«КАЗАНСКИЙ (ПРИВОЛЖСКИЙ) ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Лабораторная работа №3 по теме: «Моделирование данных с заданным законом распределения и визуализация данных»

Выполнила

Студент 3 курса

группы 09-115(3)

Зиновьев Е. А.

Преподаватель:

Шустова Е.П

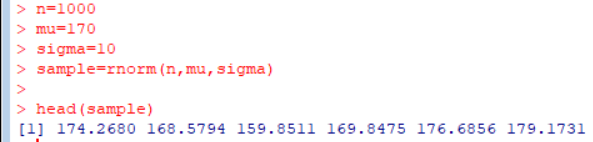
Казань 2021

**Ход работы:**

***Нормальное распределение.***

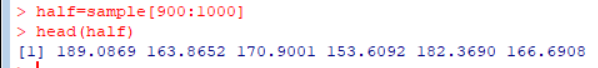
1. Пусть с.в., распределение которой можно считать близким к нормальному - это измерения роста у 1000 человек в сантиметрах: <https://allasamsonova.ru/statistika/zakon-normalnogo-raspredelenija/>

В данном случае объем выборки будет равен n=1000. Зададим параметры распределения: mu=170, sigma=10.



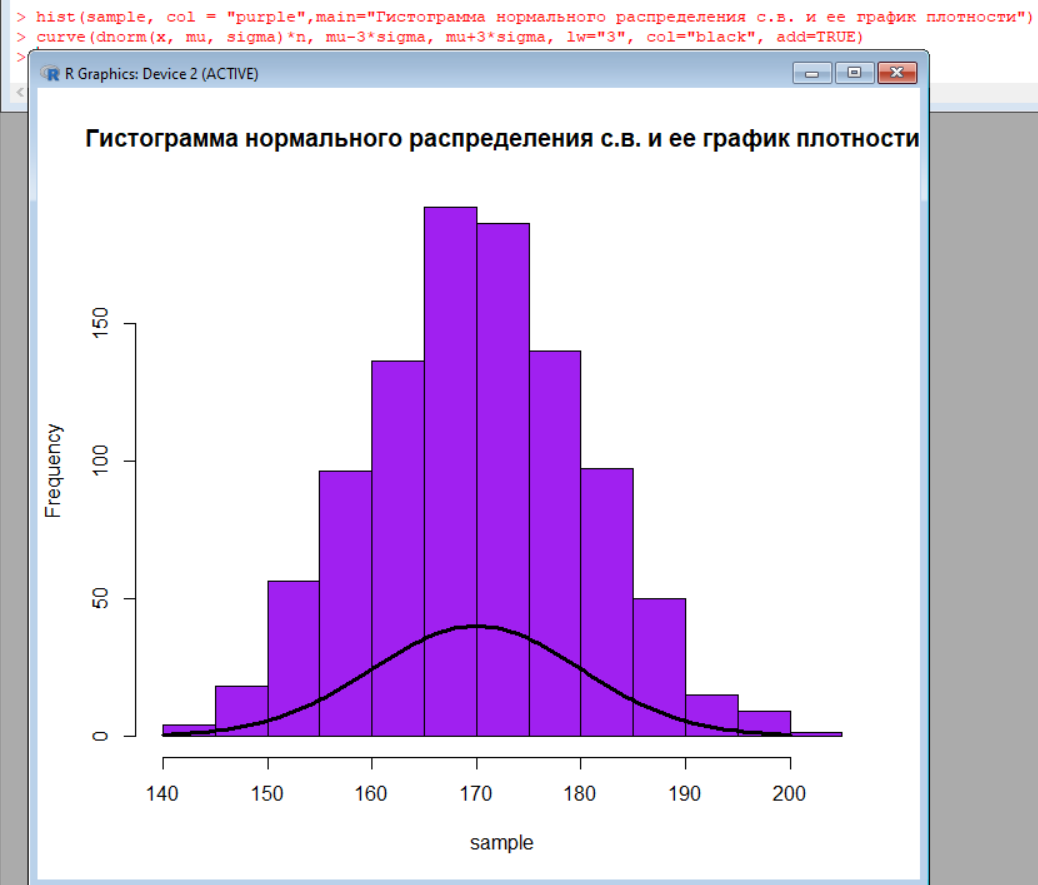
Как видим, все первые 5 значений генеральной совокупности отклоняются не больше, чем на 1 стандартное отклонение от среднего.

2. Сформируем выборку из генеральной совокупности: возьмем последние 100 значений генеральной совокупности:



Как видим, в этой выборке в первые 5 значений уже попали значения, отклоняющиеся от среднего больше, чем на 1 стандартное отклонение.

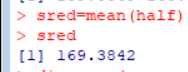
3. Построим гистограмму распределения (сиреневого цвета) и совместим ее с графиком плотности (черного цвета) на интервале [mu-3\*sigma, mu+3\*sigma] для генеральной совокупности:



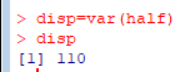
По гистограмме видим, что по 200 значений распределено между 165-170 и 170-175 (половины от 1 стандартного отклонения), то есть 400/1000 = 0.4 = 40% значений находится в половине стандартного отклонения от среднего.

4. Получим выборочные оценки параметров из распределения и сравним их с истинными значениями этих параметров, заданных при моделировании. (Для наглядности представим результаты сравнения в виде таблицы):

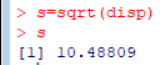
Найдем выборочное среднее:



Найдем выборочную дисперсию:



Найдем выборочное среднеквадратичное отклонение:



|  |  |  |
| --- | --- | --- |
| **Параметры** | **Истинные значения** | **Полученные оценки** |
| mu | 170 | 169.3842 |
| sigma | 10 | 10.48809 |

Таким образом, среднее значение подобранной выборки почти не изменилось (изменение составило 0.035% от среднего ген.совокупности), но зато сами значения этой выборки стали больше колебаться, то есть среднеквадратическое отклонение увеличилось (на 4.9%).

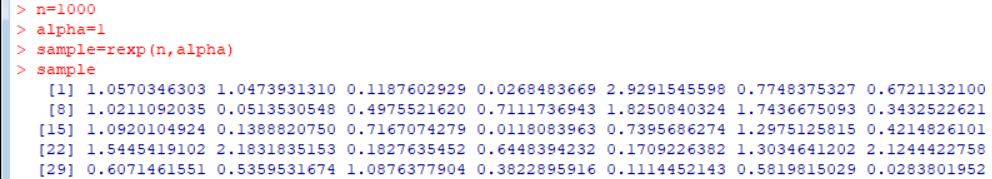
5. Используя функции rexp, rt и runif получите выборки из распределений (параметры распределений выберите произвольно):

* экспоненциального (показательного) распределения;
* распределения Стьюдента ,
* равномерного распределения.

Для каждой полученной выборки выполните пункты 3 - 4

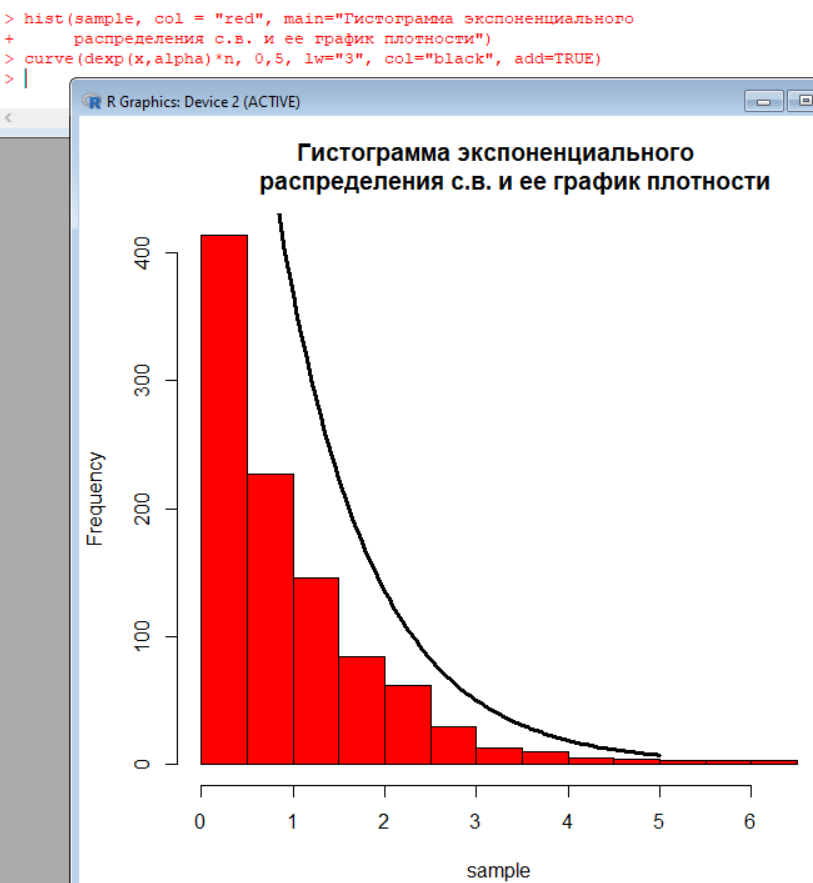
***5.1 Экспоненциальное (показательное) распределение***

1. Пусть объем выборки равен n=1000, параметр принимает значение alpha=1.



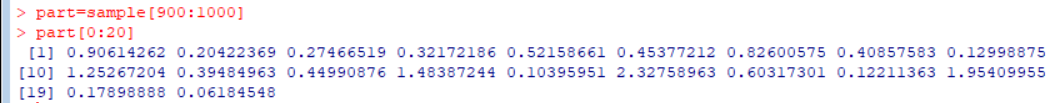
Видим, что значения, в основном, колеблются от 0 до 2.

Построим гистограмму экспоненциального распределения с.в. (красного цвета) и совместим ее с графиком плотности (черного цвета) на интервале [0;5]:



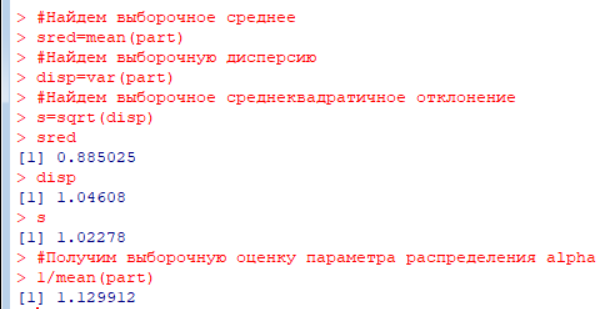
По гистограмме видим, что около 65% всех значений генеральной совокупности находятся от 0 до 1

2. Сформируем выборку из 100 последних значений генеральной совокупности:



Как видим, тут также почти все значения находятся в диапазоне от 0 до 1.

3. Получим параметр альфа из подобранной выборки и сравним ее с истинным значением параметра альфа, заданного при моделировании. (Для наглядности представим результаты сравнения в виде таблицы):

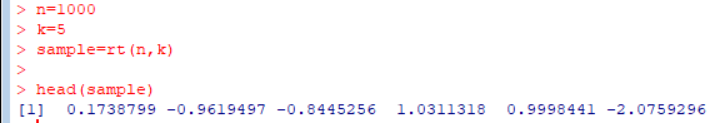


|  |  |  |
| --- | --- | --- |
| **Параметр** | **Истинное значение** | **Полученная оценка** |
| alpha | 1 | 1.1299 |

Видим, что параметр альфа сильно увеличился на 12% у выборки.

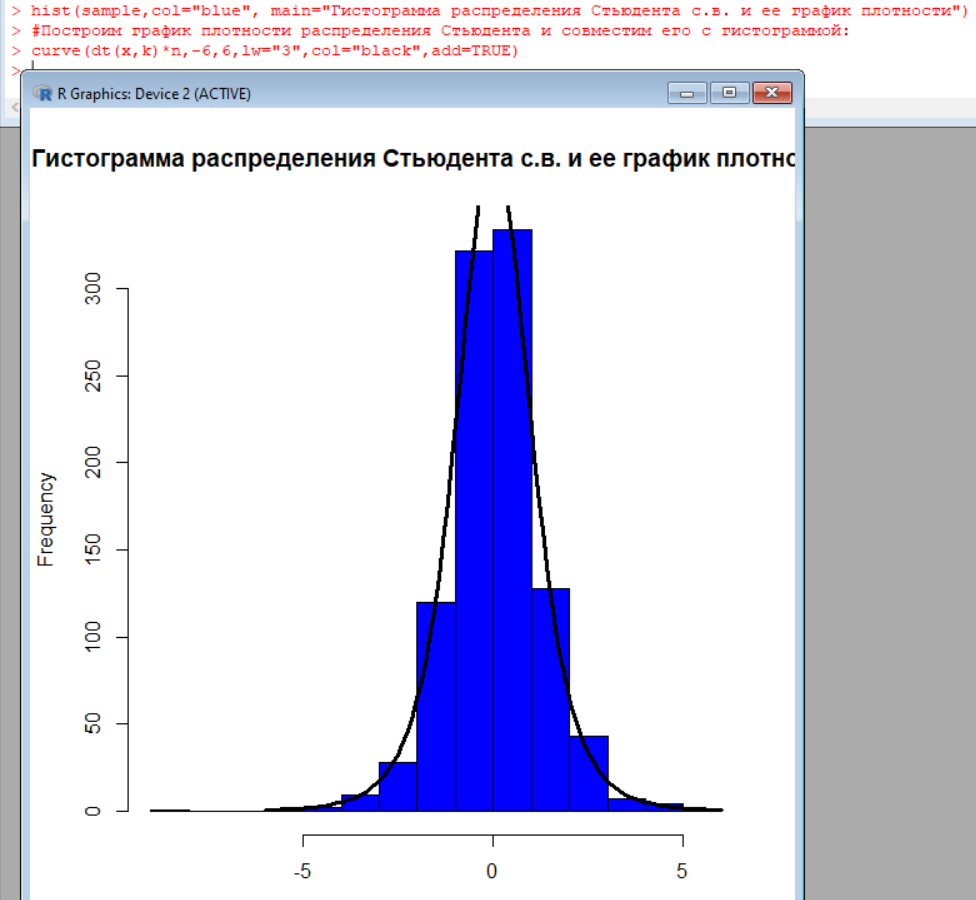
***5.2 Распределение Стьюдента***

1. Пусть объем выборки равен n =1000, параметр принимает значение k=5 (где k - число степеней свободы).



Первые 5 значений колеблются в районе от -2 до 1.

Построим гистограмму распределения Стьюдента с k степенями свободы (синего цвета) и совместим ее с графиком плотности (черного цвета) на интервале [-6;6]:



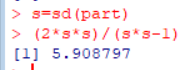
Около 65 % значений ген.совокупности лежат в диапазоне от -1 до 1.

2. Сформируем выборку из 100 последних значений генеральной совокупности:



Первые 5 значений теперь уже лежат от -1.3 до 1.76.

3. Получим параметр k из подобранной выборки и сравним ее с истинным значением параметра k, заданного при моделировании. (Для наглядности представим результаты сравнения в виде таблицы):

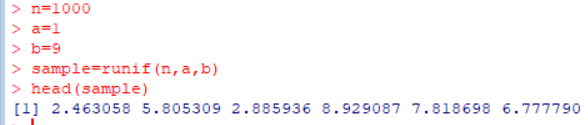


|  |  |  |
| --- | --- | --- |
| **Параметр** | **Истинное значение** | **Полученная оценка** |
| k | 5 | 5.908 |

Видим, что выборочный параметр (k) еще даже больше увеличен относительно экспоненциального распределения у выборки (на 18%).

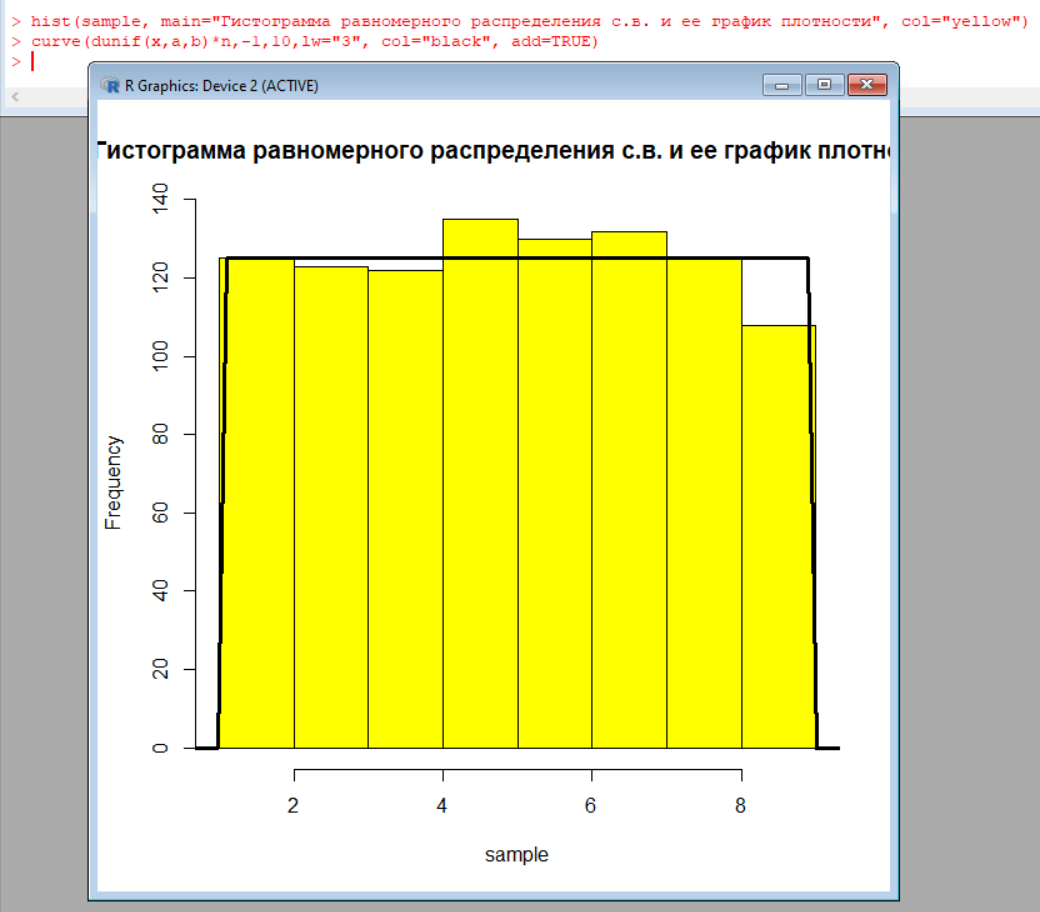
***5.3 Равномерное распределение***

1. Пусть объем выборки равен n=1000, параметры принимают значения a=1, b=9.



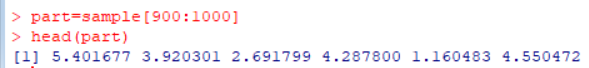
Первые 5 значений ген.совокупности лежат в диапазоне от 2.46 до 8.92.

Построим гистограмму равномерного распределения (цвет гистограммы - красный) и совместим ее с графиком плотности (синего цвета) на интервале [-1;10]:



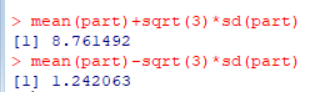
Видим, что значения от 1 до 9 равномерно распределены, то есть каждое значение встречается примерно с одинаковой частотой.

2. Сформируем выборку из 100 последних значений генеральной совокупности:



Первые 5 значений выборки лежат в диапазоне от 1.16 до 5.4.

3. Получим параметры a, b из подобранной выборки и сравним их с истинными значениями параметрами a, b, заданных при моделировании. (Для наглядности представим результаты сравнения в виде таблицы):



|  |  |  |
| --- | --- | --- |
| **Параметры** | **Истинные значения** | **Полученные оценки** |
| a | 1 | 1.242 |
| b | 9 | 8.761 |

Параметр а у выборки больше на 25 % (больше скорее связано с мелким значением), а параметр b меньше на 3 %. То есть с помощью последних ста значений удалось довольно хорошо перепостроить равномерное распределение.

**Общий вывод:** судя по выборочным параметрам распределений, с помощью последних 100 значений выборки из генеральной совокупности, состоящей из 1000 значений, удалось довольно хорошо отразить генеральную совокупность при нормальном (среднее значение - 0.035%, среднеквадратическое отклонение - 4.9%) и равномерном распределении (b – 3%), и хуже при экспоненциальном (параметр альфа – 12%) и распределении Стьюдента (параметр k – 18%).

**Код:**

#Назначим рабочий директорий и проверим его

setwd("C://Users/users/Desktop/стат")

getwd()

# Смоделируем выборку объёма 1000 из нормального распределения с параметрами: mu = 170; sigma = 10

n=1000

mu=170

sigma=10

sample=rnorm(n,mu,sigma)

#Сформируем выборку из ген. совокупности:

half=sample[900:1000]

head(half)

# Построим гистограмму:

hist(sample, col = "purple",main="Гистограмма нормального распределения с.в. и ее график плотности")

curve(dnorm(x, mu, sigma)\*n, mu-3\*sigma, mu+3\*sigma, lw="3", col="black", add=TRUE)

#Найдем выборочное среднее

sred=mean(half)

#Найдем выборочную дисперсию

disp=var(half)

#Найдем выборочное среднеквадратичное отклонение

s=sqrt(disp)

# Смоделируем выборку объёма 1000 из экспоненциального распределения с параметром alpha=1

n=1000

alpha=1

sample=rexp(n,alpha)

# Построим гистограмму:

hist(sample, col = "red", main="Гистограмма экспоненциального

распределения с.в. и ее график плотности")

# Построим график плотности экспоненциального распределения и совместим его с гистограммой:

curve(dexp(x,alpha)\*n, 0,5, lw="3", col="black", add=TRUE)

#Получим выборочную оценку параметра распределения alpha

1/mean(part)

#Смоделируем выборку объема 1000 из распределения Стьюдента с параметром k=5:

n=1000

k=5

sample=rt(n,k)

#Построим гистограмму

hist(sample,col="blue", main="Гистограмма распределения Стьюдента с.в. и ее график плотности")

#Построим график плотности распределения Стьюдента и совместим его с гистограммой:

curve(dt(x,k)\*n,-6,6,lw="3",col="black",add=TRUE)

#Сформируем выборку из ген. совокупности:

part=sample[900:1000]

#Получение выборочной оценки для k

s=sd(part)

(2\*s\*s)/(s\*s-1)

#Получим выборку из равномерного распределения

#объем выборки

n=1000

#Зададим интервал (-4;6)

#параметры

a=1

b=9

#выборка

sample=runif(n,a,b)

#Построим гистограмму

hist(sample, main="Гистограмма равномерного распределения с.в. и ее график плотности", col="yellow")

# Построим график плотности равномерного распределения (с выбранными параметрами) и совместим его с гистограммой:

curve(dunif(x,a,b)\*n,-1,10,lw="3", col="black", add=TRUE)

#Сформируем выборку из ген. совокупности:

part=sample[900:1000]

#Вычисление оценок

mean(part)+sqrt(3)\*sd(part)

mean(part)-sqrt(3)\*sd(part)