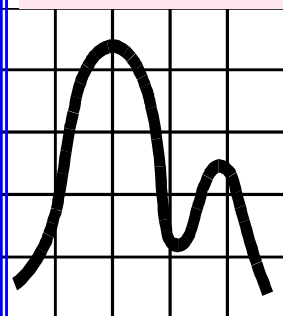


**Донецкий национальный университет
Экономический факультет**

Породников В.Д.

Кафедра
Математики и
Математических
Методов в
Экономике



StatLab

**Рекомендовано
Министерством образования
и науки Украины
в качестве учебного пособия**

© Донецкий национальный университет, 2000

© Породников В.Д., 2000

Введение

У становление закономерностей, которым подчиняются массовые случайные явления, основано на изучении статистических данных - сведений о том, какие значения принял интересующий признак. В зависимости от исходных данных существует два подхода к их статистической обработке. Первый основан на вероятностной интерпретации исходных данных, а второй - не опирается на вероятностную природу обрабатываемых данных. Поэтому прежде всего необходимо определить какова природа исходных данных, если основываться на вероятностной природе исходных данных тогда, в качестве инструмента обработки статистических данных применяются вероятностно-статистические методы. На этих моделях мы и сосредоточим наше внимание.

Настоящее пособие позволит студентам понять как применять математическую теорию на практике. Каждая лабораторная работа состоит из нескольких задач, которые предваряет краткое изложение необходимого теоретического материала и приводятся необходимые формулы, требуемые для решения задач. Теория сопровождается конкретным примером выполнения лабораторной работы. В конце заданы варианты выборок для решения задач, а так же приложения различных распределений случайных величин.

Лабораторный практикум представляет собой четыре лабораторных задания по курсу математическая статистика, которые выполняются на основе теоретического материала. Вычисления целесообразно проводить, используя электронные таблицы и встроенные функции на ЭВМ. Решения задач необходимо оформлять в специальной тетради или в виде сброшюрованных распечаток, рисунки выполнять на миллиметровой бумаге или при помощи графических пакетов прикладных программ.

Лабораторная работа №1.

ТЕМА: Первичная обработка статистических данных.

Так как основная задача математической статистики состоит в получении выводов о массовых явлениях и процессах по данным наблюдений над ними или экспериментов. Полное исследование генеральной совокупности обычно практически невозможно или не экономно. Поэтому всеобщее исследование применяют крайне редко. Будем предполагать, что мы располагаем исходными статистическими данными, характеризующими те или иные стороны интересующего нас процесса или явления.

Постановка задачи.

1. Отобразить переменные;
2. Графическое представление исходных данных.
3. Произвести статистическое описание исходных данных, с определением варьирования переменных;
4. Анализ резко выделяющихся наблюдений;
5. Восстановление пропущенных исходных данных, если это необходимо;
6. Графики вариационных рядов;
7. Экспериментальный анализ закона распределения исследуемой генеральной совокупности и параметризация сведений о природе изучаемых распределений;
8. Числовые характеристики выборки.

Пример выполнения лабораторной работы:

В таблице приведены значения X - заработной платы (в условных единицах) рабочих-сдельщиков. Другими словами дана одномерная выборка заработной платы 100 рабочих.

Исходные данные

Таблица №1.

4.2781	3.1645	1.8657	2.4554	6.1401	3.1535	7.5917	3.408	1.1001	5.7285
3.5566	2.659	3.5202	5.085	4.8022	5.6521	2.5561	3.1291	2.2626	4.2695
5.0908	3.6177	3.4169	2.1926	5.7115	4.7699	5.4602	2.9712	3.1363	4.9042
0.5653	3.5348	4.4335	1.8655	2.0584	2.3136	2.4538	3.4256	1.8667	0.8356
3.6185	3.6528	2.5759	1.6058	3.4139	2.4861	0.0044	3.3833	3.3956	0.2852
3.7213	3.2617	3.0226	4.5737	2.1159	1.879	1.7298	0.9983	2.1602	3.0121
3.0742	5.158	3.6585	0.0758	2.4228	3.9749	2.0997	1.3595	3.0733	2.6159
2.495	2.6765	1.0106	4.3656	2.5593	3.7782	5.6209	1.5937	0.8298	2.5329
1.5209	1.4857	2.958	3.2787	4.4941	4.3576	2.8779	3.0059	3.9802	7.2719
3.9614	4.7845	0.451	3.1273	3.5688	4.4303	0.7919	1.7188	4.7897	3.3967

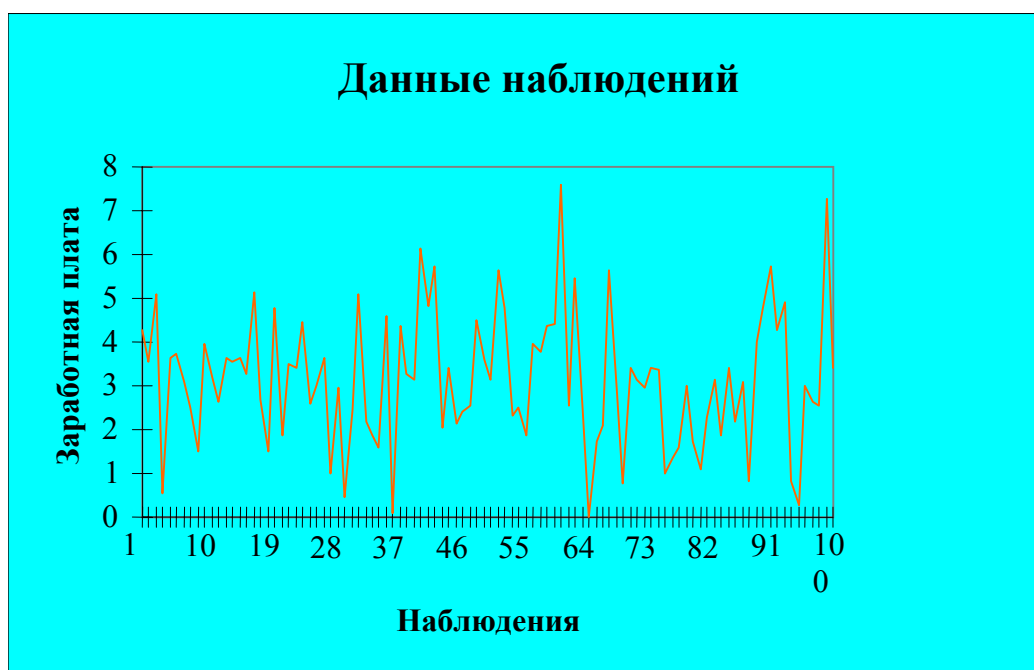


Рис. 1.1.

Выборка будет намного наглядней, если построить вариационный ряд т.е. ее элементы упорядочить по возрастанию или убыванию.

Таблица №2.

0.0044	1.1001	1.8667	2.4538	2.6765	3.1291	3.4139	3.6585	4.4303	5.0908
0.0758	1.3595	1.879	2.4554	2.8779	3.1363	3.4169	3.7213	4.4335	5.158
0.2852	1.4857	2.0584	2.4861	2.958	3.1535	3.4256	3.7782	4.4941	5.4602
0.4509	1.5209	2.0997	2.495	2.9712	3.1645	3.5202	3.9614	4.5737	5.6209
0.5653	1.5937	2.1159	2.5329	3.0059	3.2617	3.5348	3.9749	4.7699	5.6521
0.7919	1.6058	2.1602	2.5561	3.0121	3.2787	3.5566	3.9802	4.7845	5.7115
0.8298	1.7188	2.1926	2.5593	3.0226	3.3833	3.5688	4.2695	4.7897	5.7285
0.8356	1.7298	2.2626	2.5759	3.0733	3.3956	3.6177	4.2781	4.8022	6.1401
0.9983	1.8655	2.3136	2.6159	3.0742	3.3967	3.6185	4.3576	4.9042	7.2719
1.0106	1.8657	2.4228	2.659	3.1273	3.408	3.6528	4.3656	5.085	7.5917

Для составления вариационного ряда по интервалам значений (если количество вариантов слишком велико) нужно:

1. Определить размах выборки

$$\omega = \max_i(x_i) - \min_i(x_i) = 7.5917 - 0.0044 = 7.5873,$$

Определим длину интервала, используя формулу Стэрджеса:

$$h = \frac{x_{\max} - x_{\min}}{1 + 3.322 \lg n} \quad (1)$$

$$h = 7.5873 / (1 + 3.322 * \lg 100) = 0.993 \approx 1.$$

Если h оказывается дробным числом, то за величину интервала следует взять либо ближайшее целое число, либо несложную дробь.

За начало первого интервала рекомендуется принимать величину, равную $a_1 = (x_{\min} - h/2)$, $a_2 = a_1 + h$, $a_3 = a_2 + h$ и т.д. до тех пор, пока начало следующего по порядку интервала не будет равным или большим $x_{\max}$.

После установления шкалы интервалов можно приступить к группировке результатов наблюдений.

Здесь удобно за начало первого интервала удобно взять $a_1 = 0$ так как все $x_i > 0$. Тогда $a_2 = 1$; $a_3 = 2$ и т.д. (обычно начало интервала входит в интервал, а его конец - не входит).

2. Составим вариационный ряд по интервалам значений:

Таблица №3.

Интервал	0 - 1	1 - 2	2 - 3	3 - 4	4 - 5	5 - 6	6 - 7	7 - 8
n_i	9	13	22	32	13	8	1	2
n_i/n	0.09	0.13	0.22	0.32	0.13	0.08	0.01	0.02

В результате вместо многочисленных отдельных записей получается вполне обозримая статистическая таблица. Однако, следует помнить, что в зависимости от конкретного содержания задачи в приведенную схему группировки данных могут быть внесены некоторые изменения (например, длины интервалов могут быть разной длины, а также желательно чтобы количество интервалов было не меньше 8 - 10 и больше 20 - 25); выбор количества интервалов (s) зависит от объёма выборки n , можно для ориентировки пользоваться приближенной формулой:

$$s \approx \log_2 n + 1,$$

которую следует воспринимать как оценку снизу для больших n .

Для нашего примера имеем $s \approx \log_2 100 + 1 = 7.6439 \approx 8$, что вполне согласуется с (1).

3. Подсчитаем количество элементов в выборке (объём выборки - n), которое равно

$$n = \sum_{i=1}^m n_i = \sum_{i=1}^8 n_i = 9+13+22+32+13+8+1+2=100, \quad (2)$$

где m - количество вариантов в вариационном ряде, а n_i называется частотой варианта. Если (2) не выполнено, то повторить все пункты.

Очень важно, чтобы собранные данные были тщательно просмотрены и отредактированы прежде, чем к ним будет применена основная статистическая техника.

При вводе данных в ЭВМ необходимо позаботиться о том, чтобы их можно было легко редактировать или пополнить, а также возможность обнаружения аномальных, т.е. резко выделяющихся по своей величине наблюдений, которые могут быть нерепрезентативными для изучаемой популяции.

Иногда в таблице №3 помещают не только частоту варианта, а и долю её в сумме всех частот (w_i), которая равна отношению частоты (n_i) к общему числу наблюдений (n), т.е.

$$w_i = n_i / n \quad (3)$$

Такая величина называется *частотой*.

На ряду с понятием частоты и частости используют понятие накопленной частоты и накопленной частости, которые соответственно равны:

$$n_i^* = \sum_{i < x_{i+1}} n_i \quad \text{и} \quad w_i^* = \sum_{i < x_{i+1}} w_i \quad (4)$$

Замечание. В дискретном вариационном ряду накопленные частоты (частости) вычисляются для каждого варианта и являются последовательного суммирования частот (частостей), начиная с частоты (частости) первого варианта.

4. Графическое представление вариационных рядов.

Графическое изображение вариационного ряда представляет в наглядной форме закономерности варьирования значений признака. Рассмотрим некоторые из них, наиболее широко используемые: полигон, гистограмма, кумулятивная кривая и огива.

Полигон, как правило, служит для изображения дискретного вариационного ряда. Для его построения в прямоугольной системе координат нанесем точки с координатами (x, n_x) , где x - вариант, а n_x - соответствующая ему частота. Если вариационный ряд составлен по интервалам, то в качестве значений x_i следует рассматривать середины интервалов. Затем эти точки соединяют последовательно отрезками. Полученная ломаная линия называется полигоном.

Таблица №4.

X_i^*	0.5	1.5	2.5	3.5	4.5	5.5	6.5	7.5
n_i	9	13	22	32	13	8	1	2
n_i/n	0.09	0.13	0.22	0.32	0.13	0.08	0.01	0.02



Рис. 1.2.



Рис. 1.3.

Гистограмма служит для изображения интервального вариационного ряда. Для её построения в прямоугольной системе координат по оси абсцисс откладывают интервалы варьирования, и на этих отрезках, как на основании, строят прямоугольники с высотами, равными частотам (или частостям) соответствующего интервала. В результате получается фигура, состоящая из прямоугольников, которая называется гистограммой.

Если по оси абсцисс взять масштаб с единицей равной ширине интервала, а по оси ординат единицу, которая соответствует одному наблюдению, то очевидно, что площадь гистограммы равна общему числу наблюдений, если по оси ординат отложить частости-получим площадь равную единице.

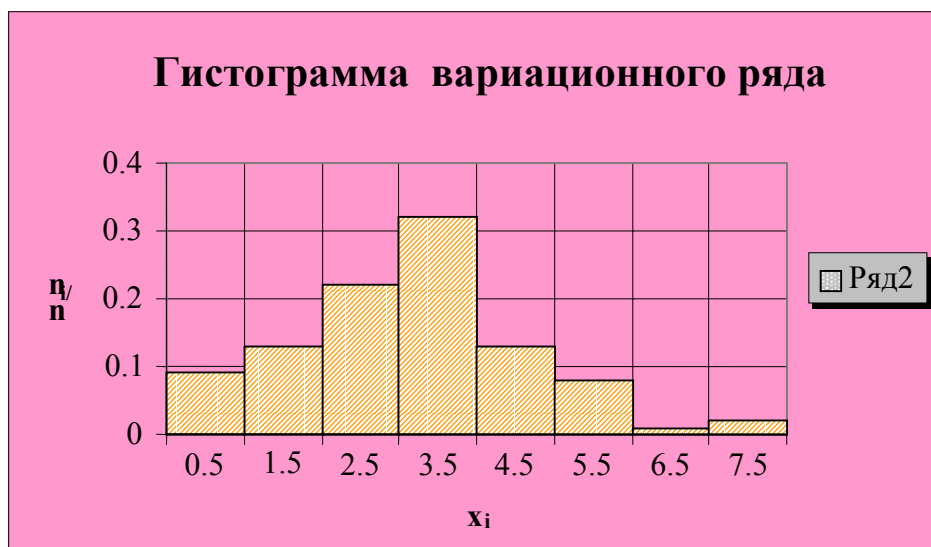


Рис. 1.4.

По вариационному ряду, в котором нет совпадающих точек, строится *эмпирическая функция распределения*

$$F_n^*(x) = \sum_{x_i < x} \frac{n_i}{n} \quad (5)$$

Эмпирическая функция распределения обладает всеми свойствами теоретической функции распределения и имеет вид ступенчатой функции со скачками, равными по величине $1/n$, в точках вариационного ряда.

$$F^*(x) = \begin{cases} 0, & x \leq x_1; \\ \frac{n_1}{n}, & x_1 < x \leq x_2; \\ \frac{n_1 + n_2}{n}, & x_2 < x \leq x_3; \\ \dots\dots\dots \\ \sum_{i=1}^{m-1} \frac{n_i}{n}, & x_{m-1} < x \leq x_m; \\ 1, & x > x_m. \end{cases}$$

Процесс нахождения эмпирической функции распределения $F^*(x)$ аналогичен процессу нахождения функции распределения $F(x)$ дискретной случайной величины, где на основании закона больших чисел Бернулли вместо вероятностей p_i , берутся относительные частоты n_i/n . Т.е. свойство статистической устойчивости относительных частот является основанием использования $F^*(x)$ и эта оценка

теоретического распределения называется *непараметрической* так как не связана с выбором общего модельного вида функции $F(x)$.

Таблица №5.

x_i	0.5	1.5	2.5	3.5	4.5	5.5	6.5	7.5
n_i/n	0.09	0.13	0.22	0.32	0.13	0.08	0.01	0.02

$$F^*(x) = \begin{cases} 0, & x \leq 0.5; \\ 0,09, & 0.5 < x \leq 1.5; \\ 0,22, & 1.5 < x \leq 2.5; \\ 0,44, & 2.5 < x \leq 3.5; \\ 0,76, & 3.5 < x \leq 4.5; \\ 0,89, & 4.5 < x \leq 5.5; \\ 0,97, & 5.5 < x \leq 6.5; \\ 0,98, & 6.5 < x \leq 7.5; \\ 1, & x > 7.5; \end{cases}$$

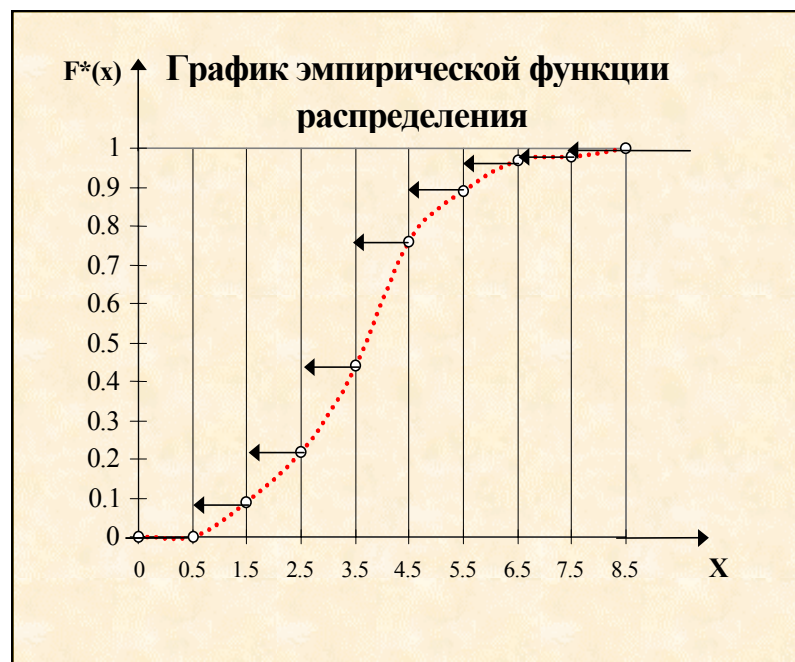


Рис. 1.5.

Построение вариационного ряда - это первый шаг к осмысливанию ряда наблюдений. Однако на практике этого недостаточно, в особенности, когда необходимо сравнивать два или более ряда.

5. Числовые характеристики выборки.

При практическом изучении генеральной совокупности иногда бывает достаточной информацией получение нескольких числовых характеристик распределения, которые позволяют оценить центр группирования значений исследуемой случайной величины, меру их случайного рассеивания и т. д.. Например, при изучении закона распределения заработной платы работников в первую очередь интересуются *средней* заработной платы и одной из мер ее

случайного рассеивания (*дисперсией, коэффициентом дифференциации*). Полезность получения числовых характеристик обусловлено тем, что большинство используемых в статистических приложениях модельных законов могут быть однозначно восстановлены по своим числовым характеристикам, например по среднему значению и дисперсии (если случайная величина $\xi \in N(a, \sigma)$, то по выборкам необходимо оценить a и σ , т.к. они полностью определяют нормальное распределение).

Числовые характеристики вариационного ряда:

1. Среднее арифметическое вариационного ряда

$$\bar{x} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \cdot \sum_{i=1}^n x_i, \quad (6)$$

где x_i - элемент выборки, n - ее объем. Если объем выборки велик нужно использовать ЭВМ или другую вычислительную технику.

Итак, средняя заработная плата рабочих-сдельщиков

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i = \underline{3.1312}.$$

2. Дисперсия выборки вычисляется по формуле:

$$S^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 \quad (7)$$

т.е. это есть средняя арифметическая квадратов отклонений от их средней арифметической. А оценку $\hat{S}^2$ называют *исправленной выборочной дисперсией*.

$$\hat{S}^2 = \frac{n}{n-1} \cdot S^2, \quad (8)$$

Дробь $\frac{n}{n-1}$ называют *поправкой Бесселя*. При малых n поправка Бесселя

значительно отличается от 1. При $n > 50$ практически нет разницы между S^2 и $\hat{S}^2$.

Для нашего примера получаем результаты:

$$S^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 = \underline{2.2605}, \text{ а } \hat{S}^2 = \frac{n}{n-1} \cdot S^2 = \underline{2.2833}.$$

Но мера рассеивания должна выражаться в тех же единицах, что и значение признака, поэтому в качестве показателя вариации чаще используют корень квадратный из дисперсии (*среднее квадратическое отклонение*)

$$S = \sqrt{\frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2} = \underline{1.5035}, \text{ а для } \hat{S} = \sqrt{\hat{S}^2} = \sqrt{\frac{n}{n-1} \cdot S^2} = \underline{1.5111}.$$

3. Наряду со средними величинами в качестве описательных характеристик вариационного ряда применяется медиана и мода.

Медианой (M_e) называется значение признака, приходящего на середину ранжированного ряда наблюдений. Если вариационный ряд содержит нечетное число элементов, то это действительное значение срединного элемента, а при n четном - среднее арифметическое двух срединных элементов.

