

*Jednotkový koreň (unit root),
diferencovanie časového radu,
unit root test (ADF test)*

Beáta Stehlíková
Časové rady, FMFI UK

Obsah

- Čo je jednotkový koreň (unit root) a čo spôsobuje
- Ak má proces jednotkový koreň - ako dáta transformovať, aby sme s nimi mohli pracovať a použiť ARMA metodológiu
- Ako z dát zistiť, či má proces jednotkový koreň → unit root testy

Príklady

- Majme proces $y_t = y_{t-1} + u_t$:
 - ◇ je to nestacionárny AR(1) proces s jednotkovým koreňom
 - ◇ pre jeho diferencie $\Delta y_t = y_t - y_{t-1}$ platí $\Delta y_t = u_t$
 - ◇ teda Δy_t je stacionárny proces

- Majme nestacionárny proces s jednotkovým koreňom

$$(1 - \frac{1}{2}L)(1 - L)x_t = 1 + (1 - \frac{1}{3}L)u_t$$

Potom pre diferencie

$$\Delta y_t = y_t - y_{t-1} = (1 - L)y_t$$

platí

$$(1 - \frac{1}{2}L)\Delta y_t = 1 + (1 - \frac{1}{3}L)u_t,$$

teda Δy_t je stacionárny proces

Príklady

- Majme nestacionárny proces s dvojnásobným jednotkovým koreňom

$$(1 - \frac{1}{2}L)(1 - L)^2 x_t = 1 + (1 - \frac{1}{3}L)u_t$$

Potom pre druhé diferencie

$$\Delta^2 y_t = \Delta(\Delta y_t) = (1 - L)(1 - L)y_t = (1 - L)^2 y_t$$

platí

$$(1 - \frac{1}{2}L)\Delta^2 y_t = 1 + (1 - \frac{1}{3}L)u_t,$$

teda $\Delta^2 y_t$ je stacionárny proces

- Vo všeobecnosti:
Ak má proces jednotkový koreň násobnosti k (a ostatné mimo jednotkového kruhu), tak jeho k -te diferencie sú stacionárne

ARIMA models

- Ak treba proces k -krát diferencovať, aby sme dostali stacionárny proces, nazýva sa **integrovaný proces rádu k** , označujeme $I(k)$
- Ak tie k -te diferencie sú ARMA(p,q), tak o pôvodnom procese hovoríme, že je **ARIMA(p,k,q)**.
- Napríklad x_t , ak

$$\left(1 - \frac{1}{2}L\right)(1 - L)^2 x_t = 1 + \left(1 - \frac{1}{3}L\right)u_t,$$

je proces ARIMA(1,2,1).

Ciel'

- Uvažujme najprv AR(1) model: $x_t = \delta + \rho x_{t-1} + u_t$
- Chceme:
 - ◇ testovať hypotézu o jednotkovom koreni (vtedy je proces nestacionárny), teda $H_0 : \rho = 1$
 - ◇ zistiť, či sa dá zamietnuť v prospech stacionarity - $H_1 : \rho < 1$

Jednotkový koreň a t-štatistika

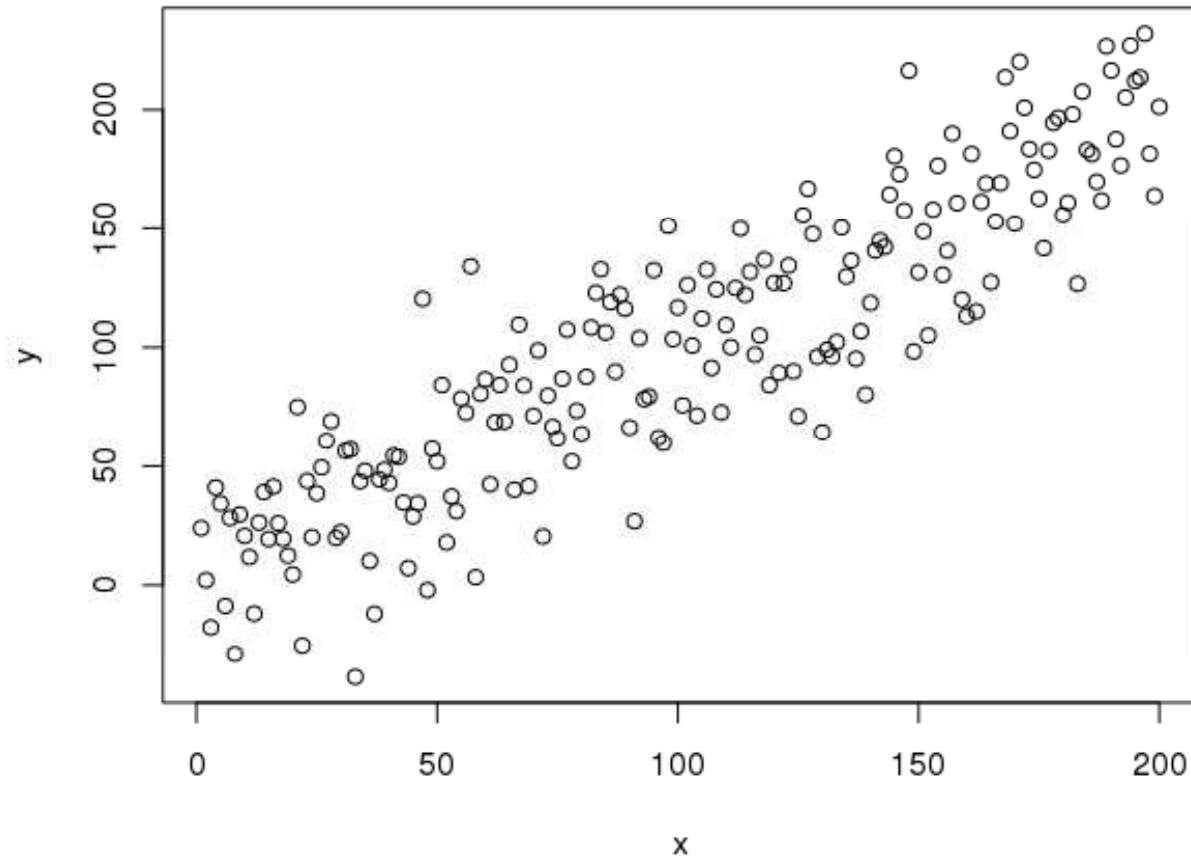
Skúsme použiť testovanie hypotéz o koeficientoch regresného modelu známe z ekonometrie.

