Modelovanie trendu: exponenciálne zhladzovanie, Holt-Wintersova metóda, Hodrick-Prescottov filter

Beáta Stehlíková

2-EFM-102 Časové rady

Fakulta matematiky, fyziky a informatiky, UK v Bratislave

Obsah

- **Exponenciálne zhladzovanie** - nájdenie strednej hodnoty v časovom rade bez tredu a sezónnosti
- **Holt - Wintersova metóda** - zovšeobecnenie pre dáta s trendom a sezónnosťou
- **Hodrick - Prescottov filter** - vyhladenie dát bez sezónnosti pomocou dvoch kritérií: zhoda s dátami a malá krivosť krivky

Exponenciálne zhladzovanie

Označenie

- Máme dáta $x_1, x_2, \dots, x_n$ a chceme predikovať hodnotu x_{n+k}
- V tejto časti predpokladáme, že v dátach **nie je ani trend, ani sezónnnosť**
- Model:

$$x_t = \mu_t + w_t,$$

  kde

  - μ_t je stredná hodnota (môže závisieť od času)
  - w_t sú nezávislé náhodné odchýlky s nulovou strednou hodnotou

- Označme a_t náš odhad strednej hodnoty μ_t

Model

- Základná myšlienka exponenciálneho zhladzovania: ďalší odhad strednej hodnoty (teda a_t) bude váženým priemerom predchádzajúceho odhadu (teda a_{t-1}) a novej realizovanej hodnoty x_t:

$$a_t = \alpha x_t + (1 - \alpha) a_{t-1}$$

- Parameter zhladzovania α:
  - $\alpha \approx 1$ - slabé zhladzovanie, $a_t \approx x_t$
  - $\alpha \approx 0$ - silné zhladzovanie, $a_t \approx a_{t-1}$
- Iný zápis a_t:

$$a_t = \alpha x_t + \alpha(1 - \alpha) x_{t-1} + \alpha(1 - \alpha)^2 x_{t-2} + \dots$$

- váhy exponenciálne klesajú, preto názov exponenciálne zhladzovanie

Predikcie a optimálna α

- Označenie: $\hat{x}_{n+k|n}$ - predikcia dát x na čas $n+k$, ak je dnešný čas n
- Keďže nemáme trend ani sezónnosť, vieme spraviť iba

$$\hat{x}_{n+k|n} = a_n$$

- Pre daný parameter α:
  - $a_1 = x_1$ a potom rekurentne
  - máme teda predikčné chyby:

$$e_t = x_t - \hat{x}_{t|t-1} = x_t - a_{t-1}$$

- Optimálny parameter α - minimalizujeme sumu štvorcov predikčných chýb:

$$\sum_{t=2}^{n} e_t^2 \to \min$$

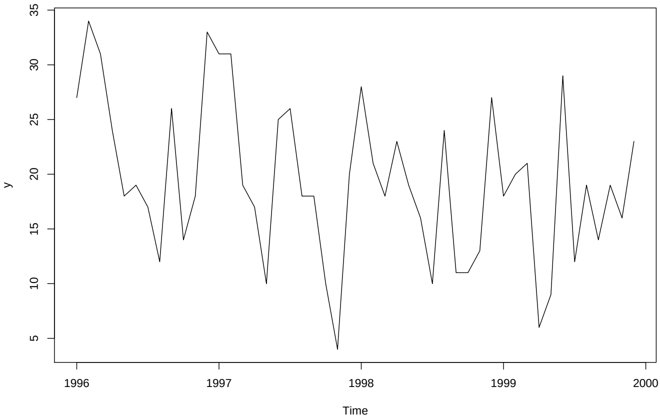
Exponenciálne zhladzovanie v R-ku

Dáta

- *P. S. P.Cowpertwait, A. V. Metcalfe: Introductory Time Series with R. Springer, 2009. Complaints to a motoring organization, pp. 56-58.*
- počet sťažností, mesačné dáta, 1996/01 - 1999/12
- dáta na stránke (`complaints.txt`)

```
y <- read.table("complaints.txt")
y <- ts(y$V1, frequency = 12, start = c(1996, 1))
```

```r
plot(y)
```



