

**Государственное автономное образовательное учреждение
Астраханской области высшего образования
«Астраханский государственный архитектурно-строительный университет»**

Кафедра систем автоматизированного проектирования и моделирования

Отчет
по лабораторной работе №2

«ОДНОМЕРНАЯ СТАТИСТИЧЕСКАЯ МОДЕЛЬ»

Выполнил ст. группы ПГ-31-18

Кузнецов К.А.

(Ф.И.О.)

Преподаватель к.т.н., доцент

Евсина Е.М.

(Ф.И.О.)

	дата	подпись
отчет		

Лабораторная работа №2

Одномерная статистическая модель.

Цель работы:

- Расчёт статистических характеристик никеля.

Оборудование: 1. MatCAD

Ответы на контрольные вопросы

1. Реализация случайной величины

Одномерная статистическая модель применяется для изучения одного свойства. Пусть имеется система, состоящая из множества однородных геологических объектов. Выборочным методом возьмем из множества n объектов и у каждого из них измерим характеристику свойства x . Результаты измерений обозначим x_1, x_2, \dots, x_n и составим из них матрицу (1.1), в которой число строк равно n , а число столбцов $k = 1$.

В основе одномерной статистической модели лежат три гипотезы: а) измеренные значения x_1, x_2, \dots, x_n носят случайный характер; б) они не зависят друг от друга; в) значения образуют однородную совокупность. Измеренные значения принято называть *реализациями* случайной величины x .

2. Размах

Размах – это разность между максимальным x_{\max} и минимальным x_{\min} значениями свойства:

$$p = x_{\max} - x_{\min}$$

3. Медиана

Медиана – средний член упорядоченного ряда значений. Для нахождения медианы нужно расположить все значения в порядке возрастания или убывания и найти средний по порядку член ряда. В случае n – четного числа в середине ряда окажутся два значения, тогда медиана будет равна их полусумме.

$$M_{\text{мед}} = \frac{\sum_{i=1}^n x_i}{2}$$

4. Мода

Мода – наиболее часто встречающееся значение случайной величины. Методику ее нахождения мы рассмотрим позднее.

5. Среднее значение

Среднее значение – это среднеарифметическое из всех измеренных значений:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

Существуют другие виды средних (среднее взвешенное, среднее геометрическое, среднее гармоническое и др.), которые вычисляются в особых случаях и здесь не рассматриваются.

6. Дисперсия

Дисперсия – это число, равное среднему квадрату отклонений значений случайной величины от ее среднего значения:

$$\sigma^2 = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$$

7. Среднеквадратичное отклонение

Среднеквадратичное отклонение – это число, равное квадратному корню из дисперсии:

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$$

Среднеквадратичное отклонение имеет размерность, совпадающую с размерностью случайной величины и среднего значения. Например, если значения случайной величины измерены в метрах, то и среднеквадратичное отклонение также будет выражаться в метрах.

8. Коэффициент вариации

Коэффициент вариации – это отношение среднеквадратичного отклонения к среднему значению:

$$V = \frac{\sigma}{\bar{x}}$$

Коэффициент вариации выражается в долях единицы или (после умножения на 100) в процентах. Вычисление коэффициента вариации имеет смысл для положительных случайных величин

9. Ассиметрия

Ассиметрия – степень асимметричности распределения значений случайной величины относительно среднего значения,

$$A = \frac{1}{n \cdot \sigma^3} \sum_{i=1}^n (x_i - \bar{x})^3$$

10. Экцесс

Экцесс – степень остро- или плосковершинности распределения значений случайной величины относительно нормального закона распределения,

$$E = \frac{1}{n \cdot \sigma^4} \sum_{i=1}^n (x_i - \bar{x})^4 - 3$$

11. Математическое ожидание

Математическое ожидание случайной величины $M(x)$ – это ее среднее значение в генеральной совокупности. Оно, за редким исключением, бывает неизвестно, и приходится пользоваться его приближенной оценкой (точечной оценкой) – выборочным средним значением \bar{x} , определяемым по формуле. При увеличении числа наблюдений выборочное среднее стремится к пределу – к математическому ожиданию.

12. Дисперсия генеральной совокупности

Дисперсия генеральной совокупности $D(x)$ – это число, равное среднему квадрату отклонений случайной величины от ее математического ожидания. Если математическое ожидание известно, то дисперсию находят по формуле.

$$D(x) = \frac{1}{n} \sum_{i=1}^n [x_i - M(x)]^2$$

13. Моменты случайной величины, их связь со статистическими характеристиками

Вычисление статистических характеристик можно производить непосредственно по формулам, но на практике характеристики обычно находят с помощью моментов.

Моментом случайной величины k -го порядка относительно постоянного параметра a называется выражение

$$m_k = \frac{1}{n} \sum_{i=1}^n (x_i - a)^k$$

Порядок k может быть выражен любым целым числом, но интерес представляют первые четыре момента (порядка).

