Камчатский государственный технический университет

Кафедра высшей математики

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Обработка статистических данных

Корреляционная зависимость

Задания и методические указания
для студентов дневного и заочного отделений

Петропавловск-Камчатский
2004

Батуев Э.Н., Сидорова А.С.

Б 28 **Математическая статистика. Обработка статистических данных. Корреляционная зависимость**. Задания и методические указания для студентов дневного и заочного отделений. – Петропавловск-Камчатский: КамчатГТУ, 2004. – 26 с.

Задания по математической статистики предназначены для студентов дневного и заочного отделений КамчатГТУ. Задания соответствуют базовым стандартам специальностей и программе, утвержденной Министерством образования Российской Федерации.

Рекомендовано к изданию решением ученого совета КамчатГТУ (протокол № 8 от 23,04,2004 г.).

Введение

Построение и анализ выборочного распределения является основным математическим способом исследования реальной случайной величины.

Рассматривается двумерная случайная величина (X, У), имеется выборка объема n (x_i, y_i), i = 1, 2, ..., n.

Работа состоит из двух частей:

1. Анализируем компоненты нашего случайного вектора. Для каждой компоненты строится выборочное распределение, определяются его числовые характеристики, которые служат точечными оценками для числовых характеристик X и У. Проводится интервальное оценивание параметров, оценивается вид закона распределения X и У.
2. Анализируется совместное распределение X и У. Проверяется гипотеза о коррелированности компонент. Строятся уравнения линейной регрессии.

Варианты выдаются преподавателем индивидуально, в противном случае студент выбирает вариант, номер которого совпадает с последней цифрой зачетной книжки.

Часть I

Задание

По результатам наблюдений над случайной величиной X и У требуется для каждой величины:

1) Построить интервальный и дискретный вариационный ряд, полигон или гистограмму в зависимости от того, дискретна или непрерывна изучаемая случайная величина.

Найти эмпирическую функцию распределения и построить ее график.

2) Найти точечные оценки параметров закона распределения случайной величины.

3) С помощью выборочных коэффициентов ассиметрии и эксцесса определить, имеется ли основание для выдвижения гипотезы о нормальности распределения.

Выбрать гипотетический закон распределения. Используя точечные оценки параметров, записать плотность и функцию распределения.

4) В случае нормальности распределения построить доверительные интервалы с надежностью γ.

4.1 для математического ожидания, считая, что $\sigma = \sqrt{S^2}$;

4.2 для математического ожидания, считая дисперсию неизвестной;

4.3 для среднего квадратического отклонения.

5) Проверить с помощью критерия согласия χ^2 гипотезу о виде закона распределения при уровне значимости β.

6) Построить график функции плотности и сравнить его с гистограммой, в случае дискретной случайной величины сравнить многоугольник распределения с полигоном.

Указания

1) Пусть $x_1, x_2, ..., x_n$ – совокупность значений случайной величины X, полученных в результате n независимых повторений некоторого эксперимента. Эта совокупность называется выборкой объема n из генеральной совокупности X.

Элементы выборки, расположенные в порядке возрастания, образуют вариационный ряд. Пусть l – число различных выборочных значений в данной выборке.

Далее через $x_1, x_2, ..., x_n$ обозначаем только различные выборочные значения, m_i – число элементов выборки, равных x_i, $\sum_{i=1}^{l} m_i = n$, $\frac{m_i}{n}$ – относительная частота появления x_i.

Таким образом, мы получаем дискретную случайную величину (выборочное распределение)

X	x_1	x_2	...	x_l
$\frac{m_i}{n}$	$\frac{m_1}{n}$	$\frac{m_2}{n}$	...	$\frac{m_l}{n}$

$$\sum_{i=1}^{l} \frac{m_i}{n} = \frac{1}{n}\sum_{i=1}^{l} m_i = 1$$

В силу закона больших чисел выборочное распределение при $n \to \infty$ сходится по вероятности к распределению генеральной совокупности X.

В декартовой системе координат изобразим точки $\left(x_i, \frac{m_i}{n}\right)$ и соединим их. Получим ломаную – приближенный график многоугольника распределения X. Он называется полигоном относительных частот.

В случае, когда случайная величина X непрерывна, а также если l велико, то весь диапазон выборочных значений разбивают на K интервалов одинаковой длины. Для определения оптимальной длины интервала можно использовать формулы

$$h = \frac{x_{max} - x_{min}}{1 + log_2 n}, \quad или$$

$$h = \frac{x_{max} - x_{min}}{1 + 3,322 \cdot lg\, n}$$

При объеме выборки порядка $60 \le n \le 200$ формулы надежны. Здесь x_{min} и x_{max} – соответственно наименьшее и наибольшее выборочное значение X. За начало первого интервала принимаем величину $a_1 = x_{min} - \frac{h}{2}$. Тогда $a_2 = a_1 + h$, $a_3 = a_2 + h$, ... Далее подсчитываем число n_i выборочных значений, попавших в каждый интервал $(a_i, a_{i+1}]$, i = 1, 2, ..., k. Получаем интервальный ряд (выборочное распределение)

$(a_i, a_{i+1}]$	$(a_1, a_2]$	$(a_2, a_3]$	...	$(a_k, a_{k+1}]$
$\frac{n_i}{n}$	$\frac{n_1}{n}$	$\frac{n_2}{n}$	...	$\frac{n_k}{n}$

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

Графически интервальный ряд – это ступенчатая фигура, состоящая из прямоугольников, основание которых – интервалы длины h, а площадь ступеньки – относительная частота попадания X в данный интервал. Соответственно, высоты определяют как $\frac{n_i}{n} : h$.

