Příklad 7.4. Jednovýběrový párový test
Načtěte datový soubor 03-paired-means-clavicle.txt a odstraňte z načtených dat NA hodnoty. Na hladině významnosti $\alpha = 0.05$ testujte, zda je u mužů délka klíční kosti na levé straně statisticky významně větší než na straně pravé.

Řešení příkladu 7.4
Nejprve stanovíme nulovou a alternativní hypotézu.

- H_0 : .... $\rightarrow$ ....
- H_1 : .... $\rightarrow$ .... (.. alternativa).
- Hladina významnosti $\alpha =$ ....
- .... test o .... když .... známe / neznáme.

Nutným předpokladem umožňujícím použití parametrického párového testu na otestování H_0 je **normalita rozdílů** mezi naměřenými hodnotami na levé a pravé straně.

Test normality rozdílů na levé a pravé straně

- H_0 : Rozdíly mezi levou a pravou stranou .... z normálního rozdělení.
- H_1 : Rozdíly mezi levou a pravou stranou .... z normálního rozdělení.

Hladina významnosti $\alpha =$............... . Nyní zjistíme rozsah náhodného výběru.

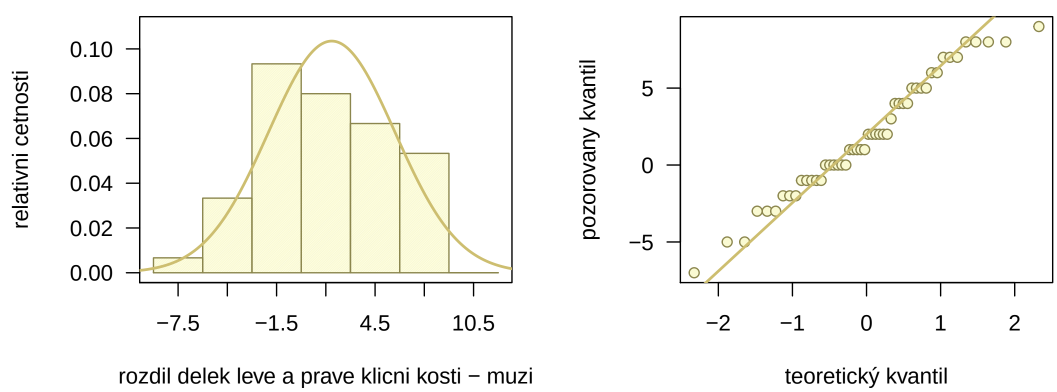
```
[1] 50
```
1

Protože náhodný výběr rozdílu naměřených délek klíčních kostí na levé a pravé straně u mužů má rozsah ...., což je .... než 30, použijeme na testování hypotézy o normalitě dat .... test.

```
[1] 0.1776153
```
2

P-hodnota vyšla .... Protože p-hodnota .... α, H_0 .... na hladině významnosti $\alpha = 0.05$.

Grafická vizualizace normality rozdílů délek pravých a levých klíčních kostí



Interpretace výsledku testu normality: Data .... z normálního rozdělení.

Párový test o střední hodnotě
V úvodu příkladu jsme si uvedli, že test o statisticky významném rozdílu mezi středními hodnotami dvou párových znaků můžeme převést na test o statisticky významném rozdílu mezi *rozdílem hodnot obou párových znaků* μ_d a nuly. Nulovou hypotézu $H_0 : \mu_d \leq 0$ oproti alternativní hypotéze $H_1 : \mu_d > 0$ testujeme nyní pomocí jednovýběrového testu o .... když .... neznáme.

a) Testování pomocí kritického oboru

```
[1] 3.41212
```
3

```
[1] 1.676551
```
4

Hodnota testovací statistiky t_0 je .... Kritický obor má tvar ....
Protože ...., H_0................................ na hladině významnosti $\alpha =$............................ .

