Министерство образования Республики Беларусь

Учреждение образования

«Брестский государственный технический университет»

Кафедра ИИТ

Лабораторная работа №7

за 7 семестр

По дисциплине: «КМиАД»

Выполнил:

Студент 4 курса

Группы ПО-4

Елисеев С.Г.

Проверил:

Чичурин А. В.

2022

**Статистический анализ данных**

Wolfram Language объединяет многие аспекты статистического анализа данных, от получения и изучения данных до построения высококачественных моделей и вывода последствий.

Wolfram Language предоставляет несколько способов получения данных, начиная со встроенных кураторских источников данных, импорта из различных форматов файлов или подключения к базам данных. Базовая обработка данных, включая вычисление статистических величин, сглаживание, тестирование и визуализацию, дает первый уровень анализа.

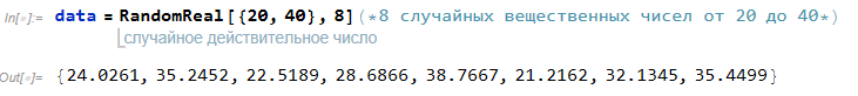
Добавляя в микс модели, такие как модели распределения или регрессии, можно ответить на более широкий круг вопросов анализа или даже предоставить возможности прогнозирования.

1. **Описательная статистика**

Описательные статистические функции Wolfram Language работают как на явных данных, так и на символьных представлениях статистических распределений.

При работе с явными данными функции регулярно обрабатывают огромные наборы данных, которые могут содержать не только числа, но и символьные элементы, представляющие, например, параметризованные или неизвестные данные.

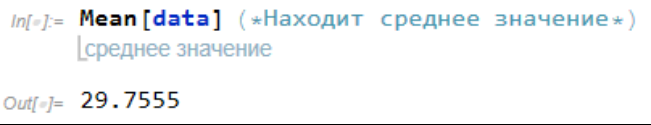
Наиболее часто статистические программы используются для генерации случайных чисел и массивов данных с ними. Для отдельных псевдослучайных реальных чисел предусмотрена функция RandomReal[]. При каждом обращении к этой функции генерируется случайное (точнее, псевдослучайное) число в интервале от 0 до 1 с равномерным распределением. Функция RandomReal[xmin, xmax] генерирует случайное число в заданном интервале изменения переменной x.



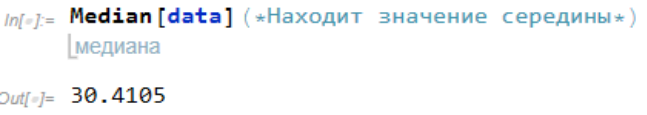
1. Статистика местоположения(**Mean**, **Median**)

Статистика местоположения описывает, где находятся данные. Наиболее распространённые функции включают меры центральной тенденции, такие как среднее и медиана.Quantile [*данные*,*q*] указывает местоположение, перед которым https://reference.wolfram.com/language/tutorial/Files/NumericalOperationsOnData.en/48.png лежит процент данных. Другими словами, Quantile дает такое значение https://reference.wolfram.com/language/tutorial/Files/NumericalOperationsOnData.en/49.png, что вероятность, которая https://reference.wolfram.com/language/tutorial/Files/NumericalOperationsOnData.en/50.png меньше или равнаhttps://reference.wolfram.com/language/tutorial/Files/NumericalOperationsOnData.en/51.png, и вероятность, котораяhttps://reference.wolfram.com/language/tutorial/Files/NumericalOperationsOnData.en/52.png больше или равна https://reference.wolfram.com/language/tutorial/Files/NumericalOperationsOnData.en/53.png.

При наличии списка с n элементами  *среднее* [значение Mean](https://reference.wolfram.com/language/ref/Mean.html)[*list*] определяется как .



*Медиана* [Median](https://reference.wolfram.com/language/ref/Median.html)[*list*] эффективно дает значение в середине отсортированной версии *списка*. Его часто считают более надежной мерой центра распределения, чем среднее, поскольку оно меньше зависит от отдаленных значений.



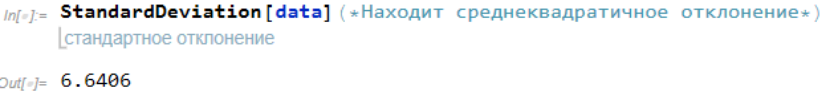
1. Статистика дисперсии(**Variance**, **StandartDeviation**)

Статистика дисперсии обобщает разброс или распространение данных. Большинство из этих функций описывают отклонение от определенного местоположения. Например, дисперсия является мерой отклонения от среднего значения, а стандартное отклонение — это просто квадратный корень дисперсии.