Для интервального вариационного ряда медиана вычисляется по следующей приближенной формуле:

$$M_1 = x_0 + k \cdot \frac{n/2 - T_{i-1}}{n_i}, \quad (9)$$

где x_0 - начало медианного интервала (интервал в котором содержится срединный элемент); k - длина медианного интервала; n - объем выборки; T_{i-1} - сумма частот признаков в интервалах, предшествующих медианному; n_i - частота медианного интервала.

$$\text{Для нашего случая } M_1 = \frac{x_{50} + x_{51}}{2} = \frac{3.1273 + 3.1291}{2} = 3.1282.$$

Модой (M_0) называется такое значение признака вариационного ряда, которое наблюдалось наибольшее число раз (имеет максимальную частоту).

Нахождение моды для дискретного вариационного ряда не требует вычислений. В случае интервального вариационного ряда мода вычисляется по следующей приближенной формуле:

$$M_0 = x_0 + k \cdot \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})}, \quad (10)$$

где x_0 - начало модального интервала (интервал в котором содержится максимальная частота); k - длина модального интервала; n_{i+1}, n_{i-1} - частоты в последующем и предшествующем интервалах; n_i - частота модального интервала.

Используя формулу (10) вычислим моду. Модальным является интервал от 3 до 4 (см. таб. №3),

Интервал	0 - 1	1 - 2	2 - 3	3 - 4	4 - 5	5 - 6	6 - 7	7 - 8
n_i	9	13	22	32	13	8	1	2

так как ему соответствует наибольшая частота, равная 32, а частоты в интервале (2;3) - 22 и в (4;5) - 13, $k = 1$.

$$M_0 = x_0 + k \cdot \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} = 3 + 1 \cdot \frac{32 - 22}{(32 - 22) + (32 - 13)} = \underline{3.3448}.$$

Лабораторная работа № 2.

ТЕМА: Статистическое оценивание параметров.

1. Понятие оценки. Одна из главных целей, которые ставит перед собой исследователь, приступая к статистической обработке исходных данных, заключается в том, чтобы представить множество обрабатываемых данных в виде сравнительно небольшого числа сводных характеристик, построенных на основании этих данных. При этом, чтобы потеря информации для принятия существенных решений была минимальной.

Распределение случайной величины ξ характеризуется целым рядом параметров. Сюда относятся числовые характеристики ξ (математическое ожидание, дисперсия, среднее квадратическое отклонение, мода, медиана и т.д.) . Эти параметры называются *параметрами генеральной совокупности*. Они могут быть найдены из закона распределения. Однако при решении многих практических задач нет необходимости характеризовать случайную величину исчерпывающим образом, а достаточно указать только отдельные числовые характеристики, которые характеризуют существенные черты распределения. Используя выборку наблюдений случайной величины, можно вычислить приближенные значения каждого из параметров, называемые в статистике *числовыми оценками* параметров или просто оценками.

Сформулируем задачу оценки параметров в общем виде. Пусть ξ - случайная величина, подчинена закону распределения $F_{\xi}(x, \theta)$, аналитическое выражение которое известно. Функция распределения $F_{\xi}(x, \theta)$ определяется параметром θ , численное значение которого неизвестно. Исследовать все элементы генеральной совокупности для вычисления параметра θ не представляется возможным, поэтому о параметре пытаются судить по выборкам из генеральной совокупности.

Пусть произведено m выборок, каждая из которых состоит из независимых наблюдений x_{ij} (i - номер выборки, j - номер элемента в выборке), т.е.

1-я	выборка	состоит	из	элементов	(	$x_{11}, x_{12}, \dots, x_{1n}$	)
2-я	-''-	-''-	-''-	-''-	(	$x_{21}, x_{22}, \dots, x_{2n}$	)
.....							
i-я	-''-	-''-	-''-	-''-	(	$x_{i1}, x_{i2}, \dots, x_{in}$	)
.....							
m-я	-''-	-''-	-''-	-''-	(	$x_{m1}, x_{m2}, \dots, x_{mn}$	) и т.д.

Но так как выборки случайны, тогда можно обозначить через ξ_1 случайную величину, принимающую значения $(x_{11}, x_{12}, \dots, x_{1n})$, через ξ_2 - случайную величину, принимающую значения $(x_{21}, x_{22}, \dots, x_{2n})$ и т.д. Тогда последовательность случайных величин $\xi_1, \xi_2, \dots, \xi_n$ (независимых) распределена по закону $F(x, \theta)$.

Всякую однозначно определенную функцию результатов наблюдений над случайной величиной ξ , с помощью которой судят о значении параметра θ , называют оценкой параметра θ . Если оценку параметра θ обозначить через $\tilde{\theta}_n$, то можно записать, что $\tilde{\theta}_n = f(\xi_1, \xi_2, \dots, \xi_n)$.

Очевидно, что $\tilde{\theta}_n$ - оценка является случайной величиной, поскольку она зависит от $\xi_i (i = \overline{1:n})$. Поэтому закон распределения $\tilde{\theta}_n$ зависит от закона распределения случайной величины ξ и от числа наблюдений n . Так, например, для пуассоновской случайной величины необходимо оценить параметр λ , определяющий это распределение. Если случайная величина $\xi \in N(a, \sigma)$, то по выборкам необходимо оценить $\theta_1 = a$ и $\theta_2 = \sigma$, т.к. они полностью определяют нормальное распределение. На практике поступают следующим образом:

- 1) x_1 - первый элемент выборки и его используют как оценку $\tilde{\theta}_1$;
- 2) эту оценку можно получить, как $\tilde{\theta}_1 = (x_{\max} - x_{\min})/2$ - среднее арифметическое максимального и минимального элементов выборки;
- 3) M_0 - в качестве оценки можно взять моду, которая при нормальном распределении равна среднему значению a ;
- 4) M_e - медиану, которая при нормальном распределении также равна среднему значению a ;
- 5) $\bar{x}$ - среднее арифметическое.

Выбор оценки, позволяющей получить хорошее приближение оцениваемого параметра, - основная задача теории оценивания. Для того чтобы установить, какая из оценок лучше надо знать основные свойства (виды) оценок.

2. Основные свойства точечных оценок. Для того чтобы оценка $\tilde{\theta}_n$ имела практическую ценность, она должна обладать следующими свойствами.

1. Оценка $\tilde{\theta}_n$ параметра θ называется *несмещенной*, если ее математическое ожидание равно оцениваемому параметру θ , т.е.

$$M\tilde{\theta}_n = \theta. \quad (1)$$

Если равенство (1) не выполняется, то оценка $\tilde{\theta}_n$ может либо завышать значение θ ($M\tilde{\theta}_n > \theta$), либо занижать его ($M\tilde{\theta}_n < \theta$). Естественно в качестве приближенного неизвестного параметра брать несмещенные оценки для того, чтобы не делать систематической ошибки в сторону завышения или занижения.

2. Оценка $\tilde{\theta}_n$ параметра θ называется *состоятельной*, если она подчиняется закону больших чисел, т.е. *сходится по вероятности* к оцениваемому параметру при неограниченном возрастании числа опытов (наблюдений) и следовательно выполняется следующее равенство :

$$\lim_{n \rightarrow \infty} P\{|\tilde{\theta}_n - \theta| < \varepsilon\} = 1 \quad (2)$$

где $\varepsilon > 0$ сколько угодно малое число.

Для выполнения (2) достаточно, чтобы дисперсия оценки стремилась к нулю при $n \rightarrow \infty$, т.е.

$$\lim_{n \rightarrow \infty} D(\tilde{\theta}_n) = 0 \quad (3)$$

И, кроме того, чтобы оценка была несмещенной. От формулы (2) легко перейти к (3), если воспользоваться неравенством Чебышева.

Итак, состоятельность оценки означает, что при достаточно большом количестве опытов и со сколько угодно большой достоверностью отклонение оценки от истинного значения параметра меньше любой наперед заданной величины. Этим оправдано увеличение объема выборки.

3. Несмещенная оценка $\tilde{\theta}_n$, которая имеет наименьшую дисперсию среди всех возможных несмещенных оценок параметра θ , вычисленных по выборкам одного и того же объема, называется *эффективной* оценкой.

Так как $\tilde{\theta}_n$ - случайная величина, значение которой изменяется от выборки к выборке, то меру ее рассеивания около математического ожидания θ будем характеризовать дисперсией $D\tilde{\theta}_n$. Пусть $\tilde{\alpha}_n^1$ и $\tilde{\alpha}_n^2$ - две несмещенные оценки параметра θ , т.е. $M\tilde{\alpha}_n^1 = \theta$ и $M\tilde{\alpha}_n^2 = \theta$, соответственно $D\tilde{\alpha}_n^1$ и $D\tilde{\alpha}_n^2$ и, если $D\tilde{\alpha}_n^1 < D\tilde{\alpha}_n^2$, то в качестве оценки принимают $\tilde{\alpha}_n^1$.

На практике при оценке параметров не всегда удается удовлетворить одновременно требованиям 1, 2, 3. Однако выбору оценки всегда должно предшествовать ее критическое рассмотрение со всех точек зрения. При выборе практических методов обработки опытных данных необходимо руководствоваться сформулированными свойствами оценок.

3. Понятие доверительного интервала. Доверительная вероятность.

Оценки, которыми мы до сих пор занимались, называются *точечными*, так как полученные оценки выражались одним числом. Однако, в ряде задач требуется не только найти для оцениваемого параметра θ подходящее числовое значение, но оценить его точность и надежность. Такого рода задачи очень важны при малом числе наблюдений, так как конечная оценка $\tilde{\theta}_n^*$ в значительной мере является случайной и приближенная замена θ на $\tilde{\theta}_n^*$ может привести к серьезным ошибкам.

Задачу интервального оценивания в самом общем виде можно сформулировать так: *по данным выборки построить числовой интервал, относительно которого с заранее выбранной вероятностью можно сказать, что внутри этого интервала находится оцениваемый параметр.*

Для определения точности оценки $\tilde{\theta}_n$ в математической статистике пользуются *доверительными интервалами*, а определения надежности - *доверительными вероятностями*. Раскроем существо этих понятий.

Доверительным интервалом для параметра θ называется такой интервал, относительно которого можно с заранее выбранной вероятностью $\beta = 1 - \alpha$, близкой к единице, утверждать, что он содержит неизвестное значение параметра θ . Пусть $\tilde{\theta}_n$ - несмещенная оценка параметра θ . Требуется оценить возможную при этом ошибку. Найдем теперь такое значение $\varepsilon > 0$, чтобы

$$P\{|\tilde{\theta}_n - \theta| < \varepsilon\} = \beta, \quad (4)$$

где β - заданная вероятность (например, $\beta = 0.99$).

Представим (4) в виде

$$P\{\tilde{\theta}_n - \varepsilon < \theta < \tilde{\theta}_n + \varepsilon\} = \beta. \quad (5)$$

Равенство (5) означает, что неизвестное значение параметра θ с вероятностью β попадает в интервал $[\theta_1; \theta_2]$, где $\theta_1 = \tilde{\theta}_n - \varepsilon$, а $\theta_2 = \tilde{\theta}_n + \varepsilon$. Заметим, что здесь θ не случайная величина, а интервал $[\theta_1; \theta_2]$ является случайной величиной.

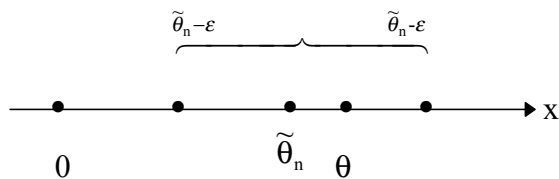


Рис. 2.1.

β - называют *доверительной вероятностью*, а значение α - *уровнем значимости*. Нижняя и верхняя граница доверительного интервала θ_1 и θ_2 определяется по результатам наблюдений. В связи с этим говорят, что доверительный интервал накрывает оцениваемый параметр с вероятностью $\beta = 1 - \alpha$ или $100 (1 - \alpha) \%$ случаев. Выбор β определяется конкретными условиями решаемой проблемы.

4. Построение доверительного интервала для математического ожидания при известной дисперсии. Пусть $\xi \in N(a; \sigma)$, причем σ - известно, а $a = \theta$ - оцениваемый параметр. Требуется оценить неизвестное математическое ожидание. Наилучшей оценкой $M\xi = a$ в смысле несмещенности, состоятельности и эффективности, как следует из теории, является выборочное среднее $\bar{x}$. Можно показать, что $\bar{x}$ распределено нормально с параметрами $M\bar{x} = a$, $D\bar{x} = \frac{\sigma^2}{n}$;

нормированное отклонение $\mu = \frac{\bar{x} - a}{\sigma / \sqrt{n}}$ распределено нормально с параметрами

$\mu \in N(0;1)$, поэтому вероятность любого отклонения $|\bar{x} - a|$ может быть вычислена по формуле

$$P\left\{\left|\frac{\bar{x} - a}{\sigma} \cdot \sqrt{n}\right| < Z_\beta\right\} = \Phi(z). \quad (6)$$

Задавая определенную вероятность $\beta = 1 - \alpha = \Phi(z)$ по таблице значений можно определить значение Z_β . Для оценки a преобразуем формулу (6)

$$P\left\{|\bar{x} - a| < Z_\beta \cdot \frac{\sigma}{\sqrt{n}}\right\} = \Phi(z).$$

или

$$P\left\{-Z_\beta \cdot \frac{\sigma}{\sqrt{n}} < \bar{x} - a < Z_\beta \cdot \frac{\sigma}{\sqrt{n}}\right\} = \Phi(z),$$

откуда получим

$$P\left\{\bar{x} - Z_\beta \cdot \frac{\sigma}{\sqrt{n}} < a < \bar{x} + Z_\beta \cdot \frac{\sigma}{\sqrt{n}}\right\} = \Phi(z).$$

Таким образом, с вероятностью (надежностью) $\Phi(z) = \beta = 1 - \alpha$ можно утверждать, что интервал

$$(\bar{x} - Z_{\beta} \cdot \frac{\sigma}{\sqrt{n}}; \bar{x} + Z_{\beta} \cdot \frac{\sigma}{\sqrt{n}})$$

является доверительным для оценки a .

З а м е ч а н и е I. Обозначим $\Delta = Z_{\beta} \cdot \frac{\sigma}{\sqrt{n}}$. Тогда:

1) при фиксированном значении Z_{β} с возрастанием n величина Δ уменьшается, следовательно, точность интервального оценивания увеличивается;

2) увеличение надежности оценки $\beta = 1 - \alpha$ ведет к увеличению функции $\Phi(z)$, так как она возрастающая, и Z_{β} , поэтому, при фиксированном объеме выборки n , величина Δ также возрастает, что ведет к увеличению доверительного интервала, и, следовательно, уменьшению точности оценки;

3) так как $\bar{x}$ - случайная величина, возможные значения которой меняются от выборки к выборке, то концы доверительного интервала

$(\bar{x} - \Delta; \bar{x} + \Delta)$ также являются случайными величинами, меняющимися от выборки к выборке.

5. Построение доверительного интервала для математического ожидания при неизвестной дисперсии. Пусть $\xi \in N(a; \sigma)$, причем среднее квадратическое отклонение этого распределения неизвестно. Требуется оценить неизвестное математическое ожидание $a = \theta$ - оцениваемый параметр.

Рассмотрим случайную величину

$$t = \frac{\bar{x} - a}{\hat{S}} \cdot \sqrt{n},$$

которая распределена по закону *Стьюдента*, поэтому, выбрав вероятность

$\beta = 1 - \alpha$ и зная объем выборки n , можно пользуясь таблицей для значений распределения Стьюдента, найти $t_{n\beta}$ такое, что

$$P\left\{\left|\frac{\bar{x} - a}{\hat{S}}\right| \cdot \sqrt{n} < t_{n\beta}\right\} = \beta. \quad (7)$$

Производя аналогичные преобразования формулы (7), как и (6) получим

$$(\bar{x} - t_{n\beta} \cdot \frac{\hat{S}}{\sqrt{n}}; \bar{x} + t_{n\beta} \cdot \frac{\hat{S}}{\sqrt{n}})$$

доверительный интервал для оценки a .

З а м е ч а н и е II. Несмотря на кажущееся сходство формул (6) и (7), между ними существуют различия, т.к. $t_{n\beta}$ зависит не только от β , но и от объема выборки. Особенно это различие заметно при малом объеме выборки.

6. Определение объема выборки n . До сих пор рассматривались выборки, которые имели фиксированный объем выборки n . Но на практике больший объем выборки требует определенных затрат для ее получения и обработки. Поэтому целесообразно использовать минимальный объем, который позволяет получить результаты нужной точности. Воспользуемся для определения объема выборки нормального распределения формулой $\Delta = Z_{\beta} \cdot \frac{\sigma}{\sqrt{n}}$. Откуда

$$n = Z_{\beta}^2 \cdot \frac{\sigma^2}{\Delta^2}. \quad (8)$$

Изучите замечание I, которое имеет практическую ценность при определении n .

Если σ неизвестно, то для выборки объема n из формулы $\Delta = t_{\beta} \frac{\hat{S}}{\sqrt{n}}$, получаем $n = t_{\beta}^2 \frac{\hat{S}^2}{\Delta^2}$, по этой формуле можно получить нужный объем выборки n после обработки уже имеющейся пробной выборки, по которой вычисляется оценка дисперсии.

Постановка задачи.

1. Вычислить по выборке несмещенные оценки параметров μ - среднего значения, σ^2 - дисперсии, σ - стандартного отклонения генеральной совокупности $\bar{x}, S^2, S$;

2. Найти доверительные интервалы для μ при различной доверительной вероятности $\beta = 0,8; 0,95; 0,99$;

3. Считая данные выборки пробными, определить минимальный объем выборки n для нахождения доверительного интервала среднего значения - μ длины 2Δ и доверительной вероятностью $\beta = 0,8; 0,9; 0,95; 0,99; 0,995$;

Пример выполнения лабораторной работы:

Используем выборку предыдущей работы:

Исходные данные

Таблица №1.

4.2781	3.1645	1.8657	2.4554	6.1401	3.1535	7.5917	3.408	1.1001	5.7285
3.5566	2.659	3.5202	5.085	4.8022	5.6521	2.5561	3.1291	2.2626	4.2695
5.0908	3.6177	3.4169	2.1926	5.7115	4.7699	5.4602	2.9712	3.1363	4.9042
0.5653	3.5348	4.4335	1.8655	2.0584	2.3136	2.4538	3.4256	1.8667	0.8356
3.6185	3.6528	2.5759	1.6058	3.4139	2.4861	0.0044	3.3833	3.3956	0.2852
3.7213	3.2617	3.0226	4.5737	2.1159	1.879	1.7298	0.9983	2.1602	3.0121
3.0742	5.158	3.6585	0.0758	2.4228	3.9749	2.0997	1.3595	3.0733	2.6159
2.495	2.6765	1.0106	4.3656	2.5593	3.7782	5.6209	1.5937	0.8298	2.5329
1.5209	1.4857	2.958	3.2787	4.4941	4.3576	2.8779	3.0059	3.9802	7.2719
3.9614	4.7845	0.451	3.1273	3.5688	4.4303	0.7919	1.7188	4.7897	3.3967

1. Здесь взять в качестве оценки параметра μ можно:

$$\Theta_1 = x_1 = 4.2781;$$

$$\Theta_2 = (x_{\max} - x_{\min})/2 = (7.5917 - 0.0044)/2 = 3.7937;$$

$$\Theta_3 = M_0 = x_0 + k \cdot \frac{n_i - n_{i-1}}{(n_i - n_{i-1}) + (n_i - n_{i+1})} = 3 + 1 \cdot \frac{32 - 22}{(32 - 22) + (32 - 13)} = 3.3448;$$

$$\Theta_4 = M_e = \frac{x_{50} + x_{51}}{2} = \frac{3.1273 + 3.1291}{2} = 3.1282;$$

$$\Theta_5 = \bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i = 3.1312.$$

Согласно теории наилучшей оценкой будет $\Theta_5 = 3,1312$. (Ответте почему). Соответственно для дисперсии и стандартного отклонения наилучшими оценками будут:

$$\text{не } S^2 = \frac{1}{n} \cdot \sum_{i=1}^n (x_i - \bar{x})^2 = \underline{2.2605}, \text{ а } \hat{S}^2 = \frac{n}{n-1} \cdot S^2 = \underline{2.2833} \text{ и}$$

$$\hat{S} = \sqrt{\hat{S}^2} = \sqrt{\frac{n}{n-1} \cdot S^2} = \underline{1.5111}.$$

2. С доверительной вероятностью $\beta = 0,8$ по формуле

$$P\left\{\bar{x} - t_\beta \cdot \frac{S}{\sqrt{n}} < \mu < \bar{x} + t_\beta \cdot \frac{S}{\sqrt{n}}\right\} = \beta,$$

найдем доверительный интервал для параметра μ :

$$\left(\bar{x} - t_\beta \cdot \frac{\hat{S}}{\sqrt{n}}; \bar{x} + t_\beta \cdot \frac{\hat{S}}{\sqrt{n}}\right).$$

а). Из таблицы распределения Стьюдента находим $t_\beta = 1.48$, а тогда

$$\Delta = 1,48 * 1,5111 / 10 = 0.2236;$$

$$3,1312 - 0,2236 \leq \mu \leq 3,1312 + 0,2236, \quad \underline{2.9076 \leq \mu \leq 3.3548}.$$

б). Если доверительная вероятность $\beta = 0.95$, тогда $t_\beta = 1.98$, а доверительный интервал для параметра будет $\underline{\mu \in (2.832; 3.4304)}$.