b) Test pomocí intervalu spolehlivosti
Proti .... alternativě postavíme .... IS.

```
[1] 0.9460859
```
5

Interval spolehlivosti má tvar .... Protože ...., H_0................................ na hladině významnosti $\alpha =$............................ .

c) Test pomocí p-hodnoty

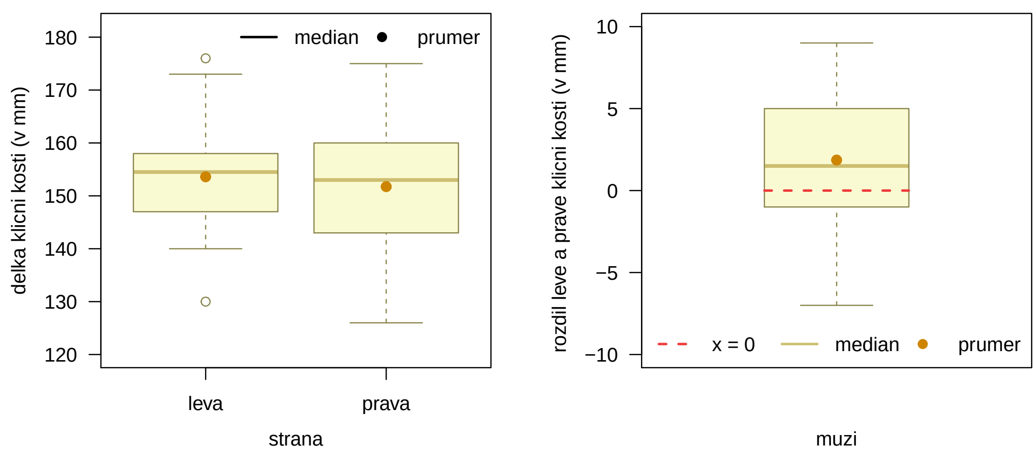
```
[1] 0.0006503568
```
6

p-hodnota vyšla .... Protože p-hodnota ...., H_0 .... na hladině významnosti $\alpha =$............................ .

Interpretace výsledků: Délka levé klíční kosti u mužů je / není statisticky významně vyšší než délka pravé líční kosti.

Příklad 7.5. Vizualizace statisticky významného rozdílu dvou náhodných výběrů
V rámci příkladu 7.4 jsme zjistili, že délka levé klíční kosti u mužů je statisticky významně vyšší než délka pravé klíční kosti u mužů. Získaný výsledek podpořte vykreslením krabicového diagramu.

Řešení příkladu 7.5



Příklad 7.6. Jednovýběrový test o parametru θ alternativního rozdělení
Určitá cestovní kancelář organizuje zahraniční zájezdy podle individuálních přání zákazníků. Z několika minulých let ví, že 30 % všech takto organizovaných zájezdů má za cíl zemi X. Po zhoršení politických podmínek v této zemi se cestovní kancelář obává, že se zájem o tuto zemi mezi zákazníky sníží. Ze 150 náhodně vybraných zákazníků v tomto roce má 38 za cíl právě zemi X. Potvrzují nejnovější data pokles zájmu o tuto zemi? Volte hladinu významnosti $\alpha = 0.05$.

Máme náhodný výběr $X_1, \ldots, X_{150}$ z rozdělení ....

- H_0 : ....
- H_1 : ....

Na testování hypotézy H_0 použijeme **jednovýběrový test o pravděpodobnosti** také nazývaný jako **test o parametru θ alternativního rozdělení**, který je **asymptotickým testem**. Před samotným testování je tedy nutné ověřit podmínku dobré aproximace $n\theta(1-\theta) > 9$. Protože $n\theta(1-\theta) =$ ....což je .... než 9, podmínka dobré aproximace je / není splněna.

a) Testování pomocí kritického oboru

```
[1] -1.247219
```
7

```
[1] -1.644854
```
8

