

ІТМО

Планирование эксперимента, методы анализа и обработки данных

Лабораторная работа 3

Обработка результатов пассивного эксперимента.
Дисперсионный, корреляционный и регрессионный анализ

Цель лабораторной работы: Получить умения и отработать навыки дисперсионного, корреляционного и регрессионного .

Задачи лабораторной работы:

1. Провести пассивные эксперименты.
2. Обработать результаты пассивных экспериментов.
3. Оформить отчет по лабораторной работе.

Дисперсионный анализ применяется для исследования влияния одной или нескольких качественных переменных (факторов) на одну зависимую количественную переменную (отклик).

Корреляционный анализ – это группа статистических методов, направленная на выявление и математическое представление структурных зависимостей между выборками.

Регрессионный анализ - это статистический метод, устанавливающий количественно форму зависимости двух случайных величин, между которыми существует корреляционная связь.

Регрессионные процедуры позволяют рассчитать модель, описываемую некоторым уравнением и отражающую функциональную зависимость между экспериментальными количественными переменными, а также проверяют гипотезу об адекватности модели экспериментальным данным. По полученным результатам можно оценить природу и степень зависимости переменных и предсказать новые значения зависимой переменной.

Анализ временных рядов применим к одиночным или связанным временным рядам и позволяет выделять различные формы периодичности и взаимовлияния временных процессов, а также осуществлять **прогнозирование** будущего поведения временного ряда.

Кластерный анализ осуществляет разбиение объектов на заданное число удаленных друг от друга классов, а также строит дерево классификаций объектов посредством иерархического объединения их в группы (кластеры).

Основной задачей **факторного анализа** является нахождение в многомерном пространстве первичных переменных (значения которых регистрируются в эксперименте), сокращенной системы вторичных переменных (факторов).

Дисперсионный анализ – это анализ изменчивости признака под влиянием каких-либо контролируемых переменных факторов. **Дисперсионный анализ** является статистическим методом, предназначенным для выявления влияния ряда отдельных факторов на результаты экспериментов.

В основе дисперсионного анализа лежит предположение о том, что одни переменные могут рассматриваться как причины (факторы, независимые переменные), а другие как следствия (зависимые переменные). Независимые переменные называют иногда регулируемыми факторами именно потому, что в эксперименте исследователь имеет возможность варьировать ими и анализировать получающийся результат.

Сущность дисперсионного анализа заключается в расчленении общей дисперсии изучаемого признака на отдельные компоненты, обусловленные влиянием конкретных факторов, и проверке гипотез о значимости влияния этих факторов на исследуемый признак.

Исходным материалом для дисперсионного анализа служат данные исследования выборок, которые могут быть как **равными**, так и **неравными** по численности, как **связными**, так и **несвязными**.

По количеству выявляемых регулируемых факторов дисперсионный анализ может быть **однофакторным** (при этом изучается влияние одного фактора на результаты эксперимента), **двухфакторным** (при изучении влияния двух факторов) и **многофакторным** (позволяет оценить не только влияние каждого из факторов в отдельности, но и их взаимодействие).

Дисперсионный анализ относится к группе параметрических методов и поэтому его следует применять только тогда, когда доказано, что распределение является нормальным.

Почему дисперсионный анализ называется дисперсионным?

При дисперсионном анализе исследуется отношение двух дисперсий и определяют удельный вес суммарного воздействия одного или нескольких факторов	
Первая дисперсия (межгрупповая) – дисперсия объясненная влиянием фактора, которая характеризует рассеивание значений между градациями фактора (группами) вокруг средней всех данных	Вторая дисперсия (внутригрупповая) – дисперсия необъясненная , которая характеризует рассеивание данных внутри градаций (групп) вокруг средних значений самих групп
Отношение этих дисперсий называется фактическим отношением (критерием) Фишера и сравнивается с критическим значением отношения (критерием) Фишера	
Если $F_{расч} < F_{табл}$, то средние классов градации отличаются друг от друга и исследуемый фактор существенно влияет на изменение данных	Если $F_{расч} > F_{табл}$, то средние классов градации не отличаются друг от друга и фактор не имеет существенного влияния

Постановка задачи:

Нулевая гипотеза H_0 : $M_1 = M_2 = \dots = M_k$, где k - число классов градации и все классы градации имеют одно значение средних.

