

**МИНИСТЕРСТВО ЗДРАВООХРАНЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное учреждение
«Национальный медицинский исследовательский центр имени В.А. Алмазова»
ИНСТИТУТ МЕДИЦИНСКОГО ОБРАЗОВАНИЯ**

Кафедра математики и естественнонаучных дисциплин

ОДОБРЕНО
Учебно-методическим советом
ФГБУ «НМИЦ им. В.А. Алмазова»
Минздрава России

Протокол № 1/2022
«25» января 2022 г.

УТВЕРЖДАЮ
Директор Института медицинского
образования
ФГБУ «НМИЦ им. В.А. Алмазова»
Минздрава России
Е.В. Пармон
«25» января 2022 г.

В.Н. Леоненко, Н.В. Михайлова

**ВЫЯВЛЕНИЕ СТАТИСТИЧЕСКОЙ СВЯЗИ
МЕЖДУ КОЛИЧЕСТВЕННЫМИ ПЕРЕМЕННЫМИ**

учебно-методическое пособие для обучающихся
по дисциплине «Основы биostatистики»
по направлению подготовки 06.04.01 Биология (уровень магистратуры)
профиль: «Медицинские лабораторные исследования»

УДК – 311
ББК – 60.6

Леоненко В.Н., Михайлова Н.В. Выявление статистической связи между количественными переменными: учебно-методическое пособие / В.Н. Леоненко, Н.В. Михайлова – СПб.: Издательство ... 2021. – 11с.: ил.

Авторы:

Леоненко В.Н. – доцент кафедры математики и естественнонаучных дисциплин ИМО ФГБУ «НМИЦ им. В.А. Алмазова» Минздрава России, к.ф.-м.н.;

Михайлова Н.В. – заведующий кафедрой математики и естественнонаучных дисциплин ИМО ФГБУ «НМИЦ им. В.А. Алмазова» Минздрава России, к.х.н., доцент.

Рецензент:

Ковальчук С.В. – доцент факультета цифровых трансформаций, старший научный сотрудник Национального центра когнитивных разработок ФГАОУ ВО «Национальный исследовательский университет ИТМО», канд.техн.наук.

В учебно-методическом пособии «Выявление статистической связи между количественными переменными» кратко изложены основные статистические методы, предназначенные для установления связи между количественными переменными.

Данное пособие направлено на развитие навыков сравнительного анализа данных, полученных в результате проведения экспериментальных исследований биологических объектов, у обучающихся по направлению подготовки 06.04.01 Биология (уровень магистратуры) профиль «Медицинские лабораторные исследования» при изучении дисциплины Основы биostatистики.

Обсуждено на заседании кафедры математики и естественнонаучных дисциплин
Института медицинского образования
ФГБУ «НМИЦ им. В.А. Алмазова» Минздрава России

ОГЛАВЛЕНИЕ

Введение.....	4
1. Коэффициент корреляции Пирсона.....	6
2. Задачи.....	10
3. Вопросы для самоподготовки.....	10
Библиографический список.....	11

ВВЕДЕНИЕ

Продолжительность изучения раздела «Выявление статистической связи между количественными переменными» – 11 часов

Из них:

Лекции – 2 часов

Семинары – 2 часа

Практические занятия – 2 часа

Самостоятельная работа – 5 часов

Цель изучения раздела:

- сформировать навыки первичного анализа связей между количественными переменными;
- изучить способы оценки меры монотонной и линейной связи между переменными;
- сформировать навыки обоснования гипотезы о наличии связи между переменными и обоснования её корректности.

Актуальность блока: данный блок вносит вклад в формирование следующих компетенций:

УК-1, ОПК-6, ОПК-8, ПК-5.