Testovací statistika t_0 nabývá hodnoty ...., kritický obor má tvar ....
Protože t_0...........W, H_0 .... na **asymptotické** hladině významnosti $\alpha =$ ....

b) Testování pomocí intervalu spolehlivosti

```
[1] 0.3117439
```
9

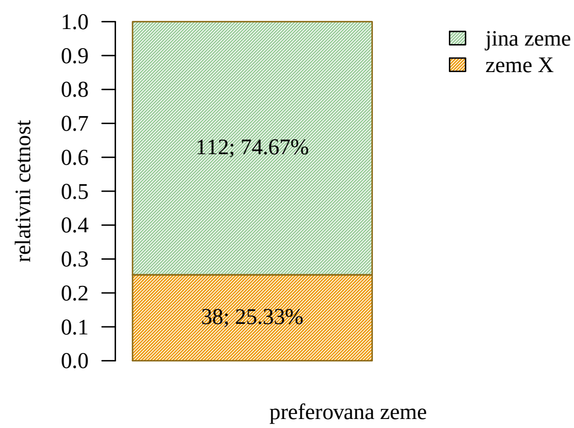
Interval spolehlivosti má tvar ....

Protože ...., H_0 .... na **asymptotické** hladině významnosti $\alpha =$ ....

c) Testování pomocí p-hodnoty

```
[1] 0.1061586
```
10

Protože p-hodnota $=$ ...., H_0 .... na **asymptotické** hladině významnosti $\alpha =$ ....



Interpretace výsledků testování: Nejnovější data potvrzují / nepotvrzují pokles zájmu zákazníků o zemi X.

8 Dvouvýběrové parametrické testy

Příklad 8.1. Dvouvýběrový tets o rozdílu středních hodnot
Načtěte datový soubor 01-one-sample-mean-skull-mf.txt a odstraňte z načtených dat NA hodnoty. Na hladině významnosti $\alpha = 0.05$ otestujte nulovou hypotézu o shodě střední hodnoty největší šířky mozkovny mužů se střední hodnotou největší šířky mozkovny u žen starověké egyptské populace.

Řešení příkladu 8.1
V rámci tohoto příkladu pracujeme se .... náhodnými výběry. První výběr obsahuje naměřené údaje o největší šířce mozkovny u starověké egyptské .... populace, druhý výběr obsahuje naměřené údaje o největší šířce mozkovny u starověké egyptské .... populace. Nejprve stanovíme nulovou a alternativní hypotézu.

- H_0 : ....
- H_1 : .... (.. alternativa).
- Hladina významnosti $\alpha =$ ....
- Test o .... když .... známe / neznáme.

Nutnými předpoklady umožňujícími použití parametrického testu na otestování nulové hypotézy je

1. .... naměřených hodnot (zvlášť v každém výběru);
2. shoda .... obou náhodných výběrů.

Test normality naměřených hodnot pro muže
Protože máme dva výběry, musíme provést test normality dat pro každý výběr zvlášť.

- H_0 : Naměřené hodnoty největší šířky mozkovny pro muže .... z normálního rozdělení.
- H_1 : Naměřené hodnoty největší šířky mozkovny pro muže .... z normálního rozdělení.

Hladina významnosti $\alpha =$............... . Nyní zjistíme rozsah náhodného výběru pro muže.

```
[1] 216
```
11

Protože náhodný výběr rozdílu naměřených šířek mozkovny u mužů má rozsah ...., což je .... než 30, použijeme na testování hypotézy o normalitě dat .... test.

```
[1] 0.07662229
```
12

P-hodnota vyšla .... Protože p-hodnota .... α, H_0 .... na hladině významnosti $\alpha = 0.05$.

Test normality naměřených hodnot pro ženy

- H_0 : Naměřené hodnoty největší šířky mozkovny pro ženy .... z normálního rozdělení.
- H_1 : Naměřené hodnoty největší šířky mozkovny pro ženy .... z normálního rozdělení.