Napríklad:

- Majme vektor $x=1:200$
- Vygenerujeme $y=x+rnorm(200)*sigma$
- Odhadneme model $y = c + \rho x + \varepsilon$
- Zaznamenávame:
 - ◇ odhad parametra ρ
 - ◇ hodnotu t-štatistiky zodpovedajúcej hypotéze $H_0 : \rho = 1$ (ktorá platí)
- Zopakujeme 10^5 krát a vykreslíme histogram

Jednotkový koreň a t -štatistika

- Príklad vygenerovaných dát:



Jednotkový koreň a t-štatistika

- Odhadnutá regresia:

```
Call:
lm(formula = y ~ x)

Residuals:
    Min       1Q   Median       3Q      Max
-75.067 -19.239   3.818  20.520  74.772

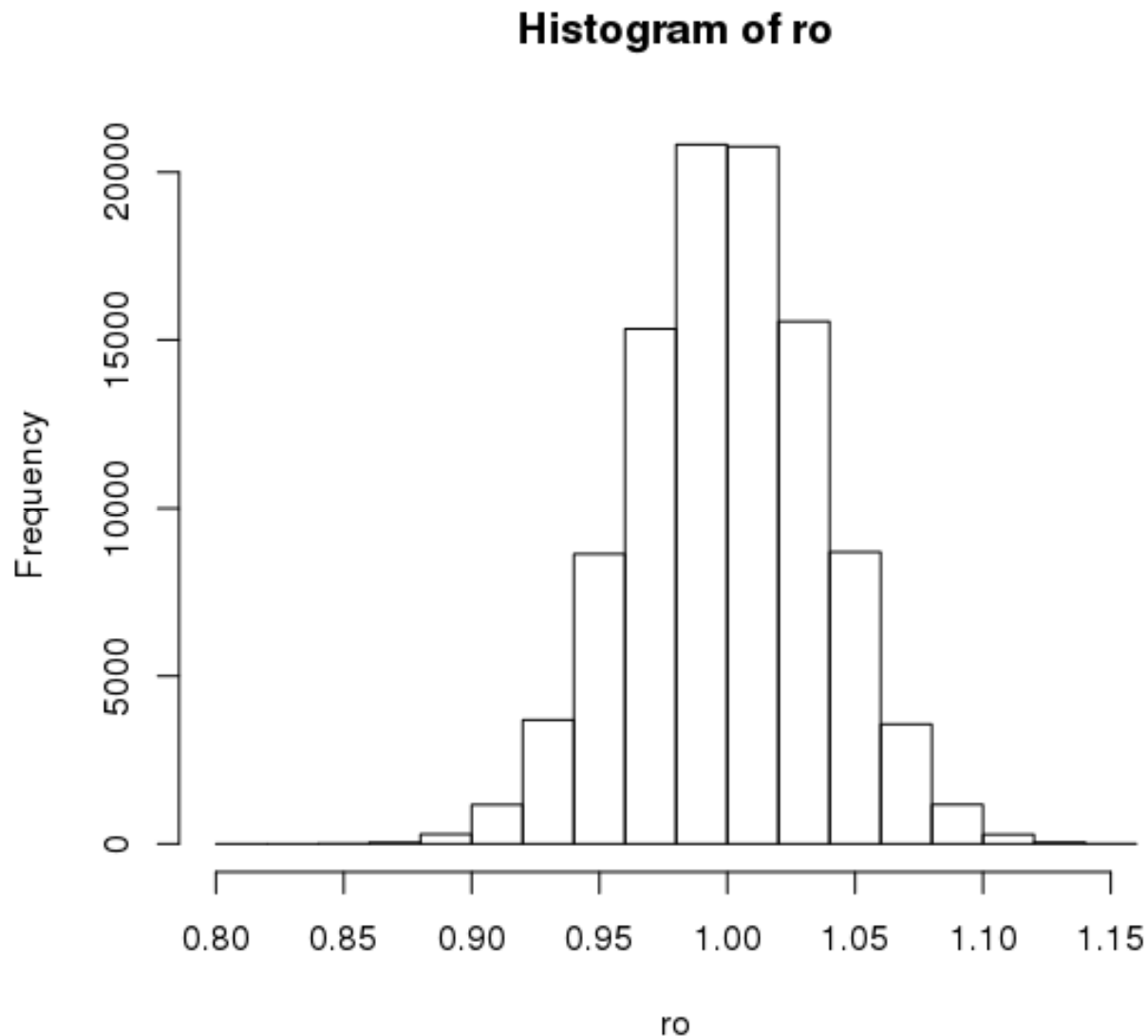
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.72079    4.00941   1.177   0.24
x            0.95689    0.03459  27.662 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 28.24 on 198 degrees of freedom
Multiple R-squared:  0.7944,    Adjusted R-squared:  0.7934
F-statistic: 765.2 on 1 and 198 DF,  p-value: < 2.2e-16
```

