# **Задания**

## **Задача 1:**

1. Загрузите данные из файла **VILLA.xls**
2. Определите тип данных, с которыми Вы работаете.
3. Проверьте нормальность распределения переменной **Price** с помощью критерия Лиллифорса.
4. Проверьте гипотезу о равенстве дисперсий Цены коттеджей (**Price**) в двух совокупностях (рядом с озером и нет).
5. В зависимости от предыдущего результата проверить гипотезу о равенстве Цены коттеджей (**Price**) в двух совокупностях (рядом с озером и нет). Сделайте выводы.

### **Работа программы:**

### **Листинг:**

## **Задача 2:**

Будем работать с базой данных psych\_survey.csv

Проверить гипотезу о равенстве среднего роста (height) у студентов с разным любимым предметом (subject)

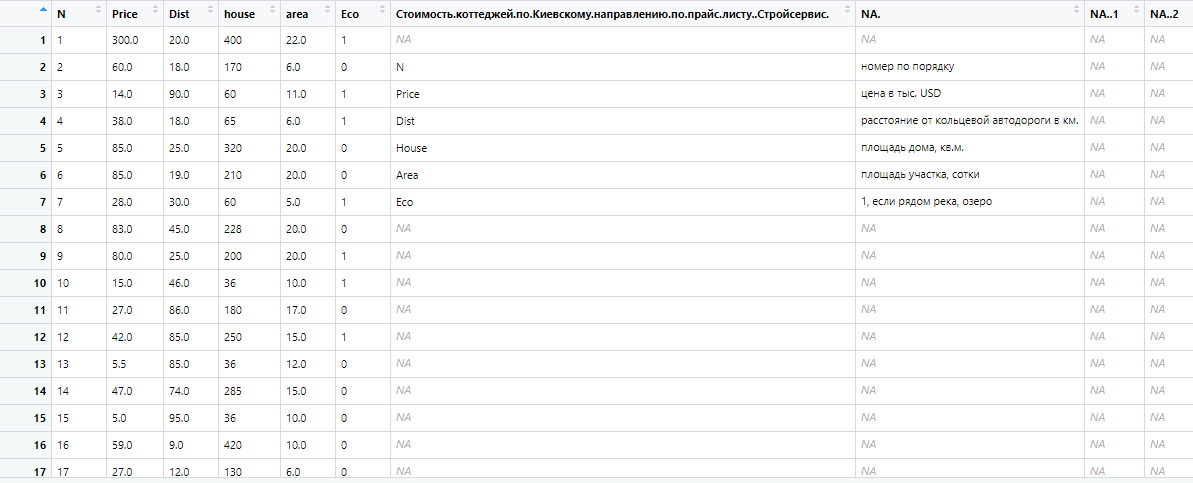
Любимый предмет. Респонденту нужно было выбрать один ответ из 5 предложенных вариантов:

1. Математика;
2. Биология;
3. Русский язык;
4. Иностранный язык;

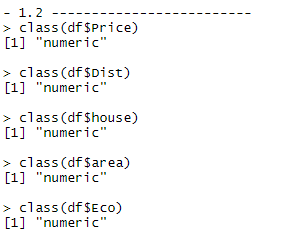
Ни один из вышеперечисленных предметов.

### **Работа программы:**

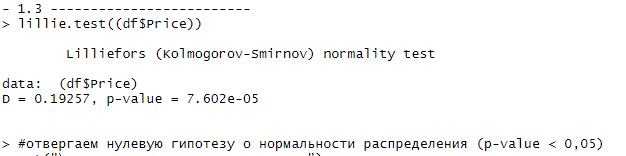
1.1



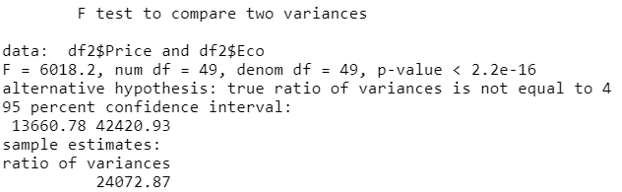
1.2



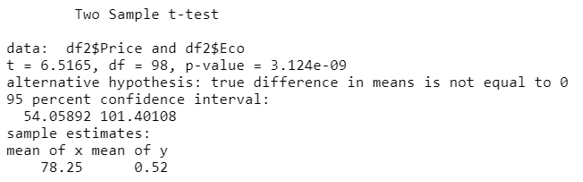
1.3



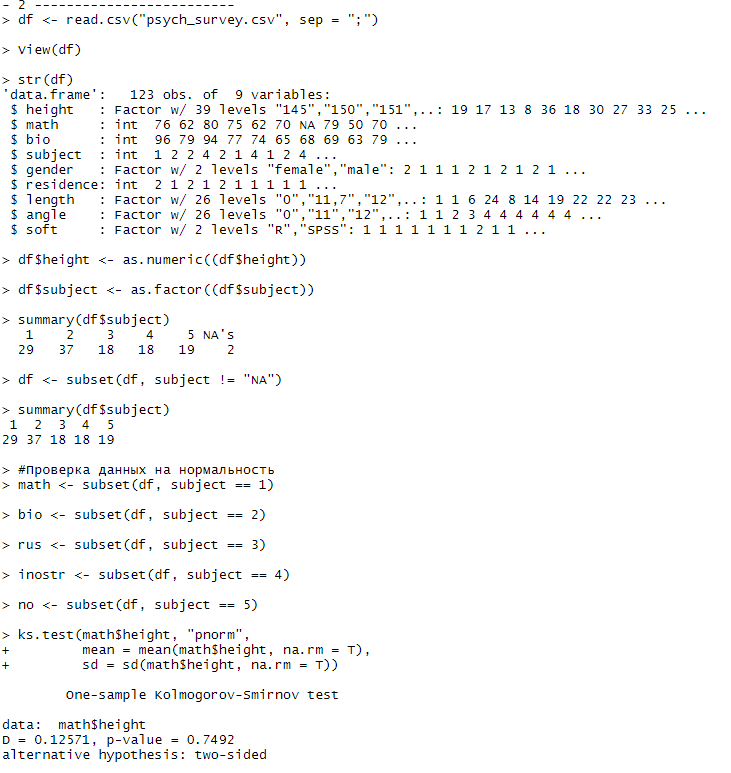
1.4

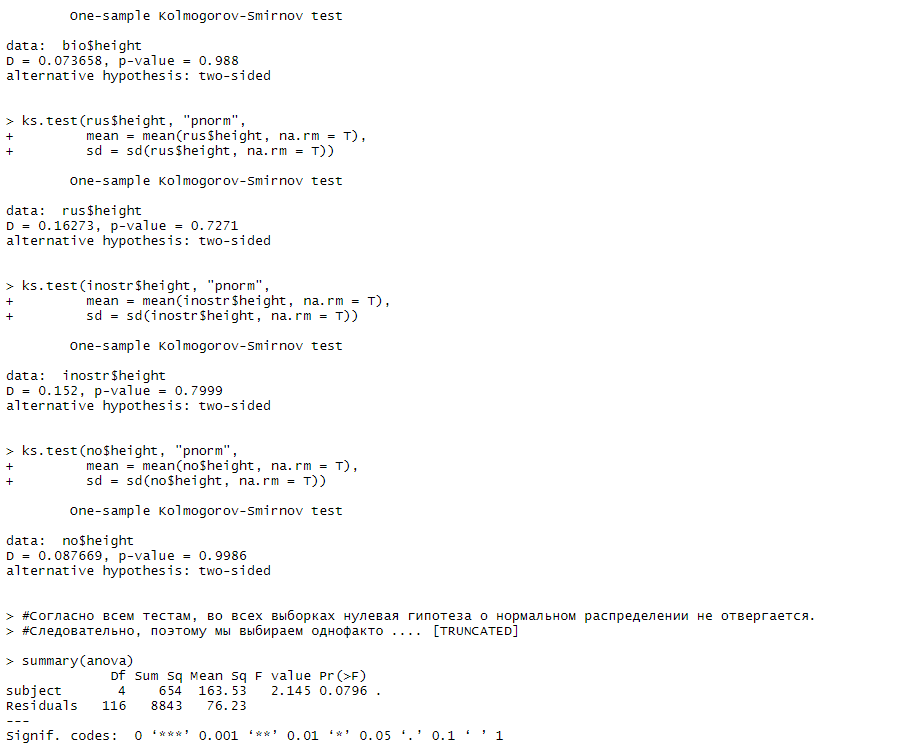


1.5



2





### **Листинг:**

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

library(nortest);

library(xlsx);

cat("\n- 1.1 -------------------------")

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

View(df);

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

cat("\n- 1.2 -------------------------")

class(df$Price)

class(df$Dist)

class(df$house)

class(df$area)

class(df$Eco)

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

cat("\n- 1.3 -------------------------")

lillie.test((df$Price))

#отвергаем нулевую гипотезу о нормальности распределения (p-value < 0,05)

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

cat("\n- 1.4 -------------------------")

var.test(df$Price ~ df$Eco, data = df, alternative = "two.sided")

#гипотеза о равенстве дисперсий отвергается (p-value < 0,05)

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

cat("\n- 1.5 -------------------------")

#исходя из результатов предыдущего номера используем двухвыборочный критерий Стьюдента равенства средних

#(t-критерий в модификации Уэлча (Welch) с неравными дисперсиями)

t.test(Price ~ Eco, df, var.equal = FALSE)

#вывод: нулевая гипотеза о равенстве средних отвергается, поскольку p-value меньше уровня значимости 0,05

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

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

df <- read.csv("psych\_survey.csv", sep = ";")

View(df)

str(df)

df$height <- as.numeric((df$height))

df$subject <- as.factor((df$subject))

summary(df$subject)

df <- subset(df, subject != "NA")

summary(df$subject)

#Проверка данных на нормальность

math <- subset(df, subject == 1)

bio <- subset(df, subject == 2)

rus <- subset(df, subject == 3)

inostr <- subset(df, subject == 4)

no <- subset(df, subject == 5)

ks.test(math$height, "pnorm",

mean = mean(math$height, na.rm = T),

sd = sd(math$height, na.rm = T))

ks.test(bio$height, "pnorm",

mean = mean(bio$height, na.rm = T),

sd = sd(bio$height, na.rm = T))

ks.test(rus$height, "pnorm",

mean = mean(rus$height, na.rm = T),

sd = sd(rus$height, na.rm = T))

ks.test(inostr$height, "pnorm",

mean = mean(inostr$height, na.rm = T),

sd = sd(inostr$height, na.rm = T))

ks.test(no$height, "pnorm",

mean = mean(no$height, na.rm = T),

sd = sd(no$height, na.rm = T))

#Согласно всем тестам, во всех выборках нулевая гипотеза о нормальном распределении не отвергается.

#Следовательно, поэтому мы выбираем однофакторный дисперсионный анализ для сравнения средних в нескольких группах.

anova <- aov(height ~ subject, data = df)

summary(anova)

#В данном случае мы не отвергаем нулевую гипотезу об отсутствии различий между всеми средними против альтернативы о том, что хотя бы одно среднее отличается.

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