

# Швейцарские кантоны в 1888 году

Хрюша

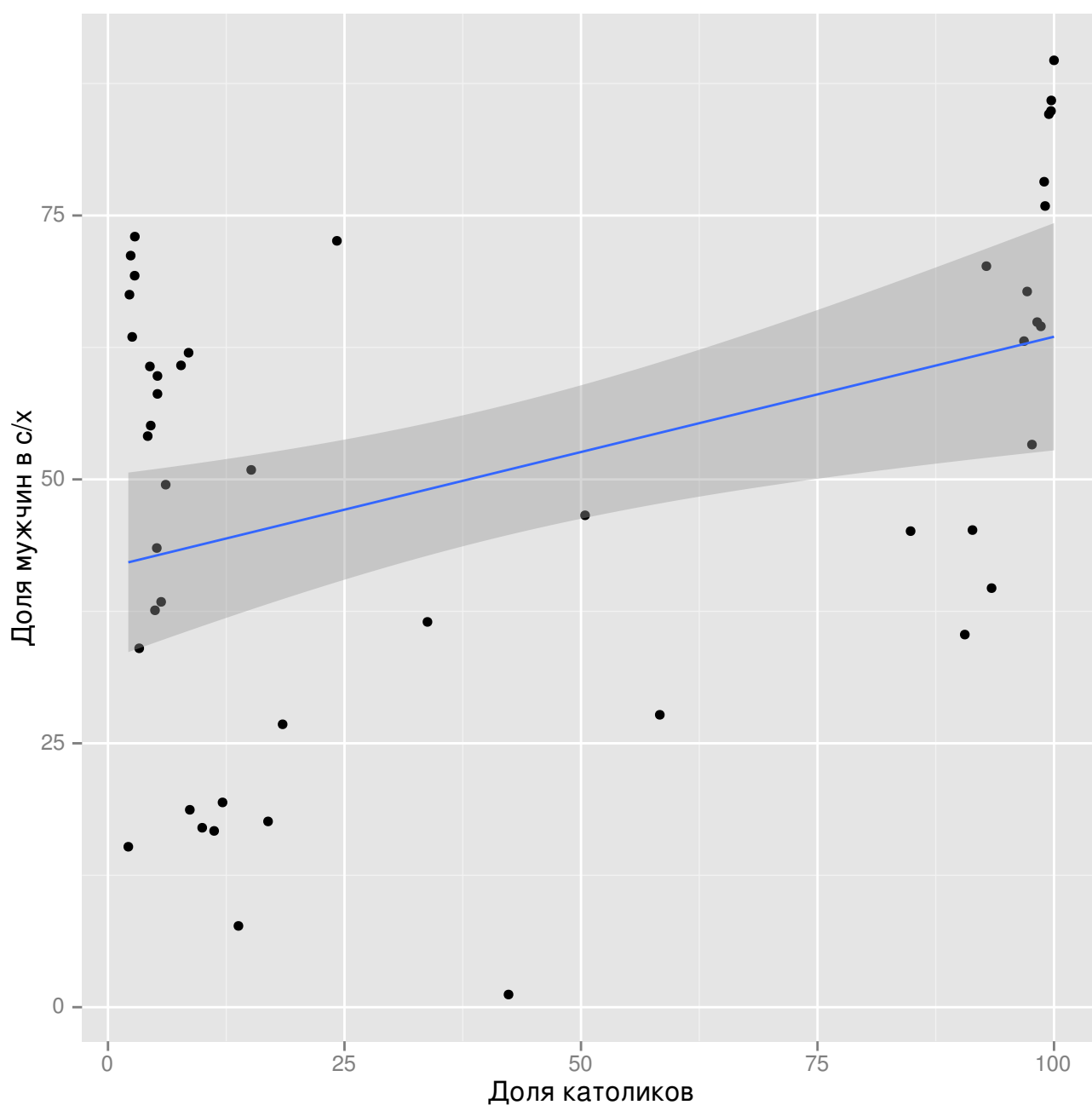
18 декабря 2012 г.

Загрузим данные по швейцарским кантонам

```
h <- swiss
```

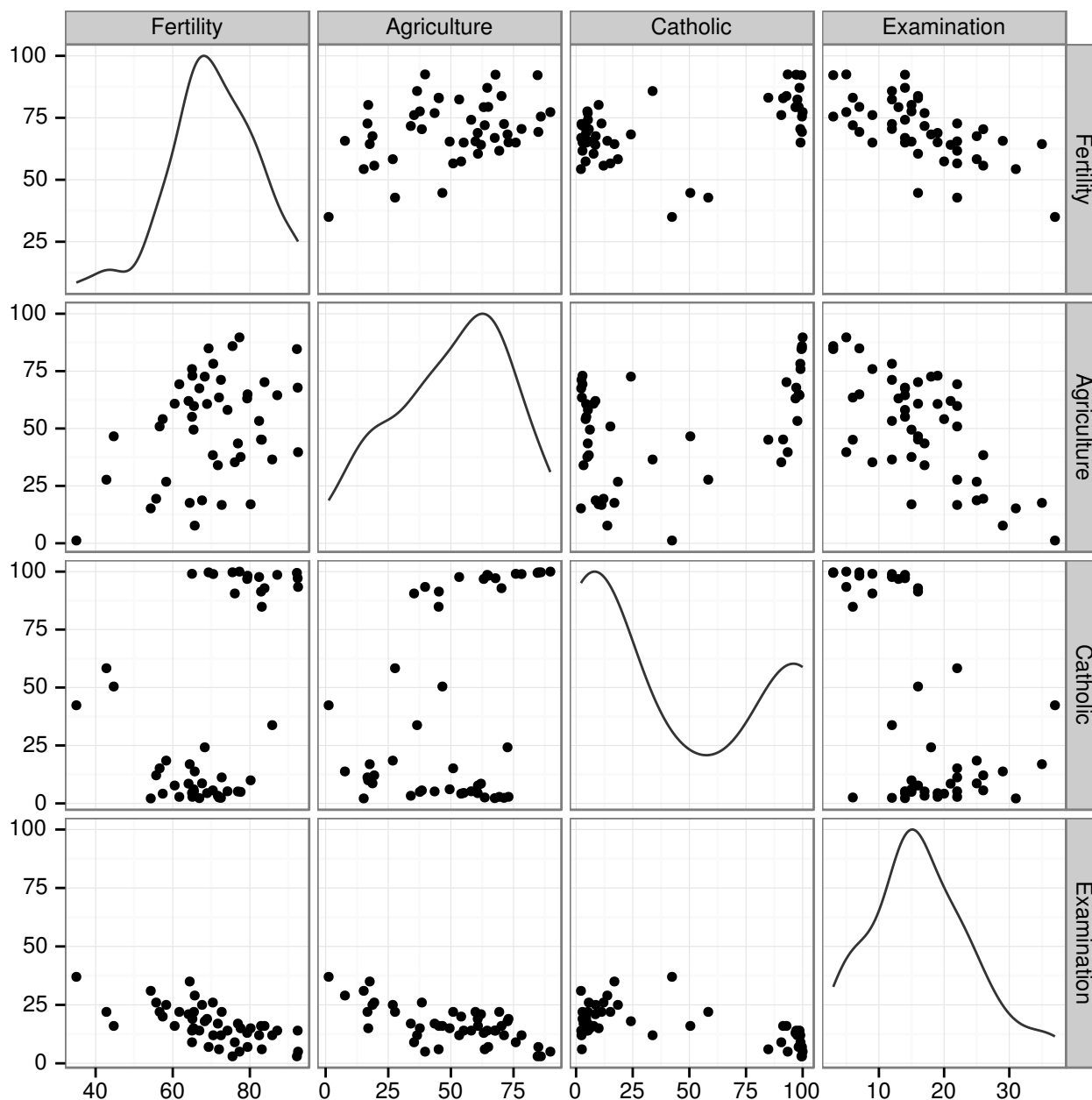
Построим несколько графиков для наглядности:

```
ggplot(h, aes(x = Catholic, y = Agriculture)) + geom_point(size = 2) + stat_smooth(method = lm) +  
  xlab("Доля католиков") + ylab("Доля мужчин в с/х")
```



Если график планируется печатать на бумаге, а не показывать на презентации, то можно указать для него чёрно-белую тему с помощью опции `theme_bw()`:

```
plotmatrix(h[, c("Fertility", "Agriculture", "Catholic", "Examination")]) +
  xlab("") + ylab("") + theme_bw()
```



Оценим зависимость доли мужчин занятых в сельском хозяйстве от доли католиков и средних результатов экзаменов.

```
model <- lm(Agriculture ~ Catholic + Examination, data = h)
report <- summary(model)

coef.table <- report$coefficients

colnames(coef.table) <- c("Оценка", "$\\hat{\\sigma}_{\\hat{\\beta}}$",
  "t-статистика", "Р-значение")
rownames(coef.table) <- c("Константа", "Доля католиков",
```

```
"Результат экзамена")
```

```
print(xtable(coef.table), sanitize.colnames.function = identity)
```

|                    | Оценка | $\hat{\sigma}_{\hat{\beta}}$ | t-статистика | P-значение |
|--------------------|--------|------------------------------|--------------|------------|
| Константа          | 82.31  | 8.72                         | 9.44         | 0.00       |
| Доля католиков     | 0.01   | 0.07                         | 0.09         | 0.93       |
| Результат экзамена | -1.94  | 0.38                         | -5.08        | 0.00       |

В этой регрессии коэффициент детерминации  $R^2$  равен 0.4714. Сумма квадратов остатков,  $SSR$ , равна 12541.1965.

```
model.p.value <- 1 - pf(report$fstatistic[1], report$fstatistic[2], report$fstatistic[3])
```

Гипотеза о незначимости регрессии в целом отвергается, т.к.  $F$ -статистика равна 19.6219 с точным  $P$ -значением равным 0.

Оценка ковариационной матрицы коэффициентов равна

```
var.hat <- vcov(model)
xtable(var.hat)
```

|             | (Intercept) | Catholic | Examination |
|-------------|-------------|----------|-------------|
| (Intercept) | 75.97       | -0.48    | -3.04       |
| Catholic    | -0.48       | 0.01     | 0.02        |
| Examination | -3.04       | 0.02     | 0.14        |

Сравним несколько моделей в одной таблице...

```
model2 <- lm(Agriculture ~ Catholic + Examination + Education, data = h)
table.comp <- texreg(list(model, model2), use.packages = FALSE, table = FALSE,
  model.names = c("Модель без Education", "Модель с Education"))
```

|             | Модель без Education | Модель с Education |
|-------------|----------------------|--------------------|
| (Intercept) | 82.31***<br>(8.72)   | 72.69***<br>(9.19) |
| Catholic    | 0.01<br>(0.07)       | 0.08<br>(0.08)     |
| Examination | -1.94***<br>(0.38)   | -0.94*<br>(0.55)   |
| Education   |                      | -0.91**<br>(0.38)  |
| $R^2$       | 0.47                 | 0.53               |
| Adj. $R^2$  | 0.45                 | 0.50               |
| Num. obs.   | 47                   | 47                 |

\*\*\*  $p < 0.01$ , \*\*  $p < 0.05$ , \*  $p < 0.1$

При написании «настоящей» работы программный код чаще всего скрывают с помощью опции `echo=FALSE`. С этой опцией будут видны только результаты работы R. В учебной работе эту опцию включать не надо, чтобы получить вкусные комментарии к своему коду.

В конце учебной работы имеет смысл дать команду

```
sessionInfo()
```

```
## R version 2.14.1 (2011-12-22)
## Platform: i686-pc-linux-gnu (32-bit)
##
```

```

## locale:
## [1] LC_CTYPE=en_GB.UTF-8      LC_NUMERIC=C
## [3] LC_TIME=ru_RU.UTF-8      LC_COLLATE=en_GB.UTF-8
## [5] LC_MONETARY=ru_RU.UTF-8  LC_MESSAGES=en_GB.UTF-8
## [7] LC_PAPER=C                LC_NAME=C
## [9] LC_ADDRESS=C              LC_TELEPHONE=C
## [11] LC_MEASUREMENT=ru_RU.UTF-8 LC_IDENTIFICATION=C
##
## attached base packages:
## [1] stats      graphics  grDevices  utils      datasets  methods    base
##
## other attached packages:
## [1] texreg_1.15      ggplot2_0.9.2.1 xtable_1.7-0    knitr_0.8
##
## loaded via a namespace (and not attached):
## [1] colorspace_1.1-1  dichromat_1.2-4   digest_0.5.2
## [4] evaluate_0.4.2    formatR_0.6       grid_2.14.1
## [7] gtable_0.1.1      labeling_0.1      MASS_7.3-16
## [10] memoise_0.1       munsell_0.3       plyr_1.7.1
## [13] proto_0.3-9.2     RColorBrewer_1.0-5 reshape2_1.2.1
## [16] scales_0.2.2      stringr_0.6.1     tools_2.14.1

```