

# Линейная регрессия

Рябенко Евгений  
riabenko.e@gmail.com

1 декабря 2016 г.

# Успеваемость школьников

Влияет ли уровень потребления алкоголя на успеваемость школьников?

Эксперимент:

- возьмём случайную выборку школьников
- назначим им случайную еженедельную дозу алкоголя
- по окончании учебного года измерим корреляцию между дозой и успеваемостью

# Успеваемость школьников

Влияет ли уровень потребления алкоголя на успеваемость школьников?

Эксперимент:

- возьмём случайную выборку школьников
- назначим им случайную еженедельную дозу алкоголя
- по окончании учебного года измерим корреляцию между дозой и успеваемостью

Неэтично!

## Успеваемость школьников

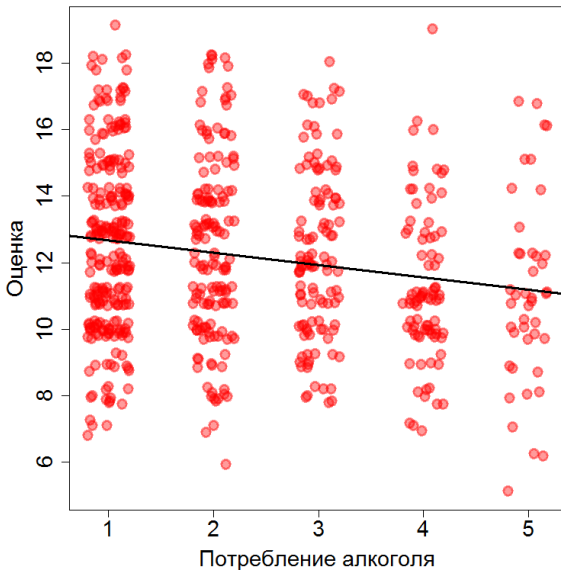
Эксперимент провести нельзя, но есть обсервационные данные.

Для 633 учеников старших классов двух португальских школ известны ряд демографических показателей и показателей успеваемости; известны уровень потребления алкоголя по выходным и финальная оценка по португальскому языку.

---

Cortez P., Silva A. (2008). *Using Data Mining to Predict Secondary School Student Performance*. Proceedings of 5th FUTURE BUSINESS TECHNOLOGY Conference, pp. 5-12.

## Успеваемость школьников



# Успеваемость школьников

У нас есть ещё 29 признаков, потенциально влияющих на успеваемость.

Если учесть их влияние, остаётся ли у потребления алкоголя предсказательная сила?

Можно ли утверждать, что повышение потребления алкоголя вызывает снижение оценок?

# Линейная регрессия

$1, \dots, n$  — объекты

$x_1, \dots, x_k$  — объясняющие переменные

$y$  — отклик

Ищем такой вектор  $\beta$ , что  $y \approx \beta x$ .

Модель линейной регрессии:

$$\mathbb{E}(y|x) = \beta_0 + \sum_{j=1}^k \beta_j x_j$$

$\beta_j$  показывает, насколько в среднем увеличивается  $y$ , если  $x_j$  увеличивается на единицу, а остальные факторы фиксированы.

⇒ регрессию можно использовать для исследования остаточного влияния признака на отклик с учётом других признаков.

# Линейная регрессия

Модель линейной регрессии:

$$\mathbb{E}(y|x) = \beta_0 + \sum_{j=1}^k \beta_j x_j$$

$$X = \begin{pmatrix} x_{10} = 1 & x_{11} & \dots & x_{1k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{n0} = 1 & x_{n1} & \dots & x_{nk} \end{pmatrix} \quad y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \quad \beta = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}$$



Метод наименьших квадратов:

$$\|y - X\beta\|_2^2 \rightarrow \min_{\beta};$$

$$\hat{\beta} = (X^T X)^{-1} X^T y,$$

$$\hat{y} = X (X^T X)^{-1} X^T y.$$

## Качество решения

$$\text{TSS} = \sum_{i=1}^n (y_i - \bar{y})^2 \quad (\text{Total Sum of Squares});$$

$$\text{ESS} = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 \quad (\text{Explained Sum of Squares});$$

$$\text{RSS} = \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (\text{Residual Sum of Squares});$$

$$\text{TSS} = \text{ESS} + \text{RSS}.$$

Коэффициент детерминации:

$$R^2 = \frac{\text{ESS}}{\text{TSS}} = 1 - \frac{\text{RSS}}{\text{TSS}}.$$

## Приведённый коэффициент детерминации

Стандартный коэффициент детерминации всегда увеличивается при добавлении регрессоров в модель.

