**Практика 3. Домашнее задание.**

**Проверка адекватности моделей для прогнозирования значений временных рядов.**

**ФИО**: Козицын Андрей Дмитриевич

**Задание**: Подобрать функцию, описывающую зависимость между средней заработной платой в России и средним чеком в «фастфуде».

**Описание данных:**

Данные взяты с ресурса <https://www.sberbank.com/ru/analytics/opendata>, выбран период с 01-01-2015 до 01-01-2019.

На графиках ниже представлено численное распределение величин.

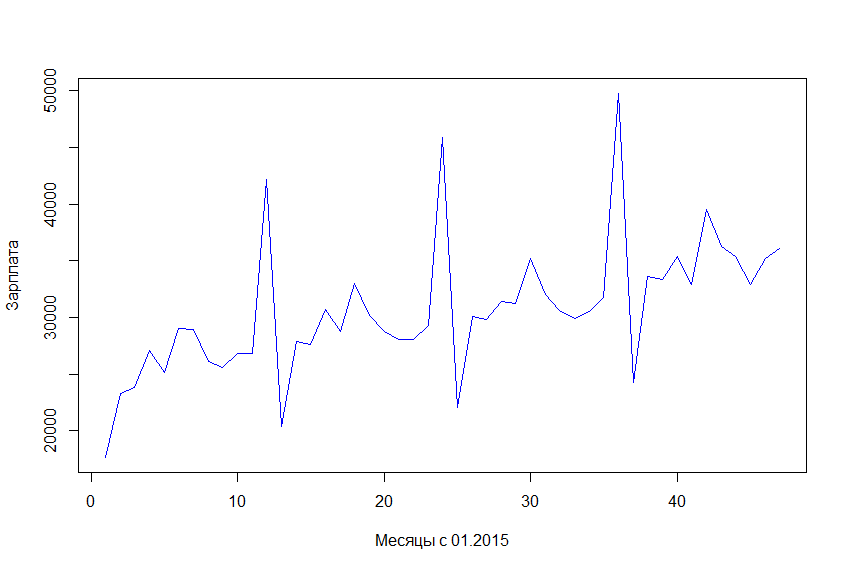


Рис.1. Средняя зарплата по месяцам с 01.2015 по 12.2018.

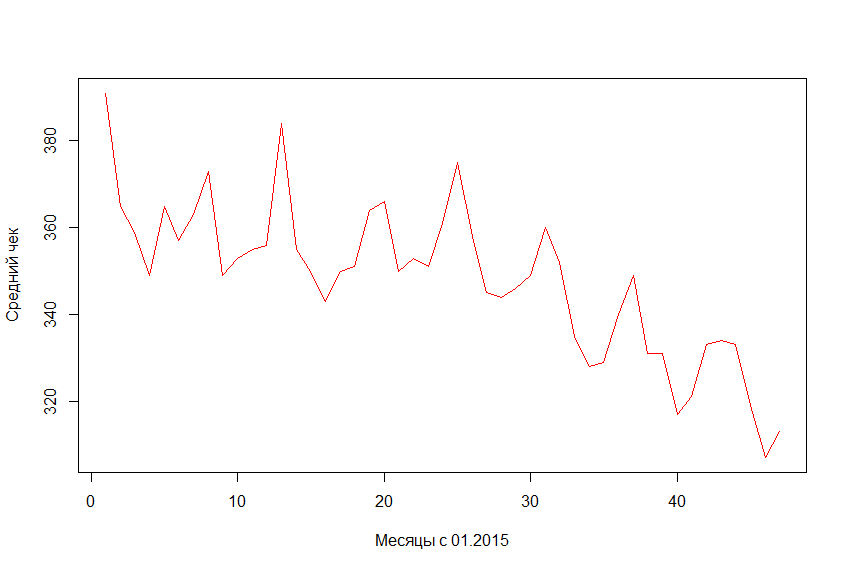


Рис.2. Средний чек за фастфуд по месяцам с 01.2015 по 12.2018.

