Прикладная статистика в R
Лекция 9. Другие регрессионные модели. Ридж-регрессия. Медианная регрессия

Елена Михайловна Парилина

д. ф.-м. н., проф.

2020

Основные предположения регрессионного анализа

Сформулируем основные предположения регрессионного анализа, которые относятся к случайным компонентам ε_i, $i = 1, \ldots, n$.

Первая группа предположений:

- Случайные величины ε_i, $i = 1, \ldots, n$ образуют так называемый слабый белый шум, т. е. последовательность центрированных ($E\varepsilon_i = 0$, $i = 1, \ldots, n$) и некоррелированных ($E(\varepsilon_l \varepsilon_u) = 0$ при $l \neq u$) случайных величин с одинаковыми дисперсиями σ^2 ($E\varepsilon_i^2 = \sigma^2$, $i = 1, \ldots n$).

Кроме первой группы предположений будем также рассматривать вторую группу, в которой сформулируем предположение о совместном распределении случайных величин ε_i, $i = 1, \ldots, n$.

Вторая группа предположений:

- Совместное распределение случайных величин ε_i, $i = 1, \ldots, n$ является нормальным распределением с нулевым вектором математических ожиданий и ковариационной матрицей $\sigma^2 E_n$, т. е. случайный вектор $\varepsilon^T = (\varepsilon_1, \ldots, \varepsilon_n) \sim N(0, \sigma^2 E_n)$, где E_n — единичная матрица порядка $n \times n$.

Ридж регрессия

Модель ридж регрессии

Модель линейной регрессии в матричной форме:

$$Y = X\beta + \varepsilon,$$

где $Y = (y_1, \ldots, y_n)^\intercal$, $\beta = (\beta_0, \beta_1, \ldots, \beta_k)^\intercal$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\intercal$,

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & \ldots & x_{2k} \\ \ldots & \ldots & \ldots & \ldots & \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nk} \end{pmatrix}.$$

$|X^\intercal X| = 0$: проблема мультиколлинеарности (две или более независимые переменные в регрессионной модели значимо коррелируют).

Модель ридж регрессии

Сформулируем альтернативную математическую задачу:

$$\min_{\beta}\left[\sum_{i=1}^{n}(y_i-(X\beta)_i)^2+\lambda\sum_{j=0}^{k}\beta_j^2\right],\quad \lambda>0.$$

Оценки: $\hat{\beta}(\lambda)=(X^TX+\lambda I_{k+1})^{-1}X^TY$.

Модель ридж регрессии: $\hat{y}(x,\lambda)=\hat{\beta}_0(\lambda)+\hat{\beta}_1(\lambda)x_1+\ldots+\hat{\beta}_k(\lambda)x_k$.

Свойства:

1. Для любой матрицы X и для любого параметра $\lambda>0$, существует матрица $(X^TX+\lambda I_{k+1})^{-1}$, тогда оценки $\hat{\beta}(\lambda)$ единственны.
2. $\hat{\beta}(\lambda)\to\hat{\beta}$, если $\lambda\to 0$.
3. $\hat{\beta}(\lambda)\to 0$, если $\lambda\to\infty$.

