**Государственный комитет Российской Федерации**
**по высшему образованию**

**Новосибирский государственный технический университет**

---



Б.Ю. Лемешко

# *Корреляционный анализ многомерных случайных величин*

Программная система

Новосибирск - 1995

УДК 519.237 (075.8)+519.214.4 (075.8)

Б.Ю. Лемешко. Корреляционный анализ многомерных случайных величин: Программная система / Новосиб. гос. техн. ун-т. - Новосибирск, 1995. - 39 с.

ISBN 5-230-12109-2

Программная система предназначена для решения совокупности задач статистического анализа наблюдений многомерных величин.

В пособии рассмотрены основные процедуры оценивания и проверки статистических гипотез, используемые при анализе многомерных величин. Изложение материала соответствует структуре программного обеспечения и перечню решаемых с его использованием задач, приведены рекомендации и инструкции по использованию системы.

Программная система предназначена для студентов, аспирантов, научных и инженерных работников, а также специалистов различных отраслей, использующих аппарат теории вероятностей и математической статистики, и рекомендуется в качестве учебного пособия для студентов, обучающихся по специальности "Прикладная математика".

Рецензент: А.А. Попов, к-т. техн. наук, доц.

Работа подготовлена на кафедре прикладной математики

ISBN 5-230-12109-2

При анализе совокупности случайных величин нас может интересовать либо взаимозависимость между несколькими случайными величинами, либо зависимость одной или большего числа величин от остальных.

Теория корреляции занимается изучением *взаимозависимости*. Именно на это нацелены алгоритмы и методы корреляционного анализа многомерных случайных величин. С помощью аппарата корреляционного анализа можно убедиться в наличии взаимозависимости величин, исследовать взаимозависимость величин при устранении влияния совокупности других (частная корреляция), рассмотреть зависимость одной величины от группы величин (множественная корреляция), прийти к выводу о наличии статистической зависимости, возможно даже предположить и наличие функциональной зависимости, скрытой за ошибками результатов измерений. Но с помощью статистических методов невозможно установить причинной связи. Исследование функциональных закономерностей, присущих соответствующей области, лежит вообще за рамками статистики. Корреляционный анализ не занимается и построением регрессионных зависимостей. Исследованием статистических зависимостей, построением моделей занимаются уже в рамках регрессионного анализа.

С использованием результатов корреляционного анализа исследователь может делать определённые выводы о наличии и характере *взаимозависимости*, что уже само по себе может представлять существенную информацию об исследуемом объекте. Результаты могут подсказать и направление дальнейших исследований, и совокупность требуемых методов, в том числе статистических, необходимых для более полного изучения объекта.

Особенно реальную пользу применение аппарата корреляционного анализа может принести на стадии ранних исследований в областях, где характеры причин определённых явлений ещё недостаточно понятны. Это может касаться изучения очень сложных систем различного характера: как технических, так и социальных.

## 1. Назначение

Комплекс программ корреляционного анализа предназначен для анализа и обработки выборок многомерных случайных величин (векторного параметра некоторой системы или некоторого технологического процесса).

Он позволяет вычислять оценки вектора математических ожиданий, ковариационной матрицы, матрицы парных корреляций, матрицы корреляционных отношений, матриц частных корреляций, вектора множественных корреляций, проверять гипотезы о значимости различных параметров или о равенстве их определенным значениям, о наличии функциональных зависимостей.

Предоставляет возможности анализа многомерных выборок на наличие аномальных измерений, на случайность, на принадлежность к нормальной генеральной совокупности.

Позволяет вычислять вероятности и квантили для ряда статистических распределений, наиболее часто используемых в задачах проверки статистических гипотез.

Использование комплекса дает возможность выявления характера статистических зависимостей между компонентами многомерной случайной величины и определения меры линейности этих связей.

Комплекс программ будет весьма полезен при обработке результатов экспериментальных исследований и анализе статистических данных в технике, медицине, биологии, при обработке социологических обследований.

## 2. Состав комплекса

В состав программного обеспечения входят 5 файлов:

*kora.exe*
*kora00.exe*
*kora01.exe*
*kr_ega.exe*
*title.leo*
*kora.hlp*

*act.txt*

Головной программой является *kora.exe*. Остальные файлы должны находиться в этой же директории.