3. Для определения объема выборки, считая, что $n = 100$ - пробная выборка определим минимальный объем. Пусть $\beta = 0,8$ и $\Delta = 0.2236$ найдем $t_\beta = 1.48$. По выборке получаем $S^2 = 2.2833$ и

$$n_{\min} = 1.48^2 * 2.2833 / 0.2236^2 = 100.03 \approx 100.$$

Для $\beta = 0,995$ и $\Delta = 0,2236$ имеем $t_\beta = 2.86$ и

$$n_{\min} = 2.86^2 * 2.2833 / 0.2236^2 = 373.55 \approx 374.$$

Таким образом, увеличивая надежность оценок или уменьшая длину интервала, мы должны увеличивать и объем выборки, например при тех же значениях возьмем $\Delta = 0.1$, тогда $n_{\min} = 2.86^2 * 2.2833 / 0.1^2 = 1867.6 \approx 1868$.

Лабораторная работа № 3.

ТЕМА: Статистические гипотезы.

1. Понятие статистической гипотезы. Под статистической проверкой гипотез мы будем понимать процедуру построения некоторых правил, позволяющих по результатам эксперимента высказывать суждение о природе явлений, обуславливающих изучаемый эксперимент.

Проверка статистических гипотез тесно связана с теорией оценивания параметров. В естествознании, технике, экономике часто для выяснения того или иного случайного фактора прибегают к высказыванию гипотез, которые можно проверить статистически, т.е. опираясь на результаты наблюдений в случайной выборке.

Пусть ξ – наблюдаемая дискретная или непрерывная случайная величина. Статистической гипотезой **H** называется предположение относительно параметров или вида распределения случайной величины ξ . Так, например, статистической является гипотеза о том, что распределение производительности труда рабочих, выполняющих одинаковую работу в одинаковых организационно-технических условиях, имеет нормальный закон распределения. Статистической будет также с гипотеза о том, что средние размеры деталей, производимые на однотипных, параллельно работающих станках не различаются между собой.

Статистическая гипотеза **H** называется простой, если она однозначно определяет распределение случайной величины ξ , в противном случае гипотеза **H** называется сложной. Например, простой гипотезой является предположение о том, что случайная величина ξ распределена по нормальному закону $\xi \in N(0;1)$, если же высказывается предположение, что случайная величина ξ имеет нормальное распределение $\xi \in N(a;1)$, где $a < \theta < b$, то это сложная гипотеза.

Сформулируем задачу статистической проверки гипотезы в общем виде. Пусть $f(x, \theta)$ – закон распределения случайной величины ξ зависящей от одного параметра θ . Предположим, что необходимо проверить гипотезу $H_0 = \{\theta = \theta_0\}$ – нулевую гипотезу. Гипотезу о том, что $\theta = \theta_1$, назовем *конкурирующей* и обозначим ее через $H_1 = \{\theta = \theta_1\}$. Заметим, что иногда гипотезу H_1 называют *альтернативной гипотезой* или альтернативой. И так как распределение случайной величины ξ известно, и по выборке нужно сделать предположение о параметре распределения, то такие гипотезы называются параметрическими.

Таким образом, перед нами стоит задача проверки гипотезы H_0 относительно конкурирующей гипотезы H_1 , на основании выборки, состоящей из n независимых наблюдений $x_1, x_2, \dots, x_n$ над случайной величиной ξ . Следовательно, все возможное множество выборок объема n можно разделить на два непересекающихся множества (обозначим их Q и W) таких, что проверяемая гипотеза H_0 должна быть отвергнута, если наблюдаемая выборка попадает в подмножество W , и принята, если выборка принадлежит Q .

W - называют критической областью;

Q - областью допустимых значений.

Правило, по которому принимается решение принять или отклонить гипотезу H_0 , называется *критерием K* .

Проверка статистической гипотезы основывается на принципе, в соответствии с которым маловероятные события считаются невозможными, а события, имеющие большую вероятность, считаются достоверными. Этот принцип можно реализовать следующим образом. Перед анализом выборки фиксируется некоторая малая вероятность α , называемая уровнем значимости.

Пусть

$$V = W \cup Q$$

множество значений статистики Z , которая выбирается подходящим образом в зависимости от K . Тогда W такое подмножество, что при условии истинности гипотезы H_0 вероятность попадания статистики критерия в W равна α ,

$$P\{Z \in W/H_0\} = \alpha.$$

Обозначим Z_v выборочное значение статистики Z , вычисленное по выборке наблюдений. Критерий формулируется следующим образом:

отклонить гипотезу H_0 , если $Z_v \in W$; принять гипотезу H_0 , если $Z_v \in Q = V \setminus W$.
Критерий, основанный на использовании заранее заданного уровня, значимости, называют *критерием значимости*.

Уровень значимости α определяет "размер" критической области W . Положение критической области на множестве статистики Z зависит от формулировки альтернативной гипотезы H_1 . Например, если проверяемая гипотеза $H_0 = \{\theta = \theta_0\}$, а альтернативная $-H_1: (\{\theta < \theta_0\}; \{\theta > \theta_0\})$, то критическая область размещается на правом или левом "хвосте" распределения статистики Z . В этом случае критерий называется *односторонним*.

Если $H_1 = \{\theta \neq \theta_0\}$, то критическая область размещается на обоих "хвостах" распределения Z , а K называют *двусторонним*.

Пусть $f(z/H_0)$ – плотность распределения статистики Z критерия при условии, что верна гипотеза H_0 . $Q = V \setminus W$ - область принятия гипотезы.

Покажем на рисунках расположение W для различных H_1 : $P\{z \in Q\} = 1 - \alpha$, $P\{z \in W\} = \alpha$.

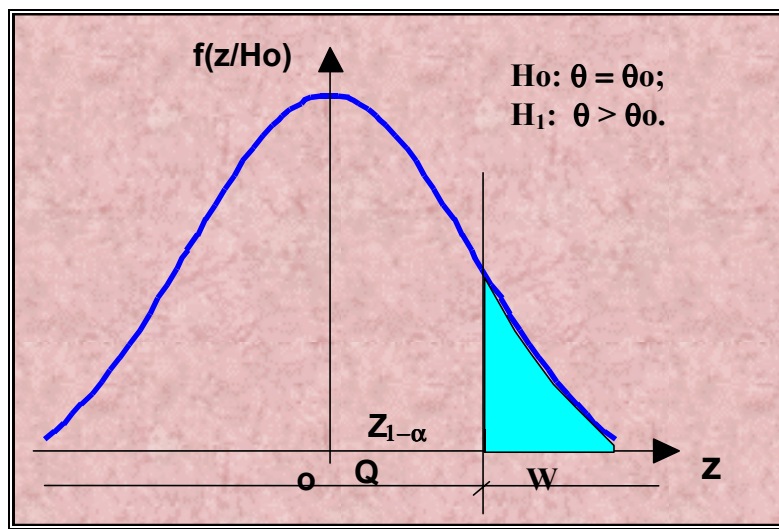


Рис.3.1.

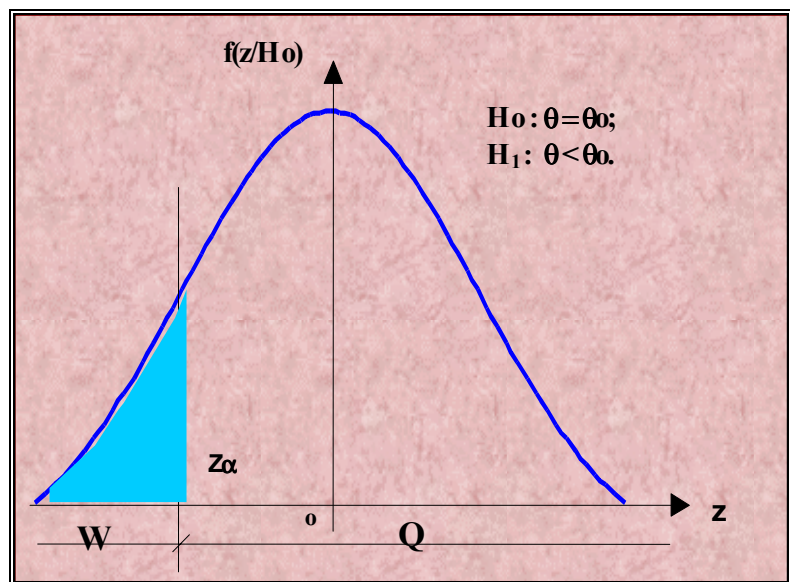


Рис.3.2.

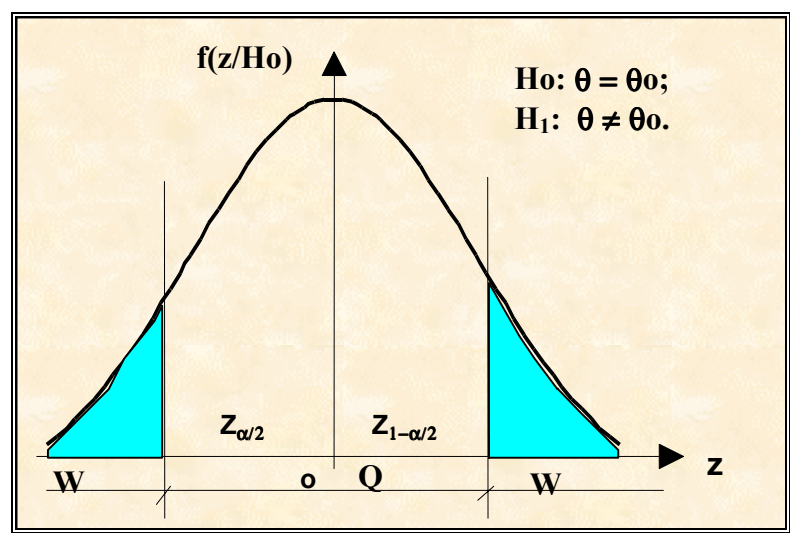


Рис. 3.3.

Таким образом, проверка параметрической статистикой гипотезы при помощи K – значимости может быть разбита на следующие этапы:

1. Сформулировать проверяемую (H_0) и альтернативную (H_1) гипотезу;
2. Назначить уровень значимости α ;
3. Выбрать статистику Z критерия для проверки гипотезы H_0 ;
4. Определить выборочное распределение статистики Z при условии, что верна гипотеза H_0 ;
5. В зависимости от формулировки H_1 определить критическую область W одним из неравенств $Z > Z_{1-\alpha}$ или их совокупностью;
6. Получить выборку наблюдений и вычислить выборочное значение статистики Z_b критерия.
- 7 . Принять статистическое решение: если $Z_b \in W$, то отклонить гипотезу H_0 , как не согласующуюся с результатами наблюдений;

если, $Z_b \in \theta = V \setminus W$, то принять гипотезу H_0 , т.е. считать, что H_0 не противоречит результатам наблюдений.

З а м е ч а н и е. При выборе критической области W следует иметь в виду, что принимая или отклоняя гипотезу H_0 , можно допустить ошибки двух видов.

Ошибка первого рода состоит в том, что H_0 - отвергается, т. е. принимается H_1 , в то время, как в действительности все же верна гипотеза H_0 .

Ошибка второго рода состоит в том, что гипотеза H_0 принимается, в то время, как верна H_1 . Тогда вероятность ошибки первого рода определяется уровнем значимости, т. к.

$$P\left\{Z \in W / H_0\right\} = \alpha,$$

а вероятность ошибки второго рода β можно вычислить (при простой альтернативной гипотезе H_1)

$$P\left\{Z \in Q / H_1\right\} = \beta.$$

Пример 3.1. (Проверка гипотезы о равенстве центров распределения двух нормальных генеральных совокупностей при известном σ).

В результате двух серий измерений с количеством измерений $n_1=25$ и $n_2 = 50$ получены следующие средние значения исследуемой величины:

$\bar{x} = 9,79$; $\bar{y} = 9,60$. Можно ли с надежностью $\beta = 1 - \alpha = 0,99$ объяснить это расхождение случайными причинами, если известно, что средние квадратические отклонения в обеих сериях измерений $\sigma_x = \sigma_y = 0,30$?

Р е ш е н и е:

Пусть ξ, η – независимые случайные величины, каждая из которых распределена по нормальному закону. Гипотеза $H_0 = \{M\xi = M\eta\}$, а $H_1 = \{|M\xi - M\eta| > 0\}$. $M\xi, M\eta$ – неизвестны, а $\sigma_x = \sigma_y = 0,3$. Тогда для проверки гипотезы H_0 используется их наилучшие оценки $\bar{x}$ и $\bar{y}$. Известно, что $\bar{x}$ и $\bar{y}$ имеют нормальный закон распределения с параметрами $N_1(M\xi; \sigma\xi; \sqrt{n_1})$ и $N_2(M\eta; \sigma\eta; \sqrt{n_2})$. Выборки независимы, поэтому $\bar{x}$ и $\bar{y}$ также независимы и случайная величина, равная разности между $\bar{x}, \bar{y}$, имеет нормальное распределение, причем

$$D(\bar{x} - \bar{y}) = D\bar{x} + D\bar{y} = \frac{\sigma\xi^2}{n_1} + \frac{\sigma\eta^2}{n_2}.$$

Если гипотеза H_0 справедлива, то

$$M(\bar{x} - \bar{y}) = M\bar{x} - M\bar{y} = 0,$$

следовательно, нормированная разность

$$|Z| = \frac{|\bar{x} - \bar{y}|}{\sqrt{\frac{\sigma\xi^2}{n_1} + \frac{\sigma\eta^2}{n_2}}} = \frac{9,79 - 9,60}{0,3 \cdot \sqrt{\frac{1}{25} + \frac{1}{50}}} = 2,59$$

подчиняется нормальному закону распределения с математическим ожиданием, равным нулю, и дисперсией, равной единице.

По таблице определим статистику Z_p , которая разделит множество Z на два пересекающихся подмножества: область допустимых значений $Z(Q)$ и критическую

область $Z (W)$. Те значения $|Z| \leq Z_p = Z_{0,99} = 2,576$, образуют область допустимых значений и $W = \{|Z| > Z_p = Z_{0,99} = 2,576\}$. Теперь имеем $2,59 > 2,576$, поэтому с надежностью 0.99 можно считать расхождение средних неслучайным (значимым), так как попадание в область значений $|Z| > 2,576$, при нашей гипотезе практически невозможна.

Замечание. Однако следует отметить, что для значений $|Z| \leq 2,576$,

еще нельзя утверждать, что гипотеза подтвердилась: можно только признать допустимость гипотезы для рассмотренных выборочных наблюдений до тех пор, пока более обстоятельные исследования не позволят сделать противоположное заключение.

Следовательно, с помощью проверки статистических гипотез можно лишь отвергнуть проверяемую гипотезу, но никогда нельзя доказать ее справедливость.

2. Критерий согласия χ^2 Пирсона (проверка гипотез о законе распределения).

Во многих практических задачах точный закон распределения исследуемой случайной величины неизвестен, т.е. является гипотезой, которая требует статистической проверки.

Пусть ξ исследуемая случайная величина. Нужно проверить гипотезу H_0 – {случайная величина ξ подчиняется закону распределения $F(x)$ }.

При получении критерия для проверки гипотезы H_0 используем меру, введенную Пирсоном, приводящую к так называемому критерию χ^2 Пирсона.

Этот критерий наиболее часто употребляется для проверки о законе распределения. Отметим, что существует несколько критериев согласия: Колмогорова, Смирнова, ω^2 Мизеса и др.

Для проверки гипотезы H_0 произведем выборку, состоящую из n независимых наблюдений над случайной величиной ξ . По выборке построим эмпирическую функцию распределения $F_n^*(x)$. Сравнение эмпирического $F_n^*(x)$ и теоретического распределения производится с помощью специально подобранной случайной величины – критерия согласия.

Разобьем множество значений ξ на r множеств $(S_1, S_2, \dots, S_r)$ без общих точек.

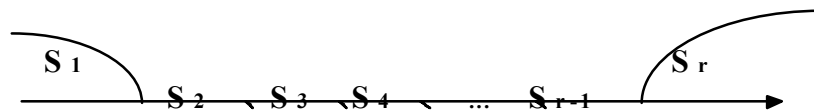


Рис. 3.4.

Подсчитаем количество элементов выборки v_i , попавших в каждый из интервалов S_i . В силу гипотезы H_0 предполагаем известным закон распределения $F(x)$, всегда можно определить

$$p_i = P \{ \xi \in S_i \}, \quad \left(\sum_{i=1}^r p_i = 1 \right).$$

Но с другой стороны

$$\sum_{i=1}^r v_i = n, \quad \sum_{i=1}^r \frac{v_i}{n} = 1.$$

И, более того, теоретическое число значений случайной величины ξ , попавших в интервал Si , можно определить по формуле $n \cdot p_i$. Если эмпирические частоты сильно отличаются от теоретических, то проверяемую гипотезу H_0 следует отвергнуть, в противном случае принять.

Сформулируем критерий, который бы характеризовал степень расхождения между эмпирическими и теоретическими частотами. Если проверяемая гипотеза H_0 верна, то случайная величина v_i , характеризующая количество попаданий в интервал Si , подчиняется биномиальному закону распределения с математическим ожиданием

$Mv_i = n \cdot p_i$ и дисперсией $Dv_i = n \cdot p_i \cdot q_i$, ($q_i = 1 - p_i$). Тогда при $n \rightarrow \infty$ случайная величина

$$y_i = \frac{v_i - n \cdot p_i}{\sqrt{n \cdot p_i \cdot q_i}},$$

распределена нормально с $My_i = 0$ и $Dy_i = 1$. Случайные величины $y_1, y_2, \dots, y_r$ связаны между собой линейной зависимостью.

В литературе по математической статистике доказывается, что при $n \rightarrow \infty$ статистика

$$\sum_{i=1}^r y_i^2 \cdot q_i = \sum_{i=1}^r \frac{(v_i - n \cdot p_i)^2}{n \cdot p_i}$$

имеет распределение χ^2 с $k = r - 1$ степенями свободы.

Однако, если параметры распределения $F(x)$ оцениваются по выборке, то при $n \rightarrow \infty$

$$\sum_{i=1}^r \frac{(v_i - n \cdot p_i)^2}{n \cdot p_i}$$

имеет χ^2 распределение с $k = r - e - 1$ степенями свободы (e – число параметров распределения $F(x)$, рассчитанных по выборке). Следовательно, в качестве меры расхождения между v_i и $n \cdot p_i$ для $i = \overline{1; r}$ используют критерий

$$\chi^2 = \sum_{i=1}^r \frac{(v_i - n \cdot p_i)^2}{n \cdot p_i}. \quad (*)$$

Правило применения критерия χ^2 сводится к следующему. Рассчитав значения χ^2 и выбрав уровень значимости критерия α , по таблице χ^2 – распределения определяется $\chi^2_{k\alpha}$. Если $\chi^2 \geq \chi^2_{k\alpha}$, то гипотеза H_0 отвергается, если $\chi^2 \leq \chi^2_{k\alpha}$, то гипотеза принимается. Очевидно, что при проверке гипотезы о законе распределения контролируется лишь ошибка первого рода.

З а м е ч а н и е. Как отмечалось раньше, статистика (*) имеет χ^2 распределение лишь при $n \rightarrow \infty$, поэтому необходимым условием применения критерия Пирсона является наличие в каждом Si по меньшей мере 5-10 наблюдений. Если v_i очень малы (1-2), то имеет смысл объединить некоторые Si .

Пример выполнения лабораторной работы:

1. При выполнении лабораторной работы №1 было найдено, что

$$\bar{x} = 3.1312, \quad S^2 = 2.2605, \quad \text{а} \quad \hat{S}^2 = \frac{n}{n-1} \cdot S^2 = 2.2833, \quad S = 1.5035,$$

а для $\hat{S} = \sqrt{\hat{S}^2} = \sqrt{\frac{n}{n-1} \cdot S^2} = 1.5111, n = 100$ (См. статистику Z - таб. №1).

Пусть задано уровень значимости $\alpha = 0.01, \sigma = 1.5, n = 100$.

Проверим гипотезу: $H_0: \mu = 3$ при $H_1: \mu > 3$. Статистика Z имеет нормальное распределение. По альтернативной гипотезе H_1 найдем правостороннюю критическую область $P(Z > z_\alpha) = \alpha$. Из таблицы получаем значение z_α , учитывая, что $P(Z < z_\alpha) = \Phi(z_\alpha) = 1 - \alpha$. В качестве статистики $Z = \frac{\bar{x} - \mu}{\sigma} \sqrt{n}$, которая при истинной гипотезе H_0 имеет нормированное нормальное распределение $Z \in N(0,1)$. $\Phi(z_\alpha) = 1 - 0.01 = 0.99$. По таблицам нормального распределения (см. приложение 2) получаем $z_\alpha = 2.33$. Отсюда следует, что критическая область имеет вид $z > 2.33$. Вычислим значение статистики Z.

$$Z = \frac{3.1312 - 3}{1.5} \cdot \sqrt{100} = 0.87.$$

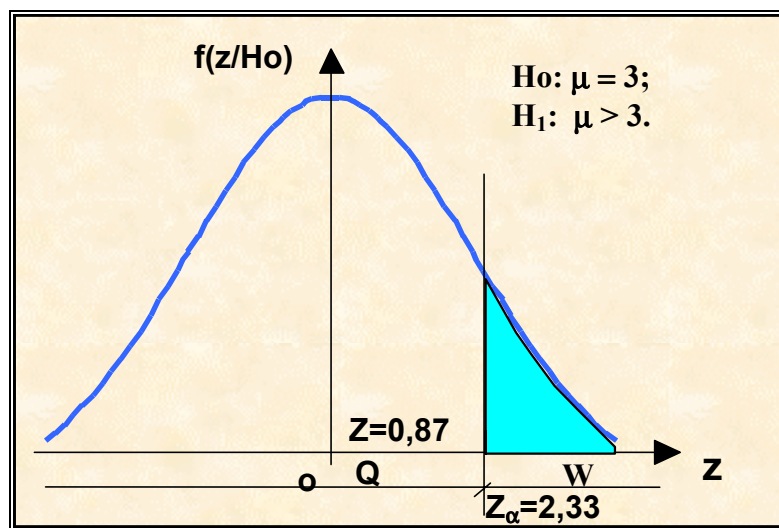


Рис. 3.5

Значение статистики не принадлежит критической области ($0.87 < 2.33$). Следовательно нет оснований отвергать гипотезу H_0 . Отвергаем гипотезу H_1 .

