
   u <- qnorm(...)   ... u_{1-α/2}
   dhj <- pj*(N/(N+u^2))+...   vektor dolních hranic skóre IS (délka = N-1)
   hhj <- pj*(N/(N+u^2))+...   vektor horních hranic skóre IS (délka = N-1)
}
```

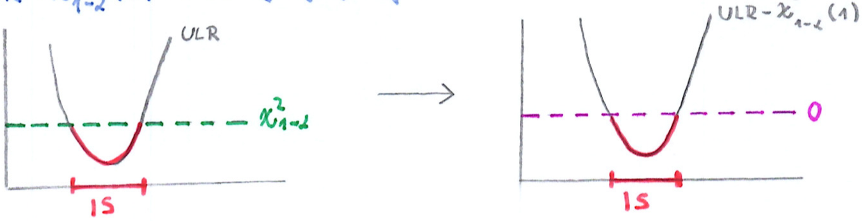
Pozn: Nezapomeňte v celém vzorci nahradit $\hat{p}$ vektorem pj. Analogicky jako jsme to provedli v příkazech počítajících vektor dolních, resp. horních hranic Waldova IS v 6.2.

Pozn: Při opisování vzorců do proměnných dhj a hhj buďte velmi opatrní a zápis si několikrát zkontrolujte. Vzorce jsou strašně složité a často se v nich při přepisu chybuje!!!

2. Doplnění výpočtu hranic věrohodnostního $(1-\alpha)100\%$ DIS ve fci PokrytiWSL( ):

Pro výpočet hranic věrohodnostního DIS explicitní vzorec neexistuje. Oblast věroh. DIS je opět definována jako množina hodnot p, pro které je test. statistika $ULR < \chi^2_{1-\alpha}(1)$. Přibližné hranice DIS bychom tedy mohli nalézt analogicky jako v příkladech 5.8, 6.6 a 7.3. Přesnost hranic však v těchto příkladech závisela na zvolené posloupnosti (μ, σ^2, resp. ϑ). Zde si ukážeme nový přístup, který nám hranice IS spočítá přesně. Stejný postup si pak můžete cvičně zkusit použít v příkladech 5.8, 6.6 a 7.3. a vypočítat si přesné hranice těchto IS.

Myšlenka nového přístupu je založena na převedení nerovnosti $ULR < \chi^2_{1-\alpha}(1)$ na nerovnost $ULR - \chi^2_{1-\alpha}(1) < 0$. Tj. graficky:



Hranice věrohodnostního $(1-\alpha)100\%$ DIS nalezneme jako kořeny rovnice $U_{LR} - \chi^2_{1-\alpha}(1) \leq 0$ např. pomocí funkce uniroot( ). Nahrazení $\leq$ za $<$ nám v tomto případě až tolik nevadí. Kulaté závorky okolo IS definují, že hraniční hodnoty do IS stejně nespadají =).

```
if(type == 'likely'){
```

Příprava fce ULRchisq(), která vrací hodnotu $U_{LR} - \chi^2_{1-\alpha}(1)$:

```
ULRchisq <- function(x, p, N, alpha=... ){
    uLRchisq <- ...
    return(...)
}
```

$$uLRchisq \leftarrow U_{LR} - \chi^2_{1-\alpha}(1) = -2\ln(\lambda(X)) - \chi^2_{1-\alpha}(1) = \underbrace{2\left(x\ln\frac{x}{Np} + (N-x)\ln\frac{N-x}{N-Np}\right)}_{-\ln(\lambda(X))} - \underbrace{\chi^2_{1-\alpha}(1)}_{\text{qchisq(...)}}$$

(odvození viz teorie, slidy, nebo kniha AŠII.)

Výpočet vektorů dolních a horních hranic IS (délka = N-1):

```
dhj <- hhj <- NULL
for(i in 1:length(pj)){
```

definováno v úvodu fce PokrytiWSL()

```
  dhj[j] <- uniroot(f = ULRchisq,
                    interval = c(1e-6, pj[j]),
                    x = xj[j], N = N, alpha = alpha)$root
```

f = ULRchisq, ... fce, jejíž kořen hledáme

interval = c(1e-6, pj[j]), interval, kde kořen leží (dolní hranice ISj bude určitě ležet mezi 0 a hodnotou pj (protože pj je střed ISj))

x = xj[j], N = N, alpha = alpha ... specifikace všech ostatních parametrů fci ULRchisq().

