МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«СЕВЕРО-КАВКАЗСКИЙ ФЕДЕРАЛЬНЫЙ УНИВЕРСИТЕТ»

Пятигорский институт (филиал) СКФУ

Методические указания

по выполнению практических работ
по дисциплине «Основы экспериментальных исследований»
для студентов направления подготовки 13.03.02 Электроэнергетика и электротехника Передача и распределение электрической энергии в системах электроснабжения

(ЭЛЕКТРОННЫЙ ДОКУМЕНТ)

СОДЕРЖАНИЕ

1. ПРЕДИСЛОВИЕ……………………………………………………………………………3
2. ОПИСАНИЕ ПРАКТИЧЕСКИХ ЗАНЯТИЙ...4
2.1 Практическое занятие №1...4
2.2 Практическое занятие №2...8
2.3 Практическое занятие №3...15
2.4 Практическое занятие №4...4
2.5 Практическое занятие №5...4
2.6 Практическое занятие №6...8
2.7 Практическое занятие №7...15
2.8 Практическое занятие №8...46
2.9 Практическое занятие №9...4
2.10 Практическое занятие №10...8
2.11 Практическое занятие №11...15
2.12 Практическое занятие №12...4
2.13 Практическое занятие №13...4
2.14 Практическое занятие №14...8
2.15 Практическое занятие №15...15
2.16 Практическое занятие №16...46
3. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ 56
3.1. Рекомендуемая литература...56
3.2. Интернет-ресурсы: 56

ПРЕДИСЛОВИЕ

Формирование набора общенаучных, профессиональных компетенций будущего бакалавра по направлению подготовки 13.03.02 «Электроэнергетика и электротехника», фундаментальных знаний в области теории эксперимента.

Основными задачами изучения дисциплины являются:

− освоение основных принципов статистической экспериментальных данных;

− получение навыков расчетов основных статистических характеристик результатов экспериментов, анализа временных рядов и прогнозирования, пользования методамифакторного, кластерного анализа, многомерного шкалирования;

− изучение правил выбор основных факторов эксперимента и построения факторных планов;

− освоение современных программных средств, позволяющих автоматизировать процесс обработки экспериментальных данных.

Определим вводные требования к изучению данной дисциплины. В результате освоения дисциплины обучающийся должен:

Знать:

-основные приемы организации инженерного эксперимента

-статистические методы обработки результатов эксперимента

- методы планирования эксперимента и регрессионный анализ

- требования по оформлению результатов научных исследований

Уметь:

- формулировать задачу экспериментального исследования технического объекта

-проводить научные эксперименты, оценивать результаты выполненной работы

-планировать и ставить задачи научного исследования

-использовать приобретенные навыки для самостоятельной обработки полученных результатов с использованием известных методов корреляционного и регрессионного анализа

Владеть:

- навыками оценки погрешностей математических моделей

- навыками выбора методов экспериментальной работы

- навыками построения математических моделей

- способностью интерпретировать и представлять результаты научных исследований

1. ОПИСАНИЕ ПРАКТИЧЕСКИХ ЗАНЯТИЙ

Практическое занятие №1

Тема занятия. Вариационные ряды и их характеристики

Цель занятия. Приобрести навыки оценки основных показателей качества объекта по экспериментальным данным.

Знания и умения, приобретаемые студентом в результате освоения темы, формируемые компетенции. Знает методы оценки показателей. Умеет выполнять оценку показателей качества объекта по результатам экспериментов. Владеет навыками анализа и оценки показателей качества.

Актуальность темы. Методы оценки качества применяется при решении многих прикладных задач.

Теоретическая часть

Каждый студент разрабатывает свой вариант программной процедуры в соответствии с индивидуальным заданием (таблица 1.1). Можно использовать любой язык программирования. При оценке заданных свойств датчика следует использовать одни и те же участки последовательности чисел.

При построении каждого вида гистограмм необходимо исследовать несколько значений объема выборки, например: минимальной – длина выборки должна обеспечить попадание в каждый разряд не менее 7 – 10 реализаций; средней – каждый разряд попадает несколько десятков реализаций; большой – несколько сотен попаданий в каждый разряд.

Построение одномерной гистограммы осуществляется обычным образом. Для построения двумерной гистограммы последовательно выбираются различные пары случайных чисел: первое число из каждой пары соответствует координате X, а второе число - координате Y . Аналогично строится трехмерная гистограмма. Целесообразно выбрать одинаковые по длине разряды.

Проверка равномерности выполняется с помощью критерия «хи-квадрат»:

a) значение хи-квадрат рассчитывается по каждой сформированной гистограмме по формуле (1.1)

$$\chi^2 = \sum_{i=1}^{\psi} \frac{(n_i - F_i)^2}{F_i}, \qquad (1.1)$$

где ψ – количество разрядов гистограммы;

n_i – фактическое количество реализаций в i-м разряде гистограммы;

F_i – теоретическое количество реализаций для i-го разряда гистограммы.

b) в соответствии с числом степеней свободы $\psi - 1$, заданным уровнем значимости 0,05 по таблице распределения критерия находится его критическое значение, сопоставляется это значение с рассчитанным значением и делается вывод о равномерности распределения.

Вместо таблиц распределения хи-квадрат можно воспользоваться соответствующей функцией из состава табличного процессора Excel. Функция позволяют найти уровень значимости по значению хи-квадрат и числу степеней свободы. В этом случае сравнению подлежат вычисленный и заданный уровень значимости.

Вычисление коэффициента автокорреляции производится по формуле (1.2)

$$r = \frac{\sum_{i=1}^{n}(x_i - 0{,}5)(x_{i+k} - 0{,}5)}{n * m_2}, \qquad (1.2)$$

где x_i, x_{i+k} – соответственно первое и второе число выборочной пары;

k – постоянный параметр, выбирается из таблицы 1;

n – количество пар чисел, используемых для расчета коэффициента;

m_2 – выборочная дисперсия совокупности сформированной последовательности чисел.

В качестве первых чисел выборочной пары выступают числа формируемой ДСЧ последовательности. В качестве вторых чисел – числа выборки, отстоящие относительно первых на k (один, два, три или четыре) шагов, в соответствии с конкретным вариантом задания, например при k = 2 суммируются парные произведения $x_1\, x_3$, $x_2\, x_4$, $x_3\, x_5$ и т. д.

Расчет необходимо произвести для выборок, полученных в пункте 3.1.

Выборочные моменты рассчитываются непосредственно по исходной выборке. Теоретические значения первых четырех моментов определяются студентами самостоятельно. Для расчета выборочных моментов можно воспользоваться формулами для нахождения начальных и центральных моментов:

Начальный момент порядка k случайной величины X

$$\alpha_k(X) = M(X^k) = \sum_{i=1}^{n} x_i^k p_i$$

Центральный момент порядка k случайной величины X

$$\mu_k(X) = M\left((X - M(X))^k\right)$$

Кроме того, необходимо вычислить первый и второй момент по одномерной гистограмме.

M = интеграл от 0 до 1 x по dx;

D = интеграл от 0 до 1(x-0.5)^2 по dx

Для выборочных значений математического ожидания следует проверить, попадает ли оно в заданный доверительный интервал уровня 0.95:

$$0{,}5 - d < \alpha_1 < 0{,}5 + d\,,$$

где 0,5 – теоретическое значение математического ожидания случайной величины;

d – доверительный интервал уровня.

Доверительный интервал определяется выражением (1.3):

$$d = \frac{t\sqrt{m_2}}{\sqrt{n}}, \qquad (1.3)$$

где t – квантиль функции распределения Лапласа (значение аргумента, при котором значение функции по условию задачи равно $\frac{0,95}{2}$, т. е. равно 0,475). Для указанного уровня квантиль равна 1,96;

m_2 – выборочная дисперсия;

n – число элементов выборки.

Задание для самостоятельного решения

Разработать программную процедуру для экспериментального исследования ДСЧ

- на равномерность и случайность формируемых чисел,
- соответствие выборочных моментов распределения теоретическим значениям.

1. Проверку равномерности провести на основе построения одномерной, двухмерной и трехмерной гистограмм с различными объемами выборок случайных чисел.
2. Проверку независимости случайных чисел провести на основе оценки коэффициента автокорреляции различных пар чисел последовательности.
3. Оценить выборочные значения математического ожидания, второго, третьего и четвертого центральных моментов и сопоставить их с теоретическими значениями.

Таблица 1.1

Варианты индивидуальных заданий

N пп	Количество разрядов гистограмм	Значение параметра k для расчета коэффициента автокорреляции
1	7, 7*7, 7*7*7	4
2	8, 8*8, 8*8*8	4
3	9, 9*9, 9*9*9	4
4	10, 10*10, 10*10*10	1
5	11, 11*11, 11*11*11	2
6	10, 9*9, 9*9*9	2
7	10, 11*11, 11*11*11	2
8	9, 12*12, 10*10*10	2
9	12, 10*10, 10*10*10	3
10	11, 8*8, 9*9*9	3
11	7, 10*10, 12*12*12	3
12	8, 8*8, 12*12*12	3
13	8, 10*10, 9*9*9	1
14	9, 8*8, 10*10*10	1
15	9, 11*11, 10*10*10	1
16	10, 8*8, 9*9*9	2
17	12, 7*7, 8*8*8	2
18	12, 9*9, 9*9*9	2
19	12, 8*8, 7*8*7	3
20	7, 7*7, 7*7*7	4
21	8, 8*8, 8*8*8	4
22	7, 12*12, 11*11*11	3
23	10, 7*8, 10*9*8	3

N пп	Количество разрядов гистограмм	Значение параметра k для расчета коэффициента автокорреляции
24	11, 7*7, 14*14*14	1
25	11, 7*7, 10*10*10	2
26	11, 7*7, 11*11*11	3
27	11, 8*8, 10*10*10	3
28	11, 8*8, 11*11*11	2
29	11, 8*8, 12*12*12	1

Примечание: во второй колонке таблицы первая цифра указывает количество разрядов для одномерной, вторая группа цифр – для двухмерной, а треть группа – для трехмерной гистограммы.

Вопросы для самопроверки

1. Дать определение понятиям: показатель качества, начальный и центральный момент
2. Привести формулы для вычисления показателей

Рекомендуемая литература:

Основная литература:

1. Мелас В.Б., Шпилев П.В. Планирование и анализ для регрессионных моделей: учеб. пособие.- СПб.:Изд-во С.-Перерб. ун-та, 2014.- 96 с.
2. Мусина О.Н. Планирование и постановка научного эксперимента: учебно-методическое пособие / О.Н. Мусина.- М.- Берлин:Директ-Медиа, 2015.- 88 с.
3. Боярский М.В. Планирование и организация эксперимента: учебное пособие / М.В. Боярский, Э.А. Анисимов.- Йошкар-Ола: Поволжский государственный технологический университет, 2015.- 168 с.

Дополнительная литература:

1. Сафин Р.Г. Основы научных исследований. Организация и планирование эксперимента: учебное пособие / Р.Г. Сафин, А.И. Иванов, Н.Ф. Тимербаев; Мин-во образование и науки России, Казан. нац. исслед. технол. ун-т.-Казань: Изд-во КНИТУ, 2013.- 156 с.
2. Костин В.Н. Теория эксперимента: учебное пособие / В.Н. Костин, В.В. паничев; Оренбургский государственный университет – Оренбург: ОГУ, 2013 .- 209 с.

Практическое занятие №2

Тема занятия. Построение вариационных рядов. Расчет числовых характеристик

Цель занятия. Приобрести навыки оценки основных показателей качества объекта по экспериментальным данным.

Знания и умения, приобретаемые студентом в результате освоения темы, формируемые компетенции. Знает методы оценки показателей. Умеет выполнять оценку показателей качества объекта по результатам экспериментов. Владеет навыками анализа и оценки показателей качества.

Актуальность темы. Вариационные ряды применяются при решении многих прикладных задач.

Теоретическая часть

Вариацией называется различие значений признака у отдельных единиц изучаемой совокупности в один и тот же период или момент времени.

Статистический анализ вариации предполагает выполнение следующих основных этапов:

1. Построение вариационного ряда.
2. Графическое изображение вариационного ряда.
3. Расчет показателей центра распределения и структурных характеристик вариационного ряда.
4. Расчет показателей размера и интенсивности вариации.
5. Оценка вариационного ряда на асимметрию и эксцесс (самостоятельно).

Пример. В большом магазине собираются данные о числе продаж, произведённых каждым из 20 продавцов.

16, 17, 6, 12, 24, 19, 9, 10, 13, 15, 17, 14, 20, 14, 16, 16, 21, 18, 18, 22 n = 20 (если x_i— число продаж). В общем виде: x_1, x_2, x_3, …x_n — называются вариантами.

Варианты, расположенные в возрастающем порядке, т.е. ранжированные, составляют вариационный ряд. Если значения вариантов встречаются несколько раз, то вариационный ряд имеет вид:

6, 9, 10, 12, 13, 14, 14, 15, 16, 16, 16, 17, 17, 18, 18, 19, 20, 21, 22, 24

Он характеризует варьирование (изменение числа продаж). Абсолютные числа, показывающие, сколько раз встречается тот или иной вариант, называется частотами и обозначаются m_1, m_2, … $m_к$ (иногда n_1, n_2, … n_k).

Частоты появления значений признака в примере 1:

Значения признака x_i (числа продаж)	6	9	10	12	13	14	15	16	17	18	19	20	21	22	24
Частóты	1	1	1	1	1	2	1	3	2	2	1	1	1	1	1

Здесь $k = 15; \quad \sum_{i=1}^{k} m_i = n; \quad \sum_{i=1}^{15} m_i = 20.$

2. Виды вариации

Вариация признака может быть *дискретной* и *непрерывной*.

Дискретной вариацией называется такая, при которой отдельные значения признака (варианты) отличаются друг от друга на некоторую конечную величину (обычно целое число), т.е. даны в виде прерывных чисел.

Непрерывной называется вариация, при которой значения признака могут отличаться одно от другого на сколь угодно малую величину.

Например: % выполнения рабочим нормы выработки, вес одного семени и т.д.

Пример 2. Менеджер большого универмага записал суммы денег, которые израсходовали 184 покупателя, посетившие отдел верхней одежды в день сезонной распродажи товаров по сниженным ценам. Данные о суммах, израсходованных на покупки, были отгруппированы по следующим категориям: от 0 до 100 рублей, от 100 до 200 рублей и так далее до 600 рублей.

Интервалы и частоты каждого интервала представлены в таблице (для простоты расчётов примем за единицу измерения 1000 рублей). Частоты обозначим латинской буквой m.

Таблица 1

Интервалы	0-100	100-200	200-300	300-400	400-500	500-600
Частоты	30	38	50	31	22	13
Центральное значение интервалов	50	150	250	350	450	550

При непрерывной вариации распределение признака называется интервальным.

Частоты относятся не к отдельному значению признака, а ко всему интервалу. Иногда за значение интервала принимают его середину (центр).

Интервальные вариационные ряды бывают с одинаковыми и неодинаковыми интервалами.

3. Частость (относительная частота или доля)

Пусть мы имеем частоты: m_1, m_2, …, m_k

$$m_1 + m_2 + m_3 + \ldots + m_k = \sum_{i=1}^{i=k} m_i$$

Отношение частоты того или иного варианта (а также интервала) к сумме всех частот называется частостью.

$$\boxed{W = \frac{m_i}{\sum_{i=1}^{k} m_i}} \qquad (1)$$

$$W_1 = \frac{30}{184} \quad W_2 = \frac{38}{184} \quad W_3 = \frac{50}{184} \quad W_4 = \frac{31}{184} \quad W_5 = \frac{22}{184} \quad W_6 = \frac{13}{184}$$

т. е. сумма всех частостей равна 1.

$$\boxed{\sum_{i=1}^{k} W_i = 1} \qquad (2)$$

Замечание: Частости иногда выражают в %. Тогда их сумма =100%.

Частоты и частости называются *весами.*

4. Границы интервалов. Величина интервала

В интервальном вариационном ряду в каждом интервале различают нижнюю и верхнюю границы интервала:

Нижняя граница интервала: $x_{i\,(min)}$

Верхняя граница интервала: $x_{i\,(max)}$

Величина интервала:

$$\boxed{K_i = x_{i\,(max)} - x_{i\,(min)}} \qquad (3)$$

При построении интервальных вариационных рядов в каждый интервал включаются варианты, числовые значения которых больше нижней границы и меньше (или равны) верхней границы (или наоборот)

Для выбора оптимальной величины интервала (при которой вариационный ряд будет не очень громоздким) применяют формулу:

$$K_i = \frac{x_{\max} - x_{\min}}{1 + 3{,}322 \lg n}, \qquad (4)$$

где n - число единиц совокупности.

Разность между экстремальными значениями вариантов называется вариационным размахом:

$$\boxed{R = x_{\max} - x_{\min}} \qquad (5)$$

И тогда

$$\boxed{K_i = \frac{x_{\max} - x_{\min}}{L}}. \qquad (6)$$

Знаменатели формул (4) и (6) – это количества интервалов, которые выбираются по формулам Стерджерса.

Подсчитывая число значений, попавших в полуинтервал $[x_i, x_{i+1})$, получим значения частот $m_i(\overline{i,k})$. В итоге интервальный ряд можно представить таблицей

Интервалы	$[x_1; x_2)$	$[x_2; x_3)$	$\cdots$	$[x_k; x_{k+1})$
Частоты	m_1	m_2	$\cdots$	m_k

Если значение варианта находится на границе интервала, то его присоединяют к правому интервалу.