При увеличении объема выборки n контур гистограммы приближается к графику функции плотности X (как и в случае полигона). Таким образом, гистограмма – приближенный график плотности X.

Естественным образом определяется эмпирическая функция распределения:

$$F^*(x) = \frac{n_x}{n},$$

где n_x – число выборочных значений X, меньших x. Понятно, что при $n \to \infty$ $F^*(x)$ по вероятности стремится к $F(x)$.

Отметим, что мы опираемся на статистическое определение вероятности (теорему Бернулли), согласно которому в качестве вероятности берется относительная частота.

2) Точечные оценки параметров.

Рассмотрим задачу определения неизвестных числовых параметров распределения X. Вспомним некоторые основные моменты.

Пусть θ – неизвестный параметр. Функция (статистика) от выборочных значений $\theta^* = f(x_1, x_2, ..., x_n)$ называется точечной оценкой θ, если она дает некоторое приближенное значение θ. Вспомним, что выборка – это последовательность n независимых случайных величин, распределенных как случайная величина X. Поэтому θ* – случайная величина.

Оценка θ* называется несмещенной, если ее математическое ожидание равно оцениваемому параметру θ, т.е. M(θ*) = θ.

Оценка θ* называется состоятельной, если она сходится по вероятности к оцениваемому параметру, т.е. $\forall\ \varepsilon > 0\ \lim_{n\to\infty} p(|\theta^* - \theta| > \varepsilon) = 0$.

Оценка θ* называется эффективной в некотором классе точечных оценок, если ее дисперсия наименьшая среди дисперсий других оценок из этого класса.

Естественно, имеет смысл использовать состоятельные и несмещенные (или хотя бы асимптотически несмещенные), и по возможности, эффективные оценки.

Некоторые точечные оценки:

- $\bar{x} = \frac{1}{n}\sum_{i=1}^{k} x_i \cdot n_i$ – выборочное среднее, оценка для M(X);
- $S^2 = \frac{1}{n}\sum_{i=1}^{k} (x_i - \bar{x}) \cdot n_i$ – выборочная дисперсия, оценка для D(X);
- $A^* = \dfrac{\frac{1}{n}\sum_{i=1}^{k} (x_i - \bar{x})^3 \cdot n_i}{S^3}$ – эмпирический коэффициент ассиметрии, характеризует ассиметрию графика функции плотности относительно графика плотности нормального распределения;
- $E^* = \dfrac{\frac{1}{n}\sum_{i=1}^{k} (x_i - \bar{x})^4 \cdot n_i}{S^4} - 3$ – эмпирический коэффициент эксцесса, характеризует «крутизну» графика плотности относительно плотности нормального распределения с тем же M(X) и D(X).

Известно, что $\overline{X}$ – состоятельная и несмещенная (и эффективна в классе линейных оценок) оценка, остальные оценки - состоятельные и асимптотически несмещенные.

В курсе лекций обосновано, что в качестве значений x_i для интервального ряда надо брать середины интервалов, т.е. $x_i = \dfrac{a_i + a_{i+1}}{2}$.

Вычислять эти оценки удобно по таблице:

x_i	n_i	$x_i \cdot n_i$	$x_i - \overline{x}$	$(x_i - \overline{x})^2 \cdot n$	$(x_i - \overline{x})^3 \cdot n$	$(x_i - \overline{x})^4 \cdot n_i$
x_1	n_1	$x_1 \cdot n_1$	$x_1 - \overline{x}$	$(x_1 - \overline{x})^2 \cdot n$	$(x_1 - \overline{x})^3 \cdot n$	$(x_1 - \overline{x})^4 \cdot n_1$
....	....	....	....	....	....	....
x_k	n_k	$x_k \cdot n_k$	$x_k - \overline{x}$	$(x_k - \overline{x})^2 \cdot$	$(x_k - \overline{x})^3 \cdot$	$(x_k - \overline{x})^4 \cdot n_k$
Σ	n	$\sum x_i \cdot$	$\sum (x_i -$	$\sum (x_i - \overline{x})^2$	$\sum (x_i - \overline{x})^3$	$\sum (x_i - \overline{x})^4$

После оценивания параметров можно определиться с гипотетической функцией плотности и распределения. В качестве параметров распределений берут их точечные оценки.

3) Приближенная проверка нормальности распределения.

Для нормального распределения коэффициент ассиметрии A и эксцесс E равны нулю. Соответственно, близость к нулю A* и E* говорит о том, что распределение X близко к нормальному. В частности, если

$$|A^*| \le 3 \cdot \sqrt{\frac{6(n-1)}{(n+1)(n+3)}} \text{ и } |E^*| \le 3 \cdot \sqrt{\frac{24n \cdot (n-2) \cdot (n-3)}{(n-1)^2 \cdot (n+3) \cdot (n+5)}},$$

то есть основание предположить, что распределение X нормально, т.е.

$$f(x) = \frac{1}{\sqrt{2\pi} \cdot S} \cdot e^{\frac{-(x-\overline{x})^2}{2S^2}},$$

$$F(x) = \int_{-\infty}^{x} f(t)\,dt .$$

4) Интервальное оценивание параметров.

Точечная оценка может значительно отличаться от истинного значения параметра. Чтобы иметь представление о точности и надежности точечной оценки, строят доверительные интервалы.

