ECONOMETRICS
Александр Левкун

Содержание

СМF

Есть КЛММР — классическая линейная модель множественной регрессии.

А есть ОЛММР — обобщенная линейная модель множественной регрессии — общий вид Ω (нарушается условие гомоскедастичности и некоррелированных ошибок).

$$\widehat{\beta}_{GLS} = (X'\Omega^{-1}X)^{-1}X'\Omega^{-1}y$$

Оценка, полученная методом ОМНК (GLS):

1. несмещенная;
2. состоятельная;
3. распределена нормально при нормальном распределении ошибок;
4. эффективная в классе несмещенных линейных оценок.

$$\mathbb{V}ar(\hat{\beta}_{OLS}) = (X'X)^{-1}X'\Omega X(X'X)^{-1}$$

Нужно уметь эффективно оценивать Ω!

В КЛММР:

$$\widehat{R}^2 = 1 - \frac{\left(y - X\hat{\beta}_{GLS}\right)'\left(y - X\hat{\beta}_{GLS}\right)}{(y - \bar{y}I)'(y - \bar{y}I)}$$

Но такая оценка R^2 является смещенной. Поэтому используем несмещенную оценку:

$$\widehat{R}^2_{adj} = 1 - \frac{\left(y - X\hat{\beta}_{GLS}\right)'\left(y - X\hat{\beta}_{GLS}\right)/(n-k)}{(y - \bar{y}I)'(y - \bar{y}I)/(n-1)}$$

Про ограниченную модель. Выдвигается какая-то гипотеза о коэффициентах: в матричном виде $H_0 : H\beta = r$.

F-статистика:

$$F = \frac{(H\hat{\beta} - r)'(H(X'X)^{-1}H')^{-1}(H\hat{\beta} - r)/q}{RSS/(n-k)}$$

Пример:

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

$$H_0 = \begin{cases} \beta_1 = 1 \\ \beta_2 = \beta_3 \end{cases}$$

Матрица H и вектор r здесь:

$H = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & -1 \end{pmatrix}$

$r = \begin{pmatrix} 1 \\ 0 \end{pmatrix}$

Проверка гипотезы: $H_0 : \beta_2 = ... = \beta_k = 0$.

$$F = \frac{R^2}{1 - R^2}\frac{n-k}{k-1}$$

При верной H_0: F-статистика имеет распределение Фишера с $k-1$ и $n-k$ степенями свободы.

Если хотим проверить, что q коэффициентов незначимы, то:

$$F = \frac{(R^2_{UR} - R^2_R)/q}{(1 - R^2_{UR})/(n-k)}$$

$$F = \frac{(RSS_R - RSS_{UR})/q}{RSS_{UR}/(n-k)}$$

Ну и оценка дисперсии:

$$\hat{\sigma}^2 = \frac{1}{n-k}\hat{\varepsilon}'\hat{\varepsilon}$$

Семинар. November, 24

Перед тем, как переходить к стохастическим регрессорам, для начала вернемся к теории вероятности:

Y — одна случайная величина; X — одна случайная величина.

Вспоминаем концепцию **условного математического ожидания**: $\mathbb{E}(Y|X)$ — случайная величина.

$$\mathbb{E}(Y|X) = \mathbb{E}(Y|X = x_i) \text{ if } X = x_i$$

Основные свойства:

- $\mathbb{E}(a|X) = a$
- $\mathbb{E}(aY|X) = a\mathbb{E}(Y|X)$
- $\mathbb{E}(f(X)|X) = f(X)$
- $\mathbb{E}\left(\mathbb{E}(Y|X)\right) = \mathbb{E}(Y)$

Задача 1. Задача на пуассоновский поток ($\lambda = 2$):

день 1: Вася ждет T до 1 автобуса

день 2: Вася ждет время T. N — кол-во автобусов за время T.

Найти: $\mathbb{E}(N)$; $\mathbb{E}(N^2)$; Правда ли, что N распределено по Пуассону?

Решение:

$$\mathbb{E}(N) = \mathbb{E}\left(\mathbb{E}(N|T)\right)$$

$$N|T \sim Poiss(2 \cdot T)$$

Учитывая то, что для пуассоновского потока $T \sim Exp(2)$:

$$\mathbb{E}(N) = \mathbb{E}\left(\mathbb{E}(N|T)\right) = \mathbb{E}(2T) = 2 \cdot \frac{1}{2} = 1$$

$$\mathbb{E}(N^2) = \mathbb{E}\left(\mathbb{E}(N^2|T)\right) = \mathbb{E}\left(\mathbb{V}ar(N|T) + (\mathbb{E}(N|T))^2\right) = \mathbb{E}\left(2T + 4T^2\right) =$$
$$= 2\mathbb{E}T + 4\mathbb{V}ar(T) + 4\left(\mathbb{E}T\right)^2 = 1 + 1 + 1 = 3$$

Правда ли, что N распределено по Пуассону? Нет! Так как $\mathbb{V}ar = 3 - 1 = 2 \neq \mathbb{E}(N)$.

Задача 2. Задача про грибы:

n грибов. Вероятности найти рыжик R, лисичку L или другой гриб соответственно равны p_r, p_l, $1 - p_r - p_l$.

Найти: $\mathbb{E}(R|L)$; $\mathbb{P}(R = 0|L)$; $\mathbb{P}(\mathbb{E}(R|L) = 0)$; $\mathbb{E}\left[(\mathbb{E}(R|L))^2\right]$.

Решение:

$$R|L \sim Bin\left(n - L, \frac{p_r}{1 - p_l}\right)$$

$$\mathbb{E}(R|L) = (n - L)\frac{p_r}{1 - p_l}$$

$$\mathbb{P}(R=0|L) = \left(1 - \frac{p_r}{1-p_l}\right)^{n-L}$$

$$\mathbb{P}(\mathbb{E}(R|L)=0) = \mathbb{P}(L=n) = p_l^n$$

$$\mathbb{E}\left[(\mathbb{E}(R|L))^2\right] = \mathbb{E}(Y^2) = \mathbb{V}ar(Y) + (\mathbb{E}(Y))^2 = \mathbb{V}ar\left((n-L)\frac{p_r}{1-p_l}\right) + (\mathbb{E}(R))^2 =$$

$$= \left(\frac{p_r}{1-p_l}\right)^2 \mathbb{V}ar(L) + n^2p_r^2 = \left(\frac{p_r}{1-p_l}\right)^2 np_l(1-p_l) + n^2p_r^2 = np_r^2\left(\frac{p_l}{1-p_l} + n\right)$$

Условная дисперсия:

$$\mathbb{V}ar(Y|X) = \mathbb{E}(Y^2|X) - (\mathbb{E}(Y|X))^2$$

Основные свойства условной дисперсии:

- $\mathbb{V}ar(aY|X) = a^2\mathbb{V}ar(Y|X)$
- $\mathbb{V}ar(f(X)|X) = 0$
- $\mathbb{V}ar(XY|X) = X^2\mathbb{V}ar(Y|X)$

Теорема Пифагора (by definition $|y|^2 = \mathbb{V}ar(y)$, скалярное произведение — $(x,y) = \mathbb{C}ov(x,y)$, критерий перпендикулярности: $\mathbb{C}ov(x,y) = 0$):

$$\mathbb{V}ar(y) = \mathbb{E}\left(\mathbb{V}ar(y|x)\right) + \mathbb{V}ar\left(\mathbb{E}(y|x)\right)$$

$\mathbb{C}ov(y - \mathbb{E}(y|x), \mathbb{E}(y|x)) = 0$ доказывается из теории меры.

Определение:

$\mathbb{E}(y|x)$ — это такая с.в. $\hat{y}$, что $\hat{y} = f(x)$ (борелевская), что: $\mathbb{E}(\hat{y}) = \mathbb{E}(y)$, $\mathbb{C}ov(y, g(x)) = \mathbb{C}ov(\hat{y}, g(x))$ (или $\mathbb{C}ov(y - \hat{y}, g(x)) = 0$), т.е. y и $\hat{y}$ неотличимы в смысле мат. ожиданий и ковариаций.

Задача 3. X_1, X_2, ..., $X_{100} \sim$ i.i.d $U[0,1]$

$M = \max\{X_1, ..., X_{100}\}$

$L = \max\{X_1, ..., X_{80}\}$

$R = \max\{X_{81}, ..., X_{100}\}$

Найти: $\mathbb{P}(L > R|L)$, $\mathbb{P}(L > R|M)$, $\mathbb{P}(L > R|R)$, $\mathbb{P}(L > R|L, M)$, $\mathbb{E}(M|L)$, $\mathbb{E}(L|M)$.

Решение:

$$\mathbb{P}(L > R|L) = \mathbb{P}(X_{81} < L|L) \cdot \mathbb{P}(X_{82} < L|L) \cdot ... \cdot \mathbb{P}(X_{100} < L|L) = F^{20}(L) = L^{20}$$

$$\mathbb{P}(L > R|R) = 1 - \mathbb{P}(L < R|R) = 1 - \mathbb{P}(X_1 < R|R) \cdot ... \cdot \mathbb{P}(X_{80} < R|R) = 1 - F^{80}(R) = 1 - R^{80}$$

Т.к. $M = \max\{L, R\}$, $\mathbb{P}(M = L) = 0.8$, $\mathbb{P}(M = R) = 0.2$, то:

$$\mathbb{P}(L > R|M) = \frac{4}{5}M^{20} + \frac{1}{5}\left(1 - M^{80}\right)$$

$$\mathbb{P}(L > R|L, M) = \mathbf{1}_{L=M}$$

Заметим, что $F_R(x) = x^{20}$, тогда $f_R(x) = 20x^{19}$. Отсюда можем найти математическое ожидание случайной величины R (этот же результат можно получить из соображений симметрии):

$$\mathbb{E}(R) = \int\limits_0^1 20x^{20}dx = \frac{20}{21}$$

$$\mathbb{E}(M|L) = \frac{4}{5}L + \frac{1}{5} \cdot \frac{20}{21} = \frac{4}{5}L + \frac{4}{21}$$

Учитывая то, что $\mathbb{E}(L|M \neq L) = 80M/81$ (из соображений симметрии):

$$\mathbb{E}(L|M) = \frac{4}{5}M + \frac{1}{5} \cdot \frac{80}{81}M = \frac{404}{405}M$$

Семинар. December, 1

Последовательность оценок $\hat{\theta}_1$, $\hat{\theta}_2$, $\hat{\theta}_3$, ... называется **состоятельной** для параметра θ, если при $n \to \infty$:

$$\mathbb{P}\left(|\hat{\theta}_n - \theta| > \varepsilon\right) \to 0$$

Обозначение состоятельности: $\text{plim}\hat{\theta}_n = \theta$ или $\hat{\theta}_n \to \theta$

Задача 1. a) X — детерменированная. Оценивается модель парной регрессии:

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

$$x_i = \begin{cases} 1, & \text{при } i = 1; \\ 2, & \text{при } i \geq 2. \end{cases}$$

$n = 3, 4, 5, 6, ...$

Верно ли, что $\text{plim}_{n\to\infty}\hat{\beta}_2^n = \beta_2$ (n здесь — не степень, а просто индекс)?

$X = \begin{pmatrix} 1 & 1 \\ 1 & 2 \\ \vdots & \vdots \\ 1 & 2 \end{pmatrix}$

$X'X = \begin{pmatrix} n & 2n-1 \\ 2n-1 & 4n-3 \end{pmatrix}$

$(X'X)^{-1} = \dfrac{1}{n-1}\begin{pmatrix} 4n-3 & -2n+1 \\ -2n+1 & n \end{pmatrix}$

Так как $\mathbb{V}ar\hat{\beta} = \sigma^2 (X'X)^{-1}$, то: $\lim\limits_{n\to\infty} \mathbb{V}ar\left(\hat{\beta}^n\right) = \sigma^2 \begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix}$

Итак, $\mathbb{V}ar\left(\hat{\beta}_2^n\right) = \sigma^2$, $\mathbb{V}ar\left(\hat{\beta}_1^n\right) = 4\sigma^2$. Достаточное условие не выполняется.

Достаточными условиями состоятельности являются:

1. $\mathbb{V}ar(\hat{\theta}_n) \to 0$
2. $\mathbb{E}(\hat{\theta}_n) = \theta$

Пример, когда без выполнения условий Чебышева выполняется состоятельность:

$\hat{\theta}_n$	$\theta - 100^n$	θ	$\theta + 100^n$
$\mathbb{P}(\hat{\theta}_n)$	0.1^n	$1 - 2 \cdot 0.1^n$	0.1^n

Оценка состоятельна, но $\mathbb{V}ar(\hat{\theta}_n) \to \infty$.

