Оглавление

[Задание 1 2](#_Toc27260497)

[Задание 2 8](#_Toc27260498)

[Задание 3 10](#_Toc27260499)

[Задание 4 13](#_Toc27260500)

[Задание 5 15](#_Toc27260501)

# **Задание 1**

1) (функция спроса)

2) (функция потребления)

3) (функция спроса-потребления)

4) (функция спроса с учетом цены на товары-заменители)

Загружаем набор данных и создаём логарифмированные значения переменных.

library("lmtest")

df <- read.table("Desktop/data.txt", header=TRUE)

df <- lapply(df, function(x) as.numeric(gsub(",", ".", gsub("\\.", "", as.character(x)))))

df <- as.data.frame(df)

for (var\_name in c('Y', paste('X',1:4,sep = ""))){

var\_log\_name <- paste(var\_name,'log', sep='\_')

df[var\_log\_name] <- log(df[var\_name])

}

Сразу построим четыре регрессионные модели:

1)Функция спроса

model1 <- lm(Y\_log ~ X2\_log, data = df)

2)Функция потребления

model2 <- lm(Y\_log ~ X1\_log, data = df)

3)Функция спроса-потребления

model3 <- lm(Y\_log ~ X1\_log + X2\_log, data = df)

4)Функция спроса с учетом цены на товары-заменители

model4 <- lm(Y\_log ~ X2\_log + X3\_log + X4\_log, data = df)

Напишем функцию, которая будет анализировать модель – проводить регрессионный анализа (в частности проверять значимость всех коэффициентов уравнения, находить оценку остаточной дисперсии ошибки по формуле, скорректированный коэффициент детерминации, F-статистику и проверять гипотезу о значимости уравнения регрессии), искать статистику Дарбина-Уотсона, среднюю относительную ошибку аппроксимации по формуле , потенцировать полученные точечные оценки коэффициентов и выводить результаты в консоль.

analyse <- function(model){

print(summary(model))

print(dwtest(model, alternative = "greater"))

cat("Average relative error of approximation:", sum(abs(predict(model) - df['Y\_log'])/(predict(model)\*length(df$t)))\*100, "%\n\n")

for (var\_name in names(model$coefficients)){

print(exp(model$coefficients[var\_name]))

}

}

Вызовем в цикле функцию анализа для каждой модели

for (model in list(model1, model2, model3, model4)){

analyse(model)

}

1. Call:

lm(formula = Y\_log ~ X2\_log, data = df)

Residuals:

Min 1Q Median 3Q Max

-0.137580 -0.049196 0.008714 0.058689 0.120432

Coefficients:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Estimate** | **Std.Error** | **tvalue** | **Pr(>|t|)** |
| (Intercept) | 1.57772 | 0.31051 | 5.081 | 9.25e-05\*\*\* |
| X2\_log | 0.55272 | 0.07989 | 6.918 | 2.49e-06\*\*\* |

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.07812 on 17 degrees of freedom

Multiple R-squared: 0.7379, Adjusted R-squared: 0.7225

F-statistic: 47.86 on 1 and 17 DF, p-value: 2.485e-06

Durbin-Watson test

data: model

DW = 0.77954, p-value = 0.000399

alternative hypothesis: true autocorrelation is greater than 0

Average relative error of approximation: 1.683106 %

(Intercept)

4.843888

X2\_log

1.737969

Y\_hat = 4.843888 \* X21.737969

Все коэффициенты значимы. Оценка остаточной дисперсии ошибки = 0.07812. Исправленный коэффициент детерминации = 0.7225. F-статистика = 47.86 больше критического значения, что означает значимость уравнения регрессии - все коэффициенты не равны нулю одновременно. Статистика Дарбина-Уотсона = 0.77954 - гипотеза об отсутствии автокорреляции остатков отвергается так как p-value = 0.000399 меньше критического уровня 0.05 - принимается альтернативная гипотеза о наличии положительной автокорреляции. Средняя относительная ошибка аппроксимации = 1.683106%

С ростом цены цыплят на 1% спрос на них в среднем увеличивается на 1.737% - это противоречит экономическому смыслу. В модели не учитывается инфляция и рост среднедушевых доходов населения за рассматриваемые 20 лет. Модель экономически не интерпретируема.

2) Call:

lm(formula = Y\_log ~ X1\_log, data = df)

Residuals:

Min 1Q Median 3Q Max

-0.072152 -0.020452 0.003223 0.025825 0.059829

Coefficients:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Estimate** | **Std.Error** | **tvalue** | **Pr(>|t|)** |
| (Intercept) | 1.74599 | 0.12292 | 14.20 | 7.33e-11\*\*\* |
| X1\_log | 0.28492 | 0.01768 | 16.12 | 9.84e-12\*\*\* |

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03781 on 17 degrees of freedom

Multiple R-squared: 0.9386, Adjusted R-squared: 0.935

F-statistic: 259.8 on 1 and 17 DF, p-value: 9.837e-12

Durbin-Watson test

data: model

DW = 0.7529, p-value = 0.0002786

alternative hypothesis: true autocorrelation is greater than 0

Average relative error of approximation: 0.7533616 %

(Intercept)

5.731563

X1\_log

1.329654

Y\_hat = 5.731563 \* X11.329654

Все коэффициенты значимы. Оценка остаточной дисперсии ошибки = 0.03781. Исправленный коэффициент детерминации = 0.935. F-статистика = 259.8 больше критического значения, что означает значимость уравнения регрессии - все коэффициенты не равны нулю одновременно. Статистика Дарбина-Уотсона = 0.7529, (p-value = 0.0002786) - гипотеза об отсутствии автокорреляции остатков отвергается так как p-value = 0.0002786 меньше критического уровня 0.05 - принимается альтернативная гипотеза о наличии положительной автокорреляции. Средняя относительная ошибка аппроксимации = 0.7533616%.

При увеличении среднедушевого дохода на 1% потребление цыплят в среднем растёт на 1.33%, но в модели не учитывается динамика цены за рассматриваемые 20 лет.