Пусть θ^* – оценка для параметра θ. Зададим достаточно большую вероятность γ (например, $\gamma = 0{,}95$ или $\gamma = 0{,}99$) и найдем такое $\varepsilon > 0$, для которого $p(|\theta - \theta^*| < \varepsilon) = \gamma$ или, что то же самое, $p(\theta^* - \varepsilon < \theta < \theta^* + \varepsilon) = \gamma$.

Это означает, что с вероятностью γ истинное значение θ покрывается интервалом со случайными концами ($\theta^* - \varepsilon$, $\theta^* + \varepsilon$).

Величина γ называется надежностью или доверительной вероятностью, $\beta = 1 - \gamma$ – уровнем значимости, а ε – точностью оценки.

В равенстве θ^* – случайная величина. Если же θ^* конкретное значение оценки, вычисленное по выборке, то мы имеем следующее: вероятность ошибки при утверждении, что $\theta^* - \varepsilon < \theta < \theta^* + \varepsilon$, равна β.

4.1 Доверительный интервал для оценки неизвестного математического ожидания нормальной случайной величины.

Точечной оценкой для математического ожидания M(X) мы взяли выборочное среднее $\overline{X}$. Следовательно, доверительный интервал имеет вид

$$\left(\overline{x} - \varepsilon, \overline{x} + \varepsilon\right)$$

Задавая надежность γ, подберем ε так, чтобы выполнялось соотношение

$$p\left(\left|\overline{X} - M(X)\right| < \varepsilon\right) = \gamma .$$

Имеем два случая: первый – дисперсия нормальной случайной величины X известна, второй – дисперсия неизвестна.

4.2 Пусть дисперсия D(X) =σ^2 известна. Для отыскания ε воспользуемся тем, что $\overline{x}$ – нормальная сл   чайная величина с математическим ожиданием M(X) и дисперсией $\frac{\sigma^2}{n}$ (центральная предельная теорема). В силу этого $p\left(\left|\overline{X} - M(X)\right| < \varepsilon\right) = 2\Phi\left(\frac{\varepsilon \cdot \sqrt{n}}{\sigma}\right)$, где Φ(t) – функция Лапласа. Обозначим $t = \frac{\varepsilon \cdot \sqrt{n}}{\sigma^2}$. По условию 2$\Phi$(t) = γ, или $\Phi(t) = \frac{\gamma}{2}$. Из таблицы значений функции Лапласа найдем t_γ, такое, что $\Phi(t_\gamma) = \frac{\gamma}{2}$. Отсюда $\varepsilon = t_\gamma \cdot \frac{\sigma}{\sqrt{n}}$ и искомый интервал имеет вид:

$$\left(\overline{x} - t_\gamma \cdot \frac{\sigma}{\sqrt{n}}, \overline{x} + t_\gamma \cdot \frac{\sigma}{\sqrt{n}}\right)$$

Таким образом, $\overline{x} - t_\gamma \cdot \frac{\sigma}{\sqrt{n}} < M(X) < \overline{x} + t_\gamma \cdot \frac{\sigma}{\sqrt{n}}$, и вероятность того, что это утверждение неверно, равно $\beta = 1 - \gamma$.

Пусть дисперсия неизвестна. В этом случае, случайная величина (статистика) $T = \frac{\overline{x} - M(X)}{S} \cdot \sqrt{n-1}$ имеет распределение Стьюдента с n – 1 степенями свободы, где $S^2 = \frac{1}{n}\sum_{i=1}^{k}\left(x_i - \overline{x}\right)^2 \cdot n_i$.

По таблице распределения Стьюдента находим t_γ такое, что

$$P\left(|T| < t_\gamma\right) = \gamma .$$

Это равенство означает, что

$$P\left(\left|\overline{x} - M(X)\right| < t_\gamma \cdot \frac{S}{\sqrt{n-1}}\right) = \gamma ,$$

отсюда доверительный интервал имеет вид

$$\left(\overline{x} - t_\gamma \cdot \frac{S}{\sqrt{n-1}}, \overline{x} + t_\gamma \cdot \frac{S}{\sqrt{n-1}}\right).$$

Замечание. В случае если имеется таблица $P(|T| > t_\gamma)$, то по надежности γ и числу степеней свободы n – 1, находим t_γ такое, что $P(|T| > t_\gamma) = \beta = 1 - \gamma$.

4.3 Доверительный интервал для неизвестной дисперсии нормальной случайной величины.

Вспомним, что точечной оценкой для дисперсии $D(X) = \sigma^2$ служит статистика S^2. Для построения доверительного интервала воспользуемся тем, что случайная величина $\frac{nS^2}{\sigma^2}$ имеет распределение хи – квадрат (χ^2) с n – 1 степенями свободы.

При $n \le 30$ по таблице распределения χ^2 найдем числа χ_1 и χ_2 такие, что

$$P(\chi^2 < \chi_1^2) = \frac{1}{2}(1 - \gamma)$$
$$P(\chi^2 < \chi_2^2) = \frac{1}{2}(1 + \gamma),$$

где γ - заданная надежность.

Тогда $P\left(\chi_1^2 < \frac{nS^2}{\sigma^2} < \chi_1^2\right) = \gamma$.

Отсюда вытекает, что $P\left(\frac{nS^2}{\chi_2^2} < \sigma^2 < \frac{nS^2}{\chi_1^2}\right) = \gamma$.