Докажем все же по-честному (выше мы просто показали, что достаточное условие состоятельности не выполняется):

$X'\varepsilon = \begin{pmatrix} \varepsilon_1 + \varepsilon_2 + ... + \varepsilon_n \\ \varepsilon_1 + 2\varepsilon_2 + ... + 2\varepsilon_n \end{pmatrix}$

Тогда:

$\lim\limits_{n\to\infty} (X'X)^{-1} X'\varepsilon = \begin{pmatrix} 4 & -2 \\ -2 & 1 \end{pmatrix} X'\varepsilon = \begin{pmatrix} 2\varepsilon_1 \\ -\varepsilon_1 \end{pmatrix}$

$\hat{\beta} = \beta + (X'X)^{-1} X'\varepsilon$
$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \to \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} 2\varepsilon_1 \\ -\varepsilon_1 \end{pmatrix}$
Из последнего следует несостоятельность оценок коэффициентов.
б) Теперь:

$$x_i = \begin{cases} 1, & \text{при } i = 2k+1; \\ 2, & \text{при } i = 2k. \end{cases}$$

$X'X = \begin{pmatrix} n & 3n/2 \\ 3n/2 & 5n/2 \end{pmatrix}$
$(X'X)^{-1} = \dfrac{4}{n^2} \begin{pmatrix} 5n/2 & -3n/2 \\ -3n/2 & n \end{pmatrix}$
$\lim\limits_{n\to\infty} \mathbb{V}ar\left(\hat{\beta}^n\right) = \sigma^2 \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}$
Оценки — состоятельны по достаточным условиям.
в) Теперь $x_i = i$.
Так как $1^2 + 2^2 + ... + n^2 = n(n+1)(2n+1)/6$ (доказывается по мат. индукции), то:
$X'X = \begin{pmatrix} n & n(n+1)/2 \\ n(n+1)/2 & n(n+1)(2n+1)/6 \end{pmatrix}$
Находим определитель:

$$det\,(X'X) = \frac{n^2(n+1)(2n+1)}{6} - \frac{n^2(n+1)^2}{4} =$$

$$n^2(n+1)\frac{4n+2-3n-3}{12} = \frac{n(n+1)(n-1)}{12}$$

$(X'X)^{-1} = \dfrac{12}{n(n+1)(n-1)} \begin{pmatrix} n(n+1)(2n+1)/6 & -n(n+1)/2 \\ -n(n+1)/2 & n \end{pmatrix} =$
$= \begin{pmatrix} 2(2n+1)/(n-1) & -6/(n-1) \\ -6/(n-1) & 12/(n+1)/(n-1) \end{pmatrix}$
Значит: $\lim\limits_{n\to\infty} \mathbb{V}ar\left(\hat{\beta}^n\right) = \sigma^2 \begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix}$
Достаточно свойство не выполняется. Докажем снова по-честному:
$X'\varepsilon = \begin{pmatrix} \varepsilon_1 + \varepsilon_2 + ... + \varepsilon_n \\ \varepsilon_1 + 2\varepsilon_2 + ... + n\varepsilon_n \end{pmatrix}$
$\lim\limits_{n\to\infty} (X'X)^{-1} X'\varepsilon = \begin{pmatrix} 4 & 0 \\ 0 & 0 \end{pmatrix} X'\varepsilon = \begin{pmatrix} 4\sum \varepsilon_i \\ 0 \end{pmatrix}$
$\hat{\beta} = \beta + (X'X)^{-1} X'\varepsilon$
$\begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \end{pmatrix} \to \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} + \begin{pmatrix} 4\sum \varepsilon_i \\ 0 \end{pmatrix}$
Видим, что оценка коэффициента β_2 — состоятельна, однако оценка коэффициента β_1 не является состоятельной.

БСХС (Большой Список Хороших Свойств):
Если:

- $y = X\beta + \varepsilon$

- $\mathbb{E}(\varepsilon|X) = 0$
- $\mathbb{V}ar(\varepsilon|X) = \sigma^2 I_n$
- $P(X$ полного ранга$) = 1$
- векторы $(x_{i2}, x_{i3}, ..., x_{ik}, y_i)$ — i.i.d.
- $n > k$
- строится регрессия y на X

то:

1. Базовые:
   - $\hat{\beta}$ — линейные по y: $\hat{\beta} = (X'X)^{-1} X'y$
   - $\mathbb{E}(\hat{\beta}|X) = \beta$
   - $\mathbb{E}(\hat{\beta}) = \beta$
   - условная эффективность среди линейных несмещенных оценок: $\mathbb{V}ar\left(\hat{\beta}_j^{OLS}|X\right) \leqslant \mathbb{V}ar\left(\hat{\beta}_j^{alt}|X\right)$
   - безусловная эффективность среди линейных несмещенных оценок (следует из условной из теоремы Пифагора): $\mathbb{V}ar\left(\hat{\beta}_j^{OLS}\right) \leqslant \mathbb{V}ar\left(\hat{\beta}_j^{alt}\right)$
   - $\mathbb{E}(\hat{\sigma^2}) = \sigma^2$
2. Асимптотические:
   - $\text{plim}\hat{\sigma^2} = \sigma^2$
   - $\text{plim}\hat{\beta} = \beta$
   - $t \sim N(0, 1)$
   - $qF \sim \chi^2_q$ (из restricted and unrestricted models)
3. При дополнительной предпосылке $\varepsilon|X \sim N(0; \sigma^2 I_n)$:
   - $t|X \sim t_{n-k}$
   - $t \sim t_{n-k}$
   - $F|X \sim F_{q,n-k}$
   - $F \sim F_{q,n-k}$
   - $\hat{\sigma^2}(n-k)/\sigma^2|X \sim \chi^2_{n-k}$
   - $\hat{\sigma^2}(n-k)/\sigma^2 \sim \chi^2_{n-k}$

Семинар. December, 8

Setup: Регрессоры стохастические, **гетероскедастичность** $\mathbb{V}ar(\varepsilon|X) \neq \sigma^2$

В случайных выборках всегда присутствует гетероскедастичность.
Если есть гетероскедастичность (а она, скорее, есть!), то:

1. $\mathbb{E}(\hat{\beta}) = \beta$
2. $\mathbb{V}ar(\hat{\beta}|X) \neq \sigma^2 (X'X)^{-1}$

$$\mathbb{V}ar(\varepsilon|X) = \mathbb{V}ar(y|X) = \Omega$$
$$\mathbb{V}ar(\hat{\beta}|X) = \mathbb{V}ar\left((X'X)^{-1}X'y|X\right) = (X'X)^{-1}X'\Omega X(X'X)^{-1}$$
$$\widehat{\mathbb{V}ar}(\hat{\beta}|X) = (X'X)^{-1}X'\widehat{\Omega}X(X'X)^{-1}$$

Уайт предложил оценки для Ω: HC0, HC1, HC2, ...
Подгружаем пакеты:

```
library("lmtest") # тест Бройша-Пагана

## Loading required package: zoo
##
## Attaching package: 'zoo'
##
## The following objects are masked from 'package:base':
##
##     as.Date, as.Date.numeric

library("sandwich") # оценка дисперсии для гетероскедастичности
library("ggplot2")
library("hexbin")
```

```
str(diamonds) # data set, встроенный в пакет ggplot2

## 'data.frame': 53940 obs. of  10 variables:
##  $ carat  : num  0.23 0.21 0.23 0.29 0.31 0.24 0.24 0.26 0.22 0.23 ...
##  $ cut    : Ord.factor w/ 5 levels "Fair"<"Good"<..: 5 4 2 4 2 3 3 3 1 3 ...
##  $ color  : Ord.factor w/ 7 levels "D"<"E"<"F"<"G"<..: 2 2 2 6 7 7 6 5 2 5 ...
##  $ clarity: Ord.factor w/ 8 levels "I1"<"SI2"<"SI1"<..: 2 3 5 4 2 6 7 3 4 5 ...
##  $ depth  : num  61.5 59.8 56.9 62.4 63.3 62.8 62.3 61.9 65.1 59.4 ...
##  $ table  : num  55 61 65 58 58 57 57 55 61 61 ...
##  $ price  : int  326 326 327 334 335 336 336 337 337 338 ...
##  $ x      : num  3.95 3.89 4.05 4.2 4.34 3.94 3.95 4.07 3.87 4 ...
##  $ y      : num  3.98 3.84 4.07 4.23 4.35 3.96 3.98 4.11 3.78 4.05 ...
##  $ z      : num  2.43 2.31 2.31 2.63 2.75 2.48 2.47 2.53 2.49 2.39 ...

model <- lm(log(price) ~ log(carat), data = diamonds)
```

```r
summary(model)

##
## Call:
## lm(formula = log(price) ~ log(carat), data = diamonds)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -1.50833 -0.16951 -0.00591  0.16637  1.33793
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept) 8.448661   0.001365  6190.9   <2e-16 ***
## log(carat)  1.675817   0.001934   866.6   <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.2627 on 53938 degrees of freedom
## Multiple R-squared:  0.933,Adjusted R-squared:  0.933
## F-statistic: 7.51e+05 on 1 and 53938 DF,  p-value: < 2.2e-16

vcov(model)

##              (Intercept)   log(carat)
## (Intercept) 1.862382e-06 1.477020e-06
## log(carat)  1.477020e-06 3.739604e-06

vcovHC(model)

##              (Intercept)   log(carat)
## (Intercept) 2.175479e-06 1.946888e-06
## log(carat)  1.946888e-06 4.112166e-06

vcovHC(model, type = "HC5") # HC5 очень похож на HC3, т.к. много наблюдений

##              (Intercept)   log(carat)
## (Intercept) 2.175367e-06 1.946783e-06
## log(carat)  1.946783e-06 4.111931e-06
```

$$\widehat{\mathbb{V}ar}(\hat{\beta}|X) = \frac{RSS}{n-k}(X'X)^{-1}$$

Heteroscedasticity consistent (HC):

$$\widehat{\mathbb{V}ar}(\hat{\beta}|X) = (X'X)^{-1}X'\widehat{\Omega}X(X'X)^{-1}$$

Следующие команды использут не HC оценки:

```r
coeftest(model)

##
```

```
## t test of coefficients:
##
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 8.4486607  0.0013647 6190.90 < 2.2e-16 ***
## log(carat)  1.6758167  0.0019338  866.59 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
confint(model)
```

```
##                2.5 %   97.5 %
## (Intercept) 8.445986 8.451336
## log(carat)  1.672026 1.679607
```

Поэтому указываем, какую ковариационную матрицу брать:

```
coeftest(model, vcov = vcovHC(model))
```

```
##
## t test of coefficients:
##
##              Estimate Std. Error t value  Pr(>|t|)
## (Intercept) 8.4486607  0.0014750  5728.1 < 2.2e-16 ***
## log(carat)  1.6758167  0.0020278   826.4 < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
confint(model, vcov = vcovHC(model))
```

```
##                2.5 %   97.5 %
## (Intercept) 8.445986 8.451336
## log(carat)  1.672026 1.679607
```

