V rámci tohoto příkladu si vyzkoušíme aplikaci Waldova a věrohodnostního testu o podílu rozptylů na reálná data. Před samotným testováním ověříme předpoklad normality obou náhodných výběrů. H_0 ze zadání následně otestujeme všemi třemi způsoby (krit. oborem, IS i p-hodnotou) pro každý test. Nezapomeneme vždy uvést zdůvodněný závěr o H_0 a nakonec interpretaci závěru testování. Nakonec vykreslíme hranice a oblast 95% věrohodnostního DIS.

---

1. V rámci příkladu máme testem porovnat (a) shodu rozptylů největší šířky mozkovny mužů a žen; (b) shodu rozptylů morfologické výšky tváře mužů a žen. Tutéž testovací proceduru máme tedy provést 2x. Abychom nemuseli uvádět 2x tentýž kód, vytvoříme si funkci FULRtest( ), která pro zadané vektory dat X a Y a hl. významnosti α vrátí souhrnnou tabulku výsledků F-testu i věrohodnostního testu, a navíc vykreslí graf s hranicemi a oblastí 95% věroh. DIS.

Příprava dat

data <- read.delim(...) ... načtení datového souboru

skull.m <- na.omit(data[...,...]) ... z tabulky data vybereme pouze řádky týkající se mužů a sloupec skull.pH (odstranění NA hodnot)

skull.f <- ... -||- -||- -||- žen -||-

face.m <- na.omit(data[... ,...]) ... z tabulky data vybereme pouze řádky týkající se mužů a sloupec face.H

face.f <- ... -||- -||- -||- žen -||- -||-

FULRtest <- function(X, Y, alpha=0.05, plot=T, xlim=c(0,2), xlab='', col.IS='lightseagreen'){ (plot=T ... vykreslení grafu na přání)

X <- skull.m ; Y <- skull.f ... do X a Y si vložte data skull.m a skull.f, abyste při programování fce mohli průběžně ověřovat, že příkazy fungují správně a dávají správné výsledky. Po naprogramování fce FULRtest( ) řádek vymažeme. =)

Příprava proměnných pro testování

X <- na.omit(X) ... automatické odstranění NA hodnot (hodí se obecně, když vytváříme fci, která pracuje s reálnými daty)

Y <- ... -||-

S1 <- sd(X) ... s_1

S2 <- ... -||- ... s_2

n1 <- length(X) ... n_1 ... rozsah 1. náh. výběru

n2 <- ... -||- n_2 ... rozsah 2. náh. výběru

Dvouvýběrový F-test

KO

fW <- s_1^2/s_2^2 ... hodnota test. statistiky F_W

W.hh.fW <- ... $F_{n_1-1,n_2-1}(\alpha/2)$ ... horní hranice krit. oboru

W.dh.fW <- ... $F_{n_1-1,n_2-1}(1-\alpha/2)$ ... dolní hranice krit. oboru

} qf(..., n1-1, n2-1)

IS

dh.fW <- $\dfrac{s_1^2/s_2^2}{F_{n_1-1,n_2-1}(1-\alpha/2)}$ ... dolní hranice IS

hh.fW <- $\dfrac{s_1^2/s_2^2}{F_{n_1-1,n_2-1}(\alpha/2)}$ ... horní hranice IS

p

p.fW <- 2·min(pf(...), 1-pf(...)) ... p-hodnota

Test poměrem věrohodnosti

Pro testování poměrem věrohodnosti si nejprve vytvoříme fci `ULRstat( )`, která pro zadané hodnoty s_1, s_2, n_1, n_2 a sigma0 vrátí hodnotu test. statistiky U_{LR}. Dále vytvoříme fci `ULRchisq( )`, která pro zadané hodnoty s_1, s_2, n_1, n_2, sigma0 a alpha vrátí hodnotu $U_{LR} - \chi^2_1(1-\alpha)$.

sigma0 je zástupný symbol pro $\sigma_0 = \frac{\sigma_1^2}{\sigma_2^2}$

Příprava fcí pro U_{LR}

`ULRstat <- function(s1, s2, n1, n2, sigma0 = 1){`
`uLR <-` $-2\left(-\frac{n_1}{2}\ln\left(\frac{n_1}{n_1+n_2}+\frac{n_2 s_2^2\cdot\sigma_0}{s_1^2(n_1+n_2)}\right)-\frac{n_2}{2}\ln\left(\frac{n_1 s_1^2}{\sigma_0\cdot s_2^2(n_1+n_2)}+\frac{n_2}{n_1+n_2}\right)\right)$
`return(uLR)`
`}`