Иногда при группировке с равными интервалами сначала определяют число интервалов (L) при заданном объёме совокупности, пользуясь формулой:

L= 2 × *ln* n (7)

И тогда

$$\boxed{K_i = \frac{R}{L}} \qquad (8)$$

Замечание: число интервалов должно быть на единицу больше размаха вариации, деленного на ширину интервала $\left(L = \frac{R}{k_i} + 1\right)$.

8.6. Плотность вариационного ряда (или плотность распределения)

Плотность вариационного ряда — это отношение частот (или частостей) к величине интервала.

Различают абсолютную плотность:

$$\boxed{f_a(i) = \frac{m_i}{k_i}} \tag{9}$$

и относительную плотность распределения

$$\boxed{f_0(i) = \frac{w_i}{k_i}} \tag{10}$$

Пример:из таблицы 1:

$$f_a(3) = \frac{50}{300-200} = 0,5$$

$$f_0(3) = \frac{50/184}{100} = \frac{1}{368} \approx 0,0027$$

6. Накопленные частоты (частости)

Пусть x - некоторое число. Тогда количество вариантов m_x, значения которых меньше x, называется *накопленной частотой*

$$\boxed{m_x = \sum_{x_i < x} m_i} \tag{11}$$

Отношение накопленной частоты к общему числу наблюдений n называется *накопленной частостью*

$$\boxed{W_x = \frac{m_x}{n} = \frac{1}{n}\sum_{x_i < x} m_i} \tag{12}$$

Накопленные частоты можно получить в *восходящем* порядке, и в *нисходящем*.

Таблица 2

Накопленные частоты для примера 2

Интервал	m_i	Накопленные частоты (ν) в восходящем порядке	Накопленные частоты (ν) в нисходящем порядке

0-100	30	30	184
100-200	38	68	154
200-300	50	118	116
300-400	31	149	66
400-500	22	171	35
500-600	13	184	13
Σ	184	–	–

Рассмотрим ещё один пример нахождения накопленных частот для дискретного вариационного ряда.

Таблица 3

x_i	m_i	Накопленные частоты	
		в восходящем порядке	в нисходящем порядке
2	3	3	10
3	6	9	7
6	1	10	1
Σ	10	–	–

Накопленной частотой (частостью) в нисходящем порядке называется суммарная частота (частость) членов статистического ряда, значения которых меньше x.

7. Графические методы изображения вариационных рядов

Правило “золотого сечения”: график должен быть расположен в прямоугольнике, в котором высота будет относиться к ширине как 5 : 8.

Вариационные ряды графически могут быть изображены в виде полигона, гистограммы, кумуляты и огивы.

Полигон распределения (многоугольник) строится в прямоугольной системе координат. Значения признака откладываются на оси абсцисс, значения частот или частостей – на оси ординат. Чаще всего полигоны применяются для изображения дискретных вариационных рядов. На оси абсцисс отмечаются точки, соответствующие величинам вариантов. Из них восстанавливаются ординаты (перпендикуляры), длина которых соответствует (пропорциональна) численности этих вариантов. Вершины ординат соединяются отрезками. В случаях построения полигона для интервальных рядов ординаты, пропорциональные частоте (или частости) интервала, восстанавливаются перпендикулярно оси абсцисс в точке, соответствующей середине данного интервала. Для замыкания крайние ординаты соединяются с серединой интервалов, в которых частоты или частости равны 0.

Гистограмма распределения строится аналогично полигону в прямоугольной системе координат. В отличие от полигона при построении гистограммы на оси абсцисс берутся не точки, а отрезки, изображающие интервал, а вместо ординат строят прямоугольники с высотой, *пропорциональной частотам, частостям* или плотностям интервалов (в случае неравенства интервалов). Если в ряду с равными интервалами соединить прямыми линиями середины верхних сторон прямоугольников, то получим полигон распределения.

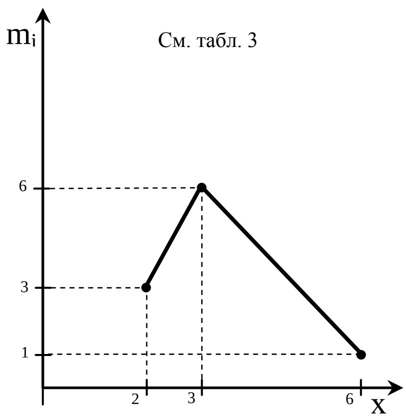

Рис.1

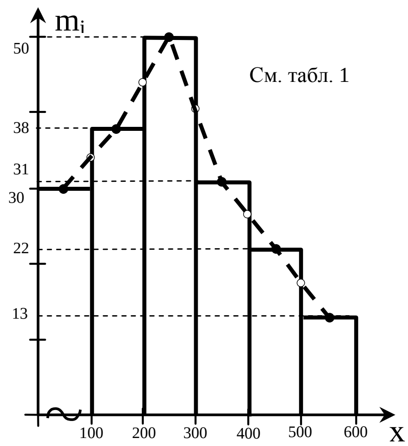

Рис. 2

Кумулятивная кривая (кривая сумм-кумулята) получается при изображении вариационного ряда с накопленными частотами или частостями в прямоугольной системе координат. При построении кумуляты дискретного признака на ось абсцисс наносят значения признака (варианты). Ординатами служат вертикальные отрезки, длина которых пропорциональна накопленной частоте (или частости) какого-либо варианта. Соединяя вершины ординат прямыми линиями, получат ломаную (кривую) кумуляту. При построении кумуляты интервального вариационного ряда нижней границе первого интервала соответствует частота, равная 0, а верхней – вся частота интервала. Верхней границе второго интервала соответствует накопленная частота первых двух интервалов (т.е.) сумма частот этих интервалов. Верхней границе последнего интервала соответствует накопленная частота, равная сумме всех частот.

Огива строится аналогично кумуляте с той лишь разницей, что на ось абсцисс наносят накопленные частоты, а на ось ординат – значения признака.

8. Средняя величина – обобщающая количественная характеристика вариационного ряда

Средняя арифметическая

Простая

$$\boxed{\overline{X}_{np} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}} \tag{13}$$

Взвешенная

$$\boxed{\overline{X}_{взв} = \frac{x_1 m_1 + x_2 m_2 + \ldots + x_k m_k}{n} = \frac{\sum_{i=1}^{k} x_i m_i}{\sum_{i=1}^{k} m_i}} \tag{14}$$

или

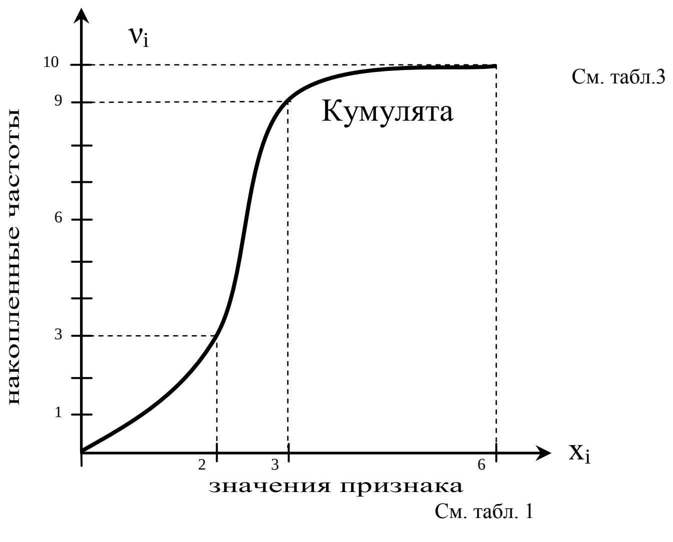

Рис. 8

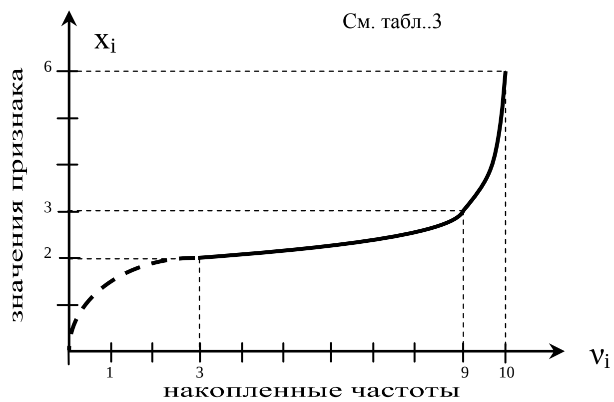

Рис. 4

$$\boxed{\overline{X}_{взв} = \frac{x_1 n_1 + x_2 n_2 + \ldots + x_m n_m}{n} = \frac{\sum_{i=1}^{m} x_i n_i}{\sum_{i=1}^{m} n_i}} \tag{14'}$$

Иногда в литературе m_i обозначают через n_i — это частота!

$$m_1 + m_2 + \ldots + m_k = n = \sum_{i=1}^{k} m_i$$

$$n_1 + n_2 + \ldots + n_k = n$$

или

$$\overline{X} = \frac{\sum xf}{f}$$

Свойства $\overline{\text{X}}$:

1. Средняя арифметическая постоянной величины равна этой постоянной, то есть $\overline{\text{C}} = \text{C}$.

2. Если из всех вариантов вычесть постоянную величину x_0 ($x_i - x_0 = x_i$) и найти из результатов вычитания среднюю $\left(\overline{X}\right)$, то она окажется меньше искомой средней на эту постоянную величину x_0.

$$\boxed{\overline{X'} = \overline{X} - x_0}, \text{ т.е. } \boxed{\overline{X} = \overline{X'} + x_0} \quad (15)$$

3. Если все варианты уменьшить (увеличить) в одно и то же число раз $\left(\frac{x_i}{k} = x_i'\right)$ и из полученных частных вычислить среднюю, то она окажется уменьшенной (увеличенной) в такое же число раз.

$$\boxed{\overline{X'} = \frac{\overline{X}}{k}} \quad \text{или} \quad \boxed{\overline{X} = \overline{X'}k} \quad (16)$$

4. Если в частотах (m_i) имеется общий множитель, то при вычислении средней его можно не принимать во внимание, то есть производить взвешивание по сокращённым частотам. Численное значение средней при этом не изменится.

5. Сумма отклонений вариантов от $\overline{X}$, умноженных на веса (частоты), равна 0.

$$\boxed{\sum_{i=1}^{k}\left(x_i - \overline{X}\right)m_i = 0} \quad (17)$$

или

$$\sum_{i=1}^{m}\left(x_i - \overline{X}\right)n_i = 0$$

6. Средняя суммы (разности) двух или нескольких величин равна сумме (разности)

их средних:

$$\boxed{\overline{X \pm Y} = \overline{X} \pm \overline{Y}}$$

7. Если находят среденюю для интервального вариационного ряда, то в качестве значения признака для каждого интервала условно принимают его середину, то есть центр.

9. Медиана дискретного вариационного ряда

В качестве описательной характеристики вариационного ряда применяется медиана (*Me*), то есть такое значение варьирующего признака, которое приходится на середину упорядоченного ряда. Если в вариационном ряду нечётное число вариантов (2m+1), то медиана вычисляется по формуле:

$$Me = x_{m+1} \tag{18}$$

Если чётное число (2m) вариантов, то:

$$\boxed{Me = \frac{x_m + x_{m+1}}{2}} \tag{19}$$

Замечание. Медиану иногда называют непараметрическим средним значением.

Рассмотрим на примере, как вычисляется *Me* дискретного вариационного ряда.

Пример: *X – возраст людей.*

x_i	8;	9;	11;	12;	15;	16;	18;	19;	21;
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9

Нечётное число вариантов: 2m+1 — формула нечётного числа.

9 =2m+1

2m = 8

m = 4

$$Me = x_{\mathrm{m+1}}$$

$$Me = x_{4+1} = x_5 = 15 \text{ (лет)}$$

Пример: *X – возраст людей.*

x_i	8;	9;	11;	12;	15;	16;	18;	19;	21;	23;	24;	26;
	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}

Чётное число вариантов: 2m — формула чётного числа.

12 = 2m

m = 6

$$Me = \frac{x_m + x_{m+1}}{2} = \frac{x_6 + x_7}{2} = \frac{16+18}{2} = 17 \text{ (лет)}$$

10. Медиана интервального вариационного ряда

$$\boxed{Me = x_{Me(\min)} + k_i \frac{0{,}5 \cdot \sum m - \nu_{Me\text{-}1}}{m_{Me}}}, \qquad (20)$$

где $x_{Me(\min)}$ – нижняя граница медианного интервала; k_i – величина медианного интервала; $\nu_{Me\text{-}1}$ – накопленная частота (или частость) интервала, предшествующего медианному; $0{,}5 \cdot \sum m$ – половина суммы всех частот (или частостей); m_{Me} – частота медианного интервала.

При исчислении медианы интервального вариационного ряда сначала находят интервал, содержащий медиану. Для этого используют накопленные частоты (или частости). Медианному интервалу соответствует первая из накопленных частот (или частостей), превышающая половину всего объёма совокупности.

Пример: *процент выполнения норм выработки (x).*

Интервалы	x_i	m_i	Накопленные частоты
90–100	95	3	3
100–110	105	8	11
110–120	115	7	18
120–130	125	2	20
Σ	–	20	–

Первая из накопленных частот превышает $0{,}5 \cdot \Sigma m_i$, т.е. 10:

$$\frac{\sum m_i}{2} = 0{,}5 \cdot \Sigma m_i, = 10.$$

Значит медианный интервал (100–110):

$x_{Me(\min)} = 100;$ $\nu_{Me\text{-}1} = 3;$

$k = 100 – 100 = 10;$ $m_{Me} = 3;$

$$Me = 100 + 10 \cdot \frac{10 - 3}{8} = 100 + 10 \cdot \frac{7}{8} = 108{,}75$$

$$\boxed{Me = 108{,}75}$$

11. Мода

В математической статистике модой называют вариант, наиболее часто встречающийся в данном вариационном ряду.

Для дискретного ряда мода определяется по наибольшей частоте и соответствует варианту с наибольшей частотой.

Мода для непрерывного (интервального с равными интервалами) ряда исчисляется по формуле:

$$\boxed{Mo = x_{Mo(\min)} + k_i \frac{m_{Mo} - m_{Mo\text{-}1}}{\left(m_{Mo} - m_{Mo\text{-}1}\right) + \left(m_{Mo} - m_{Mo+1}\right)}}, \quad (21)$$

где $x_{Mo(\min)}$ - нижняя граница модального интервала;

m_{Mo} - частота модального интервала;

m_{Mo-1} - частота интервала, предшествующего модальному;

m_{Mo+1} - частота интервала, последующего за модальным;

k_i - величина модального интервала.

Может быть: одна мода – унимодальное распределение;

две моды – бимодальное распределение;

три и более – мультимодальное распределение.

Модальный интервал определяется по набольшей частоте.

Пример:

Интервалы	m_i
90–100	3
100–110	8

110–120	7
120–130	2
Σ	20

Модальный интервал (100–110), т.к. он имеет наибольшую частоту.

$$\boxed{x_{Mo(\min)} = 100}$$

$k = 10$ $m_{Mo-1} = 3;$

$m_{Mo} = 8;$ $m_{Mo+1} = 7;$

$$Mo = 100 + 10\frac{8-3}{(8-3)+(8-7)} \approx 108{,}3$$

$$\boxed{Mo \approx 108{,}3}$$

Показатели колеблемости (вариации) признака

Такие признаки, как заработная плата, профессия, число членов семьи, возраст и т.д. — варьируют.

Для измерения вариации признака математическая статистика применяет ряд показателей.

12. Вариационный размах (R), или широта распределения

$$R = x_{max} - x_{min} \tag{22}$$

применялся в формуле (8.6)

x_{max} — наибольший вариант вариационного ряда.

x_{min} — наименьший вариант вариационного ряда.

R представляет собой величину неустойчивую, зависящую от случайных обстоятельств. Она применяется в качестве приблизительной оценки вариации.

Среднее линейное отклонение

$$\boxed{\rho = \frac{\sum_{i=1}^{n}|x_i - \overline{X}|}{n} \text{ невзвешенное}}$$

$$\rho = \frac{\sum_{i=1}^{k} \left| x_i - \overline{X} \right| m_i}{\sum_{i=1}^{k} m_i} \text{ взвешенное} \qquad (23)$$

13 Дисперсия (средний квадрат отклонения)

$$\sigma^2 = \frac{\sum_{i=1}^{n} \left(x_i - \overline{X} \right)^2}{n} \text{ невзвешенная}$$

$$\sigma^2 = \frac{\sum_{i=1}^{n} \left(x_i - \overline{X} \right)^2 m_i}{\sum_{i=1}^{k} m_i} \text{ взвешенная} \qquad (24)$$

Упрощённая формула дисперсии

$$\sigma^2 = \overline{X^2} - \overline{X}^2, \qquad (25)$$

где

$$\overline{X^2_{np}} = \frac{\sum_{i=1}^{n} x_i^2}{n}; \quad \overline{X^2_{взв}} = \frac{\sum_{i=1}^{k} x_i^2 m_i}{\sum_{i=1}^{k} m_i}$$

14. Среднее квадратическое отклонение (с.к.о.)

$$\boxed{\sigma = +\sqrt{\sigma^2}} \quad (26)$$

15. Коэффициент вариации (υ)

$$\boxed{\upsilon = \frac{\sigma}{\overline{\overline{X}}} \cdot 100\%} \quad (27)$$

Применяется только для признака, принимающего только положительные значения.