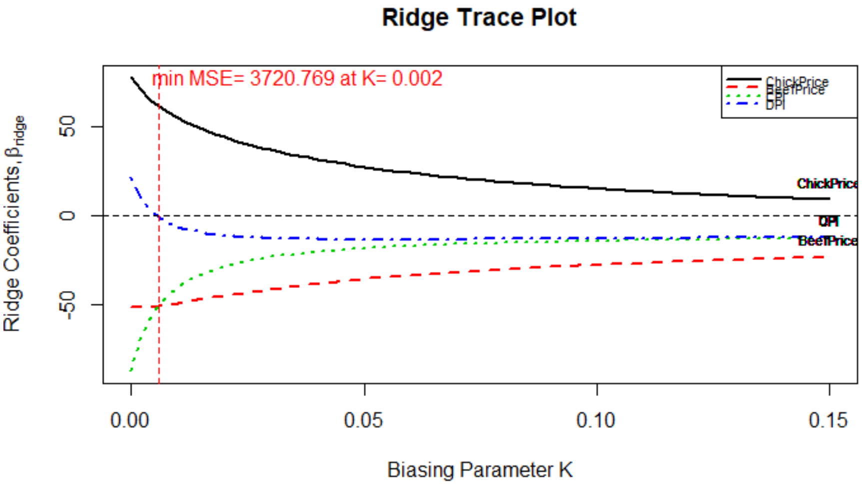
Ридж регрессия в R

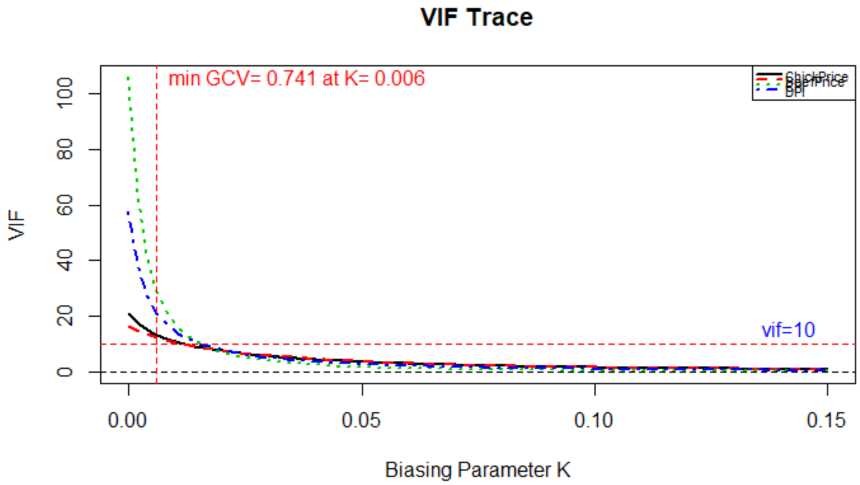
Библиотека MASS

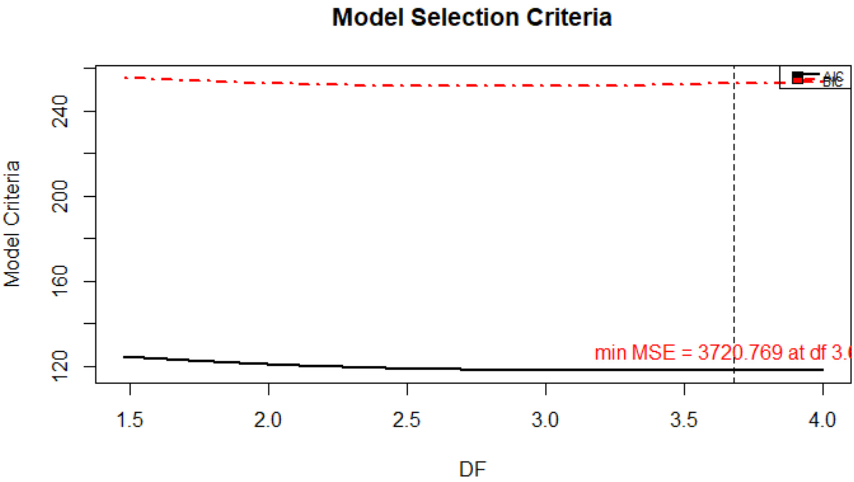
```
library(MASS)
lm.ridge(formula, data, subset, na.action, lambda = 0, model =
FALSE, x = FALSE, y = FALSE, contrasts = NULL, ...)
```

Библиотека lmridge

```
library(lmridge)
beef_fidgemod <- lmridge(beef1$BeefConsump~.,data=beef1, K =
seq(0, 0.15, 0.002))
plot(beef_fidgemod, type = "ridge") # изображаем хвосты в
зависимости от лямбда
plot(beef_fidgemod, type = "vif") # смотрим, как меняются
коэффициенты vif для каждого предиктора
info.plot(beef_fidgemod) # смотрим, как меняются коэффициенты AIC
и BIC
```


Ridge Trace Plot
min MSE= 3720.769 at K= 0.002
ChickPrice
BeefPrice
DPI
ChickPrice
DPI
BeefPrice
Ridge Coefficients, β_{ridge}
50
0
-50
0.00
0.05
0.10
0.15
Biasing Parameter K


VIF Trace
min GCV= 0.741 at K= 0.006
vif=10
VIF
0
20
40
60
80
100
0.00
0.05
0.10
0.15
Biasing Parameter K

