

Министерство образования и науки РФ
Федеральное государственное автономное образовательное
учреждения высшего образования
«Национальный исследовательский технологический университет «МИСИС»
Кафедра бизнес-информатики и систем управления производством

ПРАКТИЧЕСКАЯ РАБОТА № 2

НА ТЕМУ:

«Описательная статистика. Графический анализ»

по дисциплине

«Статистический анализ данных в принятии решений»

Направления подготовки:
38.03.05 Бизнес-информатика
Семестр 4

Москва – 2026

Оглавление

Установка библиотеки	3
1. Линейный график	3
2. График рассеяния	4
3. Гистограмма	6
4. Гистограмма накопленной частоты	8

Установка библиотеки

Сначала убедитесь, что нужная библиотека установлена. Если нет, то сделайте это с помощью команды:

```
import pandas as pd
import math
import matplotlib.pyplot as plt
```

Рисунок 1 – Импорт библиотек

1. Линейный график

Линейный график – это специальный инструмент, который помогает нам визуально представлять и понимать данные. Представьте, что у вас есть много информации, и вы хотите показать, как она меняется с течением времени или в зависимости от других факторов. Линейный график делает это очень простым и понятным!

На графике есть две оси:

1. **Горизонтальная (ось X)** показывает, как меняется время или другой важный фактор. Например, по оси X можно отложить дни недели, месяцы, периоды.
2. **Вертикальная (ось Y)** – она показывает, что мы измеряем. Например, это может быть количество конфет, баллы на тесте или температура.

Для чего нужен линейный график?

Линейные графики помогают нам быстро увидеть, как данные меняются.

Например, можно заметить, что вы читали больше книг в течение третьей недели. Или линейный график покажет, что у вас был небольшой спад в чтении на четвертой неделе.

Примеры применения линейного графика

1. **Температура:** для отслеживания температуры каждый день вы записываете, сколько градусов было. Отобразив значения на графике, вы получаете представление, как температура колебалась.
2. **Финансовые расходы:** если вы ведете учет своих расходов на продукты каждую неделю, линейный график поможет увидеть, растут ли ваши расходы или наоборот, уменьшаются.
3. **Спорт:** если вы занимаетесь спортом, например пробежками, вы можете записывать, сколько километров вы пробегаете каждую неделю. На графике будет видно, как вы становитесь все сильнее с каждой неделей.

Как строить график в Python:

```
# Строим линейный график
plt.figure(figsize=(24,12))
x = df['Дата']
y = df['Цена']
plt.plot(x, y, marker='o')
plt.title('Линейный график')
plt.xlabel('X')
plt.ylabel('Y')
plt.grid(True)
plt.show()
```

Рисунок 2 – Построение линейного графика

2. График рассеяния

График рассеяния – это способ показать, как две переменных связаны друг с другом. Например, что вы исследуете, как погода влияет на количество мороженого, которое люди покупают.

На графике рассеяния у нас есть две оси:

- Одна ось (обычно горизонтальная) показывает одно значение. Например, это может быть температура воздуха в градусах Цельсия.
- Другая ось (обычно вертикальная) показывает другое значение. Например, это может быть количество мороженого, проданного в день.

Каждая точка на графике представляет собой пару значений: одно значение из первой оси и одно из второй.

Смотрим на точки. Когда вы посмотрите на точки, можете увидеть, что они располагаются где-то на графике. Если вы заметите, что точки образуют прямую линию (или кривую), это значит, что есть связь между температурой и количеством проданного мороженого. Например, чем выше температура, тем больше мороженого покупают.

Какие бывают зависимости:

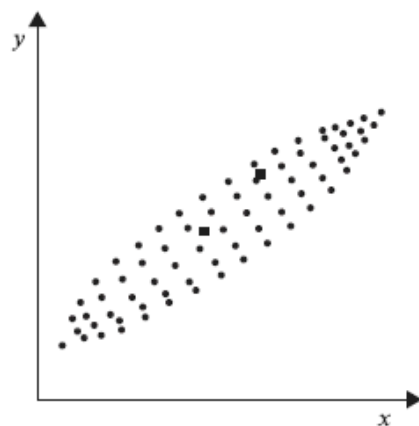
1. **Прямая зависимость:** если точки на графике образуют примерно прямую линию, это значит, что, когда одно значение увеличивается, другое тоже увеличивается. Например,

если мы рисуем точки для роста детей и их возрастов, скорее всего, мы увидим, что чем старше ребенок, тем выше он становится.