- Odhadnutý koeficient ρ je 0.95689
- T-štatistika k hypotéze $\rho = 1$ je $\frac{0.95689-1}{0.03459}$

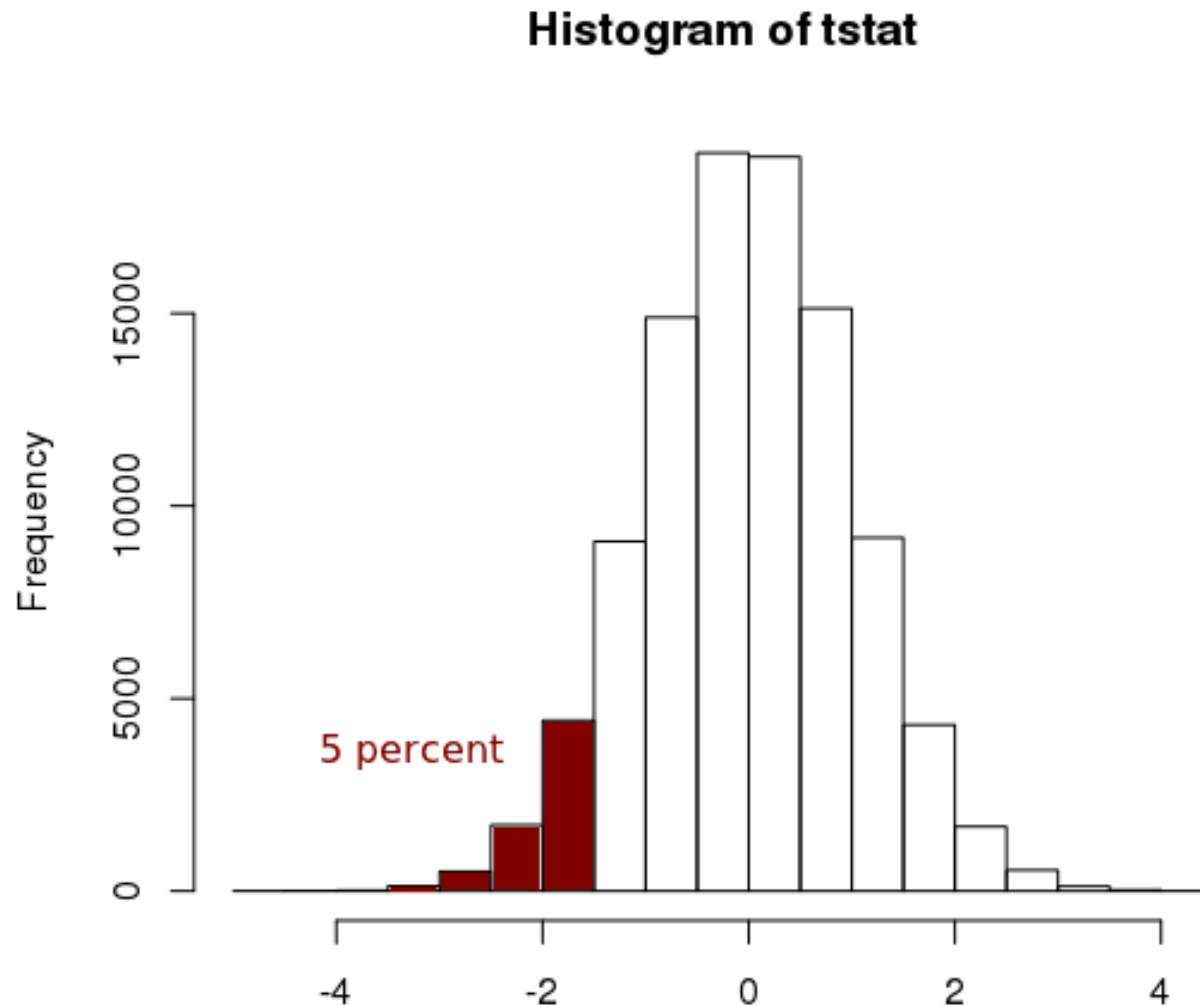
Jednotkový koreň a t -štatistika

- Odhad parametra ρ : normálne rozdelenie



Jednotkový koreň a t-štatistika

- t-štatistika k hypotéze $H_0 : \rho = 1$: t-rozdelenie



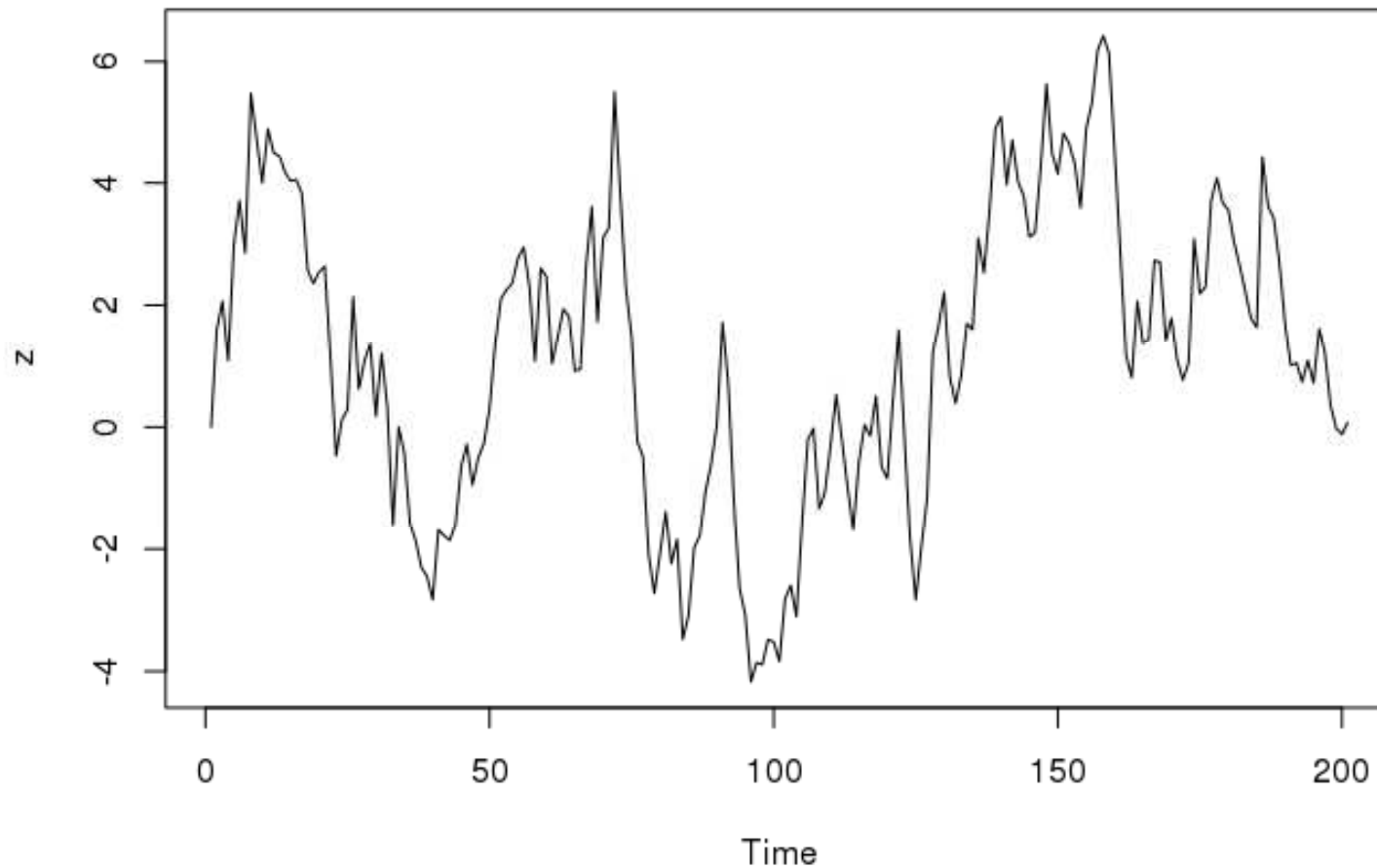
Jednotkový koreň a t-štatistika

Druhá simulácia:

- Majme vektor z vygenerovaný ako $z_t = z_{t-1} + \varepsilon_t$
- Zoberieme $x=z[1:200]$, $y=z[2:201]$, teda $y_t = z_t$,
 $x_t = z_{t-1}$
- Odhadneme model $y = c + \rho x + \varepsilon$
- Zaznamenávame:
 - ◇ odhad parametra ρ
 - ◇ hodnotu t-štatistiky zodpovedajúcej hypotéze
 $H_0 : \rho = 1$ (ktorá platí)
- Zopakujeme 10^5 krát a vykreslíme histogram