Альтернативная гипотеза \bar{H} : $M_1 \neq M_2 \neq \dots \neq M_k$, не все классы градации имеют одно значение средних (по крайней мере одно среднее значение отличается от остальных).

Однофакторный дисперсионный анализ

Основан на том, что сумму квадратов отклонений статистического комплекса возможно разделить на компоненты:

$$SS = SS_k + SS_e,$$

где $SS = \sum_i^k \sum_j^{n_i} (x_{ij} - \bar{x})^2$ - общая сумма квадратов отклонений;

$SS_k = \sum_j^{n_i} n_i (\bar{x}_i - \bar{x})^2$ - объясненная влиянием фактора K сумма квадратов отклонений;

$SS_e = \sum_i^k \sum_j^{n_i} (x_{ij} - \bar{x}_i)^2$ - необъясненная сумма квадратов отклонений или сумма квадратов отклонений ошибки;

$n = \sum_i^k n_i$; $\bar{x} = \frac{1}{kn} \sum_i^k \sum_j^{n_i} x_{ij} = \frac{1}{k} \sum_i^k \bar{x}_i$ - общее среднее наблюдений; $\bar{x}_i = \frac{1}{n_i} \sum_j^{n_i} x_{ij}$ - среднее наблюдений в каждой градации фактора K

Кроме того: $SS_e = \sum_i^k \sum_j^{n_i} (x_{ij} - \bar{x}_i)^2 = \sum_i^k (n_i - 1) S_i^2 = (n_1 - 1) S_1^2 + (n_2 - 1) S_2^2 + \dots + (n_k - 1) S_k^2$,

где S_i^2 - дисперсия градации фактора (группы).

Дисперсии рассчитываются следующим образом:

$S^2 = SS/v = SS/n-1$ - общая дисперсия;

$S_k^2 = SS_k/v_k = SS_k/k-1$ - объясненная дисперсия;

$S_e^2 = SS_e/v_k = SS_e/n-k$ - необъясненная дисперсия (дисперсия ошибки).

$$F_{\text{расч}} = \frac{S_k^2}{S_e^2} < F_{\text{табл}} = F_{\alpha; v_k; v_e}$$

Пример 1.

Дано: Для оценки работы трех отрезных станков проведены замеры на соответствия по длине полученных сортиментов установленным требованиям, при этом фиксировалась величина отклонения фактического размера сортимента от номинального размера. Результаты замеров представлены в таблице 7.1.

Таблица 7.1

i	$k = 3$	x_{ij}										n_i
		1	2	3	4	5	6	7	8	9	10	
1	Станок 1	0,4	0,5	0,5	0,3	0,4						5
2	Станок 2	0,5	0,4	0,3	0,6	0,8	0,5	0,7	0,6	0,7	0,4	10
3	Станок 3	0,6	0,4	0,7	0,5	0,4	0,6	0,5	0,8			8

Найти: Влияние различий в настройках станков на точность работы станков по отрезке сортиментов номинальной длины при уровне значимости $\alpha = 0,05$.

Решение:

i	$k = 3$	x_{ij}										n_i	n	\bar{x}_i	$\bar{\bar{x}}$	SS	SS_k	SS_e	S^2	S_k^2	S_e^2
		1	2	3	4	5	6	7	8	9	10										
1	Станок 1	0,4	0,5	0,5	0,3	0,4						5	23	0,42	0,53	0,46	0,07	0,39	0,0211	0,0363	0,0196
2	Станок 2	0,5	0,4	0,3	0,6	0,8	0,5	0,7	0,6	0,7	0,4	10		0,55							
3	Станок 3	0,6	0,4	0,7	0,5	0,4	0,6	0,5	0,8			8		0,56							

$$F_{\text{расч}} = 1,85 < F_{\text{табл}} = F_{0,05;2;20} = 3,49$$

Вывод: С вероятностью 0,95 можем принять, что настройка станков не влияет на точность их работы.