$root ... fce uniroot() vrací více výstupů najednou. Hodnota kořene je uložena ve výstupu **root**.

```
  hhj[j] <- uniroot(..., interval = c(pj[j], 1-1e-6), ...)
}
```

... vektor horních hranic věrohodn. DIS (délka = N-1).

Pozn: Ptáte se, proč jsme tento postup výpočtu přesných hranic nepoužili už dříve v příkladech 5.8, 6.6 a 7.3 ? Protože i já se průběžně učím nové metody a vylepšuji staré postupy na lepší a přesnější. A toto je přesně ten případ =).

3. Vykreslení grafů pstí pokrytí skóre a věrohodnostního 95% DIS:

Grafy vykreslíme opět pro (i) N=30; (ii) N=100; (iii) N=500, a to pomocí fce PokrytiPlot(). Pokud jsme fci naprogramovali pomocí postupu uvedeného v I-1, nemusíme už nic modifikovat. Pouze znovu spustíme kód obou funkcí PokrytiWSL() a PokrytiPlot(), abychom je aktualizovali a mohli je vykreslovat.

```
N <- c(30, 100, 500)
for (i in 1:length(N)) {PokrytiPlot(..., type = 'score')}
for (     -||-       ) {   -||-    (..., type = 'likely')}
```

... trojice grafů pstí pokrytí skóre DIS (N=30, 100, 500)

... trojice grafů pstí pokrytí věroh. DIS (-||-).

Pozn: Závěrem opět v komentářích uveďte, zda je skóre a věrohodnostní DIS konzervativní, liberální, nebo ani jedno.

Pozn: Výpočet pstí pokrytí věrohodnostního DIS je časově i výpočetně náročný. Uvědomme si, že pro situaci N=500 hledáme pro každé $p \in \{0.001, 0.003, \dots, 0.999\}$ 499 x dolní hranici (pro $\forall p_j$, $j = 1 \dots (N-1)$) a 499x horní hranici. Množina $\{0.001, 0.003, \dots, 0.999\}$ obsahuje 500 hodnot. Funkce uniroot() tedy pro N=500 proběhne 499 x 500 x 2 = 499 000x. Z toho důvodu může vykreslení grafu (v závislosti na počítači) trvat déle. (U mě např. 2-3 minuty.)

8.4

Z řešení příkladu 8.3 vidíme, že Waldův IS si ze všech tří typů intervalů vedl nejhůře. Jeho aktuální psť pokrytí je pro malá N výrazně nižší než nominální psť pokrytí. Stejně tak je aktuální psť pokrytí neuspokojivá pro velká N ($N=500$) je-li psť p moc malá ($p<0.1$) nebo moc velká ($p>0.9$). Navíc se v praxi může stát že dolní hranice Waldova IS bude menší než 0 (pokud $\hat{p}$ bude blízký 0) nebo větší než 1 (pokud $\hat{p}$ bude blízký 1). Toto všechno jsou nežádoucí vlastnosti Waldova DIS. Jeho nespornou výhodou je však výpočetní jednoduchost (znatelná např. při srovnání s věrohodnostním DIS). Snahou statistiků tedy bylo najít kompromis mezi výpočetní jednoduchostí a uspokojivými vlastnostmi DIS. Jednou z cest jsou tzv. zpětně transformované DIS. Jejich myšlenka je založena na tom, že vytvoříme vhodnou transformaci parametru p (tj. $g(p)$), pomocí Δ metody spočítáme odhad rozptylu této transformace (tj. $\widehat{Var}[g(\hat{p})]$). Vypočítáme dolní a horní hranici DIS pro tuto transformaci (výpočet DIS je založen na Waldově principu), tj. obecně:

$$(dh_g ; hh_g) = \left(g(\hat{p}) - u_{1-\alpha/2}\sqrt{\widehat{Var}[g(\hat{p})]} \; ; \; g(\hat{p}) - u_{\alpha/2}\sqrt{\widehat{Var}[g(\hat{p})]} \right)$$

a nakonec hranice IS pro transformaci zpětně transformujeme (vhodnou transformací) na hranice IS pro parametr p.