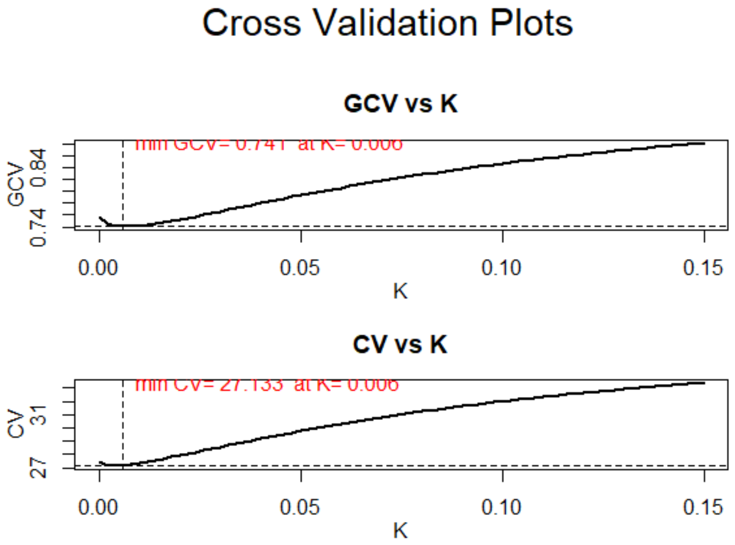

Model Selection Criteria
Model Criteria
240
200
160
120
AIC
BIC
min MSE = 3720.769 at df 3.
1.5
2.0
2.5
3.0
3.5
4.0
DF

Ридж регрессия с lmridge

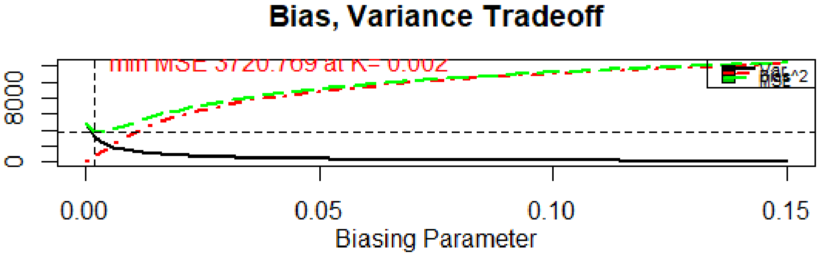
Библиотека lmridge

```
cv.plot(beef_fidgemod)
# cv.plot полезна для нахождения параметра K.
bias.plot(beef_fidgemod)
# bias.plot также используется для нахождения параметра K.
```

Ридж регрессия с lmridge



Ридж регрессия с `lmridge`



Ридж регрессия с lmridge

Построим модель ридж-регрессии с рекомендованным значением параметра лямбда, равным 0.002:

```
> beefRmod1<-lmridge(beef1$BeefConsump~., data=beef1,K=0.002)
> summary(beefRmod1)

Call:
lmridge.default(formula = beef1$BeefConsump ~ ., data = beef1,
    K = 0.002)


Coefficients: for Ridge parameter K= 0.002
            Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|)
Intercept   1.0460e+02   -1.0344e+05  3.4053e+05      -0.3038   0.7633
ChickPrice  5.3620e-01    7.0650e+01  2.0466e+01       3.4521   0.0016 **
BeefPrice  -9.8900e-02   -5.1860e+01  1.8963e+01      -2.7348   0.0102 *
CPI        -2.4810e-01   -6.9237e+01  3.8832e+01      -1.7830   0.0843 .
DPI         2.0000e-04    1.0101e+01  3.0142e+01       0.3351   0.7398
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Ridge Summary
       R2     adj-R2   DF ridge          F        AIC        BIC
  0.71250    0.68550    3.67659  21.69913 118.05405 252.88268
Ridge minimum MSE= 3720.769 at K= 0.002
P-value for F-test ( 3.67659 , 32.06058 ) = 1.751433e-08
-------------------------------------------------------------------
```

Полезные функции в `lmridge`

- `coef(mod)` : коэффициенты модели
- `predict(mod)` : предсказание значений зависимой переменной
- `residuals(mod)` : остатки модели
- `plot(mod)` : графики
- `vif(mod)` : коэффициенты VIF
- `vcov(mod)` : ковариационная матрица
- `fitted(mod)` : предсказанные значения
- `rstats1(mod)` : статистика 1 для оценки значимости модели
- `rstats2(mod)` : статистика 2 для оценки значимости модели