**Решение:**

**1. МНК.**

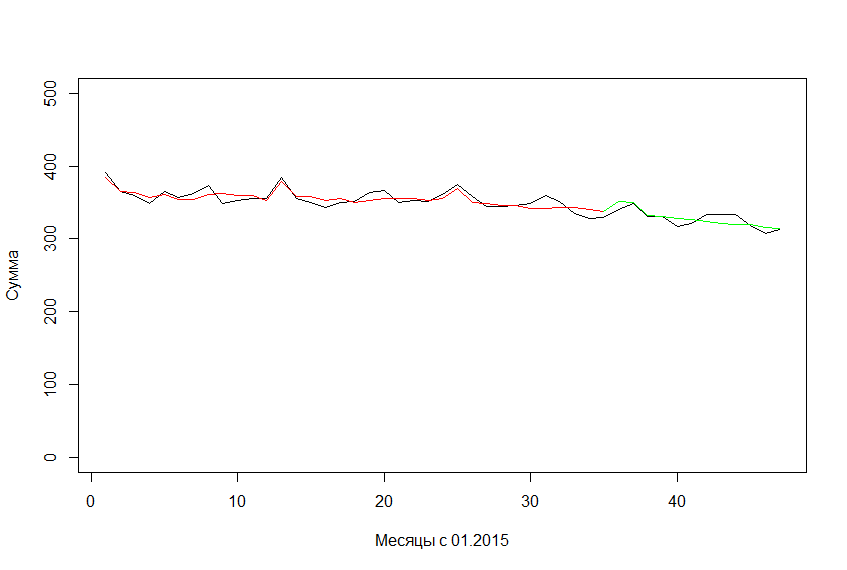


Рис.3. МНК.

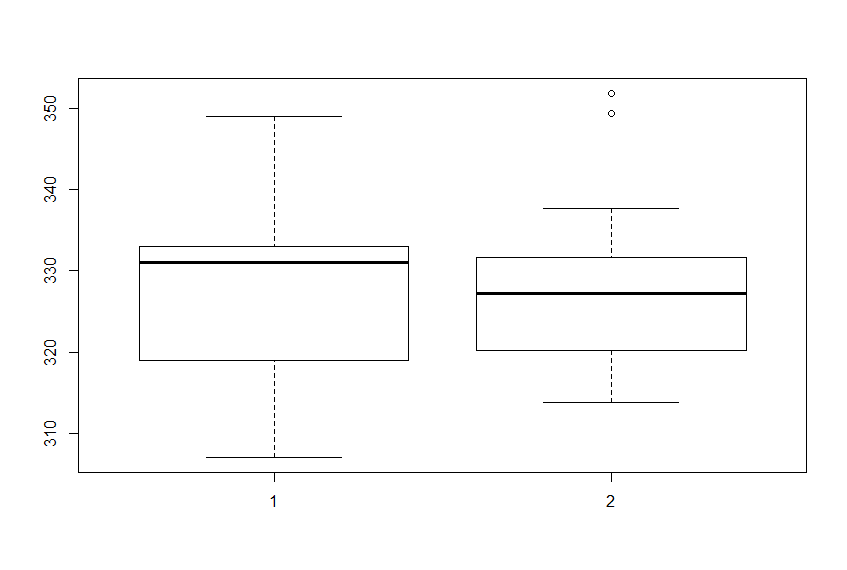


Рис.4. box-диаграмма.

Проверка критерием хи-квадрат выдала результат **p-value = 1**, а т.к. результаты построения отличаются, то результат моделирования нас не устраивает. Скорее всего ошибка.

Проверка критерием Фишера выдала результат **p-value = 0.9547**, а т.к. p>0,05, то результат моделирования значимый.

Проверка критерием Стьюдента выдала результат **p-value = 0.6215**, а т.к. p>0,05, то результат моделирования значимый.

**2. KNN.**

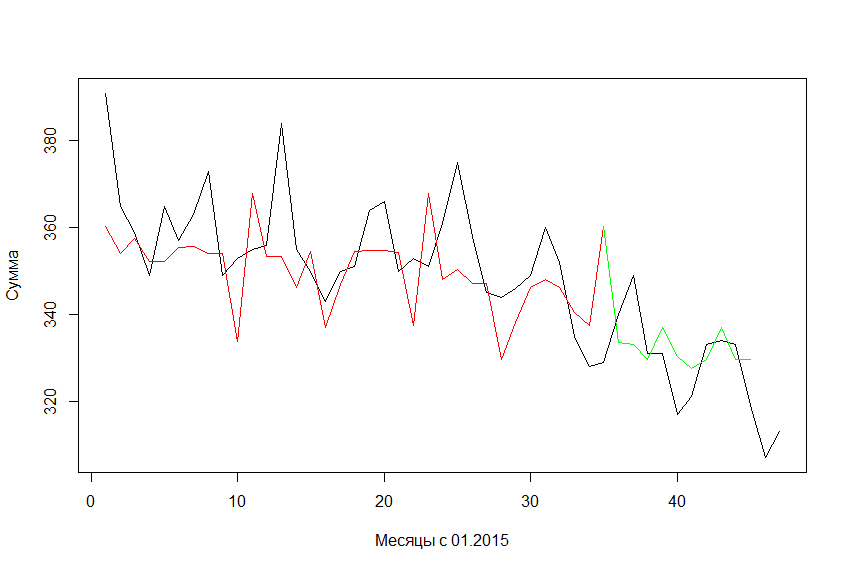


Рис.5. KNN.

