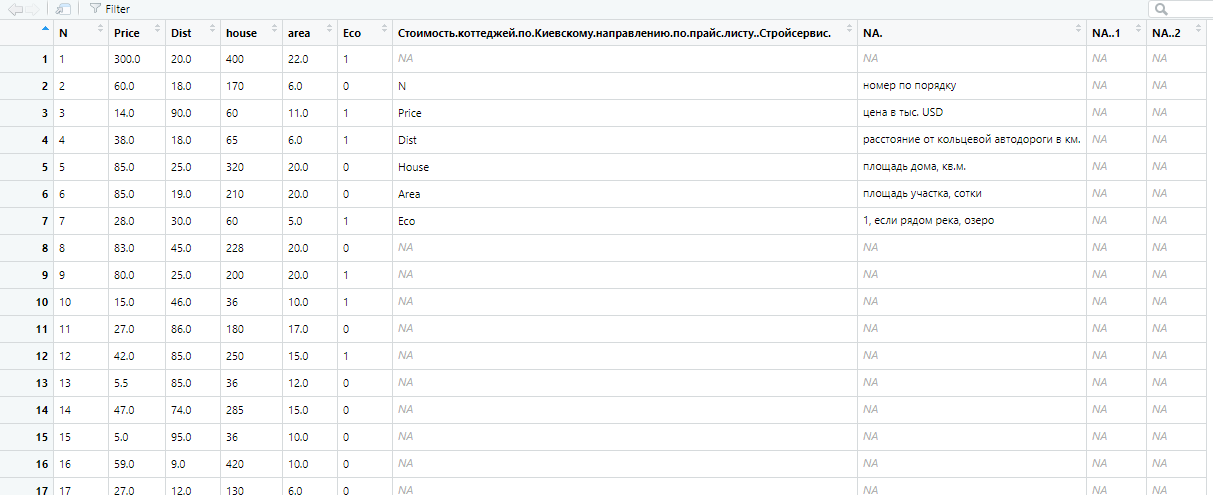
# **Задания**

## **Задача 1:**

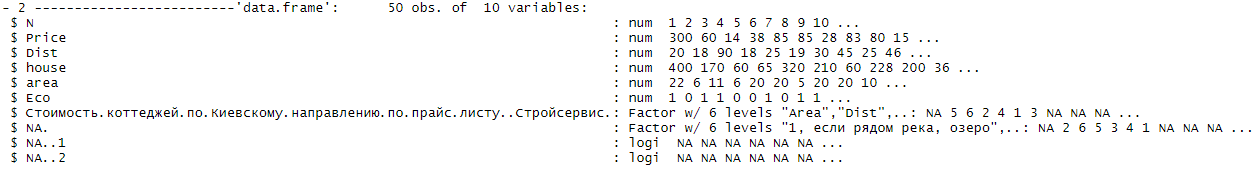
1. Загрузите данные из файла **VILLA.xls**
2. Определите тип данных, с которыми Вы работаете.
3. Рассчитайте и проинтерпретируйте описательные статистики по каждой переменной, включая фиктивную переменную.
4. Проанализируйте исходную выборку на наличие статистических выбросов, используя анализ ящичковых диаграмм. Сделайте выводы.
5. Проверьте однородность всех переменных с помощью коэффициента вариации по каждой переменной. Сделайте выводы.
6. Проверьте нормальность распределения переменной **Price** с помощью:
   1. гистограммы
   2. коэффициентов асимметрии и эксцесса
   3. графика Q-Qplot
   4. проверки гипотезы о нормально распределении ( на уровне значимости 0,05) с помощью критериев: Колмогорова-Смирнова, Шапиро-Уилка, Лиллифорса, Крамера-фон Мизеса и Андерсона-Дарлинга, Шапиро-Франсиа, хи-квадрат Пирсона. Сделайте выводы