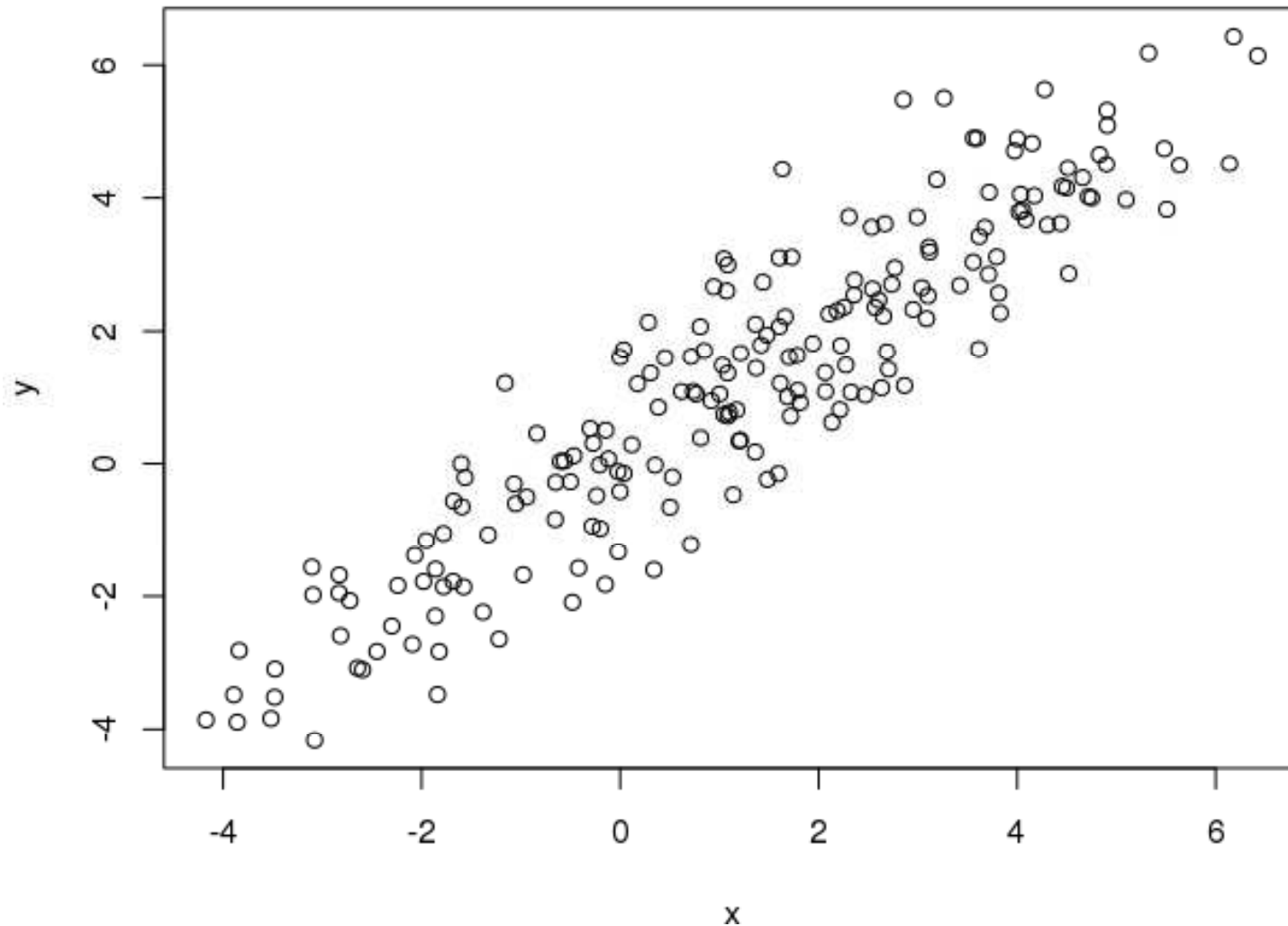
Jednotkový koreň a t -štatistika

- Príklad vygenerovaných dát - časový rad z :