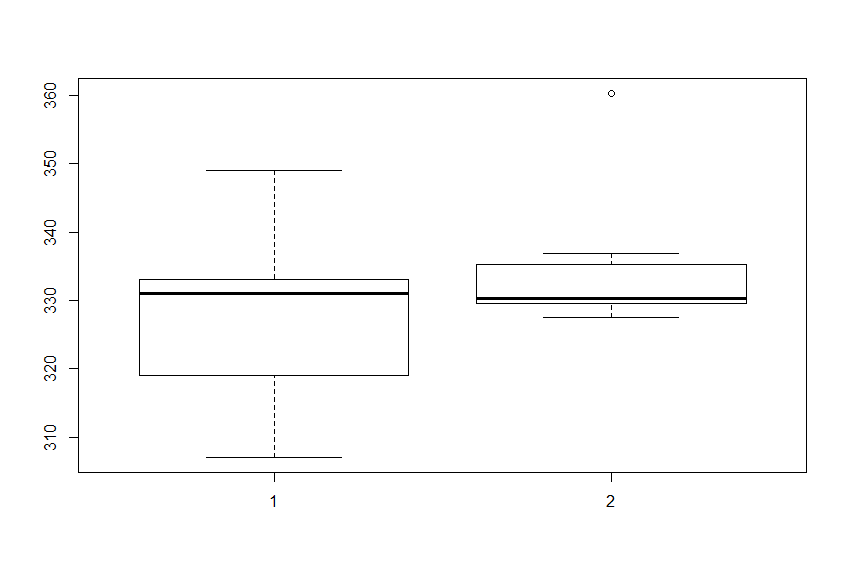


Рис.6. box-диаграмма.

Проверка критерием хи-квадрат выдала результат **p-value = 0.994**, результат моделирования близок к фактическим данным.

Проверка критерием Фишера выдала результат **p-value = 0.4752**, а т.к. p>0,05, то результат моделирования значимый.

Проверка критерием Стьюдента выдала результат **p-value = 0.3541**, а т.к. p>0,05, то результат моделирования значимый.

**3. SVM.**

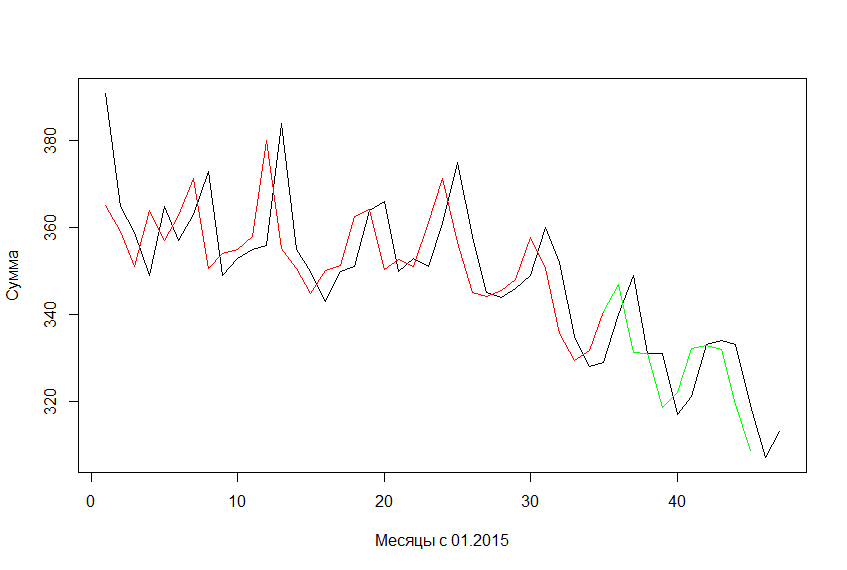


Рис.7. SVM.