Для сравнения качества моделей, содержащих разное число признаков, можно использовать приведённый коэффициент детерминации:

$$R_a^2 = \frac{ESS / (n - k - 1)}{TSS / (n - 1)} = 1 - (1 - R^2) \frac{n - 1}{n - k - 1}.$$

# Базовые предположения МНК

1 Линейность отклика:

$$y = X\beta + \varepsilon$$

# Базовые предположения МНК

- 1 Линейность отклика
- 2 Случайность выборки: наблюдения  $(x_i, y_i)$  независимы.

# Базовые предположения МНК

- 1 Линейность отклика
- 2 Случайность выборки
- 3 Полнота ранга  $X$ : ни один из признаков не является линейной комбинацией других признаков ( $\text{rank } X = k + 1$ ).

# Базовые предположения МНК

- 1 Линейность отклика
- 2 Случайность выборки
- 3 Полнота ранга  $X$
- 4 Случайность ошибок:

$$\mathbb{E}(\varepsilon | x) = 0$$

# Базовые предположения МНК

- 1 Линейность отклика
- 2 Случайность выборки
- 3 Полнота ранга  $X$
- 4 Случайность ошибок

⇒ МНК-оценки коэффициентов  $\beta$  несмещённые:

$$\mathbb{E}\hat{\beta}_j = \beta_j$$

и состоятельные:

$$\forall \gamma > 0 \quad \lim_{n \rightarrow \infty} P\left(\left|\beta_j - \hat{\beta}_j\right| < \gamma\right) = 1$$



# Базовые предположения МНК

- 1 Линейность отклика
- 2 Случайность выборки
- 3 Полнота ранга  $X$
- 4 Случайность ошибок
- 5 Гомоскедастичность ошибок: дисперсия ошибки не зависит от значений признаков:  $\mathbb{D}(\varepsilon | x) = \sigma^2$

(предположения Гаусса-Маркова)

# Базовые предположения МНК

- 1 Линейность отклика
- 2 Случайность выборки
- 3 Полнота ранга  $X$
- 4 Случайность ошибок
- 5 Гомоскедастичность ошибок

⇒ МНК-оценки имеют наименьшую дисперсию в классе оценок  $\beta$ , линейных по  $y$ .

# Дисперсия $\hat{\beta}_j$

(1)-(5)⇒

$$\mathbb{D}(\hat{\beta}_j) = \frac{\sigma^2}{TSS_j(1 - R_j^2)},$$

где  $TSS_j = \sum_{i=1}^n (x_{ij} - \bar{x}_j)^2$ ,  $R_j^2$  — коэффициент детерминации при регрессии  $x_j$  на все остальные признаки.

- Чем больше  $\sigma^2$ , тем больше дисперсия  $\hat{\beta}_j$ .
- Чем больше вариация значений  $x_j$  в выборке, тем меньше дисперсия  $\hat{\beta}_j$ .
- Чем лучше признак  $x_j$  объясняется линейной комбинацией оставшихся признаков, тем больше дисперсия  $\hat{\beta}_j$ .

# Дисперсия $\hat{\beta}_j$

$R_j^2 < 1$  по предположению (3); тем не менее, может быть  $R_j^2 \approx 1$ .

В матричном виде:

$$\mathbb{D}(\hat{\beta}) = \sigma^2 (X^T X)^{-1}.$$

Если столбцы  $X$  почти линейно зависимы, то матрица  $X^T X$  плохо обусловлена, и дисперсия оценок  $\hat{\beta}_j$  велика.

Близкая к линейной зависимость между  $x_j$  — **мультиколлинеарность**.