Лабораторная работа 3 - Microsoft Excel (Свой активации продукта)

Файл Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид

Из Интернета Из текста Из других источников Подключения Существующие подключения Обновить все Изменить связи Подключения Подключения

Сортировка Фильтр Дополнительно Сортировка и фильтр

Очистить Повторить Текст по столбцам Удалить дубликаты Проверка Консолидация Анализ "что если" Группировать Разгруппировать Промежуточный итог Структура

Отобразить детали Анализ данных Скрыть детали Анализ

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T
1	Пример 1													Однофакторный дисперсионный анализ						
2			1	2	3	4	5	6	7	8	9	10		ИТОГИ						
3	1	Станок 1	0,4	0,5	0,5	0,3	0,4							Группы	Счет	Сумма	Среднее	Дисперсия		
4	2	Станок 2	0,5	0,4	0,3	0,6	0,8	0,5	0,7	0,6	0,7	0,4		Станок 1	5	2,1	0,42	0,01		
5	3	Станок 3	0,6	0,4	0,7	0,5	0,4	0,6	0,5	0,8				Станок 2	10	5,5	0,55	0,03		
6														Станок 3	8	4,5	0,56	0,02		
7														Дисперсионный анализ						
8														Источники вариации	SS	df	MS	F	P-Значение	F критическое
9														Между группами	0,07	2	0,0363	1,85	0,18	3,49
10														Внутри групп	0,39	20	0,0196			
11														Итого	0,46	22				
12																				
13																				
14																				
15																				
16																				
17																				
18																				
19																				

Двухфакторный дисперсионный анализ без повтор... ? X

Входные данные
 Входной интервал: \$1:\$10
 Метки
 Альфа: 0,05

Параметры вывода
 Выходной интервал: \$F\$1
 Новый рабочий лист:
 Новая рабочая книга

Лист1 Пример 1 Пример 2

Укажите

17:06 31.03.2023

Двухфакторный дисперсионный анализ без повторений:

Двухфакторный дисперсионный анализ применяется для того, чтобы проверить возможную зависимость результативного признака от двух факторов - A и B . Тогда a - число градаций фактора A и b - число градаций фактора B . В статистическом комплексе сумма квадратов остатков разделяется на три компоненты:

$$SS = SS_a + SS_b + SS_e,$$

где $SS = \sum_i^a \sum_j^b (x_{ij} - \bar{x})^2$ - общая сумма квадратов отклонений;

$SS_a = b \sum_i^a (\bar{x}_i - \bar{x})^2$ - объясненная влиянием фактора A сумма квадратов отклонений;

$SS_b = a \sum_j^b (\bar{x}_j - \bar{x})^2$ - объясненная влиянием фактора B сумма квадратов отклонений;

$SS_e = \sum_i^a \sum_j^b (x_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2$ - необъясненная сумма квадратов отклонений или сумма квадратов отклонений ошибки;

$\bar{x} = \frac{1}{ab} \sum_i^a \sum_j^b x_{ij}$ - общее среднее наблюдений; $\bar{x}_i = \frac{1}{b} \sum_j^b x_{ij}$ - среднее наблюдений в каждой градации

фактора A ; $\bar{x}_j = \frac{1}{a} \sum_i^a x_{ij}$ - среднее наблюдений в каждой градации фактора B .

Дисперсии рассчитываются следующим образом:

$S^2 = SS/v = SS/ab-1$ - общая дисперсия;

$S_a^2 = SS_a/v_a = SS_a/a-1$ - дисперсия, объясненная влиянием фактора A ;

$S_b^2 = SS_b/v_b = SS_b/b-1$ - дисперсия, объясненная влиянием фактора B ;

$S_e^2 = SS_e/v_e = SS_e/(a-1)(b-1)$ - необъясненная дисперсия (дисперсия ошибки).

$$F_{\text{расч.}a} = \frac{S_a^2}{S_e^2} < F_{\text{табл}} = F_{\alpha;v_a;v_e}$$

$$F_{\text{расч.}b} = \frac{S_b^2}{S_e^2} < F_{\text{табл}} = F_{\alpha;v_b;v_e}$$

Таблица 10.1

		B – число степеней свободы		
		1-2	3-4	5-6
A – тип привода	электрический	0,1	0,3	0,5
	гидравлический	0,2	0,6	1,0
	пневматический	0,3	0,9	1,5

Пример 2.