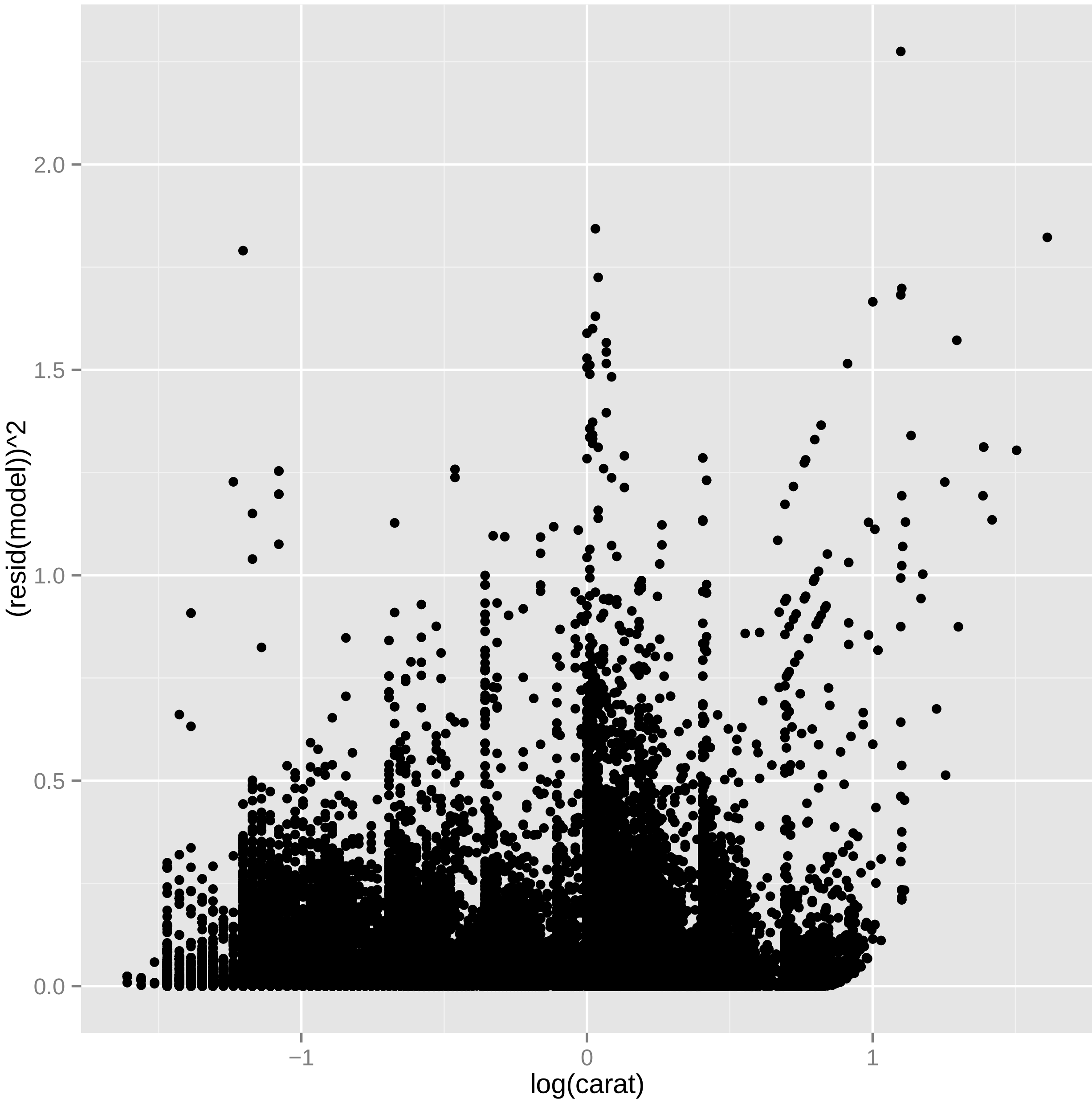
Относительно «руками» доверительный интервал считается следующим образом:

```
low <- as.vector(coef(model) - qnorm(0.975)*sqrt(diag(vcovHC(model))))
high <- as.vector(coef(model) + qnorm(0.975)*sqrt(diag(vcovHC(model))))
conf <- data.frame(low, high)
conf
```

```
##        low     high
## 1 8.445770 8.451552
## 2 1.671842 1.679791
```

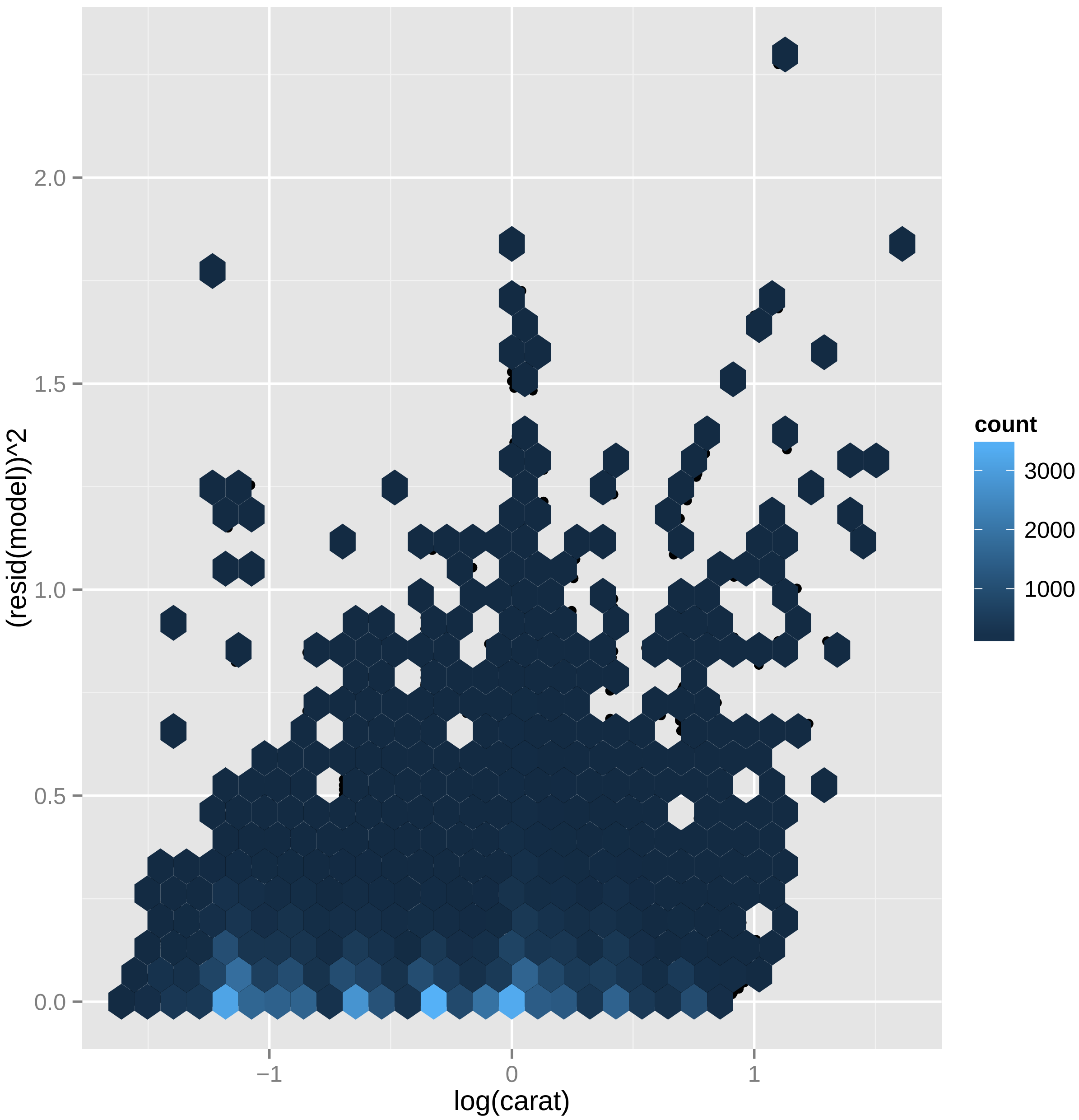
Попробуем задетектить гетероскедастичность с помощью графического анализа:

```
qplot(data = diamonds, log(carat), (resid(model))^2)
```



На таком графике ничего не видно. Попробуем увидеть черную кошку в черной комнате с помощью пакета *hexbin*:

```
qplot(data = diamonds, log(carat), (resid(model))^2) + geom_hex()
```



Видим, что с увеличением логарифма массы бриллианта растет дисперсия ошибки. Значит, есть гетероскедастичность.

Теперь попробуем выявлять и бороться с гетероскедастичностью теоретическими методами [Магнус, Пересецкий]. Итак, мы рассматриваем частный случай обобщенной регрессионной модели, а именно, модель с гетероскедастичностью. Это означает, что ковариационная матрица Ω — диагональная.

Стандартные ошибки в форме Уайта:

$$\mathbb{V}ar(\hat{\beta}|X) = (X'X)^{-1}X'\Omega X(X'X)^{-1}$$

Уайт показал, что

$$\widehat{\mathbb{V}ar}(\hat{\beta}|X) = (X'X)^{-1}X'\widehat{\Omega}X(X'X)^{-1}$$

является состоятельной оценкой ковариационной матрицы оценок коэффициентов регрессии.

Тесты на гетероскедастичность:

Во всех этих тестах проверяется основная гипотеза $H_0 : \sigma_1^2 = ... = \sigma_n^2$ против альтернативной гипотезы $H_a : H_0 \ is \ not \ true$.

Большинство тестов есть априорные структурные ограничения относительно характера гетероскедастичности (из каких-либо разумных соображений). Однако тест Уайта освобожден от этих ограничений:

1. **Тест Уайта (White)**

   Если в модели присутствуют гетероскедастичность, то очень часто это связано с тем, что дисперсии ошибок некоторым образом (возможно, сложным образом) зависят от регрессоров; гетероскедастичность должна как-то отражаться в остатках обычной регрессии исходной модели.

   Сначала к исходной модели применяется обычный МНК и находятся остатки регрессии e_t. Затем осуществляется регрессия квадратов этих остатков на все регрессоры X, их квадраты, а также попарные произведения и константу (если ее не было в составе исходных регрессоров). Тогда при верной гипотезе H_0 величина nR^2 асимптотически имеет распределение $\chi^2(N-1)$, где N — число регрессоров вспомагательной регрессии.

   Тест универсален, но в случае если H_0 отвергается, он не дает никакого указания на форму гетероскедастичности; единственный способ коррекции — использование стандартных ошибок в форме Уайта.

   [Демешев] Предпосылки:

   – H_0: $Var(\varepsilon_i) = \sigma^2$, нормальность не требуется;
   – $E(\varepsilon_i^4) = const$;
   – тест асимптотический;

   Процедура:

   – оценивается регрессия $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$;
   – строим регрессию $\hat{\varepsilon_i}^2$ на все регрессоры, их квадраты, их попарные произведения и константу (можно пойти и дальше, но разложение в Тейлора до членов, отвечающих за квадраты, уже дает хорошую точность);
   – асимптотически nR^2 имеет хи-квадрат распределение.

   Тест Уайта в R:

```
e2 <- (model$residuals)^2
logcarat <- log(diamonds$carat)
wteststat <- summary(lm(e2 ~ logcarat + (logcarat^2)))$r.squared *
  length(model$fitted.values)
p.value <- 1 - pchisq(wteststat, df = length(model$coefficients) - 1)
p.value

## [1] 0
```

2 Тест Голдфелда-Куандта (Goldfeld-Quandt)

Применяется, когда есть предположение о прямой зависимости дисперсии ошибки от величины некоторой независимой переменной. Кратко тест можно описать следующим образом:

- упорядочить данные по убыванию той независимой переменной, относительно которой есть подозрение на гетероскедастичность;
- исключить d средних в этом упорядочении наблюдений (d должно быть равно примерно четверти общего количества наблюдений);
- провести две независимые регрессии первых $n/2 - d/2$ наблюдений и последних $n/2 - d/2$ наблюдений и построить соответствующие остатки e_1 и e_2;
- составить статистику $F = e_1'e_1/e_2'e_2$. Если верна гипотеза H_0, то F имеет распределение Фишера с $(n/2 - d/2 - k, n/2 - d/2 - k)$ степенями свободы.

Большая величина этой статистики означает, что гипотезу H_0 следует отвергнуть. Формально тест работает и без исключения наблюдений, но, как показывает опыт, при этом его мощность уменьшается.

[Демешев] Предпосылки:

- нормальность остатков, $\varepsilon_i \sim N(0, \sigma_i^2)$;
- наблюдения упорядочены по возрастанию гетероскедастичности;
- тест точный (неасимптотический);

Процедура:

- упорядочить наблюдения в том порядке, в котором подозревается рост гетероскедастичности;
- выкинуть некий процент наблюдений по середине, чтобы подчеркнуть разницу в дисперсии между начальными и конечными наблюдениями;
- оценить исходную модель по наблюдениям из начала выборки и по наблюдениям конца выборки;
- получить, соответственно, RSS_1 и RSS_2;
- при верной H_0:
$$\frac{RSS_2}{RSS_1} \sim F_{r-k,r-k}$$
  где r — размер подвыборки в начале и в конце.

Тест Голдфельда-Квандта в R:

```
diamonds2 <- diamonds[order(log(diamonds$carat)), ]
# сменим порядок строк в табличке diamonds
model2 <- lm(log(price) ~ log(carat), data = diamonds2)
gqtest(model2, fraction = 0.2)
```

```
##
##  Goldfeld-Quandt test
##
## data:  model2
## GQ = 1.3634, df1 = 21574, df2 = 21574, p-value < 2.2e-16

# проведем GQ тест выкинув посередине 20% наблюдений
```

3 Тест Бреуша-Пагана (Breusch-Pagan)

[Магнус, Пересецкий] Тест применяется в тех случаях, когда априорно предполагается, что дисперсии σ_t^2 зависят от нескольких дополнительных переменных:

$$\sigma_t^2 = \gamma_0 + z_t'\gamma$$

В соответствии с этим тестом следует действовать так:

- провести обычную регрессию и получить вектор остатков e;
- построить оценку $\hat{\sigma}^2 = (1/n)\sum e_t^2$;
- провести регрессию $e_t^2/\hat{\sigma}^2 = \gamma_0 + z_t'\gamma + \nu_t$, где ν_t — белый шум, и найти для нее объясненную часть вариации ESS;
- построить статистику $ESS/2$. При верной гипотезе H_0: $ESS/2$ асимптотически имеет распределение $\chi^2(p)$, где p — длина вектора γ.