`ULRchisq <- function(s1, s2, n1, n2, sigma0, alpha = 0.05){`
`uLR.chisq <- ULRstat(s1, s2, n1, n2, sigma0) - qchisq(...)` $\chi^2_1(1-\alpha)$
`return(uLR)`
`}`

KO

`uLR <-` ... hodnota test. stat. U_{LR}
`W.dh.uLR <-` ... $\chi^2_1(1-\alpha)$ ... dolní hranice krit. oboru

IS

`s1s2 <- s1^2 / s2^2` ... $\frac{s_1^2}{s_2^2}$
`dh.uLR <- uniroot(..., c(dh.fW - 0.3, s1s2), ...)$root` ... dolní hranice IS
`hh.uLR <- uniroot(..., c(s1s2, hh.fw + 0.3), ...)$root` ... horní hranice IS

P

`p.uLR <- 1 - pchisq(...)` ... p-hodnota

`if(plot == T){` ... pokud plot == T, vykreslíme graf

Graf

`sigma0.i <- seq(...)` ... posloupnost od 0.01 do xlim[2] o délce minimálně 1000
`uLR.i <- uLRstat(..., sigma0.i)` ... vektor test. statistik uLR.i pro σ_{0i} (1000)
`if(1 < hh.uLR & 1 > dh.uLR) {col <- 'royalblue'}` ... odlišení situace, kdy $\sigma_0^2 = 1 \in IS$ (modrá)
`if(1 > hh.uLR | 1 < dh.uLR) {col <- 'red'}` ... a situace, kdy $\sigma_0^2 = 1 \notin IS$ (červená)
`plot(sigma0.i, uLR.i, ylim=c(0,15), xlim=xlim, ...)` ... ∪
`mtext(expression(...), ...)` ... popisek 'σ_1^2 / σ_2^2'
`mtext(xlab, ...)` ... popisek s názvem proměnné na ose x
`abline(...)` ... referenční čára (šedá) v hodnotě $\chi^2_1(1-\alpha)$
`lines(..., col=col.IS, ...)` ... ∪
`abline(..., col=col, ...)` ... referenční čára (svislá) v hodnotě $\sigma_0^2 = \frac{\sigma_1^2}{\sigma_2^2} = 1$ ... barva podle toho, zda $1 \in IS$ či $1 \notin IS$
`points(1, ULRstat(..., sigma0=1), ...)` ... bod se souřadnicemi $[\sigma_0^2 = 1, U_{LR}(\sigma_0^2 = 1)]$ ... barva -//-
`}`
`tab <- data.frame(...)` ... souhrnná tabulka výsledků
`return(tab)`
`}` ← zde končí tělo fce `FULRtest( )`

2.a Nyní provedeme test normality pro (a) největší šířku mozkovny mužů a žen:

```
# H0m:
# H1m:
# H0f:
# H1f:
lillie.test(...)   nebo   shapiro.test(...)
lillie.test(...)   nebo   shapiro.test(...)
# závěr pro muže:
# závěr pro ženy:
```

Interpretace výsledku testu normality pro muže:
-||- -||- pro ženy:

2.b Nyní provedeme test Ho ze zadání pro (a) a vykreslíme graf 95% věrohodnostního emp. DIS

```
FULRtest(skull.m, skull.f,...)
```

Závěr pro F-test: Vyberte 1 způsob (KO, IS, p) a pomocí něj stanovte závěr o Ho.
Závěr pro U_{LR}: Vyberte 1 způsob (KO, IS, p) a pomocí něj stanovte závěr o Ho.
Interpretace výsledků: Uveďte antropologický závěr.

3.a Analogicky jako v 2.a proveďte testy normality pro (b) morfologickou výšku tváře.
3.b Analogicky jako v 2.b proveďte test Ho ze zadání pro (b) a vykreslete graf 95% věroh. emp. DIS.

11.1

V rámci tohoto příkladu si vyzkoušíme aplikaci (1) Waldova ; (2) skóre ; (3) věrohodnostního testu na reálná data. H_0 ze zadání otestujeme všemi třemi způsoby (krit. oborem, IS i p-hodnotou) pro každý test. Nezapomeneme vždy uvést zdůvodněný závěr o H_0 a nakonec interpretaci závěru testování. Nakonec vykreslíme hranici a oblast 95% věrohodnostního DIS.