Дано: Дана информация о точности движения РТК (промышленного робота) в зависимости от типа привода и числа степеней свободы (таблице 10.1).

Найти: Зависит ли точность движения РТК (промышленного робота) от типа привода и числа степеней свободы при уровне значимости $\alpha = 0,05$.

Решение:

		B – число степеней свободы			\bar{x}_i	SS	SS _a	SS _b	SS _e	S ²	S _a ²	S _b ²	S _e ²
		1-2	3-4	5-6									
A – тип привода	электрический	0,1	0,3	0,5	0,3	1,66	0,54	0,96	0,16	0,21	0,27	0,48	0,04
	гидравлический	0,2	0,6	1,0	0,6								
	пневматический	0,3	0,9	1,5	0,9								
\bar{x}_j		0,2	0,6	1,0	$\bar{\bar{x}} = 0,6$								

$$F_{\text{расч.}a} = 6,75 < F_{\text{табл}} = F_{0,05;2;4} = 6,94$$

$$F_{\text{расч.}b} = 12,00 > F_{\text{табл}} = F_{0,05;2;4} = 6,94$$

Вывод: С вероятностью 0,95 можем принять, что тип привода не влияет, а число степеней свободы влияет на точность движения РТК (промышленного робота).

Двухфакторный дисперсионный анализ с повторениями:

Двухфакторный дисперсионный анализ с повторениями применяется для того, чтобы проверить не только возможную зависимость результативного признака от двух факторов - A и B , но и возможное взаимодействие факторов A и B . Тогда a - число градаций фактора A и b - число градаций фактора B , r - число повторений. В статистическом комплексе сумма квадратов остатков разделяется на четыре компоненты:

$$SS = SS_a + SS_b + SS_{ab} + SS_e,$$

где $SS = \sum_i^a \sum_j^b \sum_k^r (x_{ijk} - \bar{x})^2$ - общая сумма квадратов отклонений;

$SS_a = br \sum_i^a (\bar{x}_i - \bar{x})^2$ - объясненная влиянием фактора A сумма квадратов отклонений;

$SS_b = ar \sum_j^b (\bar{x}_j - \bar{x})^2$ - объясненная влиянием фактора B сумма квадратов отклонений;

$SS_{ab} = r \sum_i^a \sum_j^b (x_{ij} - \bar{x}_i - \bar{x}_j + \bar{x})^2$ - объясненная влиянием взаимодействия факторов A и B сумма квадратов отклонений;

$SS_e = \sum_i^a \sum_j^b \sum_k^r (x_{ijk} - \bar{x}_{ij})^2$ - необъясненная сумма квадратов отклонений или сумма квадратов отклонений ошибки;

$\bar{x} = \frac{1}{abr} \sum_i^a \sum_j^b \sum_k^r x_{ijk}$ - общее среднее наблюдений;

$\bar{x}_i = \frac{1}{br} \sum_j^b \sum_k^r x_{ijk}$ - среднее наблюдений в каждой градации фактора A ;

$\bar{x}_j = \frac{1}{ar} \sum_i^a \sum_k^r x_{ijk}$ - среднее наблюдений в каждой градации фактора B ;

$\bar{x}_{ij} = \frac{1}{r} \sum_k^r x_{ijk}$ - среднее число наблюдений в каждой комбинации градаций факторов A и B .

