

# Գծային ռեգրեսիա

1-ին դաս՝ նվազագույն քառակուսիների մեթոդ

Գալթոնի օրինակը

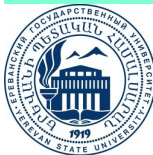
R ծրագիր

Փիրսոնի փվյալները

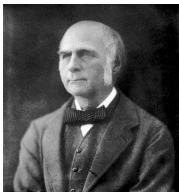
Սեդիկի գուրշու  
ճնշումը

Դյուարնի բանաձևը

Ավելորդ  
փոփոխականներ



- եզրույթը ներմուծել է Ֆրանսիս Գալթոնը, 19-րդ դարի բրիտանացի գիտնական, որը զբաղվել է ժառանգականության ուսումնասիրությամբ:



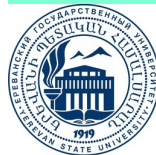
Մըր Ֆրանսիս Գալթոն 1822-1911

- Նոդվածը՝

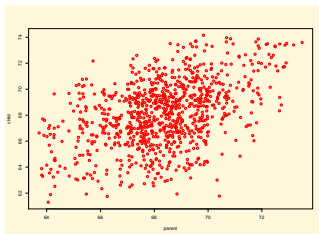
*Regression towards mediocrity in hereditary stature*  
*Journal of the Anthropological Institute 15 : 246-63 (1886)*

- Ցանկացել է նկարագրել հեղինակ կենսաբանական երևույթը: չափից դուրս բարձրահասակ ծնողների երեխաների հասակը ավելի մոտ է միջին հասակին:
- Գալթոնը վերլուծել է 928 ավյալներ, որոնցից յուրաքանչյուրը (*ծնողների միջին հասակ, երեխայի հասակ*) գույգի արեքով է: (չափված դյույմերով՝ 1 դյույմ = 2,54 ցմ) (Կանանց հասակները բազմապարկված են 1.08-ով:)

## Գալթոնի օրինակը (շարունակություն)



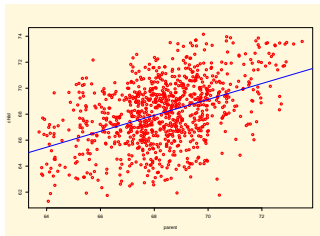
- Մտորն ներկայացված է կեփային ամպը,



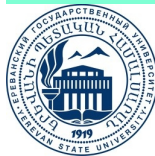
## Գալթոնի օրինակը (շարունակություն)



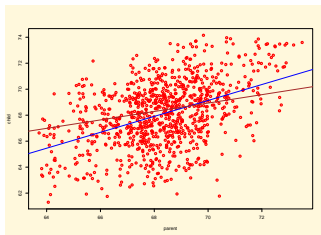
- Մտորեն ներկայացված է կեփային ամպը, ռեգրեսիայի ուղիղը *երեխայի հասակը* *ծնողի հասակի* նկատմամբ



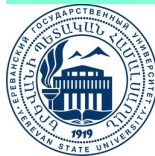
## Գալթոնի օրինակը (շարունակություն)



- Մտորն ներկայացված է կերպային ամպը, ռեգրեսիայի ուղիղը *երեխայի հասակը ծնողի հասակի* նկարմամբ ինչպես նաև ռեգրեսիայի ուղիղը *ծնողի հասակը երեխայի հասակի* նկարմամբ:



## Գալթոնի օրինակը (շարունակություն)



### Գալթոնի օրինակը

R ծրագիր

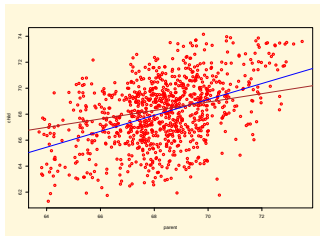
Փիրսոնի փոխանցում

Սեդիկի գոտրշու  
ճնշումը

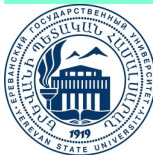
Դյուպրեի բանձնը

Ավելորդ  
փոփոխականներ

- Մտորել ներկայացված է կեփային ամպը, ռեգրեսիայի ուղիղը *երեխայի հասակը ծնողի հասակի* նկատմամբ ինչպես նաև ռեգրեսիայի ուղիղը *ծնողի հասակը երեխայի հասակի* նկատմամբ:



- Նկատենք, որ՝ սպասվածի համաձայն, երկու ուղիղներն էլ ունեն թեկվածության դրական գործակից:
- Արձանագրենք նաև որ երկու ուղիղները չեն համընկնում:
- Վարժություն**՝ Ո՞ր դեպքում երկու ուղիղները կհամընկնեն:



- Ներբեռնել այսպեղից՝  
<https://cran.gis-lab.info/bin/windows/base/>

- Կիրառությունը Գալթոնի փվյալների վրա՝

```
install.packages("UsingR")  
library(UsingR)  
data(galton)  
attach(galton)  
LinReg=lm(child ~ parent)  
plot(parent, child, bg="red")  
abline(LinReg, lwd=3, col="blue")  
summary(LinReg)
```

Գալթոնի օրինակը

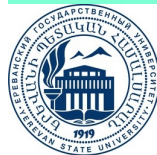
R ծրագիր

Փիրսոնի փվյալները

Սեդիկի գոլորշու  
ճնշումը

Դյուպրեի բանաձևը

Ավելորդ  
փոփոխականներ



Գալթոնի օրինակը

R ծրագիր

Փիրսոնի փվյալները

Սեդիկի գոլորշու  
ճնշումը

Դյուպրենի բանաձևը

Ավելորդ  
փոփոխականներ

## ՉՈՐՐՈՐԴ ԴԱՍ





- Գալթոնի օրինակը ավելի խորացված ձևով ուսումնասիրել է Կարլ Փիրսոնը (1857 – 1936), մաթեմատիկական վիճակագրության հիմնադիրներից մեկը: Նա վիճակագրական վերլուծության է ենթարկել ավելի մեծ քանակով տվյալներ: Այդ տվյալները գտնվում են `father.son` բազայում:

```
library(UsingR)
data(father.son)
names(father.son)
par(bg="cornsilk",pch=21)
plot(father.son,bg="red")
```

Գալթոնի օրինակը

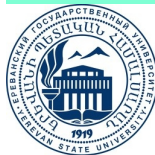
R ծրագիր

Փյունի տվյալները

Սեդիկի գոլորշու  
ճնշումը

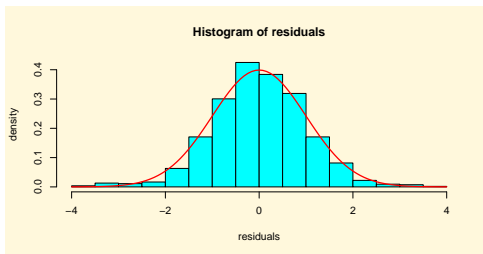
Դյուպրեի բանաձևը

Ավելորդ  
փոփոխականներ



- Ներառված է և օգտակար է նայել ռեգրեսիայի մնացքներին (սպանդարտեցնելուց հետո), հարկապես նրանց բաշխմանը:

```
llin=lm(father.son$height ~ father.son$fheight)
title = "Histogram of residuals"
hist(llin$residuals/2.437, freq=F, col="cyan",
      xlab="residuals", ylab="density", main=title)
curve(dnorm, add=T, col="red", lwd=2)
```



Գալթոնի օրինակը

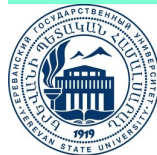
R ծրագիր

Փիրսոնի փոխանցումը

Սեդիկի գոյություն  
ճշմանը

Դյուարնի բանաձևը

Ավելորդ  
փոփոխականներ



- Մեր նպատակն է որոշել այն բանաձևը որով կարելի է հաշվել սնդիկի գոլորշու ճնշումը (չափված մմ սնդիկով)՝ ջերմաստիճանի միջոցով (չափված Ցելսիուսի աստիճանով):
- Նազեցած գոլորշու ճնշումը այն ճնշումն է, որի դեպքում նյութի (տվյալ դեպքում՝ սնդիկի) գազային վիճակը հավասարակշռության մեջ է գտնվում հեղուկ կամ պինդ վիճակի հետ:
- Այլ խոսքերով, հազեցած գոլորշու ճնշումը այն ճնշումն է, որի դեպքում տվյալ նյութը գազային վիճակից անցնում է հեղուկ վիճակի (կամ հեղուկ վիճակից գազայինի) տվյալ ջերմաստիճանի դեպքում: Նազեցած գոլորշու ճնշումը կախված է ջերմաստիճանից:

Գալթոնի օրինակը

R ծրագիր

Փիրսոնի տվյալները

Սնդիկի գոլորշու  
ճնշումը

Դյուպրեի բանաձևը

Ավելորդ  
փոփոխականներ

- Ներբեռնել և ցուցադրել տվյալները՝

```
library(datasets)
```

```
data(pressure)
```

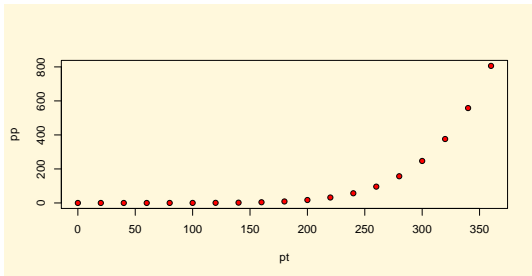
```
pp=pressure$pressure
```

```
pt=pressure$temperature
```

```
par(bg="cornsilk",pch=21)
```

```
plot(pt,pp,bg="red")
```

- Որ փարրական ֆունկցիան կարող է լավ մոտարկել այս կեփային ամպը:





- Գտնել  $\beta_0$  և  $\beta_1$  գործակիցները՝ իրականացնելով գծային ռեգրեսիա, այնպիսին որ

$$\log(\text{գոլորշու ճնշում}) = \beta_0 + \beta_1 \cdot \text{ջերմաստիճան} + \varepsilon.$$

- Կարելի է՝ արդյոք վստահել ստացված մոդելին:
- Օգտագործելով վերոնշյալ մոդելը՝ կանխատեսել սնդիկի գոլորշու ճնշումը  $T = 90; 230; 400$  ջերմաստիճանների դեպքում:

```
new=data.frame(pt=c(90,230,400))
linReg = lm(log(pp) ~ pt)
Pred=predict(linReg, new, interval="confidence")
Pred=exp(Pred)
```

Գալթոնի օրինակը

R ծրագիր

Փիրսոնի փվյակները

Սնդիկի գոլորշու ճնշումը

Դյուարեի բանաձևը

Ավելորդ փոփոխականներ



Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-6.068144	0.483831	-12.54	5.10e-10	***
pt	0.039792	0.002296	17.33	3.07e-12	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.096 on 17 degrees of freedom

Multiple R-squared: 0.9464, Adjusted R-squared: 0.9433

F-statistic: 300.3 on 1 and 17 DF, p-value: 3.07e-12

Գալթոնի օրինակը

R ծրագիր

Փիրսոնի փվյակները

Մեդիկի գոլորշու  
ճնշումը

Դյուարիի բանձնը

Ավելորդ  
փոփոխականներ



Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-6.068144	0.483831	-12.54	5.10e-10	***
pt	0.039792	0.002296	17.33	3.07e-12	***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.096 on 17 degrees of freedom  
Multiple R-squared: 0.9464, Adjusted R-squared: 0.9433  
F-statistic: 300.3 on 1 and 17 DF, p-value: 3.07e-12

Շրջանակի մեջ վերցված թվերը համապատասխանում են  $\beta_0$  և  $\beta_1$  պարամետրերի գնահատականներին:

Գալթոնի օրինակը

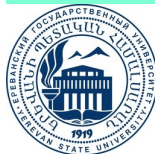
R ծրագիր

Փիրսոնի վավայները

Սնդիկի գոյորշու ճնշումը

Դյուարեի բանձնը

Ավելորդ փոփոխականներ



Coefficients:

	Estimate	Std. Error	t value	Pr(> t )							
(Intercept)	-6.068144	0.483831	-12.54	5.10e-10	***						
pt	0.039792	0.002296	17.33	3.07e-12	***						
---											
Signif. codes:	0	'***'	0.001	'**'	0.01	'*'	0.05	'.'	0.1	' '	1

**Residual standard error: 1.096** on 17 degrees of freedom  
Multiple R-squared: 0.9464, Adjusted R-squared: 0.9433  
F-statistic: 300.3 on 1 and 17 DF, p-value: 3.07e-12

Շրջանակի մեջ վերցված թիվը համապատասխանում է  $\sigma$  պարամետրիի գնահատականին:

Գալթոնի օրինակը

R ծրագիր

Փիրսոնի վիճակները

Սնդիկի գոյորշու ճնշումը

Դյուարիի բանձնը

Ավելորդ փոփոխականներ





Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-6.068144	0.483831	-12.54	5.10e-10 ***
pt	0.039792	0.002296	17.33	3.07e-12 ***

---  
 Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.096 on 17 degrees of freedom  
 Multiple R-squared: 0.9464, Adjusted R-squared: 0.9433  
 F-statistic: 300.3 on 1 and 17 DF, p-value: 3.07e-12

Շրջանակի մեջ վերցված թվերը համապատասխանում են  $\beta_0$  և  $\beta_1$  պարամետրերի գնահատականների սխալի սխալին՝  $\hat{\sigma}\sqrt{b_j}$ :

Գալթոնի օրինակը

R ծրագիր

Փիրսոնի վավայները

Սնդիկի գոլորշու ճնշումը

Դյուարեի բանձնը

Ավելորդ փոփոխականներ



Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-6.068144	0.483831	-12.54	5.10e-10	***
pt	0.039792	0.002296	17.33	3.07e-12	***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.096 on 17 degrees of freedom  
Multiple R-squared: 0.9464, Adjusted R-squared: 0.9433  
F-statistic: 300.3 on 1 and 17 DF, p-value: 3.07e-12

Շրջանակի մեջ վերցված թվերը համապատասխանում են այսպես կոչված  $T$ -վիճականու արժեքին՝  $T_j = \hat{\beta}_j / (\hat{\sigma} \sqrt{b_j})$ :

Գալթոնի օրինակը

R ծրագիր

Փիրսոնի վավայները

Սնդիկի գոյորշու ճնշումը

Դյուպրեի բանձնը

Ավելորդ փոփոխականներ



Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-6.068144	0.483831	-12.54	5.10e-10	***
pt	0.039792	0.002296	17.33	3.07e-12	***

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.096 on 17 degrees of freedom

Multiple R-squared: 0.9464, Adjusted R-squared: 0.9433

F-statistic: 300.3 on 1 and 17 DF, p-value: 3.07e-12

Շրջանակի մեջ վերցված թվերը համապատասխանում են այն պարահույթի հավանականությանը, որ  $\beta_j^*$  պարամետրի գրո լինելու դեպքում  $|T|$ -վիճականին կընդունի այդքան մեծ արժեք.

$$P_{\beta_j^*=0}(|T_j| > |T_j^{obs}|):$$

Գալթոնի օրինակը

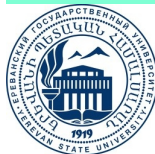
R ծրագիր

Փիրսոնի վավայները

Մեդիկի գոլորշու  
ճնշումը

Դյուպրեի բանաձևը

Ավելորդ  
փոփոխականներ



## Coefficients:

	Estimate	Std. Error	t value	Pr(> t )	
(Intercept)	-6.068144	0.483831	-12.54	5.10e-10	***
pt	0.039792	0.002296	17.33	3.07e-12	***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.096 on 17 degrees of freedom

Multiple R-squared: 0.9464, Adjusted R-squared: 0.9433

F-statistic: 300.3 on 1 and 17 DF, p-value: 3.07e-12

Շրջանակի մեջ վերցված թվերը համապատասխանում են այսպես կոչված որոշելության գործակցին, որը որոշում է մոդելի վստահելիության աստիճանը (ինչքան մեծ այնքան լավ):