Odhadnutie modelu

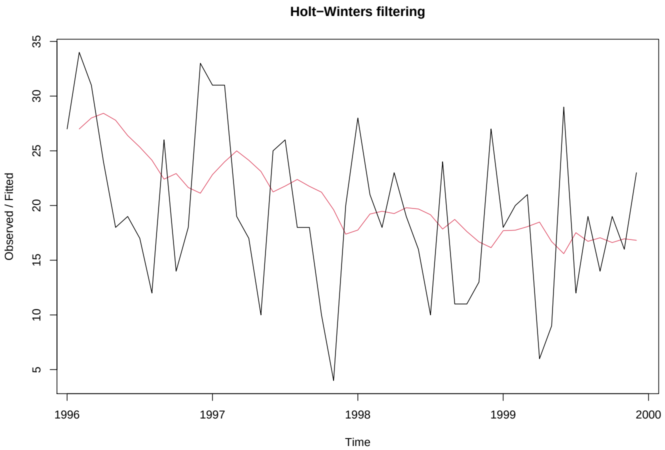
- funkcia `HoltWinters` s nastavením `beta = FALSE` a `gamma = FALSE` - je špeciálny to prípad všeobecnejšieho modelu, pre ktorý máme funkciu `HoltWinters` - uvedieme neskôr

```
model1 <-  HoltWinters(y, beta = FALSE, gamma = FALSE)
model1
```

```
## Holt-Winters exponential smoothing without trend and wit
##
## Call:
## HoltWinters(x = y, beta = FALSE, gamma = FALSE)
##
## Smoothing parameters:
##  alpha: 0.1429622
##  beta : FALSE
##  gamma: FALSE
```

Graf

```
plot(model1)
```



- Prístup k hodnote *sum of squared errors*, podľa ktorej sa vyberala optimálna α:

```
model1$SSE
```

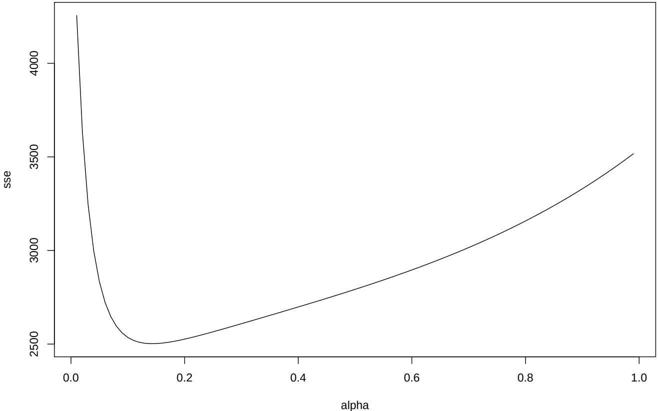
```
## [1] 2502.028
```

- Použitie našej hodnoty parametra α: napríklad

```
model_nas <- HoltWinters(y, alpha = 0.2,
                            beta = FALSE, gamma = FALSE)
```

Cvičenie: optimálna hodnota parametra α

- Vykreslíme závislosť SSE od parametra α pre naše dáta.
- Tento výpočet má potvrdiť optimálnu hodnotu α z R-ka



Holt - Wintersova metóda

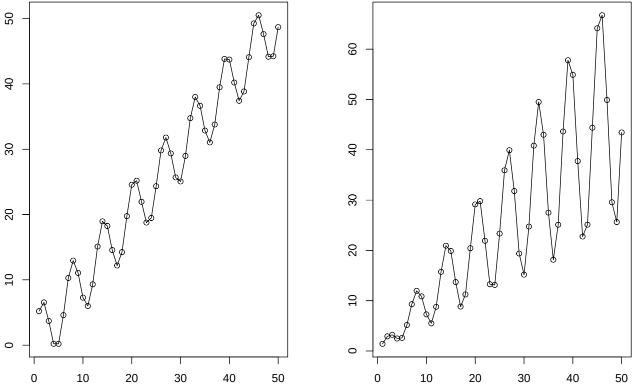
- Charakteristiky časového radu:
  - a_t = *level*, sezónne očistená stredná hodnota
  - b_t = *slope*, zmena hodnoty level z jednej periódy na druhú (zachytáva rôzne, aj krátkodobé trendy)
  - s_t = *seasonal component*, sezónna zložka (závisí napr. od mesiaca)
- Typ sezónnosti:
  - aditívna - napr. v januári je hodnota o 100 vyššia
  - multiplikatívna - napr. v januári je hodnota o 10 percent vyššia
- Predikcia pri aditívnej sezónnosti:

$$\hat{x}_{n+k|n} = a_n + kb_n + s_{n+k-p}$$

  pre $k \leq p$ (napr. $p = 12$ pri mesačných dátach)
- Pri multiplikatívnej sezónnosti:

$$\hat{x}_{n+k|n} = (a_n + kb_n)s_{n+k-p}$$

- Ukážka typického priebehu: vľavo aditívna sezónnosť, vpravo multiplikatívna:



Rekurentné vzťahy

- Analogicky ako pri exponenciálnom zhladzovaní: vážené priemery hodnôt typu *"nová hodnota"* a *"stará hodnota"*
- Pre aditívnu sezónnosť:

$$\begin{aligned} a_t &= \alpha(x_t - s_{t-p}) + (1-\alpha)(a_{t-1} + b_{t-1}) \\ b_t &= \beta(a_t - a_{t-1}) + (1-\beta)b_{t-1} \\ s_t &= \gamma(x_t - a_t) + (1-\gamma)s_{t-p} \end{aligned}$$

  kde $\alpha, \beta, \gamma \in (0, 1)$
- Analogicky pre multiplikatívnu (rovnice sú napr. aj v popise funkcie `HoltWinters`)
- Optimálne α, β, γ sa znovu určia minimalizáciou SSE

Holt-Wintersova metóda v R-ku

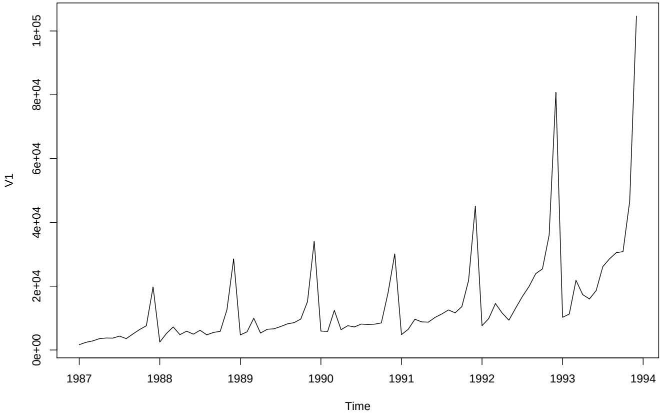
Príklad 1: predaj suvenírov

- tržby v obchode so suvenírmi v lodenici na pláži v Austrálii; mesacné dáta od januára 1987 do decembra 1993
- `suveniry.txt` na stránke

```
y <- read.table("data/suveniry.txt", header = FALSE)
y <- ts(y, frequency = 12, start = c(1987, 1))
head(y)
```

```
##            V1
## [1,] 1664.81
## [2,] 2397.53
## [3,] 2840.71
## [4,] 3547.29
## [5,] 3752.96
## [6,] 3714.74
```

```
plot(y)
```


V1
0e+00
2e+04
4e+04
6e+04
8e+04
1e+05
1987
1988
1989
1990
1991
1992
1993
1994
Time

Odhad modelu

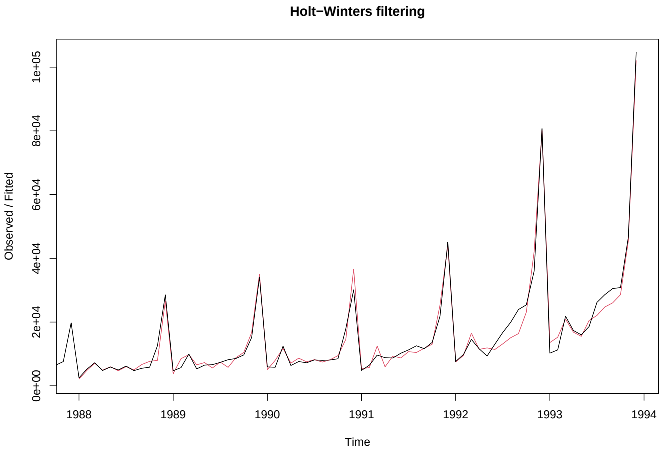
- Vidíme multiplikatívnu sezónnosť, takže:

```
HW_suveniry <- HoltWinters(y, seasonal = "multiplicative")
HW_suveniry
```

```
## Holt-Winters exponential smoothing with trend and multip
##
## Call:
## HoltWinters(x = y, seasonal = "multiplicative")
##
## Smoothing parameters:
##  alpha: 0.4889037
##  beta : 0.04653724
##  gamma: 0.947455
##
## Coefficients:
```