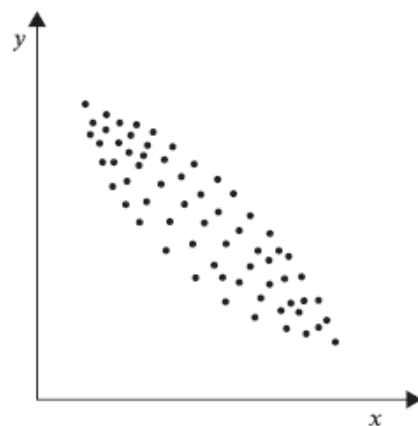
2. **Обратная зависимость:** если на графике точки идут вниз, как будто рисуем обратную линию, это значит, что, когда одно значение увеличивается, другое уменьшается. Например, если мы смотрим на количество игрушек у ребенка и его время, проведенное на улице, мы можем увидеть, что чем больше игрушек у него, тем меньше времени он проводит на улице.

3. **Отсутствие зависимости:** если точки на графике расположены случайным образом, это означает, что между переменными нет явной связи. Например, если мы посмотрим на количество облаков на небе и количество конфет, которые съели дети, смотрящие на эти облака, скорее всего, мы не увидим никакой связи.

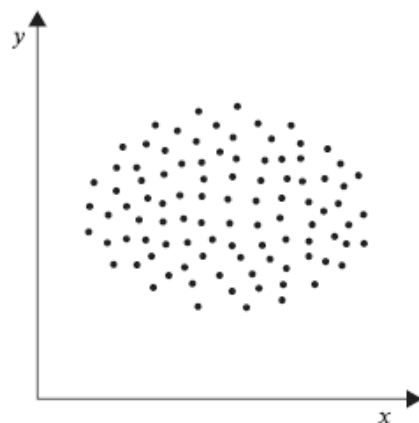
4. **Криволинейная зависимость:** иногда точки образуют какую-то кривую. Это значит, что взаимосвязь между переменными сложнее, чем просто прямая или обратная линия. Например, если мы нарисуем точки для возраста собаки и ее скорости, то можем увидеть, что собаки быстрее, когда они молодые, а потом их скорость уменьшается с возрастом.



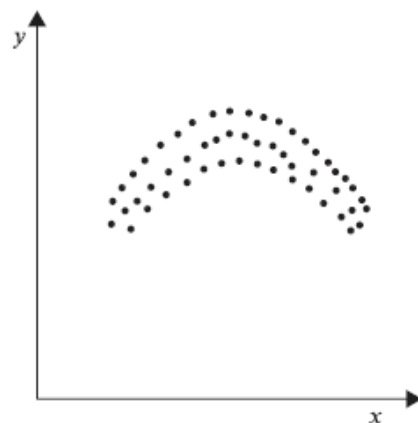
а. Прямая корреляция



б. Отрицательная корреляция



в. Отсутствие корреляции



г. Криволинейная корреляция

Рисунок 3 – Виды зависимостей

Примеры применения графиков рассеяния

1. **Связь роста и веса.** Если мы хотим понять, как рост детей влияет на их вес, мы тоже можем построить график рассеяния. На одной оси будет рост (в сантиметрах), а на другой – вес (в килограммах). Каждая точка будет представлять одного ребенка. Возможно, окажется, что чем выше ребенок, тем он тяжелее. Это значит, что рост и вес связаны.
2. **Связь между учебным временем и оценками.** Если вы собираете данные о времени, которое ученики тратят на уроки, и их оценках, вы можете построить график. Если вы увидите, что ученики, которые учатся дольше, получают более высокие оценки, это будет говорить о том, что есть связь между учебным временем и успехом в школе.
3. **Продажа игрушек и время года.** На графике рассеяния можно показать, как время года влияет на продажу игрушек. Например, у вас будет ось с месяцами (зима, весна, лето, осень) и ось с количеством проданных игрушек. Это поможет понять, в какое время года игрушки продаются лучше всего.

Как строить график в Python:

```
# Строим график рассеяния
plt.figure(figsize=(24,12))
x = df['Количество']
y = df['Цена']
plt.scatter(x, y, color='blue')
plt.title('График рассеяния')
plt.xlabel('X')
plt.ylabel('Y')
plt.grid(True)
plt.show()
```

Рисунок 4 – Построение графика рассеяния

3. Гистограмма

Гистограмма – это специальный график, который состоит из прямоугольников (столбцов), высота которых показывает, сколько значений попадает в разные группы или

диапазоны. Эти группы называются "интервалами". Гистограмма помогает нам увидеть, насколько часто возникают разные значения в данных.

Гистограммы помогают нам:

- сравнивать данные: можно быстро увидеть, в каком интервале значение встречается чаще всего;
- оценивать распределение: можно понять, есть ли в данных какие-то отклонения (например, если большинство людей имеют средний рост, а у кого-то он очень большой);
- идентифицировать тренды: если мы собираем данные о покупках из года в год, то, посмотрев на гистограмму, можно понять, изменились ли интересы и предпочтения.