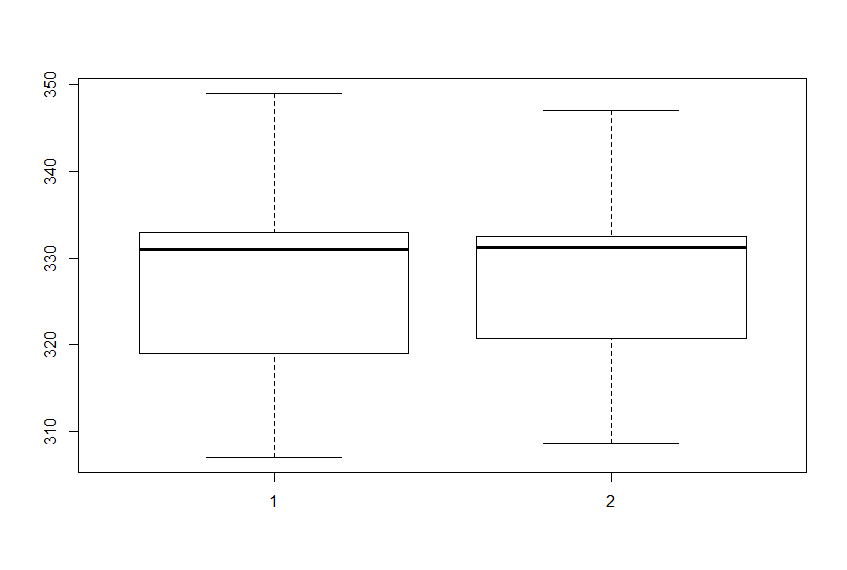


Рис. 8. Box-диаграмма.

Проверка критерием хи-квадрат выдала результат **p-value = 1**, результат моделирования близок к фактическим данным.

Проверка критерием Фишера выдала результат **p-value = 0.8416**, а т.к. p>0,05, то результат моделирования значимый.

Проверка критерием Стьюдента выдала результат **p-value = 0.5359**, а т.к. p>0,05, то результат моделирования значимый.

**4. Метод вайвлет анализа.**

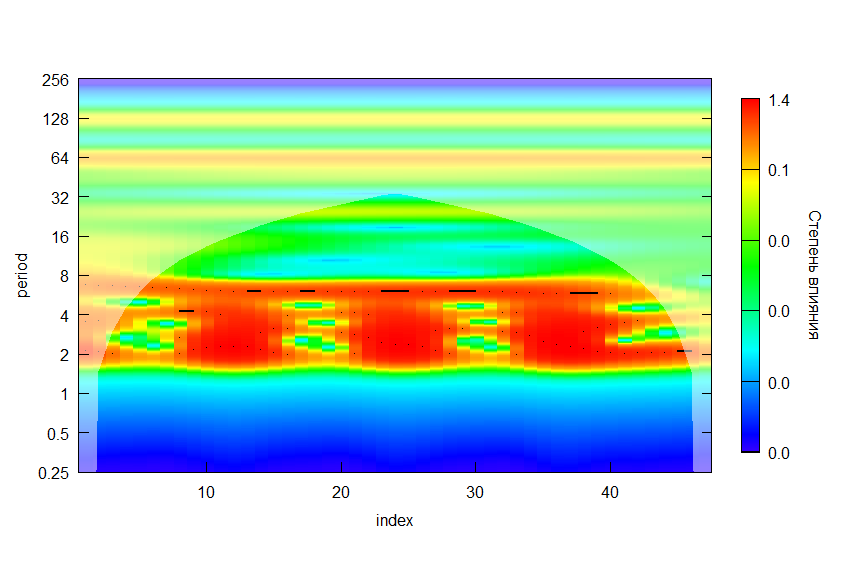


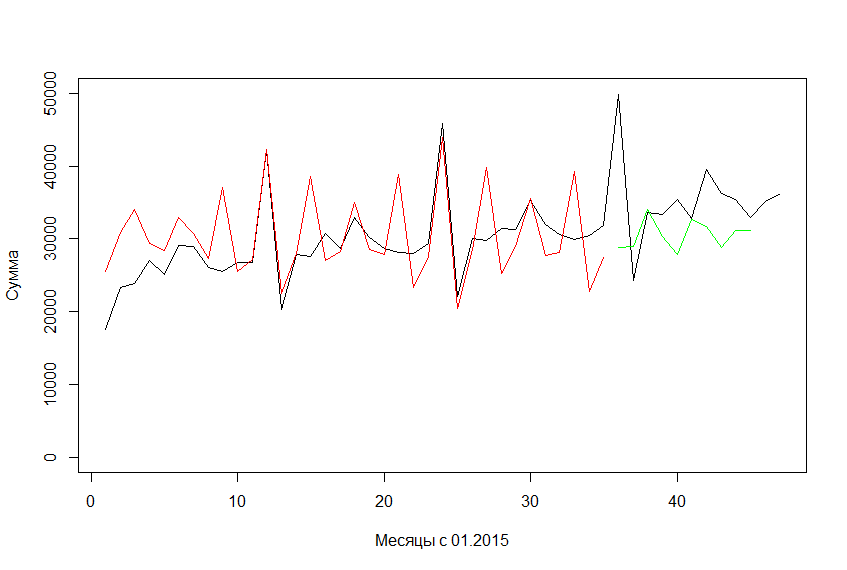
Рис. 9. Степень влияния гармоник.