*Дисперсия* [Variance](https://reference.wolfram.com/language/ref/Variance.html)[*list*] определяется как для реальных данных.

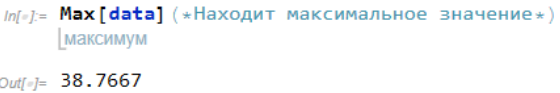


*Стандартное отклонение* [StandardDeviation](https://reference.wolfram.com/language/ref/StandardDeviation.html)[*list*] определяется как. 

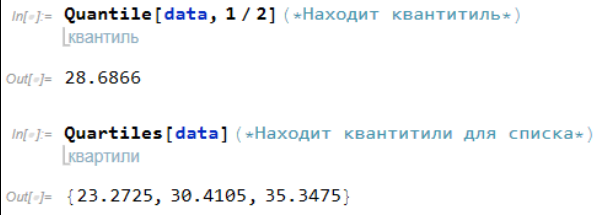


Если элементы в списке считаются выбранными случайным образом в соответствии с некоторым распределением вероятностей, то среднее дает оценку того, где находится центр распределения, в то время как стандартное отклонение дает оценку того, насколько широка дисперсия в распределении.

1. Статистика заказов(**Max**, **Quantile**, **Quartiles**)



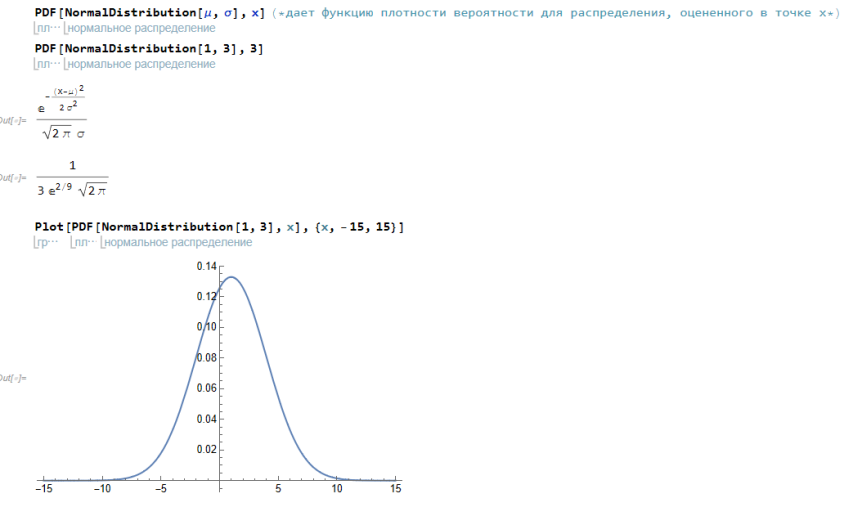
Тем https://reference.wolfram.com/language/tutorial/Files/NumericalOperationsOnData.en/7.png Й *квантиль* Quantile [*список]*,*q*]эффективно дает значение пути https://reference.wolfram.com/language/tutorial/Files/NumericalOperationsOnData.en/8.png через отсортированную версию *списка*.

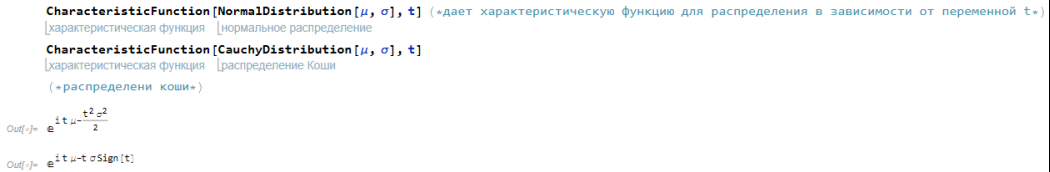


1. Символические особенности дистрибутивов(**PDF**, **CharacteristFunction**)

Дистрибутивы представлены в символической форме.

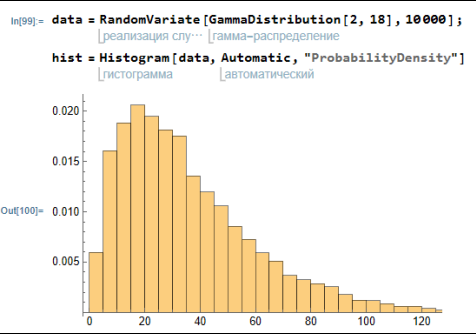
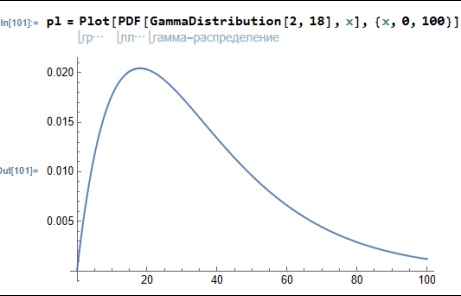
PDF[dist,x]вычисляет массовую функцию при x, если x является числовым значением, и в противном случае оставляет функцию в символьной форме, когда это возможно.