3) Call:

lm(formula = Y\_log ~ X1\_log + X2\_log, data = df)

Residuals:

Min 1Q Median 3Q Max

-0.055142 -0.018277 -0.005682 0.022947 0.051872

Coefficients:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Estimate** | **Std.Error** | **tvalue** | **Pr(>|t|)** |
| (Intercept) | 2.01848 | 0.12665 | 15.937 | 3.07e-11\*\*\* |
| X1\_log | 0.41614 | 0.04157 | 10.011 | 2.70e-08\*\*\* |
| X2\_log | -0.30482 | 0.09094 | -3.352 | 0.00405\*\* |

---

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.02988 on 16 degrees of freedom

Multiple R-squared: 0.9639, Adjusted R-squared: 0.9594

F-statistic: 213.7 on 2 and 16 DF, p-value: 2.872e-12

Durbin-Watson test

data: model

DW = 1.5656, p-value = 0.08005

alternative hypothesis: true autocorrelation is greater than 0

Average relative error of approximation: 0.6019894 %

(Intercept)

7.526901

X1\_log

1.516101

X2\_log

0.7372586

Y\_hat = 7.526901 \* X11.516101 \* X20.7372586

Все коэффициенты значимы. Оценка остаточной дисперсии ошибки = 0.02988. Средняя относительная ошибка аппроксимации = 0.6019894%. Исправленный коэффициент детерминации = 0.9594. Статистика Дарбина-Уотсона = 0.7529, (p-value = 0.08005) - гипотеза о об отсутствии автокорреляции остатков не отвергается, так как p-value превышает пороговое значение 0.05.F-статистика = 213.7 больше критического значения, что означает значимость уравнения регрессии - все коэффициенты не равны нулю одновременно.

С ростом среднедушевого дохода на 1% при неизменной стоимости цыплят их потребление в среднем увеличивается на 1.516%. Увеличение стоимости цыплят на 1% при неизменном среднедушевом доходе приводит к увеличению потребления в среднем на 0.737%

4) Call:

lm(formula = Y\_log ~ X2\_log + X3\_log + X4\_log, data = df)

Residuals:

Min 1Q Median 3Q Max

-0.06833 -0.02295 0.01050 0.02525 0.04585

Coefficients:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Estimate** | **Std.Error** | **tvalue** | **Pr(>|t|)** |
| (Intercept) | 2.31519 | 0.22155 | 10.450 | 2.79e-08 \*\*\* |
| X2\_log | -0.48745 | 0.21100 | -2.310 | 0.0355 \* |
| X3\_log | 0.23742 | 0.15562 | 1.526 | 0.1479 |
| X4\_log | 0.46005 | 0.07527 | 6.112 | 1.99e-05 \*\*\* |

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.03772 on 15 degrees of freedom

Multiple R-squared: 0.9461, Adjusted R-squared: 0.9353

F-statistic: 87.74 on 3 and 15 DF, p-value: 9.73e-10

Durbin-Watson test

data: model

DW = 1.2348, p-value = 0.01086

alternative hypothesis: true autocorrelation is greater than 0

Average relative error of approximation: 0.7583258 %

(Intercept)

10.12682

X2\_log

0.6141879

X3\_log

1.267973

X4\_log

1.584158

Y\_hat = 10.12682 \* X20.6141879 \* X31.267973 \* X41.584158

Коэффициент X3\_log незначим, остальные коэффициенты значимы. Оценка остаточной дисперсии ошибки = 0.03772. Средняя относительная ошибка аппроксимации = 0.7583258%. Исправленный коэффициент детерминации = 0.9353. Статистика Дарбина-Уотсона = 1.2348, (p-value = 0.01086) - гипотеза о об отсутствии автокорреляции остатков отвергается, так как p-value = 0.01086 меньше критического уровня 0.05 - принимается альтернативная гипотеза о наличии положительной автокорреляции. F-статистика = 87.74 больше критического значения, что означает значимость уравнения регрессии - все коэффициенты не равны нулю одновременно.

При неизменной стоимости двух товарозаменителей увеличение на 1% стоимости цыплят приводит к увеличению их потребления в среднем на 0.614%, а увеличение стоимости свинины или говядины на 1% при неизменности цен на остальные входящие в модель продукты проводит к росту потребления цыплят в среднем на 1.267973 и 1.584158 соответственно.

Работоспособны модели функций потребления (модель 2), спроса-потребления (модель 3) и спроса с учетом цен на товарозаменители (модель 4), но не функции спроса (модель 1) - она экономически не интерпретируема.

Так как из исходных данных нельзя узнать влияние инфляционных процессов на среднедушевые доходы населения, то наиболее экономически содержательными являются модели спроса-потребления и спроса с учетом цен на товарозаменители. Тем не менее, хоть 3 модель и имеет наибольший коэффициент детерминации показывающих долю объясненной дисперсии зависимой переменной, в ней присутствует автокорреляция (на уровне значимости 0.08 что не катастрофически превышает критический уровень 0.05), а в 4 модели одна из переменных незначима. Эти модели не идеальны, но они лучшие из представленных.

# **Задание 2**

fit <- lm(Y~poly(X1,4), data=df)

coef(summary(fit))

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Estimate** | **Std. Error** | **t value** | **Pr(>|t|)** |
| (Intercept) | 41.794737 | 0.3080316 | 135.683293 | 3.065404e-23 |
| poly(X1, 4)1 | 25.337077 | 1.3426785 | 18.870547 | 2.361142e-11 |
| poly(X1, 4)2 | -3.589558 | 1.3426785 | -2.673431 | 1.817756e-02 |
| poly(X1, 4)3 | 1.493198 | 1.3426785 | 1.112104 | 2.848264e-01 |
| poly(X1, 4)4 | -3.922809 | 1.3426785 | -2.921629 | 1.115476e-02 |

income.lims <- range(df$X1)

income.grid <- seq(from=income.lims[1], to=income.lims[2])

preds <- predict(fit, newdata = list(X1 = income.grid),se = TRUE)