Hladina významnosti $\alpha =$............... . Nyní zjistíme rozsahy náhodného výběru pro ženy.

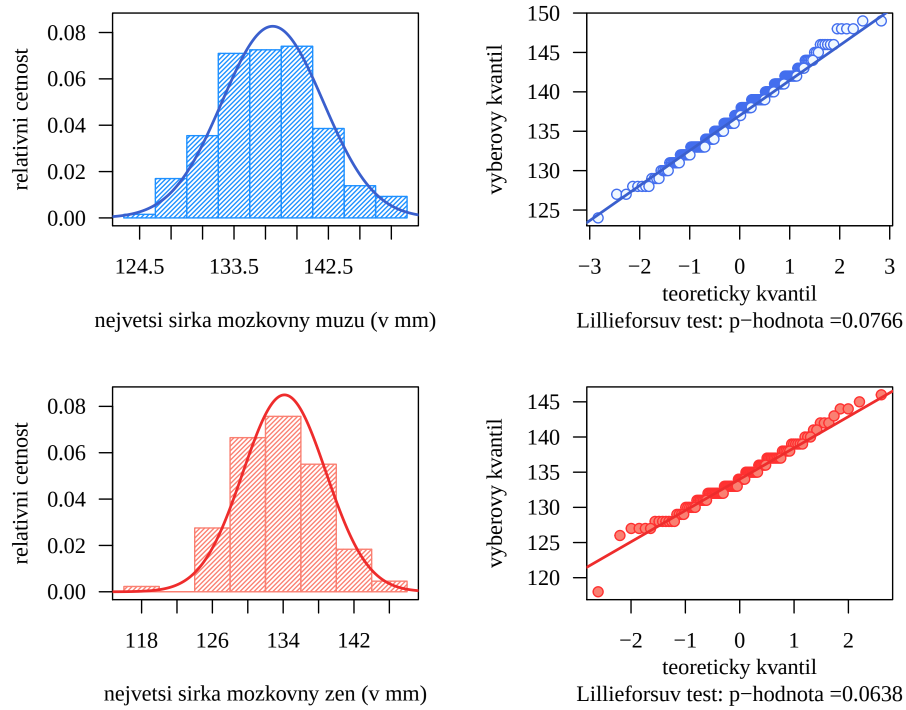
```
[1] 109
```
13

Protože náhodný výběr rozdílu naměřených šířek mozkovny u žen má rozsah ...., což je .... než 30, použijeme na testování hypotézy o normalitě dat .... test.

```
[1] 0.06380994
```
14

P-hodnota vyšla .... Protože p-hodnota .... α, H_0 .... na hladině významnosti $\alpha = 0.05$.

Interpretace výsledků: Naměřené hodnoty největší šířky mozkovny pro muže i pro ženy .... z normálního rozdělení.



Test o podílu rozptylů

První předpoklad o normalitě obou výběrů je / není splněn. Nyní ještě musíme ověřit předpoklad o shodě rozptylů obou náhodných výběrů. Nejprve stanovíme nulovou a alternativní hypotézu.

- H_0 : ....
- H_1 : .... (.. alternativa).
- Hladina významnosti $\alpha =$ ....
- Test o ....

a) Testování pomocí kritického oboru

```
[1] 1.055543
```
15

```
[1] 0.7266694
```
16

```
[1] 1.401231
```
17

Hodnota testovací statistiky t_0 je .... Kritický obor má tvar ....
Protože ...., H_0................................ na hladině významnosti $\alpha =$............................ .

b) Test pomocí intervalu spolehlivosti
Proti .... alternativě postavíme .... IS.

```
[1] 0.7532968
```
18

```
[1] 1.452576
```
19

Interval spolehlivosti má tvar .... Protože ...., H_0................................
na hladině významnosti $\alpha =$............................ .

c) Test pomocí p-hodnoty

```
[1] 0.761025
```
20

P-hodnota vyšla .... Protože p-hodnota ...., H_0 .... na hladině významnosti α =............................ .