Если ν > 40%, то это говорит о большой колеблемости признака в изучаемой совокупности (например большая колеблемость товарооборота в регионе).

$$\upsilon_R = \frac{R}{\overline{\overline{X}}} \cdot 100\%$$ – коэффициент осцилляции

$$\upsilon_\rho = \frac{\rho}{\overline{\overline{X}}} \cdot 100\%$$ – коэффициент вариации по среднему линейному отклонению.

Задание для самостоятельного решения

Разработать программную процедуру для экспериментального исследования ДСЧ

- на равномерность и случайность формируемых чисел,
- соответствие выборочных моментов распределения теоретическим значениям.

4. Проверку равномерности провести на основе построения одномерной, двухмерной и трехмерной гистограмм с различными объемами выборок случайных чисел.
5. Проверку независимости случайных чисел провести на основе оценки коэффициента автокорреляции различных пар чисел последовательности.
6. Оценить выборочные значения математического ожидания, второго, третьего и четвертого центральных моментов и сопоставить их с теоретическими значениями.

Таблица 1.1

Варианты индивидуальных заданий

N пп	Количество разрядов гистограмм	Значение параметра k для расчета коэффициента автокорреляции
1	7, 7*7, 7*7*7	4
2	8, 8*8, 8*8*8	4
3	9, 9*9, 9*9*9	4

N пп	Количество разрядов гистограмм	Значение параметра k для расчета коэффициента автокорреляции
4	10, 10*10, 10*10*10	1
5	11, 11*11, 11*11*11	2
6	10, 9*9, 9*9*9	2
7	10, 11*11, 11*11*11	2
8	9, 12*12, 10*10*10	2
9	12, 10*10, 10*10*10	3
10	11, 8*8, 9*9*9	3
11	7, 10*10, 12*12*12	3
12	8, 8*8, 12*12*12	3
13	8, 10*10, 9*9*9	1
14	9, 8*8, 10*10*10	1
15	9, 11*11, 10*10*10	1
16	10, 8*8, 9*9*9	2
17	12, 7*7, 8*8*8	2
18	12, 9*9, 9*9*9	2
19	12, 8*8, 7*8*7	3
20	7, 7*7, 7*7*7	4
21	8, 8*8, 8*8*8	4
22	7, 12*12, 11*11*11	3
23	10, 7*8, 10*9*8	3
24	11, 7*7, 14*14*14	1
25	11, 7*7, 10*10*10	2
26	11, 7*7, 11*11*11	3
27	11, 8*8, 10*10*10	3
28	11, 8*8, 11*11*11	2
29	11, 8*8, 12*12*12	1

Примечание: во второй колонке таблицы первая цифра указывает количество разрядов для одномерной, вторая группа цифр – для двухмерной, а треть группа – для трехмерной гистограммы.

Вопросы для самопроверки

1. Понятие вариационного ряда
2. Характеристики ряда

Рекомендуемая литература:

Основная литература:

1. Мелас В.Б., Шпилев П.В. Планирование и анализ для регрессионных моделей: учеб. пособие.- СПб.:Изд-во С.-Перерб. ун-та, 2014.- 96 с.
2. Мусина О.Н. Планирование и постановка научного эксперимента: учебно-методическое пособие / О.Н. Мусина.- М.- Берлин:Директ-Медиа, 2015.- 88 с.
3. Боярский М.В. Планирование и организация эксперимента: учебное пособие / М.В. Боярский, Э.А. Анисимов.- Йошкар-Ола: Поволжский государственный

технологический университет, 2015.- 168 с.

Дополнительная литература:

3. Сафин Р.Г. Основы научных исследований. Организация и планирование эксперимента: учебное пособие / Р.Г. Сафин, А.И. Иванов, Н.Ф. Тимербаев; Мин-во образование и науки России, Казан. нац. исслед. технол. ун-т.-Казань: Изд-во КНИТУ, 2013.- 156 с.
4. Костин В.Н. Теория эксперимента: учебное пособие / В.Н. Костин, В.В. паничев; Оренбургский государственный университет – Оренбург: ОГУ, 2013 .- 209 с.

Практическое занятие 3.

Тема занятия. Построение кривой нормального распределения по опытным данным. Проверка гипотезы о нормальном распределении выборки

Цель занятия. Рассмотреть соответствие выборки нормальному закону распределения

Знания и умения, приобретаемые студентом в результате освоения темы, формируемые компетенции. Знает, как проверить соответствие выборки нормальному закону распределения. Владеет способностью к самоорганизации и самообразованию.

Актуальность темы. Методы экспериментальных исследований используются в прикладных задачах.

Теоретическая часть.

Для статистической оценки показателей качества и выполнения сравнительного анализа необходимо знать закон распределения случайной влечены.

Критерий Пирсона. По этому критерию проверяют гипотезу о том, что распределение опытных данных не противоречит теоретическому распределению. В случае выполнения неравенства $\chi^2 < \chi^2_{ДОП}$ гипотеза о подчинении выборки нормальному закону распределения не отвергается.

$$\chi^2 = \sum_{i=1}^{К} \frac{(n_i^э - n_i^T)^2}{n_i^T}$$

где $n_i^э$ – частота попадания экспериментальных данных i-й интервал гистограммы; n_i^T – теоретическая частота попадания экспериментальных данных i-й интервал; К – число интервалов гистограммы распределения

Допустимая величина отклонений $\chi^2_{ДОП}$ (определяется по таблице 2.1) зависит от уровня значимости а (а=0,10) и числа степеней свободы q

$$q = К - r - 1$$

где К – число интервалов гистограммы распределения; r – число параметров закона распределения, для нормального закона распределения r равняется 2 (т.е. $\overline{X}$ и σ)

Проверку гипотезы о подчинении статистического ряда данных нормальному закону распределения проводят следующим образом.

1) Определяют максимальное и минимальное значение в выборке (X_{max}=356 X_{min}=128)
2) Для построения гистограммы распределения определяют длину интервала $D_и$ по формуле Стерджеса:

$$D_и = (X_{max} - X_{min})/(1 + 3{,}32 * \lg N)$$

N – общее число измерений физических величин, определяющее размер выборки

$$D_и = \frac{X_{max} - X_{min}}{1 + 3{,}32 * lgN} = \frac{356 - 128}{1 + 32 * lg80} = 31{,}15 \approx 32$$

3) Разбивают выборку на К интервалов длиной $D_и$. Для каждого интервала устраивают нижнюю и верхнюю границ.

Нижнюю границу 1-го интервала определяют как $X_{i=1}^{н}$, а верхнюю $X_{i=1}^{в} = X_{i=1}^{н} + D_и$
Первый должен включать X_{min} , последний интервал X_{max}

Для последующих интервалов $X_{i=1}^{н} = X_{i=1}^{в}, X_{i=1}^{в} = X_{i=1}^{н} + D_и$

Середину интервала определяют:

$$X_i^{ср} = \frac{X_{i=1}^{н} + X_{i=1}^{в}}{2}$$

4) Рассчитывают экспериментальные частоты попадания результатов измерений в каждый интервал $n_i^э$. При этом значение пренадлежит интервалу, если совпадает с его нижней границей.

5) Вычисляют статистические характеристики выборки с учётом частоты попадания значений в интервал.

Среднее арифметическое значение для середин интервалов

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{К}(X_i^{ср} * n_i^э)$$

Средние квадратичное отклонение

$$\sigma = \sqrt{\frac{\sum_{i=1}^{К}((X_i^{ср} - \overline{X})^2 * n_i^э)}{N}}$$

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{К}(X_i^{ср} * n_i^э) = \frac{17632}{80} = 220{,}4$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{К}((X_i^{ср} - \overline{X})^2 * n_i^э)}{N}} = \sqrt{\frac{170995{,}2}{80}} = 46{,}23$$

6) Вычисляют для каждого интервала значение:

$$t_i = \left|\frac{X_i^{ср} - \overline{X}}{\sigma}\right|$$

7) Устанавливаем значение плотности вероятностей нормального распределения f(ti) (по приложению 1)

8) Устанавливаем теоретические частоты попаданий результатов измерений в i-й интервал по формуле:

$$n_i^{Т} = f(t_i) * \frac{D_и * N}{\sigma}$$

9) Вычисляем значение критерия Пирсонаχ^2 по формуле:

$$\chi^2 = \sum_{i=1}^{К}\frac{(n_i^э - n_i^{Т})^2}{n_i^{Т}} = 6{,}92$$

10) Рассчитываем число степеней свободы:

$$q = К - r - 1 = 8 - 2 - 1 = 5$$

11) По таблице 2.1 определяем допустимое значение $\chi^2_{ДОП}$:
$\chi^2_{ДОП}$ (a=0,10, q=5) = 9,24>6,92 следовательно $\chi^2 < \chi^2_{ДОП}$

Вопросы

1. Нормальный закон распределения
2. Критерий Пирсона

Рекомендуемая литература:

Основная литература:

1. Мелас В.Б., Шпилев П.В. Планирование и анализ для регрессионных моделей: учеб. пособие.- СПб.:Изд-во С.-Перерб. ун-та, 2014.- 96 с.
2. Мусина О.Н. Планирование и постановка научного эксперимента: учебно-методическое пособие / О.Н. Мусина.- М.- Берлин:Директ-Медиа, 2015.- 88 с.
3. Боярский М.В. Планирование и организация эксперимента: учебное пособие / М.В. Боярский, Э.А. Анисимов.- Йошкар-Ола: Поволжский государственный технологический университет, 2015.- 168 с.

Дополнительная литература:

4. Сафин Р.Г. Основы научных исследований. Организация и планирование эксперимента: учебное пособие / Р.Г. Сафин, А.И. Иванов, Н.Ф. Тимербаев; Мин-во образование и науки России, Казан. нац. исслед. технол. ун-т.-Казань: Изд-во КНИТУ, 2013.- 156 с.
5. Костин В.Н. Теория эксперимента: учебное пособие / В.Н. Костин, В.В. Паничев; Оренбургский государственный университет – Оренбург: ОГУ, 2013 .- 209 с.

Практическое занятие 4.

Тема занятия. Проверка соответствия выборки нормальному закону распределения
Цель занятия. Рассмотреть соответствие выборки нормальному закону распределения

Знания и умения, приобретаемые студентом в результате освоения темы, формируемые компетенции. Знает, как проверить соответствие выборки нормальному закону распределения. Владеет способностью к самоорганизации и самообразованию.

Актуальность темы. Методы экспериментальных исследований используются в прикладных задачах.

Теоретическая часть.

Для статистической оценки показателей качества и выполнения сравнительного анализа необходимо знать закон распределения случайной влечены.

Критерий Пирсона. По этому критерию проверяют гипотезу о том, что распределение опытных данных не противоречит теоретическому распределению. В случае выполнения неравенства $\chi^2 < \chi^2_{\text{доп}}$ гипотеза о подчинении выборки нормальному закону распределения не отвергается.

$$\chi^2 = \sum_{i=1}^{\text{К}} \frac{(n_i^{\text{э}} - n_i^T)^2}{n_i^T}$$

где $n_i^{\text{э}}$ – частота попадания экспериментальных данных i-й интервал гистограммы; $n_i^{\text{Т}}$ – теоретическая частота попадания экспериментальных данных i-й интервал; К – число интервалов гистограммы распределения

Допустимая величина отклонений $\chi^2_{\text{доп}}$ (определяется по таблице 2.1) зависит от уровня значимости а (а=0,10) и числа степеней свободы q

$$q = К - r - 1$$

где К – число интервалов гистограммы распределения; r – число параметров закона распределения, для нормального закона распределения r равняется 2 (т.е. $\overline{X}$ и σ)

Проверку гипотезы о подчинении статистического ряда данных нормальному закону распределения проводят следующим образом.

12) Определяют максимальное и минимальное значение в выборке (X_{max}=356 X_{min}=128)

13) Для построения гистограммы распределения определяют длину интервала $D_и$ по формуле Стерджеса:

$$D_и = (X_{max} - X_{min})/(1 + 3{,}32 * \lg N)$$

N – общее число измерений физических величин, определяющее размер выборки

$$D_и = \frac{X_{max} - X_{min}}{1 + 3{,}32 * lgN} = \frac{356 - 128}{1 + 32 * lg80} = 31{,}15 \approx 32$$

14) Разбивают выборку на К интервалов длиной $D_и$. Для каждого интервала устраивают нижнюю и верхнюю границ.

Нижнюю границу 1-го интервала определяют как $X^н_{i=1}$, а верхнюю $X^в_{i=1} = X^н_{i=1} + D_и$ Первый должен включать X_{min} , последний интервал X_{max}

Для последующих интервалов $X^н_{i=1} = X^в_{i=1}, X^в_{i=1} = X^н_{i=1} + D_и$

Середину интервала определяют:

$$X_i^{ср} = \frac{X^н_{i=1} + X^в_{i=1}}{2}$$

15) Рассчитывают экспериментальные частоты попадания результатов измерений в каждый интервал $n_i^э$. При этом значение пренадлежит интервалу, если совпадает с его нижней границей.

16) Вычисляют статистические характеристики выборки с учётом частоты попадания значений в интервал.

Среднее арифметическое значение для середин интервалов

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{К}(X_i^{ср} * n_i^э)$$

Средние квадратичное отклонение

$$\sigma = \sqrt{\frac{\sum_{i=1}^{К}((X_i^{ср} - \overline{X})^2 * n_i^э)}{N}}$$

$$\overline{X} = \frac{1}{N}\sum_{i=1}^{К}(X_i^{ср} * n_i^э) = \frac{17632}{80} = 220{,}4$$

$$\sigma = \sqrt{\frac{\sum_{i=1}^{К}((X_i^{ср} - \overline{X})^2 * n_i^э)}{N}} = \sqrt{\frac{170995{,}2}{80}} = 46{,}23$$

17) Вычисляют для каждого интервала значение:

$$t_i = \left|\frac{X_i^{ср} - \overline{X}}{\sigma}\right|$$

18) Устанавливаем значение плотности вероятностей нормального распределения f(ti) (по приложению 1)

19) Устанавливаем теоретические частоты попаданий результатов измерений в i-й интервал по формуле:

$$n_i^T = f(t_i) * \frac{D_и * N}{\sigma}$$

20) Вычисляем значение критерия Пирсона χ^2 по формуле:

$$\chi^2 = \sum_{i=1}^{K} \frac{(n_i^э - n_i^T)^2}{n_i^T} = 6{,}92$$

21) Рассчитываем число степеней свободы:

$$q = K - r - 1 = 8 - 2 - 1 = 5$$

22) По таблице 2.1 определяем допустимое значение $\chi^2_{ДОП}$:

$\chi^2_{ДОП}$ (a=0,10, q=5) = 9,24>6,92 следовательно $\chi^2 < \chi^2_{ДОП}$

Вопросы

1. Нормальный закон распределения
2. Критерий Пирсона

Рекомендуемая литература:

Основная литература:

6. Мелас В.Б., Шпилев П.В. Планирование и анализ для регрессионных моделей: учеб. пособие.- СПб.:Изд-во С.-Перерб. ун-та, 2014.- 96 с.
7. Мусина О.Н. Планирование и постановка научного эксперимента: учебно-методическое пособие / О.Н. Мусина.- М.- Берлин:Директ-Медиа, 2015.- 88 с.
8. Боярский М.В. Планирование и организация эксперимента: учебное пособие / М.В. Боярский, Э.А. Анисимов.- Йошкар-Ола: Поволжский государственный технологический университет, 2015.- 168 с.

Дополнительная литература:

9. Сафин Р.Г. Основы научных исследований. Организация и планирование эксперимента: учебное пособие / Р.Г. Сафин, А.И. Иванов, Н.Ф. Тимербаев; Мин-во образование и науки России, Казан. нац. исслед. технол. ун-т.-Казань: Изд-во КНИТУ, 2013.- 156 с.
10. Костин В.Н. Теория эксперимента: учебное пособие / В.Н. Костин, В.В. Паничев; Оренбургский государственный университет – Оренбург: ОГУ, 2013 .- 209 с.

Практическое занятие №5

Тема занятия. Парная линейная корреляция

Цель занятия. Приобрести практические навыки построения уравнения парной регрессии по экспериментальным данным.

Знания и умения, приобретаемые студентом в результате освоения темы, формируемые компетенции. Знает методы построения уравнения регрессии и оценки параметров регрессии.

Актуальность темы. Использование элементов регрессионного анализа широко применяется при решении многих прикладных задач.

Теоретическая часть

Каждый студент обрабатывает свой вариант экспериментальных данных, таблицы 2.1, 2.2, 2.3. Для выполнения вычислений можно разработать соответствующие процедуры с использованием любого универсального языка программирования, можно воспользоваться возможностями табличного процессора, например Excel. Обработка данных ведется применительно к трем видам уравнений регрессии, исходные данные приведены в таблицах:

таблица 2.1 – линейная регрессия $y = a_1 x + a_0$ и параболическая регрессия второго порядка $y = a_2 x^2 + a_1 x + a_0$;

таблица 2.2 – множественная регрессия $y = a_2 x_2 + a_1 x_1 + a_0$, в таблице 2.3 приведены варианты индивидуального задания.