### **Работа программы:**

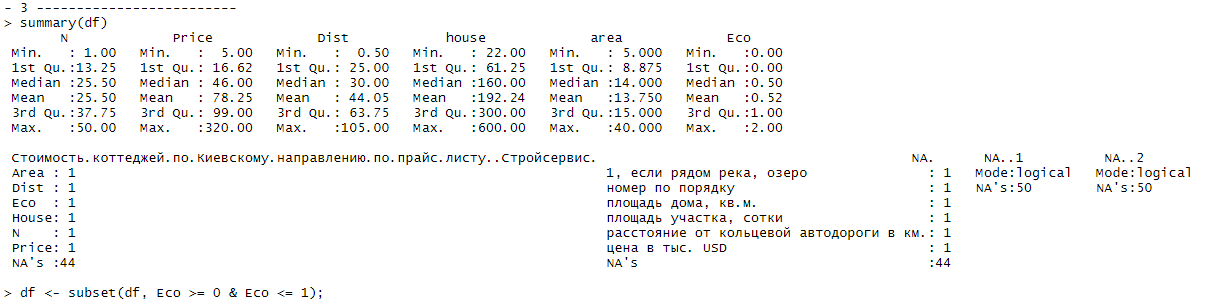
1



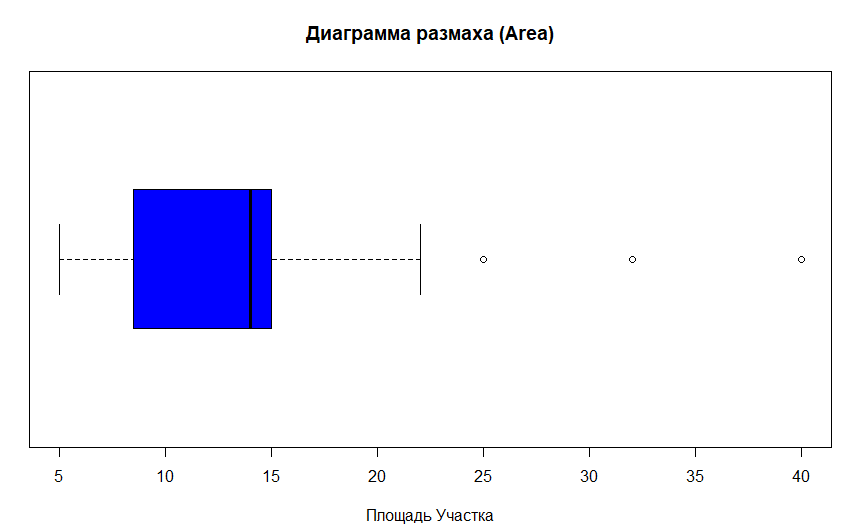