Дисперсии рассчитываются следующим образом:

$$S^2 = SS/v = SS/abr-1 - \text{общая дисперсия};$$

$$S_a^2 = SS_a/v_a = SS_a/a-1 - \text{дисперсия, объясненная влиянием фактора } A;$$

$$S_b^2 = SS_b/v_b = SS_b/b-1 - \text{дисперсия, объясненная влиянием фактора } B;$$

$$S_{ab}^2 = SS_{ab}/v_{ab} = SS_{ab}/(a-1)(b-1) - \text{дисперсия, объясненная взаимодействием факторов } A \text{ и } B$$

$$S_e^2 = SS_e/v_e = SS_e/ab(r-1) - \text{необъясненная дисперсия (дисперсия ошибки).}$$

$$F_{\text{расч.}a} = \frac{S_a^2}{S_e^2} < F_{\text{табл}} = F_{\alpha;v_a;v_e}$$

$$F_{\text{расч.}b} = \frac{S_b^2}{S_e^2} < F_{\text{табл}} = F_{\alpha;v_b;v_e}$$

$$F_{\text{расч.}ab} = \frac{S_{ab}^2}{S_e^2} < F_{\text{табл}} = F_{\alpha;v_{ab};v_e}$$

Пример 3.

Дано: Выборочные данные о пластичности материала изготовленного на различных предприятиях, в который добавлялся пластификатор (элемент) с различным процентным составом и которые подвергались различным способам химической обработки, представлены в таблице 14.1.

Найти: При уровне значимости $\alpha = 0,05$ выяснить, влияют ли на пластичность материала способы химической обработки.

Таблица 14.1

Номер предприятия	А - вид элемента	В – способ химической обработки		
		Способ 1	Способ 2	Способ 3
1	Элемент 1	21,4	20,9	17,6
2		21,2	20,3	16,6
3		20,1	19,8	17,4
1	Элемент 2	12,1	13,6	13,3
2		14,2	13,3	14,1
3		12,3	12,6	13,8
1	Элемент 3	13,5	14,0	12,4
2		11,9	15,6	13,7
3		13,4	13,8	13,1

Задача 1. На химическом заводе разработаны два новых варианта технологического процесса. Чтобы оценить, как изменится дневная производительность при переходе на работу по новым технологиям, завод в течение 10 дней работает по каждому варианту, включая существующий. Дневная производительность завода (в условных единицах) представлена в таблице 16.1.

При уровне значимости $\alpha = 0,05$ требуется выяснить, как зависит дневная производительность от технологического процесса и оценить степень этой зависимости.

Таблица 16.1

<i>День работы</i>	<i>Дневная производительность</i>		
	<i>Существующая технология</i>	<i>Вариант 1</i>	<i>Вариант 2</i>
1	119	83	76
2	66	128	149
3	96	176	97
4	174	106	77
5	71	96	114
6	101	172	197
7	135	81	179
8	171	163	154
9	148	73	112
10	84	108	109

Задача 2. В процессе производства получается побочный нежелательный продукт, процентное содержание которого измеряется четырьмя различными катализаторами при трех уровнях температуры. Данные о процентном содержании побочного продукта представлены в таблице 17.1.

Требуется при уровне значимости $\alpha = 0,05$ определить, что оказывает более значимое влияние на выявление побочного продукта (катализатор или уровень температуры).

Таблица 17.1

<i>Уровень температуры</i>	<i>Катализатор 1</i>	<i>Катализатор 2</i>	<i>Катализатор 3</i>	<i>Катализатор 4</i>
<i>Высокий</i>	1,31	1,22	1,05	1,18
<i>Средний</i>	1,13	1,09	1,26	1,24
<i>Низкий</i>	1,26	1,04	1,14	1,06

Задача 3. Оценка уровня качества средств криптографической защиты информации (СКЗИ) осуществлялась методом тестирования на трех предприятиях. Для защиты информации использовались программно-аппаратных средств различных типов. При тестировании оценивался уровень (процент) нарушения безопасности информации при ее обработке СКЗИ (таблица 18.1).

При уровне значимости $\alpha = 0,05$ выяснить, влияние программных и аппаратных средств на защиту информации.