3.1 *Парные коэффициенты корреляции* между признаками вычисляются по формуле (2.1).

$$\rho_{x,y} = \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^{N}(x_i - \mu_x)^2 \sum_{i=1}^{N}(y_i - \mu_y)^2}}. \qquad (2.1)$$

Значение коэффициента корреляции лежит в пределах от – 1 до +1. Если случайные величины независимы, то коэффициент корреляции обязательно равен нулю, обратное же утверждение неверно. Коэффициент корреляции характеризует значимость линейной связи между случайными величинами (параметрами):

- при $\rho_{xy} = 1$ значения x_i и y_i полностью совпадают. Иначе говоря, имеет место функциональная зависимость: зная значение одного параметра, можно однозначно указать значение другого параметра;
- при $\rho_x = -1$ величины x_i и y_i принимают противоположные значения. В этом случае имеет место функциональная зависимость;
- при $\rho_{xy} = 0$ величины x_i и y_i практически не связаны друг с другом линейным соотношением. Это не означает отсутствия каких-то других (например, нелинейных) связей между параметрами;
- при $0 < |\rho_{xy}| < 1$ однозначной линейной связи величин x_i и y_i нет. И чем меньше абсолютная величина коэффициента корреляции, тем в меньшей степени по значениям одного параметра можно предсказать значение другого.

Вывод соотношений для расчетов оценок параметров регрессионных моделей производится с использованием метода наименьших квадратов.

3.2 *Метод наименьших квадратов* (МНК) как вычислительная процедура был описан Лагранжем в 1806 г. Им также было предложено название этого метода.

В основе МНК лежат следующие положения:

1) значения величин ошибок и факторов независимы, а значит, и некоррелированы, т. е. предполагается, что механизмы порождения помехи не связаны с механизмом формирования значений факторов;

2) математическое ожидание ошибки ε должно быть равно нулю (постоянная составляющая входит в коэффициент a_0), иначе говоря, ошибка является центрированной величиной;

3) выборочная оценка дисперсии ошибки должна быть минимальна.

Пусть уравнение линейной регрессии имеет вид

$$y_i = \alpha_0 + \sum_{j=1}^{m} \alpha_j x_{ij} + \varepsilon_i,\ i = 1, 2, \dots, n. \qquad (2.2)$$

Обозначим $Y = (y_1, y_2, \dots, y_n)^T$, $X = \begin{pmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & \dots & x_{2p} \\ \dots & \dots & \dots & \dots \\ 1 & x_{n1} & \dots & x_{nm} \end{pmatrix}$,

$\alpha = (\alpha_0, \alpha_1, \alpha_2, \dots, \alpha_m)$, $\varepsilon = (\varepsilon_1, \varepsilon_2, \dots, \varepsilon_m)$. Тогда в матричной форме модель примет вид $Y = X\alpha + \varepsilon$. Оценкой является уравнение $Y = Xa + e$. Для оценки вектора неизвестных параметров α применяется МНК. Условие минимизации остаточной суммы может быть записано в виде

$$S = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = \sum_{i=1}^{n} e_i^2 = e^T e = (Y - Xb)^T (Y - Xb) \to \min \qquad (2.3)$$

На основании необходимого условия экстремума функции нескольких переменных необходимо приравнять частные производные по этим переменным к нулю или в матричном виде

$$\frac{\partial S}{\partial a} = \left(\frac{\partial S}{\partial a_0}, \frac{\partial S}{\partial a_1}, \dots, \frac{\partial S}{\partial a_m} \right) = 0 \qquad (2.4)$$

Оценка параметров производится путем решения систем уравнений (2.5).

$$\begin{cases} \dfrac{\partial S}{\partial a_0} = 0 \\ \dfrac{\partial S}{\partial a_1} = 0 \\ \dots \\ \dfrac{\partial S}{\partial a_m} = 0 \end{cases} \qquad (2.5)$$

Средняя относительная ошибка аппроксимации δ вычисляется по формуле (2.6).

$$\delta = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - \hat{y}_i|}{y_i}, \qquad (2.6)$$

где y_i – экспериментальные значения результативного признака,

$\hat{y}_i$ – рассчитанные по построенному уравнению регрессии значения результативного признака,

n – число элементов выборки.

3.3 Для выполнения проверки на 5%-ном уровне значимости гипотезы H_0 о равенстве нулю коэффициентов уравнения регрессии необходимо рассчитать показатели по формулам (2.7).

$$SS_{регр} = \sum_{i=1}^{n} \left(\hat{y}_i - \bar{y}\right)^2, \; SS_{ост} = \sum_{i=1}^{n} \left(\hat{y}_i - y_i\right)^2, \; SS = \sum_{i=1}^{n} \left(y_i - \bar{y}\right)^2, \qquad (2.7)$$

где y_i – экспериментальные значения результативного признака,

$\hat{y}_i$ – рассчитанные по построенному уравнению регрессии значения результативного признака,

$\bar{y}$ – среднее значение результативного признака, рассчитанное как среднее арифметическое по экспериментальным данным,

n – число элементов выборки.

Оценка коэффициента линейной детерминации, модуль оценки множественного коэффициента корреляции и оценка нормированного коэффициента линейной детерминации вычисляются по формулам (2.8), (2.9) и (2.10) соответственно.

$$\hat{R}^2 = 1 - \frac{SS_{ост}}{\mathrm{SS}} \qquad (2.8)$$

$$\hat{R} = \sqrt{\hat{R}^2} \qquad (2.9)$$

$$\tilde{R}^2 = 1 - \frac{SS_{\text{ост}}/(n-m-1)}{SS/(n-1)} \qquad (2.10)$$

Нормированный коэффициент детерминации $\tilde{R}^2$, в отличие от коэффициента линейной детерминации $\hat{R}^2$, увеличивающегося при увеличении числа m регрессоров, может и уменьшаться; чем больше его значение, тем качественнее уравнение регрессии.

Проверка гипотезы H_0: $a_1 = a_2 = \ldots = a_m = 0$ производится на основе анализа статистики, вычисляемой по формуле (2.11)

$$F_{m;n-m-1} = \frac{SS_{регр}/m}{SS_{ост}/(n-m-1)} = \frac{\hat{R}^2/m}{(1-\hat{R}^2)/(n-m-1)},$$

(2.11)

и имеющей (в предположении справедливости H_0) распределение Фишера-Снедекора с m и $(n - m - 1)$ степенями свободы. Вычисленное значение сравнивается с критическим значением $f_{0,05;\ m;\ n-m-1}$, найденным по статистическим таблицам распределения Фишера-Снедекора. Если наблюдаемое значение статистики F больше критической точки, то гипотеза H_0 отвергается на 5%-ном уровне значимости, в противном случае гипотеза принимается.

При проверке гипотез о равенстве нулю полученных коэффициентов регрессии в качестве критерия используется t-критерий Стьюдента, определяемый по формуле (2.12)

$$t_{a_i} = \frac{a_i}{\sqrt{\dfrac{\sigma^2_{\text{ост}} \sum\limits_{j=1}^{n} x_{ij}}{n \sum\limits_{j=1}^{n} (x_{ij} - \bar{x}_i)^2}}}, \qquad (2.12)$$

где t_{ai} – значения критерия Стьюдента для коэффициентов a_i;

$$\sigma_{ост}^2 = \sqrt{\frac{\sum_{i=1}^{n}(y_i - \bar{y})^2}{n-m-1}}$$ – остаточная дисперсия уравнения регрессии;

n – число элементов в выборке;

m – число переменных в уравнении регрессии; для парной линейной регрессии $m = 1$.

Полученные фактические значения критерия Стьюдента сравниваются с табличными значениями $t_{m,\alpha/2,n-m-1}$. Если оказывается, что рассчитанное значение больше критического, то соответствующий коэффициент статистически значим; в противном случае коэффициент незначим.

Оценка точности производится путем расчета суммы квадратов отклонений фактических значений от расчетных значений функций, полученных по уравнениям регрессии. Если эти отклонения много меньше (более чем на порядок) значений функции во всех точках, то уравнение регрессии хорошо описывает ЭД. При отклонениях, хотя бы в одной точке сравнимых со значениями функции, уравнения регрессии непригодны для описания всей совокупности ЭД.

По результатам выполнения двух предыдущих пунктов задания формируются выводы о возможности аппроксимации данных указанными уравнениями регрессии.

Наряду с точечными оценками a_i коэффициентов регрессии α_i регрессионный анализ позволяет получать и интервальные оценки коэффициентов с доверительной вероятностью γ.

Интервальная оценка с доверительной вероятностью γ для параметра α_i имеет вид (2.13)

$$a_i - t_{1-\gamma}\hat{s}_{a_i} \le \alpha_i \le a_i + t_{1-\gamma}\hat{s}_{a_i}, \qquad (2.13)$$

где $t_{1-\gamma}$ находят по таблице t-распределения при вероятности $1-\gamma$ и числе степеней свободы $v = n - m - 1$, а $\hat{s}_{a_{i-1}}$ вычисляется по формуле (2.14).

$$\hat{s}_{a_{i-1}} = \frac{1}{n-m-1}\left(Y-\hat{Y}\right)^T\left(Y-\hat{Y}\right)\left[\left(X^T X\right)^{-1}\right]_{ii}, \; i = 1, 2, \ldots, m+1, \qquad (2.14)$$

где $\left[\left(X^T X\right)^{-1}\right]_{ii}$ – i - ый диагональный элемент матрицы $\left[\left(X^T X\right)^{-1}\right]$.

Таблица 2.1

Линейная регрессия

x	y										
	1	2	3	4	5	6	7	8	9	10	11
0,32	2,62	2,54	2,47	3,35	3,18	3,11	3,12	13,24	8,57	7,13	3,19
0,08	1,05	0,92	0,82	1,53	1,44	1,48	1,26	12,38	7,49	6,49	1,20
1,02	7,21	7,26	7,24	8,30	8,31	8,42	8,58	15,65	12,13	8,88	8,69
0,19	1,90	1,72	1,43	2,34	2,29	2,23	2,21	12,78	8,10	6,77	2,14
0,00	0,70	0,38	0,25	0,94	0,97	0,91	0,78	12,04	7,07	6,02	0,57
1,06	7,55	7,47	7,50	8,61	8,54	8,72	8,93	15,66	12,18	9,06	8,96
0,26	2,34	2,16	2,11	2,94	2,81	2,86	2,67	12,96	8,30	6,86	2,63

0,16	1,58	1,41	1,39	2,11	2,10	2,09	1,88	12,59	7,86	6,68	1,84
0,45	3,50	3,52	3,31	4,30	4,30	4,11	4,22	13,58	9,25	7,43	4,20
1,58	10,74	11,00	11,07	12,25	12,35	12,53	12,80	17,56	14,63	10,52	13,02
1,32	9,24	9,31	9,27	10,31	10,50	10,67	10,92	16,63	13,46	9,78	11,04
1,47	10,15	10,32	10,36	11,35	11,59	11,81	12,08	17,04	14,26	10,26	12,11
0,55	4,08	4,09	3,94	4,87	4,94	4,85	4,98	13,99	9,80	7,53	4,96
0,63	4,77	4,70	4,62	5,41	5,42	5,61	5,49	14,24	10,21	7,95	5,51
1,89	12,88	13,03	13,26	14,50	14,73	15,03	15,30	18,48	16,17	11,32	15,58

Таблица 2.2

Множественная регрессия

№ п/п	Страна	Y	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$	$x^{(5)}$
1	Австралия	80	17800	15	8	16848	85
2	Австрия	79	8000	12	11	18396	58
3	Аргентина	75	33900	20	9	3408	86
4	Бангладеш	53	125000	35	11	202	16
5	Беларусь	76	10300	13	11	6500	65
6	Бельгия	79	10100	12	11	17912	96
7	Бразилия	67	156600	21	9	2354	75
8	Буркина-Фасо	50	10000	47	18	357	15
9	Великобритания	80	58400	13	11	15974	89
10	Вьетнам	68	73100	27	8	230	20
11	Гаити	47	6500	40	19	383	29
12	Германия	79	81200	11	11	17539	85
13	Гондурас	70	5600	35	6	1030	44
14	Гонконг	80	5800	13	6	14641	94
15	Египет	63	60000	29	9	748	44
16	Замбия	45	9100	46	18	573	42
17	Индия	59	911600	29	10	275	26
18	Ирландия	78	3600	14	9	12170	57
19	Испания	81	39200	11	9	13047	78
20	Италия	81	58100	11	10	17500	69
21	Канада	81	29100	14	8	19904	77
22	Китай	69	1205200	21	7	377	26
23	Колумбия	75	35600	24	6	1538	70
24	Коста-Рика	79	3300	26	4	2031	47
25	Куба	78	11100	17	7	1382	74
26	Малайзия	72	19500	29	5	2995	43
27	Марокко	70	28600	29	6	1062	46
28	Мексика	77	91800	28	5	3604	73
29	Нидерланды	81	15400	13	9	17245	89
30	Новая Зеландия	80	3524	16	8	14381	84
31	Норвегия	81	4300	13	10	17755	75
32	ОАЭ	74	2800	28	3	14193	81
33	Польша	77	38600	14	10	4429	62

№ п/п	Страна	Y	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$	$x^{(5)}$
34	Португалия	78	10500	12	10	9000	34
35	Россия	74	149200	13	11	6680	74
36	Саудовская Аравия	70	18000	38	6	6651	77
37	Северная Корея	73	23100	24	6	1000	60
38	Сингапур	79	2900	16	6	14990	100
39	США	79	260800	15	9	23474	75
40	Таиланд	72	59400	19	6	1800	22
41	Турция	73	62200	26	6	3721	61
42	Украина	75	51800	12	13	2340	67
43	Филиппины	68	69800	27	7	867	43
44	Финляндия	80	5100	13	10	15877	60
45	Франция	82	58000	13	9	18944	73
46	Чили	78	14000	23	6	2591	85
47	Швейцария	82	7000	12	9	22384	62
48	Швеция	81	8800	14	11	16900	84
49	Эфиопия	54	55200	45	14	122	12
50	ЮАР	68	43900	34	8	3128	49
51	Южная Корея	74	45000	16	6	6627	72
52	Япония	82	125500	11	7	19860	77

Здесь Y – ожидаемая продолжительность жизни женщины (в годах),
$x^{(1)}$ – численность населения (в тыс. чел.),
$x^{(2)}$ – рождаемость (на 1000 чел.),
$x^{(3)}$ – смертность (на 1000 чел.),
$x^{(4)}$ – ВВП на душу населения (в долл. США по покупательной способности валют),
$x^{(5)}$ – процент городского населения.

Таблица 2.3

Номера признаков для регрессионного анализа

Вариант	Номера факторных признаков	
1	1	2
2	1	3
3	1	4
4	1	5
5	2	3
6	2	4
7	2	5
8	3	4
9	3	5
10	4	5

Задание для самостоятельного решения

Построить регрессионные модели объектов по заданным ЭД.

Задание предусматривает построение трех регрессионных моделей:

1. построение уравнения линейной регрессии $y = a_1x + a_0$;

2. построение уравнения параболической регрессии второго порядка $y = a_2x^2 + a_1x + a_0$;

3. построение уравнения множественной регрессии $y = a_2x_2 + a_1x_1 + a_0$.

Значения результатов наблюдений величин x, y и z задаются в таблицах.

Требуется:

1. Рассчитать парные коэффициенты корреляции между признаками и сделать вывод о силе линейной связи результативного признака с регрессорами и сделать вывод о силе линейной связи между результативным признаком и регрессорами.

2. Вычислить оценки параметров модели линейной регрессии, вычислить среднюю относительную ошибку аппроксимации δ.

3. Предположив выполнение условий линейного регрессионного анализа:

a) оценить статистическую значимость уравнения регрессии (проверить на 5%-ном уровне значимости гипотезу H_0 о равенстве нулю коэффициентов уравнения регрессии);

b) проверить на 5%-ном уровне значимости гипотезы $H_0^{(j)}$: $a_j \neq 0$.

4.Систематизировать результаты, рассчитав:

a) коэффициент линейной детерминации $\hat{R}^2$ (*R*-квадрат), нормированный $\hat{R}^2$ (нормированный *R*-квадрат), ошибку аппроксимации δ, *F*-статистику и критическую точку $f_{0,05;k;\, n-k-1}$;

b) 95%-ные доверительные интервалы для коэффициентов уравнения регрессии;

c) для коэффициентов уравнения регрессии – числовые значения t-статистик и критическую точку $t_{0,05;\, n-m-1}$;

Вопросы для самопроверки

1. Привести общий вид уравнения парной регрессии
2. В чем заключается метод наименьших квадратов?

Рекомендуемая литература:

Основная литература:

1. Мелас В.Б., Шпилев П.В. Планирование и анализ для регрессионных моделей: учеб. пособие.- СПб.:Изд-во С.-Перерб. ун-та, 2014.- 96 с.
2. Мусина О.Н. Планирование и постановка научного эксперимента: учебно-методическое пособие / О.Н. Мусина.- М.- Берлин:Директ-Медиа, 2015.- 88 с.
3. Боярский М.В. Планирование и организация эксперимента: учебное пособие / М.В. Боярский, Э.А. Анисимов.- Йошкар-Ола: Поволжский государственный технологический университет, 2015.- 168 с.

Дополнительная литература:

1. Сафин Р.Г. Основы научных исследований. Организация и планирование эксперимента: учебное пособие / Р.Г. Сафин, А.И. Иванов, Н.Ф. Тимербаев; Мин-во образование и науки России, Казан. нац. исслед. технол. ун-т.-Казань: Изд-во КНИТУ, 2013.- 156 с.
2. Костин В.Н. Теория эксперимента: учебное пособие / В.Н. Костин, В.В. паничев; Оренбургский государственный университет – Оренбург: ОГУ, 2013 .- 209 с.