Рис. 10. Wavelet.

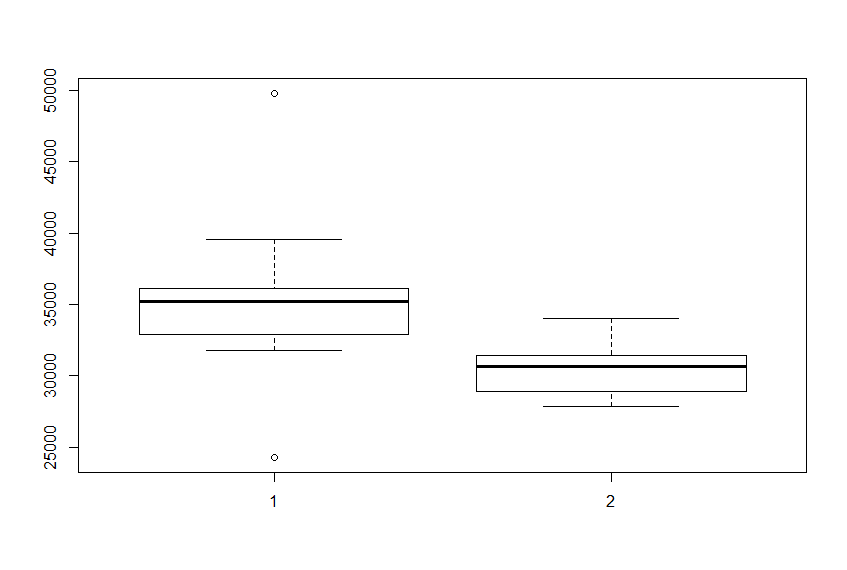


Рис. 11. Box-диаграмма.

Проверка критерием хи-квадрат выдала результат **p-value = 2.2e-16**, а т.к. p<0,05, то результат моделирования не интересен.

Проверка критерием Фишера выдала результат **p-value = 0.0005274**, а т.к. p<0,05, то результат моделирования не интересен.

Проверка критерием Стьюдента выдала результат **p-value = 0.02211**, а т.к. p<0,05, то результат моделирования не интересен.

**5. Фрактальный метод.**

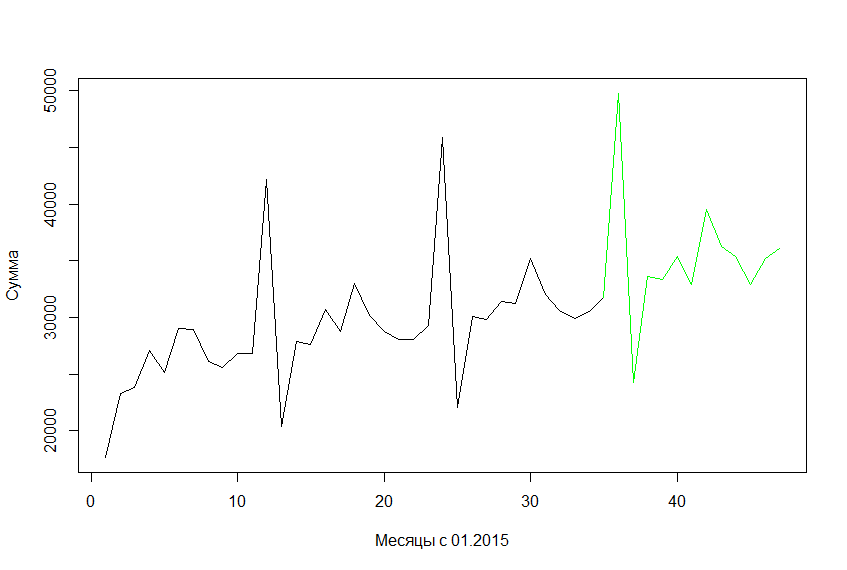


Рис. 12. Фрактальный метод.