Конкретные задачи

Название компетенции	Обучающийся должен знать	Обучающийся должен уметь	Обучающийся должен владеть
УК-1. Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	- классификацию методов сбора и анализа информации	- пользоваться учебной и научной литературой, интернет-источниками	навыками обработки и представления полученных результатов научного исследования
ОПК-6. Способен творчески применять и модифицировать современные	- классификацию методов сбора и анализа информации в	- использовать стандартные панели инструментов в	- навыками использования программного обеспечения для

компьютерные технологии, работать с профессиональными базами данных, профессионально оформлять и представлять результаты новых разработок	пакете STATISTICA	пакете STATISTICA	решения конкретных задач
ОПК-6 Способен творчески применять и модифицировать современные компьютерные технологии, работать с профессиональными базами данных, профессионально оформлять и представлять результаты новых разработок	- основные методы сбора и анализа информации, основные окна в пакете STATISTICA	- использовать стандартные панели инструментов в пакете STATISTICA	- современными компьютерными технологиями в работе с профессиональным и базами данных
ОПК-8. Способен использовать современную исследовательскую аппаратуру и вычислительную технику для решения инновационных задач в профессиональной деятельности	- классификацию методов автоматизации сбора и анализа информации, выявления связи между количественными переменными, сравнения групп по качественным признакам с использованием пакетов программ	- пользоваться основными командами пакетов программ Excel и STATISTICA для статистического анализа данных	- навыками использования программного обеспечения для решения конкретных задач
ПК-5. Способен осуществлять научные исследования в области клеточной биологии, несет ответственность за качество работ и научную достоверность результатов	классификацию типов исходных данных (количественных, качественных)	корректировать ошибки в исходных данных	- навыками использования программного обеспечения для решения конкретных задач

Место проведения занятий: компьютерный класс.

Межпредметные и внутрипредметные связи: для изучения данного блока необходимы знания, умения и навыки, приобретённые в процессе изучения дисциплин:

1. 06.03.01 Биология (уровень бакалавриата);
2. 19.03.01 Биотехнология (уровень бакалавриата).
3. Знание основ математического анализа в объеме дисциплин физика, биофизика, математика.

1. Коэффициент корреляции Пирсона

Цель занятия – изучение обучающимися понятия корреляции и способа нахождения значения корреляции Пирсона между двумя переменными интервального типа.

Введение

При проведении исследований в задачах биологии могут использоваться переменные различных типов. Типы переменных обуславливаются характером изучаемых объектов и их параметров, значения которых собираются. Наиболее встречающимися типами переменных являются следующие:

1. Переменные номинального типа. Служат для описания качественных признаков, которые можно выразить словами или символами. Например, индивид может быть «здоров» или

«болен», биологический вид может «присутствовать» или

«отсутствовать» на данной территории. Для таких переменных, как правило, не применяются математические преобразования. Понятие корреляции не применимо.

2. Переменные порядкового типа. Служат для ранжирования (упорядочения) объекта по некоторому признаку, выраженному в условных единицах. Например, размер особи какого-либо вида:

«маленькая», «средняя» и «крупная», или температура окружающей среды: «очень жарко», «жарко», «тепло»,

«холодно», «очень холодно». Для поиска монотонной связи между такими переменными может использоваться ранговая корреляция по Спирмену. Задача поиска линейной связи между переменными порядкового типа является некорректной и не может быть поставлена.

3. Переменные интервального типа. Принимают числовые значения из некоторого промежутка. Например, рост человека (формально

любое значение от 0 до 250 см), фертильность биологического вида в данных условиях обитания (формально от 0 до некоторого большого числа N). Для переменных данного типа применимы различные преобразования. В частности, возможна оценка силы линейной связи между переменными, для чего применяется понятие корреляции по Пирсону.

Общие сведения

Понятие корреляции по Пирсону используется для установления или отрицания факта связи между двумя переменными интервального типа x и y в линейной форме: $y \approx ax + b$, где a, b – некоторые константы. Если $b \neq 0$, то формула говорит о возможной пропорциональной связи между x и y .

Возможен предварительный анализ связи указанного вида с помощью построения облака точек. Форма облака точек позволяет выдвинуть гипотезу о наличии/отсутствии линейной связи между значениями x и y , а также о величине этой связи. На рисунке 1 приведён пример облака точек, на котором прослеживается линейная связь между x и y .