Практическое занятие №7

Тема занятия. Построение выборочного уравнения линии регрессии по сгруппированным данным

Цель занятия. Приобрести практические навыки построения уравнения парной регрессии по экспериментальным данным.

Знания и умения, приобретаемые студентом в результате освоения темы, формируемые компетенции. Знает методы построения уравнения регрессии и оценки параметров регрессии.

Актуальность темы. Использование элементов регрессионного анализа широко применяется при решении многих прикладных задач.

Теоретическая часть

Каждый студент обрабатывает свой вариант экспериментальных данных, таблицы 2.1, 2.2, 2.3. Для выполнения вычислений можно разработать соответствующие процедуры с использованием любого универсального языка программирования, можно воспользоваться возможностями табличного процессора, например Excel. Обработка данных ведется применительно к трем видам уравнений регрессии, исходные данные приведены в таблицах:

таблица 2.1 – линейная регрессия $y = a_1x + a_0$ и параболическая регрессия второго порядка $y = a_2x^2 + a_1x + a_0$;

таблица 2.2 – множественная регрессия $y = a_2x_2 + a_1x_1 + a_0$, в таблице 2.3 приведены варианты индивидуального задания.

3.4 *Парные коэффициенты корреляции* между признаками вычисляются по формуле (2.1).

$$\rho_{x,y} = \frac{\sum_{i=1}^{N}(x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^{N}(x_i - \mu_x)^2 \sum_{i=1}^{N}(y_i - \mu_y)^2}}. \qquad (2.1)$$

Значение коэффициента корреляции лежит в пределах от – 1 до +1. Если случайные величины независимы, то коэффициент корреляции обязательно равен нулю, обратное же утверждение неверно. Коэффициент корреляции характеризует значимость линейной связи между случайными величинами (параметрами):

- при $\rho_{xy} = 1$ значения x_i и y_i полностью совпадают. Иначе говоря, имеет место функциональная зависимость: зная значение одного параметра, можно однозначно указать значение другого параметра;
- при $\rho_x = –1$ величины x_i и y_i принимают противоположные значения. В этом случае имеет место функциональная зависимость;
- при $\rho_{xy} = 0$ величины x_i и y_i практически не связаны друг с другом линейным соотношением. Это не означает отсутствия каких-то других (например, нелинейных) связей между параметрами;

- при $0 < |\rho_{xy}| < 1$ однозначной линейной связи величин x_i и y_i нет. И чем меньше абсолютная величина коэффициента корреляции, тем в меньшей степени по значениям одного параметра можно предсказать значение другого.

Вывод соотношений для расчетов оценок параметров регрессионных моделей производится с использованием метода наименьших квадратов.

3.5 *Метод наименьших квадратов* (МНК) как вычислительная процедура был описан Лагранжем в 1806 г. Им также было предложено название этого метода.

В основе МНК лежат следующие положения:

4) значения величин ошибок и факторов независимы, а значит, и некоррелированы, т. е. предполагается, что механизмы порождения помехи не связаны с механизмом формирования значений факторов;

5) математическое ожидание ошибки ε должно быть равно нулю (постоянная составляющая входит в коэффициент a_0), иначе говоря, ошибка является центрированной величиной;

6) выборочная оценка дисперсии ошибки должна быть минимальна.

Пусть уравнение линейной регрессии имеет вид

$$y_i = \alpha_0 + \sum_{j=1}^{m} \alpha_j x_{ij} + \varepsilon_i,\ i = 1, 2, \ldots, n. \qquad (2.2)$$

Обозначим $Y = (y_1, y_2, \ldots, y_n)^T$, $X = \begin{pmatrix} 1 & x_{11} & \ldots & x_{1p} \\ 1 & x_{21} & \ldots & x_{2p} \\ \ldots & \ldots & \ldots & \ldots \\ 1 & x_{n1} & \ldots & x_{nm} \end{pmatrix}$,

$\alpha = (\alpha_0, \alpha_1, \alpha_2, \ldots, \alpha_m)$, $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_m)$. Тогда в матричной форме модель примет вид $Y = X\alpha + \varepsilon$. Оценкой является уравнение $Y = Xa + e$. Для оценки вектора неизвестных параметров α применяется МНК. Условие минимизации остаточной суммы может быть записано в виде

$$S = \sum_{i=1}^{n} (\hat{y}_i - y_i)^2 = \sum_{i=1}^{n} e_i^2 = e^T e = (Y - Xb)^T (Y - Xb) \to \min \qquad (2.3)$$

На основании необходимого условия экстремума функции нескольких переменных необходимо приравнять частные производные по этим переменным к нулю или в матричном виде

$$\frac{\partial S}{\partial a} = \left(\frac{\partial S}{\partial a_0}, \frac{\partial S}{\partial a_1}, \ldots, \frac{\partial S}{\partial a_m} \right) = 0 \qquad (2.4)$$

Оценка параметров производится путем решения систем уравнений (2.5).

$$\begin{cases} \dfrac{\partial S}{\partial a_0} = 0 \\ \dfrac{\partial S}{\partial a_1} = 0 \\ \dots \\ \dfrac{\partial S}{\partial a_m} = 0 \end{cases} \tag{2.5}$$

Средняя относительная ошибка аппроксимации δ вычисляется по формуле (2.6).

$$\delta = \frac{1}{n}\sum_{i=1}^{n}\frac{|y_i - \hat{y}_i|}{y_i}, \tag{2.6}$$

где y_i – экспериментальные значения результативного признака,

$\hat{y}_i$ – рассчитанные по построенному уравнению регрессии значения результативного признака,

n – число элементов выборки.

3.6 Для выполнения проверки на 5%-ном уровне значимости гипотезы H_0 о равенстве нулю коэффициентов уравнения регрессии необходимо рассчитать показатели по формулам (2.7).

$$SS_{регр} = \sum_{i=1}^{n}(\hat{y}_i - \bar{y})^2, \; SS_{ост} = \sum_{i=1}^{n}(\hat{y}_i - y_i)^2, \; SS = \sum_{i=1}^{n}(y_i - \bar{y})^2, \tag{2.7}$$

где y_i – экспериментальные значения результативного признака,

$\hat{y}_i$ – рассчитанные по построенному уравнению регрессии значения результативного признака,

$\bar{y}$ – среднее значение результативного признака, рассчитанное как среднее арифметическое по экспериментальным данным,

n – число элементов выборки.

Оценка коэффициента линейной детерминации, модуль оценки множественного коэффициента корреляции и оценка нормированного коэффициента линейной детерминации вычисляются по формулам (2.8), (2.9) и (2.10) соответственно.

$$\hat{R}^2 = 1 - \frac{SS_{ост}}{SS} \tag{2.8}$$

$$\hat{R} = \sqrt{\hat{R}^2} \tag{2.9}$$

$$\tilde{R}^2 = 1 - \frac{SS_{ост}/(n-m-1)}{SS/(n-1)} \tag{2.10}$$

Нормированный коэффициент детерминации $\tilde{R}^2$, в отличие от коэффициента линейной детерминации $\hat{R}^2$, увеличивающегося при увеличении числа m регрессоров, может и уменьшаться; чем больше его значение, тем качественнее уравнение регрессии.

Проверка гипотезы H_0: $a_1 = a_2 = \dots = a_m = 0$ производится на основе анализа статистики, вычисляемой по формуле (2.11)

$$F_{m;n-m-1} = \frac{SS_{регр} / m}{SS_{ост} / (n-m-1)} = \frac{\hat{R}^2 / m}{(1-\hat{R}^2)/(n-m-1)},$$

(2.11)

и имеющей (в предположении справедливости H_0) распределение Фишера-Снедекора с m и $(n - m - 1)$ степенями свободы. Вычисленное значение сравнивается с критическим значением $f_{0,05;\ m;\ n\text{-}m\text{-}1}$, найденным по статистическим таблицам распределения Фишера-Снедекора. Если наблюдаемое значение статистики F больше критической точки, то гипотеза H_0 отвергается на 5%-ном уровне значимости, в противном случае гипотеза принимается.

При проверке гипотез о равенстве нулю полученных коэффициентов регрессии в качестве критерия используется t-критерий Стьюдента, определяемый по формуле (2.12)

$$t_{a_i} = \frac{a_i}{\sqrt{\dfrac{\sigma^2_{\text{ост}} \sum\limits_{j=1}^{n} x_{ij}}{n \sum\limits_{j=1}^{n} (x_{ij} - \bar{x}_i)^2}}}, \qquad (2.12)$$

где t_{ai} – значения критерия Стьюдента для коэффициентов a_i ;

$$\sigma^2_{ост} = \sqrt{\frac{\sum\limits_{i=1}^{n} (y_i - \bar{y})^2}{n-m-1}}$$ – остаточная дисперсия уравнения регрессии;

n – число элементов в выборке;

m – число переменных в уравнении регрессии; для парной линейной регрессии $m = 1$.

Полученные фактические значения критерия Стьюдента сравниваются с табличными значениями $t_{m,\alpha/2,n\text{-}m\text{-}1}$. Если оказывается, что рассчитанное значение больше критического, то соответствующий коэффициент статистически значим; в противном случае коэффициент незначим.

Оценка точности производится путем расчета суммы квадратов отклонений фактических значений от расчетных значений функций, полученных по уравнениям регрессии. Если эти отклонения много меньше (более чем на порядок) значений функции во всех точках, то уравнение регрессии хорошо описывает ЭД. При отклонениях, хотя бы в одной точке сравнимых со значениями функции, уравнения регрессии непригодны для описания всей совокупности ЭД.

По результатам выполнения двух предыдущих пунктов задания формируются выводы о возможности аппроксимации данных указанными уравнениями регрессии.

Наряду с точечными оценками a_i коэффициентов регрессии α_i регрессионный анализ позволяет получать и интервальные оценки коэффициентов с доверительной вероятностью γ.

Интервальная оценка с доверительной вероятностью γ для параметра α_i имеет вид (2.13)

$$a_i - t_{1-\gamma}\hat{s}_{a_i} \le \alpha_i \le a_i + t_{1-\gamma}\hat{s}_{a_i}\,, \qquad (2.13)$$

где $t_{1-\gamma}$ находят по таблице t-распределения при вероятности $1-\gamma$ и числе степеней свободы $v = n - m - 1$, а $\hat{s}_{a_{i-1}}$ вычисляется по формуле (2.14).

$$\hat{s}_{a_{i-1}} = \frac{1}{n-m-1}\left(Y-\hat{Y}\right)^T\left(Y-\hat{Y}\right)\left[\left(X^T X\right)^{-1}\right]_{ii},\ i = 1, 2, \dots, m+1, \qquad (2.14)$$

где $\left[\left(X^T X\right)^{-1}\right]_{ii}$ – i - ый диагональный элемент матрицы $\left[\left(X^T X\right)^{-1}\right]$.

Таблица 2.1

Линейная регрессия

x	y										
	1	2	3	4	5	6	7	8	9	10	11
0,32	2,62	2,54	2,47	3,35	3,18	3,11	3,12	13,24	8,57	7,13	3,19
0,08	1,05	0,92	0,82	1,53	1,44	1,48	1,26	12,38	7,49	6,49	1,20
1,02	7,21	7,26	7,24	8,30	8,31	8,42	8,58	15,65	12,13	8,88	8,69
0,19	1,90	1,72	1,43	2,34	2,29	2,23	2,21	12,78	8,10	6,77	2,14
0,00	0,70	0,38	0,25	0,94	0,97	0,91	0,78	12,04	7,07	6,02	0,57
1,06	7,55	7,47	7,50	8,61	8,54	8,72	8,93	15,66	12,18	9,06	8,96
0,26	2,34	2,16	2,11	2,94	2,81	2,86	2,67	12,96	8,30	6,86	2,63
0,16	1,58	1,41	1,39	2,11	2,10	2,09	1,88	12,59	7,86	6,68	1,84
0,45	3,50	3,52	3,31	4,30	4,30	4,11	4,22	13,58	9,25	7,43	4,20
1,58	10,74	11,00	11,07	12,25	12,35	12,53	12,80	17,56	14,63	10,52	13,02
1,32	9,24	9,31	9,27	10,31	10,50	10,67	10,92	16,63	13,46	9,78	11,04
1,47	10,15	10,32	10,36	11,35	11,59	11,81	12,08	17,04	14,26	10,26	12,11
0,55	4,08	4,09	3,94	4,87	4,94	4,85	4,98	13,99	9,80	7,53	4,96
0,63	4,77	4,70	4,62	5,41	5,42	5,61	5,49	14,24	10,21	7,95	5,51
1,89	12,88	13,03	13,26	14,50	14,73	15,03	15,30	18,48	16,17	11,32	15,58

Таблица 2.2

Множественная регрессия

№ п/п	Страна	Y	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$	$x^{(5)}$
1	Австралия	80	17800	15	8	16848	85
2	Австрия	79	8000	12	11	18396	58
3	Аргентина	75	33900	20	9	3408	86
4	Бангладеш	53	125000	35	11	202	16
5	Беларусь	76	10300	13	11	6500	65
6	Бельгия	79	10100	12	11	17912	96
7	Бразилия	67	156600	21	9	2354	75
8	Буркина-Фасо	50	10000	47	18	357	15
9	Великобритания	80	58400	13	11	15974	89
10	Вьетнам	68	73100	27	8	230	20
11	Гаити	47	6500	40	19	383	29
12	Германия	79	81200	11	11	17539	85

№ п/п	Страна	Y	$x^{(1)}$	$x^{(2)}$	$x^{(3)}$	$x^{(4)}$	$x^{(5)}$
13	Гондурас	70	5600	35	6	1030	44
14	Гонконг	80	5800	13	6	14641	94
15	Египет	63	60000	29	9	748	44
16	Замбия	45	9100	46	18	573	42
17	Индия	59	911600	29	10	275	26
18	Ирландия	78	3600	14	9	12170	57
19	Испания	81	39200	11	9	13047	78
20	Италия	81	58100	11	10	17500	69
21	Канада	81	29100	14	8	19904	77
22	Китай	69	1205200	21	7	377	26
23	Колумбия	75	35600	24	6	1538	70
24	Коста-Рика	79	3300	26	4	2031	47
25	Куба	78	11100	17	7	1382	74
26	Малайзия	72	19500	29	5	2995	43
27	Марокко	70	28600	29	6	1062	46
28	Мексика	77	91800	28	5	3604	73
29	Нидерланды	81	15400	13	9	17245	89
30	Новая Зеландия	80	3524	16	8	14381	84
31	Норвегия	81	4300	13	10	17755	75
32	ОАЭ	74	2800	28	3	14193	81
33	Польша	77	38600	14	10	4429	62
34	Португалия	78	10500	12	10	9000	34
35	Россия	74	149200	13	11	6680	74
36	Саудовская Аравия	70	18000	38	6	6651	77
37	Северная Корея	73	23100	24	6	1000	60
38	Сингапур	79	2900	16	6	14990	100
39	США	79	260800	15	9	23474	75
40	Таиланд	72	59400	19	6	1800	22
41	Турция	73	62200	26	6	3721	61
42	Украина	75	51800	12	13	2340	67
43	Филиппины	68	69800	27	7	867	43
44	Финляндия	80	5100	13	10	15877	60
45	Франция	82	58000	13	9	18944	73
46	Чили	78	14000	23	6	2591	85
47	Швейцария	82	7000	12	9	22384	62
48	Швеция	81	8800	14	11	16900	84
49	Эфиопия	54	55200	45	14	122	12
50	ЮАР	68	43900	34	8	3128	49
51	Южная Корея	74	45000	16	6	6627	72
52	Япония	82	125500	11	7	19860	77

Здесь Y – ожидаемая продолжительность жизни женщины (в годах),
$x^{(1)}$ – численность населения (в тыс. чел.),
$x^{(2)}$ – рождаемость (на 1000 чел.),
$x^{(3)}$ – смертность (на 1000 чел.),

$x^{(4)}$ – ВВП на душу населения (в долл. США по покупательной способности валют),
$x^{(5)}$ – процент городского населения.

Таблица 2.3

Номера признаков для регрессионного анализа

Вариант	Номера факторных признаков	
11	1	2
12	1	3
13	1	4
14	1	5
15	2	3
16	2	4
17	2	5
18	3	4
19	3	5
20	4	5

Задание для самостоятельного решения

Построить регрессионные модели объектов по заданным ЭД.

Задание предусматривает построение трех регрессионных моделей:

4. построение уравнения линейной регрессии $y = a_1x + a_0$;
5. построение уравнения параболической регрессии второго порядка $y = a_2x^2 + a_1x + a_0$;
6. построение уравнения множественной регрессии $y = a_2x_2 + a_1x_1 + a_0$.

Значения результатов наблюдений величин x, y и z задаются в таблицах.