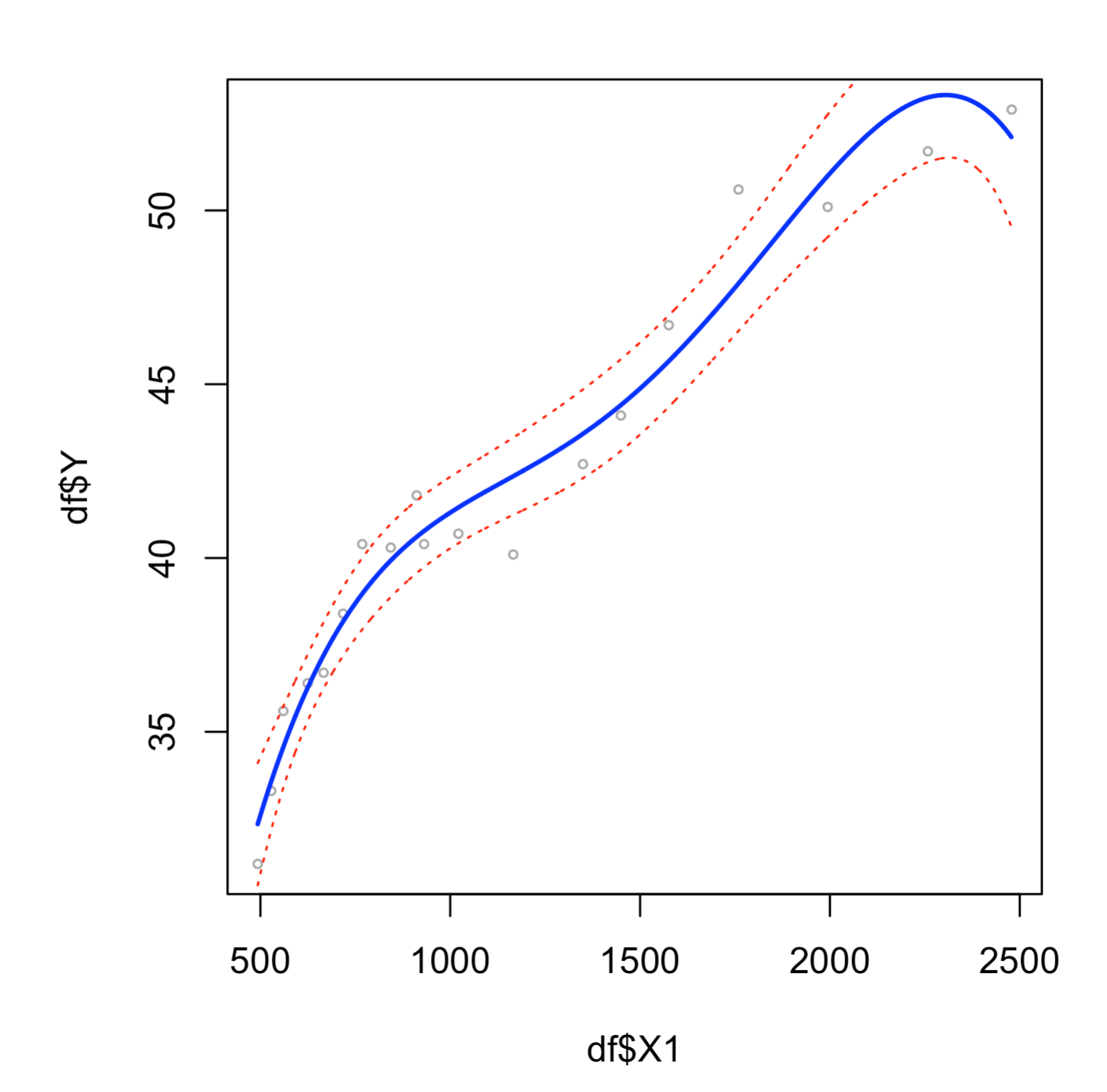
se.bands <- cbind(preds$fit + preds$se.fit \* 2, preds$fit - preds$se.fit\*2 )

plot(df$X1, df$Y, xlim = income.lims, cex=.5, col = 'darkgrey')

lines(income.grid, preds$fit, lwd = 2, col= 'blue')

matlines(income.grid, se.bands, lwd = 1, col = 'red', lty=3)

Построим график оцененная полиномиальная регрессия – она обозначается непрерывной синей линией. Точками отмечены исходные данные. Красная пунктирная линия – 95% доверительный интервал.



fit.1 <- lm(Y ~ poly(X1, 1), data = df)

fit.2 <- lm(Y ~ poly(X1, 2), data = df)

fit.3 <- lm(Y ~ poly(X1, 3), data = df)

fit.4 <- lm(Y ~ poly(X1, 4), data = df)

fit.5 <- lm(Y ~ poly(X1, 5), data = df)

fit.6 <- lm(Y ~ poly(X1, 6), data = df)

Сравним разные модели с помощью ANOVA теста

anova(fit.1,fit.2,fit.3,fit.4,fit.5,fit.6)

> coef(summary(fit))

Analysis of Variance Table

Model 1: Y ~ poly(X1, 1)

Model 2: Y ~ poly(X1, 2)

Model 3: Y ~ poly(X1, 3)

Model 4: Y ~ poly(X1, 4)

Model 5: Y ~ poly(X1, 5)

Model 6: Y ~ poly(X1, 6)

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | **Res.Df** | **RSS** | **Df** | **Sum of Sq** | **F** | **Pr(>F)** |
| 1 | 17 | 55.742 |  |  |  |  |
| 2 | 16 | 42.857 | 1 | 12.8849 | 14.2546 | 0.002645\*\* |
| 3 | 15 | 40.627 | 1 | 2.2296 | 2.4667 | 0.142266 |
| 4 | 14 | 25.239 | 1 | 15.3884 | 17.0243 | 0.001405\*\* |
| 5 | 13 | 11.355 | 1 | 13.8837 | 15.3596 | 0.002038\*\* |
| 6 | 12 | 10.847 | 1 | 0.5083 | 0.5624 | 0.467754 |

Signif. codes: 0 ‘\*\*\*’ 0.001 ‘\*\*’ 0.01 ‘\*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Наиболее хорошая их всех полиномиальных моделей - модель пятой степень, так как она имеет наибольшую степень полинома из значимых моделей по результатам дисперсионного анализа.