# Нормальность

- 1 Линейность отклика
- 2 Случайность выборки
- 3 Полнота ранга  $X$
- 4 Случайность ошибок
- 5 Гомоскедастичность ошибок
- 6 Нормальность ошибок:

$$\varepsilon | x \sim N(0, \sigma^2)$$

Эквивалентная запись:  $y | x \sim N(x\beta, \sigma^2)$

# Нормальность

- ❶ Линейность отклика
- ❷ Случайность выборки
- ❸ Полнота ранга  $X$
- ❹ Случайность ошибок
- ❺ Гомоскедастичность ошибок
- ❻ Нормальность ошибок

⇒ МНК-оценки совпадают с оценками максимального правдоподобия

# Нормальность

(1)-(6)  $\Rightarrow$  МНК-оценки совпадают с оценками максимального правдоподобия  $\Rightarrow$

- имеют наименьшую дисперсию среди всех несмещённых оценок  $\beta$
- имеют нормальное распределение  $N\left(\beta, \sigma^2 (X^T X)^{-1}\right)$
- $\hat{\sigma}^2 = \frac{1}{n-k-1} \text{RSS}$  — несмещённая оценка  $\sigma^2$ , и

$$\frac{\text{RSS}}{\sigma^2} \sim \chi_{n-k-1}^2$$

- $\forall c \in \mathbb{R}^{k+1}$

$$\frac{c^T (\beta - \hat{\beta})}{\hat{\sigma} \sqrt{c^T (X^T X)^{-1} c}} \sim St(n - k - 1)$$

# Последствия

В предположениях (1)-(6) можно строить:

- доверительные для  $\beta_j$
- доверительные интервалы для  $\mathbb{E}(y|x)$
- предсказательные интервалы для  $y|x$



## Доверительные и предсказательные интервалы

- $100(1 - \alpha)\%$  доверительный интервал для  $\sigma^2$ :

$$\frac{\text{RSS}}{\chi_{n-k-1, 1-\alpha/2}^2} \leq \sigma^2 \leq \frac{\text{RSS}}{\chi_{n-k-1, \alpha/2}^2}$$

- Возьмём  $c = \begin{pmatrix} 0 \dots 0 & 1 & 0 \dots 0 \\ & j \end{pmatrix}$ ;  $100(1 - \alpha)\%$  доверительный интервал для  $\beta_j$ :

$$\hat{\beta}_j \pm t_{n-k-1, 1-\alpha/2} \hat{\sigma} \sqrt{(X^T X)^{-1}_{jj}}$$

- Для нового объекта  $x_0$  возьмём  $c = x_0$ ;  $100(1 - \alpha)\%$  доверительный интервал для  $\mathbb{E}(y | x = x_0)$ :

$$x_0^T \hat{\beta} \pm t_{n-k-1, 1-\alpha/2} \hat{\sigma} \sqrt{x_0^T (X^T X)^{-1} x_0}$$

- Чтобы построить предсказательный интервал для  $y(x_0) = x_0^T \beta + \varepsilon(x_0)$ , учтём ещё дисперсию ошибки:

$$x_0^T \hat{\beta} \pm t_{n-k-1, 1-\alpha/2} \hat{\sigma} \sqrt{1 + x_0^T (X^T X)^{-1} x_0}$$

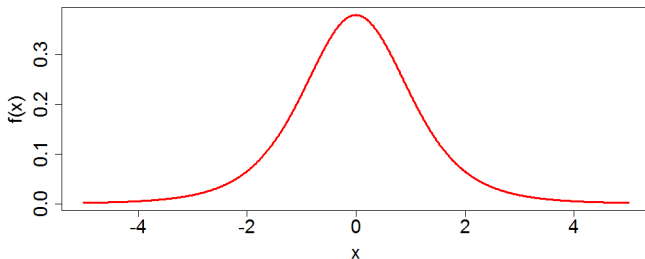
# t-критерий Стьюдента

нулевая гипотеза:  $H_0: \beta_j = 0$

альтернатива:  $H_1: \beta_j < \neq > 0$

статистика:  $T = \frac{\hat{\beta}_j}{\sqrt{\frac{RSS}{n-k-1} (X^T X)^{-1}_{jj}}}$