Файлы *kora00.exe* и *kora01.exe* осуществляют вычисления в соответствии с алгоритмами выбранных задач и запускаются непосредственно программой *kora.exe*. Управляющая информация о заданной задаче, исходной выборке, ее местонахождении и определенные пользователем известные ему параметры задачи передаются через формируемый файл с именем *KoraBase.txt*.

В процессе работы по желанию пользователя система создает файл результатов (по умолчанию *KoraRes.txt*).

## 3. Работа с меню системы

В меню системы входит ряд сервисных функций, поддерживающих выбор требуемой задачи, определение входного файла, содержащего данные с исходной выборкой, задание размерности анализируемой случайной величины и объема выборки, задание имени файла, в который будут записаны результаты вычислений, запуск выбранной задачи на выполнение. Встроенный редактор позволяет создавать исходные файлы с данными, просматривать и при необходимости корректировать файлы с результатами.

Движение по разделам меню и выбор определенной задачи или функции осуществляется перемещением указателя курсора и клавишей <*Enter*> или с помощью манипулятора "мышь". Выбор определенной задачи осуществляется повторным нажатием клавиши <*Enter*>. Для возврата на предыдущий уровень меню необходимо нажать клавишу <*Esc*>.

Для обращения за пояснениями к выбираемой задаче или функции (для обращения за справками к файлу *kora.hlp*) необходимо нажать клавишу <*F1*>. Выбор соответствующей подсказки может быть осуществлен указателем курсора и клавишей <*Enter*> или манипулятором "мышь". Отказ от помощи - манипулятором "мышь", а при ее отсутствии - клавишей <*Esc*>.

## 4. Подготовка исходных данных

Файл *act.txt* - это демонстрационный файл с исходной выборкой случайной величины размерности $m = 4$ и объемом $n = 60$ является примером подготовки исходных данных. Все приводимые ниже примеры используют в качестве исходной выборки этот файл.
Информация, содержащаяся в этом файле, имеет следующий вид:

| X1 | X2 | X3 | X4 |
|---|---|---|---|
| 154.0 | 178.0 | 59.0 | 21.0 |
| 133.0 | 164.0 | 63.0 | 27.0 |
| 58.0 | 75.0 | 36.0 | 18.0 |
| 145.0 | 161.0 | 62.0 | 10.0 |
| 94.0 | 107.0 | 48.0 | 26.0 |
| ... | ... | ... | ... |
| 85.0 | 91.0 | 48.0 | 25.0 |

Первая строка файла является символьной, она обязательно должна присутствовать, но может иметь и другой содержательный смысл (вид).

Результаты наблюдений многомерной случайной величины (выборка) должны быть представлены в исходном файле в аналогичном виде. Размерность наблюдаемой величины определит количество X-столбцов, а объем выборки - количество строк.

## 5. Вывод результатов

Результаты вычислений оценок параметров или проверки конкретных гипотез могут выводиться либо в файл на диске, либо на экран. Но в последнем случае их затруднительно фиксировать.

При выборе вывода результатов *"В файл"* система предлагает вывод в файл *"KoraRes.txt"* в текущей директории, но может быть указан любой другой файл. При решении каждой очередной задачи результаты решения предыдущей теряются, если только Вы не поменяете имя файла вывода.

После завершения очередной задачи результаты ее решения могут быть вызваны для визуального просмотра. Для этого в меню *"Выход"* должно быть выбрано подменю *"Просмотр"*, а затем команда *"Просмотр результатов"*.

## 6. Задачи корреляционного анализа

### 6.1. Проверка предпосылок

Сюда входят проверка выборки на наличие грубых, аномальных измерений, проверка гипотез о случайности выборочных данных и принадлежности выборки нормальной генеральной совокупности. Статистические выводы осуществляются отдельно по каждой компоненте наблюдённой многомерной случайной величины.

***Выделение аномальных измерений***. Резко выделяющиеся наблюдения, или выбросы, могут быть реакцией системы на некоторое резкое случайное воздействие либо быть ошибками измерений. Выброс может оказаться также одним из экстремальных значений случайной величины. В первом случае выброс должен быть отброшен, а во втором - он несёт полезную информацию о системе.

Если выборка взята из нормальной совокупности с известной дисперсией, то самый простой анализ её на наличие аномальных измерений может быть проведен по правилу "трёх сигм". Пусть $X_1, X_2, \ldots, X_n$ – выборка одномерной случайной величины, $\overline{X}$ - выборочное среднее. Тогда величины $Y_i = X_i - \overline{X}$, $i = \overline{1,n}$, распределены по нормальному закону с нулевым математическим ожиданием и дисперсией $\sigma^2(1+1/n)$. Поэтому как аномальные надо отбросить те наблюдения, для которых $|Y_i| > 3\sigma\sqrt{1+1/n}$.

***Проверка случайности выборки***. Внешним проявлением случайности наблюдений является полная непредсказуемость очередного результата. Наиболее часто при проверке случайности используется непараметрический критерий ранговой корреляции Кендалла.

*Рангом* $r_i$ $i$-го наблюдения $X_i$ из выборки $X_1, X_2, \ldots, X_n$ называется номер этого наблюдения в упорядоченной по возрастанию последо-

вательности. *Инверсией* называется ситуация, когда больший ранг предшествует меньшему: $r_i > r_j$ при $i < j$.

Статистика

$$t = 1 - \frac{4u}{n^2 - n},$$

где $u$ - число инверсий в перестановке рангов, $n$ - объём выборки, имеет нормальное распределение с нулевым математическим ожиданием и дисперсией $\tilde{\sigma}^2 = \frac{2(2n+5)}{9n(n-1)}$. Гипотеза о случайности принимается, если $|t| \le \tilde{\sigma} t_{\alpha/2}$ и $t_{\alpha/2}$ определяется из условия

$$1 - \alpha = \frac{1}{\sqrt{2\pi}} \int_{-t_{\alpha/2}}^{t_{\alpha/2}} e^{-t^2/2} dt.$$

***Проверка нормальности выборки***. При проверке гипотезы о нормальности выборки одним из критериев проверки является следующий. Вычисляется статистика

$$d = \sum_{i=1}^{n} \frac{|X_i - \overline{X}|}{nS},$$

где $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $S = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(X_i - \overline{X})^2}$, которая при $n \ge 50$ распределена приближенно по нормальному закону с математическим ожиданием $m_d = \sqrt{\frac{2}{\pi}}\left(1 + \frac{2}{8n-9}\right)$ и дисперсией $\sigma_d^2 = \left(\frac{\pi - 3}{\pi n} - \frac{1}{4\pi n^2}\right)^2$. Если выполняется неравенство

$$\frac{\left|d - \sqrt{\frac{2}{\pi}}\left(1 + \frac{2}{8n-9}\right)\right|}{\frac{\pi - 3}{\pi n} - \frac{1}{4\pi n^2}} \le t_{\alpha/2},$$

где $t_{\alpha/2}$ определяется, как и в предыдущем случае, при проверке случайности, то гипотеза о нормальности принимается.

Примечание: При выявлении аномальных измерений последние фиксируются в файле результатов, *но не исключаются из входного файла*. Поэтому следует помнить, что, для того чтобы эти измерения в дальнейшем не влияли на результаты анализа, Вы должны сами удалить указанные аномальные наблюдения из входного файла.

Ниже приведен результат решения этих задач с исходными данными из прилагаемого файла *act.txt*:

Файл ввода act.txt
В исходной выборке аномальных наблюдений нет
"Анализ выборки на случайность"
Уровень значимости alfa = 0.0500
Гипотеза о случайности выборки ПРИНИМАЕТСЯ
"Анализ выборки на нормальность"
Уровень значимости alfa = 0.0500
Гипотеза о нормальности выборки ПРИНИМАЕТСЯ

Известно, что задачи и методы корреляционного анализа опираются на многомерное нормальное распределение. И очевидно, что если гипотеза о нормальности наблюденной выборки не отвергается и выборка случайна, то дальнейшие выводы статистического анализа будут наиболее достоверными.

В то же время решение задач корреляционого анализа принесет несомненную пользу и в том случае, когда гипотеза о нормальности исходной выборки отвергается.

## 6.2. Вычисление параметров

В данном случае могут решаться 2 задачи. Первая - оценка параметров многомерного нормального распределения: вектора математических ожиданий $M = E[X]$, где $X$ - $m$-мерная случайная величина, ковариационной матрицы $\Sigma = D[X] = E[(X-M)(X-M)^{\mathrm{T}}]$ и матрицы коэффициентов парных корреляций с элементами

$$r_{ij} = \frac{E[(x_i - m_i)(x_j - m_j)]}{\sqrt{\sigma_{ii}\sigma_{jj}}},$$

где $m_i$ - математическое ожидание, а $\sigma_{ii} = \sigma_i^2$ - диагональный элемент матрицы $\Sigma$, дисперсия случайной компоненты $x_i$.