Требуется:

5. Рассчитать парные коэффициенты корреляции между признаками и сделать вывод о силе линейной связи результативного признака с регрессорами и сделать вывод о силе линейной связи между результативным признаком и регрессорами.
6. Вычислить оценки параметров модели линейной регрессии, вычислить среднюю относительную ошибку аппроксимации δ.
7. Предположив выполнение условий линейного регрессионного анализа:
   c) оценить статистическую значимость уравнения регрессии (проверить на 5%-ном уровне значимости гипотезу H_0 о равенстве нулю коэффициентов уравнения регрессии);
   d) проверить на 5%-ном уровне значимости гипотезы $H_0^{(j)}$: $a_j \neq 0$.
8. Систематизировать результаты, рассчитав:
   d) коэффициент линейной детерминации $\hat{R}^2$ (R-квадрат), нормированный $\hat{R}^2$ (нормированный R-квадрат), ошибку аппроксимации δ, F-статистику и критическую точку $f_{0,05;k;\,n-k-1}$;
   e) 95%-ные доверительные интервалы для коэффициентов уравнения регрессии;

f) для коэффициентов уравнения регрессии – числовые значения t-статистик и критическую точку $t_{0,05;\, n-m-1}$;

Вопросы для самопроверки

1. Привести общий вид уравнения парной регрессии
2. В чем заключается метод наименьших квадратов?

Рекомендуемая литература:

Основная литература:

1. Мелас В.Б., Шпилев П.В. Планирование и анализ для регрессионных моделей: учеб. пособие.- СПб.:Изд-во С.-Перерб. ун-та, 2014.- 96 с.
2. Мусина О.Н. Планирование и постановка научного эксперимента: учебно-методическое пособие / О.Н. Мусина.- М.- Берлин:Директ-Медиа, 2015.- 88 с.
3. Боярский М.В. Планирование и организация эксперимента: учебное пособие / М.В. Боярский, Э.А. Анисимов.- Йошкар-Ола: Поволжский государственный технологический университет, 2015.- 168 с.

Дополнительная литература:

4. Сафин Р.Г. Основы научных исследований. Организация и планирование эксперимента: учебное пособие / Р.Г. Сафин, А.И. Иванов, Н.Ф. Тимербаев; Мин-во образование и науки России, Казан. нац. исслед. технол. ун-т.-Казань: Изд-во КНИТУ, 2013.- 156 с.
5. Костин В.Н. Теория эксперимента: учебное пособие / В.Н. Костин, В.В. паничев; Оренбургский государственный университет – Оренбург: ОГУ, 2013 .- 209 с.

Практическое занятие №8.

Тема занятия. Парная нелинейная корреляция

Цель занятия. Рассмотреть методы нелинейной парной регрессии.

Знания и умения, приобретаемые студентом в результате освоения темы, формируемые компетенции. Знает методы построения нелинейного уравнения регрессии.

Актуальность темы. Приемы и методы нелинейной парной регрессии применяются для решения задач с прикладным содержанием.

Теоретическая часть.

Нелинейные регрессии делятся на два класса: регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам, и регрессии, нелинейные по оцениваемым параметрам.

Регрессии, *нелинейные по объясняющим переменным*:

- полиномы разных степеней $y = a + b_1 x + b_2 x^2 + b_3 x^3 + \varepsilon$;
- логарифмическая $y = a + b\ln x + \varepsilon$;
- равносторонняя гипербола $y = a + \frac{b}{x} + \varepsilon$.

Регрессии, *нелинейные по оцениваемым параметрам*:

- степенная $y = a\, x^b\, \varepsilon$;
- показательная $y = a\, b^x\, \varepsilon$;
- экспоненциальная $y = e^{a+bx}\, \varepsilon$.

Нелинейные уравнения регрессии можно разделить на два класса:

- уравнения, которые с помощью замены переменных можно привести к линейному виду в новых переменных x', y'

$$y' = a' + b'x';$$

- уравнения, для которых это невозможно. Назовем их внутренне нелинейными.

Для того чтобы оценить параметры нелинейных уравнений регрессии можно использовать МНК, но необходимо сначала преобразовать уравнение регрессии. Если уравнение не линейно по объясняющим переменным, но линейно по своим параметрам, то можно сделать замену переменной.

$$\hat{y} = a_0 + a_1/x.$$

Делается замена переменной $x' = 1/x$ и уравнение преобразовывается к виду:

$$\hat{y} = a_0 + a_1x'.$$

Параметры этого уравнения рассчитываются по обычным формулам, где используются фактические значения параметра y из выборки и рассчитанные по данным выборки с помощью значений переменной x'. Если уравнение не линейно и по объясняющим переменным и по параметрам, то его необходимо *линеаризовать*.

Линеаризующие преобразования для нелинейных моделей приведены в таблице 1

Для общей оценки качества построенной эконометрической определяются такие характеристики как коэффициент детерминации, индекс корреляции, средняя относительная ошибка аппроксимации, а также проверяется значимость уравнения регрессии с помощью F-критерия Фишера. Перечисленные характеристики являются достаточно универсальными и могут применяться как для линейных, так и для нелинейных моделей, а также моделей с двумя и более факторными переменными. Определяющее значение при вычислении всех перечисленных характеристик качества играет ряд остатков ε_i, который вычисляется путем вычитания из фактических (полученных по наблюдениям) значений исследуемого признака y_i значений, рассчитанных по уравнению модели $\hat{y}_i$.

Таблица 1 – Линеаризующие преобразования для нелинейных моделей

Зависимость	**Формула**	**Преобразование**	**Зависимость между параметрами**
Гиперболическая	$\hat{y} = a + \dfrac{b}{x}$	$y' = y,\ x' = \dfrac{1}{x}$	$a' = a,\ b' = b$
Логарифмическая	$\hat{y} = a + b\ln x$	$y' = y,\ x' = \ln x$	$a' = a,\ b' = b$
Степенная	$\hat{y} = ax^b$	$y' = \ln y,\ x' = \ln x$	$a' = \ln a,\ b' = b$
Показательная	$\hat{y} = ab^x$	$y' = \ln y,\ x' = x$	$a' = \ln a,\ b' = \ln b$
Экспоненциальная	$\hat{y} = e^{a+bx}$	$y' = \ln y,\ x' = x$	$a' = a,\ b' = b$

В случае нелинейной зависимости между показателями нельзя использовать для тесноты связи линейный парный коэффициент корреляции.

Индекс корреляции для нелинейных регрессий рассчитывается по формуле, как корень из коэффициента детерминации:

$$R = \sqrt{\frac{\sigma_{\hat{y}}^2}{\sigma_y^2}} = \sqrt{R^2}\ .$$

Эта величина всегда лежит в интервале от 0 до 1. Если между показателями x и y существует функциональная зависимость, выражаемая построенным уравнением регрессии, то объясняемая дисперсия будет равна единице ($R = 1$). Если между показателями x и y отсутствует зависимость, то объясненная дисперсия будет равна нулю ($R = 0$).

Чтобы убедиться в пригодности и надежности построенной модели для использования в прикладных целях используют F-критерий Фишера.

Задачи для самостоятельного решения

Задача 1

Получены функции:

$y = a + bx^3$, $y = a + b\ln x$, $y = a + bx^c$, $y^a = b + cx^2$, $y = 1 + a(1 - x^b)$, $y = a + bx/10$.

Определите, какие из представленных выше функций линейны по переменным, линейны по параметрам, нелинейны ни по переменным, ни по параметрам.

Задача 2

Для трех видов продукции *A*, *B* и *C* модели зависимости удельных постоянных расходов от объема выпускаемой продукции выглядит следующим образом:

$$y_A = 600;\ y_B = 80 + 0{,}7x;\ y^C = 40x^{0{,}5}.$$

Определите коэффициенты эластичности по каждому виду продукции и поясните их смысл. Определите, каким должен быть объем выпускаемой продукции, чтобы коэффициенты эластичности для *B* и *C* были равны. Сравните эластичность затрат для продукции *B* и *C* при $x = 1000$.

Задача 3

Изучается зависимость материалоемкости продукции от размера предприятия по 10 однородным заводам (см. таблицу).

Показатель	**Материалоемкость продукции по заводам**									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
Потреблено материалов на единицу продукции, кг., *y*	9	6	5	4	3,7	3,6	3,5	6	7	3,5
Выпуск продукции, тыс. ед., *x*	100	200	300	400	500	600	700	150	120	250

1. Найдите параметры уравнения $y = \alpha + \dfrac{\beta}{x} + \varepsilon$;
2. Оцените тесноту связи с помощью индекса корреляции;
3. Охарактеризуйте эластичность изменения материалоемкости продукции;
4. Сделайте вывод о значимости уравнения регрессии.

Задача 4

Некоторая организация в течении 6 кварталов вкладывала всю прибыль в свое развитие. При этом предполагается, что прибыль растет по показательному закону $y = ab^x$ (здесь фактор x – номер квартала, y – прибыль, млн. руб.). Составить уравнение регрессии, найти коэффициент нелинейной корреляции, и при $\alpha = 0{,}0.5$ проверить его значимость.

x_i, номер квартала	1	2	3	4	5	6
y_i, прибыль, млн.руб.	1	2	5	9	15	27

Задача 5

Владелец супермаркета доставил задачу определить зависимость между средней длинной очереди в кассу (фактор y, чел.) и количеством касс, обслуживающих клиентов (фактор x, шт.). По результатам наблюдений были получены выборки значений:

x_i	2	3	4	5	6	7	8
y_i	45	42	37	31	23	12	3

Предполагается, что зависимость между факторами имеет вид $y(x) = ax^2 + bx + c$. Построить уравнение параболической регрессии, найти нелинейный коэффициент парной корреляции и на уровне значимости $\alpha = 0,05$ проверить его значимость.

Вопросы для самопроверки

1. Общий вид уравнения нелинейной регрессии
2. Алгоритм построения.

Рекомендуемая литература:

Основная литература:

1. Мелас В.Б., Шпилев П.В. Планирование и анализ для регрессионных моделей: учеб. пособие.- СПб.:Изд-во С.-Перерб. ун-та, 2014.- 96 с.
2. Мусина О.Н. Планирование и постановка научного эксперимента: учебно-методическое пособие / О.Н. Мусина.- М.- Берлин:Директ-Медиа, 2015.- 88 с.
3. Боярский М.В. Планирование и организация эксперимента: учебное пособие / М.В. Боярский, Э.А. Анисимов.- Йошкар-Ола: Поволжский государственный технологический университет, 2015.- 168 с.

Дополнительная литература:

6. Сафин Р.Г. Основы научных исследований. Организация и планирование эксперимента: учебное пособие / Р.Г. Сафин, А.И. Иванов, Н.Ф. Тимербаев; Мин-во образование и науки России, Казан. нац. исслед. технол. ун-т.-Казань: Изд-во КНИТУ, 2013.- 156 с.
7. Костин В.Н. Теория эксперимента: учебное пособие / В.Н. Костин, В.В. паничев; Оренбургский государственный университет – Оренбург: ОГУ, 2013 .- 209 с.

Практическое занятие №9.

Тема занятия. Построение модельного уравнения нелинейной корреляции

Цель занятия. Рассмотреть методы нелинейной парной регрессии.

Знания и умения, приобретаемые студентом в результате освоения темы, формируемые компетенции. Знает методы построения нелинейного уравнения регрессии.

Актуальность темы. Приемы и методы нелинейной парной регрессии применяются для решения задач с прикладным содержанием.

Теоретическая часть.

Нелинейные регрессии делятся на два класса: регрессии, нелинейные относительно включенных в анализ объясняющих переменных, но линейные по оцениваемым параметрам, и регрессии, нелинейные по оцениваемым параметрам.

Регрессии, *нелинейные по объясняющим переменным*:

- полиномы разных степеней $y = a + b_1 x + b_2 x^2 + b_3 x^3 + \varepsilon$;
- логарифмическая $y = a + b\ln x + \varepsilon$;
- равносторонняя гипербола $y = a + \frac{b}{x} + \varepsilon$.

Регрессии, *нелинейные по оцениваемым параметрам*:

- степенная $y = a\, x^b\, \varepsilon$;
- показательная $y = a\, b^x\, \varepsilon$;

- экспоненциальная $y = e^{a+bx}\,\varepsilon$.

Нелинейные уравнения регрессии можно разделить на два класса:

- уравнения, которые с помощью замены переменных можно привести к линейному виду в новых переменных x', y'

$$y' = a' + b'x';$$

- уравнения, для которых это невозможно. Назовем их внутренне нелинейными.

Для того чтобы оценить параметры нелинейных уравнений регрессии можно использовать МНК, но необходимо сначала преобразовать уравнение регрессии. Если уравнение не линейно по объясняющим переменным, но линейно по своим параметрам, то можно сделать замену переменной.

$$\hat{y} = a_0 + a_1/x.$$

Делается замена переменной $x' = 1/x$ и уравнение преобразовывается к виду:

$$\hat{y} = a_0 + a_1x'.$$

Параметры этого уравнения рассчитываются по обычным формулам, где используются фактические значения параметра y из выборки и рассчитанные по данным выборки с помощью значений переменной x'. Если уравнение не линейно и по объясняющим переменным и по параметрам, то его необходимо *линеаризовать*.

Линеаризующие преобразования для нелинейных моделей приведены в таблице 1

Для общей оценки качества построенной эконометрической определяются такие характеристики как коэффициент детерминации, индекс корреляции, средняя относительная ошибка аппроксимации, а также проверяется значимость уравнения регрессии с помощью F-критерия Фишера. Перечисленные характеристики являются достаточно универсальными и могут применяться как для линейных, так и для нелинейных моделей, а также моделей с двумя и более факторными переменными. Определяющее значение при вычислении всех перечисленных характеристик качества играет ряд остатков ε_i, который вычисляется путем вычитания из фактических (полученных по наблюдениям) значений исследуемого признака y_i значений, рассчитанных по уравнению модели $\hat{y}_i$.

Таблица 1 – Линеаризующие преобразования для нелинейных моделей

Зависимость	**Формула**	**Преобразование**	**Зависимость между параметрами**
Гиперболическая	$\hat{y} = a + \dfrac{b}{x}$	$y' = y,\ x' = \dfrac{1}{x}$	$a' = a,\ b' = b$
Логарифмическая	$\hat{y} = a + b\ln x$	$y' = y,\ x' = \ln x$	$a' = a,\ b' = b$
Степенная	$\hat{y} = ax^b$	$y' = \ln y,\ x' = \ln x$	$a' = \ln a,\ b' = b$
Показательная	$\hat{y} = ab^x$	$y' = \ln y,\ x' = x$	$a' = \ln a,\ b' = \ln b$
Экспоненциальная	$\hat{y} = e^{a+bx}$	$y' = \ln y,\ x' = x$	$a' = a,\ b' = b$

В случае нелинейной зависимости между показателями нельзя использовать для тесноты связи линейный парный коэффициент корреляции.

Индекс корреляции для нелинейных регрессий рассчитывается по формуле, как корень из коэффициента детерминации:

$$R = \sqrt{\frac{\sigma_{\hat{y}}^2}{\sigma_y^2}} = \sqrt{R^2}\ .$$

Эта величина всегда лежит в интервале от 0 до 1. Если между показателями x и y существует функциональная зависимость, выражаемая построенным уравнением регрессии, то объясняемая дисперсия будет равна единице ($R = 1$). Если между показателями x и y отсутствует зависимость, то объясненная дисперсия будет равна нулю ($R = 0$).

Чтобы убедиться в пригодности и надежности построенной модели для использования в прикладных целях используют *F*-критерий Фишера.

Задачи для самостоятельного решения

Задача 1

Получены функции:

$y = a + bx^3$, $y = a + b\ln x$, $y = a + bx^c$, $y^a = b + cx^2$, $y = 1 + a(1 - x^b)$, $y = a + bx/10$.

Определите, какие из представленных выше функций линейны по переменным, линейны по параметрам, нелинейны ни по переменным, ни по параметрам.

Задача 2

Для трех видов продукции *A*, *B* и *C* модели зависимости удельных постоянных расходов от объема выпускаемой продукции выглядит следующим образом:

$$y_A = 600;\ y_B = 80 + 0{,}7x;\ y^C = 40x^{0,5}.$$

Определите коэффициенты эластичности по каждому виду продукции и поясните их смысл. Определите, каким должен быть объем выпускаемой продукции, чтобы коэффициенты эластичности для *B* и *C* были равны. Сравните эластичность затрат для продукции *B* и *C* при $x = 1000$.

Задача 3

Изучается зависимость материалоемкости продукции от размера предприятия по 10 однородным заводам (см. таблицу).

Показатель	Материалоемкость продукции по заводам									
	1	2	3	4	5	6	7	8	9	10
Потреблено материалов на единицу продукции, кг., *y*	9	6	5	4	3,7	3,6	3,5	6	7	3,5
Выпуск продукции, тыс. ед., *x*	100	200	300	400	500	600	700	150	120	250

5. Найдите параметры уравнения $y = \alpha + \frac{\beta}{x} + \varepsilon$;
6. Оцените тесноту связи с помощью индекса корреляции;
7. Охарактеризуйте эластичность изменения материалоемкости продукции;
8. Сделайте вывод о значимости уравнения регрессии.

Задача 4

Некоторая организация в течении 6 кварталов вкладывала всю прибыль в свое развитие. При этом предполагается, что прибыль растет по показательному закону $y = ab^x$ (здесь фактор *x* – номер квартала, *y* – прибыль, млн. руб.). Составить уравнение регрессии, найти коэффициент нелинейной корреляции, и при $\alpha = 0{,}0.5$ проверить его значимость.

x_i, номер квартала	1	2	3	4	5	6
y_i, прибыль, млн.руб.	1	2	5	9	15	27

Задача 5

Владелец супермаркета доставил задачу определить зависимость между средней длинной очереди в кассу (фактор *y*, чел.) и количеством касс, обслуживающих клиентов (фактор *x*, шт.). По результатам наблюдений были получены выборки значений:

x_i	2	3	4	5	6	7	8

y_i	45	42	37	31	23	12	3

Предполагается, что зависимость между факторами имеет вид $y(x) = ax^2 + bx + c$. Построить уравнение параболической регрессии, найти нелинейный коэффициент парной корреляции и на уровне значимости $\alpha = 0{,}05$ проверить его значимость.