Jednotkový koreň a t -štatistika

- Príklad vygenerovaných dát : dáta do regresie



Jednotkový koreň a t-štatistika

- Odhadnutá regresia:

```
Call:
lm(formula = y ~ x)

Residuals:
    Min       1Q   Median       3Q      Max
-2.01047 -0.64723  0.00159  0.53995  2.84017

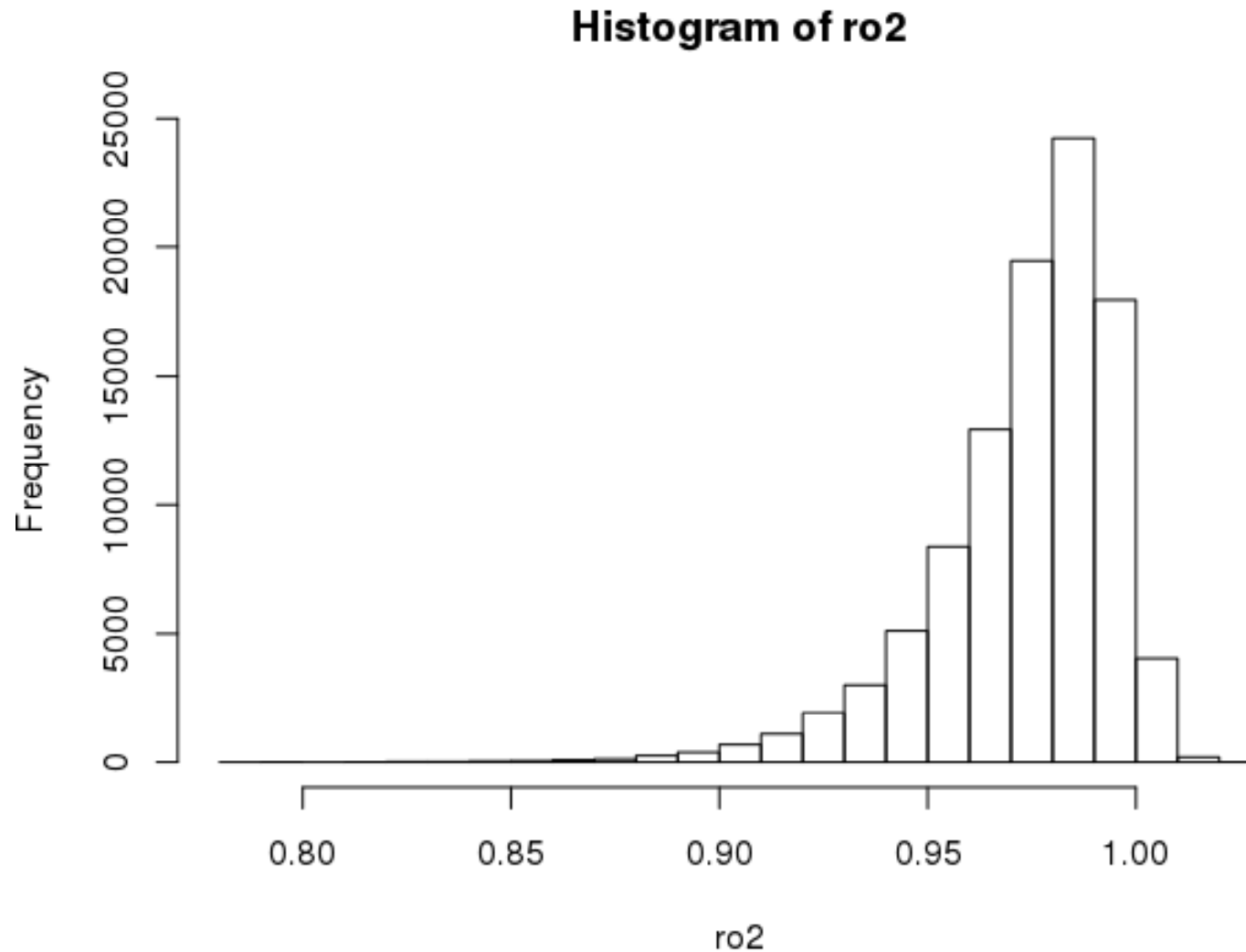
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.09589   0.07489    1.28   0.202
x            0.91962   0.02790   32.96 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9496 on 198 degrees of freedom
Multiple R-squared:  0.8458,    Adjusted R-squared:  0.845
F-statistic: 1086 on 1 and 198 DF,  p-value: < 2.2e-16
```

- Odhadnutý koeficient ρ je **0.91962**
- t-štatistika k hypotéze $\rho = 1$ je $\frac{0.91962-1}{0.02790} = -2.88$