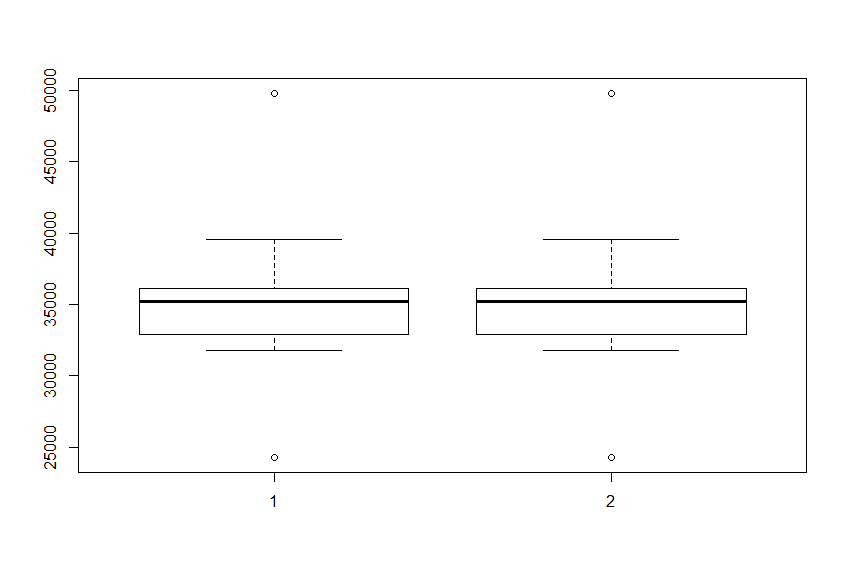


Рис. 13. Box-диаграмма.

Проверка критерием хи-квадрат выдала результат **p-value = 1**, результат моделирования соответствует фактическим данным.

Проверка критерием Фишера выдала результат **p-value = 0.8416**, результат моделирования соответствует фактическим данным.

Проверка критерием Стьюдента выдала результат **p-value = NA**. Данным методом невозможно определить результат моделирования.

**6. Распределение, построенное методом авторегрессии.**

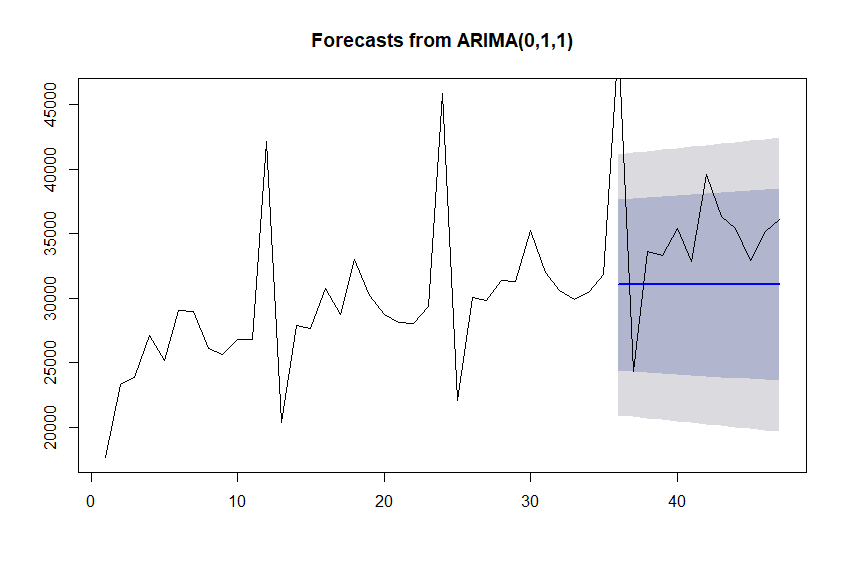


Рис. 14. Метод авторегрессии.