Вопросы для самопроверки

1. Общий вид уравнения нелинейной регрессии
2. Алгоритм построения.

Рекомендуемая литература:

Основная литература:

1. Мелас В.Б., Шпилев П.В. Планирование и анализ для регрессионных моделей: учеб. пособие.- СПб.:Изд-во С.-Перерб. ун-та, 2014.- 96 с.
2. Мусина О.Н. Планирование и постановка научного эксперимента: учебно-методическое пособие / О.Н. Мусина.- М.- Берлин:Директ-Медиа, 2015.- 88 с.
3. Боярский М.В. Планирование и организация эксперимента: учебное пособие / М.В. Боярский, Э.А. Анисимов.- Йошкар-Ола: Поволжский государственный технологический университет, 2015.- 168 с.

Дополнительная литература:

4. Сафин Р.Г. Основы научных исследований. Организация и планирование эксперимента: учебное пособие / Р.Г. Сафин, А.И. Иванов, Н.Ф. Тимербаев; Мин-во образование и науки России, Казан. нац. исслед. технол. ун-т.-Казань: Изд-во КНИТУ, 2013.- 156 с.
5. Костин В.Н. Теория эксперимента: учебное пособие / В.Н. Костин, В.В. паничев; Оренбургский государственный университет – Оренбург: ОГУ, 2013 .- 209 с.

Практическое занятие 10.

Тема занятия. Множественная линейная корреляция.

Цель занятия. Рассмотреть множественную линейную корреляцию.

Знания и умения, приобретаемые студентом в результате освоения темы, формируемые компетенции. Знает множественная линейная корреляцию и методы оценки уравнения регрессии. Владеет способностью к самоорганизации и самообразованию.

Актуальность темы. Элементы НИРС используются в прикладных задачах.

Теоретическая часть.

Для измерения тесноты связи при множественной корреляционной зависимости, то есть при исследовании трех и более признаков одновременно, вычисляется **множественный и частные коэффициенты корреляции.**

Множественный коэффициент корреляции вычисляется при наличии линейной связи между результативным и несколькими факторными признаками, а также между каждой парой факторных признаков.

Множественный коэффициент корреляции для двух факторных признаков вычисляется по формуле:

$$Ry/x_1x_2 = \sqrt{\frac{r^2{}_{yx_1} + r^2{}_{yx_2} - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r^2{}_{x_1x_2}}}, \text{ где}$$

r_{yx} - парные коэффициенты корреляции между признаками.

Множественный коэффициент корреляции изменяется в пределах от 0 до 1 и по определению положителен: $0 \le R \le 1$.

Приближение R к единице свидетельствует о сильной зависимости между признаками.

$$r_{yx_1} = \frac{\overline{yx_1} - \overline{y} \cdot \overline{x_1}}{\delta_y \cdot \delta_{x1}} \qquad r_{yx_2} = \frac{\overline{yx_2} - \overline{y} \cdot \overline{x_2}}{\delta_y \cdot \delta_{x_2}} \qquad r_{x_1x_2} = \frac{\overline{x_1 x_2} - \overline{x_1} \cdot \overline{x_2}}{\delta x_1 \cdot \delta x_2}$$

Частные коэффициенты корреляции характеризуют степень тесноты связи между двумя признаками x_1 и x_2 при фиксированном значении других (k - 2) факторных признаков, то есть когда влияние x_1 исключается, то есть оценивается связь между x_1 и x_2 в "чистом виде".

В случае зависимости y от двух факторных признаков x_1 и x_2 коэффициенты частной корреляции имеют вид:

$$r_{yx_1/x_2} = \frac{r_{yx_1} - r_{x_1x_2} \cdot r_{yx_2}}{\sqrt{(1 - r^2{}_{x_2y}) \cdot (1 - r^2{}_{x_1x_2})}}$$

$$r_{yx_2/x_1} = \frac{r_{yx_2} - r_{x_1y} \cdot r_{x_1x_2}}{\sqrt{(1 - r^2{}_{x_1y}) \cdot (1 - r^2{}_{x_1x_2})}},$$

где r – парные коэффициенты корреляции между указанными в индексе переменными.

В первом случае исключено влияние факторного признака x_2, во втором – x_1.

Вопросы

1. Множественная регрессия
2. Частные коэффициенты корреляции

Рекомендуемая литература:

Основная литература:

1. Мелас В.Б., Шпилев П.В. Планирование и анализ для регрессионных моделей: учеб. пособие.- СПб.:Изд-во С.-Переpб. ун-та, 2014.- 96 с.
2. Мусина О.Н. Планирование и постановка научного эксперимента: учебно-

методическое пособие / О.Н. Мусина.- М.- Берлин:Директ-Медиа, 2015.- 88 с.
3. Боярский М.В. Планирование и организация эксперимента: учебное пособие / М.В. Боярский, Э.А. Анисимов.- Йошкар-Ола: Поволжский государственный технологический университет, 2015.- 168 с.

Дополнительная литература:

4. Сафин Р.Г. Основы научных исследований. Организация и планирование эксперимента: учебное пособие / Р.Г. Сафин, А.И. Иванов, Н.Ф. Тимербаев; Мин-во образование и науки России, Казан. нац. исслед. технол. ун-т.-Казань: Изд-во КНИТУ, 2013.- 156 с.
5. Костин В.Н. Теория эксперимента: учебное пособие / В.Н. Костин, В.В. Паничев; Оренбургский государственный университет – Оренбург: ОГУ, 2013 .- 209 с.

Практическое занятие 11.

Тема занятия. Построение модели множественной линейной корреляции

Цель занятия. Рассмотреть множественную линейную корреляцию.

Знания и умения, приобретаемые студентом в результате освоения темы, формируемые компетенции. Знает множественная линейная корреляцию и методы оценки уравнения регрессии. Владеет способностью к самоорганизации и самообразованию.

Актуальность темы. Элементы НИРС используются в прикладных задачах.

Теоретическая часть.

Для измерения тесноты связи при множественной корреляционной зависимости, то есть при исследовании трех и более признаков одновременно, вычисляется **множественный и частные коэффициенты корреляции.**

Множественный коэффициент корреляции вычисляется при наличии линейной связи между результативным и несколькими факторными признаками, а также между каждой парой факторных признаков.

Множественный коэффициент корреляции для двух факторных признаков вычисляется по формуле:

$$Ry/x_1x_2 = \sqrt{\frac{r^2{}_{yx_1} + r^2{}_{yx_2} - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r^2{}_{x_1x_2}}}$$, где

r_{yx} - парные коэффициенты корреляции между признаками.

Множественный коэффициент корреляции изменяется в пределах от 0 до 1 и по определению положителен: $0 \leq R \leq 1$.

Приближение R к единице свидетельствует о сильной зависимости между признаками.

$$r_{yx_1} = \frac{\overline{yx_1} - \overline{y} \cdot \overline{x_1}}{\delta_y \cdot \delta_{x1}} \qquad r_{yx_2} = \frac{\overline{yx_2} - \overline{y} \cdot \overline{x_2}}{\delta_y \cdot \delta_{x_2}} \qquad r_{x_1x_2} = \frac{\overline{x_1 x_2} - \overline{x_1} \cdot \overline{x_2}}{\delta x_1 \cdot \delta x_2}$$

Частные коэффициенты корреляции характеризуют степень тесноты связи между двумя признаками x_1 и x_2 при фиксированном значении других (k - 2) факторных признаков, то есть когда влияние x_1 исключается, то есть оценивается связь между x_1 и x_2 в "чистом виде".

В случае зависимости y от двух факторных признаков x_1 и x_2 коэффициенты частной корреляции имеют вид:

$$r_{yx_1/x_2} = \frac{r_{yx_1} - r_{x_1x_2} \cdot r_{yx_2}}{\sqrt{(1 - r^2_{x_2y}) \cdot (1 - r^2_{x_1x_2})}}$$

$$r_{yx_2/x_1} = \frac{r_{yx_2} - r_{x_1y} \cdot r_{x_1x_2}}{\sqrt{(1 - r^2_{x_1y}) \cdot (1 - r^2_{x_1x_2})}},$$

где r – парные коэффициенты корреляции между указанными в индексе переменными.

В первом случае исключено влияние факторного признака x_2, во втором – x_1.

Вопросы

1. Множественная регрессия
2. Частные коэффициенты корреляции

Рекомендуемая литература:

Основная литература:

1. Мелас В.Б., Шпилев П.В. Планирование и анализ для регрессионных моделей: учеб. пособие.- СПб.:Изд-во С.-Переpб. ун-та, 2014.- 96 с.
2. Мусина О.Н. Планирование и постановка научного эксперимента: учебно-методическое пособие / О.Н. Мусина.- М.- Берлин:Директ-Медиа, 2015.- 88 с.
3. Боярский М.В. Планирование и организация эксперимента: учебное пособие / М.В. Боярский, Э.А. Анисимов.- Йошкар-Ола: Поволжский государственный технологический университет, 2015.- 168 с.

Дополнительная литература:

4. Сафин Р.Г. Основы научных исследований. Организация и планирование эксперимента: учебное пособие / Р.Г. Сафин, А.И. Иванов, Н.Ф. Тимербаев; Мин-во образование и науки России, Казан. нац. исслед. технол. ун-т.-Казань: Изд-во КНИТУ, 2013.- 156 с.
5. Костин В.Н. Теория эксперимента: учебное пособие / В.Н. Костин, В.В. Паничев; Оренбургский государственный университет – Оренбург: ОГУ, 2013 .- 209 с.

Практическое занятие 12.

Тема занятия. П а р а м е т р ы о п т и м и з а ц и и и т р е б о в а н и я к н и м

Цель занятия. Рассмотреть множественную линейную корреляцию.

Знания и умения, приобретаемые студентом в результате освоения темы, формируемые компетенции. Знает множественная линейная корреляцию и методы оценки уравнения регрессии. Владеет способностью к самоорганизации и самообразованию.

Актуальность темы. Элементы НИРС используются в прикладных задачах.

Теоретическая часть.

1 Параметры оптимизации

Выбор параметров оптимизации (критериев оптимизации) является одним из главных этапов работы на стадии предварительного изучения объекта исследования, т.к. правильная постановка задачи зависит от правильности выбора параметра оптимизации, являющегося функцией цели.

Под параметром оптимизации понимают характеристику цели, заданную количественно. Параметр оптимизации является реакцией (откликом) на воздействие факторов, которые определяют поведение выбранной системы.

2 Требования к параметру оптимизации

Параметр оптимизации - это признак, по которому оптимизируется процесс.

Он должен быть количественным, задаваться числом. Множество значений, которые может принимать параметр оптимизации, называется областью его определения. Области определения могут быть непрерывными и дискретными, ограниченными и неограниченными. Например, выход реакции - это параметр оптимизации с непрерывной ограниченной областью определения. Он может изменяться в интервале от 0 до 100%. Число бракованных изделий, число зерен на шлифе сплава, число кровяных телец в пробе крови - примеры параметров с дискретной областью определения, ограниченной снизу.

Количественная оценка параметра оптимизации на практике не всегда возможна. В таких случаях пользуются приемом, называемым ранжированием. При этом параметрам оптимизации присваиваются оценки - ранги по заранее выбранной шкале: двухбалльной, пятибалльной и т.д. Ранговый параметр имеет дискретную ограниченную область определения. В простейшем случае область содержит

10

два значения (да, нет; хорошо, плохо). Это может соответствовать, например, годной продукции и браку.

Итак, первое требование: параметр оптимизации должен быть количественным.

Второе требование: параметр оптимизации должен выражаться одним числом. Например, регистрация показания прибора. Часто приходится проводить некоторые вычисления, например, пользоваться значением отклонений (или квадратов отклонений) от заданного числа.

Третье требование, связанное с количественной природой параметра оптимизации - однозначность в статистическом смысле. Заданному набору значений факторов должно соответствовать одно значение параметра оптимизации, при этом обратное неверно: одному и тому же значению параметра могут соответствовать разные наборы значений факторов.

Четвертым - является требование к его возможности действительно эффективно оценивать функционирование системы. Представление об объекте не остается постоянным в ходе исследования. Оно меняется по мере накопления информации и в зависимости от достигнутых результатов. Это приводит к последовательному подходу при выборе параметра оптимизации. Так, например, на первых стадиях исследования технологических процессов в качестве параметра оптимизации часто используется выход продукта. Однако в дальнейшем, когда возможность повышения выхода исчерпан, начинают интересоваться такими параметрами, как себестоимость, чистота продукта и т.д. Иногда попытка добиться оптимума с учетом некоторого локального или промежуточного параметра оптимизации может оказаться неэффективной или даже привести к браку.

Пятое требование к параметру оптимизации - требование универсальности или полноты. Под универсальностью параметра оптимизации понимают его способность всесторонне охарактеризовать объект. В частности, технологические параметры недостаточно универсальны: они не учитывают экономику. Универсальностью обладают, например, обобщенные параметры оптимизации, которые строятся как функции от нескольких частных параметров.

11

Шестое требование: желательно, чтобы параметр оптимизации имел физический смысл, был простым и легко вычисляем.

Требование физического смысла связано с последующей интерпретацией результатов эксперимента. Не представляет труда объяснить, что значит максимум извлечения, максимум содержания ценного компонента.

Контрольные вопросы:

Вопросы

1 Что понимается под параметром оптимизации?

2 Что называется областью определения параметра оптимизации?

3 Какими могут быть области определения параметра оптимизации?

4 В каких случаях возможна количественная оценка параметра оптимизации и в чем она заключается?

5 Какие требования учитываются при выборе параметров оптимизации?

6 Что понимается под однозначностью параметра оптимизации в статистическом смысле?

7 Объясните требование к параметру оптимизации об эффективности оценивания функционирования системы.

8 Что подразумевается под универсальностью и полнотой параметра оптимизации?

9 Что означает физический смысл параметра оптимизации?

Рекомендуемая литература:

Основная литература:

1. Мелас В.Б., Шпилев П.В. Планирование и анализ для регрессионных моделей: учеб. пособие.- СПб.:Изд-во С.-Переб. ун-та, 2014.- 96 с.
2. Мусина О.Н. Планирование и постановка научного эксперимента: учебно-методическое пособие / О.Н. Мусина.- М.- Берлин:Директ-Медиа, 2015.- 88 с.
3. Боярский М.В. Планирование и организация эксперимента: учебное пособие / М.В. Боярский, Э.А. Анисимов.- Йошкар-Ола: Поволжский государственный технологический университет, 2015.- 168 с.

Дополнительная литература:

4. Сафин Р.Г. Основы научных исследований. Организация и планирование эксперимента: учебное пособие / Р.Г. Сафин, А.И. Иванов, Н.Ф. Тимербаев; Мин-во

образование и науки России, Казан. нац. исслед. технол. ун-т.-Казань: Изд-во КНИТУ, 2013.- 156 с.
5. Костин В.Н. Теория эксперимента: учебное пособие / В.Н. Костин, В.В. Паничев; Оренбургский государственный университет – Оренбург: ОГУ, 2013 .- 209 с.

Практическое занятие 13.

Тема занятия. Определение точности и надежности измерений

Цель занятия. Рассмотреть множественную линейную корреляцию.

Знания и умения, приобретаемые студентом в результате освоения темы, формируемые компетенции. Знает множественная линейная корреляцию и методы оценки уравнения регрессии. Владеет способностью к самоорганизации и самообразованию.

Актуальность темы. Элементы НИРС используются в прикладных задачах.

Теоретическая часть.

Для измерения тесноты связи при множественной корреляционной зависимости, то есть при исследовании трех и более признаков одновременно, вычисляется **множественный и частные коэффициенты корреляции.**

Множественный коэффициент корреляции вычисляется при наличии линейной связи между результативным и несколькими факторными признаками, а также между каждой парой факторных признаков.

Множественный коэффициент корреляции для двух факторных признаков вычисляется по формуле:

$$Ry/x_1x_2 = \sqrt{\frac{r^2{}_{yx_1} + r^2{}_{yx_2} - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r^2{}_{x_1x_2}}}$$, где

r_{yx} - парные коэффициенты корреляции между признаками.

Множественный коэффициент корреляции изменяется в пределах от 0 до 1 и по определению положителен: $0 \le R \le 1$.

Приближение R к единице свидетельствует о сильной зависимости между признаками.

$$r_{yx_1} = \frac{\overline{yx_1} - \overline{y} \cdot \overline{x_1}}{\delta_y \cdot \delta_{x1}} \qquad r_{yx_2} = \frac{\overline{yx_2} - \overline{y} \cdot \overline{x_2}}{\delta_y \cdot \delta_{x_2}} \qquad r_{x_1x_2} = \frac{\overline{x_1x_2} - \overline{x_1} \cdot \overline{x_2}}{\delta x_1 \cdot \delta x_2}$$

Частные коэффициенты корреляции характеризуют степень тесноты связи между двумя признаками x_1 и x_2 при фиксированном значении других (k - 2) факторных признаков, то есть когда влияние x_1 исключается, то есть оценивается связь между x_1 и x_2 в "чистом виде".