Таким образом, $\left(\frac{nS^2}{\chi_2^2}, \frac{nS^2}{\chi_1^2}\right)$ – доверительный интервал дисперсии σ^2, а $\left(\frac{\sqrt{n}\,S}{\chi_2}, \frac{\sqrt{n}\,S}{\chi_1}\right)$ – доверительный интервал для среднего квадратического отклонения σ.

При $n > 30$ по надежности γ и числу степеней свободы n – 1 найдем $\gamma_1 = \frac{\sqrt{n}}{\chi_2}$ и $\gamma_2 = \frac{\sqrt{n}}{\chi_1}$.

Тогда доверительный интервал для дисперсии примет вид $(\gamma_1 S^2, \gamma_2 S^2)$, а для среднего квадратического отклонения – $(\gamma_1 S, \gamma_2 S)$.

5) Проверка статистических гипотез.

Любое предположение о генеральной совокупности X, сделанные на основе выборки, – статистическая гипотеза.

Правило, показывающее, когда статистическую гипотезу надо принять, а когда отвергнуть, – статистический критерий.

Критерии, относящиеся к предположениям о виде функции распределения, называются критериями согласия.

Рассмотрим один из критериев согласия – критерий хи-квадрат (или критерий Пирсона).

Пусть выдвинута гипотеза о том, что генеральная совокупность X имеет функцию распределения F(x) или плотность f(x). Далее, пусть вся область изменения величины X разбита на k интервалов $(a_1, a_2]$, $(a_2, a_3]$, ..., $(a_k, a_{k+1}]$, и пусть P_i – вероятность для величины X при гипотетическом распределении F(x) попасть в интервал $(a_i, a_{i+1}]$:

$$P_i = P(a_i < x < a_{i+1}) = F(a_{i+1}) - F(a_i) = \int_{a_i}^{a_{i+1}} f(x)\,dx$$

Пусть n_i – число выборочных значений, попавших в интервал $(a_i, a_{i+1}]$. Тогда $\frac{n_i}{n}$ – относительная частота попадания в i-ый интервал (статистическая вероятность попадания в интервал).

Если величины $\frac{n_i}{n}$ и P_i (i = 1, 2, ..., k) мало различаются, то разумно считать, что выдвинутая гипотеза не противоречит опытным данным. Это обстоятельство лежит в основе χ^2.

Рассмотрим статистику $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - nP_i)^2}{nP_i}$.

При больших n статистика χ^2 практически не зависит от гипотетического распределения F(x) и имеет распределение χ^2 с r степенями свободы, где r = k – m – 1, где m – количество неизвестных параметров гипотетического распределения, оцениваемых по выборке.

Чтобы сформулировать правило проверки гипотезы, сначала зададим уровень значимости β (β = 0,05 или β = 0,1) и по таблице χ^2 с r степенями свободы и β находим число $\chi^2_{крит}$ такое, что выполняется равенство

$$P(\chi^2 \ge \chi^2_{крит}) = \beta .$$

Если значение χ^2, вычисляемое по выборке, окажется больше или равно $\chi^2_{крит}$, то гипотеза отвергается, в противном случае гипотеза принимается (не противоречит наблюдениям). По этому правилу верная гипотеза отвергается с вероятностью β.

<u>Схема применения критерия</u>

1. Вычисляем χ^2 для нашей выборки по формуле $\chi^2 = \sum_{i=1}^{k} \frac{(n_i - nP_i)^2}{nP_i}$.

2. По уровню значимости β и числу степеней свободы r = k – m – 1 из таблицы распределения χ^2 находим $\chi^2_{крит}$ ..

3. Если $\chi^2 \geq \chi^2_{крит}$, гипотеза отвергается, если $\chi^2 < \chi^2_{крит}$, то гипотеза считается не противоречащей наблюдениям.

6. В случае верно выбранного гипотетического распределения F(x), график выбранной функции плотности должен мало отличаться от гистограммы.

Часть II

Задание

1) По построенным интервальным рядам для компонент X и У построить двумерное выборочное распределение (X, У).

2) Найти выборочную ковариацию. Если ковариация не равна нулю, то в генеральной совокупности компоненты X и У коррелированны.

3) Найти условные выборочные средние $\overline{y}_x$ u $\overline{x}_y$, построить на одном чертеже ломанные регрессии $\overline{y}_x$ на x и $\overline{x}_y$ на y.

4) Найти выборочный коэффициент корреляции $\rho_в$, проверить гипотезу о значимости коэффициента корреляции при уровне значимости β.

5) Найти уравнения линейной регрессии $\overline{y}_x$ на x и $\overline{x}_y$ на y, построить полученные линии регрессии на том же чертеже, что и ломанные.

Указания

Величины У и X могут быть связаны функциональной зависимостью y = f(x). Если закон распределения случайной величины У зависит от X, то имеется стохастическая зависимость. Если условное математическое ожидание У функционально зависит от X, то имеется корреляционная зависимость – частный случай стохастической.

Представление о двумерной случайной величине (X, У) дает нам двумерное выборочное распределение.

1) Строим корреляционную таблицу (фактически двумерную случайную величину).