2. Пусть задана выборка объема $n = 20$ из генеральной совокупности:

-246.269	-231.067	-279.336	-245.459	-300.443
-293.724	-252.592	-237.984	-263.868	-247.673
-260.437	-262.819	-282.876	-261.361	-277.027
-255.574	-224.518	-263.938	-252.302	-286.421

Вычисляем $\bar{x} = -261.3$, $S^{2\wedge} = 439.45$, а $S^{\wedge} = 20.96$. Дано $\alpha = 0.1$, $n = 20$.

Проверим гипотезу: $H_0: \mu = -267$ при $H_1: \mu \neq -267$. При неизвестном σ используем статистику $t = \frac{\bar{x} - \mu}{S^{\wedge}} \sqrt{n}$, которая, если верна гипотеза H_0 , имеет t -распределение Стьюдента с числом степеней свободы $n-1 = 19$. Используем таблицы t -распределение Стьюдента (см. приложение 3) для нахождения правосторонней и левосторонней критических областей $t_{\alpha} = 2.86$. Отсюда критические области $|t| > t_{\alpha}$ (см. Рис. 3.6).

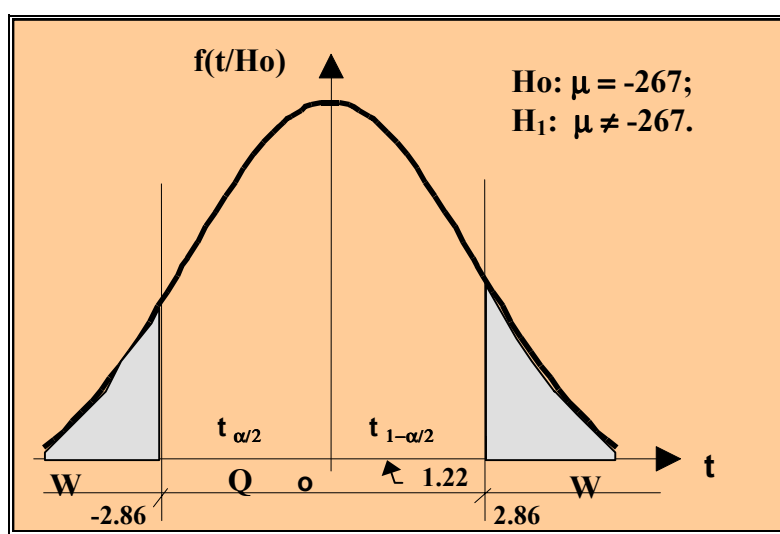


Рис. 3.6

Вычислим значение статистики t по выборке $t = \frac{-261.3 + 267}{20.96} \sqrt{20} = 1.22$.

Так как значение статистики не принадлежит критической области, то нет оснований отвергать гипотезу $H_0: \mu = -267$.

3. При решении в пункте 2 было получено, что $S^{2\wedge} = 439.45$.

Дано $\alpha = 0.01$, $n = 20$.

Проверим гипотезу: $H_0: \sigma^2 = 529$ при $H_1: \sigma^2 < 529$.

Статистика $\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$, которая, если верна гипотеза H_0 , имеет χ^2 -

распределение Пирсона с числом степеней свободы $n-1$. По альтернативной гипотезе H_1 найдем левостороннюю критическую область, удовлетворяющую условию $P(\chi^2 < \chi_{\alpha}^2) = \alpha = 0.01$. По таблице χ^2 -распределение Пирсона с числом

степеней свободы $n-1$ имеем $\chi_\alpha = 7.633$ (отметим, что таблицы составлены для противоположных условий т.е. $P(\chi^2 > \chi_\alpha) = 1 - \alpha = 0.99$).

Теперь вычислим статистику $\chi^2 = \frac{(20-1) \cdot 439.45}{529} = 15.78$.

Это значит, что значение не принадлежит критической области и гипотеза $H_0: \sigma^2 = 529$ не будет отвергнута ($15.78 > 7.633$).

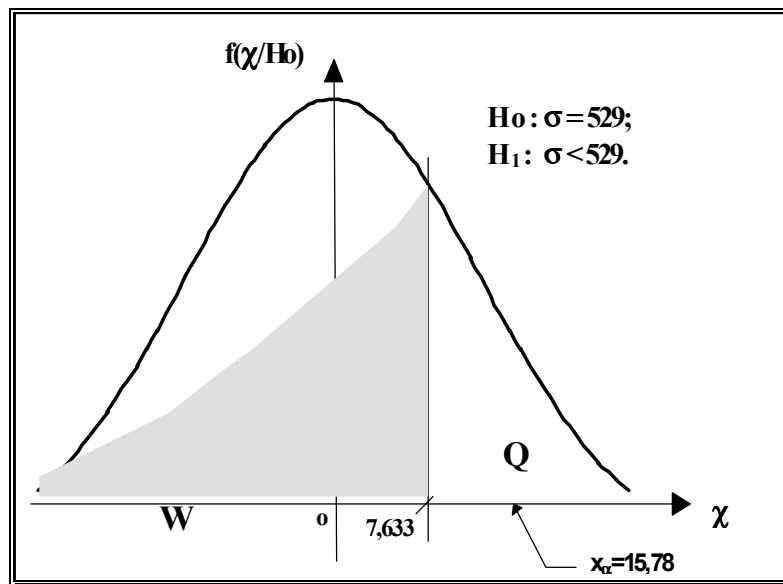


Рис. 3.7

4. При решении задач в предыдущих пунктах были использованы выборки из генеральной совокупности и получены вариационные ряды (см. табл. №2 из лабораторной работы №1). С уровнем значимости $\alpha = 0,01$ проверим гипотезу о *нормальном распределении* (H_0) генеральной совокупности. Напомним , что полученные значения $\bar{x} = 3.1312$, $S^2 = 2.28$, $S = 1.5$. Частоты последних значений вариационного ряда малы, поэтому объединяем их в один интервал (≥ 5). Количество интервалов $k = 6$.

Таблица №3.2

X_i^*	0.5	1.5	2.5	3.5	4.5	5.5
n_i	9	13	22	32	13	11
n_i/n	0.09	0.13	0.22	0.32	0.13	0.11

$$\text{Статистика } Q^2 = \sum_{i=1}^r \frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$$

имеет χ^2 распределение с $k - r - 1 = 6 - 2 - 1 = 3$ степенями свободы (r – число параметров распределения $F(x) \in N(x, s)$, рассчитанных по выборке). Критическая область имеет вид

$$P(Q^2 > \chi_\alpha) = 0.01, \chi_\alpha = 11.341, Q^2 > 11.341.$$

Вычисления выполним при помощи электронных таблиц, а результаты несем в таблицу №3.3 Вероятности p_i получим по таблицам нормального распределения (см. приложение 1).

Таблицу №3.3

X_i^*	n_i	p_i	$n \cdot p_i$	$n_i - n \cdot p_i$	$(n_i - n \cdot p_i)^2$	$\frac{(n_i - n \cdot p_i)^2}{n \cdot p_i}$
0.5	9	0.0593	5.93	3.07	9.4249	1.5894
1.5	13	0.1575	15.75	-2.75	7.5625	0.4802
2.5	22	0.2485	24.85	-2.85	8.1225	0.3269
3.5	32	0.2751	27.51	4.49	20.1601	0.7328
4.5	13	0.1784	17.84	-4.84	23.4256	1.3131
5.5	11	0.0773	7.73	3.27	10.6929	1.3833
Σ	100	0.9961	99.61			5.8257

Значение статистики $Q^2 = 5,8257$ не принадлежит критической области, гипотеза о нормальном распределении генеральной совокупности подтвердилась и нет оснований ее отвергать.

Лабораторная работа № 4.

ТЕМА: Корреляционный анализ.

Стохастическая связь состоит в том, что одна случайная переменная реагирует на изменение другой изменением своего закона распределения.

В практике статистических исследований часто рассматривают частный случай такой связи, называемый статистической связью. Изучение статистических зависимостей основывается на исследовании таких связей между случайными переменными, при которых значение одной случайной переменной, в среднем, изменяется в зависимости от того, какие значения принимает другая случайная переменная.

Чтобы изучить статистическую зависимость, нужно знать условное математическое ожидание случайной переменной. Для его оценки необходимо знать вид двумерного распределения (ξ, η) . Однако, суждение об аналитическом виде двумерного распределения, сделанное по отдельной ограниченной по объему выборке, может привести к серьезным ошибкам. Поэтому идут на упрощение и переходят от условного математического ожидания случайной переменной к условному среднему значению, т.е. принимается

$$M(\eta / \xi = x) = \bar{y}(x),$$

где $M(\eta / \xi = x)$ – математическое ожидание случайной переменной η при условии, что случайная величина ξ приняла значение x . Зависимость между одной случайной переменной и условным средним значением другой случайной переменной называется корреляционной зависимостью. Корреляционная зависимость имеет в исследовательской работе широкое применение. Она характеризуется формой и теснотой связи.

1. Определение формы связи. Понятие регрессии (линейной). Определить форму связи – это значит выявить механизм получения зависимой случайной переменной. При изучении математических зависимостей форму связи можно характеризовать функцией регрессии (линейной, квадратной, показательной и т. д.). Регрессия – это односторонняя стохастическая зависимость. Она устанавливает соответствие между случайными переменными. Например, при изучении потребления энергии (y) в зависимости от объема производства (X) речь идет об определении односторонней связи, следовательно, о регрессии.

Односторонняя стохастическая зависимость выражается с помощью функции, которая, для отличия ее от строгой математической функции, называется функцией регрессии или просто регрессией.

Отметим, что функция регрессии не обратима, если за стохастической связью не скрывается подлинная функциональная зависимость.

Для характеристики формы связи при изучении корреляционной зависимости пользуются понятием кривой регрессии. Кривой регрессии η по ξ

(или η на ξ) называется условное среднее значение случайной переменной η , рассматриваемой как функции от x , т.е.

$$\bar{y}(x) = f(x) = M(\eta / \xi = x).$$

Функция $\bar{y}(x)$ даст наименьшую среднюю погрешность оценки прогноза. Определение корреляционной связи было дано безотносительно к совместному закону распределения переменных (ξ, η) , т.е. корреляционную связь можно исследовать при любом законе распределения (ξ, η) . В дальнейшем мы остановимся на линейной регрессии.

Линейная регрессия или линейная форма связи между случайными переменными занимает особое место в теории корреляции. При такой форме связи $\bar{y}(x)$ есть линейная функция от x , имеет вид:

$$\bar{y}(x) = a_0 + a_1 \cdot x,$$

где a_0, a_1 - коэффициенты регрессии; x - независимая переменная.

Коэффициенты регрессии определяются по способу наименьших квадратов. В случае линейной регрессии за теоретическое значение принимается значение $\bar{y}(x)$, получаемое по формуле (15.2), т.е. ищется такая прямая линия, сумма квадратов отклонений измеренных значений по результатам опыта y_i от которой была бы минимальной

$$Q = \sum_{i=1}^n [y_i - \bar{y}(x)]^2 = Q_{\min}, \quad (*)$$

где $y_i (i = \overline{1:n})$ - значения η по результатам опыта.

Как известно, минимум функции можно найти, приравняв к нулю ее первую производную. Запишем (*) в виде

$$Q = \sum_{i=1}^n [y_i - a_0 - a_1 \cdot x_i]^2,$$

где $(x_i, y_i), (i = \overline{1:n})$ - наблюдения за (ξ, η) .

Находим частные производные функций $Q(a_0, a_1)$ по a_0 и a_1 приравниваем их нулю.

$$\begin{cases} \frac{\partial Q}{\partial a_0} = -2 \cdot \sum_{i=1}^n (y_i - a_0 - a_1 \cdot x_i) = 0, \\ \frac{\partial Q}{\partial a_1} = -2 \cdot \sum_{i=1}^n (y_i - a_0 - a_1 \cdot x_i) \cdot x_i = 0. \end{cases}$$

Сокращая оба уравнения на -2 и производя почленное суммирование, получим

$$\begin{cases} \sum_{i=1}^n y_i - \sum_{i=1}^n a_0 - \sum_{i=1}^n a_1 \cdot x_i = 0, \\ \sum_{i=1}^n x_i \cdot y_i - \sum_{i=1}^n a_0 \cdot x_i - \sum_{i=1}^n a_1 \cdot x_i^2 = 0. \end{cases}$$

Величины a_0, a_1 являются постоянными и следовательно

$$\begin{cases} n \cdot a_0 + a_1 \cdot \sum_{i=1}^n x_i = \sum_{i=1}^n y_i, \\ a_0 \cdot \sum_{i=1}^n x_i + a_1 \cdot \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i \cdot y_i. \end{cases} \quad (**)$$

Система (**) называется системой *нормальных уравнений*. Разрешив ее относительно a_0 и a_1 , получим значения коэффициентов регрессии:

$$\begin{aligned} a_0 &= \frac{\sum_{i=1}^n y_i \cdot \sum_{i=1}^n x_i^2 - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i \cdot x_i}{n \cdot \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}, \\ a_1 &= \frac{n \cdot \sum_{i=1}^n y_i \cdot x_i - \sum_{i=1}^n x_i \cdot \sum_{i=1}^n y_i}{n \cdot \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2}. \end{aligned} \quad (***)$$

Воспользуемся тем, что $\sum_{i=1}^n x_i = n \cdot \bar{x}$, следствие из

$$\bar{x} = \frac{1}{n} \cdot \sum_{i=1}^n x_i, \quad \sum_{i=1}^n y_i = n \cdot \bar{y}.$$

$$\sum_{i=1}^n x_i^2 = n \cdot \overline{x^2} - \text{следствие из } \overline{x^2} = \frac{1}{n} \cdot \sum_{i=1}^n x_i^2.$$

$$\sum_{i=1}^n x_i \cdot y_i = \sum_{i=1}^n n_{x_i \cdot y_i} \cdot (x_i \cdot y_i) \quad (\text{учтено, что пара чисел } (x, y) \text{ наблюдалась } n_{x_i \cdot y_i} \text{ раз})$$

$$\text{Обозначим через } \sigma_x^2 = \overline{x^2} - (\bar{x})^2.$$

Тогда из (**) a_1 имеет вид:

$$a_1 = \frac{\sum_{i=1}^n n_{x_i y_i} \cdot x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y}}{n \cdot \sigma_x^2}.$$

Умножим обе части последнего равенства на $\frac{\sigma_x}{\sigma_y}$:

$$a_1 \cdot \frac{\sigma_x}{\sigma_y} = \frac{\sum_{i=1}^n n_{x_i y_i} \cdot x_i \cdot y_i - n \cdot \bar{x} \cdot \bar{y}}{n \cdot \sigma_x \cdot \sigma_y},$$

a_1 - называется *коэффициентом регрессии*;

$a_1 \cdot \frac{\sigma_x}{\sigma_y} = r_b$ - *выборочный коэффициент корреляции*.

Тогда выборочное уравнение прямой линии регрессии η на ξ можно записать в виде

$$\bar{y}_x - \bar{y} = r_b \cdot \frac{\sigma_y}{\sigma_x} \cdot (x - \bar{x}).$$

Полученные здесь формулы пригодны для вычислений с помощью ЭВМ или микрокалькулятора, но вычисления “вручную” трудоемки для их упрощения используется корреляционная таблица.

З а м е ч а н и е 1. Аналогично находят выборочное уравнение прямой линии регрессии ξ на η

вида:

$$\bar{x}_y - \bar{x} = r_b \cdot \frac{\sigma_x}{\sigma_y} \cdot (y - \bar{y}),$$

где $r_b \cdot \frac{\sigma_x}{\sigma_y}$ – коэффициент регрессии.

З а м е ч а н и е 2 .

$$1) \quad |r_b| \leq 1$$

2) Если $r_b = 0$, и выборочные линии регрессии – прямые, то ξ и η не связаны линейной корреляционной зависимостью, возможно они связаны нелинейной или даже функциональной зависимостью.

3) а) Таким образом выборочный коэффициент корреляции характеризует тесноту линейной связи между количественными признаками в выборке: чем ближе $|r_b|$ к 1, тем связь сильнее и при $|r_b| = 1$ переходит в функциональную зависимость;

б) Знак выборочного коэффициента корреляции совпадает со знаком выборочных коэффициентов регрессии

$$r_b^2 = \sqrt{a_1^{(x)} \cdot a_1^{(y)}},$$

где $a_1^{(x)}$ и $a_1^{(y)}$ – соответственно коэффициенты регрессии.

Пример выполнения лабораторной работы:

В значительной мере на качество результатов исследования влияет статистическая обработка данных. Но в зависимости от цели и задач исследования устанавливают принцип отбора данных. Очень важно принять решение о проведении исследования по всей генеральной совокупности или по выборке из нее. Если не будет обеспечено необходимое количество статистических данных, то следует возвратиться к постановке задачи. Необходимо составить детальный план сбора исходной информации с учетом полной схемы статистического анализа.

Иногда задача ставиться в упрощенном виде, где исследуется влияние одного фактора на результирующий признак. А потом постепенно добавляются

последовательно другие факторы. Отбор факторов, оказывающих наиболее существенное влияние на результативный признак, проводится на основе исследования связности (тесноты, силы, строгости, интенсивности) двух и более явлений.

Рассмотрим, пример, зависимости производительности труда (y) от уровня механизации работ (x) по данным 14 промышленных предприятий (см. табл.4.1).

Таблица №4.1.

Предприятие	Производительность труда	Коэффициент механизации (%)
i	y_i	x_i
1	20	32
2	24	30
3	28	36
4	30	40
5	31	41
6	33	47
7	34	56
8	37	54
9	38	60
10	40	55
11	41	61
12	43	67
13	45	69
14	48	76

Построим корреляционное поле, чтобы определить тенденцию изменения зависимой переменной в зависимости от фактора (x_i - уровень механизации).

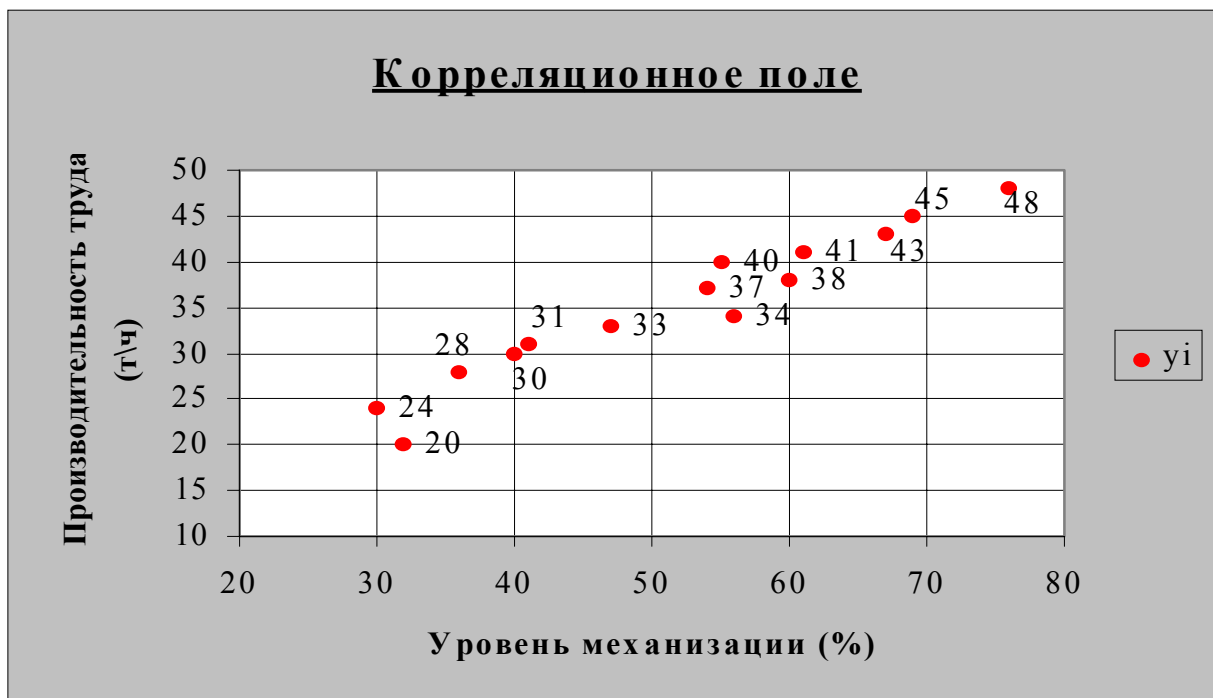


Рис. 4.1.

Скопление точек определяет в некотором смысле зависимость двух переменных. Систематизация опытного материала в виде диаграммы дает возможность высказать гипотезу о зависимости переменных. В этом примере эта тенденция имеет явно линейный характер. Поэтому попытаемся аппроксимировать рассматриваемую зависимость линейной функцией регрессии.