Ридж регрессия с lmridge

```
> coef(beefRmod1)
 Intercept ChickPrice  BeefPrice        CPI        DPI
 104.59867    0.53620   -0.09893   -0.24806    0.00024
> residuals(beefRmod1)
      K=0.002
1 -11.9700082
2  -8.0981765
3  -3.9966087
4  -1.0929719
5  -0.1164550
6   3.9691389
7   2.9484402
```

Ридж регрессия с lmridge

```
> beefRmod2<-lmridge(beef1$BeefConsump~., data=beef1,K=0.02)
> summary(beefRmod2)

Call:
lmridge.default(formula = beef1$BeefConsump ~ ., data = beef1,
    K = 0.02)


Coefficients: for Ridge parameter K= 0.02
             Estimate Estimate (Sc) StdErr (Sc) t-value (Sc) Pr(>|t|)
Intercept    108.5533   133938.5577 160538.4500       0.8343   0.4102
ChickPrice     0.3361       44.2842     13.9976       3.1637   0.0034 **
BeefPrice     -0.0860      -45.0855     14.0885      -3.2002   0.0031 **
CPI           -0.1024      -28.5757     13.4522      -2.1242   0.0414 *
DPI           -0.0003      -11.1920     14.2094      -0.7876   0.4366
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Ridge Summary
       R2     adj-R2   DF ridge          F        AIC        BIC
  0.64120    0.60760    2.62960   20.62801 118.56214 251.73284
Ridge minimum MSE= 6522.165 at K= 0.02
P-value for F-test ( 2.6296 , 32.76251 ) = 2.674541e-07
-------------------------------------------------------------------

> vif(beefRmod1)
        ChickPrice BeefPrice      CPI      DPI
k=0.002     17.229  14.79141 62.02766 37.37336
> vif(beefRmod2)
       ChickPrice BeefPrice     CPI     DPI
k=0.02    7.66175   7.76152 7.07623 7.89536
```

Ридж регрессия с lmridge

Анализируем две модели, выбираем наиболее приемлемую для нас:

```
> rstats1(beefRmod1)

Ridge Regression Statistics 1:

        Variance   Bias^2      MSE rsigma2       F     R2 adj-R2       CN
K=0.002 3194.945 525.8239 3720.769 24.3107 21.6991 0.7125 0.6855 465.2656

> rstats1(beefRmod2)

Ridge Regression Statistics 1:

       Variance   Bias^2      MSE rsigma2      F     R2 adj-R2       CN
K=0.02 777.2882 5744.877 6522.165  25.573 20.628 0.6412 0.6076 148.5429
```

Ридж регрессия с lmridge

Сравним ридж модель с линейной моделью:

```
> beeflin<-lm(beef1$BeefConsump~., data=beef1)
> summary(beeflin)

Call:
lm(formula = beef1$BeefConsump ~ ., data = beef1)

Residuals:
    Min      1Q  Median      3Q     Max 
-11.901  -3.248   0.178   2.971  12.861 

Coefficients:
              Estimate Std. Error t value Pr(>|t|)    
(Intercept)  1.039e+02  5.280e+00  19.679  < 2e-16 ***
ChickPrice   5.907e-01  1.748e-01   3.380  0.00197 ** 
BeefPrice   -9.864e-02  3.863e-02  -2.553  0.01582 *  
CPI         -3.146e-01  1.842e-01  -1.707  0.09774 .  
DPI          5.101e-04  9.025e-04   0.565  0.57599    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.002 on 31 degrees of freedom
Multiple R-squared:  0.7312,    Adjusted R-squared:  0.6965 
F-statistic: 21.08 on 4 and 31 DF,  p-value: 1.765e-08
```