Таблица 18.1

		<i>Техническое средство 1</i>	<i>Техническое средство 2</i>	<i>Техническое средство 3</i>
<i>Предприятие 1</i>	<i>Программа 1</i>	10	9	10
<i>Предприятие 2</i>		8	12	14
<i>Предприятие 3</i>		4	8	8
<i>Предприятие 1</i>	<i>Программа 2</i>	9	6	14
<i>Предприятие 2</i>		6	14	9
<i>Предприятие 3</i>		12	11	8
<i>Предприятие 1</i>	<i>Программа 3</i>	18	7	11
<i>Предприятие 2</i>		14	12	15
<i>Предприятие 3</i>		7	9	6

Основными задачами корреляционного анализа являются выявление связи между переменными и оценка тесноты этой связи.

Корреляционная зависимость – это зависимость между одной случайной переменной и условным средним значением другой переменной.

В корреляционном анализе существуют следующие допущения:

- 1) переменные x и y являются случайными величинами, распределенными по нормальному закону;
- 2) некоторому значению x можно поставить в соответствие одно или несколько значений y ;
- 3) по данному нормальному распределению случайных величин x и y можно определить выборочное среднее и среднеквадратичное отклонение величин x и y .

$$\text{Коэффициент корреляции: } r_{xy} = \frac{cov_{xy}}{S_y S_x} = \frac{1}{n-1} \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{S_y S_x} = \frac{\sum_{i=1}^n (y_i - \bar{y})(x_i - \bar{x})}{\sqrt{\sum_{i=1}^n (y_i - \bar{y})^2 \sum_{i=1}^n (x_i - \bar{x})^2}}$$

$cov_{xy} = M[(x - M_x)(y - M_y)] = M_{xy} - M_x - M_y$ - **ковариация** (или **корреляционный момент**), которая является совместным центральным второго порядка и определяется как математическое ожидание отклонений произведения отклонений случайных величин.

Коэффициент детерминации (R^2) — это доля дисперсии зависимой переменной, объясняемая рассматриваемой моделью зависимости, то есть объясняющими переменными.

$$R^2 = 1 - \frac{S^2}{S_y^2} = 1 - \frac{S_{\text{уст}}^2}{S_y^2}$$

где S^2 ($S_{\text{уст}}^2$) - условная (по факторам x) дисперсия зависимой переменной (дисперсия ошибки модели).

S_y^2 - дисперсия случайной величины y .

Если использовать выборочную оценку значений соответствующих дисперсий, то получим формулу для выборочного коэффициента детерминации (который обычно и подразумевается под R^2):

$$R^2 = 1 - \frac{SS_{res}}{SS_{tot}}$$

Или (в смысле дисперсионного анализа) R^2 - это доля объяснённой суммы квадратов в общей: $R^2 = \frac{SS_{reg}}{SS_{tot}}$

SS_{res} - сумма квадратов остатков (отклонений ошибки) ;

SS_{reg} - объясненная сумма квадратов отклонений;

SS_{tot} - общая сумма квадратов отклонений.

Решение:

Лабораторная работа 3 - Microsoft Excel (Своей активации продукта)

Номер предприятия	А - вид элемента	В - способ химической обработки			ИТОГИ				
		Способ 1	Способ 2	Способ 3	Способ 1	Способ 2	Способ 3	Итого	
1	Элемент 1	21,4	20,9	17,6	Счет	3	3	3	9
2		21,2	20,3	16,6	Сумма	62,7	61	51,6	175,3
3		20,1	19,8	17,4	Среднее	20,90	20,33	17,20	19,48
1	Элемент 2	12,1	13,6	13,3	Дисперсия	0,49	0,30	0,28	3,25
2		14,2	13,3	14,1					
3		12,3	12,6	13,8					
1	Элемент 3	13,5	14	12,4	Счет	3	3	3	9
2		11,9	15,6	13,7	Сумма	38,6	39,5	41,2	119,3
3		13,4	13,8	13,1	Среднее	12,87	13,17	13,73	13,26
					Дисперсия	1,34	0,26	0,16	0,59
					Счет	3	3	3	9
					Сумма	38,8	43,4	39,2	121,4
					Среднее	12,93	14,47	13,07	13,49
					Дисперсия	0,80	0,97	0,42	1,09
					Итого				
					Счет	9	9	9	
					Сумма	140,1	143,9	132	
					Среднее	15,57	15,99	14,67	
					Дисперсия	16,66	11,32	3,91	