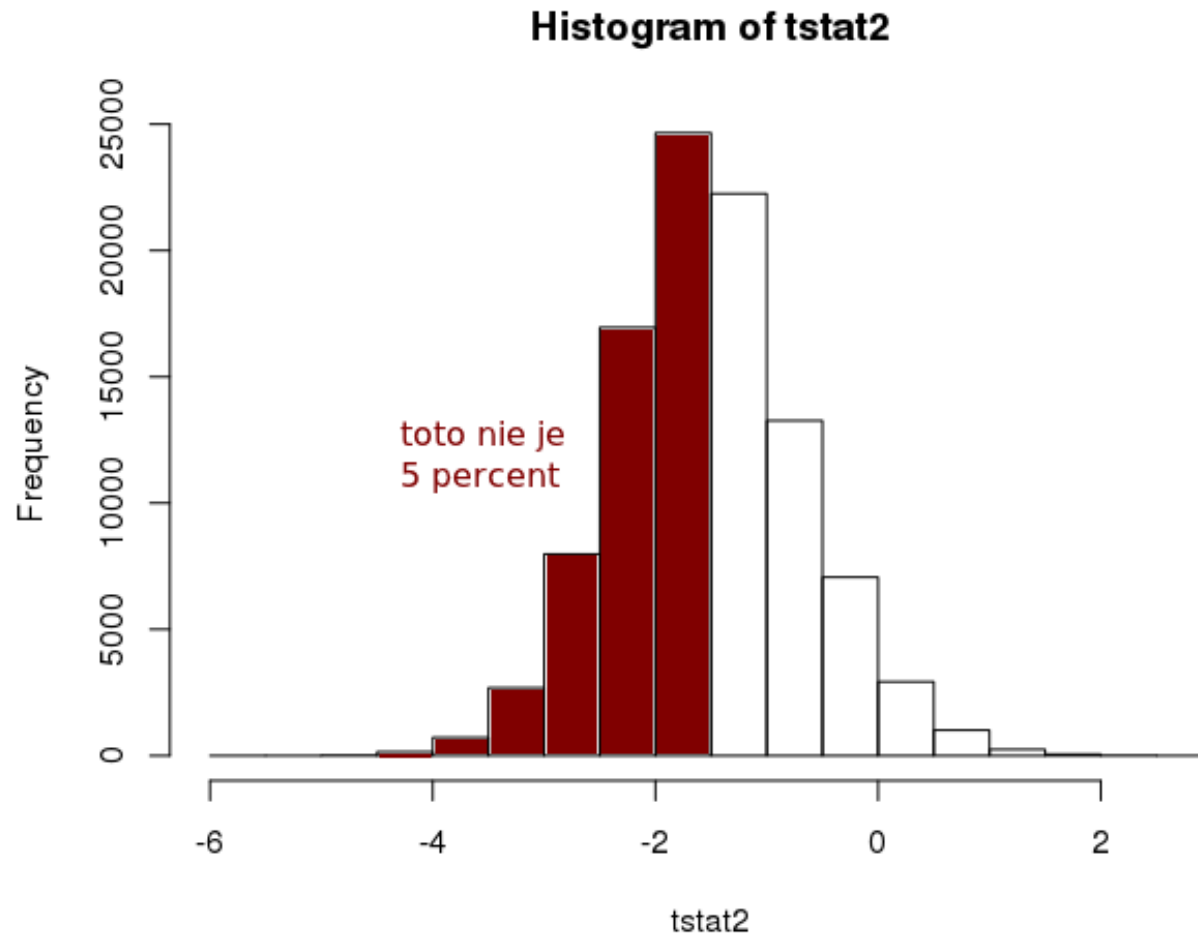
Jednotkový koreň a t -štatistika

- Odhady parametra ρ : nemá normálne rozdelenie



Jednotkový koreň a t-štatistika

- "t-štatistika" k hypotéze $H_0 : \rho = 1$: nemá t-rozdelenie



- Na testovanie hypotézy **nemôžeme použiť kritické hodnoty t-rozdelenia**

Jednotkový koreň a t-štatistika

- Riešenie - základná myšlienka:
 - ◇ ponecháme výpočet testovacej štatistiky
 - ◇ ale budeme používať iné kritické hodnoty
- Kvantily z našej simulácie:

```
> quantile(tstat2,0.05)
 5%
-2.866481
> quantile(tstat2,0.01)
 1%
-3.462751
```

- približne také by mali byť kritické hodnoty

- Otázka:
 - ◇ Čo ak nemáme AR(1) proces, ale všeobecnejší?

Testovanie jednotkového koreňa

- AR(1) proces:

$$(1) \quad y_t = \rho y_{t-1} + u_t$$

jednotkový koreň znamená, že $\rho = 1$.

- Ekvivalentne:

$$\Delta y_t = (\rho - 1)y_{t-1} + u_t$$

a zaujíma nás t-štatistika zo signifikancie koeficienta pri y_{t-1} - ale s inou kritickou hodnotou

- Tá kritická hodnota
 - ◇ závisí od počtu dát
 - ◇ zmení sa, ak rovnica (1) obsahuje konštantu a/alebo lineárny trend
- Vo všeobecnosti: $\Delta y_t = \alpha + \beta t + (\rho - 1)y_{t-1} + u_t$

Testovanie jednotkového koreňa

- AR(p) proces: $y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + u_t$
jednotkový koreň $\rightarrow \alpha_1 + \dots + \alpha_p = 1$.

- Upravíme do tvaru:

$$y_t = \rho y_{t-1} + \theta_1 \Delta y_{t-1} + \theta_2 \Delta y_{t-2} + \dots + \theta_{p-1} \Delta y_{t-p+1} + u_t,$$

kde $\rho = \sum_{j=1}^p \alpha_j$, $\theta_i = -\sum_{j=i+1}^p \alpha_j$ pre $i = 1, \dots, p-1$

- Ekvivalentne:

$$\Delta y_t = (\rho - 1) y_{t-1} + \theta_1 \Delta y_{t-1} + \theta_2 \Delta y_{t-2} + \dots + \theta_{p-1} \Delta y_{t-p+1} + u_t$$

a zaujíma nás **t-štatistika z koeficienta pri y_{t-1}**

- Vo všeobecnosti: y môže mať trend a/alebo intercept
 $\Rightarrow \Delta y_t = \alpha + \beta t + (\rho - 1) y_{t-1} + \theta_1 \Delta y_{t-1} + \theta_2 \Delta y_{t-2} + \dots$
 $+ \theta_{p-1} \Delta y_{t-p+1} + u_t$

Augmented Dickey-Fuller test (ADF)

Wayne A. Fuller (1976)

David A. Dickey, Wayne A. Fuller (1979, 1981)

- Odhadujeme rovnicu

$$\Delta y_t = \alpha + \beta t + (\rho - 1)y_{t-1} + \theta_1 \Delta y_{t-1} + \dots + \theta_k \Delta y_{t-k} + u_t$$

pričom musíme

- ◇ rozhodnúť, či zahrnúť konštantu α a/alebo lineárny trend β (podľa toho, či ich obsahuje proces y)
 - ◇ určiť k
- Zaujímá nás potom t-štatistika zo signifikancie koeficienta pri y_{t-1} , ale so správnymi kritickými hodnotami

ADF test - kritické hodnoty

- James G. MacKinnon (1991) - dostupné ako súčasť doplnenej verzie z roku 2010:

James G. MacKinnon: **Critical Values for Cointegration Tests**. Queen's Economics Department Working Paper No. 1227, 2010..