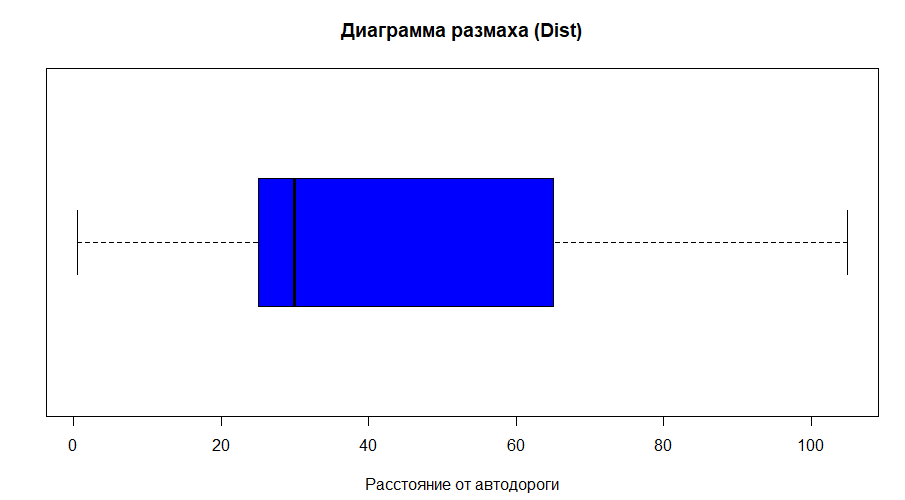
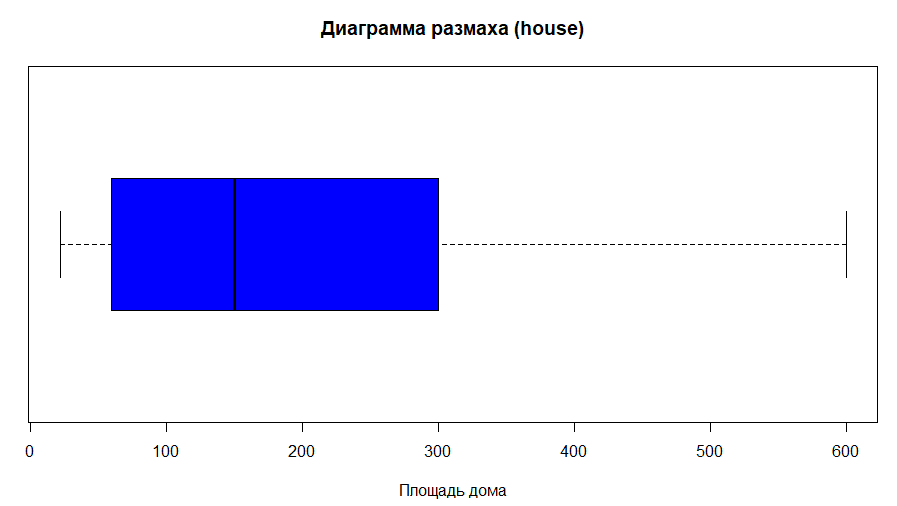
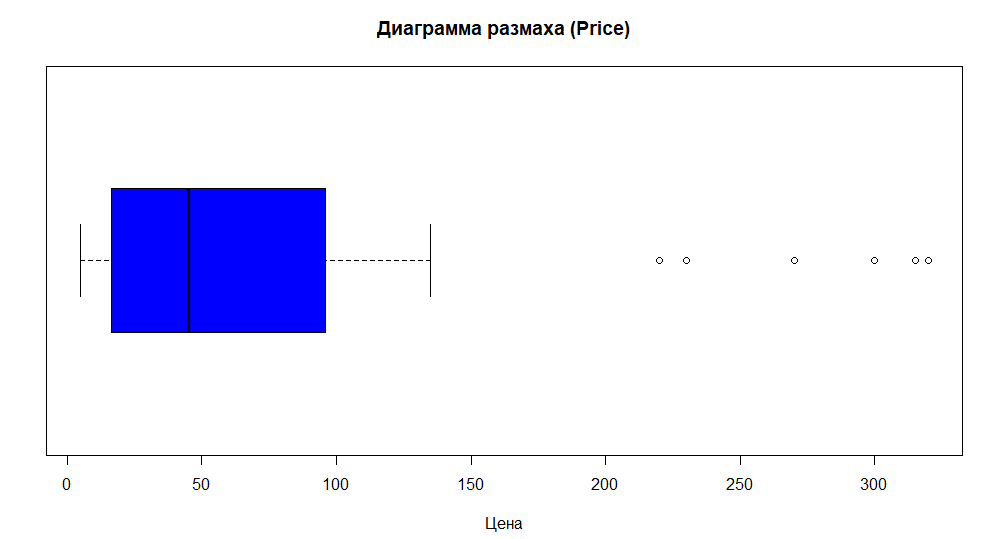
2