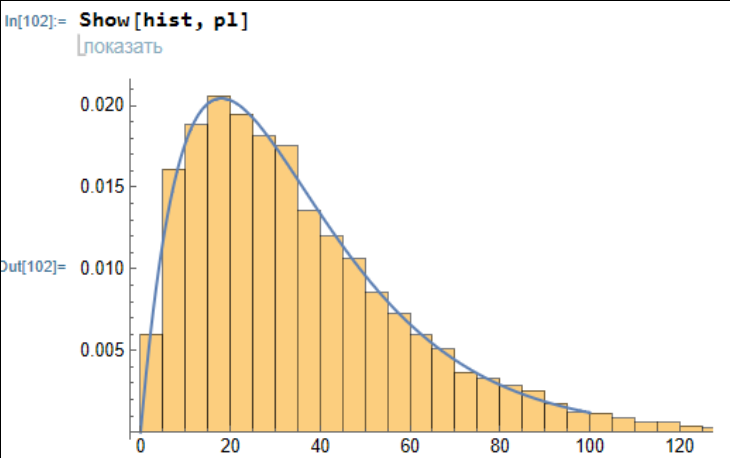


**CharacteristFunction**

1. Статистическая визуализация(**Histogram&Plot**)

Массивы данных часто представляются гистограммами по столбцам или строками. Если массив одномерный — его представляют столбиковыми гистограммами. При этом данные по горизонтали разбиваются на N участков по числу столбцов (или строк) будущей гистограммы. В каждом из частичных участков подсчитывается сумма данных — число, и оно определяет высоту (или длину) столбца. Таким образом, проводится первичная статистическая обработка данных.



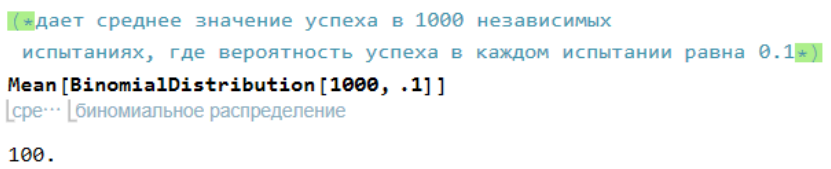
1. **Дискретные распределения**

Функции, описанные здесь, являются одними из наиболее часто используемых дискретных одномерных статистических распределений. Можно вычислить их плотности, средние значения, дисперсии и другие связанные свойства. Сами дистрибутивы представлены в символической форме name[param 1,param2,...]. Такие функции, как Mean, которые дают свойства статистических распределений, принимают символьное представление распределения в качестве аргумента.

Большинство общих дискретных статистических распределений можно понять, рассмотрев последовательность испытаний, каждое из которых имеет два возможных исхода, например, успех и неудачу.

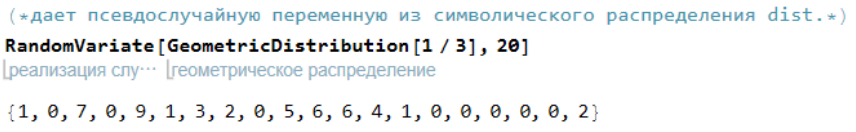
1. BinomialDitribution

Биномиальное распределение[n,p]— распределение числа успехов, которые происходят в n независимых испытаниях, где вероятность успеха в каждом испытании равна p.



1. GeometricDistribution

Геометрическое распределение[p] — это распределение общего числа испытаний до первого успеха, где вероятность успеха в каждом испытании равна p.

