

Паспорт расчетно-графического задания

по дисциплине «Теория вероятностей и математическая статистика», 4 семестр

1. Методика оценки

В рамках расчетно-графического задания по дисциплине студенты статистический анализ и обоснование закономерностей в эмпирических данных в соответствии с исходными данными.

При выполнении расчетно-графического задания студенты должны провести анализ выборок, выбрать и обосновать способ статистических оценок параметров распределений и функций регрессии, проверить основные статистические гипотезы на основе собственно разработанных программ.

Обязательные структурные части РГЗ.

1. Статистический анализ и сравнение одномерных выборок.
2. Статистический анализ и построение регрессионных зависимостей эмпирических данных.

Оцениваемые позиции: качество статистических оценок и проверки статистических гипотез, эффективность программных средств анализа и визуализации данных, умение подготовки отчетных аналитических материалов.

2. Критерии оценки

Максимальное количество баллов 30 баллов.

- Работа считается **не выполненной**, если выполнены не все части РГЗ, а выполненные содержат ошибки, нет убедительного обоснования расчетных соотношений и основных результатов оценок и проверки статистических гипотез, характеристик программной реализации и качества результатов вычислений, оценка составляет до 5 баллов.
- Работа считается выполненной **на пороговом** уровне, если части РГЗ выполнены формально: задание выполнено в не полном объеме и без грубых ошибок, убедительного обоснования расчетных соотношений, основных оценок и проверки статистических гипотез мало убедительны, отсутствует четкий выбор характеристик программной реализации и качества результатов вычислений, оценка составляет 6-13 баллов.
- Работа считается выполненной **на базовом** уровне, если задание выполнено в полном объеме с без ошибок и с обоснованием расчетных соотношений и основных характеристик качества результатов вычислений и программ, оценка составляет 14-22 баллов.
- Работа считается выполненной **на продвинутом** уровне, если задание выполнено в полном объеме, с четким обоснованием расчетных соотношений и характеристик качества результатов вычислений, разработанные программные средства позволяют реализовывать высоко эффективные вычисления и вычислительные эксперименты для исследования свойств средств статистического анализа, оценка составляет 23-30 баллов.

3. Шкала оценки

В общей оценке по дисциплине баллы за РГЗ учитываются в соответствии с правилами балльно-рейтинговой системы, приведенными в рабочей программе дисциплины.

Максимальное значение текущего рейтинга складывается: КР до 30 баллов, РГЗ - до 30 баллов.

Итоговая экзаменационная оценка определяется: менее 50 баллов - неудовлетворительно; 50-72 (удовлетворительно), 73-87 (хорошо), 88-100 (отлично).

4. Примерный перечень тем РГЗ

ТЕМА: Применение статистических методов для анализа и обоснования закономерностей в эмпирических данных.

Часть 1. Статистический анализ и сравнение одномерных выборок

1. Цель и задачи части 1

Цель: Изучение и приобретение практических навыков применения основных методов математической статистики для определения законов распределения, оценок характеристик и проверки статистических гипотез о свойствах одномерных выборок.

Задачи:

- Предварительной обработки и графического представления статистических данных.
- Выбора необходимых методов и средств для вычисления основных выборочных характеристик и оценок параметров распределения одномерных выборок.
- Решения прикладных задач на основе проверки статистических гипотез.

2. Общее описание задания 1

При выполнении задания 1 необходимо провести обработку статистических данных с целью получения ответа на те вопросы, которые были поставлены перед статистиком-аналитиком в каждом конкретном случае (в соответствии с вариантом задания).

Исходные данные представлены результатами наблюдений в виде простой не сгруппированной выборки или в виде интервального вариационного ряда (табл.1)

Таблица 1

X_j	X_1	X_2		X_k
N_j	n_1	n_2		n_k

По данным наблюдений таблиц или исходных выборок необходимо:

1. Провести предварительную обработки и графически представить статистические данные задания
 - 1.1. Вычислить статистики (оценки) положения, рассеяния; коэффициенты формы распределения.
 - 1.2. Провести идентификацию статистической модели и подобрать аппроксимирующие кривые распределения, используя нормальный и другие законы распределения, например, равномерный и/или показательный.
 - 1.3. Проанализировать исходные данные и результаты расчетов, сделать предварительные (интуитивные) выводы о свойствах выборок задания.
2. Осуществить поиск ответов на вопросы прикладных задач на основе проверки статистических гипотез
 - 2.1. Количественными методами оценить качество и достоверность предварительные выводы о свойствах выборок.
 - 2.2. Сформулировать и провести проверку статистических гипотез об оценках параметров нормального и согласия эмпирических распределений теоретическим, достаточных для ответов на содержательные (прикладные) вопросы задания.
 - 2.3. Провести анализ результатов расчетов, сделать выводы о проделанном статистическом анализе и

сравнении одномерных выборок на основе закономерностей эмпирических данных.