Примеры применения гистограмм

1. **Оценка успеваемости в школе:** учитель может собрать оценки своих учеников за контрольную и построить гистограмму, чтобы увидеть, сколько учеников получили низкие оценки. Это поможет определить, с какой темой нужно поработать больше.
2. **Изучение погоды:** метеоролог может построить гистограмму, показывающую, сколько дней в месяце были разные температуры. Это поможет увидеть, было ли в месяце много холодных или жарких дней.
3. **Спорт:** тренер может анализировать время, за которое спортсмены пробегают дистанцию, и построить гистограмму. Это поможет понять, какой результат у большинства спортсменов и как они улучшились со временем.

Как определить количество групп в гистограмме?

Определить количество групп в гистограмме можно с помощью метода Сторджеса, ниже приведена формула подсчета:

$$k = 1 + (3.322 \times \lg N),$$

где: N – количество наблюдений; k – количество групп.

Как строить график в Python:

```
# Строим гистограмму
x = 1 + (3.322 * math.log(37))
print(x)
plt.hist(df['Цена'], bins= int(x), color='orange', edgecolor='black')
plt.title('Гистограмма')
plt.xlabel('Значения')
plt.ylabel('Частота')
plt.grid(True)
plt.show()
```

Рисунок 5 – Построение гистограммы

НО стоит понимать, что значение `bins` может применять только целые значения поэтому необходимо округлить число с помощью метода `int()`.

4. Гистограмма накопленной частоты

Гистограмма накопленной частоты – это инструмент, который помогает понять, как часто встречаются разные значения в наборе данных. Она позволяет нам увидеть, сколько значений находится ниже или равно определённому числу.

Что такое гистограмма накопленной частоты

Представьте, что вы провели опрос среди своих друзей о том, сколько книг они прочитали за месяц. У вас есть такие данные:

- 1 книга – 2 друга
- 2 книги – 3 друга
- 3 книги – 5 друзей
- 4 книги – 4 друга
- 5 книг – 1 друг

Теперь вы можете создать гистограмму:

1. обозначим количество книг на горизонтальной оси;
2. обозначим количество друзей на вертикальной оси;
3. для каждой категории (количество книг) нарисуем столбик, который будет показывать, сколько друзей прочитали это количество книг.

Теперь получаем накопленную частоту. Мы считаем, сколько друзей прочитали:

- 1 или меньше книг (2 друга)
- 2 или меньше книг ($2 + 3 = 5$ друзей)
- 3 или меньше книг ($2 + 3 + 5 = 10$ друзей)
- 4 или меньше книг ($2 + 3 + 5 + 4 = 14$ друзей)
- 5 или меньше книг ($2 + 3 + 5 + 4 + 1 = 15$ друзей)

Далее мы можем нарисовать гистограмму накопленной частоты:

- после 1 книги - 2
- после 2 книг - 5
- после 3 книг - 10
- после 4 книг - 14
- после 5 книг - 15

Каждый столбик будет подниматься до значения, показывающего количество друзей, прочитавших определённое количество книг или меньше.

Зачем это нужно?

Гистограмма накопленной частоты помогает нам быстро увидеть, как распределены данные. Например, если ты видишь, что 10 друзей прочитали 3 или меньше книг, это может говорить о том, что большинство твоих друзей читают немного. Это первый шаг к оценки вида распределения данных.

Пример из жизни

Допустим, в классе учатся 20 учеников. Если вы хотите узнать, как много времени ученики тратят на домашние задания, вот какие данные вы собрали:

- 0-1 час – 5 учеников
- 1-2 часа – 8 учеников
- 2-3 часа – 4 ученика
- 3-4 часа – 3 ученика

Для накопленной частоты мы посчитаем:

- 0-1 час – 5
- 0-2 часа – $5 + 8 = 13$
- 0-3 часа – $13 + 4 = 17$
- 0-4 часа – $17 + 3 = 20$

Так мы можем увидеть, что 17 учеников делают домашние задания не более 3 часов.

Как строить график в Python:

```
# Строим гистограмму накопленной частоты
x = 1 + (3.322 * math.log(37))
plt.hist(df['Цена'], bins= int(x), cumulative=True, color='green', edgecolor='black')
plt.title('Гистограмма накопленной частоты')
plt.xlabel('Значения')
plt.ylabel('Накопленная частота')
plt.grid(True)
plt.show()
```

Рисунок 5 – Построение гистограммы с накопленной частотой

Задание

1. Постройте линейный график для каждого показателя, для которого это будет иметь аналитический смысл. Дайте краткий вывод на основе полученного результата.

2. Постройте график рассеяния для каждой пары показателей, для которых это будет иметь аналитический смысл. Дайте краткий вывод на основе полученного результата.

3. Постройте гистограмму для каждого показателя, для которого это будет иметь аналитический смысл. Дайте краткий вывод на основе полученного результата.

3. Постройте гистограмму с накопленной частотой для каждого показателя, для которого это будет иметь аналитический смысл. Дайте краткий вывод на основе полученного результата.