**Текст программы:**

*#путь*

*setwd("H:/R/R/Labs/Lab3")*

*#читаем файл*

*myData<- read.csv("fastfood.csv", header = T, sep = ";")*

*x=1:47*

*View(myData)*

*#выделяем набор данных, для которого будем строить регрессионную модель*

*e <- data.frame(y=myData$wage[1:47], x=x[1:47], z=myData$fastfood[1:47])*

*e2 <- data.frame(y=myData$wage[1:47], x=x[1:47], z=myData$fastfood[1:47], z2=myData$fastfood[1:47])*

*y <- myData$wage[1:47]*

*x2 <- 2:47*

*y2 <- myData$wage[2:47]*

*z2 <- myData$fastfood[1:46]*

*#чертим 2 графика*

*plot(x[1:47], myData$wage[1:47], type = "l", col="blue", xlab="Месяцы с 01.2015", ylab = "Сумма", ylim=c(0,500))*

*lines(x[1:47], myData$fastfood[1:47], type = "l", col="black")*

*#Пересчитываем значения Z-функции с добавленной второй температурой*

*res <- nls(z~a+b\*x+c\*x^2+f\*y+g\*y^2, data=e, start=list(a=0.1, b=0.1, c=0.1, f=0.1, g=0.1))*

*coef\_func <- coef(res)*

*zfunc <- coef\_func[1]+ coef\_func[2]\*x + coef\_func[3]\*x^2 + coef\_func[4]\*y + coef\_func[5]\*y^2*

*lines(x[1:35], zfunc[1:35], type = "l", col="red")*

*lines(x[35:47], zfunc[35:47], type = "l", col="green")*

*#КРИТЕРИИ*

*#тестовая box-диаграмма*

*boxplot(myData$fastfood[35:47], zfunc[35:47])*

*#Критерий ХИ-квадрат*

*test0 <- matrix(c(myData$fastfood[35:47], zfunc[35:47]), nrow = 2, byrow = TRUE)*

*chisq.test(test0)*

*#Критерий Фишера*

*var.test(myData$fastfood[35:47], zfunc[35:47])*

*#Критерий Стьюдента*

*t.test(myData$fastfood[35:47], zfunc[35:47], paired = TRUE)*

*#---------------------------------------------------------------*

*#-- МЕТОД KNN---------------------------------------------------*

*#---------------------------------------------------------------*

*library(caret) #библиотека позволяет использовать методы машинного обучения*

*mod1<-train (z~x+y+z2, data=e2,method = "knn")*

*test<-data.frame(y=y2[2:47],x=x2[2:47], z2=z2[2:47])*

*res1<-predict(mod1, test)*

*#рисуем график*

*plot(x[1:47], myData$fastfood[1:47], type = "l", xlab="Месяцы с 01.2015", ylab = "Сумма")*

*lines(x[1:35], res1[1:35], type = "l", col="red")*

*lines(x[35:47], res1[35:47], type = "l", col="green")*

*#КРИТЕРИИ*

*#тестовая box-диаграмма*

*boxplot(myData$fastfood[35:47], res1[35:47])*

*#Критерий ХИ-квадрат*

*test1 <- matrix(c(myData$fastfood[35:45], res1[35:45]), nrow = 2, byrow = TRUE)*

*chisq.test(test1)*

*#Критерий Фишера*

*var.test(myData$fastfood[35:47], res1[35:47])*

*#Критерий Стьюдента*

*t.test(myData$fastfood[35:47], res1[35:47], paired = TRUE)*

*#-- МЕТОД SVM*

*library(caret) #подключаем библиотеку, позволяющую использовать методы машинного обучения*

*mod2<-train (z~x+y+z2, data=e2, method = "svmPoly") #можно использовать так же линейный - svmLinear; или svmPoly*

*res2<-predict(mod2, test)*

*#Рисуем график*

*plot(x[1:47], myData$fastfood[1:47], type = "l", xlab="Месяцы с 01.2015", ylab = "Сумма")*

*lines(x[1:35], res2[1:35], type = "l", col="red")*

*lines(x[35:47], res2[35:47], type = "l", col="green")*

*#КРИТЕРИИ*

*#тестовая box-диаграмма*

*boxplot(myData$fastfood[35:47], res2[35:47])*

*#Критерий ХИ-квадрат*

*test2 <- matrix(c(myData$fastfood[35:47], res2[35:45]), nrow = 2, byrow = TRUE)*