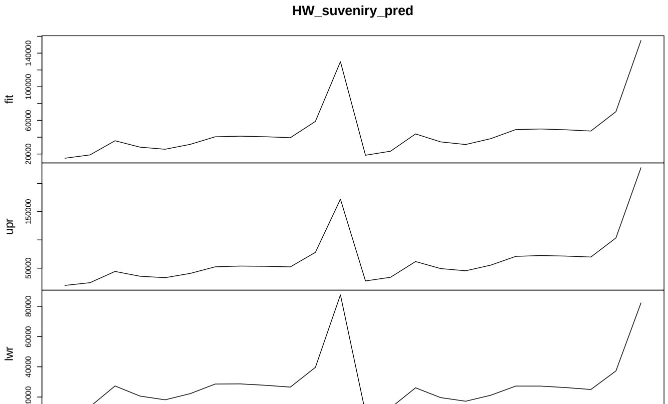
Grafické znázornenie

```
plot(HW_suveniry)
```



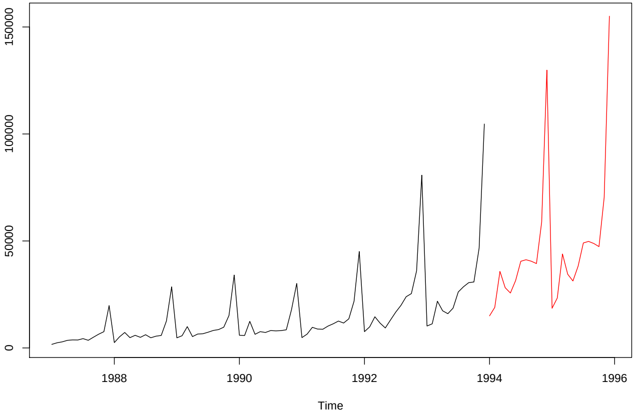
Predikcie

```
HW_suveniry_pred <- predict(HW_suveniry, n.ahead = 24,
                            prediction.interval = TRUE)
plot(HW_suveniry_pred)
```



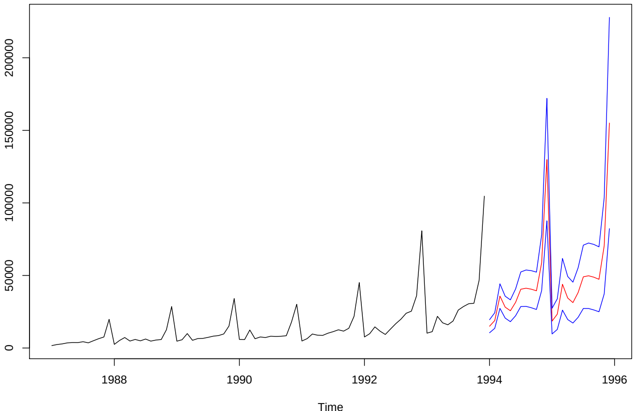
V jednom grafe spolu dáta a predikcie:

```
ts.plot(y, HW_suveniry_pred[, "fit"], col = c("black", "rec
```



V jednom grafe spolu dáta, predikcie a intervaly:

```
ts.plot(y, HW_suveniry_pred,
        col = c("black", "red", "blue", "blue"))
```

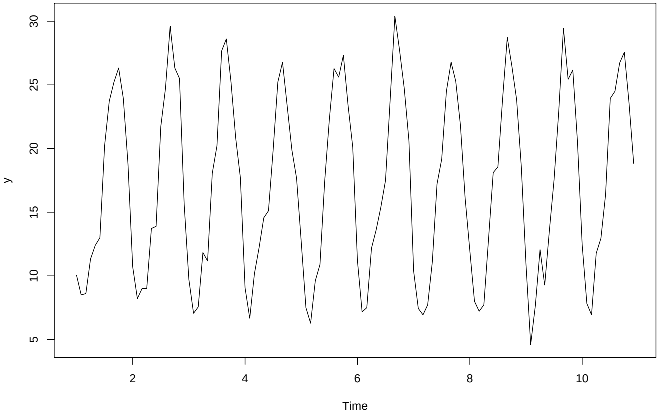


Príklad 2: teplota vzduchu (Lake Shasta)