3. Вариант задания

Процесс извлечения гелия.

Исследуется работа промышленных агрегатов по процессу извлечения гелия (He) из природного Оренбургского газа. Целью работы установки является более полное (близкое к 100%), извлечение гелия. Технологический процесс считается отлаженным, если в готовом продукте (баллоне с гелием) содержится не менее 99% He (или 0,99 He и только 0,01 примесей); качество продукта считается очень хорошим, если извлечено более 99,8% He. Испытываются два технологических режима N1 и N2 для того, чтобы выбрать лучший по признаку наибольшего процента извлечения гелия. Кроме того, необходимо повысить точность измерений, поэтому испытанию подвергаются два способа измерений: А и В. Еще одной целью исследования является установление стандарта на процент гелия (интервальная оценка математического ожидания), который может быть задан для изучаемого производственного процесса.

Результаты наблюдений в различных ситуациях представлены таблицами 1.2 - 1.5 (данные представлены в абсолютных единицах).

Таблица 1.2

Технология N1, способ А; N₁ = 210

%He; X _j	98,50	98,75	98,80	98,93	98,99	99,10	99,20	99,32	99,40	99,55	99,65	99,90	99,98
n _j	10	10	15	20	25	30	20	20	15	15	5	15	10

Таблица 1.3

Технология N2, способ В; N₂ = 120

%He; X _j	98,30	98,50	98,72	98,91	99,00	99,15	99,2	99,50	99,72	99,85	99,86
n _j	2	2	4	10	6	10	24	30	26	4	2

Таблица 1.4

Технология N1, способ В; N₃=42

%He; X _j	0,9843	0,9850	0,9871	0,9882	0,9922	0,9954	0,9973	0,9992
n _j	1	2	10	6	12	6	4	1

Таблица 1.5

Технология N2, способ А; N₄=25

%He; X _j	0,9862	0,9883	0,9900	0,9945	0,9962	0,9971	0,9984	0,9990	0,9993
n _j	1	3	3	5	6	3	2	1	1

Сформулируйте и проверьте статистические гипотезы, на основании которых можно выяснить:

- отличаются ли технологические режимы N1 и N2 и если да, то какой из них лучше?
- отличаются ли способы измерений А и В, какой из них точнее?
- какой стандарт можно установить на процентное содержание гелия?
- какое число измерений достаточно для того, чтобы определить, соответствует ли продукция стандарту?

Условия выполнения вариантов 1.2-1.5 (табл.1.6).

Таблица 1.6

Условия	1.1	1.2	1.3	1.4
Уровень значимости α	0,01; 0,05	0,005; 0,05	0,02; 0,05	0,025; 0,05
Точность вычислений (число знаков после запятой)	3	4	3	4

Часть 2. Статистический анализ и построение регрессионных зависимостей эмпирических данных

1. Цель и задачи части 2.

Цель: Изучение и приобретение практических навыков применения основных приемов регрессионного анализа для создания и применения математической (факторной) модели эмпирических многомерных данных в виде функциональной зависимости.

Задачи:

- Предварительной обработки и графического представления многомерных статистических данных для определения факта существования функциональной зависимости в исходных данных.
- Выбора вида и определение параметров функциональной (регрессионной) зависимости.
- Проверка адекватности и качества оценок параметров модели.
- Отбор для прогнозирования лучшей модели регрессионной зависимости из заданного класса.
-

Замечание. Важно иметь в виду, что "анализ и построение функциональной зависимости по эмпирическим данным" – обычно это следующий этап статистического исследования после "анализа одномерных выборок".

2. Общее описание задания 2

Объект исследования может быть представлен факторной моделью (рис.1).

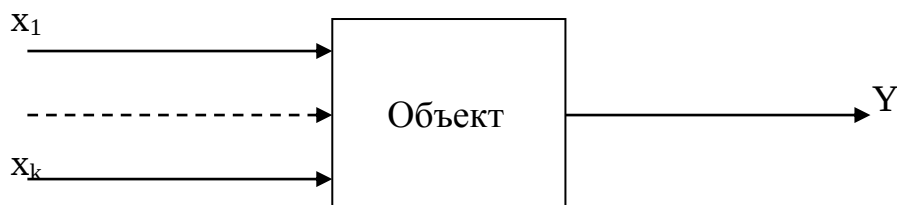


Рис. 1. Объект исследования

На рис. 1: $X = \{x_j\}$, $j=1..k$ - входы объекта (независимые факторы, заданы в относительных или абсолютных единицах измерений); Y - выход объекта (некоторый технический, производственный, экономический, социальный, экологический и т.п. показатель функционирования, задан в абсолютных единицах).