В зависимости от выбора параметра a различают начальные и центральные моменты. В первом случае a выбирается произвольно, что имеет смысл для ускорения вычислений. Часто полагают $a = 0$, и формула начальных моментов приобретает вид

$$m_k = \frac{1}{n} \sum_{i=1}^n x_i^k$$

Во втором случае принимают $a = \bar{x}$ и получают центральные моменты

$$\mu_k = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^k$$

От начальных моментов можно перейти к центральным

$$\mu_1 = 0$$

$$\mu_2 = m_2 - m_1^2$$

$$\mu_3 = m_3 - 3 m_2 m_1 - 2 m_1^3$$

$$\mu_4 = m_4 - 4 m_3 m_1 - 6 m_2 m_1^2 - 3 m_1^4$$

Зная моменты случайной величины, можно найти ее статистические характеристики по формулам

$$\bar{x} = m_1; \sigma^2 = \mu_2; \sigma = \sqrt{\mu_2};$$

$$V = \frac{\sigma}{\bar{x}}; S^2 = \frac{1}{n-1} \mu_2;$$

$$A = \frac{\mu_3}{\sigma^3}; E = \frac{\mu_4}{\sigma^4} - 3;$$

14. Группировка исходных данных

При большом числе исходных данных ($n > 50$) расчет статистических характеристик с помощью таблиц становится громоздким, поэтому применяется компактный метод расчета с предварительной группировкой данных. Для этого весь диапазон исходных значений от x_{\min} до x_{\max} разбивается на равные интервалы

(классы), границы которых удобно брать округленными, хотя это не имеет принципиального значения. С округленными границами удобнее работать.

Число классов зависит от числа исходных данных. Обычно принимается от 6 до 20 классов, но можно использовать и больше. Для определения числа классов рекомендуется эмпирическая формула $N_{\text{кл}} = 16[0,4\ln(n) - 1]$. Далее подсчитывают число исходных значений, попавших в каждый класс, и результаты сводят в таблицу.

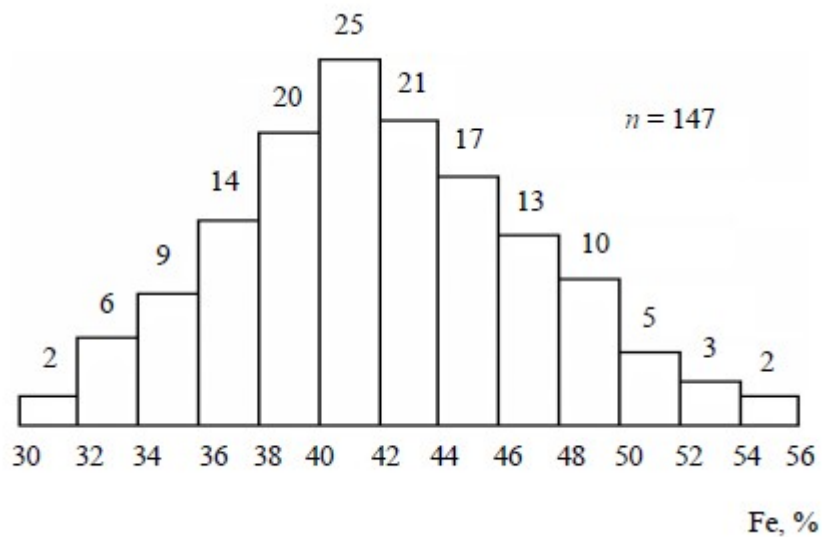
Некоторая трудность возникает в том случае, когда отдельные значения попадают на границу классов. Их можно относить в старший класс либо пытаться распределить примерно поровну между смежными классами.

Число значений в классе называется частотой. Если выразить частоту в относительных долях к общему числу значений, то получим частость. Ее можно выразить в процентах.

15. Построение гистограммы

Данные табл.2.4 позволяют построить гистограмму значений случайной величины (рис.2.1). По оси абсцисс откладывают классы, а по оси ординат – частоту или частость в виде ступенек. Для удобства обозрения над ступеньками выписана частота, а рядом с гистограммой указано суммарное значение n .

Гистограмма дает наглядное представление о поведении случайной величины. На ней видны размах и частота значений. Полезную информацию несет и форма гистограммы; она может быть симметричной и асимметричной, с одним, двумя и более максимумами частот.



Наличие нескольких максимумов свидетельствует о неоднородности изучаемой совокупности и позволяет ставить вопрос о выделении однородных совокупностей. В некоторых случаях отдельные частоты резко преобладают, это чаще всего связано с дефектами измерений. Например, при химическом анализе часто встречаются округленные значения и гораздо реже – промежуточные между ними. Чтобы устранить влияние подобных погрешностей, следует увеличить размер классов и построить гистограмму снова.