X \ Y		y_1	y_2	…	y_l	
		$(b_1, b_2]$	$(b_2, b_3]$		$(b_k, b_{k+1}]$	n_i
x_1	$(a_1, a_2]$	n_{11}	n_{12}		n_{1k}	$\sum n_{1j}$
x_2	$(a_2, a_3]$	n_{21}	n_{22}		n_{2k}	$\sum n_{2j}$
…	…	…	…	…	…	…
x_l	$(a_l, a_{l+1}]$	n_{l1}	n_{l2}	…	n_{lk}	…
	n_j	$\sum n_i$	$\sum n_i$	…	…	$\sum n_{ij}$

В клетках таблицы записываем числа n_{ij} – частота попадания выборочных значений (x_t, y_t) в прямоугольник со сторонами

$(a_i, a_{i+1}]$ и $(b_j, b_{j+1}]$, $\sum n_{ij} = n$. В последней строке таблицы – числа n_j – частоты попадания Y в интервалы $(b_j, b_{j+1}]$, соответственно в последнем столбце – частоты попадания X в интервалы $(a_i, a_{i+1}]$.

Первая и последняя строка – интервальный ряд для У, первый и последний столбец – интервальный ряд для X (см. часть I).

2) По определению считаем выборочную ковариацию:

$$cov(x,y) = \overline{xy} - \overline{x}\cdot\overline{y}, \quad где\ \overline{xy} = \frac{1}{n}\sum x_i y_j n_{ij}$$

3) Находим условные средние:

$$\overline{y}_{x=x_i} = \frac{1}{n_i}\sum_j y_j \cdot n_{ij}, \quad i = 1,2,...l$$

$$\overline{x}_{y=y_i} = \frac{1}{n_j}\sum_i x_i \cdot n_{ij}, \quad j = 1,2,...k$$

По точкам (x, $\overline{y}_x$) строим ломанную регрессии $\overline{y}_x$ на x, и соответственно ломанную регрессии $\overline{x}_y$ на y.

4) Выборочный коэффициент корреляции

$$\rho_в = \frac{cov(x,y)}{S(x)\cdot S(y)}$$

характеризует тесноту линейной связи: чем ближе $|\rho_в|$ к единице, тем сильнее линейная зависимость.

5) Методом наименьших квадратов определяем параметры уравнения регрессии $\overline{y}_x = \alpha + \beta x$:

$$\beta = \rho_в \cdot \frac{S(y)}{S(x)}, \quad \alpha = \overline{y} - \beta \overline{x}$$

Соответственно, $\overline{x}_y = \gamma + \delta x$,

$$\delta = \rho_в \cdot \frac{S(x)}{S(y)}, \quad \gamma = \overline{x} - \delta \overline{y}$$

Строим графики полученных линий регрессий.

Индивидуальные задания

Вариант 1

X	Y	X	Y	X	Y	X	Y
59	57	61	46	63	53	62	58
71	67	63	55	65	50	62	54
61	43	59	57	64	61	64	60
67	60	64	49	68	50	64	63
62	55	63	55	70	63	62	43
62	59	62	47	87	41	56	38
61	59	68	67	65	53	65	49
59	54	65	57	69	66	66	49
65	49	68	56	62	59	64	44
63	47	65	45	65	64	61	54
65	62	62	46	66	62	62	42
62	56	68	50	66	65	69	54
62	49	64	55	62	45	65	61
65	55	60	47	60	44	62	45
67	50	64	59	64	44	66	62
63	50	67	59	60	40	63	48
58	54	68	58	60	55	57	39
64	53	62	54	63	43	55	51
64	45	67	50				

Вариант 2

X	Y	X	Y	X	Y	X	Y
31	-37	23	-32	28	-34	31	-38
28	-36	28	-30	26	-29	25	-30
30	-32	26	-35	29	-38	27	-31
23	-30	26	-34	26	-30	26	-32
25	-32	28	-29	23	-33	27	-31
25	-33	27	-35	27	-30	28	-38
25	-27	25	-34	29	-39	24	-32
27	-28	26	-30	24	-25	24	-32
31	-40	27	-35	27	-32	26	-28
25	-33	28	-30	25	-35	25	-34
28	-31	27	-35	29	-30	22	-29
25	-31	29	-36	29	-34	27	-35
30	-39	24	-33	25	-33	26	-32
28	-30	23	-24	25	-31	24	-30
28	-34	27	-29	30	-34	25	-29
31	-37	27	-32	25	-29	22	-31

Вариант 3

X	Y	X	Y	X	Y	X	Y
77	72	74	66	67	57	72	58
76	58	77	68	69	56	66	49
62	45	69	66	66	62	80	63
83	80	72	60	71	53	72	62
83	71	76	58	71	69	68	61
79	70	69	60	77	60	76	57
77	61	84	76	78	72	81	79
81	66	77	61	77	73	69	55
71	51	72	59	67	63	83	64
83	70	76	56	73	65	77	70
73	56	78	61	67	53	69	62
81	75	87	69	83	69	81	80
82	71	79	70	67	53	65	62
77	60	80	68	76	66	71	55
82	63	70	54	71	59	69	65
83	77	80	70	74	60	85	68
80	61	75	71	73	58	81	62
77	60	69	54	75	60	72	56
69	57	77	61	80	61	84	77
78	74						

Вариант 4

X	Y	X	Y	X	Y	X	Y
13	88	17	117	13	76	20	134
11	72	16	110	16	102	15	94
18	123	17	117	15	97	16	110
13	90	16	97	22	149	16	101
15	93	17	117	14	85	10	67
21	135	13	82	15	95	15	102
17	112	18	112	12	64	23	153
13	75	14	88	12	74	15	93
16	110	13	71	11	63	17	105
13	74	11	60	16	109	17	102
19	115	15	102	17	105	17	109
17	116	10	69	15	92	19	119
15	87	15	93	14	83	16	102
12	76	14	93	13	82	11	64
17	118	9	59	17	101	13	75
20	120	19	127	15	91	17	101
17	107	18	107	14	91	20	133
18	115	15	101	12	78	19	132
17	116	15	92	12	72		

