Short-demo

Boris Demeshev

7 January 2015

If you can't see cyrillic letters in Rstudio, select File — Reopen with encoding... — UTF-8 — Set as default encoding for source files — Ok.

Перед компиляцией надобно установить пакеты: knitr, ggplot2, pander, memisc, psych

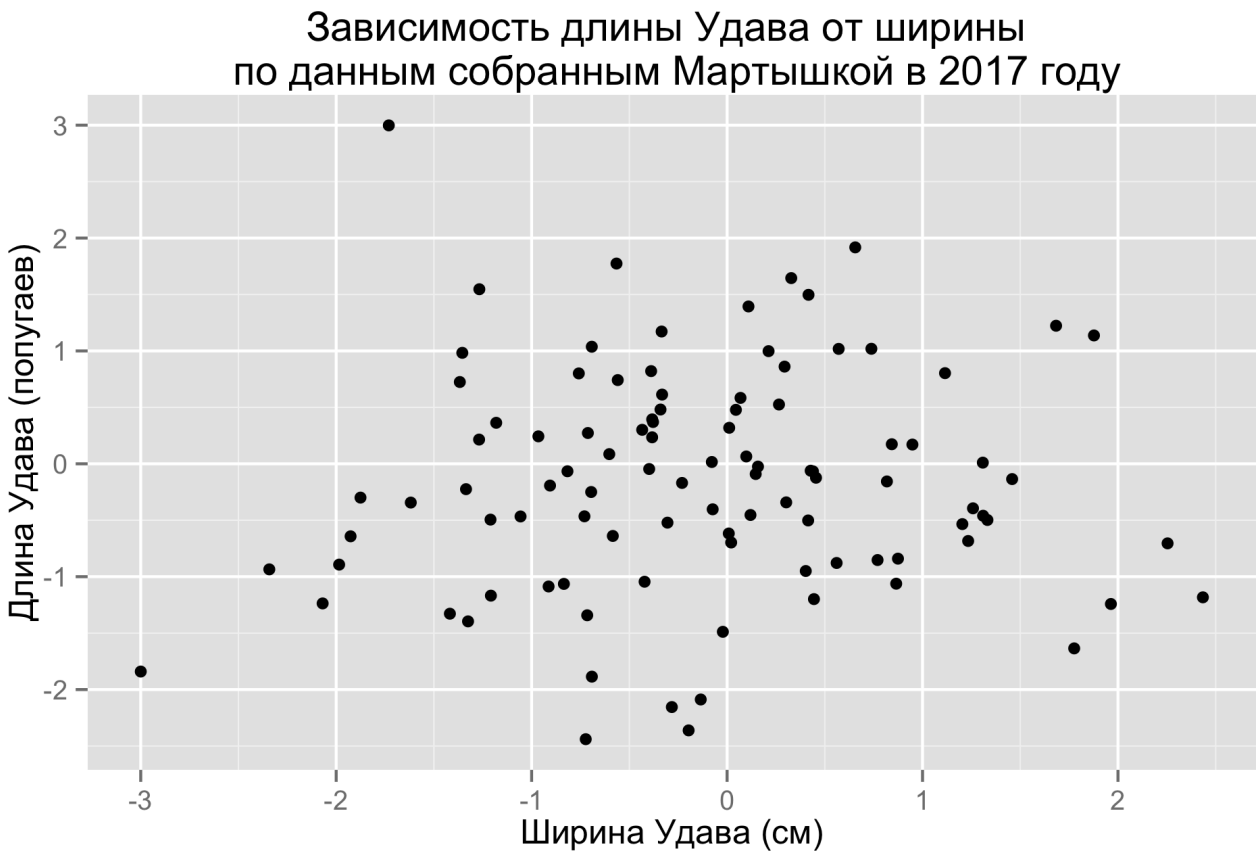
Поехали...

```
library("knitr")
opts_chunk$set(dev='png' , dpi=300, warning=FALSE, message=FALSE)
```

```
library("ggplot2")
library("pander")
library("memisc")
library("psych")
```

Просто график с русскими буквами

```
qplot(x=rnorm(100), y=rnorm(100), main="Зависимость длины Удава от ширины \n по данным собранным Мартышк
```



Начало набора данных:

```
h <- swiss
head(h)
```

```
##              Fertility Agriculture Examination Education Catholic
## Courtelary        80.2        17.0          15        12     9.96
## Delemont          83.1        45.1           6         9    84.84
## Franches-Mnt      92.5        39.7           5         5    93.40
## Moutier           85.8        36.5          12         7    33.77
## Neuveville        76.9        43.5          17        15     5.16
## Porrentruy        76.1        35.3           9         7    90.57
##              Infant.Mortality
## Courtelary               22.2
## Delemont                 22.2
## Franches-Mnt             20.2
## Moutier                  20.3
## Neuveville               20.6
## Porrentruy               26.6
```

То же начало, только красиво:

```
pander(head(h))
```

	Fertility	Agriculture	Examination
Courtelary	80.2	17	15
Delemont	83.1	45.1	6
Franches-Mnt	92.5	39.7	5
Moutier	85.8	36.5	12
Neuveville	76.9	43.5	17
Porrentruy	76.1	35.3	9

Таблица 1: Table continues below

	Education	Catholic	Infant.Mortality
Courtelary	12	9.96	22.2
Delemont	9	84.84	22.2
Franches-Mnt	5	93.4	20.2
Moutier	7	33.77	20.3
Neuveville	15	5.16	20.6
Porrentruy	7	90.57	26.6

Описательные статистики:

```
describe(h)
```

```
##                  vars  n  mean    sd median trimmed   mad   min   max
## Fertility           1 47 70.14 12.49  70.40   70.66 10.23 35.00  92.5
## Agriculture         2 47 50.66 22.71  54.10   51.16 23.87  1.20  89.7
## Examination         3 47 16.49  7.98  16.00   16.08  7.41  3.00  37.0
## Education           4 47 10.98  9.62   8.00    9.38  5.93  1.00  53.0
## Catholic            5 47 41.14 41.70  15.14   39.12 18.65  2.15 100.0
## Infant.Mortality    6 47 19.94  2.91  20.00   19.98  2.82 10.80  26.6
##                  range  skew kurtosis   se
## Fertility        57.50 -0.46    -0.26 1.82
## Agriculture      88.50 -0.32    -0.89 3.31
## Examination      34.00  0.45    -0.14 1.16
## Education        52.00  2.27     6.14 1.40
## Catholic         97.85  0.48    -1.67 6.08
## Infant.Mortality 15.80 -0.33     0.78 0.42
```

Часть описательных статистик в красивой табличке:

```
all_stats <- describe(h)
class(all_stats) <- "data.frame"
some_stats <- all_stats[,c("mean","median","min","max","sd")]
pander(some_stats)
```

	mean	median	min	max	sd
Fertility	70.14	70.4	35	92.5	12.49
Agriculture	50.66	54.1	1.2	89.7	22.71
Examination	16.49	16	3	37	7.978
Education	10.98	8	1	53	9.615
Catholic	41.14	15.14	2.15	100	41.7
Infant.Mortality	19.94	20	10.8	26.6	2.913

Оценим две модели

```
m1 <- lm(data=h, Fertility~Agriculture)
m2 <- lm(data=h, Fertility~Agriculture+Catholic)
```

Сравним просто текстом:

```
mtable("Ограниченная модель"=m1,"Неограниченная модель"=m2,
       summary.stats=c("R-squared","Deviance","N"))
```

```
##
## Calls:
## Ограниченная модель: lm(formula = Fertility ~ Agriculture, data = h)
## Неограниченная модель: lm(formula = Fertility ~ Agriculture + Catholic, data = h)
```

```
##
## ========================================================
##                Ограниченная модель  Неограниченная модель
## --------------------------------------------------------
## (Intercept)        60.304***            59.864***
##                   (4.251)              (3.988)
## Agriculture         0.194*               0.110
##                   (0.077)              (0.078)
## Catholic                                 0.115*
##                                         (0.043)
## --------------------------------------------------------
## R-squared             0.125                0.248
## Deviance           6283.116             5395.825
## N                    47                   47
## ========================================================
```

Красивая табличка:

```
comparison <- mtable("Ограниченная модель"=m1,"Неограниченная модель"=m2,
       summary.stats=c("R-squared","Deviance","N"))
pander(comparison)
```

	Ограниченная модель	Неограниченная модель
(Intercept)	60.304*** (4.251)	59.864*** (3.988)
Agriculture	0.194* (0.077)	0.110 (0.078)
Catholic		0.115* (0.043)
R-squared	0.125	0.248
Deviance	6283.116	5395.825
N	47	47

И пример красиво набранной формулы Стирлинга:

$$n! \sim \left(\frac{n}{e}\right)^n \sqrt{2\pi n}$$

Или уравнение модели

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$$

Или оценённое уравнение

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + \hat{\beta}_3 z_i$$