Příprava hodnot

n <- ... 0, 1, ..., 5 ... vektor n (délka = 6)
mn <- ... 109, 65, ..., 0 vektor m_n (6)

X <- rep(n, mn) ... vektor s počty úrazů v 200 případech : $\underbrace{0,\dots,0}_{109}, \underbrace{1,\dots,1}_{65}, \dots, \underbrace{5,\dots,5}_{0}$
N <- ... délka vektoru X (číslo)
m <- mean(X) ... $\bar{x}$ (číslo)
lambda0 <- ... λ_0
alpha <- ... α

Waldův test

H_0:
H_1:

KO

zW <- $\dfrac{\bar{x}-\lambda_0}{\sqrt{\bar{x}/N}}$ ... test. statistika z_W
W.hh.zW <- ... $u_{\alpha/2}$ ... horní hranice krit. oboru
W.dh.zW <- ... $u_{1-\alpha/2}$ ... dolní hranice krit. oboru
# závěr:

IS

dh.zW <- $\bar{x} - u_{1-\alpha/2}\cdot\sqrt{\bar{x}/N}$ ... dolní hranice IS
hh.zW <- $\bar{x} - u_{\alpha/2}\cdot\sqrt{\bar{x}/N}$ ... horní hranice IS
# závěr:

p

p.zW <- 2·min(pnorm(...), 1-pnorm(...)) ... p-hodnota
# závěr:

Skóre test

KO

uS <- $\dfrac{(\bar{x}-\lambda_0)^2}{\lambda_0/N}$ ... skóre test. statistika U_S

W.dh.uS <- ... $\chi^2_1(1-\alpha)$ ... dolní hranice krit. oboru
# závěr:

IS

q <- qnorm(...)
dh.uS <- $\bar{x} + \dfrac{1}{2}\dfrac{u^2_{1-\alpha/2}}{N} - u_{1-\alpha/2}\sqrt{\dfrac{1}{N}\left(\bar{x}+\dfrac{u^2_{1-\alpha/2}}{4N}\right)}$ ... dolní hranice IS

hh.uS <- $\bar{x} + \dfrac{1}{2}\dfrac{u^2_{1-\alpha/2}}{N} + u_{1-\alpha/2}\sqrt{\dfrac{1}{N}\left(\bar{x}+\dfrac{u^2_{1-\alpha/2}}{4N}\right)}$ ... horní hranice IS
# závěr:

p

p.uS <- 1-pchisq(...) ... p-hodnota
# závěr:

Věrohodnostní test:

Příprava fcí pro U_{LR} test

ULRstat <- function(N, X, lambda0){
uLR <- $2N\underbrace{\left(\bar{x}\ln\left(\frac{\bar{x}}{\lambda_0}\right) - \bar{x} + \lambda_0\right)}_{-\ln(\lambda(*))}$

return(uLR)
}

ULRchisq <- function(N, X, lambda0, alpha = 0.05){
uLR.chisq <- ULRstat(...) - $\underset{\downarrow \chi^2_1(1-\alpha)}{\text{qchisq(...)}}$
return(uLR.chisq)
}

KO

```
uLR <- ULRstat(...) ... test. statistika U_LR
W.dh.uLR <- ... χ²_1(1-α) ... dolní hranice krit. oboru
#Závěr:
```

IS

```
dh.uLR <- uniroot(..., c(0,m), ...)$root ... dolní hranice IS
hh.uLR <- uniroot(..., c(m,1), ...)$root ... horní hranice IS
#Závěr:
```

p

```
p.uLR <- 1-pchisq(...) ... p-hodnota
#Závěr:
```

```
tab <- data.frame(...) ... souhrnná tabulka výsledků
```

#Interpretace výsledků: Uveďte praktický závěr výsledků testování. Co jsme se dozvěděli o λ a co nám to říká o počtu úmrtí kopnutím koněm?

Příprava na graf

```
lambda0.i <- seq(..) ... posl. λ0,i od 0.2 do 1 o délce min. 1000
uLR.i <- ULRstat(...) ... vektor test. statistik U_LRi (1000)
```

Graf

```
par(...)
plot(lambda0.i, uLR.i, ylim=c(0,15), ...) ...
lines(...) ...
abline(...) ... horiz. referenční čára v hodnotě χ²_1(1-α)
```