В случае зависимости у от двух факторных признаков x_1 и x_2 коэффициенты частной корреляции имеют вид:

$$r_{yx_1/x_2} = \frac{r_{yx_1} - r_{x_1x_2} \cdot r_{yx_2}}{\sqrt{(1-r^2_{x_2y}) \cdot (1-r^2_{x_1x_2})}}$$

$$r_{yx_2/x_1} = \frac{r_{yx_2} - r_{x_1y} \cdot r_{x_1x_2}}{\sqrt{(1-r^2_{x_1y}) \cdot (1-r^2_{x_1x_2})}},$$

где r – парные коэффициенты корреляции между указанными в индексе переменными.

В первом случае исключено влияние факторного признака x_2, во втором – x_1.

Вопросы

1. Множественная регрессия
2. Частные коэффициенты корреляции

Рекомендуемая литература:

Основная литература:

1. Мелас В.Б., Шпилев П.В. Планирование и анализ для регрессионных моделей: учеб. пособие.- СПб.:Изд-во С.-Перерб. ун-та, 2014.- 96 с.
2. Мусина О.Н. Планирование и постановка научного эксперимента: учебно-методическое пособие / О.Н. Мусина.- М.- Берлин:Директ-Медиа, 2015.- 88 с.
3. Боярский М.В. Планирование и организация эксперимента: учебное пособие / М.В. Боярский, Э.А. Анисимов.- Йошкар-Ола: Поволжский государственный технологический университет, 2015.- 168 с.

Дополнительная литература:

4. Сафин Р.Г. Основы научных исследований. Организация и планирование эксперимента: учебное пособие / Р.Г. Сафин, А.И. Иванов, Н.Ф. Тимербаев; Мин-во образование и науки России, Казан. нац. исслед. технол. ун-т.-Казань: Изд-во КНИТУ, 2013.- 156 с.
5. Костин В.Н. Теория эксперимента: учебное пособие / В.Н. Костин, В.В. Паничев; Оренбургский государственный университет – Оренбург: ОГУ, 2013 .- 209 с.

Практическое занятие 14.

Тема занятия. Проверка однородности дисперсий и расчет дисперсии воспроизводимости

Цель занятия. Рассмотреть множественную линейную корреляцию.

Знания и умения, приобретаемые студентом в результате освоения темы, формируемые компетенции. Знает множественная линейная корреляцию и методы оценки уравнения регрессии. Владеет способностью к самоорганизации и самообразованию.

Актуальность темы. Элементы НИРС используются в прикладных задачах.

Теоретическая часть.

Для измерения тесноты связи при множественной корреляционной зависимости, то есть при исследовании трех и более признаков одновременно, вычисляется **множественный и частные коэффициенты корреляции.**

Множественный коэффициент корреляции вычисляется при наличии линейной связи между результативным и несколькими факторными признаками, а также между каждой парой факторных признаков.

Множественный коэффициент корреляции для двух факторных признаков вычисляется по формуле:

$$Ry/x_1x_2 = \sqrt{\frac{r^2_{yx_1} + r^2_{yx_2} - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r^2_{x_1x_2}}}$$, где

r_{yx} - парные коэффициенты корреляции между признаками.

Множественный коэффициент корреляции изменяется в пределах от 0 до 1 и по определению положителен: $0 \le R \le 1$.

Приближение R к единице свидетельствует о сильной зависимости между признаками.

$$r_{yx_1} = \frac{\overline{yx_1} - \overline{y} \cdot \overline{x_1}}{\delta_y \cdot \delta_{x1}} \qquad r_{yx_2} = \frac{\overline{yx_2} - \overline{y} \cdot \overline{x_2}}{\delta_y \cdot \delta_{x_2}} \qquad r_{x_1x_2} = \frac{\overline{x_1 x_2} - \overline{x_1} \cdot \overline{x_2}}{\delta x_1 \cdot \delta x_2}$$

Частные коэффициенты корреляции характеризуют степень тесноты связи между двумя признаками x_1 и x_2 при фиксированном значении других (k - 2) факторных признаков, то есть когда влияние x_1 исключается, то есть оценивается связь между x_1 и x_2 в "чистом виде".

В случае зависимости y от двух факторных признаков x_1 и x_2 коэффициенты частной корреляции имеют вид:

$$r_{yx_1/x_2} = \frac{r_{yx_1} - r_{x_1x_2} \cdot r_{yx_2}}{\sqrt{(1-r^2_{x_2y}) \cdot (1-r^2_{x_1x_2})}}$$

$$r_{y\,x_2/x_1} = \frac{r_{yx_2} - r_{x_1y} \cdot r_{x_1x_2}}{\sqrt{(1-r^2_{x_1y}) \cdot (1-r^2_{x_1x_2})}},$$

где r – парные коэффициенты корреляции между указанными в индексе переменными.

В первом случае исключено влияние факторного признака x_2, во втором – x_1.

Вопросы

1. Множественная регрессия
2. Частные коэффициенты корреляции

Рекомендуемая литература:

Основная литература:

1. Мелас В.Б., Шпилев П.В. Планирование и анализ для регрессионных моделей: учеб. пособие.- СПб.:Изд-во С.-Переpб. ун-та, 2014.- 96 с.
2. Мусина О.Н. Планирование и постановка научного эксперимента: учебно-методическое пособие / О.Н. Мусина.- М.- Берлин:Директ-Медиа, 2015.- 88 с.
3. Боярский М.В. Планирование и организация эксперимента: учебное пособие / М.В. Боярский, Э.А. Анисимов.- Йошкар-Ола: Поволжский государственный технологический университет, 2015.- 168 с.

Дополнительная литература:

4. Сафин Р.Г. Основы научных исследований. Организация и планирование эксперимента: учебное пособие / Р.Г. Сафин, А.И. Иванов, Н.Ф. Тимербаев; Мин-во образование и науки России, Казан. нац. исслед. технол. ун-т.-Казань: Изд-во КНИТУ, 2013.- 156 с.
5. Костин В.Н. Теория эксперимента: учебное пособие / В.Н. Костин, В.В. Паничев; Оренбургский государственный университет – Оренбург: ОГУ, 2013 .- 209 с.

Практическое занятие 15.

Тема занятия. Проверка адекватности уравнения

Цель занятия. Рассмотреть множественную линейную корреляцию.

Знания и умения, приобретаемые студентом в результате освоения темы, формируемые компетенции. Знает множественная линейная корреляцию и методы оценки уравнения регрессии. Владеет способностью к самоорганизации и самообразованию.

Актуальность темы. Элементы НИРС используются в прикладных задачах.

Теоретическая часть.

Для измерения тесноты связи при множественной корреляционной зависимости, то есть при исследовании трех и более признаков одновременно, вычисляется **множественный и частные коэффициенты корреляции.**

Множественный коэффициент корреляции вычисляется при наличии линейной связи между результативным и несколькими факторными признаками, а также между каждой парой факторных признаков.

Множественный коэффициент корреляции для двух факторных признаков вычисляется по формуле:

$$Ry/x_1x_2 = \sqrt{\frac{r^2{}_{yx_1} + r^2{}_{yx_2} - 2r_{yx_1} \cdot r_{yx_2} \cdot r_{x_1x_2}}{1 - r^2{}_{x_1x_2}}}$$, где

r_{yx} - парные коэффициенты корреляции между признаками.

Множественный коэффициент корреляции изменяется в пределах от 0 до 1 и по определению положителен: $0 \le R \le 1$.

Приближение R к единице свидетельствует о сильной зависимости между признаками.

$$r_{yx_1} = \frac{\overline{yx_1} - \overline{y} \cdot \overline{x_1}}{\delta_y \cdot \delta_{x1}} \qquad r_{yx_2} = \frac{\overline{yx_2} - \overline{y} \cdot \overline{x_2}}{\delta_y \cdot \delta_{x_2}} \qquad r_{x_1x_2} = \frac{\overline{x_1 x_2} - \overline{x_1} \cdot \overline{x_2}}{\delta x_1 \cdot \delta x_2}$$

Частные коэффициенты корреляции характеризуют степень тесноты связи между двумя признаками x_1 и x_2 при фиксированном значении других (k - 2) факторных признаков, то есть когда влияние x_1 исключается, то есть оценивается связь между x_1 и x_2 в "чистом виде".

В случае зависимости y от двух факторных признаков x_1 и x_2 коэффициенты частной корреляции имеют вид:

$$r_{yx_1/x_2} = \frac{r_{yx_1} - r_{x_1x_2} \cdot r_{yx_2}}{\sqrt{(1 - r^2{}_{x_2y}) \cdot (1 - r^2{}_{x_1x_2})}}$$

$$r_{y\,x_2/x_1} = \frac{r_{yx_2} - r_{x_1y} \cdot r_{x_1x_2}}{\sqrt{(1 - r^2{}_{x_1y}) \cdot (1 - r^2{}_{x_1x_2})}},$$

где r – парные коэффициенты корреляции между указанными в индексе переменными.

В первом случае исключено влияние факторного признака x_2, во втором – x_1.

Вопросы

1. Множественная регрессия
2. Частные коэффициенты корреляции

Рекомендуемая литература:

Основная литература:

1. Мелас В.Б., Шпилев П.В. Планирование и анализ для регрессионных моделей: учеб. пособие.- СПб.:Изд-во С.-Перерб. ун-та, 2014.- 96 с.
2. Мусина О.Н. Планирование и постановка научного эксперимента: учебно-методическое пособие / О.Н. Мусина.- М.- Берлин:Директ-Медиа, 2015.- 88 с.
3. Боярский М.В. Планирование и организация эксперимента: учебное пособие / М.В. Боярский, Э.А. Анисимов.- Йошкар-Ола: Поволжский государственный технологический университет, 2015.- 168 с.

Дополнительная литература:

4. Сафин Р.Г. Основы научных исследований. Организация и планирование эксперимента: учебное пособие / Р.Г. Сафин, А.И. Иванов, Н.Ф. Тимербаев; Мин-во образование и науки России, Казан. нац. исслед. технол. ун-т.-Казань: Изд-во КНИТУ, 2013.- 156 с.
5. Костин В.Н. Теория эксперимента: учебное пособие / В.Н. Костин, В.В. Паничев; Оренбургский государственный университет – Оренбург: ОГУ, 2013 .- 209 с.

Практическое занятие 16

Тема занятия. Информационные технологии в экспериментальных исследованиях.

Цель занятия. Рассмотреть применение информационных технологии в экспериментальных исследованиях.

Знания и умения, приобретаемые студентом в результате освоения темы, формируемые компетенции. Знает технологию работы с литературой при выполнении НИРС. Владеет способностью к самоорганизации и самообразованию.

Актуальность темы. Элементы НИРС используются в прикладных задачах.

Теоретическая часть.

Информационные технологии (ИТ, от англ. *information technology*, *IT*) — широкий класс дисциплин и областей деятельности, относящихся к технологиям управления и обработки данных, в том числе, с применением вычислительной техники.

Сегодня под информационными технологиями чаще всего понимают *компьютерные технологии*. В частности, ИТ имеют дело с использованием компьютеров и программного обеспечения для хранения, преобразования, защиты, обработки, передачи и получения информации. Специалистов по компьютерной технике и программированию часто называют ИТ-специалистами.

Согласно определению, принятому ЮНЕСКО, ИТ — это комплекс взаимосвязанных, научных, технологических, инженерных дисциплин, изучающих методы эффективной организации труда людей, занятых обработкой и хранением информации; вычислительную технику и методы организации и взаимодействия с людьми и производственным оборудованием, их практические приложения, а также связанные со всем этим социальные, экономические и культурные проблемы. Сами ИТ требуют сложной подготовки, больших первоначальных затрат и наукоемкой техники. Их введение должно начинаться с создания математического обеспечения, формирования информационных потоков в системах подготовки специалистов.

Под информационными технологиями (ИТ) следует понимать использование вычислительной техники и систем связи для создания, сбора, обработки, хранения, передачи информации для всех сфер общественной жизни [2]. Информационные технологии включают в себя главные составляющие современного информационного бизнеса: компьютеры, терминалы, компьютерное оборудование, оптическую аппаратуру, микрофильмы, лазерные диски, печатное оборудование и ксерокопирование.

Признавая поразительные технологические достижения наступившей эры информационных технологий, многие специалисты спрогнозировали дальнейший прогресс в этой области. Движущей силой этого прогресса являются пять базовых тенденций:

1) возрастание роли информационного продукта;

2) развитие способности к взаимодействию (совместимости);

3) ликвидация промежуточных звеньев (непосредственность);

4) глобализация;

5) конвергенция.

Информационным технологиям принадлежит определяющая роль в научно-техническом развитии. Эти технологии позволяют оптимизировать самые разнообразные информационные процессы, обеспечивают информационное взаимодействие путем использования различных электронных телекоммуникаций.

Производство средств информатики и связи относится к наиболее наукоемким отраслям промышленности, мировой опыт свидетельствует об их динамичном развитии. В России информационное обеспечение экономики в настоящее время находится на низком уровне.

Электроника в современном обществе обеспечивает техническую базу и реализацию информационных систем. Основным направлением развития современной электроники стала микроэлектроника, обеспечивающая разработку, производство и применение в системах обработки информации полупроводниковых интегральных схем микропроцессоров, оперативной памяти, микроконтроллеров и функциональных узлов аппаратуры обработки информации. Дальнейшее развитие микроэлектроники связывается с непрерывным уменьшением размеров электронных приборов.

Комплексное решение проблем микро- и наноэлектроники требует параллельного и взаимного развития технологий, технологического оборудования, конструкций интегральных схем, методов их проектирования, математического моделирования технологических процессов, оборудования и приборов

В области микро- и наноэлектроники должна получить опережающее развитие выработка комплекса наукоемких технологий для создания элементной базы новых поколений, в т.ч. субмикронных сверхбольших интегральных схем для супермикропроцессоров и мегабайтных схем памяти.

В ходе реализации мероприятий в рамках этого направления будет обеспечено:

создание комплексов ЭВМ с новой архитектурой и на новых принципах, пакетов программ для систем искусственного интеллекта, программ для анализа и обработки изображений и моделей для оценки и прогнозирования крупномасштабных социально-экономических и технических проектов;

создание новых видов телекоммуникационных услуг и единой коммутируемой сети связи и передачи информации с пропускной способностью, превышающей существующую в 10—100 раз;

повышение качества управления объектами хозяйственной деятельности, ускорение процессов создания и эффективной эксплуатации сложных интегрированных информационных систем;

информационное обеспечение развития науки, техники, производства и социальной сферы на базе создаваемых государственных ресурсов научно-технической информации с использованием современных информационных технологий, баз данных и знаний;

развитие микроэлектроники и элементной базы радиоэлектроники, повышение уровня радиоэлектронной аппаратуры различного применения, создание приборов на квантовых и квантоворазмерных эффектах и осуществление перехода к атомарной сборке изделий электронной техники.

Вопросы

1. Информационные технологии: понятие
2. ИТ в экспериментальных исследованиях

Рекомендуемая литература:

Основная литература:

1. Мелас В.Б., Шпилев П.В. Планирование и анализ для регрессионных моделей: учеб. пособие.- СПб.:Изд-во С.-Перерб. ун-та, 2014.- 96 с.
2. Мусина О.Н. Планирование и постановка научного эксперимента: учебно-методическое пособие / О.Н. Мусина.- М.- Берлин:Директ-Медиа, 2015.- 88 с.
3. Боярский М.В. Планирование и организация эксперимента: учебное пособие / М.В. Боярский, Э.А. Анисимов.- Йошкар-Ола: Поволжский государственный технологический университет, 2015.- 168 с.

Дополнительная литература:

4. Сафин Р.Г. Основы научных исследований. Организация и планирование эксперимента: учебное пособие / Р.Г. Сафин, А.И. Иванов, Н.Ф. Тимербаев; Мин-во образование и науки России, Казан. нац. исслед. технол. ун-т.-Казань: Изд-во КНИТУ, 2013.- 156 с.
5. Костин В.Н. Теория эксперимента: учебное пособие / В.Н. Костин, В.В. Паничев; Оренбургский государственный университет – Оренбург: ОГУ, 2013 .- 209 с.

3. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

3.1. Рекомендуемая литература

3.1.1. Основная литература:

1. Мелас В.Б., Шпилев П.В. Планирование и анализ для регрессионных моделей: учеб. пособие.- СПб.:Изд-во С.-Перерб. ун-та, 2014.- 96 с.
2. Мусина О.Н. Планирование и постановка научного эксперимента: учебно-методическое пособие / О.Н. Мусина.- М.- Берлин:Директ-Медиа, 2015.- 88 с.
3. Боярский М.В. Планирование и организация эксперимента: учебное пособие / М.В. Боярский, Э.А. Анисимов.- Йошкар-Ола: Поволжский государственный технологический университет, 2015.- 168 с.

3.1.2. Дополнительная литература:

1. Сафин Р.Г. Основы научных исследований. Организация и планирование эксперимента: учебное пособие / Р.Г. Сафин, А.И. Иванов, Н.Ф. Тимербаев; Мин-во образование и науки России, Казан. нац. исслед. технол. ун-т.-Казань: Изд-во КНИТУ, 2013.- 156 с.
2. Костин В.Н. Теория эксперимента: учебное пособие / В.Н. Костин, В.В. паничев; Оренбургский государственный университет – Оренбург: ОГУ, 2013 .- 209 с.