[Демешев] Предпосылки:

- нормальность остатков, $\varepsilon_i \sim N(0, \sigma_i^2)$;
- $\sigma_i^2 = h(\alpha_1 + \alpha_2 z_{i2} + \ldots + \alpha_p z_{ip})$;
- у функции $h(\cdot)$ существуют первая и вторая производные;
- тест асимптотический.

Суть теста: Используя метод максимального правдоподобия посчитаем LM-статистику. При верной H_0 она имеет хи-квадрат распределение с $p-1$ степенью свободы.

Оказывается, что LM-статистику можно получить с помощью вспомогательной регрессии. Авторская процедура:

Оценивается регрессия $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$.

Переходим к $g_i = n\hat{\varepsilon}_i/RSS$.

Строим регрессию g_i на $\alpha_1 + \alpha_2 z_{i2} + \ldots + \alpha_p z_{ip}$.

$LM = ESS/2$.

Современная модификация выглядит (неизвестный рецензент Коэнкера) так:

Оценивается регрессия $y_i = \beta_1 + \beta_2 x_i + \varepsilon_i$.

Оценивается регрессия $s_i^2 = \alpha_1 + \alpha_2 x_i + \varepsilon_i$.

При верной H_0 асимптотически:

$$nR^2 \sim \chi^2_{p-1}$$

где p — число оцениваемых коэффициентов во вспомогательной регрессии. По смыслу $(p-1)$ — это количество факторов, от которых потенциально может зависеть дисперсия $\mathbb{V}ar(\varepsilon_i)$.

Тест Бройша-Пагана в R:

```
bptest(model)


##
##  studentized Breusch-Pagan test
##
## data:  model
## BP = 411.6534, df = 1, p-value < 2.2e-16
```

Если хотя бы в одном из тестов H_0 отвергается, делаем вывод о гетероскедастичности.
Про прогнозирование:
КЛММР:

$$y = X\beta + \varepsilon;\ \varepsilon = 0,\ \mathbb{V}ar(\varepsilon) = \sigma^2 I$$

Предположим теперь, что есть еще один набор объясняющих переменных x_{n+1} и известно, что соответствующая зависимая переменная удовлетворяет модели:

$$y_{n+1} = x'_{n+1}\beta + \varepsilon_{n+1};\ \mathbb{E}\varepsilon_{n+1} = 0,\ \mathbb{V}ar(\varepsilon)_{n+1} = \sigma^2 I,\ \mathbb{C}ov(\varepsilon_{n+1}, \varepsilon) = 0$$

В качестве оценки y_{n+1} возьмем:

$$\hat{y}_{n+1} = x'_{n+1}\hat{\beta}$$

Стоит отметить, что где-то в этом временном промежутке необходимо было выполнить большое Домашнее задание № 1 про RLMS и гетероскедастичность.

Семинар. January, 19

$l(\theta)$ — логарифмическая функция максимального правдоподобия.

Рассмотрим три базовых теста, используемых для проверки ограничений на параметры статистических моделей. Все тесты асимптотические!

1 **Тест отношения правдоподобия (LR-тест)**:

При верной H_0:

$$LR = 2\left(l(\hat{\theta}_{UR}) - l(\hat{\theta}_R)\right) \sim \chi^2_q$$

Но на практике этот тест самый временезатратный, так как требует вычисления и $\hat{\theta}_{UR}$, и $\hat{\theta}_R$.

2 **Тест множителей Лагранжа (LM-тест)**:

Оценка информационной матрицы Фишера (матрица Гессе с минусом):

$$\widehat{I} = -\frac{\partial^2 l}{\partial\theta\partial\theta'}\left(\hat{\theta}\right)$$

При верной H_0:

$$LM = \left(\frac{\partial l}{\partial\theta}\left(\hat{\theta}_R\right)\right)' \cdot \widehat{I}^{-1}\left(\hat{\theta}_R\right) \cdot \left(\frac{\partial l}{\partial\theta}\left(\hat{\theta}_R\right)\right) \sim \chi^2_q$$

3 **Тест Вальда (Wald test)**:

$$H_0 : \begin{cases} g_1(\theta_1, ..., \theta_p) = 0 \\ ... \\ g_q(\theta_1, ..., \theta_p) = 0 \end{cases}$$

При верной H_0:

$$W = \left(g\left(\hat{\theta}_{UR}\right)\right)' \cdot \left(\frac{\partial g}{\partial\theta'}\left(\hat{\theta}_{UR}\right) \cdot \widehat{I}^{-1}\left(\hat{\theta}_{UR}\right) \cdot \frac{\partial g'}{\partial\theta}\left(\hat{\theta}_{UR}\right)\right)^{-1} \cdot g\left(\hat{\theta}_{UR}\right) \sim \chi^2_q$$

Доказано, что тест Вальда (W), тест отношения правдоподобия (LR) и тест множителей Лагранжа (LM) — асимптотически эквивалентные тесты ($LM = LR = W$). Тем не менее для конечных выборок значения статистик не совпадают. Для линейных ограничений вкупе с рядом предпосылок доказано неравенство $LM \leqslant LR \leqslant W$. Тем самым тест Вальда будет чаще других тестов отвергать нулевую гипотезу об ограничениях. В случае нелинейных ограничений первая часть неравенства выполняется, а вторая - вообще говоря, нет.

Задача 1. $X_1, ..., X_n$ — независимы, 100 наблюдений;

$$f(x) = \frac{\theta e^{-\theta^2/2x}}{\sqrt{2\pi x^3}},\ x > 0$$

$$\sum_{x=1}^{100} \frac{1}{x_i} = 12$$

а) $\hat{\theta}_{UR}$ — ?

б) $\widehat{I}\left(\hat{\theta}_{UR}\right)$ — ?

в) На уровне значимости $\alpha = 5\,\%$ с помощью базовых тестов проверить гипотезу $H_0 : \theta = 1$

Решение.

а)

$$l(\theta) = \log\left(f(x_1) \cdot ... \cdot f(x_n)\right) = \sum \log f(x_i) = \sum \left[\log\theta - \frac{\theta^2}{2x_i} - \frac{1}{2}\log(2\pi) - \frac{3}{2}\log(x_i)\right] =$$

$$= n\log(\theta) - \theta^2 \sum \frac{1}{2x_i} - \frac{n}{2}\log(2\pi) - \frac{3}{2}\sum \log(x_i)$$

$$\frac{\partial l(\theta)}{\partial\theta} = \frac{n}{\theta} - 2\theta \sum \frac{1}{2x_i} = \frac{100}{\theta} - 12\theta$$

$$\frac{100}{\hat{\theta}_{UR}} - 12 \cdot \hat{\theta}_{UR} = 0$$

$$\hat{\theta}_{UR} = \sqrt{100/12} \approx 2.887$$

Проделаем тоже самое в R с помощью пакета *maxLik*:

```
library(maxLik)

## Loading required package: miscTools
##
## Please cite the 'maxLik' package as:
## Henningsen, Arne and Toomet, Ott (2011). maxLik: A package for maximum likelihood
estimation in R. Computational Statistics 26(3), 443-458. DOI 10.1007/s00180-010-0217-1.
##
## If you have questions, suggestions, or comments regarding the 'maxLik' package,
please use a forum or 'tracker' at maxLik's R-Forge site:
## https://r-forge.r-project.org/projects/maxlik/
```

Создаем случайную выборку:

```
a <- runif(n = 100, 0, 1)
```

Для выполнения ограничения масштабируем ее:

```
b <- 12*a/sum(a)
sum(b)

## [1] 12

x <- 1/b
head(x)

## [1]   8.995076  13.194501  33.557125   7.560366  11.230284 191.314488

sum(1/x)

## [1] 12
```

Функция, возвращающая значение логарифмической функции максимального правдоподобия:

```
lik <- function(theta, data){
  n <- length(data)
  ans <- n * log(theta) - theta^2*sum(1/(2*data)) - n*log(2*pi)/2 -
    3*sum(log(data))/2
  return(ans)
}
lik(1, x)

## [1] -463.4543
```

Максимизация:

```
model <- maxLik(lik, start = 1, data = x)
summary(model)

## --------------------------------------------
## Maximum Likelihood estimation
## Newton-Raphson maximisation, 6 iterations
## Return code 1: gradient close to zero
## Log-Likelihood: -401.4411
## 1  free parameters
## Estimates:
##      Estimate Std. error t value Pr(> t)
## [1,]   2.8868     0.2042   14.14  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## --------------------------------------------
```

$\hat{\theta}_{UR}$:

```
model$estimate

## [1] 2.886751
```

б)
Гессиан:

$$\frac{\partial^2 l}{\partial \theta^2}\left(\hat{\theta}_{UR}\right) = -\frac{n}{\hat{\theta}_{UR}^2} - \sum \frac{1}{x_i} = \frac{100}{2.887^2} - 12 \approx -24$$

Или так:

```
model$hessian

##             [,1]
## [1,] -23.98792
```

Тогда:

$$\widehat{I}\left(\hat{\theta}_{UR}\right) = 24$$

в)
$\hat{\theta}_{UR}$:

```
theta_ur <- model$estimate
```

$\hat{\theta}_R$ берем из гипотезы:

```
theta_r <- 1
```

LR-тест:

```
LR <- -2*(lik(theta_r, data = x) - lik(theta_ur, data = x))
LR

## [1] 124.0264

qcrit <- qchisq(0.95, df = 1); qcrit

## [1] 3.841459

LR < qcrit

## [1] FALSE
```

Гипотеза отвергается.
Wald test:

```
I_ur <- -model$hessian
g_ur <- theta_ur - 1
dg_dtheta_ur <- 1
W <- t(g_ur) %*% solve(dg_dtheta_ur %*% solve(I_ur) %*%
                         t(dg_dtheta_ur)) %*% g_ur
W <- as.integer(W); W

## [1] 85

W < qcrit

## [1] FALSE
```

Гипотеза отвергается.

LM-тест:

```
dl_dtheta_r <- 100/theta_r - 12*theta_r
LM <- t(dl_dtheta_r) %*% solve(I_ur) %*% dl_dtheta_r
LM <- as.integer(LM); LM

## [1] 322

LM < qcrit

## [1] FALSE
```

Гипотеза отвергается.

Как видим, здесь не выполняется соотношение между LR, LM и W, о котором шла речь выше. Проблема в том, что здесь $|W/n| < 1$ (подробнее в Ullah A., Zinde-Walsh V. On the robustness of LM, LR, and W tests in regression models //Econometrica: Journal of the Econometric Society. – 1984.)

Задача 2. $x_1, ..., x_{100}$ — i. i. d. $\mathcal{N}(\mu, \sigma^2)$

$$\sum x_i = 10$$

$$\frac{\sum(x_i - \bar{x})^2}{n-1} = 5 = s_x^2$$

а) $H_0 : \mu = \sigma^2$

б) $H_0 : \begin{cases} \mu = 4 \\ \sigma^2 = 7 \end{cases}$

Решение.

а)

Функция плотности нормального распределения:

$$f(x_i) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i-\mu)^2}{2\sigma^2}}$$

Тогда логарифмическая функция максимального правдоподобия здесь:

$$l(\mu, \sigma) = \log\left(f(x_1) \cdot ... \cdot f(x_{100})\right) = \sum \log f(x_i) =$$

$$= \sum \left[-\log\left(\sqrt{2\pi}\right) - \log\sigma - \frac{(x_i - \mu)^2}{2\sigma^2}\right]$$

Берем первые производные:

$$\frac{\partial l(\cdot)}{\partial \mu} = -\frac{2\sum(x_i - \mu)}{2\sigma^2} = -\frac{\sum x_i - n\mu}{\sigma^2}$$

$$\frac{\partial l(\cdot)}{\partial \sigma} = -\frac{n}{\sigma} + \frac{\sum(x_i - \mu)^2}{\sigma^3} = -\frac{n}{\sigma} + \frac{\sum(x_i - \bar{x})^2 + 2\sum(x_i - \bar{x})(\bar{x} - \mu) + \sum(\bar{x} - \mu)^2}{\sigma^3}$$

Из F.O.C.:

$$\widehat{\mu} = \frac{\sum x_i}{n} = \bar{x} = 0.1$$

$$\widehat{\sigma}^2 = \frac{\sum (x_i - \bar{x})^2}{n} = \frac{n-1}{n} s_x^2 = 4.95$$

Теперь найдем оценки максимального правдоподобия при ограничении из гипотезы $\mu = \sigma^2$:

$$l(\mu, \sigma) = \sum \left[-\log\left(\sqrt{2\pi}\right) - \frac{1}{2} \log \mu - \frac{(x_i - \mu)^2}{2\mu} \right]$$