Dostupné online: <http://ideas.repec.org/p/qed/wpaper/1227.html>

- Simulačne získané hodnoty:

Table 1. Response Surface Estimates of Critical Values

N	Variant	Level	Obs.	β_∞	(s.e.)	β_1	β_2
1	no constant	1%	600	-2.5658	(0.0023)	-1.960	-10.04
		5%	600	-1.9393	(0.0008)	-0.398	
		10%	560	-1.6156	(0.0007)	-0.181	
1	no trend	1%	600	-3.4336	(0.0024)	-5.999	-29.25
		5%	600	-2.8621	(0.0011)	-2.738	-8.36
		10%	600	-2.5671	(0.0009)	-1.438	-4.48
1	with trend	1%	600	-3.9638	(0.0019)	-8.353	-47.44
		5%	600	-3.4126	(0.0012)	-4.039	-17.83
		10%	600	-3.1279	(0.0009)	-2.418	-7.58

ADF test - kritické hodnoty

- Ak v regresii používame T dát, kritická hodnota je $\beta_\infty + \beta_1/T + \beta_2/T^2$
- V našom príklade zo simulácií: konštanta bez trendu, $T = 200$:

N	Variant	Level	Obs.	β_∞	(s.e.)	β_1	β_2
1	no trend	1%	600	-3.4336	(0.0024)	-5.999	-29.25
		5%	600	-2.8621	(0.0011)	-2.738	-8.36
		10%	600	-2.5671	(0.0009)	-1.438	-4.48

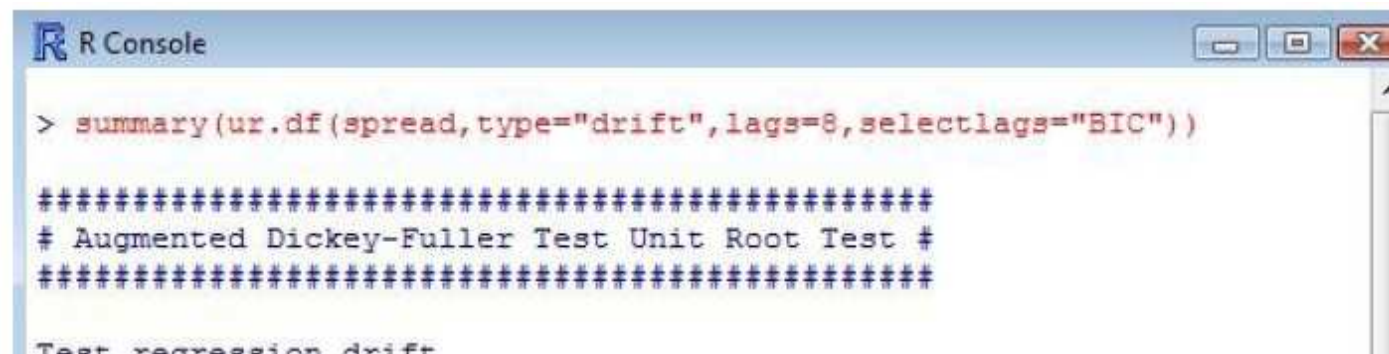
- ◇ pre 1 percento:
 $-3.4336 - 5.999/200 - 29.25/200^2 = -3.451$
- ◇ pre 5 percent:
 $-2.8621 - 2.738/200 - 8.36/200^2 = -2.879$
- Porovnajme s t-rozdelením (úplne iné) a s kvantilmi zo simulácií (ok)

ADF test v R-ku

- Knižnica `urca`
- Funkcia `ur.df` (ur - unit root, df - Dickey-Fuller) s parametrami:
 - ◇ `type`: možnosti sú `drift` (konštanta bez lineárneho trendu), `trend` (konštanta aj lineárny trend), `none` (nič)
 - ◇ `lags`: maximálny počet lagov
 - ◇ `selectlags`: kritérium, podľa ktorého sa vyberá počet lagov (informačné kritériá: AIC, BIC)

ADF test v R-ku

- Príklad: **spread** z predchádzajúcich prednášok (rozdiel dlhodobej a krátkodobej úrokovej miery)
- Príkaz:
`summary(ur.df(spread,type="drift",lags=8,selectlags="BIC"))`
- `summary` preto, aby sme dostali aj kritické hodnoty, nielen testovaciu štatistiku



```
R Console
> summary(ur.df(spread,type="drift",lags=8,selectlags="BIC"))

#####
# Augmented Dickey-Fuller Test Unit Root Test #
#####

Test regression drift
```

ADF test v R-ku

- Výstup: odhadnutá regresia a z nej získaná hodnota testovacej štatistiky:

```
Call:
lm(formula = z.diff ~ z.lag.1 + 1 + z.diff.lag)

Residuals:
    Min       1Q   Median       3Q      Max
-2.62148 -0.37475 -0.01138  0.35785  2.57280

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.10035    0.05592   1.794 0.074240 .
z.lag.1      -0.10720    0.02741  -3.911 0.000125 ***
z.diff.lag   0.29007    0.06706   4.326 2.38e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.7006 on 204 degrees of freedom
Multiple R-squared:  0.1216,    Adjusted R-squared:  0.113
F-statistic: 14.12 on 2 and 204 DF,  p-value: 1.806e-06

Value of test-statistic is: -3.9112 7.6595
```

ADF test v R-ku

- Výstup: hodnota testovacej štatistiky a kritické hodnoty:

```
Value of test-statistic is: -3.9112 7.6595
Critical values for test statistics:
      1pct  5pct 10pct
tau2 -3.46 -2.88 -2.57
```

- Kritérium: Hypotézu o jednotkovom koreni zamietame, ak je štatistika menšia ako kritická hodnota.