Գալթոնի օրինակը

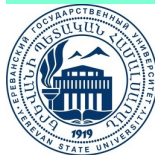
R ծրագիր

Փիրսոնի վավայները

Սնդիկի գոյորշու ճնշումը

Դյուարեի բանաձևը

Ավելորդ փոփոխականներ



- Ֆիզիկոսները պնդում են, որ առավել ճշգրիտ բանաձևը ունի հետևյալ տեսքը՝

$$P = \alpha_1 T^{\alpha_2} e^{\alpha_3/T} \quad (1)$$

որտեղ  $P$ -ն գոլորշու ճնշումն է,  $T$ -ն ջերմաստիճանը, իսկ  $\alpha_1, \alpha_2, \alpha_3$ -ը հաստատուններ են (կախված նյութից):

- Ցույց տալ, որ այս բանաձևի մեջ մտնող հաստատունները կարելի է որոշել իրականացնելով գծային ռեգրեսիա՝ որպես բացառության առարկա ընտրելով  $\log(P)$  փոփոխականը, իսկ որպես բացառող փոփոխականներ ընտրելով  $1/T$ -ն և  $\log(T)$ -ն:
- Որոշել  $\alpha_j, j = 1, 2, 3$  պարամետրերի գնահատականների արժեքները: Քանի որ  $\log$  ֆունկցիան սահմանված չէ 0 կետում, կարող ենք մի փոքր փոխել ավյալները փոխարինելով temperature փոփոխականի առաջին արժեքը 0.1-ով: Դրա համար կարող ենք օգտագործել edit հրամանը:
- Կարելի՞ է արդյոք պնդել, որ հետևյալ մոդելը

$$\log(P) = \beta_0 + \beta_1 \log(T) + \beta_2/T + \varepsilon \quad (2)$$

ավելի վստահելի է քան այն, որը որոշեցինք նախորդ էջին:

Գալթոնի օրինակը

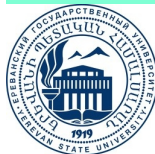
R ծրագիր

Փիրսոնի ավյալները

Ստեղծի գոլորշու ճնշումը

Դյուպրեի բանաձևը

Ավելորդ փոփոխականներ



Call:

```
lm(formula = log(pp) ~ log(pt) + I(1/pt))
```

Residuals:

Min	1Q	Median	3Q	Max
-0.83669	-0.46600	-0.04595	0.36489	1.75775

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	-23.9829	1.0605	-22.61	1.43e-13 ***
log(pt)	5.0949	0.2083	24.46	4.22e-14 ***
I(1/pt)	4.7353	0.3109	15.23	6.07e-11 ***

---  
Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6767 on 16 degrees of freedom  
Multiple R-squared: 0.9808, Adjusted R-squared: 0.9784  
F-statistic: 408.5 on 2 and 16 DF, p-value: 1.852e-14

Շրջանակի մեջ վերցված է այն հրամանը, որի միջոցով սրահցված է այս արդյունքը:

Գալթոնի օրինակը

R ծրագիր

Փիրսոնի ավյակները

Սեդիկի գոլորշու  
ճնշումը

Դիուարեի բանաձևը

Ավելորդ  
փոփոխականներ



```
Call:
lm(formula = log(pp) ~ log(pt) + I(1/pt))

Residuals:
    Min       1Q   Median       3Q      Max
-0.83669 -0.46600 -0.04595  0.36489  1.75775

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -23.9829     1.0605  -22.61 1.43e-13 ***
log(pt)       5.0949     0.2083   24.46 4.22e-14 ***
I(1/pt)       4.7353     0.3109   15.23 6.07e-11 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6767 on 16 degrees of freedom
Multiple R-squared:  0.9808,    Adjusted R-squared:  0.9784
F-statistic: 408.5 on 2 and 16 DF,  p-value: 1.852e-14
```

Գալթոնի օրինակը

R ծրագիր

Փիրսոնի փվայելը

Սեդիկի գոյորշու  
ճնշումը

Դյուարեի բանաձևը

Ավելորդ  
փոփոխականներ

Շրջանակի մեջ վերցված թիվը արտահայտում է սրացված մոդելի վստահելիության աստիճանը:



- Իրականացնել գծային ռեգրեսիա՝ որպես բացառության ենթակա փոփոխական ընտրելով  $\log(P)$ -ն, իսկ որպես բացառական փոփոխականներ ընտրելով  $\log(T)$ ,  $1/T$ ,  $1/T^2$  և  $T$ : Կարո՞ղ ենք եզրակացնել, որ  $T$  փոփոխականն անօգուտ է (5%-ի շեմով):
- Իրականացնել գծային ռեգրեսիա՝ որպես բացառության ենթակա փոփոխական ընտրելով  $\log(P)$ -ն, իսկ որպես բացառական փոփոխականներ ընտրելով  $\log(T)$ ,  $\exp(T)$ ,  $1/T^2$  և  $T$ : Կարո՞ղ ենք եզրակացնել, որ  $\exp(T)$  փոփոխականն անօգուտ է (5%-ի շեմով):

Գալթոնի օրինակը

R ծրագիր

Փիրսոնի փվյակները

Սեդիկի գոլորշու  
ճնշումը

Դյուպրեի բանձնը

Ավելորդ  
փոփոխականներ





```
Call:
lm(formula = log(pp) ~ log(pt) + I(1/pt) + pt + I(1/pt^2))

Residuals:
    Min       1Q   Median       3Q      Max
-0.126251 -0.046925  0.000001  0.053747  0.128212

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -29.354528   1.153649  -25.445 4.02e-13 ***
log(pt)      5.775785   0.253749   22.762 1.85e-12 ***
I(1/pt)      106.151437   9.178228   11.566 1.50e-08 ***
pt           0.005135    0.001122    4.576 0.000432 ***
I(1/pt^2)   -20.025005   1.774906  -11.282 2.06e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.07929 on 14 degrees of freedom
Multiple R-squared:  0.9998,    Adjusted R-squared:  0.9997
F-statistic: 1.516e+04 on 4 and 14 DF,  p-value: < 2.2e-16
```

Գալթոնի օրինակը

R ծրագիր

Փիրսոնի ավայելերը

Սեդիկի գոլորշու  
ճնշումը

Դյուարեի բանձնը

Ավելորդ  
փոփոխականներ

Կարմիր շրջանակի մեջ վերցված է պարամետրի գնահատականը: Չնայած այն չափ մոտ է զրոյի, պարամետրի զրո լինելու վարկածը այս դեպքում չի ընդունվում: Դա է վկայում կապույտ շրջանակի մեջ վերցված թվի 0.05-ից փոքր լինելը:



```
Call:
lm(formula = log(pp) ~ log(pt) + I(1/pt) + exp(pt) + I(1/pt^2))

Residuals:
    Min       1Q   Median       3Q      Max
-0.13704 -0.08267  0.00000  0.05772  0.24017

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -3.432e+01  5.504e-01 -62.349 < 2e-16 ***
log(pt)      6.888e+00  9.811e-02  70.213 < 2e-16 ***
I(1/pt)      1.428e+02  6.678e+00  21.385 4.33e-12 ***
exp(pt)      3.141e-158  6.174e-158  0.509  0.619
I(1/pt^2)    -2.709e+01  1.309e+00 -20.688 6.80e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.1241 on 14 degrees of freedom
Multiple R-squared:  0.9994,    Adjusted R-squared:  0.9993
F-statistic: 6187 on 4 and 14 DF,  p-value: < 2.2e-16
```

Գալթոնի օրինակը

R ծրագիր

Փիրսոնի կոլյակները

Ստեղծի գոլորշու  
ճնշումը

Դյուարեի բանձնը

Ավելորդ  
փոփոխականներ

Կարմիր շրջանակի մեջ վերցված է պարամետրի գնահատականը: Այն շատ մոտ է զրոյի և այս դեպքում պարամետրի զրո լինելու վարկածը ընդունվում է: Դա է վկայում կապույտ շրջանակի մեջ վերցված թվի 0.05-ից մեծ լինելը: