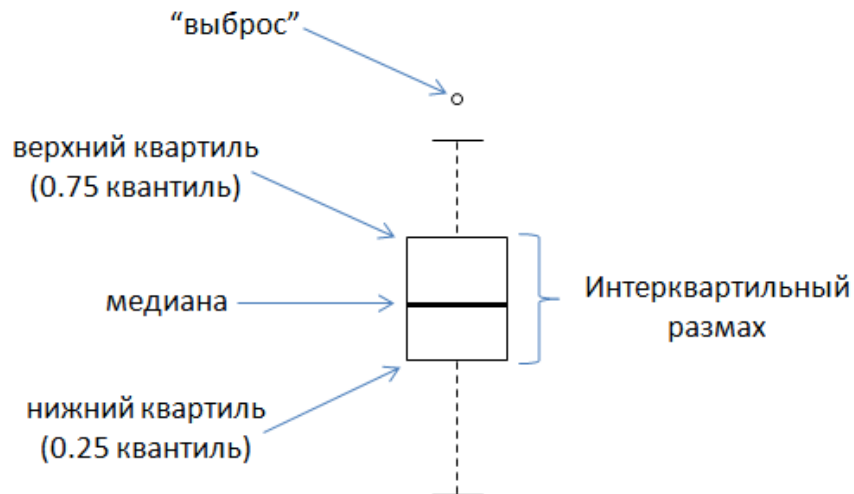


Для действий первой выборки:

#R version 3.4.4

#Узнать, если ли выбросы в данных, используя boxplot

Пример общего строения диаграммы размахов:



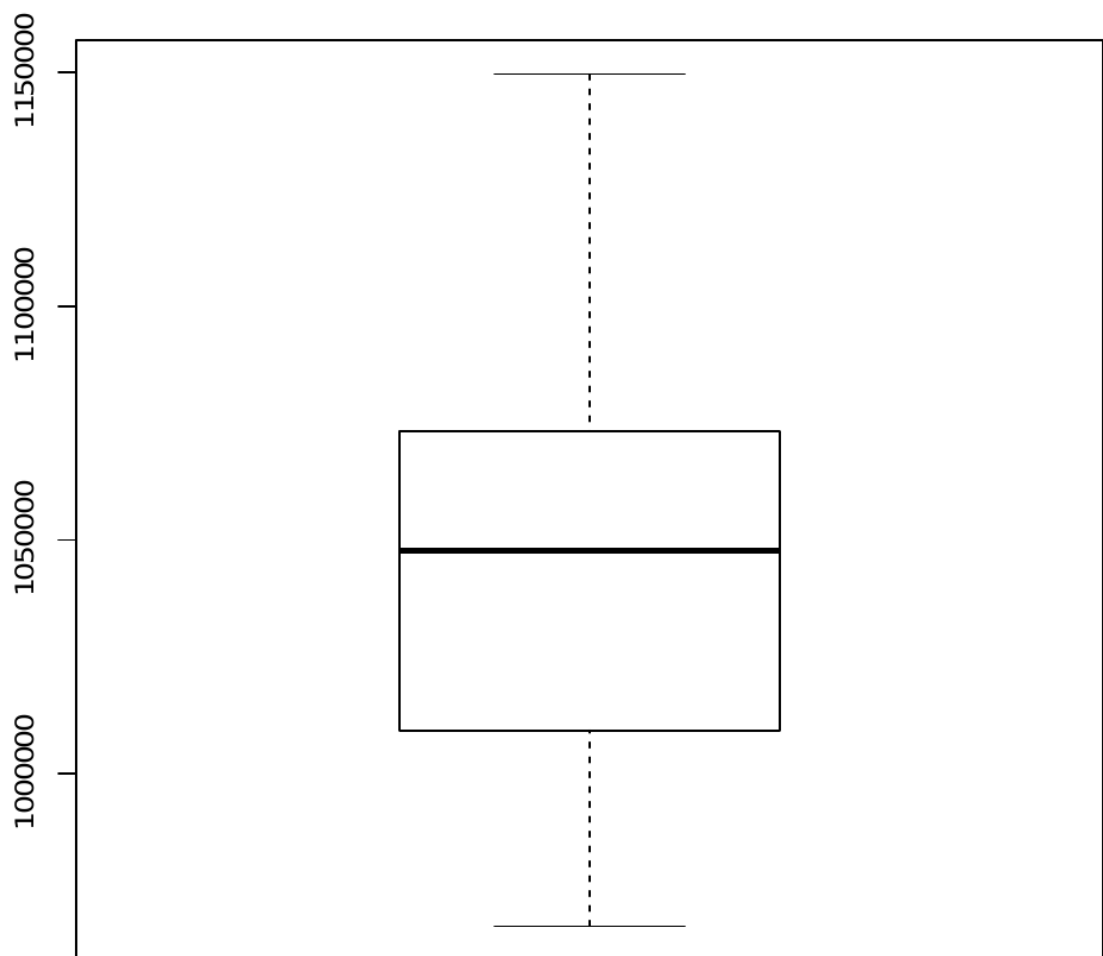
```
test_v1 <-
```

```
c(1048186.20,1111170.00,1081608.60,1035531.60,1069407.00,1073191.50,1149624.90,102971  
7.00,1037918.10,1072230.60,998883.30,1045217.70,979078.20,1114703.70,1086142.50,10616  
15.70,1000016.40,1047059.70,1009198.50,967327.20,997769.70,1061553.00)
```

```
boxplot(test_v1)
```

Выбросов нет, так как нет точек лежащих за максимальным и мин значениями

Выбросов нет, так как нет точек лежащих за максимальным и мин значениями



Выбросов нет, так как нет точек, лежащих за максимальным и мин значениями

#Узнать, является ли наша выборка нормально распределенной, используя тест Шапиро  
#Уилко

В общем виде проверяемую при помощи этих тестов нулевую гипотезу можно сформулировать так: "Анализируемая выборка происходит из генеральной совокупности, имеющей нормальное распределение". Если получаемая при помощи того или иного теста вероятность ошибки  $P$  оказывается меньше некоторого заранее принятого уровня значимости (например, 0.05), нулевая гипотеза отклоняется.

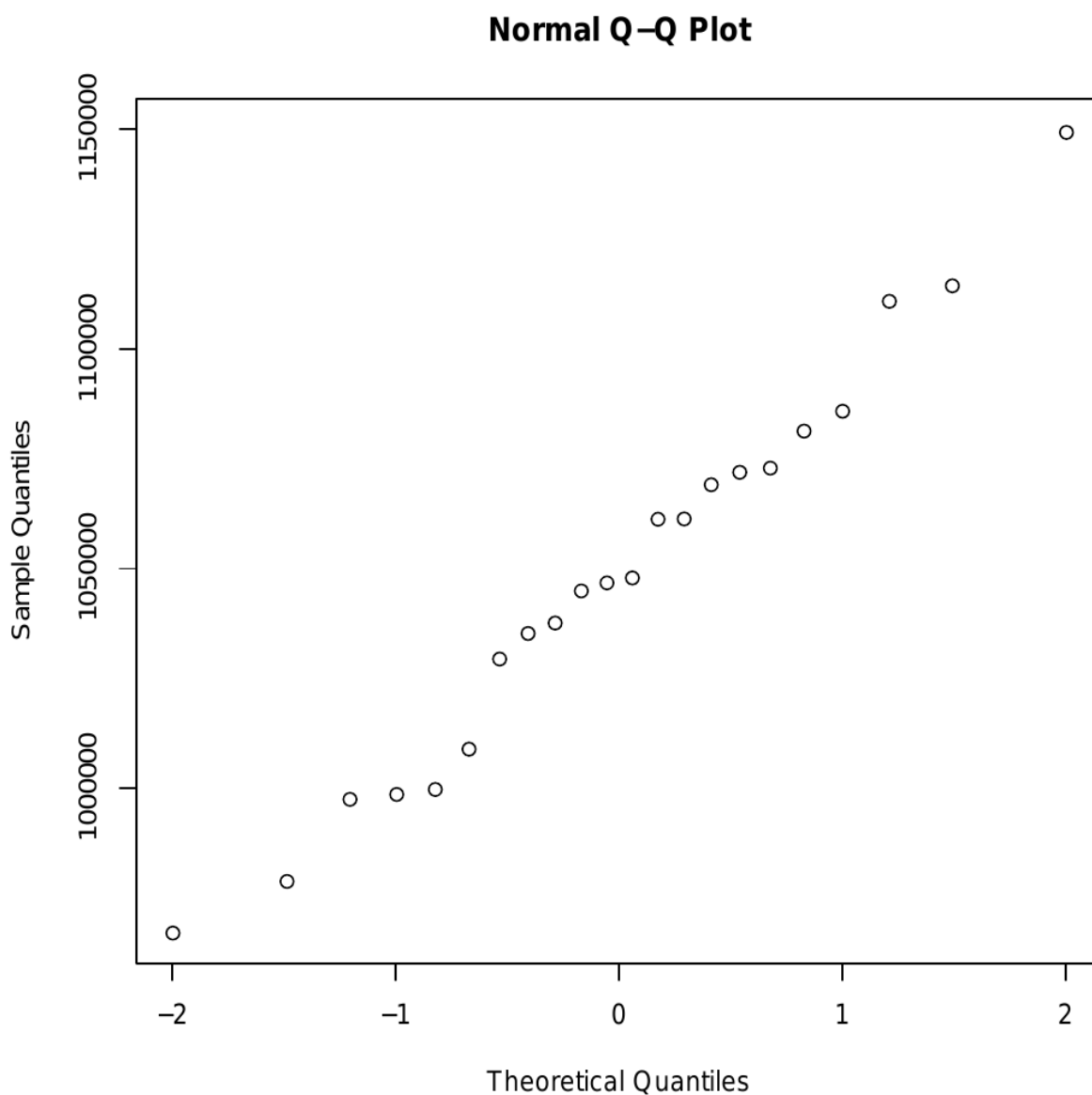
```
shapiro.test(test_v1)
```

```
Shapiro-Wilk normality test
data:  test_v1
W = 0.98164, p-value = 0.9397
```

*#  $P > 0.05$  - нулевая гипотеза не отвергается, значит рассматриваемая выборка из ген совокупности с норм распред*

При нормальном распределении проверяемой переменной точки на графике квантилей должны выстраиваться в прямую линию, исходящую под углом 45 градусов из левого нижнего угла графика. Графики квантилей особенно полезны при работе с небольшими по размеру совокупностями, для которых невозможно построить гистограммы, принимающие какую-либо выраженную форму.

qqnorm(test\_v1)



# точки выстроены вокруг прямой 45 град - можно считать норм распр

#Выборка 2:

```
test_v2 <-  
c(1746977.00,1851950.00,1802681.00,1725886.00,1782345.00,1788652.50,1916041.50,171619  
5.00,1729863.50,1787051.00,1664805.50,1742029.50,1631797.00,1857839.50,1810237.50,176  
9359.50,1666694.00,1745099.50,1681997.50,1612212.00,1662949.50,1769255.00)
```

```
boxplot(test_v2)
```

```
shapiro.test(test_v2)
```

```
qqnorm(test_v2)
```

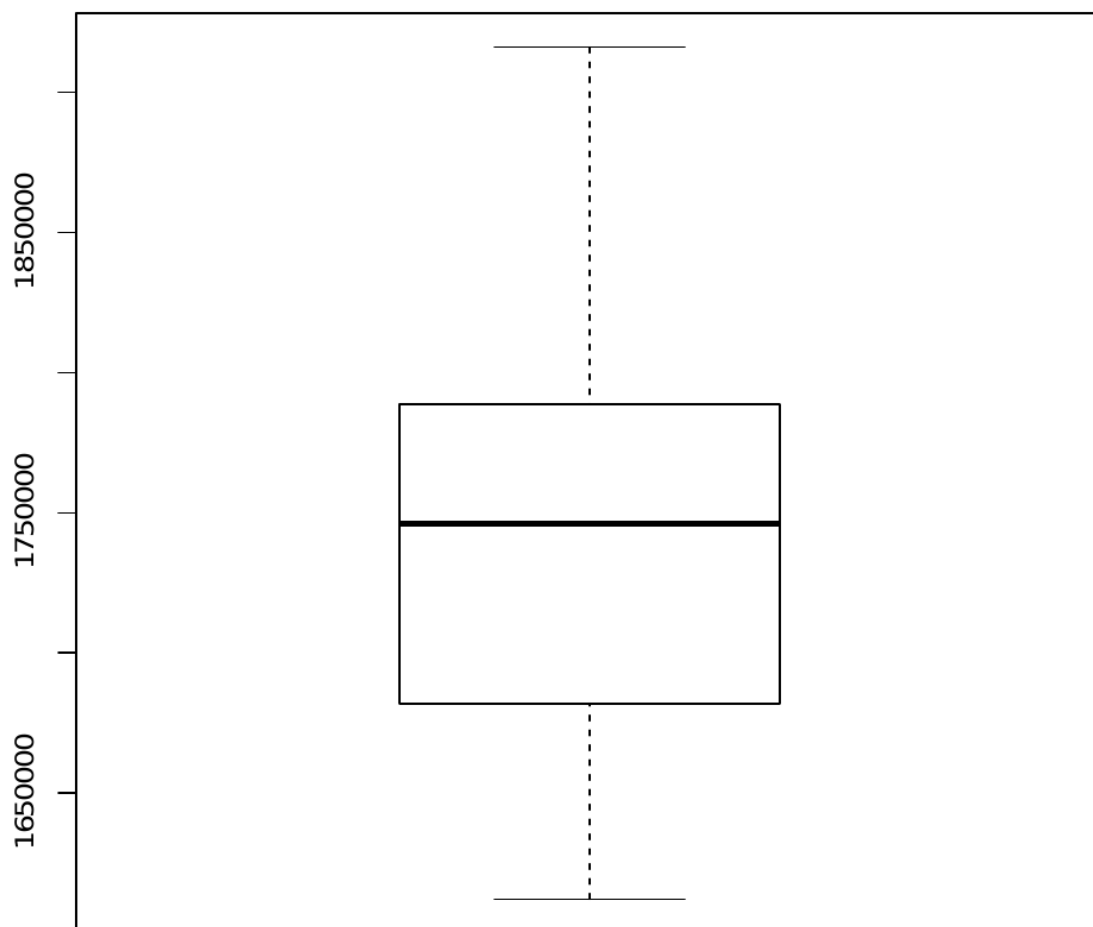
*#результаты:*

```
Shapiro-Wilk normality test
```

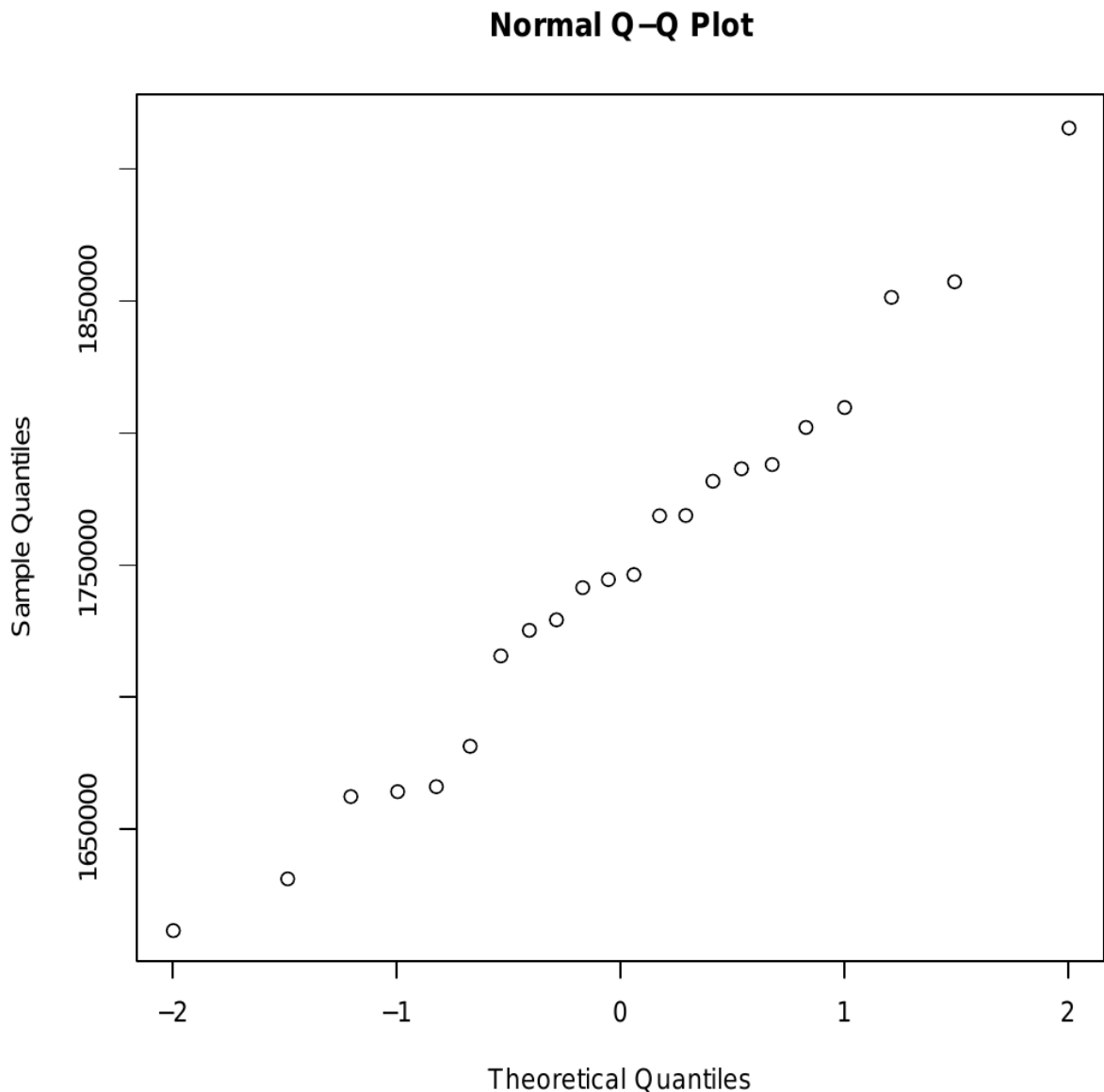
```
data: test_v2
```

```
W = 0.98164, p-value = 0.9397
```

*# P > 0.05 - нулевая гипотеза не отвергается, значит рассматриваемая выборка из ген совокупности с норм распред*



Выбросов нет, так как нет точек лежащих за максимальным и мин значениями



# точки выстроены вокруг прямой 45 град - можно считать норм распр

#доп проверка выбросов:

```
test_v1 <-
```

```
c(1048186.20,1111170.00,1081608.60,1035531.60,1069407.00,1073191.50,1149624.90,102971
7.00,1037918.10,1072230.60,998883.30,1045217.70,979078.20,1114703.70,1086142.50,10616
15.70,1000016.40,1047059.70,1009198.50,967327.20,997769.70,1061553.00)
```

```
test_v2 <-
```

```
c(1746977.00,1851950.00,1802681.00,1725886.00,1782345.00,1788652.50,1916041.50,171619
5.00,1729863.50,1787051.00,1664805.50,1742029.50,1631797.00,1857839.50,1810237.50,176
9359.50,1666694.00,1745099.50,1681997.50,1612212.00,1662949.50,1769255.00)
```

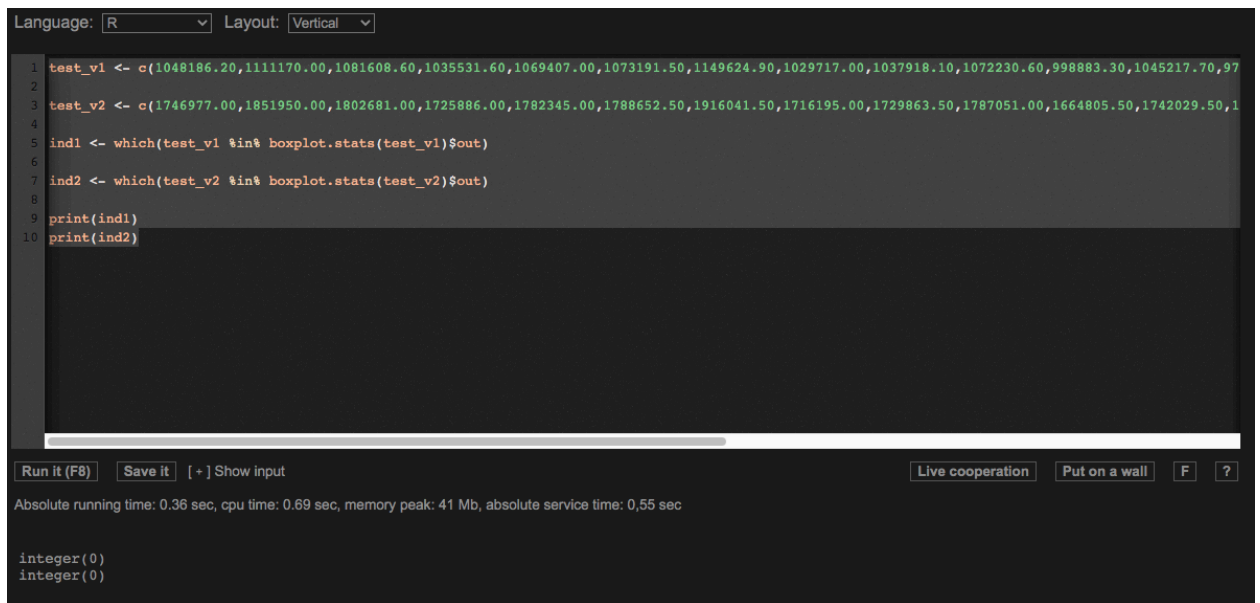
```
ind1 <- which(test_v1 %in% boxplot.stats(test_v1)$out)
```

```
ind2 <- which(test_v2 %in% boxplot.stats(test_v2)$out)
```

```
print(ind1)
```

```
print(ind2)
```

Дополнительно проверим наличие выбросов – выбросов нет



```
Language: R Layout: Vertical
1 test_v1 <- c(1048186.20,1111170.00,1081608.60,1035531.60,1069407.00,1073191.50,1149624.90,1029717.00,1037918.10,1072230.60,998883.30,1045217.70,97
2
3 test_v2 <- c(1746977.00,1851950.00,1802681.00,1725886.00,1782345.00,1788652.50,1916041.50,1716195.00,1729863.50,1787051.00,1664805.50,1742029.50,1
4
5 ind1 <- which(test_v1 %in% boxplot.stats(test_v1)$out)
6
7 ind2 <- which(test_v2 %in% boxplot.stats(test_v2)$out)
8
9 print(ind1)
10 print(ind2)
```

Run it (F8) Save it [ + ] Show input Live cooperation Put on a wall F ?

Absolute running time: 0.36 sec, cpu time: 0.69 sec, memory peak: 41 Mb, absolute service time: 0.55 sec

```
integer(0)
integer(0)
```

В качестве метода дальнейшего исследования можно использовать Метод Стьюдента (t-распределение). t-критерий Стьюдента используется для определения статистической значимости различий средних величин. Для применения t-критерия Стьюдента необходимо, чтобы исходные данные имели нормальное распределение. Используется для проведения t-теста: теста на равенство матожидания выборки стандартно нормально распределённых случайных величин некоторому значению, либо равенства матожиданий двух нормальных выборок с одинаковой дисперсией (равенство дисперсий необходимо проверять f-тестом).