# **Задание 3**

a)

d1 <- density(c(3,5,7,9,9,10,11,13,17,19), kernel="epanechnikov")

plot(d1)

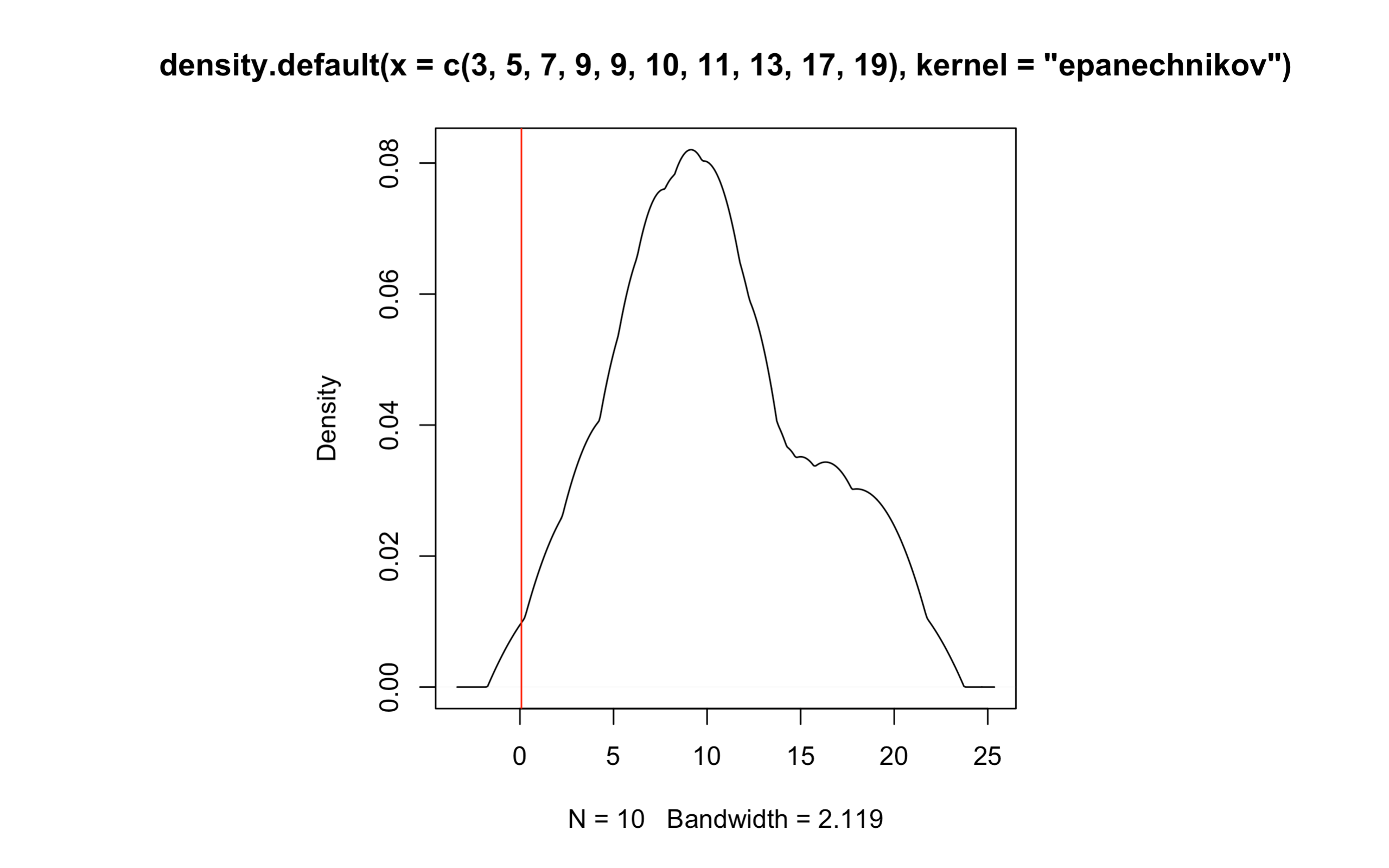
dens\_5 <- approx(d1$x, d1$y, xout=5)$y

abline(v = dens\_10, col = "red")

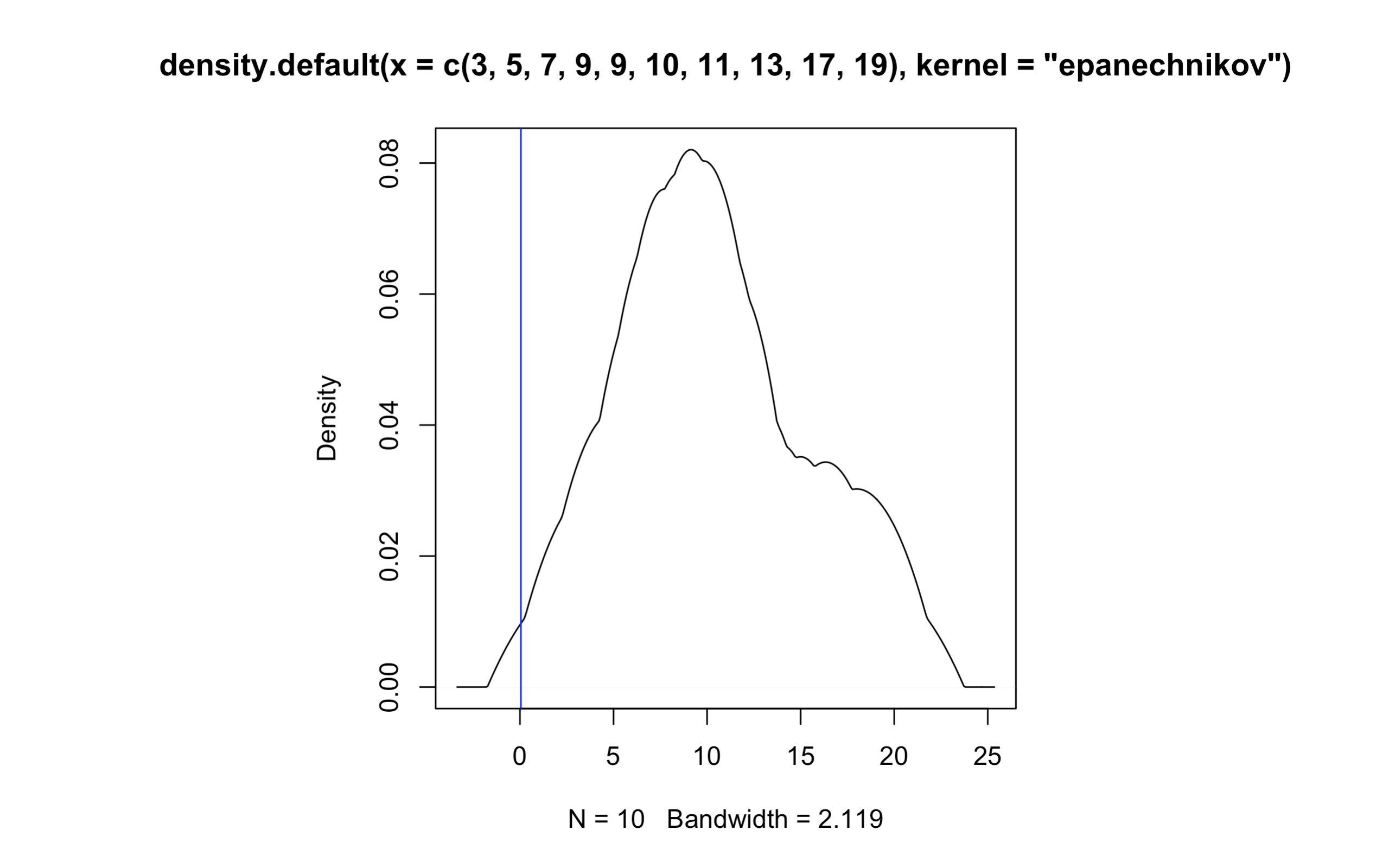
plot(d1)

dens\_10 <- approx(d1$x, d1$y, xout=10)$y

abline(v = dens\_5, col = "blue")



f (5) = 0.05098513



f (10) = 0.0801827

б)

d2 <- density(c(3,5,7,9,9,10,11,13,17,19), kernel="gaussian")

plot(d2)

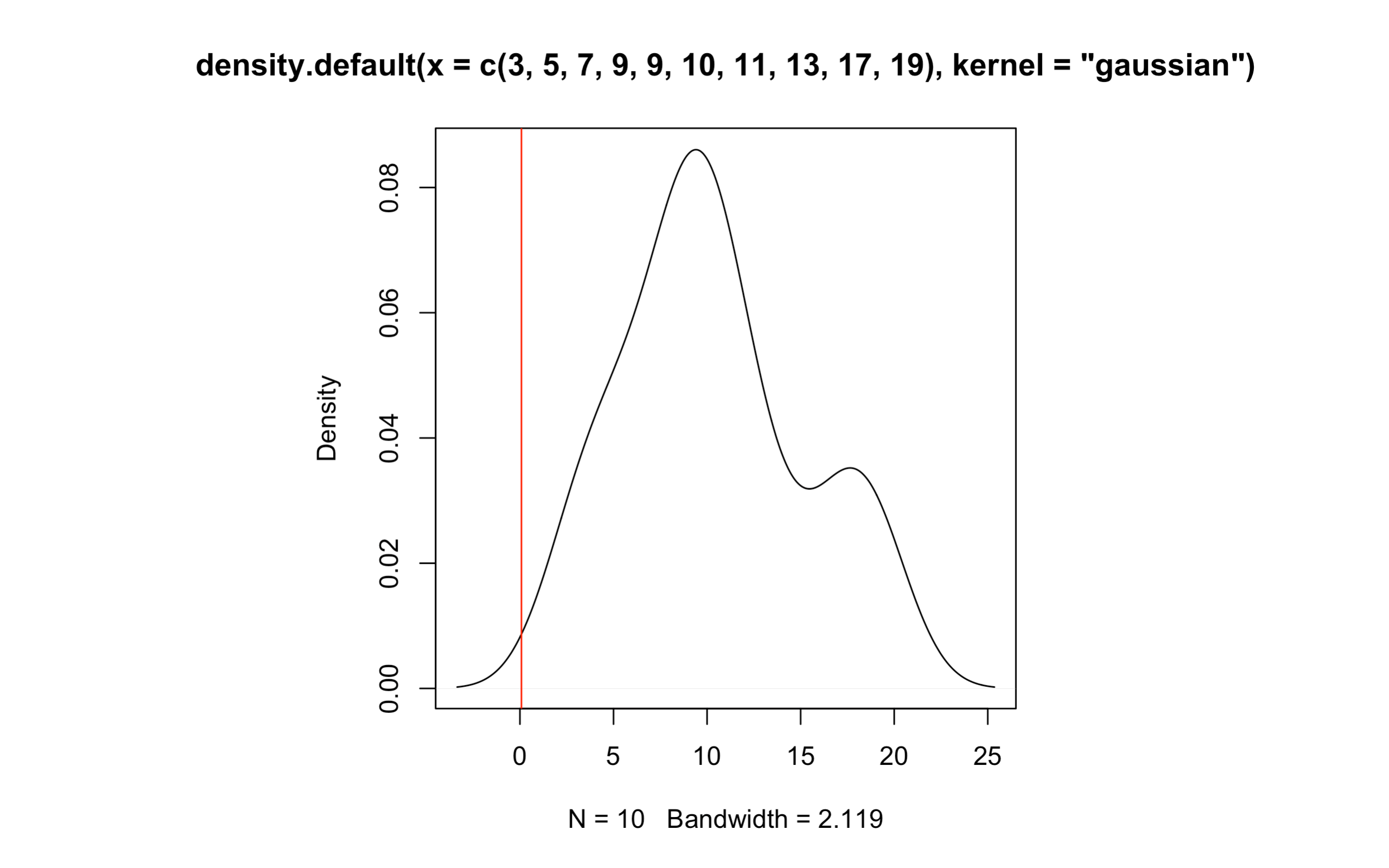
dens\_5 <- approx(d2$x, d2$y, xout=5)$y

abline(v = dens\_10, col = "red")

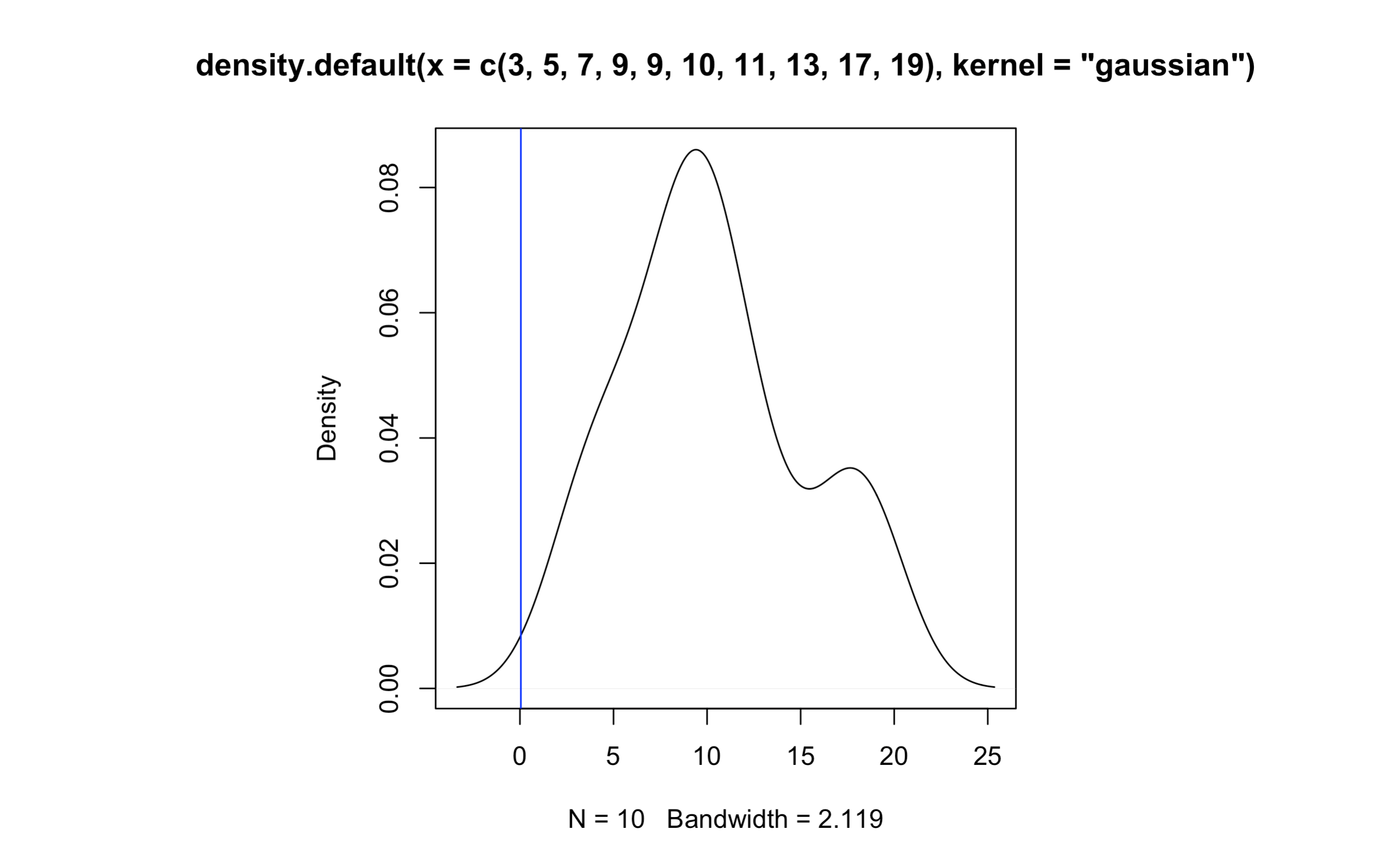
plot(d2)

dens\_10 <- approx(d2$x, d2$y, xout=10)$y

abline(v = dens\_5, col = "blue")



f (5) = 0.05085784



f(10) = 0.08453672

График, построенный с использованием ядра Гаусса, получился более сглаженным при одинаковой ширине окна в сравнении с графиков построенным с помощью ядра Епанечникова. При это значения оценок функции плотности распределения практически не отличаются в обоих методах.

# **Задание 4**

1)

install.packages('ISLR')

library('ISLR')

data(Smarket)

tendency <- as.numeric(Smarket[,9]) - 1

smarket <- Smarket[1:8]

cor(smarket, method = "pearson")

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | **Year** | **Lag1** | **Lag2** | **Lag3** |
| Year | 1.00000000 | 0.029699649 | 0.030596422 | 0.033194581 |
| Lag1 | 0.02969965 | 1.000000000 | -0.026294328 | -0.010803402 |
| Lag2 | 0.03059642 | -0.026294328 | 1.000000000 | -0.025896670 |
| Lag3 | 0.03319458 | -0.010803402 | -0.025896670 | 1.000000000 |
| Lag4 | 0.03568872 | -0.002985911 | -0.010853533 | -0.024051036 |
| Lag5 | 0.02978799 | -0.005674606 | -0.003557949 | -0.018808338 |
| Volume | 0.53900647 | 0.040909908 | -0.043383215 | -0.041823686 |
| Today | 0.03009523 | -0.026155045 | -0.010250033 | -0.002447647 |
|  | **Lag4** | **Lag5** | **Volume** | **Today** |
| Year | 0.035688718 | 0.029787995 | 0.53900647 | 0.030095229 |
| Lag1 | -0.002985911 | -0.005674606 | 0.04090991 | -0.026155045 |
| Lag2 | -0.010853533 | -0.003557949 | -0.04338321 | -0.010250033 |
| Lag3 | -0.024051036 | -0.018808338 | -0.04182369 | -0.002447647 |
| Lag4 | 1.000000000 | -0.027083641 | -0.04841425 | -0.006899527 |
| Lag5 | -0.027083641 | 1.000000000 | -0.02200231 | -0.034860083 |
| Volume | -0.048414246 | -0.022002315 | 1.00000000 | 0.014591823 |
| Today | -0.006899527 | -0.034860083 | 0.01459182 | 1.000000000 |

Большинство признаков имеют слабую очень корреляцию между собой – менее 0.05 по модулю. Умеренную корреляционную связь умеют только переменные Year и Volume, равную 0.53900647.

2)

model\_bin<-glm(Direction~Lag1+Lag2+Lag3+Lag4+Lag5+Volume,

family = binomial(link = 'logit'),

data=Smarket)

summary(model\_bin)

Deviance Residuals:

Min 1Q Median 3Q Max

-1.446 -1.203 1.065 1.145 1.326

Coefficients:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Variable** | **Estimate** | **Std. Error** | **Z value** | **Pr(>|z|)** |
| (Intercept) | -0.126000 | 0.240736 | -0.523 | 0.601 |
| Lag1 | -0.073074 | 0.050167 | -1.457 | 0.145 |
| Lag2 | -0.042301 | 0.050086 | -0.845 | 0.398 |
| Lag3 | 0.011085 | 0.049939 | 0.222 | 0.824 |
| Lag4 | 0.009359 | 0.049974 | 0.187 | 0.851 |
| Lag5 | 0.010313 | 0.049511 | 0.208 | 0.835 |
| Volume | 0.135441 | 0.158360 | 0.855 | 0.392 |

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1731.2 on 1249 degrees of freedom

Residual deviance: 1727.6 on 1243 degrees of freedom

AIC: 1741.6

Number of Fisher Scoring iterations: 3

, где z – уравнение регрессии вида b1 • X1 + b2 • Х2 + ... + bn • Xn + a

Чем больше значение коэффициента, тем больший вклад он вносит в вероятность благоприятного исхода события.

(Для каждой переменной значения рассчитаны при нулевом влиянии других факторов.)

При нулевом значении всех переменных вероятность роста рынка была бы в среднем на уровне 1 / (1 + exp(-(-0.073074))) = 0.46854160844368303

Увеличение Lag1 (доходности в предыдущий день) на единицу в среднем увеличивает вероятность роста акций на 1 / (1 + exp(-(-0.073074))) = 0.5182603751460386

Увеличение Lag2(доходности в день перед предыдущим) на единицу в среднем увеличивает вероятность роста акций на 1 / (1 + exp(-(-0.042301)))= 0.48942632664152613,

Увеличение Lag3 на единицу в среднем увеличивает вероятность роста акций на 1 / (1 + exp(-(0.011085))) = 0.5027712216233895,

Увеличение Lag4 на единицу в среднем увеличивает вероятность роста акций на 1 / (1 + exp(-(0.009359))) = 0.5023397329217526,

Увеличение Lag5 на единицу в среднем увеличивает вероятность роста акций на 1 / (1 + exp(-(0.010313))) = 0.5025782271487902,

Увеличение Volume (объемов торгов в предыдущий день) на единицу в среднем увеличивает вероятность роста акций на 1 / (1 + exp(-(0.135441))) = 0.5338085829954674

3)

fitted<-predict(model\_bin,type = 'response')

fitted - содержит информацию о росте или падении рынка каждый день

fitted.res<-ifelse(fitted>0.5,1,0)

as.numeric(summary(fitted.res == tendency)[["TRUE"]])/length(tendency)

Доля верных прогнозов - 0.5216. Модель имеет достаточно слабую предсказательную силу – смогла предсказать чуть более половины результатов.

# **Задание 5**

install.packages("psych")

library("psych")

hdi.initial <- read.table("Desktop/hdi.txt", header=TRUE)

str(hdi.initial)

describe(hdi.initial)

Переменные имеют разный разброс - при применении метода главных комнонент их нужно стандартизировать - мы ищем переменные, наиболее информативно описывающие разницу hdi между странами

hdi.num <- lapply(hdi.initial, function(x) as.numeric(gsub(",", ".", gsub("\\.", "", as.character(x)))))

hdi <- as.data.frame(hdi.num[3:5])

cor(hdi)

|  |  |  |  |
| --- | --- | --- | --- |
| **Variable name** | **sub1** | **sub2** | **sub3** |
| sub1 | 1.0000000 | 0.5918565 | 0.6425139 |
| sub2 | 0.5918565 | 1.0000000 | 0.9022894 |
| sub3 | 0.6425139 | 0.9022894 | 1.0000000 |

Присутствуют довольно сильно коррелирующие переменные, например

R(sub2,sub3)=0.9022894, а остальные коэффициенты около 0.6: R(sub1,sub2) = 0.5918565, R(sub1,sub3) = 0.6425139.

hdi.pca <- prcomp(hdi, scale=TRUE)

hdi.pca

pca1 <- hdi.pca$x[,1]

Первый столбик новых индексов развития человеческого потенциала

v1 <- hdi.pca$rotation[,1]

Веса первой компоненты - веса, с которыми старые переменные входят в новую синтетическую переменную

v1

|  |  |  |
| --- | --- | --- |
| **sub1** | **sub2** | **sub3** |
| -0.5201059 | -0.5982757 | -0.6095540 |

Наиболее важные переменная в первой компоненте - sub2 и sub3 - их вес -0.5982757 и -0.6095540 соответственно. Немногим меньше вес sub1 равный -0.5201059.

summary(hdi.pca)

Importance of components:

|  |  |  |  |
| --- | --- | --- | --- |
|  | **PC1** | **PC2** | **PC3** |
| Standard deviation | 1.5601 | 0.6864 | 0.30831 |
| Proportion of Variance | 0.8113 | 0.1570 | 0.03169 |
| Cumulative Proportion | 0.8113 | 0.9683 | 1.00000 |

Первая главная компонента объясняет 0.8113 (81.13%) совокупной дисперсии исходного набора данных, а первая и вторая в сумме объясняют 96.83%. Три компоненты в сумме объясняют 100%.

cor(hdi.num$hdi, pca1)

Результат сравнения полученного результата и фактического значения индекса - очень высокая - практически функциональная корреляция (с коэффициентом корреляции равном -0.98) - между индексом развития человеческого потенциала и первой главной компонентой.

plot(hdi.pca)

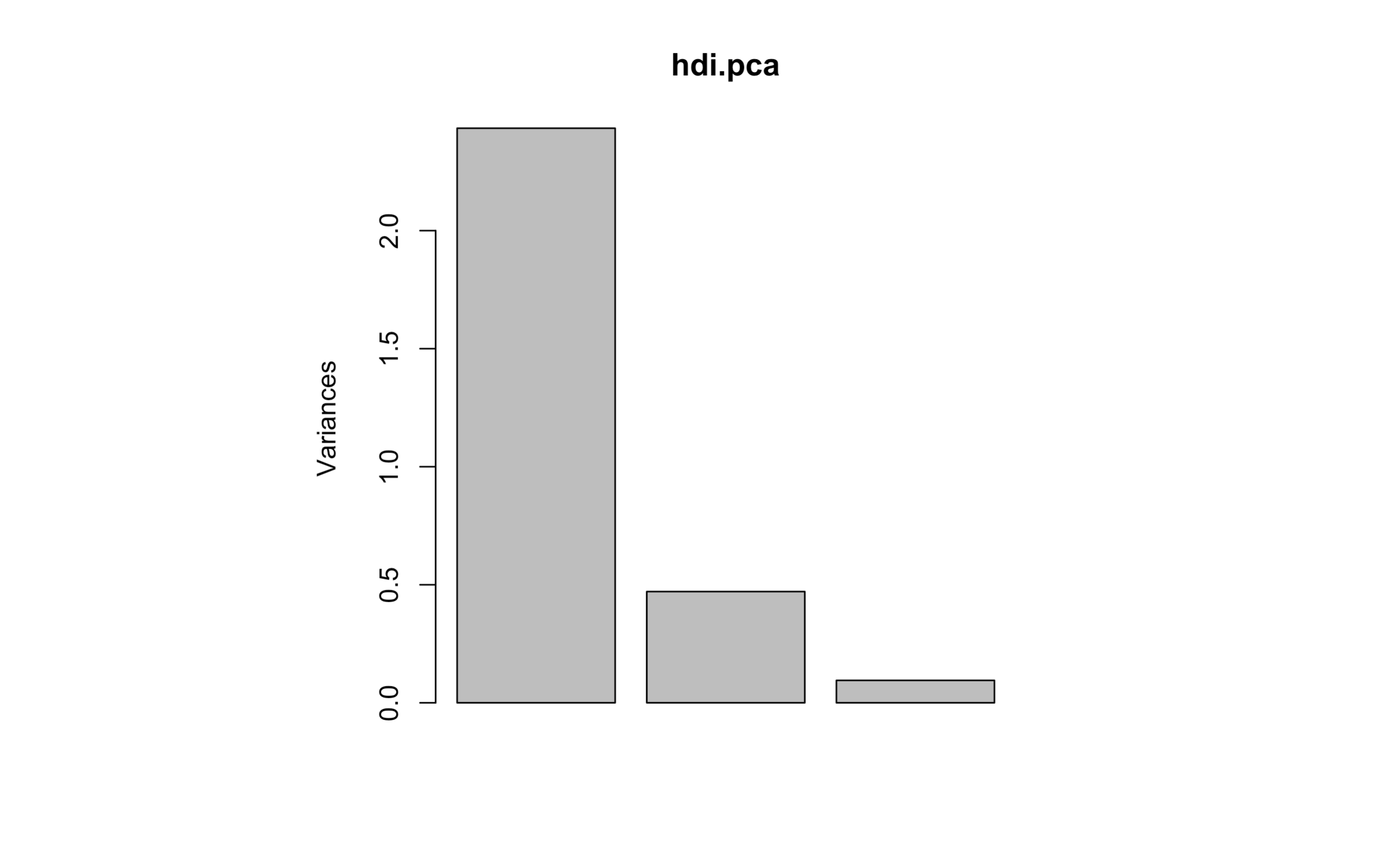
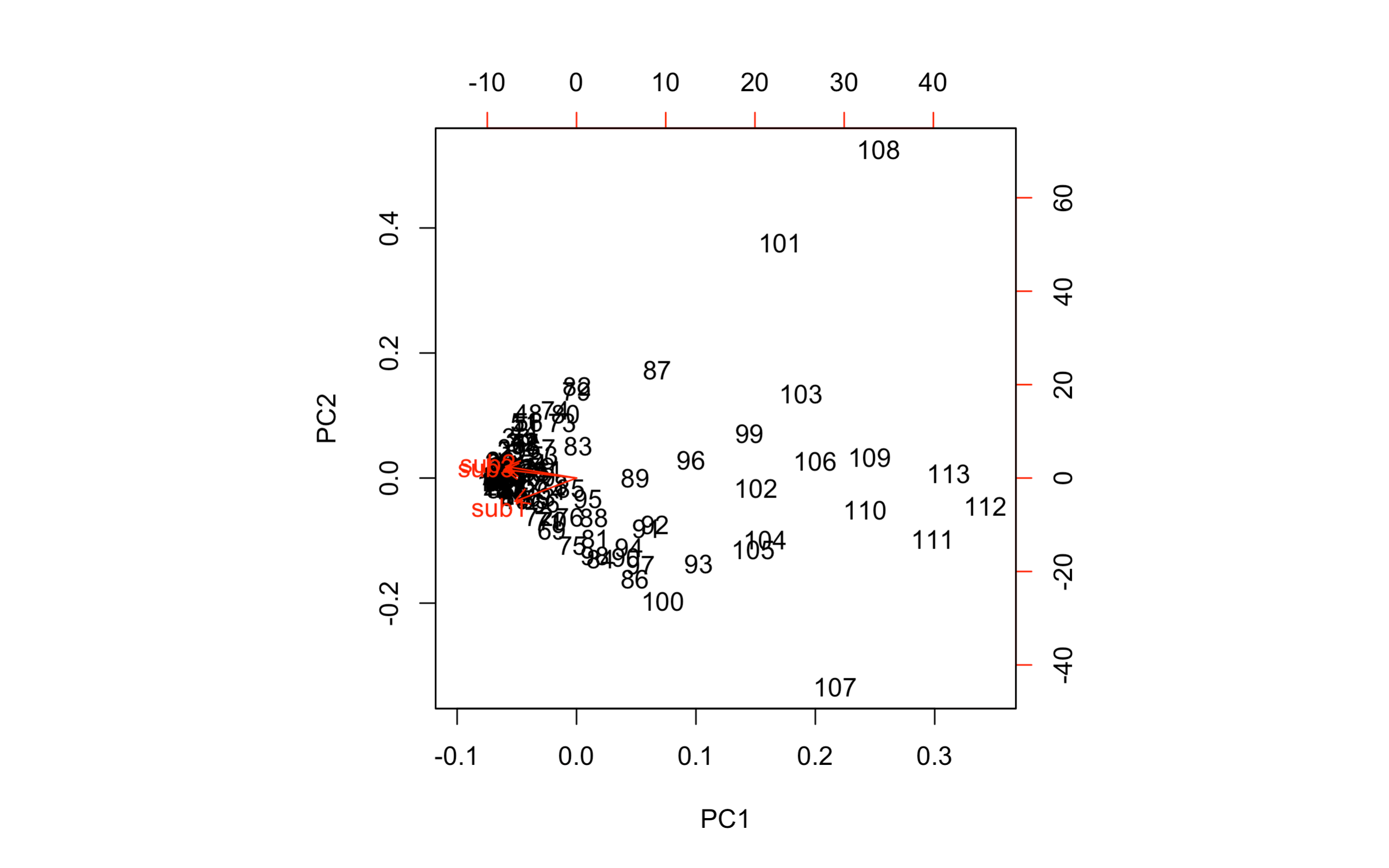


График показывает долю главной компоненты, объясняемая разными компонентами

biplot(hdi.pca, xlim=c(-0.1, 0.35))



На графике по оси абсцисс отложена первая главная компонента, а по оси ординат - вторая главная компонента. Можно увидеть наблюдения, которые существенно отличаются от всех остальных - например 108 и 107.Первая главная компонента включает в себя с большим весом все три переменные. Также можно наблюдать корреляцию между переменными - видна сильна корреляция между sub2 и sub3.