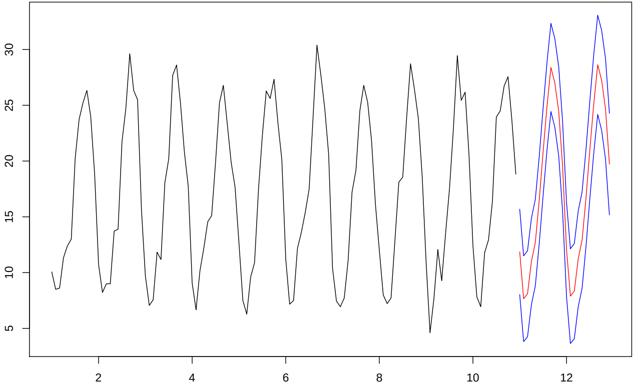
```r
library(astsa)
data("climhyd") # mesacne data
head(climhyd)
```

```
##    Temp     DewPt CldCvr   WndSpd  Precip   Inflow
## 1  5.94  1.436366   0.58 1.219485 160.528 156.1173
## 2  8.61 -0.284660   0.47 1.148620  65.786 167.7455
## 3 12.28  0.856728   0.49 1.338430  24.130 173.1567
## 4 11.61  2.696482   0.65 1.147778 178.816 273.1516
## 5 20.28  5.699536   0.33 1.256730   2.286 233.4852
## 6 23.83  8.275339   0.28 1.104325   0.508 128.4859
```

```
N <- nrow(climhyd) # pouzijeme na vyber 10 rokov
y <- ts(climhyd[(N-119):N, "Temp"], frequency = 12)
plot(y)
```



```
hw <- HoltWinters(y, seasonal = "additive")
pr <- predict(hw, n.ahead = 24, prediction.interval = TRUE)
ts.plot(y, pr[, "fit"], pr[, "lwr"], pr[, "upr"],
        col = c("black", "red", "blue","blue"))
```



Hodrick-Prescottov filter: odhadovanie trendu

Model

- Predpoklad: v dátach nie je sezónnosť
- Cieľ: chceme odhadnúť trendovú zložku dát
- Myšlienka: potrebujeme dosiahnuť dve kritériá, ktoré sú v protiklade, preto im priradíme váhy:
  - vyhladené hodnoty by mali byť blízko skutočných
  - malá krivosť grafu vyhladených hodnôt (nie veľké fluktuácie), tú vieme merať druhými diferenciami (analógia druhej derivácie)
- Optimalizačná úloha, kde $y_1, \dots, y_n$ sú dáta, $\lambda > 0$ je parameter a $\tilde{y}_1, \dots, \tilde{y}_n$ sú vyhladené hodnoty

$$\sum_{t=1}^{n}(y_t - \tilde{y}_t)^2 + \lambda \sum_{t=2}^{n-1}(\tilde{y}_{t+1} - 2\tilde{y}_t + \tilde{y}_{t-1})^2 \to \min_{\tilde{y}_1,\dots,\tilde{y}_n}$$

Hodrick-Prescottov filter v R-ku

- Balík `mFilter`
- Funkcia `hpfilter`, napr.

```
hp <- hpfilter(data, freq = 100, type = "lambda")
```

- Odhadnutý trend je potom v `hp$trend`

Príklad: vplyv parametra λ

Zoberme dáta o sťažnostiach

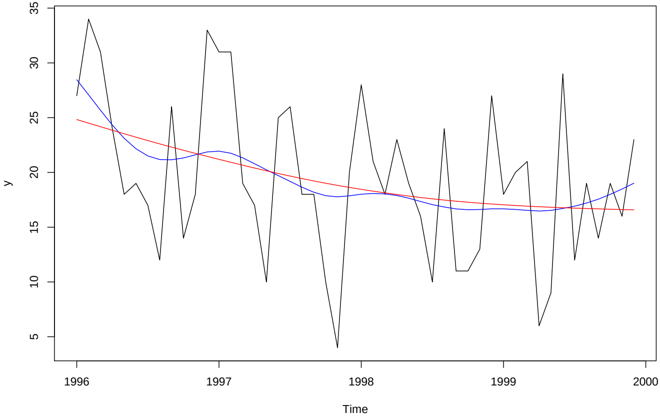
```
y <- read.table("complaints.txt")
y <- ts(y$V1, frequency = 12, start = c(1996, 1))
```