Interpretace výsledků: Mezi rozptylem největší šířky mozkovny u mužů a rozptylem největší šířky mozkovny u žen starověké egyptské populace existuje / neexistuje statisticky významný rozdíl.

Test o rozdílu středních hodnot

Protože (1) předpoklad normality obou náhodných výběrů i (2) předpoklad shody rozptylů obou náhodných výběrů jsou / nejsou splněny, můžeme provést parametrický test o .... středních hodnot když rozptyly σ_1^2 a σ_2^2 ...., ale víme, že jsou shodné.

a) Testování pomocí kritického oboru

```
[1] 5.407945
```
21

```
[1] -1.967336
```
22

```
[1] 1.967336
```
23

Hodnota testovací statistiky t_0 je .... Kritický obor má tvar ....
Protože ...., H_0................................ na hladině významnosti α =............................ .

b) Test pomocí intervalu spolehlivosti
Proti .... alternativě postavíme .... IS.

```
[1] 1.93307
```
24

```
[1] 4.143723
```
25

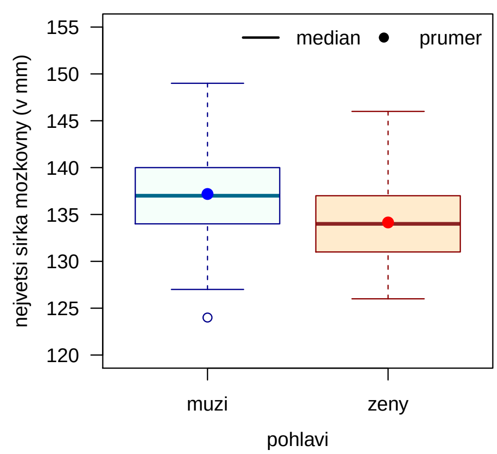
Interval spolehlivosti má tvar .... Protože ...., H_0................................ na hladině významnosti α =............................ .

c) Test pomocí p-hodnoty

```
[1] 1.242523e-07
```
26

P-hodnota vyšla .... Protože p-hodnota ...., H_0 .... na hladině významnosti α =............................ .

Vizualizace statisticky významného rozdílu dvou náhodných výběrů



Interpretace výsledků: Mezi největší šířkou mozkovny u mužů starověké egyptské populace a u žen starověké egyptské populace existuje / neexistuje statisticky významný rozdíl.

Příklad 8.2. Test o rozdílu středních hodnot
Mějme datový soubor 19-more-samples-correlations-skull.txt a proměnnou nose.B popisující šířku nosu v mm. Předpokládejme, že náhodná veličina X, popisující *šířku nosu bantuské populace*, pochází z normálního rozdělení $N(\mu_1, \sigma_1^2)$, a že náhodná veličina Y, popisující *šířku nosu čínské populace*, pochází z normálního rozdělení $N(\mu_2, \sigma_2^2)$. Na hladině významnosti $\alpha = 0.05$ testujte nulovou hypotézu o tom, že šířka nosu mužů čínské populace je menší nebo rovna šířce nosu mužů bantuské populace.

Řešení příkladu 8.2
V rámci tohoto příkladu pracujeme se .... náhodnými výběry. První výběr obsahuje naměřené údaje o šířce nosu mužů .... populace, druhý výběr obsahuje naměřené údaje o šířce nosu mužů .... populace. Nejprve stanovíme nulovou a alternativní hypotézu.

- H_0 : ....
- H_1 : .... (.. alternativa).
- Hladina významnosti $\alpha =$ ....
- Test o .... když .... známe / neznáme.

Nutnými předpoklady umožňujícími použití parametrického testu na otestování nulové hypotézy je

1. .... naměřených hodnot (zvlášť v každém výběru);
2. shoda .... obou náhodných výběrů.