Z teorie již známe tři (více či méně vhodné) transformace parametru p:

1. podíl šancí:

$$g(p) = \frac{p}{1-p} \qquad \widehat{Var}[g(\hat{p})] = \frac{\hat{p}}{N(1-\hat{p})^3} \qquad \text{Zpětná transformace: } (dh ; hh) = \left(\frac{dh_g}{1+dh_g} \; ; \; \frac{hh_g}{1+hh_g} \right)$$

2. logaritmus podílu šancí:

$$g(p) = \ln\left(\frac{p}{1-p}\right) \qquad \widehat{Var}[g(\hat{p})] = \frac{1}{N\hat{p}(1-\hat{p})} \qquad \text{Zpětná transformace: } (dh, hh) = \left(\frac{e^{dh_g}}{1+e^{dh_g}} \; ; \; \frac{e^{hh_g}}{1+e^{hh_g}} \right)$$

3. arcussinus:

$$g(p) = \arcsin(\sqrt{p}) \qquad \widehat{Var}[g(\hat{p})] = \frac{1}{4N} \qquad \text{Zpětná transformace: } (dh, hh) = \left(\sin^2(dh_g) \; ; \; \sin^2(hh_g) \right)$$

Jak kvalitní jsou vlastnosti všech tří výše uvedených DIS z hlediska psti pokrytí si ukážeme v tomto příkladu. =)

Tip: Procvičte si odvození $\widehat{Var}[g(\hat{p})]$ pomocí Δ metody ve všech 3 případech.
Zde si pro ukázku odvodíme $\widehat{Var}[g(\hat{p})]$ pro $g(p) = \ln\left(\frac{p}{1-p}\right)$:

$$L(p|x) = p^x (1-p)^{N-x}$$

$$\ell(p|x) = x \ln p + (N-x)\ln(1-p)$$

$$\frac{\partial}{\partial p}\ell(p|x) = \frac{x}{p} - \frac{N-x}{1-p}$$

$$\frac{\partial^2}{\partial p^2}\ell(p|x) = -\frac{x}{p^2} - \frac{N-x}{(1-p)^2}$$

$$I(p) = -\frac{\partial^2}{\partial p^2}\ell(p|x) = \frac{x}{p^2} + \frac{N-x}{(1-p)^2}$$

$$I(\hat{p}) = \frac{x}{\hat{p}^2} + \frac{N-x}{(1-\hat{p})^2} \qquad / x = N\hat{p}$$

$$= \frac{N\hat{p}}{\hat{p}^2} + \frac{N-N\hat{p}}{(1-\hat{p})^2} = \frac{N}{\hat{p}} + \frac{N}{1-\hat{p}}$$

$$= \frac{N(1-\hat{p}) + N\hat{p}}{\hat{p}(1-\hat{p})} = \frac{N}{\hat{p}(1-\hat{p})}$$

$$g(p) = \ln\left(\frac{p}{1-p}\right)$$

$$(1-p)^{-1} + p(1-p)^{-2} = \frac{1}{1-p} + \frac{p}{(1-p)^2}$$

$$\frac{\partial}{\partial p}\ln\left(\frac{p}{1-p}\right) = \frac{1}{\frac{p}{1-p}} \cdot \left(p(1-p)^{-1}\right)' = \frac{1-p}{p}\left(\frac{1-p+p}{(1-p)^2}\right)$$

$$= \frac{1-p}{p} \cdot \frac{1}{(1-p)^2} = \frac{1}{p(1-p)}$$

Δ metoda:

$$\widehat{Var}[g(\hat{\theta})] = \left[\frac{\partial}{\partial\theta} g(\theta)\Big|_{\theta=\hat{\theta}}\right]^2 \Big/ I(\hat{\theta})$$

$$\widehat{Var}[g(\hat{p})] = \left[\frac{\partial}{\partial p} g(p)\Big|_{p=\hat{p}}\right]^2 \Big/ I(\hat{p})$$

$$= \left[\frac{\partial}{\partial p}\ln\left(\frac{p}{1-p}\right)\Big|_{p=\hat{p}}\right]^2 \Big/ I(\hat{p})$$

$$= \left[\frac{1}{\hat{p}(1-\hat{p})}\right]^2 \Big/ \left(\frac{N}{\hat{p}(1-\hat{p})}\right)$$

$$= \frac{1}{\hat{p}^2(1-\hat{p})^2} \cdot \frac{\hat{p}(1-\hat{p})}{N} = \boxed{\frac{1}{N\hat{p}(1-\hat{p})}}$$

1. Řešení příkladu je podobné řešením uvedeným v příkladech 8.1, 8.2 a 8.3. Nejprve vytvoříme fci PokrytiZT( ), která pro zadané N, p, α a zvolenou transformaci vrátí pst. pokrytí.

