

Практическая работа №12

Корреляция и линейная регрессия

Цель: научиться вычислять коэффициент корреляции Пирсона, строить диаграмму рассеивания, вычислять коэффициент ранговой корреляции Спирмена, определять значимость полученных коэффициентов, научиться использовать возможности R для вычисления коэффициентов корреляции, делать прогнозы с использованием линейной регрессии, научиться использовать R для проверки статистических гипотез.

Задания

Часть 1. Корреляция

1. Определить два вектора, представляющие собой число автомобилей, припаркованных в течении 5 рабочих дней у бизнес-центра на уличной стоянке и в подземном гараже.

День	Улица	Гараж
Понедельник	80	100
Вторник	98	82
Среда	75	105
Четверг	91	89
Пятница	78	102

Найти и интерпретировать корреляцию между переменными «Улица» и «Гараж» (подсчитать корреляцию по Пирсону).

2. Построить диаграмму рассеяния.
3. Пусть количество автомобилей, припаркованных у аэропорта Шереметьево в течении 15 дней исследуемого месяца, распределено следующим образом:

День	Многоуровневая парковка	Открытая парковка	VIP-парковка
1	3915	3701	745
2	3833	3734	747
3	3987	3745	749

4	3912	3747	701
5	3989	3750	750
6	3975	3699	723
7	3945	3712	734
8	3996	3744	750
9	3956	3745	749
10	3977	3750	740
11	3982	3723	739
12	3999	3733	741
13	3989	3728	743
14	3983	3747	744
15	3988	3750	749

Ввести данные в Excel и сохранить под именем parkovki.csv.

- Загрузить данные в таблицу tab1.
- Вычислить коэффициент корреляции между количеством автомобилей, припаркованных на многоуровневой парковке и на открытой парковке. Проанализировать полученный результат.
- Провести корреляционный анализ для всей матрицы данных.
- Воспользуйтесь пакетом **ellipse** для визуализации корреляционной матрицы в виде эллипсов.
- Заполнить таблицу по результатам, сделанных расчетов:

Корреляция между	Коэффициент корреляции Пирсона	p-значение	Вывод (значима/незначима корреляция)
Многоуровневая парковка			
Открытая парковка			
VIP-парковка			

- Дана таблица данных, состоящая из двух переменных: в первом столбце указаны баллы, полученные на вступительном экзамене по специальности в

аспирантуру университета, во втором столбце – на выпускном экзамене по специальности. Ввести данные в таблицу в Excel, сохранить их в формате .csv с разделителями запятые.

Тест на вступительном экзамене	Тест на выпускном экзамене
67	95
63	100
71	102
81	103
82	106
74	104
74	106
80	107
84	109
80	110
81	111
85	110
80	112
90	111
94	117

10. Импортировать данные опроса в переменную **tab2**.
11. Подсчитать коэффициент корреляции **Спирмена**.
12. Вывести полную статистику расчетов, используя функцию **cor.test** и метод (**method**) корреляции Спирмена.
13. Вычислить корреляцию Кенделла (**method="kendall"**)

Часть 2. Линейная регрессия

1. Создать 2 переменные: **iqr** (IQ родителей) и **iqc** (IQ детей) со следующими значениями:

$$\text{iqr} = (126, 119, 111, 105, 106, 96, 95, 90, 81, 76)$$

$$\text{iqc} = (110, 106, 95, 126, 120, 106, 74, 95, 90, 81)$$
2. Создать таблицу IQ, состоящую из значений двух этих переменных.
3. Получить уравнение регрессии.
4. Построить диаграмму рассеяния.

5. Вычислить коэффициент корреляции **Пирсона**.
6. Вывести уравнение регрессии, связывающее IQ родителей и детей.

Контрольные вопросы:

1. Дайте определение статистической, корреляционной зависимости.
2. Какой анализ называется однофакторным?
3. В чем основная задача однофакторной линейной регрессии?
4. Какой основной метод нахождения прямой линии, ближе всего расположенной к точкам корреляционного поля?
5. Какая формула проверяет правильность подсчета параметров уравнения регрессии?
6. Какая функция используется для посчета матрицы взаимных корреляций всех переменных между собой? Как округлить результат до определенного количества цифр после запятой?
7. Дайте определение коэффициента корреляции Пирсона.
8. Начертите график сильной отрицательной связи с помощью эллипса.
9. Как рассчитать коэффициенты корреляции Спирмена, Пирсона, Кенделла?
10. Как определить значимость коэффициента корреляции?
11. Перечислите способы ввода данных?
12. Какой функцией необходимо воспользоваться, чтобы ввести данные с клавиатуры?
13. Как интерпретируются результаты сравнения по критерию Стьюдента?

Отчет по практической работе №12

1. Титульный лист: название работы, вариант, ФИО, учебная группа.
2. Оглавление.
3. Отчет оформлять на каждый пункт задания, указывая следующее:
 - 3.1. Задание;
 - 3.2. Решение с кодом;
 - 3.3. Результат;
 - 3.4. Вывод.
4. Заполнить таблицу по результатам, сделанных расчетов:

Корреляция между	Коэффициент корреляции Пирсона	p-значение	Вывод (значима/незначима корреляция)
Многоуровневая парковка			
Открытая парковка			
VIP-парковка			

Домашнее задание

1. Руководству обувной фабрики необходима информация о том, сколько пар обуви различного фасона, цвета и другим характеристикам будет продано. Нельзя допускать перепроизводства, чтобы не продавать обувь по заниженным ценам. Нельзя допускать нехватку, чтобы не выпускать больше изделий, чем было запланировано к определенному сроку. Если знать прогноз через шесть недель после поступления в продажу первого наименования, можно изготавливать к намеченному сроку необходимое число продукции. В таблице приведены данные о суммарном объеме проданной обуви каждого фасона, размера, цвета и данные за шесть недель, полученные в начале исследования. Представьте графически данные и получите оценки наименьших квадратов в модели $y=b_0+b_1x$.

Продано в первые шесть недель	Продано всего
234	393
121	190
33	73
196	306
120	201
126	184
178	290
124	192
158	313
75	141
141	236
126	189
88	141
138	254
71	97
63	95

2. Выполните задание из пункта 8.8. теоретического материала и заполните таблицу по результатам расчетов:

Мера	Группа1	Группа2
Среднее		

t-статистика		
Число степеней свободы		
p -уровень значимости		
Доверительный интервал		