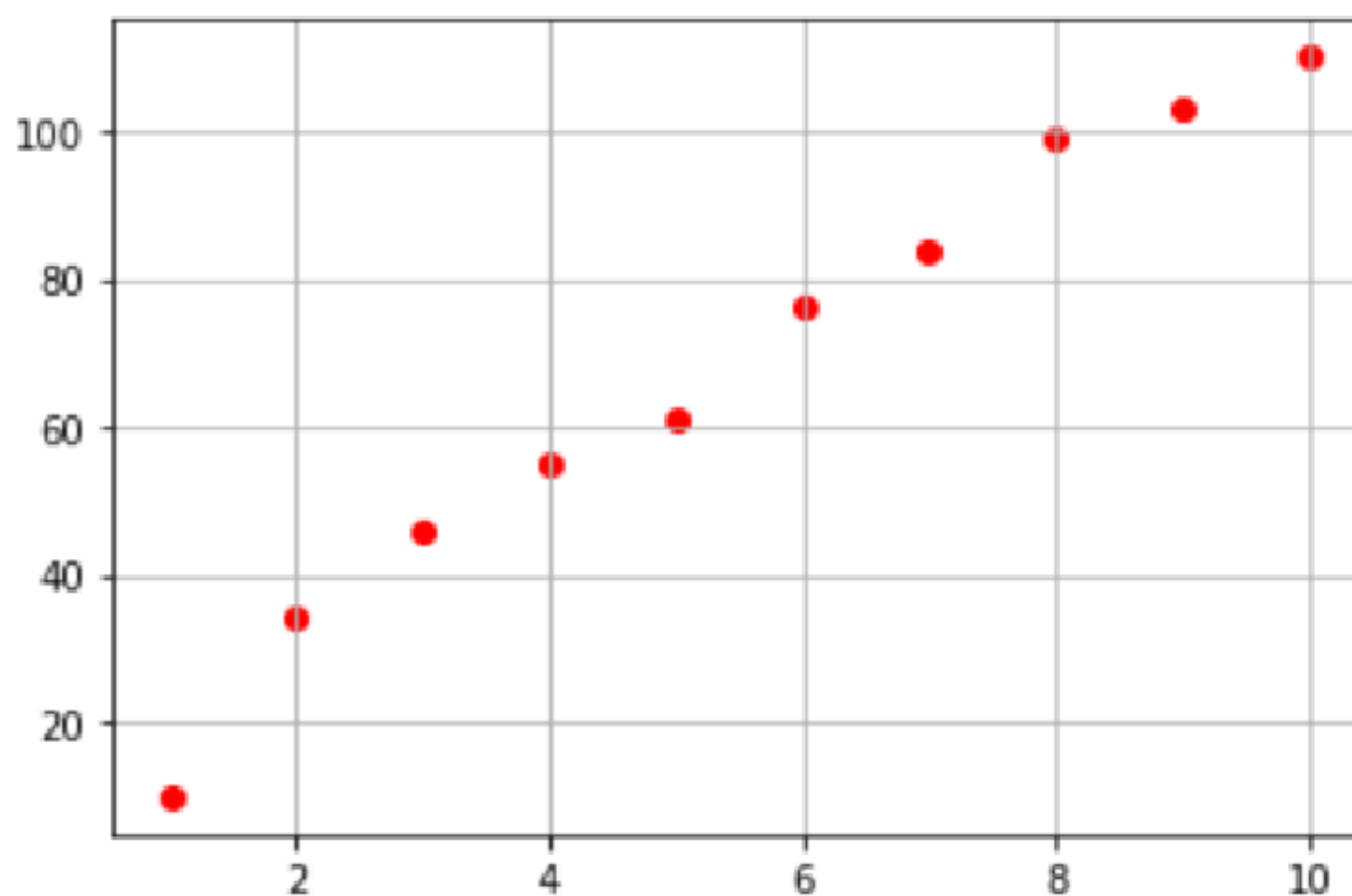


Рис. 1. Пример вида облака точек для переменных x и y интервального типа при наличии сильной линейной связи между переменными.

На рисунке 2, напротив, приведён пример облака точек, на котором линейная связь между переменными не прослеживается.