нулевое распределение:  $St(n - k - 1)$



# t-критерий Стьюдента

**Пример:** 12 испытуемых,  $x$  — результат прохождения испытуемым составного теста скорости реакции,  $y$  — результат его теста на симулятора транспортного средства. Проведение составного теста значительно проще и требует меньших затрат, поэтому ставится задача предсказания  $y$  по  $x$ ; строится линейная регрессия

$$y = \beta_0 + \beta_1 x + \varepsilon.$$

Значима ли переменная  $x$  для предсказания  $y$ ?

$$H_0: \beta_1 = 0$$

$$H_1: \beta_1 \neq 0 \Rightarrow p = 2.2021 \times 10^{-5}.$$

# Критерий Фишера

$$X_{n \times (k+1)} = \begin{pmatrix} X_1 & X_2 \\ n \times (k+1-k_1) & n \times k_1 \end{pmatrix}$$

нулевая гипотеза:

альтернатива:

статистика:

нулевое распределение:

$$\beta^T_{(k+1) \times 1} = \begin{pmatrix} \beta_1^T_{(k+1-k_1) \times 1} & \beta_2^T_{k_1 \times 1} \end{pmatrix}^T$$

$$H_0: \beta_2 = 0$$

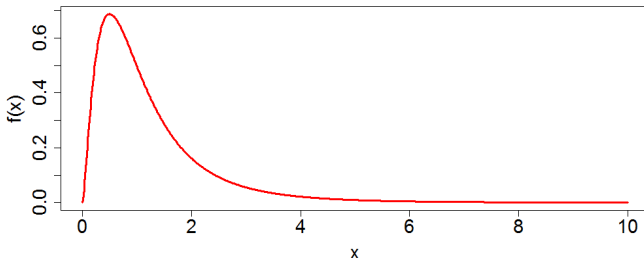
$$H_1: H_0 \text{ неверна}$$

$$RSS_r = \|y - X_1 \beta_1\|_2^2$$

$$RSS_{ur} = \|y - X \beta\|_2^2$$

$$F = \frac{(RSS_r - RSS_{ur})/k_1}{RSS_{ur}/(n-k-1)}$$

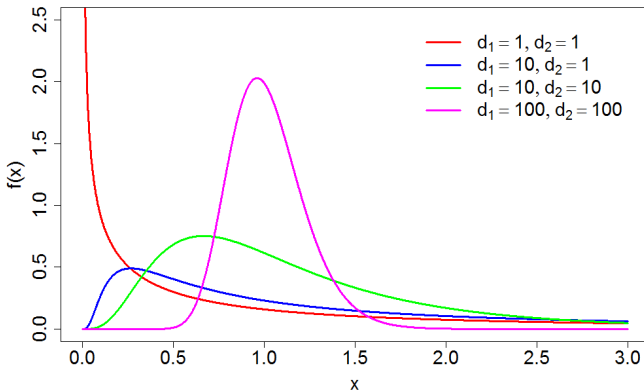
$$F(k_1, n - k - 1)$$



# Критерий Фишера

$X_1 \sim \chi_{d_1}^2, X_2 \sim \chi_{d_2}^2$  независимы,

$X = \frac{X_1/d_1}{X_2/d_2} \sim F(d_1, d_2)$  — распределение Фишера с  $d_1, d_2$  степенями свободы.



# Критерий Фишера

**Пример:** по данным о 1191 детей построена модель:

$$weight = \beta_0 + \beta_1 cigs + \beta_2 parity + \beta_3 inc + \beta_4 med + \beta_5 fed + \varepsilon$$

*weight* — вес ребёнка при рождении,

*cigs* — среднее число сигарет за один день беременности,

*parity* — номер ребёнка у матери,

*inc* — среднемесячный доход семьи,

*med* — длительность получения образования матерью, *fed* — отцом.

Зависит ли вес ребёнка при рождении от уровня образования родителей?

$$H_0: \beta_4 = \beta_5 = 0.$$

$H_1: H_0$  неверна.

Критерий Фишера:  $p = 0.2421$ .