*chisq.test(test2)*

*#Критерий Фишера*

*var.test(myData$fastfood[35:47], res2[35:47])*

*#Критерий Стьюдента*

*t.test(myData$fastfood[35:47], res2[35:47], paired = TRUE)*

*#-- МЕТОД ВАЙЛЕТ АНАЛИЗА*

*library(forecast)*

*library(WaveletComp)*

*library(biwavelet)*

*mod4<- analyze.wavelet(e2, my.series = 1,*

*loess.span = 0,*

*dt = 1, dj = 1/298,*

*lowerPeriod = 1/4,*

*upperPeriod = 256,*

*make.pval = F,*

*n.sim = 10)*

*wt.image(mod4, n.levels = 337,*

*legend.params = list(lab="Степень влияния"))*

*res4\_1<- reconstruct(mod4, sel.period = 3, show.legend = F)*

*res4\_2<- reconstruct(mod4, sel.period = 6, show.legend = F)*

*res4\_3<- reconstruct(mod4, sel.period = 16, show.legend = F)*

*res4\_4<- reconstruct(mod4, sel.period = 64, show.legend = F)*

*res4\_5<- reconstruct(mod4, sel.period = 256, show.legend = F)*

*plot(x[1:47], myData$wage[1:47], type = "l", xlab="Месяцы с 01.2015", ylab = "Сумма", ylim=c(0,50000))*

*lines(x[1:35], res4\_1$series$y.r[1:35], type = "l", col="red")*

*#прогноз по каждой гармонике отдельно*

*mod4\_rec4 <- auto.arima(res4\_1$series$y.r)*

*res4\_1\_1 <- forecast(mod4\_rec4, h=12)*

*lines(x[36:45], res4\_1\_1$mean[1:10], type = "l", col="green")*

*#КРИТЕРИИ*

*#тестовая box-диаграмма*

*boxplot(myData$wage[35:47], res4\_1\_1$mean[1:12])*

*#Критерий ХИ-квадрат*

*test1 <- matrix(c(myData$wage[35:47], res4\_1\_1$mean[1:12]), nrow = 2, byrow = TRUE)*

*chisq.test(test1)*

*#Критерий Фишера*

*var.test(myData$wage[35:47], res4\_1\_1$mean[1:12])*

*#Критерий Стьюдента*

*t.test(myData$wage[36:47], res4\_1\_1$mean[1:12], paired = TRUE)*

*#-- ФРАКТАЛЬНЫЙ МЕТОД*

*library(fractaldim)*

*values <- 12*

*endingIndex <- 35*

*total\_error <- 0*

*error\_per\_prediction <- c()*

*method <- "rodogram"*

*random\_sample\_count <- values*

*Sn1 <- as.data.frame(myData$wage[35:47], row.names = NULL) #test*

*Sn2 <- as.data.frame(e$y, row.names = NULL) #train*

*for(i in 1:values){*

*delta <- c()*

*for(j in 2:35){*

*delta <-rbind(delta, (e$y[j]-e$y[j-1]))*

*}*

*Std\_delta <- apply(delta, 2, sd)*

*V\_Reference <- fd.estimate(e$y, method=method, trim=TRUE)$fd*

*Sn\_guesses <- rnorm(random\_sample\_count, mean=e$y[35], sd=Std\_delta)*

*minDifference = 1000000*

*for(j in 1:length(Sn\_guesses)){*

*new\_Sn <- rbind(Sn2, Sn\_guesses[j])*

*new\_V\_Reference <- fd.estimate(new\_Sn$e, method=method, trim=TRUE)$fd*

*if(abs(new\_V\_Reference - V\_Reference) < minDifference){*

*Sn\_prediction <- Sn\_guesses[j]*

*minDifference = abs(new\_V\_Reference - V\_Reference)*

*}*

*}*

*Sn2 <- rbind(Sn2, Sn\_prediction)*

*}*

*s<- Sn2[,1]*

*ss<-s[(endingIndex+1):(endingIndex+values)]*

*plot(x[1:47], myData$wage[1:47], type="l", xlab="Месяцы с 01.2015", ylab = "Сумма")*

*lines(x[35:47], s[35:47], type = "l", col="green")*

*#КРИТЕРИИ*

*#тестовая box-диаграмма*

*boxplot(myData$wage[35:47], s[35:47])*

*#Критерий ХИ-квадрат*

*test0 <- matrix(c(myData$wage[35:47], s[35:47]), nrow = 2, byrow = TRUE)*

*chisq.test(test0)*

*#Критерий Фишера*

*var.test(myData$wage[35:47], s[35:47])*

*#Критерий Стьюдента*

*t.test(myData$wage[35:47], s[35:47], paired = TRUE)*

*#-- МЕТОД АВТОРЕГРЕССИИ(Непараметрический)*

*library(forecast)*

*e <- data.frame(y=myData$wage[1:35], x=x[1:35], z=myData$fastfood[1:35])*

*mod3<- auto.arima(e$y)*

*res3<- forecast(mod3, h=12) #230 прогнозируемых значений*

*plot(res3)*

*lines(x[35:47], myData$wage[35:47], type = "l", col="black")*