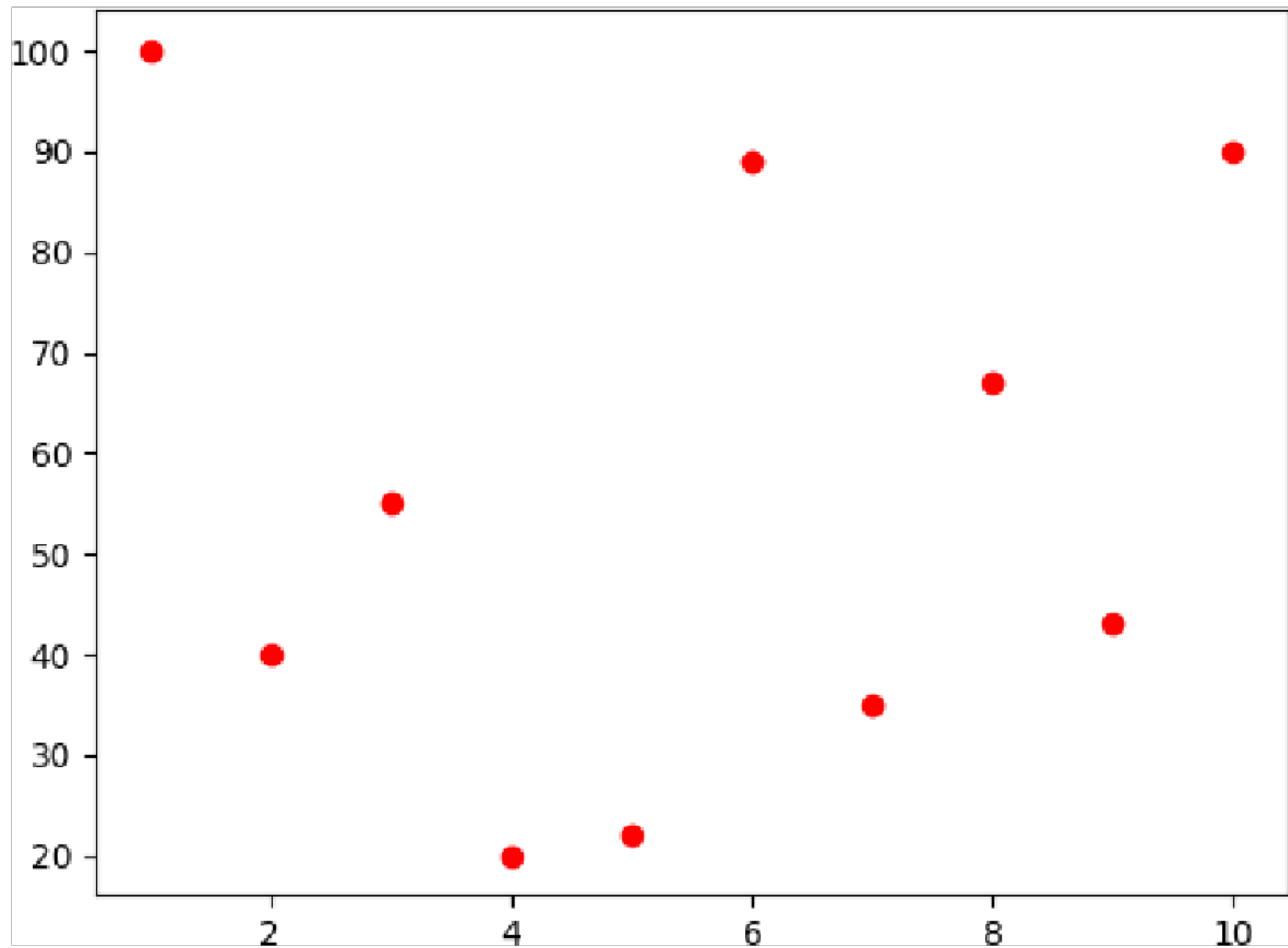


Рис. 2. Пример вида облака точек для переменных x и y интервального типа при отсутствии линейной связи между переменными.

Коэффициент корреляции по Пирсону может быть вычислен по формуле

$$r_{xy} = \frac{E(xy) - E(x) \cdot E(y)}{\sigma(x) \cdot \sigma(y)},$$

где $E(x)$ – математическое ожидание переменной x ,

$\sigma(x)$ – стандартное отклонение переменной x .

Коэффициент r_{xy} изменяется от -1 до 1 и служит мерой линейной связи между x и y . Если $r_{xy} = 1$ или $r_{xy} = -1$, то $y = ax + b$, где a и b – некоторые константы, т.е. y однозначно определяется линейной функцией от x (см. пример на рисунке 3).

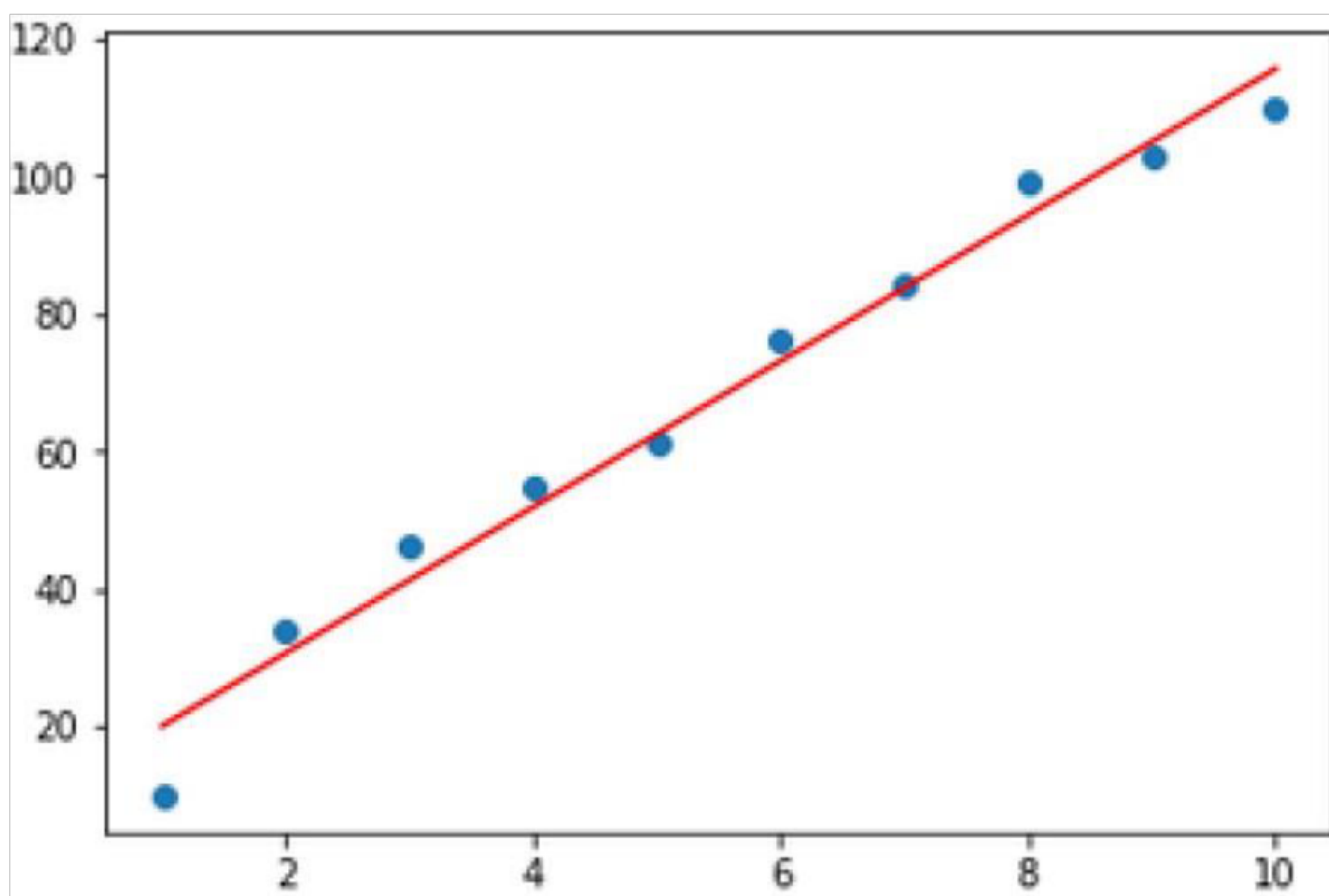


Рис. 3. Линейная функция, определяющая связь между переменными x и y интервального типа.