Формально коэффициенты корреляции могут быть вычислены для любой двумерной системы наблюдений; они являются измерителями степени тесноты линейной статистической связи между анализируемыми признаками. Однако только в случае *совместной нормальной распределенности* исследуемых величин коэффициент корреляции имеет четкий смысл как характеристика степени тесноты связи между ними. В остальных случаях коэффициент корреляции можно использовать в качестве одной из возможных характеристик.

Введем исходные данные в ЭВМ и получим статистическое описание исходных совокупностей (см. табл. №4.2). Аспекты реализации обработки экономической информации на ЭВМ здесь не рассматриваются так, как к настоящему времени накоплен определенный опыт по программному обеспечению прикладной математической статистике и можно использовать любой пакет прикладных программ.

Таблица №4.2.

Описательная статистика	y_i	x_i
Среднее	35.14286	51.71429
Стандартная ошибка	2.158187	3.846546
Медиана	35.5	54.5
Стандартное отклонение	8.075196	14.39246
Дисперсия выборки	65.20879	207.1429
Эксцесс	-0.55977	-1.07092
Асимметричность	-0.25996	-0.00504
Интервал	28	46
Минимум	20	30
Максимум	48	76
Сумма	492	724
Счет	14	14
Наибольший(1)	48	76
Наименьший(1)	20	30
Уровень надежности(95%)	4.229962	7.539081

По результатам вычислений характеристик выборочных наблюдений можно делать вывод о соответствующих параметрах генеральной совокупности.

Теперь вычислив параметры линейной регрессии и получим уравнение:

$$y = 0.5435x + 7.0361$$

Нанесем уравнение линейной регрессии на корреляционное поле

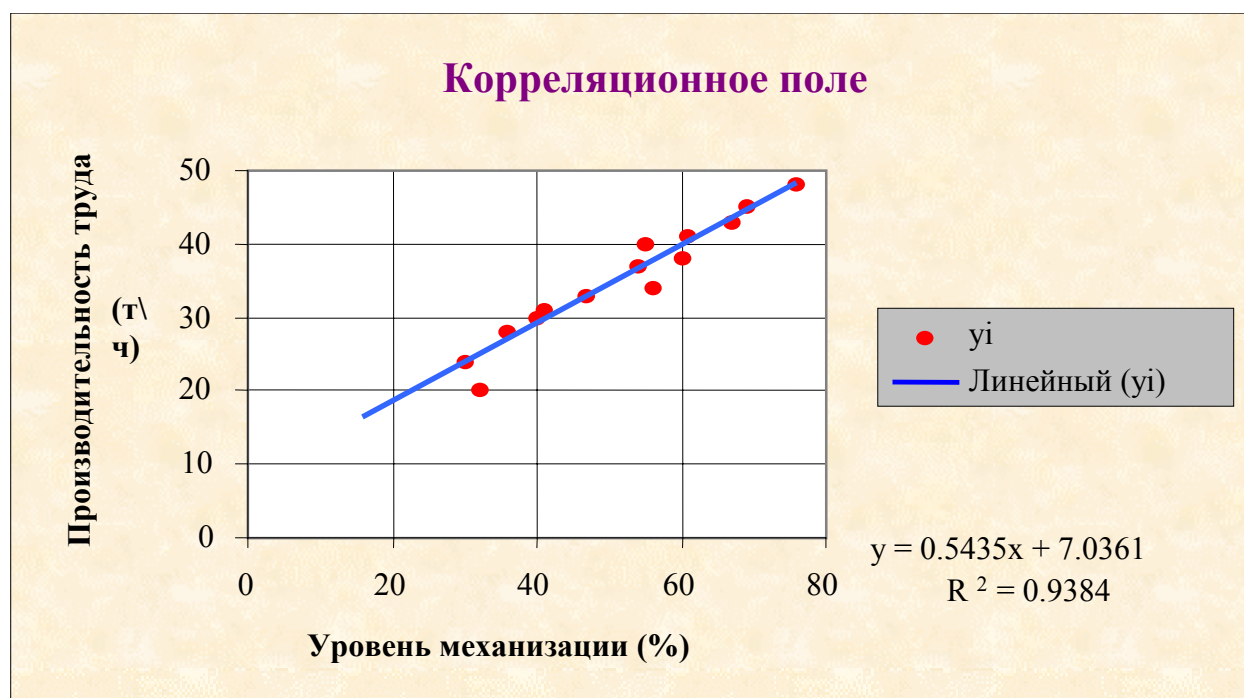


Рис. 4.2.

В качестве характеристики точности оценки регрессии или степени согласованности расчетных значений найдем значения регрессии в точках x_i и определим остатки.

Таблиц № 4.3.

Предприятие	Производи- тельность труда	Производи- тельность труда	Остатки $y_i - y_i^{\wedge}$
i	$y_i^{\wedge}$	y_i	
1	24.753	20	-4.753
2	23.666	24	0.334
3	26.927	28	1.073
4	29.101	30	0.899
5	29.6445	31	1.3555
6	32.9055	33	0.0945
7	37.797	34	-3.797
8	36.71	37	0.29
9	39.971	38	-1.971
10	37.2535	40	2.7465
11	40.5145	41	0.4855
12	43.7755	43	-0.7755
13	44.8625	45	0.1375
14	48.667	48	-0.667

Рассматривая остатки как отклонения i -х наблюдений от значений, которые следует ожидать в среднем, можно сделать ряд практических выводов, например, провести дополнительные исследования по предприятиям 1,3,5,7,9,10.

Примечание. “Ручной ” метод здесь не рассматривается, выполнение этой работы предполагается на ЭВМ. Использование корреляционной таблицы для нахождения выборочного коэффициента корреляции и коэффициентов функции линейной регрессии рассмотрено в приложении 6.

Задания

к лабораторной работе №1.

В работе, используя выборку, решить следующие задачи:

- 1. Построить вариационный ряд (дискретный и интервальный);**
- 2. Вычислить относительные частоты (частости) и накопленные частоты;**
- 3. Построить графики вариационного ряда (полигоны и гistogramмы относительных и абсолютных частот);**
- 4. Составить эмпирическую функцию распределения, построить ее график;**
- 5. Вычислить числовые характеристики вариационного ряда:**
 - *среднее арифметическое,*
 - *дисперсию,*
 - *стандартное отклонение,*
 - *моду,*
 - *медиану.*

Варианты

к лабораторной работе №1.

Вариант №1.

10.4	10.4	9.75	10.5	9.86	10.6	10.7	9.79	10.6	10
10.1	10.1	10.9	9.77	9.81	11.2	8.93	10.6	10.1	9.52
10.5	10	9.22	9.5	10.2	10.5	9.92	9.28	10	10.5
9.58	9.94	9.66	9.96	10.1	9.75	9.8	9.02	10.1	10.2
10.2	9.42	9.86	9.94	10.3	9.88	9.77	9.39	10.3	9.21
10.5	10.1	10.2	10.3	9.16	9.72	10.8	9.46	11.2	9.61
10.1	10.2	9.2	9.76	9.74	9.88	10.9	9.83	9.37	9.69
9.71	9.13	9.48	9.91	9.97	9.89	9.04	9.04	10.6	9.59
10.2	9.73	9.43	9.42	11.3	9.42	9.55	9.92	10.4	10.5
9.41	9.99	9.99	9.5	8.71	10.2	9.27	8.96	9.86	9.91

Вариант №2.

26.7	24.6	24.6	25.4	25.1	25.3	24.8	25.8	26	23.6
24.9	26.8	25.5	24.6	24.9	25.9	25.3	23.3	25.3	23.6
28.1	24.4	23.7	26.4	24.6	25.2	27.4	24.4	25.7	24.4
26.4	25.6	24.6	27.1	24.1	24.5	25.1	21.2	23.8	25.8
23.3	24.1	27.2	25.5	25.7	22.4	25.3	25.5	25.2	26.8
22.7	25.8	23.7	25.7	25.9	24.5	25.7	25	25.4	25.4
25.4	23	23.6	27.8	26.2	27.3	26.5	25.7	24.7	26.5
25.6	24.5	26.2	25.5	23.3	24	23.9	26.6	21.9	25.1
24.4	26	24.9	27.7	23	24.8	26.2	23.9	26.1	27.2
24.6	24.5	24.9	25.3	24.5	24.5	25.4	23.9	25.3	24.1

Вариант №3.

28.3	31.4	31.4	31.5	32.3	30.2	30.5	29.5	30.4	29.5
30.6	28.6	32.7	25.5	30.2	30	33.8	30.7	30.4	32.7
32.1	29.7	29.1	29.2	28.7	34.2	27.7	30.6	29.6	28
27.7	32	26.5	31.3	31.8	31.4	30.7	29.6	33.1	27.5
30.9	30.3	31.3	30.4	32.3	32.4	32.2	30.3	28.2	28
31.2	27.5	28.6	30.4	31.2	30.1	30.4	29.4	30.5	27
28.5	29.3	30.4	31.8	31.5	30.5	27.3	30.2	27.5	31.9
28.3	28.5	29.3	31.4	31.6	29.7	30.7	30	29.8	30.3
31.2	31.1	30.7	30.1	30	28.4	29.2	28.3	30.1	29.9
31.3	28.4	30.3	31.7	29.8	29	27.9	31.8	29.1	31.7

Вариант №4.

81.8	80.2	81.4	77.5	80.6	81.7	80.5	77.3	80.1	80.7
77.8	80.3	80.9	80.4	80.8	80.7	77.7	77.3	77.9	80.8

80.2	83.4	80	74.9	80.5	78.3	81.2	78.5	82.7	80.7
78.4	83.1	81.3	78.4	79.2	79.2	78.8	79.2	77.6	82.8
80.3	79.8	83.9	79.3	76.9	83.5	82.4	79.3	77.2	80.6
81.3	80.1	79.1	79.5	79.2	80.2	80.6	76	83	80.6
79	78.9	80.2	78.6	80.4	81.8	78.4	80.4	79.8	79.1
80.5	80.6	79.4	80.6	77.3	80.6	79	78.2	79	78.5
79.3	80.4	83	82.2	80	84.5	84.4	80.5	78.4	78.1
79.2	76.6	77.4	79.6	79.3	78.7	78.9	79.1	78.4	79.4

Вариант №5.

13.5	14	14.6	13.6	14.9	15.2	14.5	15.8	15.6	15.3
16.6	15	15.6	15.1	15.2	14.6	15	15.8	14.9	14.5
14.4	15.3	15.5	14.7	16.1	15.3	14.6	16.1	16	13.9
14.3	14.3	14.6	15.7	13.9	15.2	15.4	14.6	14.3	14.2
14.6	15.5	15.4	16.5	16.3	15.9	14.4	15.7	15.8	15.8
14.8	15.2	15.2	14.1	14.8	16	15.1	14.9	14.1	15.2
15.5	15.2	14.7	14.1	14.3	14.7	14.3	15.8	14.4	13.5
15.4	14.4	15.4	15.3	14.4	15.2	15.8	15.2	15.4	15.3
15.1	14.5	15.6	14.1	15.7	15	14.6	14	13.7	15
14.5	15.5	14.7	15.1	14.3	15.8	14.9	13.2	14.4	14.3

Вариант №6.

126	126	125	126	125	125	125	125	125	125
125	125	125	125	126	125	126	124	125	125
125	125	125	125	125	125	125	125	125	125
125	126	126	125	125	125	125	126	125	125
125	125	126	125	125	125	125	125	125	126
125	125	124	124	124	125	124	126	126	125
125	126	125	126	124	125	125	125	125	125
124	124	126	126	126	125	125	125	125	124
125	125	125	125	125	125	125	125	125	125
124	125	124	125	125	125	124	125	126	124

Вариант №7.

17.6	20.1	21.2	25.2	20.4	19.7	19.1	21.5	22.5	19.2
20.7	20.8	16.3	18.8	25	19.1	19.7	16.9	15.6	22.1
22.2	26.9	19	22.6	21.5	15.5	18	18	15.5	21
22.4	18	17.7	15.8	19.2	15.4	18.9	20.2	24.4	17.4
23.2	18.5	18.1	20.3	17.1	22.9	22.6	18.4	20.8	22.3
13.9	14.2	23.2	20.2	18.6	16.1	19.3	17.4	20.8	23.4
22.5	21.7	19.9	22.1	18.2	17.3	18.9	26.2	19.4	22.1
24.5	19.4	20.6	23.3	18.8	22.3	24.5	24.9	21.5	21.3
20.6	18.6	15.8	16.4	25.4	23.9	19.1	19.9	16.4	19.2
20.2	18	20.3	22.6	20.7	24.7	17.8	18.3	23.3	22

Вариант №8.

199	200	200	198	200	197	199	200	201	195
204	196	200	202	204	202	201	200	198	200
198	199	204	201	201	203	198	199	203	198
202	199	202	201	204	200	200	201	199	204
197	201	199	196	197	198	197	201	201	198
202	196	202	199	203	200	198	200	200	201
203	196	201	197	201	203	200	200	203	199
198	199	201	203	201	200	197	200	202	205
200	206	202	198	202	201	198	201	196	200
197	198	198	200	202	200	200	200	200	201

Вариант №9.

552	549	549	552	550	552	549	549	550	549
550	550	550	547	550	551	551	550	549	552
547	550	551	552	549	551	550	550	551	549
552	550	550	552	548	549	549	551	551	550
549	549	549	549	550	547	548	549	552	554
551	549	551	549	549	550	552	548	549	550
551	549	549	550	551	549	551	550	552	550
549	549	551	551	550	550	549	550	550	550
551	547	548	552	551	550	551	550	550	551
551	552	551	549	551	547	550	551	548	549

Вариант №10.

57.6	54.5	53.6	57.5	54.2	56.1	53.3	54.5	57.3	54.9
54.2	55.7	56	55.3	52	58.7	53.8	54.8	56.4	54.2
50.8	52.8	55.8	56	53.1	54.1	57.6	55.5	55.9	54.5
56.5	54.6	57.4	54	52.1	58.5	54	54.6	55.9	55.8
54.7	54.1	55.8	52.3	51.6	55.5	54.3	55.3	55.1	56.7
56.9	55.4	55.1	55.7	52.5	54.4	56.2	55.9	55.4	53.2
54.5	53.9	51.6	53.5	53.6	54.8	55.4	55.4	55.3	54.7
56.2	54.2	56	53.8	58.8	54.2	57	55.6	53.5	58.9
55.1	55	55.4	55.5	52.9	56.3	54	52.5	53.9	55.7
55.8	53.7	53.5	53.9	55.1	51	58.2	55.6	55	55

Вариант №11.

71	74.9	78.7	76.2	75.1	73	76.8	73.5	72.8	72
73	76	77.8	74.5	75.5	77.1	73	74.7	75.5	74.5
78.1	76.6	74.6	74.9	78.2	74.4	77.6	73.5	78.1	79.6
76.7	73.7	75.8	77.9	73.3	79.3	75.5	73.1	76.3	76.3
75.5	72.3	77.7	73.2	75.3	73.2	75.9	76.3	72.9	72.7
74.7	74.6	74.9	73.5	75.9	76.3	75.3	73.2	77.8	73.9
73.9	75.9	71.8	77	74.8	74.3	70.2	75.6	75.5	76.6
74.8	72.8	77.6	72.7	76.2	74.8	77	76.3	75.1	75.2
76	74.7	75.6	72.2	71.8	76.9	73.5	76.5	72.6	77.3
76.8	76.2	74.4	73.3	73.5	80.6	74.4	73.4	76.7	75.5

Вариант №12.

143	144	145	144	145	145	144	146	146	146
146	146	144	145	143	144	145	146	145	145
147	146	144	146	146	145	145	145	145	145
144	144	144	144	145	146	145	145	145	145
144	145	144	147	146	145	145	144	145	144
144	144	144	145	145	146	145	145	146	145
144	145	145	145	144	144	144	144	146	146
146	146	146	145	145	147	143	147	146	144
145	145	145	146	146	145	146	145	145	148
144	146	144	145	145	145	143	146	146	147

Вариант №13.

44.3	40	47.1	44.8	48.6	47.9	44.3	46.6	46.3	42.7
50.7	46.5	42.7	44.5	46.2	45.3	38.8	45.7	41.3	43.6
43.8	45.9	46	47	44	40.5	44.2	46.5	42.4	42.9
44	42.5	47.3	42.8	43.5	41.7	48.8	43.5	44.8	40.9
44.8	44.9	44.6	43	45.4	41	49.7	45.3	47.4	43.2
43.1	41.3	47	48.9	45.1	43	45.5	45.9	42.5	47.6
45.7	46.4	42.6	43.2	45.2	43.3	43.3	44.9	45.3	44.9
46.6	46	44.8	46.8	45.9	44.2	44.4	42.7	44.5	42.5
43.8	47.1	43.7	43.5	43.3	46.1	44.7	46.6	44.7	48
49.5	47.3	45	46.3	44.4	45.3	43.4	43.5	47.1	44.5

Вариант №14.

344	344	350	347	344	346	342	343	342	347
347	345	349	345	343	345	348	344	347	351
345	350	349	346	346	344	348	344	345	347
347	342	346	349	345	349	346	345	345	340
346	346	347	346	346	343	343	348	350	345
346	349	347	351	349	343	343	346	346	343
343	347	343	347	347	341	342	345	345	346
345	347	344	344	346	347	348	349	341	342
344	346	342	348	345	345	345	347	345	344
346	349	344	343	347	348	344	343	347	343

Вариант №15.

81.8	80.2	81.4	77.5	80.6	81.7	80.5	77.3	80.1	80.7
77.8	80.3	80.9	80.4	80.8	80.7	77.7	77.3	77.9	80.8
80.2	83.4	80	74.9	80.5	78.3	81.2	78.5	82.7	80.7
78.4	83.1	81.3	78.4	79.2	79.2	78.8	79.2	77.6	82.8
80.3	79.8	83.9	79.3	76.9	83.5	82.4	79.3	77.2	80.6
81.3	80.1	79.1	79.5	79.2	80.2	80.6	76	83	80.6
79	78.9	80.2	78.6	80.4	81.8	78.4	80.4	79.8	79.1
80.5	80.6	79.4	80.6	77.3	80.6	79	78.2	79	78.5
79.3	80.4	83	82.2	80	84.5	84.4	80.5	78.4	78.1
79.2	76.6	77.4	79.6	79.3	78.7	78.9	79.1	78.4	79.4

Задания

к лабораторной работе №2.

В работе, используя выборку лабораторной работы №1, решить следующие задачи:

1. Вычислить по выборке различные оценки параметра μ - среднего значения, определить наилучшие оценки параметров μ - среднего значения, σ^2 - дисперсии, σ - стандартного отклонения генеральной совокупности $\bar{x}$, S^2 , S ;

2. Найти доверительные интервалы для μ при различной доверительной вероятности $\beta = 0,8; 0,95; 0,99$;

3. Считая данные выборки пробными, определить минимальный объем выборки n для нахождения доверительного интервала среднего значения - μ длины 2Δ (изменяя Δ : увеличить и уменьшить) и доверительной вероятностью $\beta = 0,8; 0,9; 0,95; 0,99; 0,995$;

Рекомендации. При обработке данных необходимо использовать электронные таблицы, например, Excel_5.

Задания

к лабораторной работе №3.

В задачах этой работы используются выборки из генеральных совокупностей с нормальным распределением $X \in N(\mu, \sigma)$ (за исключением задачи 3.5).

По данной выборке при уровне значимости α проверить гипотезы в задачах:

3.1. Пологаем, что σ известно и нужно проверить гипотезу $H_0: \mu = \mu_0$, а в качестве альтернативной гипотезы можно использовать одну из следующих гипотез $H_1: \mu < \mu_0$, $H_1: \mu > \mu_0$ или $H_1: \mu \neq \mu_0$.

Критическая область определяется с помощью таблицы Функции распределения нормированного нормального распределения $\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$.

Решить задачу, используя выборку и полученные результаты лабораторной работы №1, μ_0 взять равным ближайшему целому числу к $\bar{x}$, а $\alpha = 0,01$.

3.2. Проверим гипотезу: $H_0: \mu = \mu_0$ при $H_1: \mu \neq \mu_0$. При неизвестном σ используем статистику $t = \frac{\bar{x} - \mu}{S/\sqrt{n}}$.

Критическая область определяется с помощью таблицы Функции t -распределения Стьюдента с числом степеней свободы $n-1$, где n -объем выборки.

3.3. Проверим гипотезу о дисперсии нормального распределения $X \in N(\mu, \sigma)$.

$H_0: \sigma^2 = \sigma_0^2$ при $H_1: \sigma^2 < \sigma_0^2$.

Критическая область определяется с помощью таблицы Функции χ^2 -распределение Пирсона с числом степеней свободы $n-1$, где n -объем выборки.

3.4. Проверить гипотезу о распределении генеральной совокупности X . Необходимо по выборке или учитывая другие соображения составить гипотезу H_0 о распределении генеральной совокупности. Для проверки гипотезы использовать χ^2 -критерий согласия Пирсона.

3.5. Проверка гипотезы о равенстве центров распределения двух нормальных генеральных совокупностей при известном σ .

Критическая область определяется с помощью таблицы Функции t -распределения Стьюдента с числом степеней свободы $n_1 + n_2 - 2$, где n_1, n_2 -объемы выборок.

Варианты

к лабораторной работе №3.

Задачи 3.2 и 3.3:

Вариант №1. ($\mu_0 = 1$; $\sigma_0^2 = 4$; $\alpha = 0,02$).