## Критерии Фишера и Стьюдента

- При  $k_1 = 1$  критерий Фишера эквивалентен критерию Стьюдента для двусторонней альтернативы
- Иногда критерий Фишера отвергает гипотезу незначимости признаков  $X_2$ , а критерий Стьюдента не признаёт значимым ни один из них. Возможные объяснения:
  - отдельные признаки из  $X_2$  недостаточно хорошо объясняют  $y$ , но совокупный эффект значим
  - признаки в  $X_2$  мультиколлинеарны
- Иногда критерия Фишера не отвергает гипотезу незначимости признаков  $X_2$ , а критерий Стьюдента признаёт значимыми некоторые из них. Возможные объяснения:
  - незначимые признаки в  $X_2$  маскируют влияние значимых
  - значимость отдельных признаков в  $X_2$  — результат множественной проверки гипотез

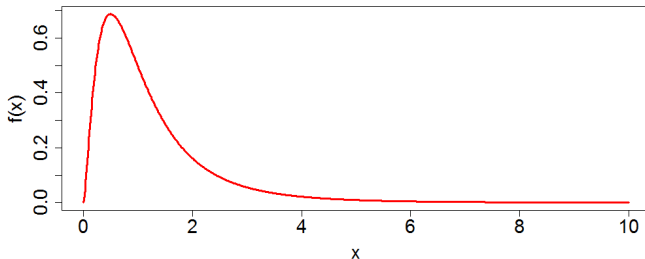
# Критерий Фишера

нулевая гипотеза:  $H_0: \beta_1 = \dots = \beta_k = 0$

альтернатива:  $H_1: H_0$  неверна

статистика:  $F = \frac{R^2/k}{(1-R^2)/(n-k-1)}$

нулевое распределение:  $F(k, n - k - 1)$





# Критерий Фишера

**Пример:** имеет ли вообще смысл модель веса ребёнка при рождении, рассмотренная выше?

$$H_0: \beta_1 = \dots = \beta_5 = 0.$$

$H_1: H_0$  неверна.

Критерий Фишера:  $p = 6 \times 10^{-9}$ .

## Отбор признаков

Незначимые признаки можно исключать из модели — доказательств тому, что они влияют на  $y$ , нет!

**Недоопределение:** если зависимая переменная определяется моделью

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{j-1} x_{j-1} + \beta_j x_j + \beta_{j+1} x_{j+1} + \cdots + \beta_k x_k + \varepsilon,$$

а вместо этого используется модель

$$y = \beta_0 + \beta_1 x_1 + \cdots + \beta_{j-1} x_{j-1} + \beta_{j+1} x_{j+1} + \cdots + \beta_k x_k + \varepsilon,$$

то МНК-оценки  $\hat{\beta}_0, \dots, \hat{\beta}_{j-1}, \hat{\beta}_{j+1}, \dots, \hat{\beta}_k$  являются смещёнными и несостоятельными оценками  $\beta_0, \dots, \beta_{j-1}, \beta_{j+1}, \dots, \beta_k$ .

**Переопределение:** если признак  $x_j$  не влияет на  $y$ , т.е.  $\beta_j = 0$ , то МНК-оценка  $\hat{\beta}$  остаётся несмещённой состоятельной оценкой  $\beta$ , но дисперсия её возрастает.

⇒ исключая незначимые признаки, мы рискуем получить смещённые оценки оставшихся коэффициентов, но уменьшаем их дисперсию.

# Best subset

Для каждого  $k$  полным перебором по  $R^2$  можно выбрать лучшую модель с  $k$  признаками, затем по  $R_a^2$  среди них можно выбрать одну лучшую модель.

Полный перебор требует больших вычислительных затрат.

# Пошаговая регрессия

- **Шаг 0.** Настраивается модель с одной только константой, а также все модели с одной переменной. Рассчитывается  $F$ -статистика каждой модели и достигаемый уровень значимости. Выбирается модель с наименьшим достигаемым уровнем значимости. Соответствующая переменная  $X_{e1}$  включается в модель, если этот достигаемый уровень значимости меньше порогового значения  $p_E = 0.05$ .
- **Шаг 1.** Рассчитывается  $F$ -статистика и достигаемый уровень значимости для всех моделей, содержащих две переменные, одна из которых  $X_{e1}$ . Аналогично принимается решение о включении  $X_{e2}$ .
- **Шаг 2.** Если была добавлена переменная  $X_{e2}$ , возможно,  $X_{e1}$  уже не нужна. В общем случае просчитываются все возможные варианты исключения одной переменной, рассматривается вариант с наибольшим достигаемым уровнем значимости, соответствующая переменная исключается, если он превосходит пороговое значение  $p_R = 0.1$ .
- ...