Если $r_{xy} = 0$, то отсутствует влияние переменной x на переменную y в линейной форме. В этом случае говорят, что переменные не коррелированы. При этом данная зависимость может существовать в другой форме.

В практических исследованиях теоретическое значение r_{xy} не известно и его оценивают на основе выборок значений x и y . Выборочный коэффициент корреляции по Пирсону может быть вычислен по формуле:

$$\bar{r}_{xy} = \frac{\sum_{i=1}^n (x_i \cdot y_i) - n\bar{x}\bar{y}}{s_x \cdot s_y},$$

где произведения $x_i y_i$ задаются по значениям выборок.

Следует отметить, что теоретическое значение r_{xy} может быть достаточно точно оценено через выборочный коэффициент корреляции только в случае, когда n велико. В силу этого для выборок не очень большого объёма можно говорить о наличии линейной связи между x и y в случае, когда выборочный коэффициент корреляции близок по модулю к единице (например, принимает значения $-0,9, -0,8, -0,75, 0,75, 0,8, 0,9$).

2. Задачи

- В пакете STATISTICA задать переменные x, w, y с объёмом выборок $n=75$ по следующим формулам:
 - $x = 0,1 * V0 + Vnormal(rnd(1); 1; 0,1)$
 - $w = 4 + 10 * rnd(1)$
 - $y = 5 + 2 * x + Vnormal(rnd(1); 0; 2)$
- Построить облака точек для пар значений переменных x и w , x и y . С помощью графического анализа выдвинуть гипотезы о наличии или отсутствии линейной связи между указанными парами переменных.
- Найти выборочный коэффициент корреляции по Пирсону между парами переменных (x, w) и (x, y) с помощью пункта меню «Парные и частные корреляции» («Анализ» → «Основные статистики и таблицы»). Проверить справедливость выдвинутых ранее гипотез.

3. Вопросы для самоподготовки

1. В чём заключается понятие корреляции по Пирсону? В какой форме ищется связь между переменными?
2. Приведите формулу теоретического коэффициента корреляции между двумя переменными x и y интервального типа. Объясните смысл каждой переменной формулы.
3. Приведите формулу выборочного коэффициента корреляции. При каких условиях значение теоретического коэффициента корреляции может быть оценено через значение выборочного коэффициента корреляции?

Библиографический список

1. Здоровоохранение и общественное здоровье: учебник/под ред. Г.Н. Царик – М.: ГЭОТАР-Медиа, 2018. – Текст: электронный//URL: <https://www.rosmedlib.ru/book/ISBN9785970443279.html>
2. Информатика и медицинская статистика/под ред. Г.Н. Царик – М.: ГЭОТАР-Медиа, 2017. – Текст: электронный//URL: <https://www.rosmedlib.ru/book/ISBN9785970442432.html>
3. Информационные технологии в профессиональной деятельности: учебник/В.П. Омельченко, А.А. Демидова. – Москва: ГЭОТАР-Медиа, 2020. – Текст: электронный//URL: <https://www.rosmedlib.ru/book/ISBN9785970454992.html>
4. Медицинская информатика: учебник/В.П. Омельченко, А.А. Демидова. - М.: ГЭОТАР-Медиа, 2016. – Текст: электронный//URL: <https://www.rosmedlib.ru/book/ISBN9785970436455.html>
5. Медицинская информатика: учебник/под общ. ред. Т.В. Зарубиной, Б.А. Кобринского. – М.: ГЭОТАР-Медиа, 2016. – Текст: электронный//URL: <https://www.rosmedlib.ru/book/ISBN9785970436899.html>
6. Основы высшей математики и математической статистики/И.В. Павлушков и др. – М.: ГЭОТАР-Медиа, 2012. – Текст: электронный//URL: <https://www.rosmedlib.ru/book/ISBN9785970415771.html>
7. Основы статистического анализа в медицине: Учебное пособие/под ред. проф., д.м.н. В.А. Решетникова / ООО «Издательство «Медицинское информационное агентство», 2020. – <https://www.medlib.ru/library/library/books/36720>