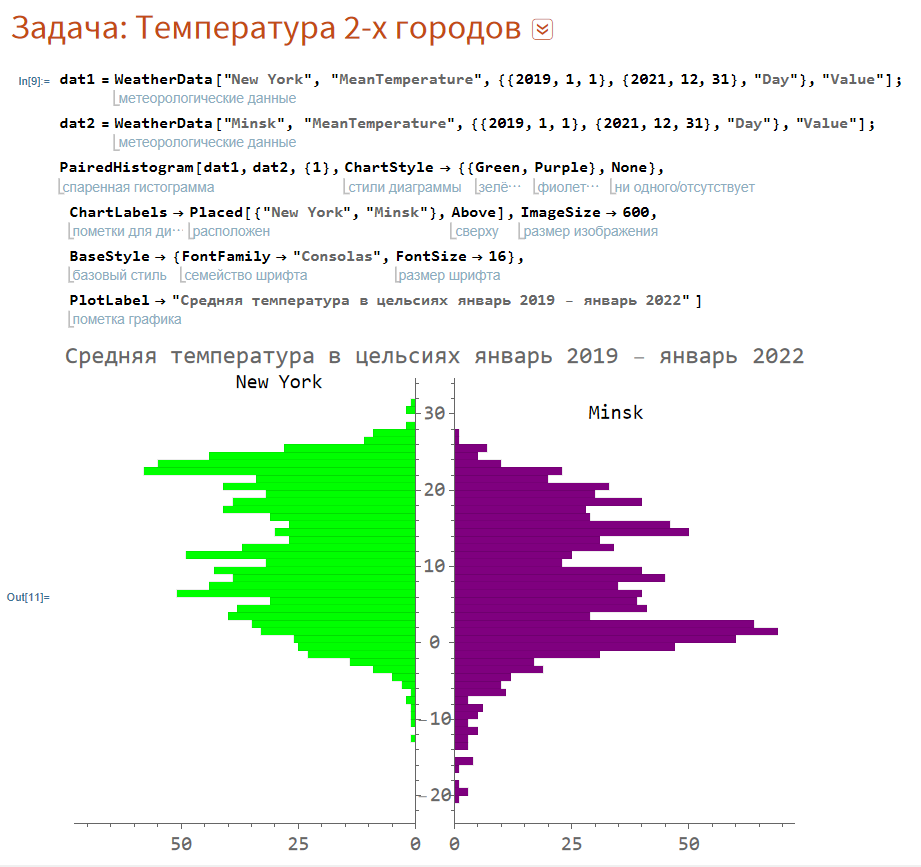


1. **Задача\_1**

Сравнение данных о средней температуре в двух городах — New York и Minsk.

В этом случае удобнее оказались сдвоенные горизонтальные гистограммы.

Левая гистограмма относится к одному городу, правая — к другому.



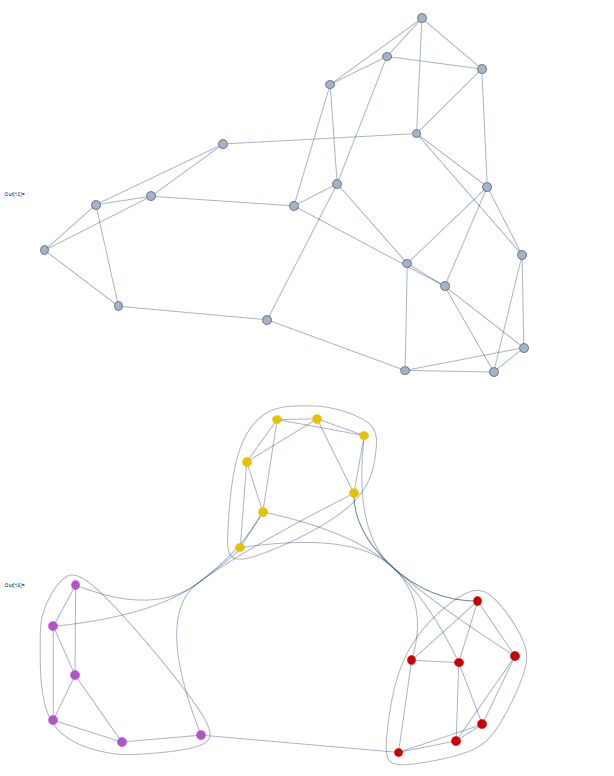
1. **Задача\_2**

Данные часто группируются в некоторых областях, объединяющих их по определенным признакам. Такие области данных называют кластерами.

На результате ниже показана кластеризация рандомно сгенерированного графа. Выявление кластеров в данном случае ,например, разбиение клиентов  
всемирной сети на сообщества, имеющие сходные интересы и имеющие близкое географическое положение, может увеличить производительность сервисов в сети; нахождение сообществ покупателей со схожими интересами и анализ покупок, совершенных ими, позволяет эффективно строить рекомендации; кластеризация больших графов позволяет строить эффективную визуализацию и быстро находить пути в графе.

RandomGraph[WattsStrogatzGraphDistribution[20,0.2,2]]

CommunityGraphPlot[%,FindGraphCommunities[%]]



**Вывод:**

Изучил теоретический материал и провёл статистический анализ данных в системе Wolfram Mathematica.