## Эксперимент Фридмана

(Freedman, 1983): best subset и пошаговая регрессия несовместимы с проверкой гипотез о значимости коэффициентов: критерии Фишера и Стьюдента антиконсервативны, если вычисляются на той же самой выборке, на которой настраивалась модель.

Если мы хотим считать значимость признаков, признаки должны отбираться не слишком интенсивно (или на другой выборке).

## Проверка предположений регрессии

- 1 Линейность отклика
- 2 Случайность выборки
- 3 Полнота ранга  $X$
- 4 Случайность ошибок
- 5 Гомоскедастичность ошибок
- 6 Нормальность ошибок

# 1. Линейность отклика

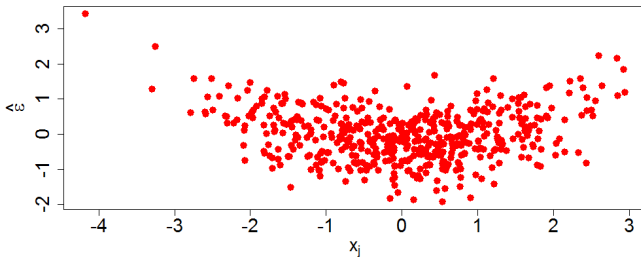
$$y = X\beta + \varepsilon$$

В точности не выполняется никогда — все модели неверны.

Чтобы убедиться в отсутствии больших отклонений от линейности, нужно анализировать остатки:

$$\hat{\varepsilon}_i = y_i - \hat{y}_i, \quad i = 1, \dots, n.$$

## 1. Линейность отклика



Стоит добавить квадрат признака  $x_j$



## 2. Случайность выборки

Наблюдения  $(x_i, y_i)$  независимы.

- Если наблюдения зависимы, дисперсия недооценивается, а критерии не работают
- Фильтровать выборку по признаку  $z$  можно только если  $\mathbb{E}(y | x, z) = \mathbb{E}(y | x)$

### 3. Полнота ранга

$$\text{rank } X = k + 1$$

- Если есть линейно зависимые признаки, то дисперсия оценки коэффициентов при них будет бесконечной
- Никакого one-hot encoding!

# Фиктивные переменные

Если признак  $x_j$  принимает  $m$  различных значений, то его нужно кодировать  $m - 1$  фиктивной переменной.

Пусть  $y$  — уровень заработной платы,  $x$  — должность.

	Dummy-кодирование	
Тип должности	$x_1$	$x_2$
рабочий	0	0
инженер	1	0
управляющий	0	1

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2$$

$\beta_1, \beta_2$  оценивают среднюю разницу в уровнях зарплат инженера и управляющего с рабочим.

## 4. Случайность ошибок

$$\mathbb{E}(\varepsilon | x) = 0$$

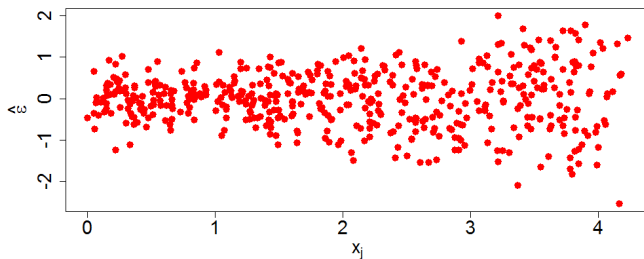
Гипотезу  $H_0: \mathbb{E}(\varepsilon | x) = 0$  можно проверить по остаткам критерием Стьюдента.