Первая производная:

$$\frac{\partial l(\cdot)}{\partial \mu} = -\frac{n}{2\mu} - \frac{1}{2} \sum \frac{-2(x_i - \mu)\mu - (x_i - \mu)^2}{\mu^2} =$$

$$= -\frac{n}{2\mu} - \frac{1}{2} \frac{-2\mu \sum x_i + 2n\mu^2 - \sum \left[(x_i - 0.1)^2 + 2(x_i - 0.1)(0.1 - \mu) + (0.1 - \mu)^2 \right]}{\mu^2} =$$

$$= -\frac{100}{2\mu} - \frac{1}{2} \frac{-20\mu + 200\mu^2 - 495 - 100(0.1 - \mu)^2}{\mu^2}$$

F.O.C.:

$$-100\widehat{\mu} + 20\widehat{\mu} - 200\widehat{\mu}^2 + 495 + 1 - 20\widehat{\mu} + 100\widehat{\mu}^2 = 0$$

$$\widehat{\mu}^2 + \widehat{\mu} - 4.96 = 0$$

Учитывая, что $\mu \geqslant 0$:

$$\widehat{\mu} = \frac{-1 + \sqrt{1 + 4 \cdot 4.96}}{2} \approx 1.7825$$

Итак:

$$\widehat{\mu}_{UR} = 0.1,\ \widehat{\sigma}^2_{UR} = 4.95,\ \widehat{\mu}_R = \widehat{\sigma}^2_R = 1.7825$$

$$l\left(\widehat{\mu}_{UR}, \widehat{\sigma}_{UR}\right) = -100 \log(\sqrt{2\pi}) - 100 \log \sqrt{4.95} - \frac{\sum (x_i - 0.1)^2}{2 \cdot 4.95} \approx -221.86$$

$$l\left(\widehat{\mu}_R, \widehat{\sigma}_R\right) = -100 \log(\sqrt{2\pi}) - 100 \log \sqrt{1.7825} - \frac{\sum (x_i - 1.7825)^2}{2 \cdot 1.7825} \approx$$

$$\approx -120.79 - \frac{\sum \left[(x_i - 0.1)^2 - 2 \cdot 1.6825 \cdot (x_i - 0.1) + 1.6825^2 \right]}{3.565} =$$

$$\approx -120.79 - \frac{497.83}{3.565} \approx -260.43$$

Применив тест отношения правдоподобия, получим:

$$LR = 2\left(l\left(\widehat{\mu}_{UR}, \widehat{\sigma}_{UR}\right) - l\left(\widehat{\mu}_R, \widehat{\sigma}_R\right)\right) = 77.14$$

P-value:

```
1 - pchisq(77.14, df = 2)

## [1] 0
```

б)

Здесь:

$$\widehat{\mu}_{UR} = 0.1,\ \widehat{\sigma}^2_{UR} = 4.95,\ \widehat{\mu}_R = 4,\ \widehat{\sigma}^2_R = 7$$

$$l\left(\widehat{\mu}_{UR}, \widehat{\sigma}_{UR}\right) = -100\log(\sqrt{2\pi}) - 100\log\sqrt{4.95} - \frac{\sum\left(x_i - 0.1\right)^2}{2\cdot 4.95} \approx -221.86$$

$$l\left(\widehat{\mu}_{R}, \widehat{\sigma}_{R}\right) = -100\log(\sqrt{2\pi}) - 100\log\sqrt{7} - \frac{\sum\left(x_i - 4\right)^2}{2\cdot 7} \approx$$

$$\approx -189.19 - \frac{\sum\left[\left(x_i - 0.1\right)^2 - 2\cdot 3.9\cdot\left(x_i - 0.1\right) + 3.9^2\right]}{14} =$$

$$= -189.19 - \frac{5\cdot 99 + 0 + 100\cdot 3.9^2}{14} = -333.19$$

Применив тест отношения правдоподобия, получим:

$$LR = 2\left(l\left(\widehat{\mu}_{UR}, \widehat{\sigma}_{UR}\right) - l\left(\widehat{\mu}_{R}, \widehat{\sigma}_{R}\right)\right) = 222.66$$

P-value:

```
1 - pchisq(222.66, df = 2)

## [1] 0
```

Гипотеза отвергается.

Семинар. January, 26

Метод максимального правдоподобия

Оценки максимального правдоподобия, вообще говоря, могут быть смещёнными, но являются состоятельными, асимптотически эффективными и асимптотически нормальными оценками. Асимптотическая нормальность означает, что:

$$\sqrt{n}(\hat{\theta}-\theta) \xrightarrow{d} N(0, I_{\infty}^{-1})$$

где $I_{\infty} = -\lim_{n\to\infty}(\mathbb{E}(H)/n)$ — асимптотическая информационная матрица.

Асимптотическая эффективность означает, что асимптотическая ковариационная матрица I_{∞}^{-1} является нижней границей для всех состоятельных асимптотически нормальных оценок.

Если $\hat{\theta}$ — оценка метода максимального правдоподобия параметров θ, то $g(\hat{\theta})$ является оценкой максимального правдоподобия для $g(\theta)$, где $g(\cdot)$ — непрерывная функция (функциональная инвариантность). Таким образом, законы распределения данных можно параметризовать различным образом.

Задачу ограниченной оптимизации лучше приводить к задаче неограниченной оптимизации.

Например:

Ограничение $p \in [0, 1]$ с помощью сигмоиды приводим к $p = 1/\left(1 + e^{-t}\right)$, где $t \in (-\infty, +\infty)$.

Ограничение $\sigma \in (0, +\infty)$ заменяем на $\sigma = e^t$, где $t \in (-\infty, +\infty)$.

Задача 1.

Если $X \sim Pois(\lambda)$, то $\mathbb{E}(X) = \lambda$, $\mathbb{V}ar(X) = \lambda$.

Но на реальных данных $\mathbb{E}(X) \neq \mathbb{V}ar(X)$ из-за того, что нули попадаются слишком часто.

Пусть X_i — число кусков мяса в плове для i-го посетителя столовой. Например, Аня может быть любителем плова с вероятностью p, в этом случае $X_i \sim Pois(\lambda)$; а может быть принципиальной его ненавистницей с вероятностью $1-p$, тогда $X_i = 0$.

а) Найти $\widehat{p}$, $\widehat{\lambda}$ с помощью метода максимального правдоподобия

б) Построить доверительные интервалы для полученных оценок

в) С помощью базовых тестов проверить гипотезу H_0: $\lambda = 1$

г) С помощью базовых тестов проверить гипотезу H_0: $p = \lambda$

Решение.

а) Для начала нужно понять, что мы будем оптимизировать:

$$\mathbb{P}(X_i = x) = \begin{cases} (1-p) + pe^{-\lambda}, & x = 0 \\ pe^{-\lambda}\lambda^x/x!, & x > 0 \end{cases}$$

Taking logs:

$$\log\left(\mathbb{P}(X_i = x)\right) = \begin{cases} \log\left(1 - p + pe^{-\lambda}\right), & x = 0 \\ \log p - \lambda + x\log\lambda - \log x!, & x > 0 \end{cases}$$

Например, если в среднем в порцию плова кидают два куска мяса:

```r
library(maxLik)

set.seed(4) # for reproducible results
# more on
# http://stackoverflow.com/questions/13605271/reasons-for-using-the-set-seed-function

X <- rpois(n = 100, lambda = 2)
head(X)

## [1] 2 0 1 1 3 1
```

Функция, которую будем оптимизировать:

```r
loglik <- function(par, data) {
  p <- par[1]
  lambda <- par[2]
  n <- length(data)
  n_zero <- sum(data == 0)
  ans <- n_zero*log(1 - p + p*exp(-lambda))
  nonzero_data <- data[data > 0]
  ans <- ans + (n - n_zero)*(log(p) - lambda) +
    log(lambda)*sum(nonzero_data)
  return(ans)
}
```

Например, для $p = 0.5$, $\lambda = 1$ значение логарифмической функции правдоподобия для данных X будет равно:

```r
loglik(c(0.5, 1), X)

## [1] -152.2423
```

Но мы знаем, что есть ограничения $p \in [0, 1]$ и $\lambda \in (0, +\infty)$. То есть фактически мы имеем дело с условной оптимизацией. Что ж, перейдем к безусловной оптимизации с помощью функции-трансформера:

```r
transformer <- function(par, data) {
        p <- 1/(1 + (exp(par[1]))^(-1)) # сигмоида
        lambda <- exp(par[2]) # экспонента
        return(c(p, lambda))
}

loglik2 <- function(par, data) {
        return(loglik(transformer(par), data))
}

result <- maxLik(logLik = loglik2, start = c(0,0), data = X)
report <- summary(result)
report
```

```
## --------------------------------------------
## Maximum Likelihood estimation
## Newton-Raphson maximisation, 9 iterations
## Return code 1: gradient close to zero
## Log-Likelihood: -41.27094
## 2  free parameters
## Estimates:
##      Estimate Std. error t value Pr(> t)
## [1,]  3.22389    1.15525   2.791 0.00526 **
## [2,]  0.85439    0.07701  11.094 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
## --------------------------------------------
```

```
result$estimate

## [1] 3.223893 0.854393

transformer(result$estimate)

## [1] 0.9617236 2.3499476
```

Итак: $\widehat{p} = 0.9617236$, $\widehat{\lambda} = 2.3499476$

б) Строим $95-\%$ доверительные интервалы для полученных оценок:

```
high <- transformer(result$estimate + qnorm(0.975)*report$estimate[,2])
low <- transformer(result$estimate - qnorm(0.975)*report$estimate[,2])
conf <- data.frame(low, high, row.names = c('p', 'lambda'))
conf

##              low      high
## p      0.7230517 0.9958815
## lambda 2.0207196 2.7328154
```

в) Мы уже знаем, что:

$$\widehat{p}_{UR} = 0.9617236, \ \widehat{\lambda}_{UR} = 2.3499476$$

При этом:

```
loglik(c(transformer(result$estimate)[1], transformer(result$estimate)[2]), X)

## [1] -41.27094
```

По условию $\widehat{\lambda}_R = 1$. Нам нужно найти $\widehat{p}_R$ посредством максимизации логарифмической функции правдоподобия с подставлением в нее $\widehat{\lambda}_R = 1$.

Нужно будет немного переписать функции:

```
loglik_restr <- function(p, data) {
  n <- length(data)
  n_zero <- sum(data == 0)
  ans <- n_zero*log(1 - p + p*exp(-1))
  nonzero_data <- data[data > 0]
  ans <- ans + (n - n_zero)*(log(p) - 1)
  return(ans)
}
```

```
transformer_restr <- function(p, data) {
        p_log <- 1/(1 + exp(p)^(-1)) # сигмоида
        return(p_log)
}

loglik2_restr <- function(p, data) {
        return(loglik_restr(transformer_restr(p), data))
}

result_restr <- maxLik(logLik = loglik2_restr, start = 0, data = X)
report_restr <- summary(result_restr)
report_restr

## --------------------------------------------
## Maximum Likelihood estimation
## Newton-Raphson maximisation, 10 iterations
## Return code 1: gradient close to zero
## Log-Likelihood: -100
## 1  free parameters
## Estimates:
##      Estimate Std. error t value Pr(> t)
## [1,]     8962        Inf       0       1
## --------------------------------------------
```

```
transformer_restr(result_restr$estimate)

## [1] 1
```

Значение логарифмической функции распределения в ограниченной модели:

```
loglik_restr(transformer_restr(result_restr$estimate), X)

## [1] -100
```

Для проверки гипотезы H_0: $\lambda = 1$ ограничимся тестом отношения правдоподобия:

```
LR <- 2*( loglik(c(transformer(result$estimate)[1],
                  transformer(result$estimate)[2]), X) -
           loglik_restr(transformer_restr(result_restr$estimate), X) )
LR
```

```
## [1] 117.4581

p.value <- 1 - pchisq(LR, df = 1)
p.value

## [1] 0
```

Гипотеза отвергается.