2.52	1.25	3.65	5.70	0.60	0.03	-0.44	4.81	1.66	0.37
-0.76	2.21	-0.69	1.44	2.94	-1.37	3.33	1.00	0.13	-1.30

Вариант №2. ($\mu_0 = 3$; $\sigma_0^2 = 1$; $\alpha = 0,05$).

3.59	3.85	2.60	3.63	3.43	2.52	3.46	2.92	1.47	3.08
3.66	2.57	2.73	1.70	2.41	2.28	2.58	2.99	1.94	3.33

Вариант №3. ($\mu_0 = 5$; $\sigma_0^2 = 9$; $\alpha = 0,01$).

2.84	6.63	10.55	7.86	6.10	1.88	1.09	5.13	3.88	7.07
8.05	6.49	4.11	1.97	8.23	-2.12	4.41	4.50	6.73	5.43

Вариант №4. ($\mu_0 = 45$; $\sigma_0^2 = 4$; $\alpha = 0,02$).

49.38	50.25	46.17	43.90	45.43	47.55	45.98	43.29	48.82	45.70
46.50	45.99	43.18	45.90	43.28	39.90	49.60	41.90	45.26	44.77

Вариант №5. ($\mu_0 = 50$; $\sigma_0^2 = 1$; $\alpha = 0,2$).

50.32	49.46	49.08	49.63	50.93	51.45	50.05	49.85	50.01	49.51
49.87	50.86	50.53	51.20	49.78	49.46	49.52	50.92	49.21	49.34

Вариант №6. ($\mu_0 = 135$; $\sigma_0^2 = 9$; $\alpha = 0,002$).

134	138	133	137	133	137	133	131	139	133
133	138	133	135	133	140	130	131	133	135

Вариант №7. ($\mu_0 = 15$; $\sigma_0^2 = 2$; $\alpha = 0,01$).

13	16	13	15	15	13	16	17	14	16
17	16	15	16	15	18	17	16	14	15

Вариант №8. ($\mu_0 = 25$; $\sigma_0^2 = 9$; $\alpha = 0,05$).

28	17	31	25	32	26	31	26	26	25
27	25	27	25	25	25	25	24	25	25

Вариант №9. ($\mu_0 = 105$; $\sigma_0^2 = 3$).

106	104	107	104	107	104	105	103	104	107
104	104	104	105	105	105	105	103	106	107

Вариант №10. ($\mu_0 = 250$; $\sigma_0^2 = 4$; $\alpha = 0,01$).

252	247	252	248	249	250	250	251	250	247
247	254	245	251	250	248	252	247	249	251

Вариант №11. ($\mu_0 = 500$; $\sigma_0^2 = 16$; $\alpha = 0,1$).

495	499	495	496	496	508	501	490	505	507
499	510	499	504	501	509	496	497	503	505

Вариант №12. ($\mu_0 = 80$; $\sigma_0^2 = 1$; $\alpha = 0,002$).

79	79	79	80	82	77	80	79	80	81
81	78	79	81	80	80	80	79	81	78

Вариант №13. ($\mu_0 = 12$; $\sigma_0^2 = 9$; $\alpha = 0,02$).

10.02	13.63	11.90	11.89	11.57	5.88	13.96	13.12	8.21	8.80
10.55	12.96	11.26	9.67	15.64	12.78	14.98	17.05	14.19	11.04

Вариант №14. ($\mu_0 = 74$; $\sigma_0^2 = 2$; $\alpha = 0,001$).

74.97	75.80	74.82	76.38	73.93	76.28	73.79	76.72	73.85	74.78
76.24	73.83	73.29	74.23	73.80	74.01	76.13	71.30	73.42	75.66

Вариант №15 ($\mu_0 = 320$; $\sigma_0^2 = 0,25$; $\alpha = 0,005$).

320	320	321	320	320	320	320	321	320	320
320	320	320	320	319	320	320	320	320	319

Задача 3.4. ($\alpha = 0,05$)

Вариант №1.(Г)

29.71	30.14	29.83	30.36	29.71	29.94	30.76	29.61	30.75	30.93
30.16	30.52	30.17	29.80	30.21	30.07	30.26	29.68	29.65	30.18
30.79	30.27	30.67	30.29	29.68	29.99	29.89	30.93	29.00	31.10
30.68	29.24	30.41	29.69	29.31	29.02	30.05	29.06	31.00	29.92
30.23	29.91	30.77	29.27	29.97	29.91	31.05	30.02	29.50	30.17

Вариант № 2.(П)

1	2	2	2	0	1	1	0	3	3
4	4	2	5	1	0	0	2	0	4
3	1	4	2	1	1	1	2	1	2
2	4	1	2	3	2	3	3	2	0
2	2	3	1	3	0	3	0	4	3

Вариант № 3.(Р)

10.01	12.82	10.97	14.04	12.93	12.40	11.75	14.48	14.11	11.70
10.87	14.29	13.55	12.57	11.52	10.07	10.46	11.82	10.74	13.52
14.94	12.23	10.60	10.02	10.04	11.89	12.66	12.86	13.01	10.76
10.83	13.32	12.25	11.76	10.29	13.04	13.92	14.01	12.60	12.38
14.38	13.63	14.78	14.63	12.70	10.71	12.31	11.18	14.31	14.15

Вариант № 4.(Г)

70.20	70.22	68.16	69.78	70.59	69.35	68.56	71.19	69.26	69.68
69.23	70.44	71.18	70.55	68.78	69.82	70.06	70.76	70.27	70.19
68.39	69.46	70.84	68.72	70.77	68.49	69.83	71.47	69.16	70.13
70.61	68.70	69.51	69.57	70.06	67.80	70.46	70.52	69.90	71.27
69.79	69.10	71.96	71.60	70.09	69.79	69.75	72.11	71.00	69.36

Вариант № 5.(П)

8.00	2.00	7.00	3.00	2.00	4.00	3.00	4.00	6.00	3.00
2.00	1.00	1.00	2.00	3.00	2.00	4.00	0.00	8.00	6.00
6.00	5.00	4.00	2.00	1.00	4.00	4.00	6.00	6.00	2.00
3.00	3.00	6.00	2.00	2.00	5.00	4.00	2.00	1.00	4.00

5.00	5.00	2.00	2.00	5.00	6.00	2.00	11.00	7.00	3.00
------	------	------	------	------	------	------	-------	------	------

Вариант № 6.(P)

50.03	72.29	68.46	63.58	72.50	64.99	61.13	70.17	58.17	63.97
54.24	71.51	59.65	50.46	74.53	63.60	72.75	73.17	55.08	72.10
58.42	72.95	64.43	55.47	61.36	67.88	54.55	51.85	72.67	55.73
56.45	57.70	71.90	53.42	64.95	64.18	71.15	55.50	66.49	64.44
71.27	56.09	51.26	57.64	52.86	66.41	63.79	63.03	62.33	72.46

Вариант № 7.(Би 0,25: 100)

16.00	31.00	20.00	25.00	25.00	15.00	22.00	24.00	22.00	30.00
30.00	24.00	25.00	27.00	26.00	28.00	22.00	23.00	29.00	23.00
18.00	19.00	22.00	23.00	20.00	28.00	32.00	33.00	28.00	26.00
21.00	30.00	20.00	20.00	22.00	31.00	27.00	26.00	25.00	30.00
29.00	21.00	23.00	16.00	30.00	24.00	21.00	32.00	23.00	27.00

Вариант № 8.(H)

171	168	170	168	175	177	168	174	171	170
165	170	169	169	171	165	171	171	170	173
164	171	167	172	172	170	168	169	175	168
165	173	174	169	167	168	171	172	170	174
165	173	175	165	173	168	174	167	167	174

Вариант № 9.(П)

12	6	6	12	15	4	9	4	5	8
8	10	11	10	12	8	7	7	5	11
11	13	11	15	9	8	12	8	6	7
9	4	8	13	9	11	7	16	8	11
14	9	10	9	8	13	8	8	12	13

Вариант № 10.(Г)

452	449	450	448	452	449	448	452	451	451
454	448	449	449	449	451	451	451	451	450
454	450	452	454	455	452	450	449	450	446
448	448	450	449	450	451	448	448	454	451
452	451	449	449	450	453	453	449	447	449

Вариант № 11.(P)

152	156	156	154	159	167	154	151	168	172
161	162	168	160	167	175	155	163	164	173

175	167	174	170	167	158	150	167	163	162
166	170	156	154	160	156	154	164	163	151
157	167	171	161	168	173	161	165	166	154

Вариант № 12.(Н)

46	44	44	44	46	45	47	45	46	46
45	47	44	43	45	47	47	45	46	45
43	45	45	46	45	45	46	44	45	44
44	44	45	45	45	45	46	46	45	46
43	47	46	46	46	45	48	46	44	46

Вариант № 13.(Б 0,51:100)

51	58	61	60	61	45	58	55	48	56
60	49	52	48	48	55	50	48	55	50
55	54	57	52	52	52	50	49	53	46
50	60	56	42	45	45	52	56	58	50
43	41	53	50	58	53	50	47	60	52

Вариант № 14.(Г)

145.4	144.5	144.4	146.3	145.3	144.7	146.6	144.6	143.0	142.7
143.4	145.9	144.7	146.2	144.3	147.7	148.7	142.7	146.3	146.0
142.7	142.7	144.6	147.4	145.8	146.0	145.5	144.5	147.8	148.1
146.1	144.7	144.9	143.2	144.3	147.0	144.2	143.1	146.5	145.0
144.9	142.6	145.3	147.8	144.6	146.5	144.9	141.1	146.9	145.9

Вариант № 15.(П)

3.0	4.0	1.0	0.0	2.0	3.0	0.0	1.0	1.0	0.0
0.0	0.0	2.0	1.0	1.0	3.0	4.0	1.0	2.0	1.0
0.0	4.0	2.0	0.0	1.0	2.0	2.0	1.0	2.0	2.0
1.0	8.0	2.0	1.0	3.0	1.0	1.0	2.0	1.0	1.0
6.0	0.0	3.0	2.0	0.0	2.0	0.0	2.0	1.0	0.0

Задача 3.5. ($\alpha = 0,01$)

Вариант №1. ($\sigma_1 = \sigma_2 = 2$; $n_1 = 20$, $n_2 = 40$).

X

148.3	147.1	147.5	141.3	147.7	146.3	147.6	144.6	142.0	145.2
141.9	149.9	146.0	144.5	144.7	146.0	142.6	142.7	145.5	146.9

Y

143.1	145.4	143.8	141.5	143.3	145.0	143.5	148.2	144.3	143.4
147.3	145.2	145.7	146.0	143.7	147.0	144.7	146.1	143.8	139.7
142.5	145.8	143.0	145.3	146.0	149.2	144.6	146.0	146.1	144.5
148.8	150.2	144.9	145.2	143.8	146.3	144.4	147.3	145.3	148.6

Вариант №2. ($\sigma_1 = \sigma_2 = 3$; $n_1 = 20$, $n_2 = 40$).

X

44.2	48.2	43.2	46.2	47.9	49.1	48.5	43.1	46.9	42.7
38.8	49.5	41.3	45.8	51.7	49.0	45.2	43.2	46.9	48.9

Y

46.0	46.0	47.1	48.8	47.6	44.7	40.9	45.1	43.7	42.8
50.1	44.7	49.9	45.1	42.4	39.0	44.5	46.5	45.8	46.9
44.4	44.9	49.5	45.0	44.8	40.4	46.6	47.4	41.8	46.1
43.0	43.9	44.2	50.4	43.3	43.1	43.8	50.3	42.8	44.7

Вариант №3. ($\sigma_1 = \sigma_2 = 1$; $n_1 = 20$, $n_2 = 40$).

X

25.8	26.9	24.7	23.5	25.0	24.6	24.9	24.4	26.2	25.0
24.0	26.0	23.9	26.0	23.9	24.5	24.6	24.2	24.2	24.8

Y

24.3	24.6	26.0	25.0	25.3	22.5	25.3	24.4	26.0	23.7
25.3	25.1	24.8	25.1	24.6	24.2	26.4	27.5	22.6	24.1
24.2	24.4	24.2	24.6	25.8	25.5	25.1	24.1	24.7	24.8
23.9	25.7	23.9	25.2	23.1	26.2	25.9	24.1	26.2	25.7

Вариант №4. ($\sigma_1 = \sigma_2 = 4$; $n_1 = 20$, $n_2 = 40$).

X

123.0	125.9	128.9	127.5	125.8	122.8	127.6	129.1	123.9	120.9
129.8	123.2	126.5	124.4	128.2	129.3	120.0	117.9	117.7	122.7

Y

121.1	131.7	128.1	123.7	124.6	125.8	122.6	125.8	121.4	125.1
128.0	120.5	128.0	121.0	122.3	129.0	121.6	126.1	126.4	131.7
123.7	127.4	123.6	121.9	128.0	127.7	123.7	118.4	128.4	120.7
125.5	128.3	122.0	118.8	124.2	118.1	129.0	117.3	127.4	122.5

Вариант №5. ($\sigma_1 = \sigma_2 = 5$; $n_1 = 20$, $n_2 = 40$).

X

250	251	254	244	250	257	253	247	256	254
244	248	255	251	253	246	247	243	254	241

Y

247	256	253	254	252	242	245	248	258	249
248	252	246	252	257	250	257	254	247	243
243	242	257	248	245	248	247	261	242	247
249	251	241	255	253	257	257	253	247	253

Вариант №6. ($\sigma_1 = \sigma_2 = 3$; $n_1 = 20$, $n_2 = 40$).

X

118	120	122	118	119	117	124	119	124	121
126	115	115	122	120	122	116	122	122	116

Y

117	121	116	118	119	117	121	118	126	117
122	121	117	124	114	123	122	117	122	118
116	122	120	117	117	118	123	123	114	123
125	120	112	121	120	122	129	115	123	120

Вариант №7. ($\sigma_1 = \sigma_2 = 2$; $n_1 = 20$, $n_2 = 40$).

X

718	720	722	720	718	723	716	723	720	719
719	723	721	718	720	720	718	721	721	720

Y

721	717	719	717	723	718	720	721	720	723
719	722	720	719	720	723	718	719	719	722
723	725	718	717	722	722	717	720	720	719
721	722	722	721	719	722	719	719	718	719

Вариант №8. ($\sigma_1 = \sigma_2 = 3$; $n_1 = 20$, $n_2 = 40$).

X

132	134	135	124	131	129	129	136	131	127
133	130	130	133	135	133	128	135	133	130

Y

129	128	132	137	133	130	135	129	131	133
135	130	131	128	132	131	134	134	132	132
135	133	132	133	130	132	134	130	135	132
134	131	137	132	130	132	132	130	128	127

Вариант №9. ($\sigma_1 = \sigma_2 = 4$; $n_1 = 20$, $n_2 = 40$).

X

623	626	624	624	623	623	632	629	624	626
624	625	624	627	618	627	629	620	623	628

Y

625	638	627	614	626	632	622	630	625	627
628	624	622	624	633	623	618	622	623	630
632	622	628	621	626	622	623	630	616	626
623	622	626	627	626	627	625	626	633	623

Вариант №10. ($\sigma_1 = \sigma_2 = 2$; $n_1 = 20$, $n_2 = 40$).

X

14.0	19.8	22.2	22.2	20.7	22.9	19.8	20.1	18.7	22.3
19.0	17.0	18.4	20.1	18.1	22.0	19.0	18.2	22.2	23.6

Y

20.4	18.0	19.3	20.6	20.1	19.8	24.1	18.0	18.5	20.7
18.4	20.0	19.3	19.9	22.2	19.1	17.0	19.5	22.9	22.4
17.8	18.8	23.2	21.5	20.9	17.8	21.9	23.7	18.0	22.8
19.5	18.7	20.3	17.5	24.2	20.8	18.8	18.3	19.4	18.5

Вариант №11. ($\sigma_1 = \sigma_2 = 1$; $n_1 = 20$, $n_2 = 40$).

X

65.0	64.4	65.3	65.7	65.3	64.8	64.8	65.8	64.6	66.0
63.5	65.4	63.6	64.4	63.7	63.5	65.2	66.2	67.0	63.4

Y

63.5	63.7	64.2	67.5	67.7	67.0	65.7	64.5	64.9	65.0
63.4	63.7	65.7	64.1	65.8	63.1	64.9	66.2	64.4	65.5
66.7	63.1	65.9	65.3	65.7	65.5	66.3	65.0	64.7	63.5
65.3	66.5	63.1	65.2	64.1	64.9	66.0	65.0	64.5	65.5

Вариант №12. ($\sigma_1 = \sigma_2 = 5$; $n_1 = 20$, $n_2 = 40$).

X

601	604	591	601	597	594	601	600	596	608
595	609	600	597	601	603	599	600	592	599

Y

591	596	603	593	595	596	599	593	605	599
603	605	603	597	604	600	602	604	599	604
599	593	595	601	598	598	605	601	607	596
602	607	600	605	600	599	592	598	605	594

Вариант №13. ($\sigma_1 = \sigma_2 = 1.5$; $n_1 = 20$, $n_2 = 40$).

X

30.74	29.82	29.77	30.51	31.29	28.79	30.45	30.55	31.72	30.65
28.50	31.72	29.68	29.66	29.99	33.00	30.91	29.49	28.89	30.70

Y

33.17	26.95	31.90	29.99	29.50	27.09	28.96	29.46	29.30	29.19
30.95	28.98	28.36	29.63	34.36	28.72	28.60	31.61	27.06	32.12
30.66	29.84	28.83	27.88	31.38	30.71	30.30	28.20	28.74	27.95
32.06	31.26	31.04	28.48	29.88	32.42	27.87	29.85	29.30	31.78

Вариант №14. ($\sigma_1 = \sigma_2 = 3$; $n_1 = 20$, $n_2 = 40$).

X

155	148	149	156	156	154	159	157	151	154
156	156	149	159	158	157	153	156	158	152

Y

155	158	154	153	160	162	154	155	154	154
161	155	159	157	152	151	155	158	158	156
162	152	149	154	160	162	156	155	156	153
152	154	155	151	154	159	156	157	155	149

Вариант №15. ($\sigma_1 = \sigma_2 = 2$; $n_1 = 20$, $n_2 = 40$).

X

320	321	321	319	319	322	321	326	321	323
319	323	321	318	319	319	320	322	319	317

Y

316	321	321	323	320	321	322	321	320	322
319	321	320	322	322	320	318	320	320	318
321	322	320	324	317	320	315	323	322	319
320	322	320	325	319	322	325	321	322	322

Задания

к лабораторной работе №4.

В работе приведены результаты наблюдений за парой признаков (X,Y). По данным наблюдений требуется:

- построить корреляционное поле;
- найти коэффициент корреляции между признаками X и Y;
- найти уравнение линейной регрессии $y = f(x)$ и дать пояснения полученного результата, используя экономический смысл данных;
- нанести график прямой регрессии на корреляционное поле и сделать предварительный прогноз.

Варианты

к лабораторной работе №4.

Вариант №1.

(X - основные фонды (усл. ед.); Y - объем производства (усл. ед.)).

X	51.55	12.87	41.03	17.54	37.89	19.71	36.05	24.87	29.37
Y	326	184	311	214	297	228	285	241	268
X	9.18	45.59	16.27	39.60	18.93	36.45	24.62	30.92	25.42
Y	150	315	201	301	222	291	233	281	246
X	48.58	13.67	39.97	18.28	37.61	23.33	32.58	25.40	28.81
Y	324	187	306	215	296	229	283	244	266
X	11.72	42.33	16.93	39.05	19.56	36.10	24.66	30.46	26.03
Y	157	314	212	298	226	287	241	277	247
X	46.67	16.24	39.63	18.66	37.59	23.76	32.42	25.40	28.33
Y	322	195	303	221	293	232	283	246	259

Вариант №2 .

(X - средний возраст работников (лет); Y - процент выполнения нормы (%)).

X	44	30	41	31	40	32	38	34	37
Y	107	113	109	112	110	112	110	111	110
X	23	42	31	40	32	39	34	38	35
Y	114	108	112	109	112	110	112	110	111
X	43	30	41	32	39	34	38	34	37
Y	108	113	109	112	110	112	110	111	110
X	29	42	31	40	32	38	34	37	35
Y	114	109	112	109	112	110	112	110	111
X	43	30	41	32	39	34	38	35	37
Y	108	113	109	112	110	112	110	111	110

Вариант №3.

(X - производительность труда (усл. ед.); Y-себестоимость продукции (усл. ед.)).

X	28	31	29	31	29	31	29	30	30
Y	9	45	14	42	16	39	23	33	26
X	33	29	31	29	31	29	30	30	30
Y	48	12	43	16	42	21	34	25	30
X	28	31	29	31	29	31	29	30	30
Y	11	44	14	42	17	38	25	33	27
X	31	29	31	29	31	29	30	30	30
Y	45	13	43	16	41	23	34	26	30
X	28	31	29	31	29	31	29	30	30

у	11	44	15	42	21	35	25	32	27
----------	----	----	----	----	----	----	----	----	----

Вариант №4.

(X-количество азотных удобрений (усл.ед.); У-урожайность зерновых (усл. ед.)).

X	1.8	1.3	1.6	1.4	1.6	1.4	1.5	1.5	1.5
у	90	25	72	41	64	46	60	50	55
X	1.3	1.6	1.4	1.6	1.4	1.6	1.5	1.5	1.5
у	7	74	35	70	43	62	49	56	52
X	1.7	1.4	1.6	1.4	1.6	1.4	1.5	1.5	1.5
у	80	26	71	41	62	49	58	51	55
X	1.3	1.6	1.4	1.6	1.4	1.6	1.5	1.5	1.5
у	25	73	41	64	45	61	49	56	52
X	1.6	1.4	1.6	1.4	1.6	1.5	1.5	1.5	1.5
у	75	35	71	42	62	49	57	51	55

Вариант №5.