PokrytiZT <- function (p, N, alpha = ... , type = '...'){

xj <- ... posl. 1, 2, ..., (N-1)

pj <- ... vektor pstí pj (délka = N-1)

if(type == 'or') {

dhg <- $\dfrac{p_j}{1-p_j} - u_{1-\alpha/2}\sqrt{\dfrac{p_j}{N(1-p_j)^3}}$ ... vektor dolních hranic ISj pro podíl šancí (délka = N-1)

hhg <- $\dfrac{p_j}{1-p_j} - u_{\alpha/2}\sqrt{\dfrac{p_j}{N(1-p_j)^3}}$ ... vektor horních hranic ISj pro podíl šancí (N-1)

dh <- $\dfrac{dhg}{1+dhg}$ ... (zpětná transformace); vektor dolních hranic zpětně transf. IS pro parametr p (N-1)

hh <- $\dfrac{hhg}{1+hhg}$ ... (zpětná transf.); vektor horních hranic zpětně transf. IS pro parametr p (N-1)

}

if(type == 'log){

dhg <- ... vektor dolních hranic ISj pro logaritmus podílu šancí (N-1)

hhg <- ... vektor horních hranic ISj pro logaritmus podílu šancí (N-1)

dh <- ... (zpětná transf.); vektor dolních hranic zpětně transf. IS pro parametr p (N-1)

hh <- ... (zpětná transf.); vektor horních hranic zpětně transf. IS pro parametr p (N-1)

}

if(type == 'arcsin'){

dhg <- asin(sqrt(pj)) - ... vektor dolních hranic ISj pro arcsin($\sqrt{\ }$) (délka = N-1)

hhg <- asin (sqrt (pj)) - ... -||- horních hranic -||- (N-1)

dh <- ... (zpětná transf.); vektor dolních hranic zpětně transf. IS pro parametr p (N-1)

hh <- ... (zpětná transf.); -||- horních hranic -||- (N-1)

}

stejné jako v PokrytiWSL().

pr.X.Npj <- dbinom(...) ... $Pr(X = Np_j)$ za předpokladu, že $X \sim Bin(N, p)$ ... (délka = N-1)

p.in.ISj <- ... $p \in IS_j$ ... vektor o délce N+1

pst.pokryti <- sum($Pr(X=Np_j)$ * ($p \in IS_j$)) ... výpočet pstí pokrytí (1 číslo)

return (pst.pokryti)

}

2. Vytvoříme fci PokrytiZTPlot( ), která pro zadané hodnoty N, α a zvolený typ transformace vrátí příslušný graf.

Tip: funkce PokrytiZTPlot( ) je analogická fci PokrytiPlot( ). Proto si ji zkuste vytvořit sami. Stačí kód fce PokrytiPlot( ) zkopírovat a vhodně modifikovat. =)

Tip2: Nechce-li se vám vytvářet dvě nové fce PokrytiZT( ) a PokrytiZTPlot( ), můžete zkusit pouze rozšířit fci PokrytiWSL( ) a PokrytiPlot( ) o kód s transformacemi a grafy pstí pokrytí zpětně transf. DIS vygenerovat pomocí nich.

3. Vykreslíme grafy pstí pokrytí zpětně transf. DIS s transformací (a) podílu šancí; (b) logaritmu podílu šancí; (c) arcsin($\sqrt{\ }$) pro (i) $N=30$; (ii) $N=100$; (iii) $N=500$. (Viz 8.3). + Nezapomeňte na komentáře! =)