г)

Для проверки данной гипотезы снова придется переписать функцию, которую мы будем оптимизировать:

```
loglik_restr2 <- function(p, data) {
  n <- length(data)
  n_zero <- sum(data == 0)
  ans <- n_zero*log(1 - p + p*exp(-p))
  nonzero_data <- data[data > 0]
  ans <- ans + (n - n_zero)*(log(p) - p) +
    log(p)*sum(nonzero_data)
  return(ans)
}

loglik2_restr2 <- function(p, data) {
        return(loglik_restr2(transformer_restr(p), data))
}
result_restr2 <- maxLik(logLik = loglik2_restr2, start = 0, data = X)
report_restr2 <- summary(result_restr2)
report_restr2

## --------------------------------------------
## Maximum Likelihood estimation
## Newton-Raphson maximisation, 11 iterations
## Return code 1: gradient close to zero
## Log-Likelihood: -100
## 1  free parameters
## Estimates:
##      Estimate Std. error t value Pr(> t)
## [1,]     3477        Inf       0       1
## --------------------------------------------
```

```
transformer_restr(result_restr2$estimate)

## [1] 1
```

Снова применим LR-тест, и снова гипотеза будет отвергнута:

```
LR <- 2*( loglik(c(transformer(result$estimate)[1],
                  transformer(result$estimate)[2]), X) -
          loglik_restr2(transformer_restr(result_restr2$estimate), X) )
LR
```

```
## [1] 117.4581
```

```r
p.value <- 1 - pchisq(LR, df = 2)
p.value
```

```
## [1] 0
```

Семинар. February, 2

Метод максимального правдоподобия применительно к logit и probit моделям

Для начала нужно познакомиться с логистическим распределением:

$X \sim L(0,1)$, если

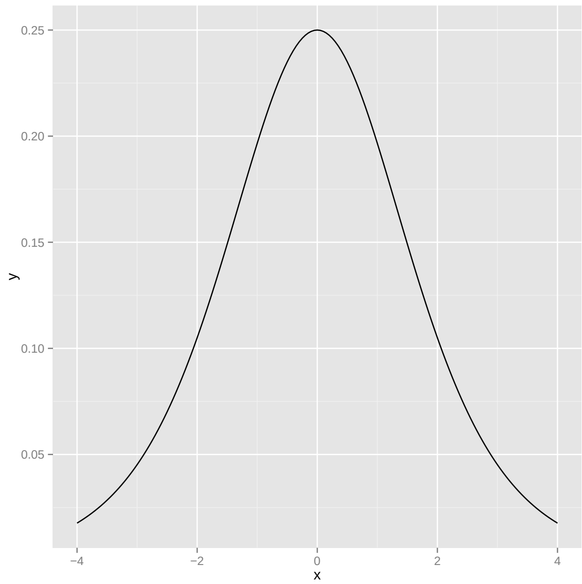
$$f_X(x) = \frac{e^{-x}}{(1+e^{-x})^2}$$

В общем виде: $X \sim L(\mu, s)$, если:

$$f_X(x) = \frac{e^{-(x-\mu)/s}}{s\left(1+e^{-(x-\mu)/s}\right)^2} = \frac{1}{4s} \cdot \operatorname{sech}^2\left(\frac{x-\mu}{2s}\right)$$

где μ — параметр сдвига, s — параметр масштаба. Построим функцию плотности логистического распределения:

```
library(ggplot2)
x <- seq(-4, 4, by = 0.05)
y <- dlogis(x, location = 0, scale = 1)
qplot(x, y, geom = 'line')
```



Легко показать, что $f_X(x)$ является четной:

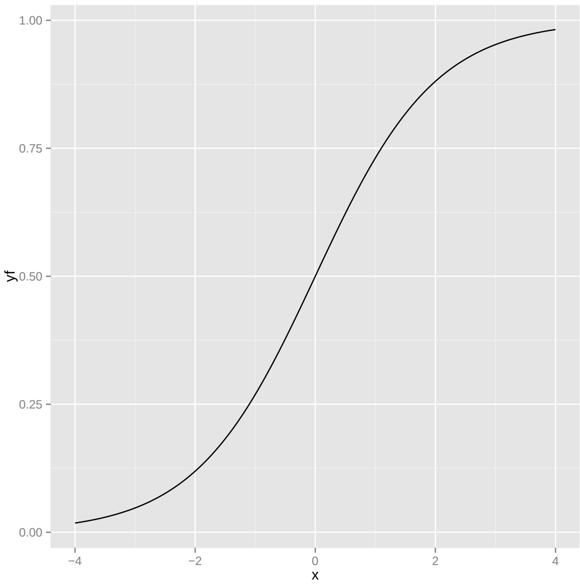
$$f(-x) = \frac{e^{x}}{(1+e^{x})^2} = \frac{e^{-x}}{(1+e^{-x})^2} = f(x)$$

Выведем функцию распределения логистического распределения:

$$F_X(x) = \int\limits_{-\infty}^{x} \frac{e^{-x}}{(1+e^{-x})^2} dx = \frac{1}{1+e^{-x}}\Bigg|_{-\infty}^{x} = \frac{1}{1+e^{-x}}$$

Достоинством логистического распределения является интегрируемость функции плотности. Заметим, что функция распределения логистического распределения — стандартная сигмоида:

```
yf <- plogis(x)
qplot(x, yf, geom = 'line')
```



Квантильная функция:

$$F^{-1}(p) = \log\left(\frac{p}{1-p}\right)$$

Так как $f(x)$ является четной, то функция $xf(x)$ является нечетной, что значит, что $\mathbb{E}X = 0$.

Если $Z = \sqrt{3}/\pi \cdot X$, то есть параметр масштаба $s = \sqrt{3}/\pi$, то:

$$f_Z(z) = \frac{\pi}{\sqrt{3}} \cdot \frac{e^{-\pi z/\sqrt{3}}}{\left(1+e^{-\pi z/\sqrt{3}}\right)^2}$$

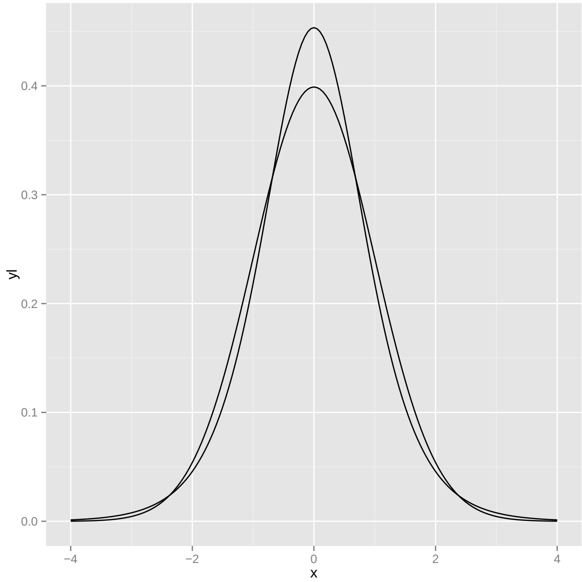
Можно показать, что $\mathbb{V}arX = \pi^2/3$. Например, так:

```
r <- rlogis(10^6)
var(r); pi^2/3

## [1] 3.286736
## [1] 3.289868
```

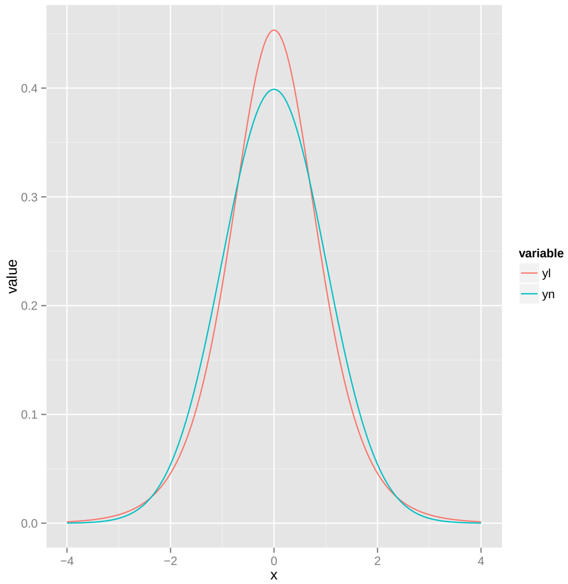
Значит, $\mathbb{V}arX = 1$. Функция плотности случайной величины Z является хорошей аппроксимацией функции плотности стандартного нормального распределения.

Покажем это, двумя способами организовав данные:

```
yl <- dlogis(x, location = 0, scale = sqrt(3)/pi)
yn <- dnorm(x)
data <- data.frame(x = x, yl = yl, yn = yn)
qplot(data = data, x, yl, geom = 'line') + geom_line(aes(y = yn))
```




```
library(reshape2) # перевод данных из широкого формата в длинный

data_long <- melt(data = data, id.vars = 'x')
head(data_long)

##       x variable       value
## 1 -4.00       yl 0.001279628
## 2 -3.95       yl 0.001400914
## 3 -3.90       yl 0.001533677
## 4 -3.85       yl 0.001678999
## 5 -3.80       yl 0.001838062
## 6 -3.75       yl 0.002012160

qplot(data = data_long, x = x, y = value, col = variable, geom = 'line')
```



Длинный формат очень удобен для автоматической обработки данных. Из длинного формата в широкий переводит команда *cast*.

Логит и пробит модели

$y_i \in \{0,1\}$ — прогулял лекцию или нет (наблюдаема);

y^* — склонность прогулять лекцию (ненаблюдаема);

Построим следующую модель:

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

Отличие логит модели от пробит модели:

- Логит: $\varepsilon_i \sim L(0,1)$
- Пробит: $\varepsilon_i \sim N(0,1)$

Для логистической модели:

$$\mathbb{P}(y_i = 1) = \mathbb{P}(\beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i \geqslant 0) = \mathbb{P}\left(\varepsilon_i \geqslant -\left(\beta_1 + \beta_2 x_i + \beta_3 z_i\right)\right) = \frac{e^{\beta_1+\beta_2 x_i+\beta_3 z_i}}{1 + e^{\beta_1+\beta_2 x_i+\beta_3 z_i}}$$

$$\mathbb{P}(y_i = 0) = 1 - \mathbb{P}(y_i = 1) = \frac{1}{1 + e^{\beta_1+\beta_2 x_i+\beta_3 z_i}}$$

Когда имеешь дело с логистической моделью, необходимо мыслить в отношениях шансов:

$$\log\left(\frac{\mathbb{P}(y_i = 1)}{\mathbb{P}(y_i = 0)}\right) = \beta_1 + \beta_2 x_i + \beta_3 z_i$$

Задача 6.6.

Винни-Пух знает, что мёд бывает правильный, $honey_i = 1$, и неправильный, $honey_i = 0$. Пчёлы также бывают правильные, $bee_i = 1$, и неправильные, $bee_i = 0$. По 100 своим попыткам добыть мёд Винни-Пух составил таблицу сопряженности:

	$honey_i = 1$	$honey_i = 0$
$bee_i = 1$	12	36
$bee_i = 0$	32	20

Используя метод максимального правдоподобия, Винни-Пух хочет оценить логит модель для прогнозирования правильности мёда с помощью правильности пчёл:

$$\log\left(\frac{\mathbb{P}(honey_i = 1)}{\mathbb{P}(honey_i = 0)}\right) = \beta_1 + \beta_2 bee_i$$

а) Выпишите функцию правдоподобия для оценки параметров β_1 и β_2

б) Оцените неизвестные параметры

в) С помощью теста отношения правдоподобия проверьте гипотезу о том, правильность пчёл не связана с правильностью мёда на уровне значимости 5 %

г) Держась в небе за воздушный шарик, Винни-Пух неожиданно понял, что перед ним неправильные пчёлы. Помогите ему оценить вероятность того, что они делают неправильный мёд

Решение.

а) Будем оценивать склонность меда к правильности от правильности пчел:

$$honey_i^* = \beta_1 + \beta_2 bee_i + \varepsilon_i$$

Логарифмическая функция правдоподобия:

$$l(\beta_1, \beta_2, data) = \sum \log \mathbb{P}\left(honey_i(\omega) = honey_i \,|\, bee_i\right)$$

То есть мы ввели случайную величину — условную правильность меда.
Найдем все необходимые условные вероятности:

$$\mathbb{P}\left(honey_i = 1 \,|\, bee_i = 1\right) = \frac{e^{\beta_1+\beta_2}}{1+e^{\beta_1+\beta_2}}$$

$$\mathbb{P}\left(honey_i = 1 \,|\, bee_i = 0\right) = \frac{e^{\beta_1}}{1+e^{\beta_1}}$$

$$\mathbb{P}\left(honey_i = 0 \,|\, bee_i = 1\right) = \frac{1}{1+e^{\beta_1+\beta_2}}$$

$$\mathbb{P}\left(honey_i = 0 \,|\, bee_i = 0\right) = \frac{1}{1+e^{\beta_1}}$$

Учитывая данные из таблички выше:

$$l(\beta_1, \beta_2, data) = 12\log\left(\frac{e^{\beta_1+\beta_2}}{1+e^{\beta_1+\beta_2}}\right) + 32\log\left(\frac{e^{\beta_1}}{1+e^{\beta_1}}\right) +$$

$$+36\log\left(\frac{1}{1+e^{\beta_1+\beta_2}}\right) + 20\log\left(\frac{1}{1+e^{\beta_1}}\right) =$$

$$= 44\beta_1 + 12\beta_2 - 48\log\left(1+e^{\beta_1+\beta_2}\right) - 52\log\left(1+e^{\beta_1}\right)$$

б) Найдем сначала первые производные функции $l(\cdot)$:

$$\frac{\partial l(\cdot)}{\partial \beta_1} = 44 - \frac{48e^{\beta_1+\beta_2}}{1+e^{\beta_1+\beta_2}} - \frac{52e^{\beta_1}}{1+e^{\beta_1}}$$

$$\frac{\partial l(\cdot)}{\partial \beta_2} = 12 - \frac{48e^{\beta_1+\beta_2}}{1+e^{\beta_1+\beta_2}}$$