Вариант 5

X	Y	X	Y	X	Y	X	Y
46	44	57	55	56	46	68	59
40	39	45	39	48	47	47	39
58	49	55	50	44	36	40	38
35	25	46	41	49	44	35	29
53	47	53	47	50	40	60	53
47	42	45	44	46	36	37	35
40	32	43	38	46	41	41	37
60	53	48	43	47	41	46	36
39	29	50	47	39	31	46	38
41	39	44	35	49	43	53	45
58	54	38	33	44	37	43	34
59	55	68	66	49	45	53	48
57	50	54	44	53	51	54	48
65	55	57	54	44	41	50	44
34	28	34	28	47	43	45	39
50	42	59	52	51	41	50	44
45	36	57	50	56	52	25	23
50	47	57	50	39	33	44	43
55	52	49	41	46	38	54	47

Вариант 6

X	Y	X	Y	X	Y	X	Y
48	520	61	664	52	564	50	542
40	435	42	456	32	345	49	538
52	564	55	601	48	518	54	588
50	541	52	567	40	430	45	485
39	424	44	475	40	436	36	387
47	516	41	450	46	504	41	441
38	413	47	516	41	450	46	503
46	505	43	468	54	584	38	416
47	514	55	602	48	521	44	483
44	479	43	472	53	580	35	376
45	490	49	538	47	509	47	512
44	483	42	459	50	542	44	474
53	577	31	333	46	498	57	624
52	569	40	432	56	608	53	579
45	493	47	516	42	461	51	554
42	461	43	471	41	445	48	519
45	486	48	527	55	595	46	496
45	492	44	483	40	431	43	467

Вариант 7

X	Y	X	Y	X	Y	X	Y
7	38	6	28	6	35	11	58
11	63	5	27	10	55	9	45
10	55	10	56	14	79	12	64
4	16	10	55	12	65	10	58
11	64	13	74	6	29	7	33
12	70	15	84	6	35	8	46
9	50	6	31	8	47	9	46
5	22	11	58	6	28	9	50
8	44	10	54	6	31	11	63
13	76	14	78	13	77	13	70
12	71	8	47	11	59	9	52
10	58	14	79	9	45	9	49
9	45	6	35	10	58	10	57
9	49	16	95	8	45	13	77
11	65	8	39	12	66	7	32
9	44	6	33	14	75	9	50
10	55	8	41	14	76	8	40
10	54	8	40	7	38	16	92
13	75	7	39	16	90	12	65
15	89	8	38	11	63	15	81
4	22	10	59	11	59	13	75
9	48	12	69	12	70	11	62
14	76						

Вариант 8

X	Y	X	Y	X	Y	X	Y
40	272	40	212	41	279	40	270
40	260	39	204	40	275	40	266
42	281	41	224	41	276	41	274
41	271	42	223	40	278	40	278
42	289	41	215	41	273	41	278
41	277	42	214	42	282	40	263
40	264	40	210	42	287	41	283
41	280	41	207	40	267	39	257
41	280	42	215	40	275	41	279
39	267	40	203	41	277	41	267
38	256	41	207	40	270	40	264
39	263	40	212	41	271	41	284
42	293	42	226	41	267	39	257
40	277	40	204	40	273	41	278
40	261	42	221	41	284	40	276
41	280	41	220	39	255		

Вариант 9

X	Y	X	Y	X	Y	X	Y
37	58	41	72	42	72	40	73
39	63	31	51	39	60	35	56
30	59	44	71	34	56	24	31
32	59	38	63	36	71	37	69
40	67	28	55	31	52	42	78
34	53	42	81	29	44	38	63
33	53	31	46	41	77	35	69
40	69	26	42	27	47	38	74
38	68	32	58	27	51	32	58
35	55	31	59	37	66	38	60
37	67	40	61	38	72	37	71
35	53	31	43	47	90	41	81
26	37	34	53	40	67	33	60
29	45	45	74	33	65	32	58
37	73	36	65	36	62	42	72
31	50	38	63	36	69	28	50
38	66	35	32	41	81	44	73
38	59	32	51	41	75	42	75
33	54	37	63	35	52	35	63
36	68	40	67	39	70	33	60
37	58	31	59	33	49	35	61
48	93						

Вариант 10

X	Y	X	Y	X	Y	X	Y
17	97	15	80	16	77	16	92
16	94	18	104	16	84	17	94
17	97	16	86	16	94	16	91
16	90	18	94	17	91	18	103
15	75	18	91	18	103	15	76
17	99	15	70	16	94	16	89
17	88	18	99	17	97	17	100
16	82	17	85	14	79	19	107
17	83	18	101	17	90	15	80
16	81	15	85	15	76	16	77
16	85	17	100	18	94	15	79
18	95	17	84	17	84	18	98
15	73	16	80	16	77	15	85
16	82	18	105	18	94	16	93
16	93	16	88	15	76	15	73
7	89	18	100				

Приложение 1: χ^2 – распределение (распределение Пирсона)
$P(\chi^2 > x_\alpha) = \alpha$

k / α	0,99	0,95	0,90	0,80	0,20	0,10	0,05	0,02	0,01
1	0,00016	0,0063	0,393	0,0642	1,642	2,706	3,841	5,412	6,635
2	0,0201	0,0404	0,103	0,446	3,219	4,605	5,991	7,824	9,210
3	0,115	0,185	0,352	1,005	4,642	6,251	7,815	9,837	11,341
4	0,297	0,429	0,711	1,649	5,989	7,779	9,488	11,668	13,277
5	0554	0,752	1,145	2,343	7,289	9,236	11,070	13,388	15,086
6	0,872	1,134	1,635	3,070	8,558	10,645	12,592	15,033	16,812
7	1,239	1,564	2,167	3,822	9,803	12,017	14,067	16,622	18,475
8	1,646	2,032	2,733	4,594	11,030	13,362	15,507	18,168	20,090
9	2,088	2,532	3,325	5,380	12,242	14,684	16,919	19,679	21,666
0	2,588	3,059	3,940	6,179	13,442	15,987	18,307	21,161	23,209
1	3,053	3,609	4,575	6,989	14,631	17,275	19,675	22,618	24,725
2	3,571	4.178	5,226	7,807	15,812	18,549	21,026	24,054	26,217
3	4,107	4,765	5,892	8,634	16,985	19,812	22,362	25,472	27,688
4	4,660	5,368	6,571	9,467	18,151	21,064	23,685	26,873	29,141
5	5,229	5,985	7,261	10,307	19,311	22,307	24,996	28,259	30,578
6	5,812	6,614	7,962	11,152	20,465	23,542	26,296	29,633	32,000
17	6,408	7,255	8,672	12,002	21,615	24,769	27,587	30,995	33,409
18	7,015	7,906	9,390	12,857	22,760	25,989	28,869	32,346	34,805
19	7,633	8,567	10,117	13,716	23,900	27,204	30,144	33,687	36,191
20	8,260	9,237	10,851	14,578	25,038	28,412	31,410	35,020	37,566
21	8,897	9,915	11,591	15,445	26,171	29,615	32,671	36,343	38,932
22	9,542	10,600	12,338	16,314	27,301	30,813	33,924	37,659	40,289
23	10,196	11,293	13,091	17,187	28,429	32,007	35,172	38,968	41,638
24	10,856	11,992	13,848	18,062	29,553	33,196	36,415	40,270	42,980
25	11,524	12,697	14,611	18,940	30,675	34,382	37,652	41,566	44,314
26	12,198	13,409	15,379	19,820	31,795	35,563	38,885	42,856	45,642
27	12,879	14,125	16,151	20,703	32,912	36,741	40,112	44,140	46,963
28	13,565	14,847	16,928	21,588	34,027	37,916	41,337	45,419	48,278
29	14,256	15,574	17,708	22,475	35,139	39,087	42,557	46,693	49,588
30	14,953	16,306	18,493	23,364	36,250	40,256	43,773	47,962	50,892

Приложение 2: t – распределение (распределение Стьюдента)

$P(t > t_{\alpha}) = \alpha$ и $P(\,|t| > t_{\alpha}) = \alpha$

	Односторонняя критическая область (α)							
	0,1	0,05	0,025	0,01	0,005	0,0025	0,001	0,0005
	Двусторонняя критическая область (α)							
k	0,2	0,1	0,05	0,02	0,01	0,005	0,002	0,001
1	3,08	6,31	12,71	31,82	63,66	127,32	318,30	636,61
2	1,89	2,92	4,30	6,96	9,92	14,09	22,33	31,60
3	1,64	2,35	3,18	4,54	5,84	7,45	10,21	12,92
4	1,53	2,13	2,78	3,75	4,60	5,60	7,17	8,61
5	1,48	2,02	2,57	3,36	4,03	4,77	5,89	6,87
6	1,44	1,94	2,45	3,14	3,71	4,32	5,21	5,96
7	1,41	1,89	2,36	3,00	3,50	4,03	4,79	5,41
8	1,40	1,86	2,31	2,90	3,36	3,83	4,50	5,04
9	1,38	1,83	2,26	2,82	3,25	3,69	4,30	4,78
10	1,37	1,81	2,23	2,76	3,17	3,58	4,14	4,59
11	1,36	1,80	2,20	2,72	3,11	3,50	4,02	4,44
12	1,36	1,78	2,18	2,68	3,05	3,43	3,93	4,32
13	1,35	1,77	2,16	2,65	3,01	3,37	3,85	4,22
14	1,34	1,76	2,14	2,62	2,98	3,33	3,79	4,14
15	1,34	1,75	2,13	2,60	2,95	3,29	3,73	4,07
16	1,34	1,75	2,12	2,58	2,92	3,25	3,69	4,02
17	1,33	1,74	2,11	2,57	2,90	3,22	3,65	3,97
18	1,33	1,73	2,10	2,55	2,88	3,20	3,61	3,92
19	1,33	1,73	2,09	2,54	2.86	3,17	3,58	3,88
20	1,33	1,72	2,09	2,53	2,85	3,15	3,55	3,85
21	1,32	1,72	2,08	2,52	2,83	3,14	3,53	3,82
22	1,32	1,72	2,07	2,51	2,82	3,12	3,51	3,79
23	1,32	1,71	2,07	2,50	2,81	3,10	3,48	3,77
24	1,32	1,71	2,06	2,49	2,80	3,09	3,47	3,75
25	1,32	1,71	2,06	2,49	2,79	3,08	3,45	3,73
26	1,32	1,71	2,06	2,48	2,78	3,07	3,44	3,71
27	1,31	1,70	2,05	2,47	2,77	3,06	3,42	3,69
28	1,31	1,70	2,05	2.47	2,76	3,05	3,41	3,67
29	1,31	1,70	2,05	2,46	2,76	3,04	3,40	3,66
30	1,31	1,70	2,04	2,46	2,75	3,03	3,39	3,65
40	1,30	1,30	1,68	2,02	2,42	2,70	2,97	3,31
60	1,30	1,30	1,67	2,00	2,39	2,66	2,91	3,23
120	1,29	1,66	1,98	2,36	2,62	2,85	3,16	3,37
∞	1,28	1,64	1,96	2,33	2,58	2,81	3,09	3,29

Приложение 3: Функция распределения 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
00,0	50000	50399	50798	51197	51595	51994	52392	52790	53188	53586
0,1	53983	54380	54776	55172	55567	55962	56356	56749	57142	57535
0,2	57926	58317	58706	59095	59483	59871	60257	60642	61026	61409
0,3	61791	62172	62552	62930	63307	63683	64058	64431	64803	65173
0,4	65542	65910	66276	66640	67003	67364	67724	68082	68439	68793
0,5	69146	69497	69847	70194	70540	70884	71226	71566	71904	72240
0,6	72575	72907	73237	73565	73891	74215	74537	74857	75175	75490
0,7	75804	76115	76424	76730	77035	77337	77637	77935	78230	78524
0,8	78814	79103	79389	79673	79955	80234	80511	80785	81057	81327
0,9	51594	81859	82121	82381	82639	82894	83147	83398	83646	83891
1,0	84134	84375	84614	84850	85083	85314	85543	85769	85993	86214
1,1	86433	86650	86864	87076	87286	87493	87698	87900	88100	88298
1,2	88493	88686	88877	89065	89251	89435	89617	89796	89973	90147
1,3	90320	90490	90658	90824	90988	91149	91308	91466	91621	91774
1,4	91924	92073	92220	92364	92507	92647	92786	92922	93056	93189
1,5	93319	93448	93574	93699	93822	93943	94062	94179	94295	94408
1,6	94520	94630	94738	94845	94950	95053	95154	95254	95352	95449
1,7	95543	95637	95728	95818	95907	95994	96080	96164	96246	96327
1,8	96407	96485	96562	96638	96712	96784	96856	96926	96995	97062
1,9	97128	97193	97257	97320	97381	97441	97500	97558	97615	97670
2,0	97725	97778	97831	97882	97932	97982	98030	98077	98124	98169
2,1	98214	98257	98300	98341	98382	98422	98461	98500	98537	98574
2,2	98610	98645	98679	98713	98745	98778	98809	98840	98870	98899
22,3	98928	98956	98983	99010	99036	99061	99086	99111	99134	99158
22,4	99180	99202	99224	99245	99266	99286	99305	99324	99343	99361
22,5	99379	99396	99413	99430	99446	99461	99477	99492	99506	99520
22,6	99534	99547	99560	99573	99585	99598	99609	99621	99632	99643
2,7	99653	99664	99674	99683	99693	99702	99711	99720	99728	99736
22,8	99744	99752	99760	99767	99774	99781	99788	99795	99801	99807
22,9	99813	99819	99825	99831	99836	99841	99846	99851	99856	99861
33,0	99865	99869	99874	99878	99882	99886	99889	99893	99896	99900
3,1			99910	99913	99916	99918	99921	99924	99926	99929
33,2	99931	99934	99936	99938	99940	99942	99944	99946	99948	99950
33,3	99952	99953	99955	99957	99958	99960	99961	99962	99964	99965

33,4	99966	99968	99969	99970	99971	99972	99973	99974	99975	99976
33,5	99977	99978	99978	99979	99980	99981	99981	99982	99983	99983
33,6	99984	99985	99985	99986	99986	99987	99987	99988	99988	99989
33,7	99989	99990	99990	99990	99991	99991	99992	99992	99992	99992
33,8	99993	99993	99993	99994	99994	99994	99994	99995	99995	99995
33,9	99995	99995	99996	99996	99996	99996	99996	99996	99997	99997
44,0	99997	99998	99999	99999	99999	–	–	–	–	–

Рекомендуемая литература

1. Гмурман В.Е. Руководство к решению задач по теории вероятностей и математической статистике. – М.: Высш. шк., 1998.
2. Гмурман В.Е. Теория вероятностей и математическая статистика. Учеб. Пособие для вузов. – М.: Высш. шк., 1998
3. Кремер Н.Ш. Математика для экономистов.
4. Колде Я.К. Практикум по теории вероятностей и математической статтистики: Учеб. пособие для техникумов. – М.: Высш. шк., 1991

Батуев Эрдэмто Николаевич
Сидорова Александра Сергеевна

МАТЕМАТИЧЕСКАЯ СТАТИСТИКА

Обработка статистических данных

Корреляционная зависимость

Задания и методические указания
для студентов дневгого и заочного отделений

В авторской редакции
Набор, верстка Сидорова А.С.
Лицензия ИД №02187 от 30.06.00 г. Подписано в печать____
Формат 61*86/16.Печать офсетная. Гарнитура Times New Roman
Усл. печ. л. ___
Тираж __ экз. Заказ №

Редакционно-издательский отдел
Камчатского государственного технического университета

Отпечатано полиграфическим участком РИО КамчатГТУ
683003, г. Петропавловск-Камчатский, ул. Ключевская, 35