(X - накладные расходы (млн. руб.); У - объем выполненных работ (млн. руб.)).

X	45.2	52.8	47.0	51.7	48.2	50.8	48.6	50.1	49.1
у	450	541	471	522	479	518	484	512	490
X	53.3	46.8	51.9	47.9	51.2	48.4	50.4	49.0	49.8
у	548	466	529	476	520	480	513	488	510
X	45.8	52.0	47.6	51.4	48.3	50.6	48.9	50.0	49.2
у	451	538	472	522	480	517	485	512	494
X	53.2	47.0	51.8	48.0	51.0	48.4	50.4	49.0	49.8
у	546	468	529	478	520	482	512	490	507
X	46.5	51.9	47.8	51.3	48.4	50.6	49.0	50.0	49.3
у	453	531	474	521	480	513	487	510	494

Вариант №6.

(X - объем сбыта товаров (млн. руб.); У - относительный уровень издержек обращения (%)).

X	2554.8	2441.8	2540.1	2467.1	2533.8	2480.5	2521.4	2485.5	2508.8
у	3.7	5.5	4.5	5.4	4.6	5.2	4.8	5.1	4.9
X	2425.9	2548.4	2461.1	2536.4	2475.9	2523.3	2482.9	2510.5	2490.3
у	6.3	4.4	5.4	4.5	5.2	4.7	5.1	4.9	5.1
X	2551.4	2454.5	2537.6	2468.0	2532.5	2480.6	2515.7	2486.5	2502.1
у	4.2	5.5	4.5	5.3	4.7	5.2	4.8	5.1	4.9
X	2434.5	2547.5	2463.2	2535.7	2477.0	2522.6	2484.3	2509.1	2490.4
у	6.1	4.4	5.4	4.6	5.2	4.7	5.1	4.9	5.0
X	2551.0	2458.8	2537.1	2471.8	2529.7	2481.5	2510.8	2487.0	2498.7
у	4.2	5.4	4.5	5.3	4.7	5.1	4.9	5.1	4.9

Вариант №7.

(X - энерговооруженность (усл. ед.); Y - себестоимость продукции (усл. ед.)).

X	74	60	69	61	68	62	67	64	66
Y	583	697	619	680	630	674	635	666	643
X	57	70	61	68	62	67	63	66	64
Y	707	608	685	626	678	633	670	641	656
X	73	61	69	62	68	62	67	64	66
Y	587	690	623	679	630	671	635	660	643
X	60	70	61	68	62	67	63	66	65
Y	698	618	681	629	675	634	670	642	655
X	70	61	68	62	67	63	67	64	65
Y	604	687	625	679	631	670	639	657	647

Вариант №8.

(X - оснащенность основными фондами (усл. ед.); Y - себестоимость продукции (усл. ед.)).

X	31	25	30	26	29	27	29	27	28	28
Y	17	19	17	18	18	18	18	18	18	18
X	25	30	26	29	27	29	27	28	27	28
Y	19	17	19	18	18	18	18	18	18	18
X	31	25	30	26	29	27	29	27	28	28
Y	17	19	17	18	18	18	18	18	18	18
X	25	30	26	29	27	29	27	28	28	28
Y	19	17	18	18	18	18	18	18	18	18
X	31	26	29	27	29	27	29	27	28	28
Y	17	19	17	18	18	18	18	18	18	18

Вариант №9.

(X - количество осадков (мм); Y - урожайность зерновых (ц/га)).

X	21	15	20	15	19	15	19	16	18	17
Y	39	23	35	26	35	27	34	29	32	31
X	14	20	15	20	15	19	16	18	16	17
Y	21	36	24	35	27	34	27	33	30	32
X	21	15	20	15	19	16	19	16	17	17
Y	37	23	35	26	35	27	33	29	32	31
X	15	20	15	19	15	19	16	18	17	17
Y	23	36	25	35	27	34	29	32	30	32
X	20	15	20	15	19	16	18	16	17	17
Y	37	24	35	27	34	27	33	30	32	32

Вариант №10.

(X - энерговооруженность (усл. ед.); Y - производительность труда (усл. ед.)).

X	248	58	186	77	171	88	151	113	138	125
Y	30	10	26	14	25	16	22	18	21	20
X	40	194	68	175	85	154	106	148	121	133
Y	6	28	13	25	15	23	18	22	19	21
X	237	59	180	82	158	91	151	114	136	126
Y	30	11	26	15	25	17	22	18	21	20
X	41	188	70	175	85	152	110	144	122	128
Y	9	28	13	25	16	22	18	22	19	20
X	205	62	178	84	157	94	151	115	135	128
Y	29	11	26	15	24	17	22	19	21	21

Вариант № 11.

(X - средняя выработка продукции в единицу рабочего времени (изд./ч.);

Y - коэф. механизации работ (%)).

X	153	127	142	130	139	132	137	133	136	135
Y	74	45	62	48	61	51	60	52	57	55
X	125	143	128	140	131	138	132	136	134	136
Y	41	67	47	61	51	60	52	58	54	56
X	148	127	141	130	139	132	137	134	136	135
Y	70	45	62	49	61	52	58	53	57	55
X	126	142	130	140	132	137	133	136	135	135
Y	44	63	47	61	51	60	52	58	54	55
X	144	128	141	131	138	132	136	134	136	136
Y	67	46	62	49	60	52	58	53	57	55

Вариант №12.

(X - толщина хлопковой нити (мкр.); Y - ее крепость (Кг)).

X	5 132	4 259	4 952	4 406	4 877	4 523	4 755	4 615	4 698	4 638
Y	6.8	4.1	5.9	4.5	5.4	4.6	5.3	4.7	5.1	4.9
X	4 170	5 043	4 365	4 905	4 518	4 831	4 576	4 717	4 625	4 677
Y	3.3	6.0	4.5	5.4	4.5	5.4	4.7	5.1	4.9	5.0
X	5 088	4 285	4 909	4 438	4 843	4 526	4 739	4 617	4 693	4 647
Y	6.1	4.2	5.5	4.5	5.4	4.6	5.2	4.8	5.0	4.9
X	4 195	5 020	4 386	4 892	4 521	4 796	4 613	4 709	4 635	4 651
Y	3.9	6.0	4.5	5.4	4.6	5.3	4.7	5.1	4.9	4.9
X	5 044	4 351	4 907	4 479	4 841	4 527	4 720	4 618	4 680	4 653
Y	6.0	4.4	5.5	4.5	5.4	4.7	5.2	4.9	5.0	4.9

Вариант №13.

(X - среднесписочный состав работников (чел.); Y - розничный товароборот (млн. крб.)).

X	172	99	148	104	144	110	134	116	129	125
Y	35.2	14.1	26.4	16.3	25.1	17.8	23.2	19.1	21.5	19.7
X	91	154	104	144	107	137	113	130	124	126
Y	12.5	28.1	15.6	25.7	17.4	23.5	18.2	21.6	19.5	20.4
X	159	101	146	105	142	112	133	121	129	125
Y	28.5	14.6	26.2	16.5	24.8	18.0	21.7	19.2	21.1	20.0
X	96	150	104	144	107	136	115	130	124	125
Y	13.5	26.7	16.2	25.5	17.6	23.4	18.7	21.6	19.5	20.0
X	159	102	144	105	142	112	132	123	129	126
Y	28.2	15.5	25.8	16.8	24.5	18.2	21.6	19.4	20.5	20.2

Вариант №14.

(X - качество кормов (усл. ед.); Y - удой молока одной коровы (Кг)).

X	26	21	25	21	24	22	24	23	23	23
Y	3 645	3 025	3 497	3 077	3 450	3 123	3 362	3 193	3 297	3 206
X	21	25	21	24	22	24	22	24	23	23
Y	2 780	3 551	3 066	3 465	3 112	3 405	3 187	3 316	3 197	3 264
X	25	21	25	22	24	22	24	23	23	23
Y	3 638	3 034	3 480	3 081	3 447	3 175	3 354	3 193	3 293	3 207
X	21	25	21	24	22	24	22	23	23	23
Y	2 914	3 524	3 068	3 460	3 121	3 403	3 189	3 307	3 203	3 228
X	25	21	25	22	24	22	24	23	23	23
Y	3 553	3 039	3 477	3 083	3 420	3 178	3 341	3 197	3 271	3 242

Вариант №15.

(X - накладные расходы (млн. крб.); Y - себестоимость объекта (млн. крб.)).

X	387	61	334	131	292	176	276	205	255	227
Y	3 645	3 025	3 497	3 077	3 450	3 123	3 362	3 193	3 297	3 206
X	24	337	116	306	166	280	193	266	220	246
Y	2 780	3 551	3 066	3 465	3 112	3 405	3 187	3 316	3 197	3 264
X	377	85	328	133	291	177	268	205	249	231
Y	3 638	3 034	3 480	3 081	3 447	3 175	3 354	3 193	3 293	3 207
X	53	335	123	304	166	277	203	260	224	234
Y	2 914	3 524	3 068	3 460	3 121	3 403	3 189	3 307	3 203	3 228
X	370	95	324	143	281	191	267	218	248	235
Y	3 553	3 039	3 477	3 083	3 420	3 178	3 341	3 197	3 271	3 242

ПРИЛОЖЕНИЯ

1. Нормальное распределение

Плотность вероятности нормированного нормального распределения

$$\varphi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

x	0	1	2	3	4	5	6	7	8	9
0	0,39894	0,39892	0,39886	0,39876	0,39862	0,39844	0,39822	0,39797	0,39767	0,39733
0,1	0,39695	0,39654	0,39608	0,39559	0,39505	0,39448	0,39387	0,39322	0,39253	0,39181
0,2	0,39104	0,39024	0,3894	0,38853	0,38762	0,38667	0,38568	0,38466	0,38361	0,38251
0,3	0,38139	0,38023	0,37903	0,3778	0,37654	0,37524	0,37391	0,37255	0,37115	0,36973
0,4	0,36827	0,36678	0,36526	0,36371	0,36213	0,36053	0,35889	0,35723	0,35553	0,35381
0,5	0,35207	0,35029	0,34849	0,34667	0,34482	0,34294	0,34105	0,33912	0,33718	0,33521
0,6	0,33322	0,33121	0,32918	0,32713	0,32506	0,32297	0,32086	0,31874	0,31659	0,31443
0,7	0,31225	0,31006	0,30785	0,30563	0,30339	0,30114	0,29887	0,29659	0,29431	0,292
0,8	0,28969	0,28737	0,28504	0,28269	0,28034	0,27798	0,27562	0,27324	0,27086	0,26848
0,9	0,26609	0,26369	0,26129	0,25888	0,25647	0,25406	0,25164	0,24923	0,24681	0,24439
1	0,24197	0,23955	0,23713	0,23471	0,2323	0,22988	0,22747	0,22506	0,22265	0,22025
1,1	0,21785	0,21546	0,21307	0,21069	0,20831	0,20594	0,20357	0,20121	0,19886	0,19652
1,2	0,19419	0,19186	0,18954	0,18724	0,18494	0,18265	0,18037	0,1781	0,17585	0,1736
1,3	0,17137	0,16915	0,16694	0,16474	0,16256	0,16038	0,15822	0,15608	0,15395	0,15183
1,4	0,14973	0,14764	0,14556	0,1435	0,14146	0,13943	0,13742	0,13542	0,13344	0,13147
1,5	0,12952	0,12758	0,12566	0,12376	0,12188	0,12001	0,11816	0,11632	0,1145	0,1127
1,6	0,11092	0,10915	0,10741	0,10567	0,10396	0,10226	0,10059	0,09893	0,09728	0,09566
1,7	0,09405	0,09246	0,09089	0,08933	0,0878	0,08628	0,08478	0,08329	0,08183	0,08038
1,8	0,07895	0,07754	0,07614	0,07477	0,07341	0,07206	0,07074	0,06943	0,06814	0,06687
1,9	0,06562	0,06438	0,06316	0,06195	0,06077	0,05959	0,05844	0,0573	0,05618	0,05508
2	0,05399	0,05292	0,05186	0,05082	0,0498	0,04879	0,0478	0,04682	0,04586	0,04491
2,1	0,04398	0,04307	0,04217	0,04128	0,04041	0,03955	0,03871	0,03788	0,03706	0,03626
2,2	0,03547	0,0347	0,03394	0,03319	0,03246	0,03174	0,03103	0,03034	0,02965	0,02898
2,3	0,02833	0,02768	0,02705	0,02643	0,02582	0,02522	0,02463	0,02406	0,02349	0,02294
2,4	0,02239	0,02186	0,02134	0,02083	0,02033	0,01984	0,01936	0,01888	0,01842	0,01797
2,5	0,01753	0,01709	0,01667	0,01625	0,01585	0,01545	0,01506	0,01468	0,01431	0,01394
2,6	0,01358	0,01323	0,01289	0,01256	0,01223	0,01191	0,0116	0,0113	0,011	0,01071
2,7	0,01042	0,01014	0,00987	0,00961	0,00935	0,00909	0,00885	0,00861	0,00837	0,00814
2,8	0,00792	0,0077	0,00748	0,00727	0,00707	0,00687	0,00668	0,00649	0,00631	0,00613
2,9	0,00595	0,00578	0,00562	0,00545	0,0053	0,00514	0,00499	0,00485	0,0047	0,00457
3	0,00443	0,0043	0,00417	0,00405	0,00393	0,00381	0,0037	0,00358	0,00348	0,00337
3,1	0,00327	0,00317	0,00307	0,00298	0,00288	0,00279	0,00271	0,00262	0,00254	0,00246
3,2	0,00238	0,00231	0,00224	0,00216	0,0021	0,00203	0,00196	0,0019	0,00184	0,00178
3,3	0,00172	0,00167	0,00161	0,00156	0,00151	0,00146	0,00141	0,00136	0,00132	0,00127
3,4	0,00123	0,00119	0,00115	0,00111	0,00107	0,00104	0,001	0,00097	0,00094	0,0009
3,5	0,00087	0,00084	0,00081	0,00079	0,00076	0,00073	0,00071	0,00068	0,00066	0,00063
3,6	0,00061	0,00059	0,00057	0,00055	0,00053	0,00051	0,00049	0,00047	0,00046	0,00044
3,7	0,00042	0,00041	0,00039	0,00038	0,00037	0,00035	0,00034	0,00033	0,00031	0,0003
3,8	0,00029	0,00028	0,00027	0,00026	0,00025	0,00024	0,00023	0,00022	0,00021	0,00021
3,9	0,0002	0,00019	0,00018	0,00018	0,00017	0,00016	0,00016	0,00015	0,00014	0,00014

2. Нормальное распределение

Функция нормированного нормального распределения

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{t^2}{2}} dt$$

x	0	1	2	3	4	5	6	7	8	9
0	0,5	0,50399	0,50798	0,51197	0,51595	0,5199	0,52392	0,5279	0,5319	0,53586
0,1	0,53983	0,5438	0,54776	0,55172	0,55567	0,5596	0,56356	0,56749	0,5714	0,57535
0,2	0,57926	0,58317	0,58706	0,59095	0,59483	0,5987	0,60257	0,60642	0,6103	0,61409
0,3	0,61791	0,62172	0,62552	0,6293	0,63307	0,6368	0,64058	0,64431	0,648	0,65173
0,4	0,65542	0,6591	0,66276	0,6664	0,67003	0,6736	0,67724	0,68082	0,6844	0,68793
0,5	0,69146	0,69497	0,69847	0,70194	0,7054	0,7088	0,71226	0,71566	0,719	0,7224
0,6	0,72575	0,72907	0,73237	0,73565	0,73891	0,7422	0,74537	0,74857	0,7517	0,7549
0,7	0,75804	0,76115	0,76424	0,7673	0,77035	0,7734	0,77637	0,77935	0,7823	0,78524
0,8	0,78814	0,79103	0,79389	0,79673	0,79955	0,8023	0,80511	0,80785	0,8106	0,81327
0,9	0,81594	0,81859	0,82121	0,82381	0,82639	0,8289	0,83147	0,83398	0,8365	0,83891
1	0,84134	0,84375	0,84614	0,84849	0,85083	0,8531	0,85543	0,85769	0,8599	0,86214
1,1	0,86433	0,8665	0,86864	0,87076	0,87286	0,8749	0,87698	0,879	0,881	0,88298
1,2	0,88493	0,88686	0,88877	0,89065	0,89251	0,8944	0,89617	0,89796	0,8997	0,90147
1,3	0,9032	0,9049	0,90658	0,90824	0,90988	0,9115	0,91308	0,91466	0,9162	0,91774
1,4	0,91924	0,92073	0,9222	0,92364	0,92507	0,9265	0,92785	0,92922	0,9306	0,93189
1,5	0,93319	0,93448	0,93574	0,93699	0,93822	0,9394	0,94062	0,94179	0,9429	0,94408
1,6	0,9452	0,9463	0,94738	0,94845	0,9495	0,9505	0,95154	0,95254	0,9535	0,95449
1,7	0,95543	0,95637	0,95728	0,95818	0,95907	0,9599	0,9608	0,96164	0,9625	0,96327
1,8	0,96407	0,96485	0,96562	0,96638	0,96712	0,9678	0,96856	0,96926	0,9699	0,97062
1,9	0,97128	0,97193	0,97257	0,9732	0,97381	0,9744	0,975	0,97558	0,9761	0,9767
2	0,97725	0,97778	0,97831	0,97882	0,97932	0,9798	0,9803	0,98077	0,9812	0,98169
2,1	0,98214	0,98257	0,983	0,98341	0,98382	0,9842	0,98461	0,985	0,9854	0,98574
2,2	0,9861	0,98645	0,98679	0,98713	0,98745	0,9878	0,98809	0,9884	0,9887	0,98899
2,3	0,98928	0,98956	0,98983	0,9901	0,99036	0,9906	0,99086	0,99111	0,9913	0,99158
2,4	0,9918	0,99202	0,99224	0,99245	0,99266	0,9929	0,99305	0,99324	0,9934	0,99361
2,5	0,99379	0,99396	0,99413	0,9943	0,99446	0,9946	0,99477	0,99492	0,9951	0,9952
2,6	0,99534	0,99547	0,9956	0,99573	0,99585	0,996	0,99609	0,99621	0,9963	0,99643
2,7	0,99653	0,99664	0,99674	0,99683	0,99693	0,997	0,99711	0,9972	0,9973	0,99736
2,8	0,99744	0,99752	0,9976	0,99767	0,99774	0,9978	0,99788	0,99795	0,998	0,99807
2,9	0,99813	0,99819	0,99825	0,99831	0,99836	0,9984	0,99846	0,99851	0,9986	0,99861
3	0,99865	0,99869	0,99874	0,99878	0,99882	0,9989	0,99889	0,99893	0,999	0,999
3,1	0,99903	0,99906	0,9991	0,99913	0,99916	0,9992	0,99921	0,99924	0,9993	0,99929
3,2	0,99931	0,99934	0,99936	0,99938	0,9994	0,9994	0,99944	0,99946	0,9995	0,9995
3,3	0,99952	0,99953	0,99955	0,99957	0,99958	0,9996	0,99961	0,99962	0,9996	0,99965
3,4	0,99966	0,99968	0,99969	0,9997	0,99971	0,9997	0,99973	0,99974	0,9997	0,99976
3,5	0,99977	0,99978	0,99978	0,99979	0,9998	0,9998	0,99981	0,99982	0,9998	0,99983
3,6	0,99984	0,99985	0,99985	0,99986	0,99986	0,9999	0,99987	0,99988	0,9999	0,99989
3,7	0,99989	0,9999	0,9999	0,9999	0,99991	0,9999	0,99992	0,99992	0,9999	0,99992
3,8	0,99993	0,99993	0,99993	0,99994	0,99994	0,9999	0,99994	0,99995	0,9999	0,99995
3,9	0,99995	0,99995	0,99996	0,99996	0,99996	1	0,99996	0,99996	1	0,99997
4	0,99997	0,99997	0,99997	0,99997	0,99997	1	0,99998	0,99998	1	0,99998
4,1	0,99998	0,99998	0,99998	0,99998	0,99998	1	0,99998	0,99998	1	0,99999
4,2	0,99999	0,99999	0,99999	0,99999	0,99999	1	0,99999	0,99999	1	0,99999
4,3	0,99999	0,99999	0,99999	0,99999	0,99999	1	0,99999	0,99999	1	0,99999
4,4	0,99999	0,99999	1							

3. χ^2 -распределение

Значения $\chi^2_{k\alpha}$, соответствующие вероятности $p = P\{\chi^2_k > \chi^2_{k,\alpha}\}$,