Из условий первого порядка задачи максимизации:

$$42 - \frac{52e^{\widehat{\beta}_1}}{1+e^{\widehat{\beta}_1}} = 12$$

Отсюда:

$$\widehat{\beta}_1 = \log\left(\frac{8}{5}\right) \approx 0.47$$

$$\widehat{\beta}_2 = \log\left(\frac{5}{24}\right) \approx -1.57$$

в) Можем сформулировать нулевую гипотезу: $H_0 : \beta_2 = 0$

Нужно максимизировать логарифмическую функцию правдоподобия по β_1 при ограничении $\beta_2 = 0$:

$$l^R(\beta_1, data) = 44\beta_1 - 100\log\left(1+e^{\beta_1}\right)$$

$$\frac{\partial l(\cdot)}{\partial\beta_1} = 44 - \frac{100e^{\beta_1}}{1+e^{\beta_1}}$$

$$44 - \frac{100e^{\widehat{\beta}_1}}{1+e^{\widehat{\beta}_1}} = 0$$

$$\widehat{\beta}_1 = \log\left(\frac{11}{14}\right) \approx -0.24$$

Здесь:

$$\widehat{\beta}_{1,UR} = 0.47,\ \widehat{\beta}_{2,UR} = -1.57,\ \widehat{\beta}_{1,R} = -0.24,\ \widehat{\beta}_{2,R} = 0$$

Применив тест отношения правдоподобия, получим:

$$LR = 2\left(l\left(\widehat{\beta}_{1,UR}, \widehat{\beta}_{2,UR}\right) - l\left(\widehat{\beta}_{1,R}, \widehat{\beta}_{2,R}\right)\right) = 2\left(-61.63857 + 68.59298\right) \approx 13.91$$

```
p.value <- 1 - pchisq(13.91, df = 2)
p.value

## [1] 0.0009538539

p.value < 0.05

## [1] TRUE
```

Гипотеза отвергается. То есть на самом деле правильность меда связана с правильностью пчел.

г) Используем оценки коэффициентов для оценки необходимой вероятности:

$$\widehat{\mathbb{P}}\left(honey_i = 0 \,|\, bee_i = 0\right) = \frac{1}{1+e^{\widehat{\beta}_{1,UR}}} = \frac{1}{1+e^{0.47}} \approx 0.38$$

Задача 6.5.

При оценке логит модели

$$\mathbb{P}(y_i = 1) = \Lambda\left(\beta_1 + \beta_2 x_i\right)$$

оказалось, что $\widehat{\beta}_1 = 0.7$ и $\widehat{\beta}_2 = 3$. Найдите максимальный предельный эффект роста x_i на вероятность $\mathbb{P}(y_i = 1)$.

Решение.

$$\mathbb{P}\widehat{(y_i = 1)} = \Lambda\left(\widehat{\beta}_1 + \widehat{\beta}_2 x_i\right) = \frac{1}{\left(1+e^{-\left(\widehat{\beta}_1+\widehat{\beta}_2 x_i\right)}\right)} = \frac{1}{1+e^{-\left(\widehat{\beta}_1+\widehat{\beta}_2 x_i\right)}}$$

Предельный эффект роста x_i на вероятность $\mathbb{P}(y_i = 1)$:

$$\frac{\partial\mathbb{P}\widehat{(y_i = 1)}}{\partial x_i} = \frac{\widehat{\beta}_2 e^{-\left(\widehat{\beta}_1+\widehat{\beta}_2 x_i\right)}}{\left(1+e^{-\left(\widehat{\beta}_1+\widehat{\beta}_2 x_i\right)}\right)^2}$$

Максимизируем:

$$\frac{\partial^2\mathbb{P}(y_i = 1)}{\partial x_i^2} = \widehat{\beta}_2\frac{-\widehat{\beta}_2 e^{-\left(\widehat{\beta}_1+\widehat{\beta}_2 x_i\right)}\left(1+e^{-\left(\widehat{\beta}_1+\widehat{\beta}_2 x_i\right)}\right)^2 + 2\left(1+e^{-\left(\widehat{\beta}_1+\widehat{\beta}_2 x_i\right)}\right)e^{-\left(\widehat{\beta}_1+\widehat{\beta}_2 x_i\right)}\widehat{\beta}_2}{\left(1+e^{-\left(\widehat{\beta}_1+\widehat{\beta}_2 x_i\right)}\right)^4} = 0$$

$$1+e^{-\left(\widehat{\beta_1}+\widehat{\beta_2}x_i\right)}=2$$
$$\widehat{\beta}_1+\widehat{\beta}_2x_i=0$$
$$x_i=-\frac{\widehat{\beta}_1}{\widehat{\beta}_2}$$

Итак:

$$\hat{x}_i=-\frac{\widehat{\beta}_1}{\widehat{\beta}_2}\approx -0.23$$

Максимальный предельный эффект роста x_i на вероятность:

$$\frac{\partial\widehat{\mathbb{P}(y_i=1)}}{\partial x_i}=\frac{3e^{-(0.7+3\cdot(-0.23))}}{\left(1+e^{-(0.7+3\cdot(-0.23))}\right)^2}\approx 0.75$$

Задача 6.4.

Как известно, Фрекен Бок любит пить коньяк по утрам. За прошедшие 4 дня она записала, сколько рюмочек коньяка выпила утром, x_i, и видела ли она в этот день приведение, y_i:

y_i	1	0	1	0
x_i	2	1	3	0

Зависимость между y_i и x_i описывается логит моделью:

$$\log\left(\frac{\mathbb{P}(y_i=1)}{\mathbb{P}(y_i=0)}\right)=\beta_1+\beta_2x_i$$

а) Выпишите в явном виде логарифмическую функцию максимального правдоподобия

б) Найдите оценки параметров β_1 и β_2

Решение.

а) Логарифмическая функция максимального правдоподобия выглядит следующим образом:

$$l(\beta_1,\beta_2,data)=\sum\log\mathbb{P}\left(y_i\,|\,x_i\right)$$

Выпишем все вероятности, соответствующие данным из таблицы:

$$\mathbb{P}\left(y_i=1\,|\,x_i=2\right)=\frac{e^{\beta_1+2\beta_2}}{1+e^{\beta_1+2\beta_2}}$$
$$\mathbb{P}\left(y_i=0\,|\,x_i=1\right)=\frac{1}{1+e^{\beta_1+\beta_2}}$$
$$\mathbb{P}\left(y_i=1\,|\,x_i=3\right)=\frac{e^{\beta_1+3\beta_2}}{1+e^{\beta_1+3\beta_2}}$$
$$\mathbb{P}\left(y_i=0\,|\,x_i=0\right)=\frac{1}{1+e^{\beta_1}}$$

$$l(\beta_1,\beta_2,data)=2\beta_1+5\beta_2-\log\left(1+e^{\beta_1+2\beta_2}\right)-\log\left(1+e^{\beta_1+\beta_2}\right)-$$
$$-\log\left(1+e^{\beta_1+3\beta_2}\right)-\log\left(1+e^{\beta_1}\right)$$

б)

```r
loglik <- function(par) {
  b1 <- par[1]
  b2 <- par[2]
  ans <- 2*b1 + 5*b2 - log(1 + exp(b1 + 2*b2)) -
    log(1 + exp(b1 + b2)) -
    log(1 + exp(b1 + 3*b2)) -
    log(1 + exp(b1))
  return(ans)
}
```

```r
result <- maxLik(logLik = loglik, start = c(0,0))
report <- summary(result)
report

## --------------------------------------------
## Maximum Likelihood estimation
## Newton-Raphson maximisation, 6 iterations
## Return code 1: gradient close to zero
## Log-Likelihood: -1.136868e-13
## 2  free parameters
## Estimates:
##      Estimate Std. error t value Pr(> t)
## [1,]   -369.7        Inf       0       1
## [2,]    267.0        Inf       0       1
## --------------------------------------------

result$estimate

## [1] -369.7451  267.0360
```

Наблюдается ситуация perfect prediction. Простое правило — если $x_i > 1.5$, то $y_i = 1$. Ситуация иногда наблюдается в логит моделях, когда много объясняющих переменных.

Грамотный выход из ситуации — Байесовский подход или получить дополнительное наблюдение (с последним сложнее).

Байесовский подход: β_1 и β_2 случайны.

$$\beta_i \sim N(0, 10^6)$$

Оценки получить апостериорно, они будут существовать.

Семинар. February, 9

Нам потребуются функции $glm(\cdot)$ (Generalized Linear Model) и $maBina$ (Marginal Effect for Binary Probit and Logit Model) из пакета $erer$:

```
library(erer) # glm
```

Загрузим данные из задачи 6.4., но с дополнительным наблюдением $x_5 = 3$, $y_5 = 0$ для того, чтобы избежать проблему perfect prediction:

```
df <- data.frame(x = c(2, 1, 3, 0, 3),
                 y = c(1, 0, 1, 0, 0))
```

Оценим по этим данным логит и пробит модели. Для этого нужно выбрать аргумент $link$, соответствующий определенной модели классификации:

```
logit <- glm(data = df, y ~ x,
             family = binomial(link = "logit"))
summary(logit)

##
## Call:
## glm(formula = y ~ x, family = binomial(link = "logit"), data = df)
##
## Deviance Residuals:
##       1        2        3        4        5
##  1.3443  -0.6454   0.9000  -0.3893  -1.4829
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)   -2.542      2.592  -0.981    0.327
## x              1.079      1.091   0.989    0.323
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6.7301  on 4  degrees of freedom
## Residual deviance: 5.3844  on 3  degrees of freedom
## AIC: 9.3844
##
## Number of Fisher Scoring iterations: 4
```

```
probit <- glm(data = df, y ~ x,
              family = binomial(link = "probit"))
summary(probit)

##
## Call:
## glm(formula = y ~ x, family = binomial(link = "probit"), data = df)
##
```

```
## Deviance Residuals:
##       1        2        3        4        5
##  1.3544  -0.6199   0.9030  -0.3300  -1.4793
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  -1.6164     1.4617  -1.106    0.269
## x             0.6810     0.6258   1.088    0.277
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 6.7301  on 4  degrees of freedom
## Residual deviance: 5.3314  on 3  degrees of freedom
## AIC: 9.3314
##
## Number of Fisher Scoring iterations: 6
```

```
library(memisc)

## Loading required package: lattice
## Loading required package: MASS
##
## Attaching package: 'memisc'
##
## The following objects are masked from 'package:stats':
##
##    contr.sum, contr.treatment, contrasts
##
## The following object is masked from 'package:base':
##
##    as.array

mtable(logit, probit) # a table of estimates for several models

##
## Calls:
## logit: glm(formula = y ~ x, family = binomial(link = "logit"), data = df)
## probit: glm(formula = y ~ x, family = binomial(link = "probit"), data = df)
##
## =====================================
##                        logit  probit
## -------------------------------------
## (Intercept)           -2.542  -1.616
##                       (2.592) (1.462)
## x                      1.079   0.681
##                       (1.091) (0.626)
## -------------------------------------
## Aldrich-Nelson R-sq.   0.212   0.219
## McFadden R-sq.         0.200   0.208
## Cox-Snell R-sq.        0.236   0.244
```

```
## Nagelkerke R-sq.        0.319   0.330
## phi                     1.000   1.000
## Likelihood-ratio        1.346   1.399
## p                       0.246   0.237
## Log-likelihood         -2.692  -2.666
## Deviance                5.384   5.331
## AIC                     9.384   9.331
## BIC                     8.603   8.550
## N                       5       5
## ========================================
```

Заметим, что так как $L(0,1) \approx N(0, \pi^2/3)$, то отношение коэффициентов в логит модели к коэффициентам в пробит модели приблизительно равно $\sqrt{\pi}/3$.

Задача 1.

$$y_i = \begin{cases} 1, & \text{если } y_i^* \geqslant 0 \\ 0, & \text{если } y_i^* < 0 \end{cases}$$

Оценена логит модель:

$$\hat{y}_i^* = -2.542 + 1.079 x_i$$

Прогноз на завтра: $x_F = 2$.

а) Найти $\mathbb{P}(\widehat{y_F = 1})$

б) Построить 95-% доверительный интервал для $\mathbb{P}(\widehat{y_F = 1})$ с использованием дельта-метода и без использования дельта-метода

Решение.

а)

$$\mathbb{P}(\widehat{y_F = 1}) = F\left(\hat{\beta}_1 + \hat{\beta}_2 x_F\right)$$

$F\left(\hat{\beta}_1 + \hat{\beta}_2 x_F\right)$ для логит модели — сигмоида в точке $\hat{\beta}_1 + \hat{\beta}_2 x_F$, для пробит модели — функция Лапласа в точке $\hat{\beta}_1 + \hat{\beta}_2 x_F$.