Ридж регрессия с ridge

Ридж регрессия в библиотеке ridge

```
library(ridge)
linRidgeMod <- linearRidge(formula, data=dataset)
summary(linRidgeMod)
```

```
> library(ridge)
> beefRidge2<-linearRidge(beef1$BeefConsump~., data=beef1)
> summary(beefRidge2)

Call:
linearRidge(formula = beef1$BeefConsump ~ ., data = beef1)


Coefficients:
              Estimate Scaled estimate Std. Error (scaled) t value (scaled)
(Intercept)  1.137e+02              NA                  NA               NA
ChickPrice   1.361e-01       1.794e+01           7.701e+00            2.329
BeefPrice   -5.690e-02      -2.982e+01           7.820e+00            3.814
CPI         -5.400e-02      -1.507e+01           4.787e+00            3.149
DPI         -3.116e-04      -1.312e+01           6.474e+00            2.027
            Pr(>|t|)
(Intercept)       NA
ChickPrice  0.019849 *
BeefPrice   0.000137 ***
CPI         0.001638 **
DPI         0.042671 *
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Ridge parameter: 0.08438167, chosen automatically, computed using 1 PCs

Degrees of freedom: model 1.76 , variance 1.22 , residual 2.3
```

Квантильная регрессия

Модель квантильной регрессии

Запишем модель регрессии в матричной форме:

$$Y = X\beta + \varepsilon,$$

где $Y = (y_1, \ldots, y_n)^\intercal$, $\beta = (\beta_0, \beta_1, \ldots, \beta_k)^\intercal$, $\varepsilon = (\varepsilon_1, \ldots, \varepsilon_n)^\intercal$,

$$X = \begin{pmatrix} 1 & x_{11} & x_{12} & \ldots & x_{1k} \\ 1 & x_{21} & x_{22} & \ldots & x_{2k} \\ \ldots & \ldots & \ldots & \ldots & \\ 1 & x_{n1} & x_{n2} & \ldots & x_{nk} \end{pmatrix}.$$

Будем решать оптимизационную задачу:

$$\min_{\beta(\tau)} \left[\sum_{i\colon\, y_i \geq (X\beta)_i} \tau \left| y_i - (X\beta)_i \right| + \sum_{i\colon\, y_i < (X\beta)_i} (1-\tau) \left| y_i - (X\beta)_i \right| \right], \quad \tau \in (0,1).$$

LAD (Least absolute deviations) оценки или оценки метода наименьших модулей: $\hat{\beta}(\tau)$.

Квантильная регрессия: $\hat{y}(x, \tau) = \hat{\beta}_0(\tau) + \hat{\beta}_1(\tau)x_1 + \ldots + \hat{\beta}_k(\tau)x_k$.

Медианная регрессия: $\hat{y}(x, \frac{1}{2}) = \hat{\beta}_0(\frac{1}{2}) + \hat{\beta}_1(\frac{1}{2})x_1 + \ldots + \hat{\beta}_k(\frac{1}{2})x_k$.

Квантильная регрессия в R

Функции в библиотеке "quantreg"

```
install.packages("quantreg")
library(quantreg)
data(engel)
```

Данные `engel` содержат 235 наблюдений с 2 переменными: расходы семьи на питание и доход семьи.

Задача

Построить модель медианной и нескольких квантильных регрессий для различных τ, где зависимой будет переменная расходов на питание, а независимой — доход семьи. Сравнить модели между собой и с моделью линейной регрессии.

Функция "rq"

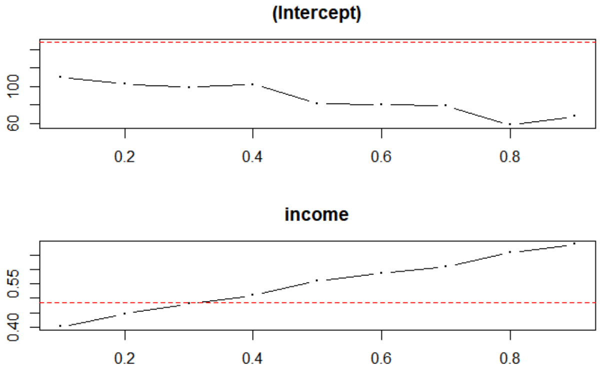
qreg1 < − rq(foodexp ∼ income, tau = .5, data = engel)
summary(myqreg)

