|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ | | | | | |
| **ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ** | | | | | |
| **УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ** | | | | | |
| **«СЕВАСТОПОЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»** | | | | | |
|  | | | | | |
| Институт информационных технологий и управления в технических системах | | | | | |
| (полное название института) | | | | | |
|  | | | | | |
| кафедра «Информационные системы» | | | | | |
| (полное название кафедры) | | | | | |
|  | | | | | |
|  | | | | | |
| **Лабораторные работы** | | | | | |
| по дисциплине “Интеллектуальный анализ данных” | | | | | |
| студента группы ИС/б-32-о  **Волобуева Юрия Сергеевича** | | | | | |
| № | Выполнение | Работу принял | | | |
| Дата | Дата | Оценка | Ф.И.О. | Подпись |
| 1 |  |  |  |  |  |
| 2 |  |  |  |  |  |
| 3 |  |  |  |  |  |
| 4 |  |  |  |  |  |
| 5 |  |  |  |  |  |
| 6 |  |  |  |  |  |
| 7 |  |  |  |  |  |
| 8 |  |  |  |  |  |
| 9 |  |  |  |  |  |
| 10 |  |  |  |  |  |

|  |  |  |
| --- | --- | --- |
| 20 | 20 | г. |

Лабораторная работа №1

“Исследование возможностей языка R для статистического анализа данных”

1 ЦЕЛЬ РАБОТЫ

− изучить основные особенности языка R;

− исследовать возможности языка R для работы с графикой.

2 ПОСТАНОВКА ЗАДАЧИ

1. Исследовать команду 'demo()', полученные результаты вставить в отчет
2. Исследовать основные функции и команды языка R, представленные в данной лабораторной работе, полученные результаты вставить в отчет.

3 ХОД РАБОТЫ

1. Результаты выполнения команды “demo(glm.vr)”

> # Copyright (C) 1997-2009 The R Core Team

>

> #### -\*- R -\*-

> require(stats)

> Fr <- c(68,42,42,30, 37,52,24,43,

+ 66,50,33,23, 47,55,23,47,

+ 63,53,29,27, 57,49,19,29)

> Temp <- gl(2, 2, 24, labels = c("Low", "High"))

> Soft <- gl(3, 8, 24, labels = c("Hard","Medium","Soft"))

> M.user <- gl(2, 4, 24, labels = c("N", "Y"))

> Brand <- gl(2, 1, 24, labels = c("X", "M"))

> detg <- data.frame(Fr,Temp, Soft,M.user, Brand)

> detg.m0 <- glm(Fr ~ M.user\*Temp\*Soft + Brand, family = poisson, data = detg)

> summary(detg.m0)

Call:

glm(formula = Fr ~ M.user \* Temp \* Soft + Brand, family = poisson,

data = detg)

Deviance Residuals:

Min 1Q Median 3Q Max

-2.20876 -0.99190 -0.00126 0.93542 1.97601

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 4.01524 0.10034 40.018 < 2e-16 \*\*\*

M.userY -0.21184 0.14257 -1.486 0.13731

TempHigh -0.42381 0.15159 -2.796 0.00518 \*\*

SoftMedium 0.05311 0.13308 0.399 0.68984

SoftSoft 0.05311 0.13308 0.399 0.68984

BrandM -0.01587 0.06300 -0.252 0.80106

M.userY:TempHigh 0.13987 0.22168 0.631 0.52806

M.userY:SoftMedium 0.08323 0.19685 0.423 0.67245

M.userY:SoftSoft 0.12169 0.19591 0.621 0.53449

TempHigh:SoftMedium -0.30442 0.22239 -1.369 0.17104

TempHigh:SoftSoft -0.30442 0.22239 -1.369 0.17104

M.userY:TempHigh:SoftMedium 0.21189 0.31577 0.671 0.50220

M.userY:TempHigh:SoftSoft -0.20387 0.32540 -0.627 0.53098

---

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 118.627 on 23 degrees of freedom

Residual deviance: 32.826 on 11 degrees of freedom

AIC: 191.24

Number of Fisher Scoring iterations: 4

> detg.mod <- glm(terms(Fr ~ M.user\*Temp\*Soft + Brand\*M.user\*Temp,

+ keep.order = TRUE),

+ family = poisson, data = detg)

> summary(detg.mod)

Call:

glm(formula = terms(Fr ~ M.user \* Temp \* Soft + Brand \* M.user \*

Temp, keep.order = TRUE), family = poisson, data = detg)

Deviance Residuals:

Min 1Q Median 3Q Max

-0.91365 -0.35585 0.00253 0.33027 0.92146

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 4.14887 0.10603 39.128 < 2e-16 \*\*\*

M.userY -0.40521 0.16188 -2.503 0.01231 \*

TempHigh -0.44275 0.17121 -2.586 0.00971 \*\*

M.userY:TempHigh -0.12692 0.26257 -0.483 0.62883

SoftMedium 0.05311 0.13308 0.399 0.68984

SoftSoft 0.05311 0.13308 0.399 0.68984

M.userY:SoftMedium 0.08323 0.19685 0.423 0.67245

M.userY:SoftSoft 0.12169 0.19591 0.621 0.53449

TempHigh:SoftMedium -0.30442 0.22239 -1.369 0.17104

TempHigh:SoftSoft -0.30442 0.22239 -1.369 0.17104

M.userY:TempHigh:SoftMedium 0.21189 0.31577 0.671 0.50220

M.userY:TempHigh:SoftSoft -0.20387 0.32540 -0.627 0.53098

BrandM -0.30647 0.10942 -2.801 0.00510 \*\*

M.userY:BrandM 0.40757 0.15961 2.554 0.01066 \*

TempHigh:BrandM 0.04411 0.18463 0.239 0.81119

M.userY:TempHigh:BrandM 0.44427 0.26673 1.666 0.09579 .

---

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 118.627 on 23 degrees of freedom

Residual deviance: 5.656 on 8 degrees of freedom

AIC: 170.07

Number of Fisher Scoring iterations: 4

> summary(detg.mod, correlation = TRUE, symbolic.cor = TRUE)

Call:

glm(formula = terms(Fr ~ M.user \* Temp \* Soft + Brand \* M.user \*

Temp, keep.order = TRUE), family = poisson, data = detg)

Deviance Residuals:

Min 1Q Median 3Q Max

-0.91365 -0.35585 0.00253 0.33027 0.92146

Coefficients:

Estimate Std. Error z value Pr(>|z|)

(Intercept) 4.14887 0.10603 39.128 < 2e-16 \*\*\*

M.userY -0.40521 0.16188 -2.503 0.01231 \*

TempHigh -0.44275 0.17121 -2.586 0.00971 \*\*

M.userY:TempHigh -0.12692 0.26257 -0.483 0.62883

SoftMedium 0.05311 0.13308 0.399 0.68984

SoftSoft 0.05311 0.13308 0.399 0.68984

M.userY:SoftMedium 0.08323 0.19685 0.423 0.67245

M.userY:SoftSoft 0.12169 0.19591 0.621 0.53449

TempHigh:SoftMedium -0.30442 0.22239 -1.369 0.17104

TempHigh:SoftSoft -0.30442 0.22239 -1.369 0.17104

M.userY:TempHigh:SoftMedium 0.21189 0.31577 0.671 0.50220

M.userY:TempHigh:SoftSoft -0.20387 0.32540 -0.627 0.53098

BrandM -0.30647 0.10942 -2.801 0.00510 \*\*

M.userY:BrandM 0.40757 0.15961 2.554 0.01066 \*

TempHigh:BrandM 0.04411 0.18463 0.239 0.81119

M.userY:TempHigh:BrandM 0.44427 0.26673 1.666 0.09579 .

---

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

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 118.627 on 23 degrees of freedom

Residual deviance: 5.656 on 8 degrees of freedom

AIC: 170.07

Number of Fisher Scoring iterations: 4

Correlation of Coefficients:

(Intercept) 1

M.userY , 1

TempHigh , . 1

M.userY:TempHigh . , , 1

SoftMedium , . . 1

SoftSoft , . . . 1

M.userY:SoftMedium . , . , . 1

M.userY:SoftSoft . , . . , . 1

TempHigh:SoftMedium . , . . . . 1

TempHigh:SoftSoft . , . . . . . 1

M.userY:TempHigh:SoftMedium . . . . , . , . 1

M.userY:TempHigh:SoftSoft . . . . . , . , . 1

BrandM . 1

M.userY:BrandM . , 1

TempHigh:BrandM . . . . 1

M.userY:TempHigh:BrandM . . . . , 1

attr(,"legend")

[1] 0 ‘ ’ 0.3 ‘.’ 0.6 ‘,’ 0.8 ‘+’ 0.9 ‘\*’ 0.95 ‘B’ 1

> anova(detg.m0, detg.mod)

Analysis of Deviance Table

Model 1: Fr ~ M.user \* Temp \* Soft + Brand

Model 2: Fr ~ M.user \* Temp \* Soft + Brand \* M.user \* Temp

Resid. Df Resid. Dev Df Deviance

1 11 32.826

2 8 5.656 3 27.17

1. Результаты выполнения функций и команд из методических указаний

- Арифметические действия

> 3/7

[1] 0.4285714

> 3/7-0.4285714

[1] 2.857143e-08

> sqrt(2)\*sqrt(2)

[1] 2

> (sqrt(2)\*sqrt(2))-2

[1] 4.440892e-16

- Команда Round()

> round(pi, 3)

[1] 3.142

> round(pi, digits = 10)

[1] 3.141593

> round(pi, d = 5)

[1] 3.14159

> round(digits = 5, pi)

[1] 3.14159

- Сохранение и вывод данных

> number <- 10

> number

[1] 10

> (number <- 10)

[1] 10

> string <- "Hello"

> string

[1] "Hello"

- Работа с векторами

> v1 <- c(2, 3, 4, 6, 10)

> v1

[1] 2 3 4 6 10

> v1[3]

[1] 4

> v1[3:5]

[1] 4 6 10

> v2 <- c(1:5)

> v2

[1] 1 2 3 4 5

> s <- rep("a", 4)

> s

[1] "a" "a" "a" "a"

> rep(1:4, 2)

[1] 1 2 3 4 1 2 3 4

> rep(1:4, each = 2)

[1] 1 1 2 2 3 3 4 4

> (v1 + v2) \* v2

[1] 3 10 21 40 75

> crossprod(v1, v2)

[,1]

[1,] 94

> v1[v1 > 4]

[1] 6 10

> length(v1)

[1] 5

> mean(v1)

[1] 5

> var(v1)

[1] 10

- Работа с матрицами

> args(matrix)

function (data = NA, nrow = 1, ncol = 1, byrow = FALSE, dimnames = NULL)

NULL

> function (data = NA, nrow = 1, ncol = 1, byrow = FALSE, dimnames =

+ NULL)

+ matrix(data = 1:5, nrow = 5, ncol = 5, byrow = FALSE)

function (data = NA, nrow = 1, ncol = 1, byrow = FALSE, dimnames =

NULL)

matrix(data = 1:5, nrow = 5, ncol = 5, byrow = FALSE)

> matrix(data = 1:5, nrow = 5, ncol = 5, byrow = TRUE)

[,1] [,2] [,3] [,4] [,5]

[1,] 1 2 3 4 5

[2,] 1 2 3 4 5

[3,] 1 2 3 4 5

[4,] 1 2 3 4 5

[5,] 1 2 3 4 5

> matrix(data = NA, nrow = 3, ncol = 3)

[,1] [,2] [,3]

[1,] NA NA NA

[2,] NA NA NA

[3,] NA NA NA

> m <- cbind(v1, v2) # Создаем матрицу

> m

v1 v2

[1,] 2 1

[2,] 3 2

[3,] 4 3

[4,] 6 4

[5,] 10 5

> typeof(m)

[1] "double"

> class(m)

[1] "matrix"

> is.matrix(m)

[1] TRUE

> is.vector(m)

[1] FALSE

> dim(m)

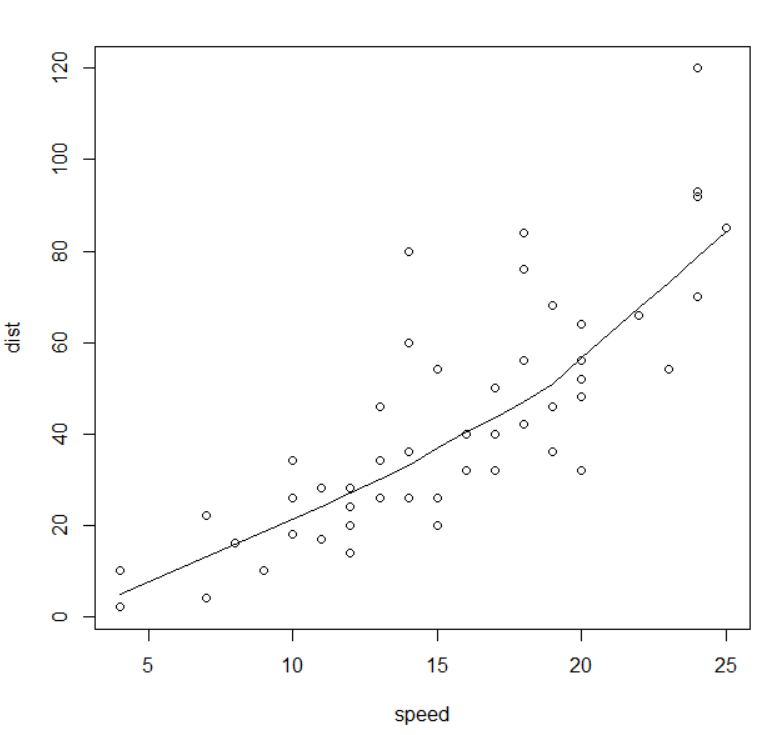
[1] 5 2

- Работа с графиками

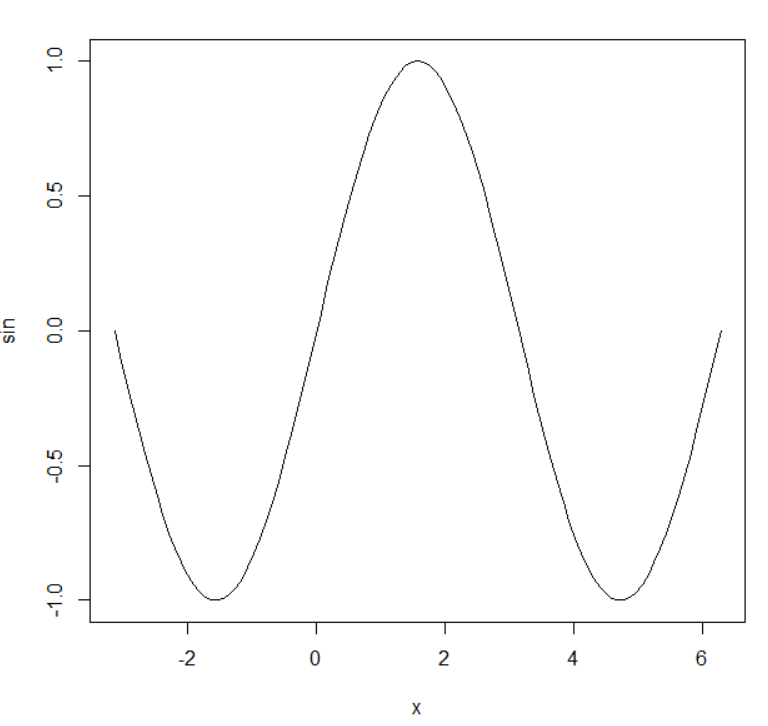
> require(stats)

> plot(cars)

> lines(lowess(cars))

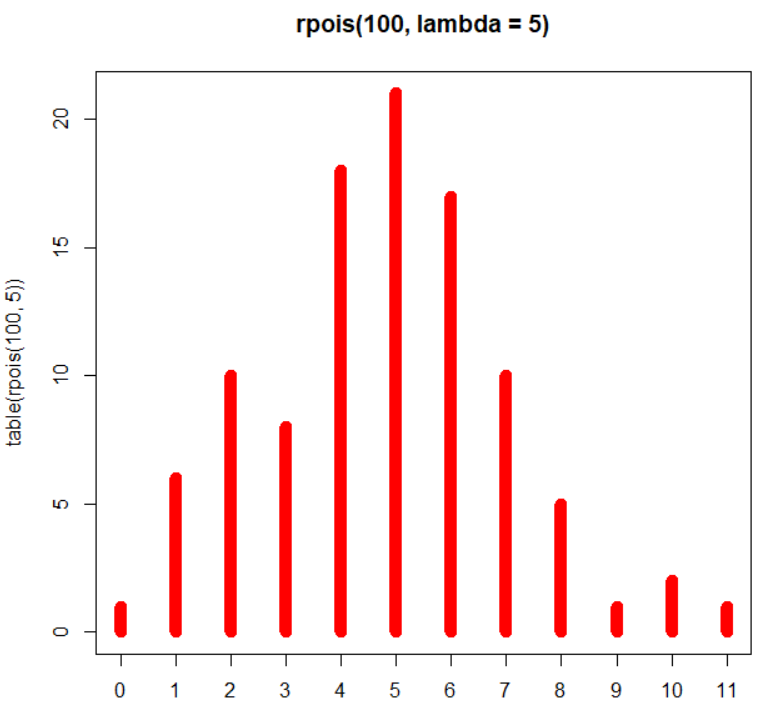


> plot(sin, -pi, 2\*pi)



> plot(table(rpois(100, 5)), type = "h", col = "red", lwd = 10,

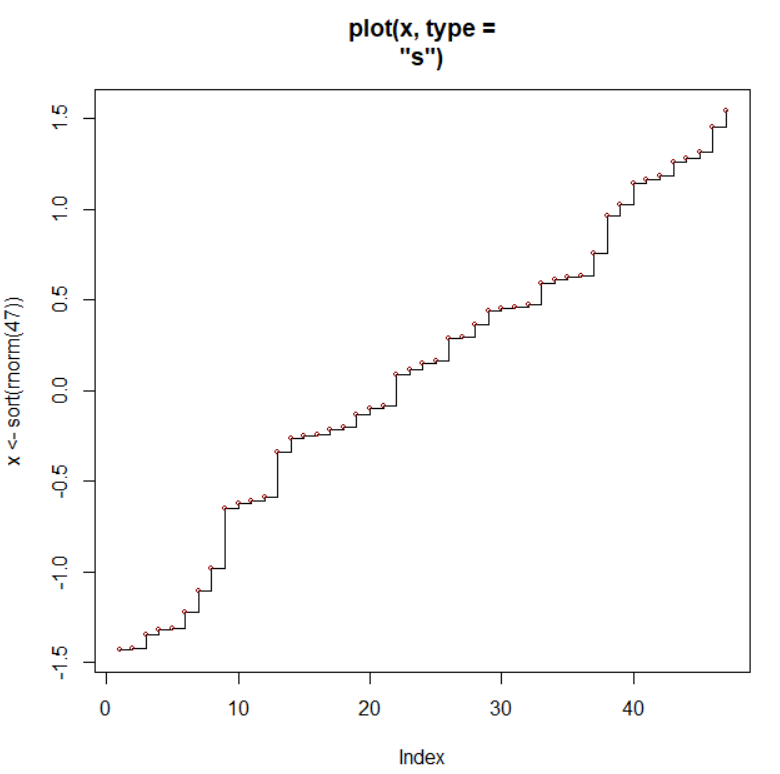
+ main = "rpois(100, lambda = 5)")



> plot(x <- sort(rnorm(47)), type = "s", main = "plot(x, type =

+ \"s\")")

> points(x, cex = .5, col = "dark red")



ВЫВОД

В ходе выполнения лабораторной работы было получено общее представление о статистической системе анализов R. Изучены особенности интерфейса, функциональные основные возможности, сформировались навыки практической работы в среде R-Studio.

Сделаны примеры использования функций и команд, приведенных в методических указаниях.