Объект подвергается сериям испытаний или наблюдению в процессе его естественного функционирования. В результате накапливаются данные по $i=1..N$ опытам: $X = \{x_{ij}\}$ и соответствующие им Y_i (табл.2).

Таблица 2

№ опыта	Факторы	Значение показателя Y
1	$x_{1,1}...x_{1,k}$	Y_1
·	· ... ·	·
·	· ... ·	·
N	$x_{N,1}...x_{N,k}$	Y_N

Целью наблюдений является построение по данным таблицы 2 регрессионной модели

$$\hat{Y} = f(\vec{B}, \vec{X}),$$

где $f(B, X)$ - функция отклика, чаще всего имеющая вид полинома:

$$\hat{Y} = b_0 + \sum_{j=1}^k b_j x_j + \sum_{\substack{j,i=1 \\ j \neq i}}^k b_{ij} x_i x_j + \sum_{j=1}^k b_{jj} x_j^2 + \dots,$$

b - оценки коэффициентов β "истинной" регрессии:

$$\hat{Y} = b_0 + \sum_{j=1}^k \beta_j x_j + \sum_{\substack{j,i=1 \\ j \neq i}}^k \beta_{ij} x_i x_j + \sum_{j=1}^k \beta_{jj} x_j^2 + \dots$$

Регрессионные модели могут быть построены и использованы, если выполнены постулаты регрессионного анализа (условия Гаусса-Маркова). Для проверки выполнения постулатов используются данные табл.3.

Таблица 3

Фиксированные наборы значений $X=\{x_j\}$	Значение показателя Y при $j=1..m$ опытах
$x_1 = \{x_{1,1}, x_{1,2}, \dots, x_{1,k1}\}$ $x_2 = \{x_{2,1}, x_{2,2}, \dots, x_{2,k2}\}$ $\dots$	$y_1 = \{y_{1,1}, y_{1,j}, \dots, y_{1,m1}\}$ $y_2 = \{y_{2,1}, y_{2,j}, \dots, y_{2,m2}\}$ $\dots$

Регрессионная модель в зависимости от входящих в нее факторов x_j может быть использована:

- для управления объектом;
- для прогнозирования его состояния;
- для объяснения физики явлений.

Прежде чем использовать модель, необходимо провести ее анализ: проверить адекватность модели, значимость коэффициентов регрессии, выбрать наилучшее уравнение регрессии, дать физическую интерпретацию модели и показать возможности ее применения для решения прикладных задач.

3. Варианты задания

Процесс извлечения гелия.

Исследуется работа промышленных агрегатов по процессу извлечения гелия (He) из природного Оренбургского газа. Целью исследования является определение зависимости выходного показателя Y (% извлечения гелия, %He), от технологических переменных X_1 (температура, °C) и X_2 (давление, атм.). В реальных условиях процент извлечения гелия зависит от сотен технологических факторов. В настоящей работе исследуется одна из простейших производственных ситуаций. Из многих переменных выбраны только две, используемые на последних этапах процесса извлечения. Данные наблюдений приведены в таблице. Величины Y заданы в абсолютных единицах, X_2 и X_1 - в относительных (абсолютные значения: $X_1 \in [-250^\circ\text{C}; -230^\circ\text{C}]$; $X_2 \in [4,4 \text{ атм.}; 5,0 \text{ атм.}]$). Определенные наборы значений X_1 и X_2 при фиксированных значениях других параметров соответствуют различным технологиям производства. Предполагается, что измерения достаточно точны и погрешность невелика. Необходимо найти зависимость $Y = f(X_1, X_2)$, оцениваемую результатами наблюдений, и определить значения X_1 и X_2 , обеспечивающие заданный номинал $Y_{\text{ном.}} = 99,8; 99,5; 99,9; 99,85$. Определить точность ε , с которой достигается это значение.

Таблица 4

№	X_{1i}	X_{2i}	Y_i
1.	3	7	99,25
2.	3	7	99,35
3.	3	7	99,10
4.	6	3	98,50
5.	6	3	98,70
6.	5	6	99,00
7.	7	4	98,50
8.	7	4	98,00
9.	7	4	98,40
10.	4	6	98,75
11.	5	5	98,75
12.	5	5	98,60
13.	5	5	98,55
14.	3	8	99,75
15.	3	8	99,60

16.	10	1	97,50
17.	1	10	99,98
18.	1	10	99,50
19.	8	2	97,80
20.	2	9	99,85
21.	2	9	99,70
22.	9	1	97,75
23.	9	1	98,25
24.	9	1	97,90
25.	2	10	99,95
26.	2	10	99,85
27.	6	8	99,25
28.	6	8	99,35
29.	3	6	99,20
30.	3	6	99,40