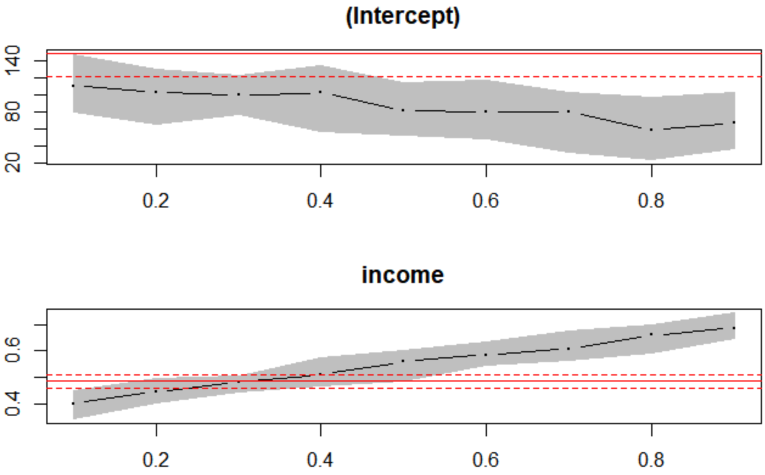
```
> qreg1 <- rq(foodexp ~ income, tau = .5, data = engel)
> summary(qreg1)

Call: rq(formula = foodexp ~ income, tau = 0.5, data = engel)

tau: [1] 0.5

Coefficients:
            coefficients lower bd  upper bd 
(Intercept)  81.48225     53.25915 114.01156
income        0.56018      0.48702   0.60199
```

Построим квантильные регрессии для разных значений параметра τ от 0.1 до 0.9 с шагом 0.1. Нарисуем графики изменения коэффициентов регрессии:

Функция rq

```
myqreg <- rq(foodexp ~ income, tau = 1:9/10, data = engel)
plot(myqreg)
plot(summary(myqreg))
```




(Intercept)
20
80
140
0.2
0.4
0.6
0.8
income
0.4
0.6
0.2
0.4
0.6
0.8

Сплошная красная линия — это коэффициент линейной регрессии (МНК), а пунктирные красные линии — границы доверительных интервалов для коэффициентов линейной регрессии. Каждая черная точка — это коэффициент в модели квантильной регрессии, где квантиль указывается на оси абсцисс. Светло-серая область вокруг черных точек — доверительный интервал для коэффициентов квантильной регрессии. Нижняя граница доверительного интервала для коэффициентов квантильной регрессии значительно ниже соответствующей границы доверительного интервала для коэффициентов линейной регрессии, а верхние границы значительно выше соответствующих верхних границ для коэффициентов линейной регрессии.

Использование модели для прогнозирования: функция predict

```
predict(object, newdata, interval = c("none" , "confidence"),
level = .95, na.action = na.pass, ...)
```

Сравнение медианной и линейной регрессий

Модель линейной регрессии

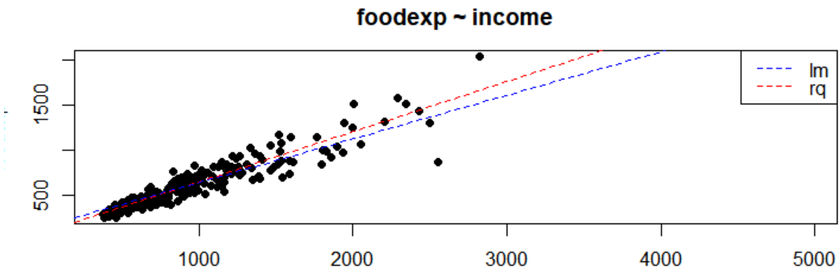
```
lin1<-lm(foodexp ~ income, data = engel)
```

```
> summary(lin1)

Call:
lm(formula = foodexp ~ income, data = engel)

Residuals:
    Min      1Q  Median      3Q     Max
-725.70  -60.24   -4.32   53.41  515.77

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept) 147.47539   15.95708   9.242   <2e-16 ***
income        0.48518    0.01437  33.772   <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 114.1 on 233 degrees of freedom
Multiple R-squared:  0.8304,    Adjusted R-squared:  0.8296
F-statistic:  1141 on 1 and 233 DF,  p-value: < 2.2e-16
```

Сравнение медианной и линейной регрессий

```
> plot(foodexp ~ income, data = engel, pch = 16, main = "foodexp ~ income")
> abline(qreg1, col = "red", lty = 2)
> abline(lin1, col = "blue", lty = 2)
```

Красная линия: медианная регрессия,
Синяя линия: линейная регрессия:

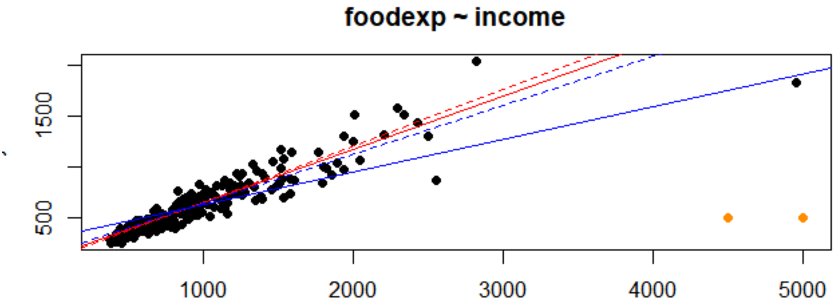


Линейная vs медианная регрессия

Медианная регрессия (т.е. 0.5-квантильная регрессия) иногда предпочтительнее линейной регрессии, потому что она «устойчива к выбросам». Продемонстрируем это на примере.

```
> y <- c(engel$foodexp, 500, 500)
> x <- c(engel$income, 4500, 5000)
> plot(y ~ x, pch = 16, main = "foodexp ~ income")
> points(c(4500, 5000), c(500, 500), pch = 16, col = "dark orange")
> abline(qreg1, col = "red", lty = 2)
> abline(rq(y ~ x), col = "red")
> abline(lin1, col = "blue", lty = 2)
> abline(lm(y ~ x), col = "blue")
```

1. Добавим два выброса к данным (окрашены оранжевым цветом) и посмотрим, как это повлияет на оба уравнения регрессии.
2. Изобразим уравнения линейной регрессии голубым цветом (до включения выбросов: пунктирная линия, после включения выбросов: сплошная линия).
3. Изобразим уравнения медианной регрессии красным цветом (до включения выбросов: пунктирная линия, после включения выбросов: сплошная линия).



Нелинейная регрессия

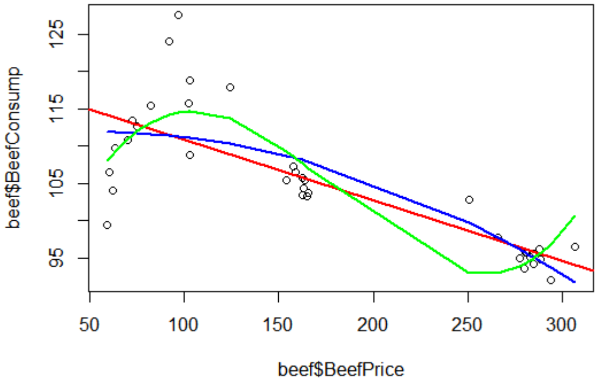
Полиномиальные модели регрессии

Квадратичная и кубическая модели в R

```
beef_quadratic <- lm(beef$BeefConsump ~ beef$BeefPrice +
I(beef$BeefPrice^2), data = beef)
beef_cubic <- lm(beef$BeefConsump ~ poly(beef$BeefPrice, degree =
3, raw = TRUE), data = beef)
order_id <- order(beef$BeefPrice)
lines(x = beef$BeefPrice[order_id], y =
fitted(beef_quadratic)[order_id], col = "blue" , lwd = 2)
plot(beef$BeefPrice,beef$BeefConsump)
abline(beef_lin, col = "red" , lwd = 2)
lines(x = beef$BeefPrice[order_id], y =
fitted(beef_quadratic)[order_id], col = "blue" , lwd = 2)
lines(x = beef$BeefPrice[order_id], y =
fitted(beef_cubic)[order_id], col = "green" , lwd = 2)
```

Анализ моделей

Красный: линейная регрессия,
Синий: квадратичная регрессия,
Зеленый: кубическая регрессия.



Анализ моделей

```
> load("~/beef.RData")
> beef_lin<-lm(beef$BeefConsump~beef$BeefPrice)
> summary(beef_lin)

Call:
lm(formula = beef$BeefConsump ~ beef$BeefPrice)

Residuals:
     Min       1Q   Median       3Q      Max
-14.7115  -2.0022  -0.4843   0.6484  16.3302

Coefficients:
                Estimate Std. Error t value Pr(>|t|)
(Intercept)    119.05052    2.08600  57.071  < 2e-16 ***
beef$BeefPrice  -0.08133    0.01069  -7.606 7.75e-09 ***
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.605 on 34 degrees of freedom
Multiple R-squared:  0.6298,    Adjusted R-squared:  0.619
F-statistic: 57.85 on 1 and 34 DF,  p-value: 7.747e-09
```

Анализ моделей

```
> summary(beef_quadratic)

Call:
lm(formula = beef$BeefConsump ~ beef$BeefPrice + I(beef$BeefPrice^2),
    data = beef)

Residuals:
     Min       1Q   Median       3Q      Max
-12.3527  -2.4654  -0.9538   1.5945  16.2355

Coefficients:
                      Estimate Std. Error t value Pr(>|t|)
(Intercept)          1.110e+02  5.104e+00  21.742   <2e-16
beef$BeefPrice       3.325e-02  6.733e-02   0.494   0.6247
I(beef$BeefPrice^2) -3.132e-04  1.818e-04  -1.722   0.0944

(Intercept)         ***
beef$BeefPrice
I(beef$BeefPrice^2) .
---
Signif. codes:
0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 5.45 on 33 degrees of freedom
Multiple R-squared:  0.6604,    Adjusted R-squared:  0.6398
F-statistic: 32.08 on 2 and 33 DF,  p-value: 1.826e-08
```

Анализ моделей

```
> summary(beef_cubic)

Call:
lm(formula = beef$BeefConsump ~ poly(beef$BeefPrice, degree = 3,
    raw = TRUE), data = beef)

Residuals:
    Min      1Q  Median      3Q     Max
-8.4885 -3.3005 -0.3731  1.4229 12.9806

Coefficients:
                                            Estimate
(Intercept)                                7.068e+01
poly(beef$BeefPrice, degree = 3, raw = TRUE)1  9.742e-01
poly(beef$BeefPrice, degree = 3, raw = TRUE)2 -6.553e-03
poly(beef$BeefPrice, degree = 3, raw = TRUE)3  1.205e-05
                                           Std. Error
(Intercept)                                1.191e+01
poly(beef$BeefPrice, degree = 3, raw = TRUE)1  2.650e-01
poly(beef$BeefPrice, degree = 3, raw = TRUE)2  1.722e-03
poly(beef$BeefPrice, degree = 3, raw = TRUE)3  3.313e-06
                                           t value
(Intercept)                                  5.937
poly(beef$BeefPrice, degree = 3, raw = TRUE)1   3.676
poly(beef$BeefPrice, degree = 3, raw = TRUE)2  -3.805
poly(beef$BeefPrice, degree = 3, raw = TRUE)3   3.638
```

Анализ моделей

```
                                              Pr(>|t|)
(Intercept)                                    1.3e-06
poly(beef$BeefPrice, degree = 3, raw = TRUE)1 0.000861
poly(beef$BeefPrice, degree = 3, raw = TRUE)2 0.000605
poly(beef$BeefPrice, degree = 3, raw = TRUE)3 0.000958

(Intercept)                                   ***
poly(beef$BeefPrice, degree = 3, raw = TRUE)1 ***
poly(beef$BeefPrice, degree = 3, raw = TRUE)2 ***
poly(beef$BeefPrice, degree = 3, raw = TRUE)3 ***
---
Signif. codes:
0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 4.655 on 32 degrees of freedom
Multiple R-squared:  0.7597,    Adjusted R-squared:  0.7372
F-statistic: 33.73 on 3 and 32 DF,  p-value: 5.014e-10
```

Уравнения полиномиальных регрессий

Линейная регрессия

$$\widehat{BeefConsump} = 119.05 - 0.08133 \times BeefPrice.$$

Квадратичная регрессия

$$\widehat{BeefConsump} = 111.0 + 0.03325 \times BeefPrice - 0.0003132 \times BeefPrice^2.$$

Кубическая регрессия

$$\begin{aligned}\widehat{BeefConsump} =& 70.68 + 0.9742 \times BeefPrice - 0.006553 \times BeefPrice^2 \\ & - 0.00001205 \times BeefPrice^3.\end{aligned}$$

Итоги

Что мы узнали на Лекции 9?

- какие предположения регрессионного анализа могут нарушаться.
- как построить медианную регрессию.
- как построить ридж-регрессию и зачем это нужно.
- как построить нелинейную модель регрессии.

Что мы узнаем на Лекции 10?

Мы узнаем,

- как решается задача классификации объектов с помощью алгоритма «случайных лесов».
- что такое задача кластеризации и какие существуют методы ее решения.
- как реализовать алгоритм k средних для задачи кластеризации.
- как решать перечисленные задачи в R.

Спасибо за внимание!