Test normality naměřených hodnot pro muže čínské populace
Protože máme dva výběry, musíme provést test normality dat pro každý výběr zvlášť.

- H_0 : Naměřené hodnoty šířky nosu mužů čínské populace .... z normálního rozdělení.
- H_1 : Naměřené hodnoty šířky nosu mužů čínské populace .... z normálního rozdělení.

Hladina významnosti $\alpha =$............... . Nyní zjistíme rozsah náhodného výběru pro muže čínské populace.

```
[1] 19
```
27

Protože náhodný výběr rozdílu naměřených šířek nosu mužů čínské populace má rozsah ...., což je .... než 30, použijeme na testování hypotézy o normalitě dat .... test.

```
[1] 0.1173442
```
28

P-hodnota vyšla .... Protože p-hodnota .... α, H_0 .... na hladině významnosti $\alpha = 0.05$.

Test normality naměřených hodnot pro muže bantuské populace

- H_0 : Naměřené hodnoty šířky nosu u mužů bantuské populace .... z normálního rozdělení.
- H_1 : Naměřené hodnoty šířky nosu u mužů bantuské populace .... z normálního rozdělení.

Hladina významnosti $\alpha =$............... . Nyní zjistíme rozsahy náhodného výběru pro muže bantuské populace.

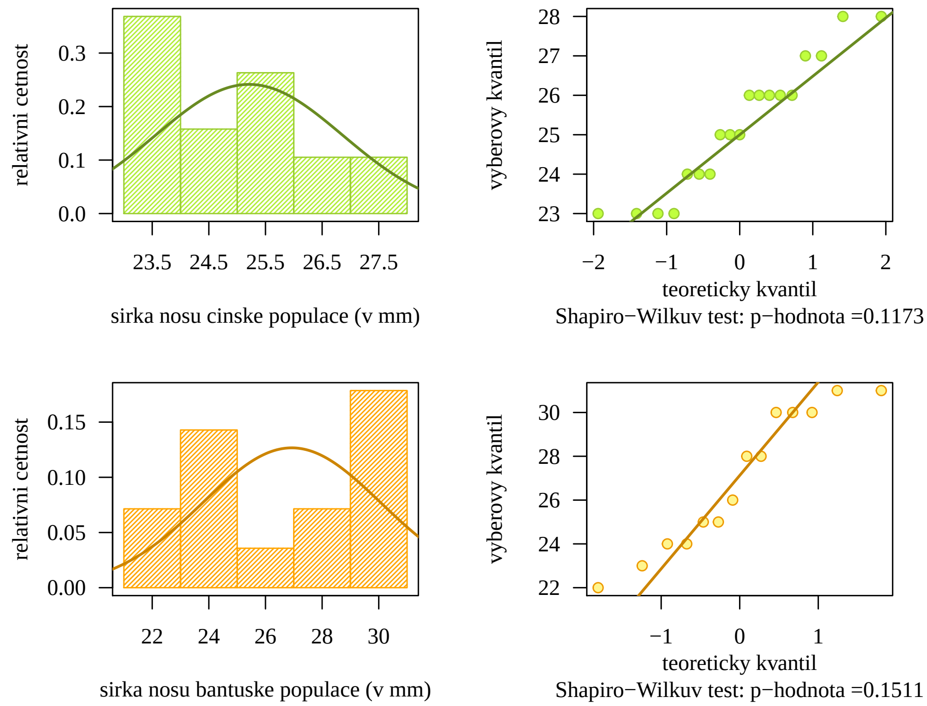
```
[1] 14
```
29

Protože náhodný výběr rozdílu naměřených šířek nosu mužů bantuské populace má rozsah ...., což je .... než 30, použijeme na testování hypotézy o normalitě dat .... test.

```
[1] 0.1511379
```
30

P-hodnota vyšla .... Protože p-hodnota .... α, H_0 .... na hladině významnosti $\alpha = 0.05$.