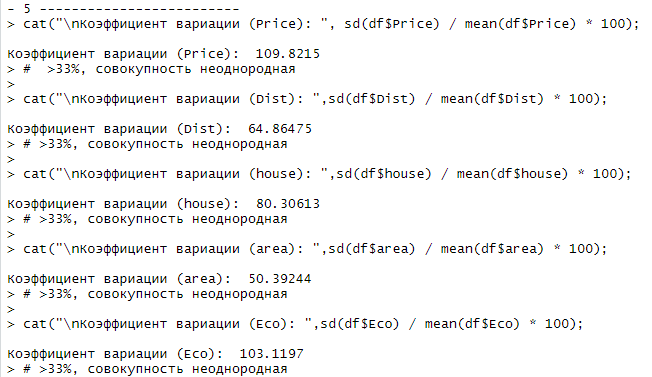
3



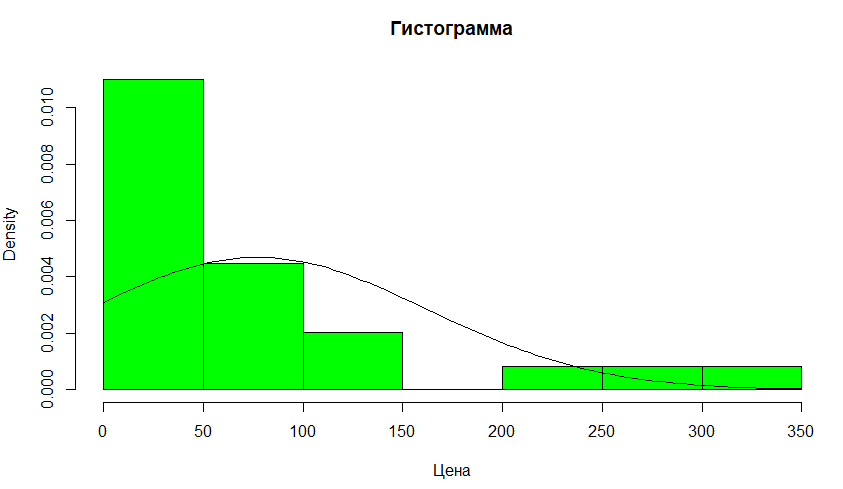
4



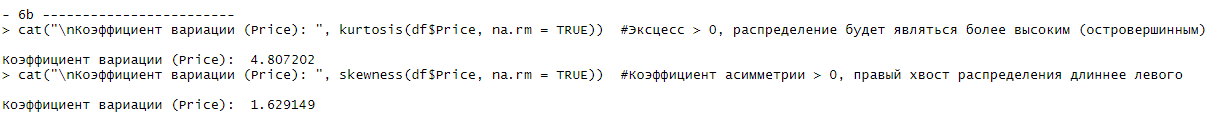
5



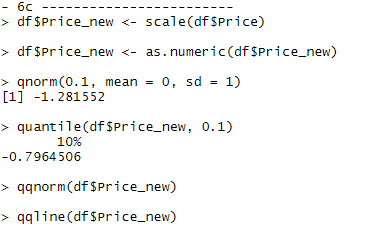
6a

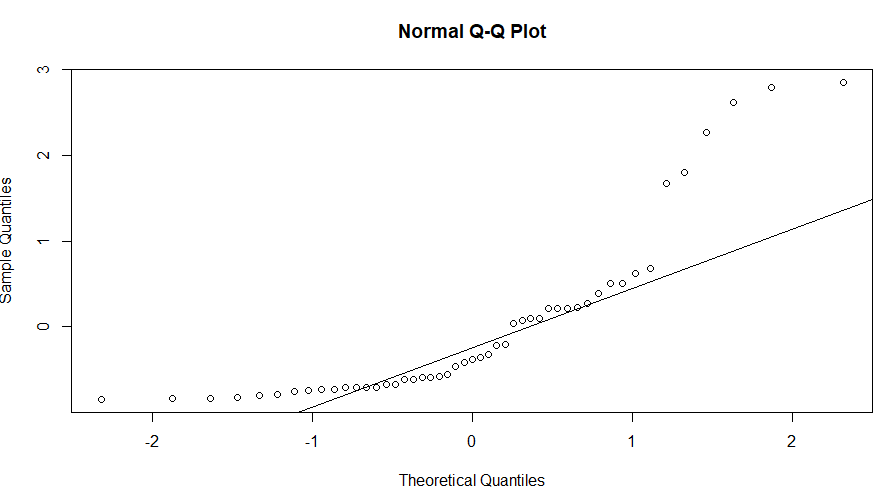


6b

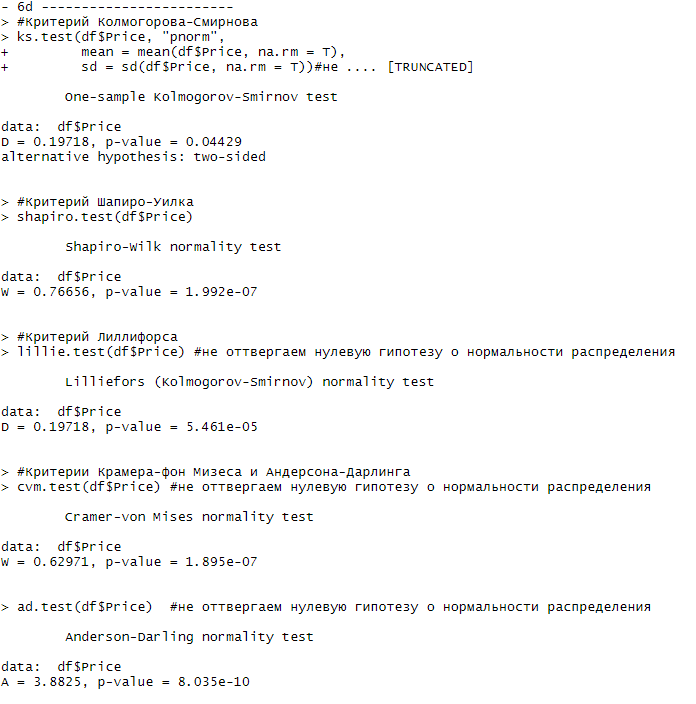


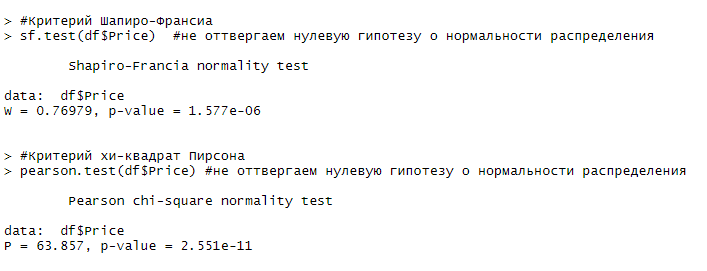
6c





6d





### **Листинг:**

#install.packages(c("xlsx", "moments", "nortest"));

library(nortest);

library(moments);

library(xlsx);

cat("\n- 1 -------------------------")

df <- read.xlsx("villa\_new.xlsx", 1,encoding = "UTF-8");