где χ^2_k имеет χ^2 -распределение с k степенями свободы

k	α										
	0,99	0,95	0,9	0,5	0,25	0,1	0,05	0,025	0,01	0,005	0,001
1	0	0	0,02	0,45	1,32	2,71	3,84	5,02	6,63	7,88	10,8
2	0,02	0,1	0,21	1,39	2,77	4,61	5,99	7,38	9,21	10,6	13,8
3	0,11	0,35	0,58	2,37	4,11	6,25	7,81	9,35	11,3	12,8	16,3
4	0,3	0,71	1,06	3,36	5,39	7,78	9,49	11,1	13,3	14,9	18,5
5	0,55	1,15	1,61	4,35	6,63	9,24	11,1	12,8	15,1	16,7	20,5
6	0,87	1,64	2,2	5,35	7,84	10,6	12,6	14,4	16,8	18,5	22,5
7	1,24	2,17	2,83	6,35	9,04	12	14,1	16	18,5	20,3	24,3
8	1,65	2,73	3,49	7,34	10,2	13,4	15,5	17,5	20,1	22	26,1
9	2,09	3,33	4,17	8,34	11,4	14,7	16,9	19	21,7	23,6	27,9
10	2,56	3,94	4,87	9,34	12,5	16	18,3	20,5	23,2	25,2	29,6
11	3,05	4,57	5,58	10,3	13,7	17,3	19,7	21,9	24,7	26,8	31,3
12	3,57	5,23	6,3	11,3	14,8	18,5	21	23,3	26,2	28,3	32,9
13	4,11	5,89	7,04	12,3	16	19,8	22,4	24,7	27,7	29,8	34,5
14	4,66	6,57	7,79	13,3	17,1	21,1	23,7	26,1	29,1	31,3	36,1
15	5,23	7,26	8,55	14,3	18,2	22,3	25	27,5	30,6	32,8	37,7
16	5,81	7,96	9,31	15,3	19,4	23,5	26,3	28,8	32	34,3	39,3
17	6,41	8,67	10,1	16,3	20,5	24,8	27,6	30,2	33,4	35,7	40,8
18	7,01	9,39	10,9	17,3	21,6	26	28,9	31,5	34,8	37,2	42,3
19	7,63	10,1	11,7	18,3	22,7	27,2	30,1	32,9	36,2	38,6	43,8
20	8,26	10,9	12,4	19,3	23,8	28,4	31,4	34,2	37,6	40	45,3
21	8,9	11,6	13,2	20,3	24,9	29,6	32,7	35,5	38,9	41,4	46,8
22	9,54	12,3	14	21,3	26	30,8	33,9	36,8	40,3	42,8	48,3
23	10,2	13,1	14,8	22,3	27,1	32	35,2	38,1	41,6	44,2	49,7
24	10,9	13,8	15,7	23,3	28,2	33,2	36,4	39,4	43	45,6	51,2
25	11,5	14,6	16,5	24,3	29,3	34,4	37,7	40,6	44,3	46,9	52,6
26	12,2	15,4	17,3	25,3	30,4	35,6	38,9	41,9	45,6	48,3	54,1
27	12,9	16,2	18,1	26,3	31,5	36,7	40,1	43,2	47	49,6	55,5
28	13,6	16,9	18,9	27,3	32,6	37,9	41,3	44,5	48,3	51	56,9
29	14,3	17,7	19,8	28,3	33,7	39,1	42,6	45,7	49,6	52,3	58,3
30	15	18,5	20,6	29,3	34,8	40,3	43,8	47	50,9	53,7	59,7

4. Распределение Стьюдента

Значения $t_{k\beta}$, соответствующие вероятности $\beta = P\{|t_k| > t_{k,\beta}\}$,

где случайная величина t_k имеет распределение Стьюдента с k степенями свободы

k	β							
	0,2	0,1	0,05	0,02	0,01	0,005	0,002	0,001
1	3,078	6,314	12,71	31,82	63,66	127,3	318,3	636,6
2	1,886	2,92	4,303	6,965	9,925	14,09	22,33	31,6
3	1,638	2,353	3,182	4,541	5,841	7,453	10,21	12,92
4	1,533	2,132	2,776	3,747	4,604	5,598	7,173	8,61
5	1,476	2,015	2,571	3,365	4,032	4,773	5,894	6,869
6	1,44	1,943	2,447	3,143	3,707	4,317	5,208	5,959
7	1,415	1,895	2,365	2,998	3,499	4,029	4,785	5,408
8	1,397	1,86	2,306	2,896	3,355	3,833	4,501	5,041
9	1,383	1,833	2,262	2,821	3,25	3,69	4,297	4,781
10	1,372	1,812	2,228	2,764	3,169	3,581	4,144	4,587
11	1,363	1,796	2,201	2,718	3,106	3,497	4,025	4,437
12	1,356	1,782	2,179	2,681	3,055	3,428	3,93	4,318
13	1,35	1,771	2,16	2,65	3,012	3,372	3,852	4,221
14	1,345	1,761	2,145	2,624	2,977	3,326	3,787	4,14
15	1,341	1,753	2,131	2,602	2,947	3,286	3,733	4,073
16	1,337	1,746	2,12	2,583	2,921	3,252	3,686	4,015
17	1,333	1,74	2,11	2,567	2,898	3,222	3,646	3,965
18	1,33	1,734	2,101	2,552	2,878	3,197	3,61	3,922
19	1,328	1,729	2,093	2,539	2,861	3,174	3,579	3,883
20	1,325	1,725	2,086	2,528	2,845	3,153	3,552	3,85
21	1,323	1,721	2,08	2,518	2,831	3,135	3,527	3,819
22	1,321	1,717	2,074	2,508	2,819	3,119	3,505	3,792
23	1,319	1,714	2,069	2,5	2,807	3,104	3,485	3,768
24	1,318	1,711	2,064	2,492	2,797	3,091	3,467	3,745
25	1,316	1,708	2,06	2,485	2,787	3,078	3,45	3,725
26	1,315	1,706	2,056	2,479	2,779	3,067	3,435	3,707
27	1,314	1,703	2,052	2,473	2,771	3,057	3,421	3,689
28	1,313	1,701	2,048	2,467	2,763	3,047	3,408	3,674
29	1,311	1,699	2,045	2,462	2,756	3,038	3,396	3,66
30	1,31	1,697	2,042	2,457	2,75	3,03	3,385	3,646
40	1,303	1,684	2,021	2,423	2,704	2,971	3,307	3,551
60	1,296	1,671	2	2,39	2,66	2,915	3,232	3,46
120	1,289	1,658	1,98	2,358	2,617	2,86	3,16	3,373
10000	1,282	1,645	1,96	2,327	2,576	2,808	3,091	3,291

5. Распределение Пуассона

$$P(k) = \frac{\lambda^k}{k!} \cdot e^{-\lambda}$$

$k \backslash \lambda$	0.1	0.2	0.3	0.4	0.5	0.6	0.7	0.8	0.9	
0	0.9048	0.8187	0.7408	0.6703	0.6065	0.5488	0.4966	0.4493	0.4066	
1	0.0905	0.1637	0.2222	0.2681	0.3033	0.3293	0.3476	0.3595	0.3659	
2	0.0045	0.0164	0.0333	0.0536	0.0758	0.0988	0.1217	0.1438	0.1647	
3	0.0002	0.0011	0.0033	0.0072	0.0126	0.0198	0.0284	0.0383	0.0494	
4	0.0000	0.0001	0.0003	0.0007	0.0016	0.0030	0.0050	0.0077	0.0111	
5	0.0000	0.0000	0.0000	0.0001	0.0002	0.0004	0.0007	0.0012	0.0020	
6	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0002	0.0003	
$k \backslash \lambda$	1	2	3	4	5	6	7	8	9	10
0	0.3679	0.1353	0.0498	0.0183	0.0067	0.0025	0.0009	0.0003	0.0001	0.0000
1	0.3679	0.2707	0.1494	0.0733	0.0337	0.0149	0.0064	0.0027	0.0011	0.0005
2	0.1839	0.2707	0.2240	0.1465	0.0842	0.0446	0.0223	0.0107	0.0050	0.0023
3	0.0613	0.1804	0.2240	0.1954	0.1404	0.0892	0.0521	0.0286	0.0150	0.0076
4	0.0153	0.0902	0.1680	0.1954	0.1755	0.1339	0.0912	0.0573	0.0337	0.0189
5	0.0031	0.0361	0.1008	0.1563	0.1755	0.1606	0.1277	0.0916	0.0607	0.0378
6	0.0005	0.0120	0.0504	0.1042	0.1462	0.1606	0.1490	0.1221	0.0911	0.0631
7	0.0001	0.0034	0.0216	0.0595	0.1044	0.1377	0.1490	0.1396	0.1171	0.0901
8	0.0000	0.0009	0.0081	0.0298	0.0653	0.1033	0.1304	0.1396	0.1318	0.1126
9	0.0000	0.0002	0.0027	0.0132	0.0363	0.0688	0.1014	0.1241	0.1318	0.1251
10	0.0000	0.0000	0.0008	0.0053	0.0181	0.0413	0.0710	0.0993	0.1186	0.1251
11	0.0000	0.0000	0.0002	0.0019	0.0082	0.0225	0.0452	0.0722	0.0970	0.1137
12	0.0000	0.0000	0.0001	0.0006	0.0034	0.0113	0.0263	0.0481	0.0728	0.0948
13	0.0000	0.0000	0.0000	0.0002	0.0013	0.0052	0.0142	0.0296	0.0504	0.0729
14	0.0000	0.0000	0.0000	0.0001	0.0005	0.0022	0.0071	0.0169	0.0324	0.0521
15	0.0000	0.0000	0.0000	0.0000	0.0002	0.0009	0.0033	0.0090	0.0194	0.0347
16	0.0000	0.0000	0.0000	0.0000	0.0000	0.0003	0.0014	0.0045	0.0109	0.0217
17	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0006	0.0021	0.0058	0.0128
18	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0009	0.0029	0.0071
19	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004	0.0014	0.0037
20	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002	0.0006	0.0019
21	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0003	0.0009
22	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001	0.0004
23	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0002
24	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0001

6. Пример вычисления выборочного коэффициента корреляции и коэффициентов функции линейной регрессии по корреляционной таблице.

Рассмотрим пример зависимости Y от X по данным 100 наблюдений (см. табл.1.). Найдем выборочное уравнение прямой линии $\bar{y}_x - \bar{y} = r_b \cdot \frac{\sigma_y}{\sigma_x} \cdot (x - \bar{x})$ регрессии Y на X .

Таблица №1.

X	40	30	35	30	40	35	35	40	35	30
Y	56	36	56	36	46	36	46	36	46	46
X	30	25	30	25	30	25	30	30	30	30
Y	36	16	36	16	36	26	36	26	26	26
X	30	40	30	40	35	35	40	35	40	35
Y	36	56	36	46	36	46	36	46	36	46
X	20	30	25	30	25	30	25	30	30	30
Y	16	36	16	36	26	36	26	36	26	26
X	40	30	40	30	40	40	35	40	35	30
X	30	20	30	25	30	25	30	30	30	30
Y	36	16	36	16	36	26	36	26	36	26
Y	56	36	46	36	46	36	46	36	46	46
X	30	40	30	40	35	35	40	35	40	30
X	20	30	25	30	25	30	25	30	30	30
Y	16	36	16	36	26	36	26	36	26	26
Y	36	56	36	46	36	46	36	46	36	46
X	40	30	40	30	35	40	35	40	35	30
Y	56	36	46	36	46	36	46	36	46	46
X	30	20	30	25	30	25	30	25	30	30
Y	36	16	36	16	36	26	36	26	36	26

Получим здесь формулы пригодные для вычислений с помощью ЭВМ или микрокалькулятора, но так как вычисления “вручную” трудоемки для их упрощения используется корреляционная таблица.

По таблице наблюдений (см. табл. №1) невозможно судить о распределении случайных величин X и Y , а также об их общем распределении для этого составим *корреляционную таблицу*.

Варианты или интервалы значений одной случайной величины (Y) запишем в первый столбец корреляционной таблицы, а варианты или интервалы другой (X) - в первую строку. По каждой паре значений (x_i, y_i) считаем число повторений (частоту) и заносим в клетку соответствующего столбца и строки (n_{xy}). Далее суммируя числа n_{xy} по строкам и столбцам находим частоты n_x и n_y . Объем выборки можно получить с помощью сумм (см. табл.№2).

$$n = \sum_i n_{x_i} = \sum_j n_{y_j} = \sum_i \sum_j n_{x_i y_j}$$

Таблица №2.

Y	X					n _y
	20	25	30	35	40	
16	4	6				10
26		8	10			18
36			32	3	9	44
46			4	12	6	22
56				1	5	6
n _x	4	14	46	16	20	n = 100

Для упрощения “ручного” расчета составим корреляционную табл. №2 в условных вариантах, выбрав в качестве ложных нулей $C_1 = 30$ и $C_2 = 36$ (каждая из этих вариант расположена в середине соответствующего вариационного ряда).

Таблица №3.

v	u					n _v
	-2	-1	0	1	2	
-2	4	6				10
-1		8	10			18
0			32	3	9	44
1			4	12	6	22
2				1	5	6
n _u	4	14	46	16	20	n = 100

Найдем $\bar{u}$ и $\bar{v}$:

$$\bar{u} = \frac{\sum_i n_{u_i} u_i}{n} = \frac{4(-2) + 14(-1) + 46 \cdot 0 + 16 \cdot 1 + 20 \cdot 2}{100} = 0.34;$$

$$\bar{v} = \frac{\sum_j n_{v_j} v_j}{n} = \frac{10(-2) + 18(-1) + 44 \cdot 0 + 22 \cdot 1 + 6 \cdot 2}{100} = -0.04.$$

Найдем вспомогательные величины $\overline{u^2}$ и $\overline{v^2}$:

$$\overline{u^2} = \frac{\sum_i n_{u_i} u_i^2}{n} = \frac{4(-2)^2 + 14(-1)^2 + 46 \cdot 0 + 16 \cdot 1 + 20 \cdot 2^2}{100} = 1.26;$$

$$\overline{v^2} = \frac{\sum_j n_{v_j} v_j^2}{n} = \frac{10 \cdot 4 + 18 \cdot 1 + 44 \cdot 0 + 22 \cdot 1 + 6 \cdot 4}{100} = 1.04.$$

Найдем σ_u и σ_v :

$$\sigma_u = \sqrt{\overline{u^2} - (\bar{u})^2} = \sqrt{1.26 - (0.34)^2} \approx 1.07;$$

$$\sigma_v = \sqrt{v^2 - (\bar{v})^2} = \sqrt{1.04 - (0.04)^2} \approx 1.02.$$

Найдем теперь $\sum n_{uv} u \cdot v$ методом *четырёх полей*, для чего составим расчетную таблицу №4.

Таблица №4.

u							
v	-2	-1	0	1	2	I	II
-2	4 ⁴	6 ²				28	
-1		8 ¹				8	
0						III	IV
1				12 ¹	6 ²		24
2				1	5 ⁴		22
I	16	20	III			36	0
II			IV	14	32	0	46

Указания к составлению таблицы №4. Произведение вариантов $v_i u_j$ записываем в правом верхнем углу соответствующей клетки, содержащей частоту. Сложив произведения $n_{ij} v_i u_j$ по строке или по столбцу, запишем результат в соответствующую результирующую (итоговую) клетку поля (I, II, III, IV).

Сложив числа итоговых клеток (четыре клетки в нижнем правом углу таблицы, получим $\sum n_{uv} u \cdot v = 36 + 46 = 82$.

Для контроля расчета суммы произведений $n_{ij} v_i u_j$ по строке и по столбцу должны совпадать.

Найдем выборочный коэффициент корреляции:

$$r_b = \frac{\sum n_{uv} uv - n \cdot \bar{u} \cdot \bar{v}}{n \cdot \sigma_u \sigma_v} = \frac{82 - 100 \cdot 0.34 \cdot (-0.04)}{100 \cdot 1.07 \cdot 1.02} = 0.76.$$

Теперь перейдем к старым переменным. Определим шаги h_1 и h_2 (разности между любыми двумя соседними вариантами в табл. №2):

$$h_1 = 25 - 20 = 5; \quad h_2 = 26 - 16 = 10.$$

Найдем $\bar{x}$ и $\bar{y}$, учитывая, что $C_1 = 30$, $C_2 = 36$:

$$\bar{x} = \bar{u} \cdot h_1 + C_1 = 0.34 \cdot 5 + 30 = 31.70;$$

$$\bar{y} = \bar{v} \cdot h_2 + C_2 = (-0.04) \cdot 10 + 36 = 35.60.$$

Найдем σ_x и σ_y :

$$\sigma_x = h_1 \cdot \sigma_u = 5 \cdot 1.07 = 5.35;$$

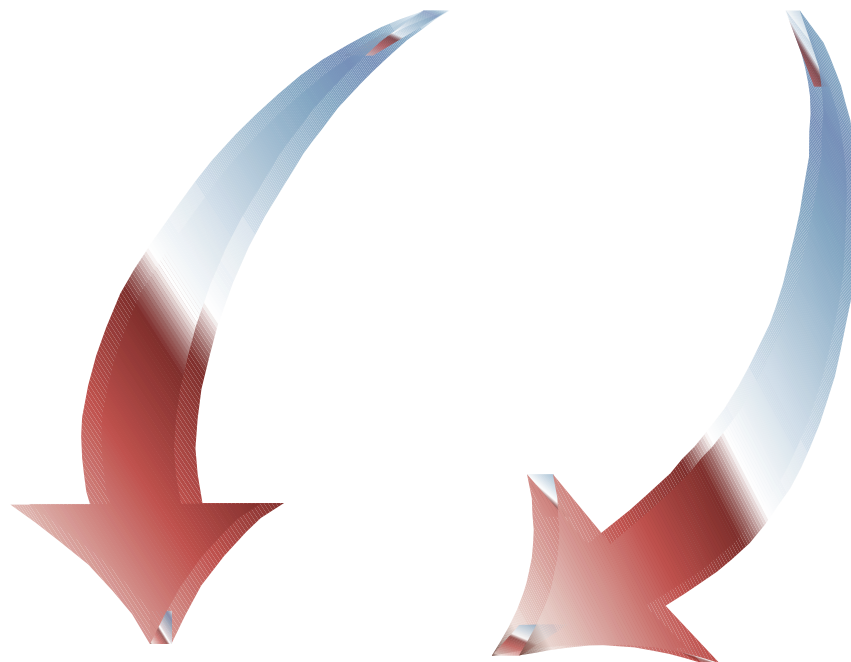
$$\sigma_y = h_2 \cdot \sigma_v = 10 \cdot 1.02 = 10.2.$$

Подставив найденные величины в уравнение линии получим:

$$\bar{y}_x - 35.6 = 0.76 \cdot \frac{10.2}{5.35} \cdot (x - 31.7)$$

или окончательно

$$\bar{y}_x = 1.45x - 10.369).$$



О г л а в л е н и е

В в е д е н и е	3
1. Лабораторная работа №1.....	4
Тема: Первичная обработка статистических данных	
1.1 Пример выполнения лабораторной работы	4
1. Отобразить переменные. 2. Графическое представление исходных данных. 3. Произвести статистическое описание исходных данных, с определением варьирования переменных. 4. Анализ резко выделяющихся наблюдений. 5. Восстановление пропущенных исходных данных. 6. Графики вариационных рядов. 7. Экспериментальный анализ закона распределения исследуемой генеральной совокупности и параметризация сведений о природе изучаемых распределений. 8. Числовые характеристики выборки.	
2. Лабораторная работа №2.....	11
Тема: Статистическое оценивание параметров	
1. Понятие оценки. 2. Основные свойства точечных оценок. 3. Понятие доверительного интервала. Доверительная вероятность. 4. Построение доверительного интервала для математического ожидания при известной дисперсии. 5. Построение доверительного интервала для математического ожидания при неизвестной дисперсии. 6. Определение объема выборки n .	
2.1 Пример выполнения лабораторной работы.....	16
1. Вычислить по выборке несмещенные оценки параметров μ - среднего значения, σ^2 - дисперсии, σ - стандартного отклонения генеральной совокупности $\bar{x}$, S^2 , S ; 2. Найти доверительные интервалы для μ при различной доверительной вероятности $\beta = 0,8; 0,95; 0,99$; 3. Считая данные выборки пробными, определить минимальный объем выборки n для нахождения доверительного интервала среднего значения - μ длины 2Δ и доверительной вероятностью $\beta = 0,8; 0,9; 0,95; 0,99; 0,995$;	

3.	Лабораторная работа №3.....	18
	Тема: Статистические гипотезы.	
	1.Понятие статистической гипотезы. Пример 3.1. (Проверка гипотезы о равенстве центров распределения двух нормальных генеральных совокупностей при известном σ) 2. Критерий согласия χ^2 Пирсона (проверка гипотез о законе распределения).	
3.1	Пример выполнения лабораторной работы.....	24
	Лабораторная работа №4.....	27
	Тема: Корреляционный анализ	
	1. Определение формы связи. Понятие регрессии (линейной).	
4.1	Пример выполнения лабораторной работы.....	31
5.	Задания к лабораторной работе №1.....	35
6.	Варианты к лабораторной работе №1.....	35
7.	Задания к лабораторной работе №2.....	39
8.	Варианты к лабораторной работе №2.....	39
9.	Задания к лабораторной работе №3.....	39
10.	Варианты к лабораторной работе №3.....	40
11.	Задания к лабораторной работе №4.....	48
12.	Варианты к лабораторной работе №4.....	49
13.	П р и л о ж е н и я	54
	1. Плотность вероятности нормированного нормального распределения.....	54
	2. Функция распределения нормированного нормального распределения.....	55
	3. χ^2 - распределение (распределение Пирсона).....	56
	4. t - распределение (распределение Стьюдента).....	57
	5. Распределение Пуассона.....	58
	6. Пример вычисления выборочного коэффициента корреляции и коэффициентов функции линейной регрессии по корреляционной таблице.....	59

Учебное издание
Породников Виктор Дмитриевич

Практикум по математической статистике

Практикум предназначен для выполнения самостоятельных занятий по математической статистике. Он состоит из четырех лабораторных работ и шести приложений. В начале каждой работы дается теоретический материал и формулы для решения задач. Приведены 15 вариантов заданий и решение одной работы. При решении задач предусматривается использование микрокалькулятора и ЭВМ.