Двухфакторный дисперсионный анализ с повтор... ? X

Входные данные
Входной интервал: \$B\$2:\$E\$10
Число строк для выборки: 3
Дельта: 0,05

Параметры вывода
 Выходной интервал: \$G\$1
 Новый рабочий лист:
 Новая рабочая книга

Дисперсионный анализ						
Источники вариации	SS	df	MS	F	P-Значение	F критическое
Выборка	223,91	2	111,96	199,79	5,1382E-13	3,55
Столбцы	8,21	2	4,10	7,33	0,004703925	3,55
Взаимодействие	21,11	4	5,28	9,42	0,000273755	2,93
Внутри	10,09	18	0,56			
Итого	263,32	26				

Вывод: с вероятностью 0,95 можем принять, что:

- 1) вид пластификатора (элемента) влияет на пластичность материала;
- 2) способ химической обработки влияет на пластичность материала;
- 3) результативность различных видов пластификаторов (элементов) изменяется в зависимости от способа химической обработки.
- 4) выборочный коэффициент детерминации для А:

$$R^2 = \frac{SS_{\text{выборка}}}{SS_{\text{итого}}} = \frac{223,91}{263,32} = 0,85$$

показывает, что на 85% всей выборочной вариации пластичности материала влияет вид пластификатора.

В случае линейной зависимости R^2 является квадратом так называемого множественного коэффициента корреляции между зависимой переменной и объясняющими переменными.

В частности, для модели парной линейной регрессии $R^2 = r_{xy}^2$.

Метод вычисления коэффициента корреляции зависит от вида шкалы, к которой относятся переменные.

Ранговая корреляция — это взаимосвязь между ранжированием различных порядковых переменных или различными ранжированиями одной и той же переменной, где под «ранжированием» понимается присвоение меток порядка «первый», «второй» и т.д. к различным наблюдениям за конкретной переменной.

Коэффициент ранговой корреляции измеряет степень сходства между двумя рейтингами и может использоваться для оценки значимости связи между ними.

Задача 4. На основе результатов ультразвуковых исследований оценить степень корреляции следующих характеристик:

- U_0 - величина зондирующего ультразвукового импульса;
- U - величина отраженного ультразвукового импульса;
- c - скорость распространения ультразвуковой волны;
- t - время прихода ультразвукового сигнала;
- h - толщина объекта контроля.

При изучении статистической зависимости обычно ограничиваются исследованием усредненной зависимости (Как в среднем будет изменяться значение одной величины при изменении другой?). Такая зависимость называется **регрессионной**.

Более строго, **регрессионная зависимость между двумя случайными величинами** — это функциональная зависимость между значениями одной из них и условным математическим ожиданием другой.

На практике по экспериментальным данным можно найти только оценку (приближенное выражение) функции регрессии, т.е. по выборке определяют **выборочное (эмпирическое) уравнение регрессии**.

Как правило, до проведения эксперимента выбирают вид выборочной функции регрессии с точностью до нескольких параметров, а значения параметров определяют по выборке.

Вид эмпирической функции регрессии определяют исходя из:

- 1) теоретических соображений о физической сущности исследуемой зависимости;
- 2) опыта предыдущих исследований;
- 3) характера расположения точек на **корреляционном поле**, которое получается, если отметить на плоскости все точки с координатами $(x_i; y_i)$, соответствующие наблюдениям.

Наибольший интерес представляет линейное эмпирическое уравнение регрессии, поскольку:

- 1) это наиболее простой случай для расчетов и анализа;
- 2) при нормальном распределении модельная функция регрессии является линейной.

Спасибо
за внимание!

ITMO *re than a*
UNIVERSITY

Федоров Алексей Владимирович
avfedorov@itmo.ru