Porovnáme:

```
plot(y)

hpf1 <- hpfilter(y, freq = 500, type = "lambda")
lines(hpf1$trend, col="blue")

hpf2 <- hpfilter(y, freq = 10000, type = "lambda")
lines(hpf2$trend, col="red")
```



Vyskúšajte iné hodnoty. Čo sa deje pre $\lambda \to 0$ a pre $\lambda \to \infty$?

Odporúčané hodnoty λ

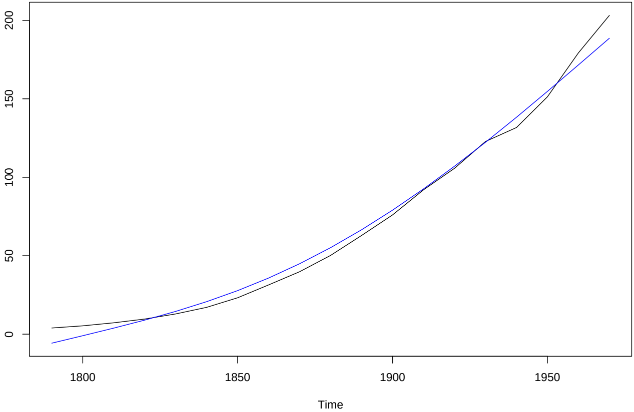
- $\lambda = 100$ pre ročné dáta
- $\lambda = 1600$ pre kvartálne dáta
- $\lambda = 14400$ pre mesačné dáta

Vyskúšajte pre naše dáta o sťažnostiach

Príklad 1: Veľkosť populácie

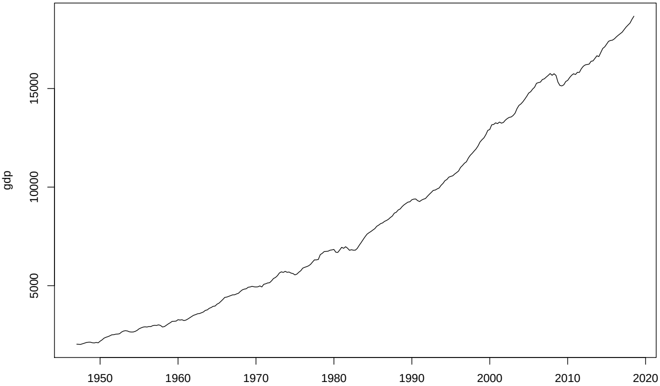
```
library(datasets); data("uspop") # rocne
plot(uspop)
```

```r
hpf <- hpfilter(uspop, freq = 100, type = "lambda")
ts.plot(uspop, hpf$trend, col = c("black", "blue"))
```

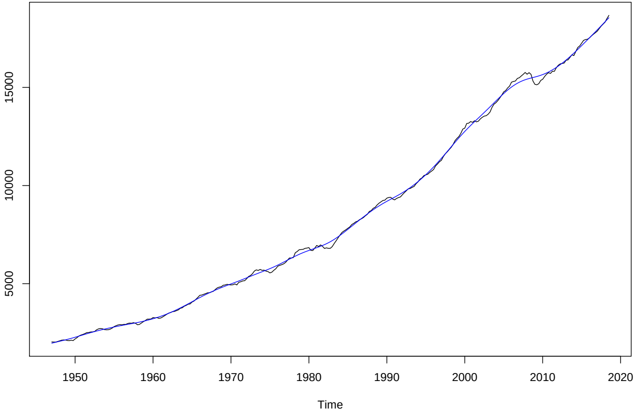


Príklad 2: HDP

```
library(astsa); data("gdp") # kvartalne, sezonne ocistene
plot(gdp)
```



```r
hpf <- hpfilter(gdp, freq = 1600, type = "lambda")
ts.plot(gdp, hpf$trend, col = c("black", "blue"))
```



```r
plot(hpf$trend - gdp, ylab = "produkcna medzera")
abline(h = 0, col = "grey", lty = 2)
```