## 5. Гомоскедастичность ошибок

$$\mathbb{D}(\varepsilon | x) = \sigma^2$$

Проверка:

- визуальный анализ:



- критерий Бройша-Пагана

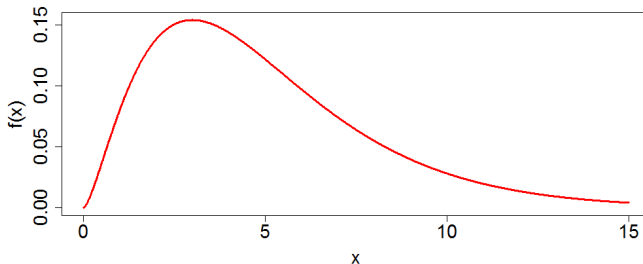
# Критерий Бройша-Пагана

нулевая гипотеза:  $H_0: \mathbb{D}\varepsilon = \sigma^2$

альтернатива:  $H_1: H_0$  неверна

статистика:  $LM = nR_{\varepsilon^2}^2$ ,  $R_{\varepsilon^2}^2$  — коэффициент детерминации при регрессии  $\hat{\varepsilon}^2$  на  $x$

нулевое распределение:  $\chi_k^2$



# Гетероскедастичность

Гетероскедастичность может быть следствием недоопределения модели.

Последствия гетероскедастичности:

- МНК-оценки  $\beta$  и  $R^2$  остаются несмещёнными и состоятельными
- нарушаются предположения критериев Стьюдента и Фишера и методов построения доверительных интервалов для  $\sigma$  и  $\beta$  (независимо от объёма выборки)

Варианты:

- переопределить модель, добавить признаки, преобразовать отклик
- использовать модифицированные оценки дисперсии коэффициентов

## Устойчивая оценка дисперсии Уайта

Если не удаётся избавиться от гетероскедастичности, при анализе моделей (далее) можно использовать устойчивые оценки дисперсии.

White's heteroscedasticity-consistent estimator (HCE):

$$\mathbb{D}(\hat{\beta} | X) = (X^T X)^{-1} (X^T \text{diag}(\hat{\varepsilon}_1^2, \dots, \hat{\varepsilon}_n^2) X) (X^T X)^{-1}.$$

Асимптотика устойчивой оценки:

$$\sqrt{n}(\beta - \hat{\beta}) \xrightarrow{d} N(0, \Omega),$$

$$\hat{\Omega} = n (X^T X)^{-1} (X^T \text{diag}(\hat{\varepsilon}_1^2, \dots, \hat{\varepsilon}_n^2) X) (X^T X)^{-1}.$$



## Другие устойчивые оценки дисперсии

Элементы диагональной матрицы могут задаваться разными способами:

const	$\hat{\sigma}^2$
HC0	$\hat{\varepsilon}_i^2$
HC1	$\frac{n}{n-k} \hat{\varepsilon}_i^2$
HC2	$\frac{\hat{\varepsilon}_i^2}{1-h_i}$
HC3	$\frac{\hat{\varepsilon}_i^2}{(1-h_i)^2}$
HC4	$\frac{\hat{\varepsilon}_i^2}{(1-h_i)^{\min\left(4, \frac{nh_i}{k}\right)}}$

const — случай гомоскедастичной ошибки,

HC0 — оценка Уайта,

HC1–HC3 — модификации МакКиннона-Уайта,

HC4 — модификация Крибари-Нето.

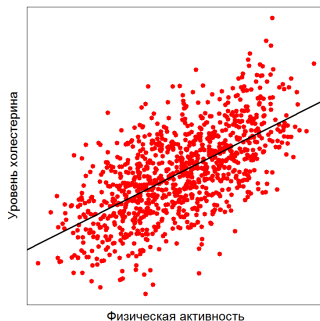
## 6. Нормальность ошибок

$$\varepsilon | x \sim N(0, \sigma^2)$$

Проверка:

- ку-ку график
- критерий Шапиро-Уилка

# Упражнения и холестерин



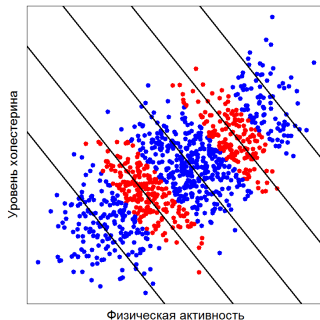
$$chol = \beta_0 + \beta_1 ex$$

$$H_0: \beta_1 = 0$$

$$H_0: \beta_1 > 0$$

Критерий Стьюдента:  $p = 2 \times 10^{-16}$ .