Для логит модели:

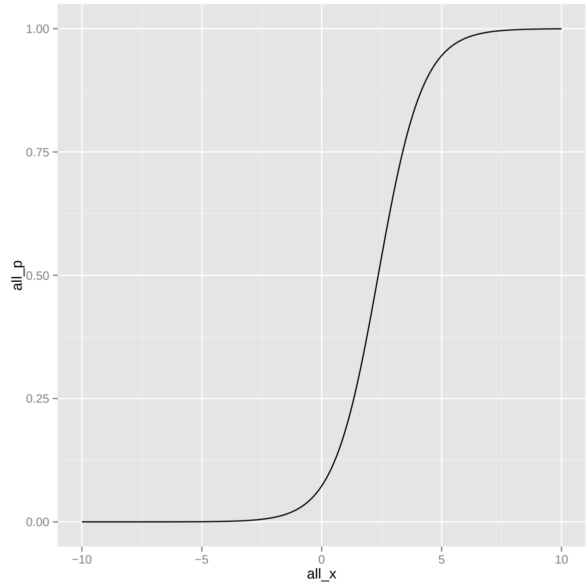
$$\mathbb{P}(\widehat{y_F = 1}) = \frac{e^{\hat{\beta}_1 + \hat{\beta}_2 x_F}}{1 + e^{\hat{\beta}_1 + \hat{\beta}_2 x_F}} \approx 0.405$$

```r
hb1_l <- as.numeric(coef(logit)[1])
hb2_l <- as.numeric(coef(logit)[2])
hb1_p <- as.numeric(coef(probit)[1])
hb2_p <- as.numeric(coef(probit)[2])
# hb --- hat beta, l --- logit, p --- probit

xf <- 2

q_l <- hb1_l + hb2_l*xf
plogis(q_l)

## [1] 0.4050925
```

Построим графики $\mathbb{P}(\widehat{y_F = 1})$ и предельного эффекта x_F:

График $\mathbb{P}(\widehat{y_F = 1})$:

```
all_x <- seq(-10, 10, by = 0.1)
all_p <- plogis(hb1_l + hb2_l*all_x)
qplot(all_x, all_p, geom = "line")
```

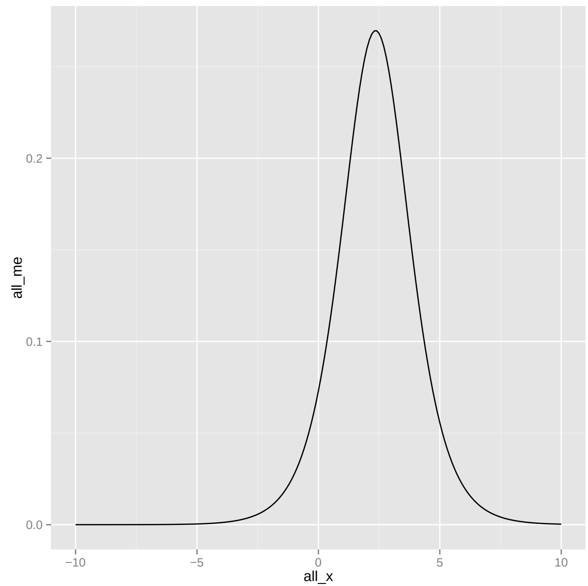


Предельный эффект x_F:

$$\frac{d\mathbb{P}(\widehat{y_F = 1})}{dx_F} = f_{\varepsilon_i}\left(\hat{\beta}_1 + \hat{\beta}_2 x_F\right) \cdot \hat{\beta}_2$$

График предельного эффекта:

```
all_me <- dlogis(hb1_l + hb2_l*all_x)*hb2_l
qplot(all_x, all_me, geom = "line")
```



б) Дельта-метод

Если $\hat{\theta}_1$, ..., $\hat{\theta}_n$ — оценки для параметра θ, $g(\cdot)$ дифференцируема в θ и

$$\sqrt{n}\left(\hat{\theta}_n - \theta\right) \xrightarrow{D} \mathcal{N}(0, \sigma^2)$$

то:

$$\sqrt{n}\left(g\left(\hat{\theta}_n\right) - g\left(\theta\right)\right) \xrightarrow{D} \mathcal{N}(0, \sigma^2 \cdot \left(g'\left(\theta\right)\right)^2)$$

Для нашей задачи:

$$\hat{\beta} = \begin{pmatrix} -2.542 \\ 1.019 \end{pmatrix}$$

$$g(\hat{\beta}) = \widehat{\mathbb{P}(y_F = 1)} = \Lambda\left(\hat{\beta}_1 + \hat{\beta}_2 x_F\right)$$

При этом:

$$g(\beta) = \Lambda\left(\beta_1 + \beta_2 x_F\right)$$

Настало время дельта-метода:

$$\mathbb{V}ar\left(g(\hat{\beta})\right) \approx \left(\frac{dg}{d\beta}\left(\hat{\beta}\right)\right)' \widehat{V}\left(\hat{\beta}\right)\left(\frac{dg}{d\beta}\left(\hat{\beta}\right)\right)$$

$$\frac{dg}{d\beta}\left(\hat{\beta}\right) = \begin{pmatrix} \frac{d\Lambda}{d\beta_1}(\hat{\beta}_1 + \hat{\beta}_2 x_F) \\ \frac{d\Lambda}{d\beta_2}(\hat{\beta}_1 + \hat{\beta}_2 x_F) \end{pmatrix} = \begin{pmatrix} f_{\varepsilon_i}(\hat{\beta}_1 + \hat{\beta}_2 x_F) \cdot 1 \\ f_{\varepsilon_i}(\hat{\beta}_1 + \hat{\beta}_2 x_F) \cdot x_F \end{pmatrix}$$

```
dgdb1 <- dlogis(hb1_l + hb2_l*xf)*1; dgdb1

## [1] 0.2409926

dgdb2 <- dlogis(hb1_l + hb2_l*xf)*xf; dgdb2

## [1] 0.4819851

grad <- c(dgdb1, dgdb2)
grad

## [1] 0.2409926 0.4819851

vcov(logit)

##             (Intercept)         x
## (Intercept)    6.719864 -2.584741
## x             -2.584741  1.189710

Var_g <- t(grad) %*% vcov(logit) %*% grad
Var_g <- as.numeric(Var_g); Var_g

## [1] 0.06619309

sqrt(Var_g)

## [1] 0.2572802
```

Проверим то, что мы получили с помощью функции *deltamethod*$(\cdot)$ из пакета *msm*, позволяющую сразу найти стандартное отклонение оценки:

```
library(msm) # delta method
deltamethod(~ 1/(1 + exp(-x1 - x2*xf)), c(hb1_l, hb2_l),
            vcov(logit))

## [1] 0.2572802
```

Теперь легко построить доверительный интервал, учитывая, что по дельта-методу оценка сходится по распределению к нормальному распределению:

```
low <- plogis(q_l) - qnorm(0.975)*sqrt(Var_g)
high <- plogis(q_l) + qnorm(0.975)*sqrt(Var_g)
conf <- data.frame(low, high)
conf

##             low      high
## 1 -0.09916737 0.9093524
```

Видим, что доверительный интервал выходит из области допустимых значений и имеет большой разброс. Но этот результат на основе дельта-метода стоило ожидать, так как очень мало наблюдений.

Семинар. February, 16

Продолжение задачи с прошлого семинара.

Коэффициенты и ковариационная матрица коэффициентов для оцененной модели:

```
logit$coefficients

## (Intercept)           x
##   -2.541897    1.078803

vcov(logit)

##             (Intercept)         x
## (Intercept)    6.719864 -2.584741
## x             -2.584741  1.189710
```

Прогнозируем в случае $x_F = 2$ или $x_F = 4$.

```
new <- data.frame(x = c(2, 4))
```

Можем получить оценку ненаблюдаемой склонности увидеть привидение:

```
pred <- predict(logit, newdata = new, se.fit = T)
pred

## $fit
##          1          2
## -0.3842904  1.7733160
##
## $se.fit
##        1        2
## 1.067586 2.253285
##
## $residual.scale
## [1] 1
```

Чтобы привести ее к вероятности, нужно применить функцию распределения логистического распределения:

```
plogis(pred$fit)

##         1         2
## 0.4050925 0.8548696
```

А можем сразу получить вероятность, прописав аргумент *type*:

```
pred2 <- predict(logit, newdata = new, type = 'response')
pred2

##         1         2
## 0.4050925 0.8548696
```

Построим доверительный интервал для прогноза для ненаблюдаемой склонности:

$$\hat{y}_F^* = \hat{\beta}_1 + \hat{\beta}_2 x_F$$

$$\mathbb{E}\left(\hat{y}_F^* | x_F\right) = \beta_1 + \beta_2 x_F$$
$$\mathbb{V}ar\left(\hat{y}_F^* | x_F\right) = \mathbb{V}ar(\hat{\beta}_1) + x_F^2 \mathbb{V}ar(\hat{\beta}_2) + 2x_F \mathbb{C}ov(\hat{\beta}_1, \hat{\beta}_2)$$
$$\hat{y}_F^* | x_F \sim \mathcal{N}\left(\beta_1 + \beta_2 x_F, \mathbb{V}ar\left(\hat{y}_F^* | x_F\right)\right)$$

Например, для $x_F = 2$:

```
var.hat <- vcov(logit)[1, 1] + 2^2*vcov(logit)[2, 2] + 2*2*vcov(logit)[2, 1]
var.hat

## [1] 1.139739

sqrt(var.hat)

## [1] 1.067586
```

То же самое мы получили в нашем прогнозе:

```
pred_df <- as.data.frame(pred)
pred_df

##          fit   se.fit residual.scale
## 1 -0.3842904 1.067586              1
## 2  1.7733160 2.253285              1
```

Новые значения объясняющей переменной:

```
pred_df$x <- c(2, 4)
pred_df

##          fit   se.fit residual.scale x
## 1 -0.3842904 1.067586              1 2
## 2  1.7733160 2.253285              1 4
```

Строим доверительный интервал для склонности увидеть привидение и несимметричный доверительный интервал для вероятности увидеть привидение:

```
library(dplyr)

##
## Attaching package: 'dplyr'
##
## The following objects are masked from 'package:memisc':
##
##    collect, query, rename
##
## The following object is masked from 'package:MASS':
##
```

```
##     select
##
## The following object is masked from 'package:stats':
##
##     filter
##
## The following objects are masked from 'package:base':
##
##     intersect, setdiff, setequal, union
```

```
pred_df <- pred_df %>%
  mutate(left_ci = fit - qnorm(0.975)*se.fit,
         right_ci = fit + qnorm(0.975)*se.fit) %>%
  mutate(left_ci_p = plogis(left_ci),
         right_ci_p = plogis(right_ci),
         p_hat = plogis(fit))
pred_df
```

Теперь построим симметричный доверительный интервал для вероятности, используя дельта-метод:

```
pred_df <- pred_df %>%
  mutate(se_DM = se.fit*dlogis(fit)) %>%
  mutate(left_ci_p_DM = p_hat - qnorm(0.975)*se_DM,
         right_ci_p_DM = p_hat + qnorm(0.975)*se_DM)
pred_df
```

```
##           fit   se.fit residual.scale x   left_ci right_ci  left_ci_p
## 1 -0.3842904 1.067586              1 2 -2.476720 1.708139 0.07750642
## 2  1.7733160 2.253285              1 4 -2.643042 6.189674 0.06641918
##   right_ci_p     p_hat     se_DM left_ci_p_DM right_ci_p_DM
## 1  0.8465947 0.4050925 0.2572802  -0.09916737     0.9093524
## 2  0.9979537 0.8548696 0.2795597   0.30694269     1.4027964
```

Посмотрим на предельный эффект x_i на вероятность увидеть привидение:

```
library(erer)
logit_w_memory <- glm(data = df, y ~ x,
                      family = binomial(link = "logit"),
                      x = TRUE)
```

Предельный эффект для среднего потребления коньяка:

$$maBina = f_{\varepsilon_i}\left(\beta_1 + \beta_2 \frac{\sum x_i}{n}\right)\beta_2$$

```
maBina(logit_w_memory)

##             effect error t.value p.value
## (Intercept) -0.582 0.477  -1.218   0.310
## x            0.247 0.233   1.060   0.367
```

Средний предельный эффект:

$$maBina(x.mean = FALSE) = \frac{1}{n} \sum f_{\varepsilon_i} \left(\beta_1 + \beta_2 x_i \right) \beta_2$$

```
maBina(logit_w_memory, x.mean = F)

##             effect error t.value p.value
## (Intercept) -0.460 0.378  -1.218   0.310
## x            0.195 0.184   1.060   0.367
```