Министерство науки и высшего образования Российской Федерации

Федеральное государственное бюджетное образовательное учреждение

высшего образования

**ИРКУТСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ**

Институт информационных технологий и анализа данных

наименование института

ОТЧЕТ  
к лабораторной работе по дисциплине

Моделирование процессов и систем

«Программные средства статистического анализа»

Наименование темы

Выполнил студент группы ИСМб 19–1 Вовиков Д.Е.

шифр Фамилия И.О.

Проверил Бучнев О.С.

Фамилия И.О.

Содержание отчета на 7 стр.

Иркутск 2021 г.

Содержание

[1 Введение 3](#_Toc66545860)

[2 Ход выполнения лабораторной работы 4](#_Toc66545861)

[Вывод 9](#_Toc66545862)

[Список использованных источников 10](#_Toc66545863)

1 Введение

**Цель работы**

1. Ознакомление с программными средствами статистического анализа и обработки наблюдений;
2. Повторение базовых понятий теории вероятностей и математической статистки.

**Вариант индивидуального задания 6**

2 Ход выполнения лабораторной работы

Выборка из генеральной совокупности согласно 6 варианту. Выборка приведена для использования на языке R и записана в переменную x

x<-c(-0.087902,0.013424,0.098910,0.656652,-0.854133,-3.365868,0.325508,-2.459858,-2.718851,0.128654,7.761721,0.589181,0.130496,-0.014648,0.950373,15.252800,29.097190,0.072971,0.129118,0.908724,1.245623,0.412579,-1.377286,-1.646605,-0.385762,-0.137182,3.577536,0.732370,-0.796749,0.199767,0.871844,0.996863,-0.104056,-0.223416,1.806323,0.175750,-0.168530,0.285408,0.338004,0.429857,-0.542026,-0.068744,-1.764305,-0.482814,0.124820,1.009103,39.486750,2.048678,-1.222162,0.108109,-0.715133,-0.512997,-0.755282,-1.183587,0.160644,0.171939,-3.632062,0.463528,6.400114,10.852150,-1.576288,-0.374543,-0.287290,0.029972,0.305377,0.140491,0.945222,0.432636,0.054965,0.350172,-0.180790,1.772126,-0.265042,-0.280971,0.329976,-0.376820,-0.531885,-0.303357,3.744286,-0.228976,-1.145483,0.221435,-1.462524,-0.261592,-0.559662,-1.800746,-0.541298,-2.646112,0.499927,-17.764750,-1.053972,0.436956,1.481638,-1.133699,0.361200,-0.329416,0.923937,0.193188,3.884292,0.495872)

Используя команду построения гистограммы hist() построена гистограмма частот (рисунок 2.1).

Для построения гистограммы был использован следующий код:

par(mfrow=c(1, 2))

p1<-hist(x, breaks = 6, freq = FALSE, col = "lightgreen")

lines(density(x),col="red")

p1

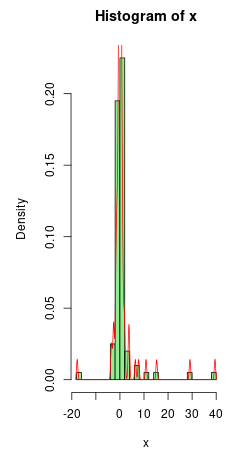


Рисунок 2.1 -- Гистограмма частот

**Случайная величина** – это величина, которая в результате опыта может принять какое-либо значение, какое именно мы не знаем. Закон распределения нужен для связи случайной величины и вероятности ее выпадения.

**Гистограмма частот** – столбчатая фигура, показывающая сколько значений попадает в каждый под интервал.

**Плотность распределения** – связывает возможные значения случайной величины и вероятности их выпадения (оценивается по гистограмме частот).

Построив гистограмму, используем команду получения гистограммы построенных частот cumsum() (рисунок 2.2).

Для построения гистограммы был использован следующий код:

p1$counts <- cumsum(p1$counts)

plot(p1,col = "red")

p1

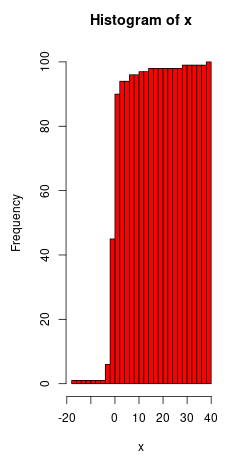


Рисунок 2.2 -- Гистограмма накопленных частот

**Гистограмма накопленных частот** – столбчатая фигура, показывающая сколько значений попадает в каждый под интервал и соседний.

**Функцией распределения случайной величины** X называется функция F(x), определяющая для каждого значения x вероятность того, что случайная величина X примет значение меньшее, чем x, то есть. F(x) = P(X < x).

Оценки основных числовых характеристик получим с использованием следующих команд mean(), var(), sd() (рисунок 2.3).

Для построения числовых характеристик был использован следующий код:

m\_x=mean(x)

m\_x

disp=var(x)

disp

sqo=sd(x)

sqo

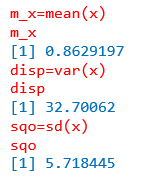


Рисунок 2.3 – Оценки основных числовых характеристик

**Оценка** – любое значение искомого параметра, вычисленное на основе ограниченного числа опытов, всегда будет содержать элемент случайности.

**Среднее квадратическое отклонение** равно квадратному корню из среднего квадрата отклонений отдельных значений признака от средней арифметической.

**Дисперсия** - представляет собой средний квадрат отклонений индивидуальных значений признака от их средней величины.

**Стандартное отклонение** - оценка среднеквадратического отклонения случайной величины *x* относительно её математического ожидания на основе несмещённой оценки её дисперсии

Для получения информации о доверительных интервалах подключим пакет DescTools, и вызовем команды MeanCI(), VarCI()(Рисунок 2.4).

Для получения информации о доверительных интервалах был использован следующий код:

#install.packages("DescTools")

library("DescTools")

MeanCI( x,conf.level = 0.95)

VarCI(x,conf.level = 0.95)

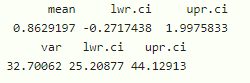


Рисунок 2.4 – Доверительные интервалы

**Доверительный интервал** представляет собой диапазон, для которого можно утверждать, с заданной вероятностью называемой степенью доверия, что он будет содержать оцениваемый параметр, входящий в генеральную совокупность.

Выдвинем гипотезу о том, что выборка извлечена из генеральной совокупности с законом распределения вероятностей Коши. Для проверки этой гипотезы посчитаем фактические частоты:

k<-50 # число интервалов

int<-seq(min(x),max(x),(max(x)-min(x))/k) #интервалы для расчета частот

x.fact<-hist(x,breaks=int,plot=FALSE)

Затем посчитаем теоретические частоты:

int[1]<-(-Inf)

int[k+1]<-(Inf) #границы

x.theor<-pcauchy(int,location = 0, scale = 0.5)

Затем выполним проверку гипотезы:

x.theor<-(x.theor[2:(k+1)]-x.theor[1:k])

chisq.test(x.fact$counts,p=x.theor, simulate.p.value=TRUE)

Результаты проверки гипотезы представлены на рисунке 2.5. Поскольку значение p-value (вероятность того, что расхождение теоретических и эмпирических частот при повторении эксперимента при тех же условиях окажется больше наблюдаемого) больше уровня значимости, гипотезу следует принять.



Рис. 2.5 – Проверка гипотезы о равномерном законе

распределения вероятностей

**Х2 (Критерий согласия Пирсона)** – разница между ожидаемой частотой и наблюдаемой.

**Нулевая гипотеза** заключается в том, что частоты согласованы, то есть фактические данные не противоречат ожидаемым. Альтернативная гипотеза – отклонения в частотах выходят за рамки случайных колебаний, расхождения статистически значимы.



Рис. 2.6 – Таблица ошибок Первого и Второго рода

**Ошибка 1 рода** происходит, когда мы отвергаем нулевую гипотезу (принимаем альтернативную), когда она правильная.

**Ошибка 2 рода** происходит, когда мы не отвергаем нулевую гипотиреозу, когда альтернативная гипотеза верна.

Альтернативная гипотеза (H1) — это единственное утверждение, являющееся логическим отрицанием нулевой гипотезы

Вывод

В ходе выполнения лабораторной работы было произведено ознакомление с программными средствами статистического анализа и обработки наблюдений. Были повторены базовые понятия теории вероятностей и математической статистики.

Список использованных источников

1. Гнеденко Б.В. Курс теории вероятностей. – М.: Высшая школа, 1974.

2. Гмурман В.Е. Теория вероятностей и математическая статистика: Учебное пособие для вузов. - 6-е изд., стер. - М.: Высшая школа, 1998. - 478с., граф.

3. Вентцель Е. С. Теория вероятностей: Учебник для вузов. - 5-е изд., стер. -М.: Высшая школа, 1998. - 575с., ил.

4. Вентцель Е.С. Теория случайных процессов и ее инженерные при-ложения : учебное пособие / Е. С. Вентцель, Л. А. Овчаров. — 5-е изд., стер. — М. : КНОРУС, 2013. — 448 с.

5. Венэбльз У. Н., Смит Д. М. и Рабочая группа разработки R. Введе-ние в R. Заметки по R: среда программирования для анализа дан-ных и графики. Версия 2.15.0 (2012-03-30)