Interpretace výsledků: Naměřené hodnoty šířky nosu mužů čínské i bantuské populace .... z normálního rozdělení.



Test o podílu rozptylů

První předpoklad o normalitě obou výběrů je / není splněn. Nyní ještě musíme ověřit předpoklad o shodě rozptylů obou náhodných výběrů. Nejprve stanovíme nulovou a alternativní hypotézu.

- H_0 : ....
- H_1 : .... (... alternativa).
- Hladina významnosti $\alpha =$ ....
- Test o ....

a) Testování pomocí funkce var.test()

```

        F test to compare two variances

data:  nose.Bc and nose.Bb
F = 0.27537, num df = 18, denom df = 13, p-value = 0.01258
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.09230537 0.75180759
sample estimates:
ratio of variances
         0.2753689
```

```
[1] 0.3662758
```

```
[1] 2.983239
```

Hodnota testovací statistiky t_0 je .... Kritický obor má tvar .... Protože ...., H_0................................ na hladině významnosti $\alpha =$............................ .

Interval spolehlivosti má tvar .... Protože ...., H_0................................ na hladině významnosti $\alpha =$............................ .

P-hodnota vyšla .... Protože p-hodnota ...., H_0 .... na hladině významnosti $\alpha =$............................ .

Interpretace výsledků: Mezi rozptylem šířky nosu mužů čínské populace a rozptylem šířky nosu mužů bantuské populace existuje / neexistuje statisticky významný rozdíl.

Test o rozdílu středních hodnot
Náhodné výběry pochází z normálních rozdělení s vzájemně shodnými / odlišnými rozptyly, proto na otestování hypotézy ze zadání použijeme klasický / Welchův dvouvýběrový t-test o rozdílu středních hodnot $\mu_1 - \mu_2$, tj. parametrický test o .... středních hodnot když neznámé rozptyly σ_1^2 a σ_2^2 jsou / nejsou shodné.

a) Testování pomocí funkce t.test()

```

        Welch Two Sample t-test

data:  nose.Bc and nose.Bb
t = -1.8611, df = 18.268, p-value = 0.9606
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -3.317511       Inf
sample estimates:
mean of x mean of y
 25.21053  26.92857
```

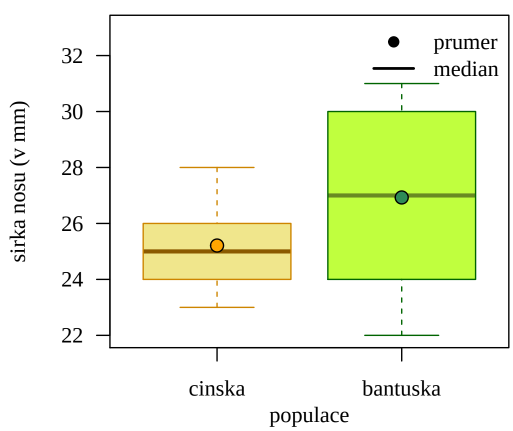
```
[1] 1.732686
```

Hodnota testovací statistiky t_0 je .... Kritický obor má tvar .... Protože ...., H_0................................ na hladině významnosti $\alpha =$............................ .

Interval spolehlivosti má tvar .... Protože ...., H_0................................ na hladině významnosti $\alpha =$............................ .

P-hodnota vyšla .... Protože p-hodnota ...., H_0 .... na hladině významnosti $\alpha =$............................ .

Grafická vizualizace výsledků testování



Interpretace výsledků: Na základě všech tří způsobů testování .... hypotézu o tom, že střední hodnota šířky nosu čínské populace je menší nebo rovna střední hodnotě šířky nosu bantuské populace. Máme / Nemáme tedy dostatek indicií k zamítnutí tvrzení, že šířka nosu čínské populace je statisticky významně menší nebo rovná šířce nosu bantuské populace.