# Упражнения и холестерин



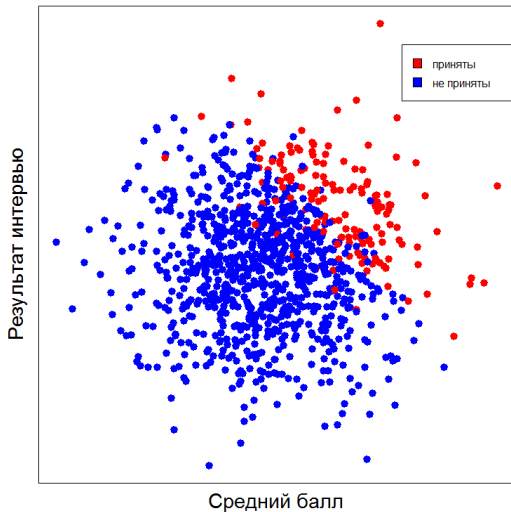
$$chol = \beta_0 + \beta_1 ex + \beta_2 age$$

$$H_0: \beta_1 = 0$$

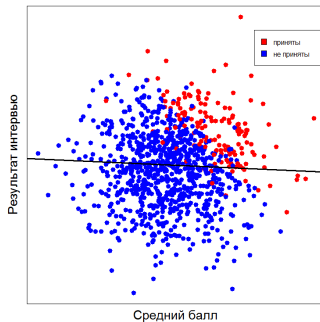
$$H_0: \beta_1 < 0$$

Критерий Стьюдента:  $p = 2 \times 10^{-16}$ .

## Средний балл и мотивация



# Средний балл и мотивация



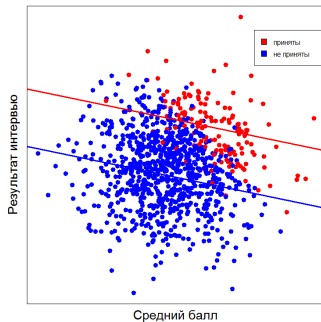
$$mot = \beta_0 + \beta_1 SAT$$

$$H_0: \beta_1 = 0$$

$$H_0: \beta_1 \neq 0$$

Критерий Стьюдента:  $p = 0.1452$ .

# Средний балл и мотивация



$$mot = \beta_0 + \beta_1 SAT + \beta_2 acc$$

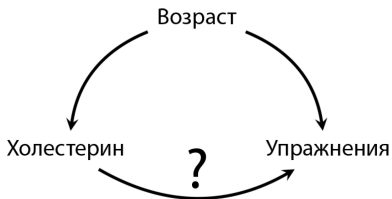
$$H_0: \beta_1 = 0$$

$$H_0: \beta_1 \neq 0$$

Критерий Стьюдента:  $p = 2 \times 10^{-16}$ .

# В чём разница?

Вилка:



Коллайдер:





## Причинно-следственная связь

$\hat{\beta}_1$  — оценка среднего эффекта от увеличения  $x_1$  на единицу, если среди  $x_2, \dots, x_k$ :

- содержатся все признаки, являющиеся причинами  $x_1$
- не содержится признаков, являющихся следствиями одновременно  $x_1$  и  $y$
- иногда регрессия позволяет обнаруживать причинно-следственные связи!
- плохо подобранные признаки могут привести к противоположным выводам

# Литература

- линейная регрессия в целом — Wooldridge (много примеров, без матричной алгебры);
- преобразование Бокса-Кокса (Box-Cox transformation) — Дрейпер, гл. 14;
- устойчивые оценки дисперсии — White;
- расстояние Кука (Cook's distance) — Cook.

Дрейпер Н.Р., Смит Г. *Прикладной регрессионный анализ*, 2007.

Cook D.R., Weisberg S. *Residuals and influence in regression*, 1982.

White H. (1980). *A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity*. *Econometrica: Journal of the Econometric Society*, 48(4), 817–838.

Wooldridge J. *Introductory Econometrics: A Modern Approach*, 2016.