View(df);

cat("\n-----------------------------")

cat("\n- 2 -------------------------")

str(df);

cat("\n-----------------------------")

cat("\n- 3 -------------------------")

summary(df)

df <- subset(df, Eco >= 0 & Eco <= 1);

cat("\n-----------------------------")

cat("\n- 4 -------------------------")

boxplot(df$Price, data = df, xlab = "Цена", main = "Диаграмма размаха (Price)", col = "blue", horizontal = TRUE);

# 6 выбросов

boxplot(df$Dist, data = df, xlab = "Расстояние от автодороги", main = "Диаграмма размаха (Dist)", col = "blue", horizontal = TRUE);

# нет выбросов

boxplot(df$house, data = df, xlab = "Площадь дома", main = "Диаграмма размаха (House)", col = "blue", horizontal = TRUE);

# нет выбросов

boxplot(df$area, data = df, xlab = "Площадь Участка", main = "Диаграмма размаха (Area)", col = "blue", horizontal = TRUE);

# нет выбросов

cat("\n-----------------------------")

cat("\n- 5 -------------------------")

cat("\nКоэффициент вариации (Price): ", sd(df$Price) / mean(df$Price) \* 100);

# >33%, совокупность неоднородная

cat("\nКоэффициент вариации (Dist): ",sd(df$Dist) / mean(df$Dist) \* 100);

# >33%, совокупность неоднородная

cat("\nКоэффициент вариации (house): ",sd(df$house) / mean(df$house) \* 100);

# >33%, совокупность неоднородная

cat("\nКоэффициент вариации (area): ",sd(df$area) / mean(df$area) \* 100);

# >33%, совокупность неоднородная

cat("\nКоэффициент вариации (Eco): ",sd(df$Eco) / mean(df$Eco) \* 100);

# >33%, совокупность неоднородная

cat("\n-----------------------------")

cat("\n- 6a ------------------------")

hist(df$Price)

K <- round(1 + 3.32 \* log(nrow(df),10),0)

hist(df$Price, breaks = K, freq = FALSE, col = "green",

xlab = "Цена",

main = "Гистограмма")

curve(dnorm(x, mean(df$Price), sd = sd(df$Price)), add = TRUE)

cat("\n-----------------------------")

cat("\n- 6b ------------------------")

cat("\nКоэффициент вариации (Price): ", kurtosis(df$Price, na.rm = TRUE)) #Эксцесс > 0, распределение будет являться более высоким (островершинным)

cat("\nКоэффициент вариации (Price): ", skewness(df$Price, na.rm = TRUE)) #Коэффициент асимметрии > 0, правый хвост распределения длиннее левого

cat("\n-----------------------------")

cat("\n- 6c ------------------------")

df$Price\_new <- scale(df$Price)

df$Price\_new <- as.numeric(df$Price\_new)

qnorm(0.1, mean = 0, sd = 1)

quantile(df$Price\_new, 0.1)

qqnorm(df$Price\_new)

qqline(df$Price\_new)

cat("\n-----------------------------")

cat("\n- 6d ------------------------")

#Критерий Колмогорова-Смирнова

ks.test(df$Price, "pnorm",

mean = mean(df$Price, na.rm = T),

sd = sd(df$Price, na.rm = T))#не оттвергаем нулевую гипотезу о нормальности распределения

#Критерий Шапиро-Уилка

shapiro.test(df$Price)

#Критерий Лиллифорса

lillie.test(df$Price) #не оттвергаем нулевую гипотезу о нормальности распределения

#Критерии Крамера-фон Мизеса и Андерсона-Дарлинга

cvm.test(df$Price) #не оттвергаем нулевую гипотезу о нормальности распределения

ad.test(df$Price) #не оттвергаем нулевую гипотезу о нормальности распределения

#Критерий Шапиро-Франсиа

sf.test(df$Price) #не оттвергаем нулевую гипотезу о нормальности распределения

#Критерий хи-квадрат Пирсона

pearson.test(df$Price) #не оттвергаем нулевую гипотезу о нормальности распределения

cat("\n-----------------------------")