Оценка максимального правдоподобия (ОМП) математического ожидания многомерной случайной величины $X$ вычисляется по формуле

$$\hat{M} = \frac{1}{n}\sum_{i=1}^{n} X_i,$$

где $X_1, X_2, \ldots, X_n$ - выборка $m$-мерной случайной величины $X$ объёмом $n$. Несмещенная ОМП ковариационной матрицы $\Sigma$ определяется соотношением

$$\hat{\Sigma} = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \hat{M}\right)\left(X_i - \hat{M}\right)^{\mathrm{T}}.$$

Элементы матрицы парных коэффициентов корреляции оцениваются в соответствии с выражением

$$\hat{r}_{ij} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}},$$

где $\hat{\sigma}_{ij}$ - элементы матрицы $\hat{\Sigma}$.

Результаты оценивания параметров по выборке, содержащейся в файле *act.txt*, имеют вид:

Файл ввода act.txt

Вектор математического ожидания

| N | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 9.603333e+01 | 1.151667e+02 | 5.070000e+01 | 2.093333e+01 |

Ковариационная матрица

| N | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.0911510e+03 | 1.063689e+03 | 3.008068e+02 | -2.218081e+00 |
| 2 | 1.063689e+03 | 1.185294e+03 | 3.135423e+02 | 7.231637e+00 |
| 3 | 3.008068e+02 | 3.135423e+02 | 1.423831e+02 | -4.986442e+00 |
| 4 | -2.218081e+00 | 7.231637e+00 | -4.986442e+00 | 6.3351410e+01 |

Матрица парных коэффициентов корреляции

| N | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.000000e+00 | 9.353179e-01 | 7.631606e-01 | -8.436385e-03 |
| 2 | 9.353179e-01 | 1.000000e+00 | 7.632276e-01 | 2.639037e-02 |

| | | | | |
|---|---|---|---|---|
| 3 | 7.631606e-01 | 7.632276e-01 | 1.000000e+00 | -5.250291e-02 |
| 4 | -8.436385e-03 | 2.639037e-02 | -5.250291e-02 | 1.000000e+00 |

Этот же пункт меню предусматривает проверку гипотез о равенстве математического ожидания некоторому известному вектору, т.е. проверку гипотез вида: $M = M_0$.

Такая задача очень часто возникает на практике, когда, например, на основании наблюдений некоторого многомерного показателя технологического процесса желают убедиться, что эти показатели равны номинальному значению, т.е. процесс протекает нормально, а отклонения наблюденных значений от номинальных объясняются лишь ошибками наблюдений (измерений).

При решении этой задачи возможны две ситуации: ковариационная матрица $\Sigma$ может быть известна из ранее проводимых экспериментов, по другим выборкам (в этом случае пользователь в процессе диалога должен будет ввести элементы этой матрицы), или неизвестна, тогда в процессе вычислений она должна быть оценена.

В случае проверки гипотезы $H_0$: $M = M_0$ при *известной ковариационной матрице* вычисляется статистика

$$X_m^2 = n\left(\hat{M} - M_0\right)^{\mathrm{T}} \Sigma^{-1}\left(\hat{M} - M_0\right).$$

При справедливой гипотезе $H_0$ эта статистика имеет $\chi^2$-распределение с $m$ степенями свободы. Таким образом, гипотеза $H_0$ принимается, если

$$X_m^2 < \chi_{m,\alpha}^2,$$

где $\alpha$ - уровень значимости и

$$1 - \alpha = P\left\{X_m^2 > \chi_{m,\alpha}^2\right\} = \frac{1}{2^{m/2}\Gamma(m/2)} \int_0^{\chi_{m,\alpha}^2} s^{m/2-1} e^{-s/2} ds.$$

В случае проверки гипотезы $H_0$: $M = M_0$ при *неизвестной ковариационной матрице* вычисляемая статистика имеет вид

$$T^2 = \frac{mn}{(n-1)(n-m)}\left(\hat{M} - M_0\right)^{\mathrm{T}} \hat{\Sigma}^{-1}\left(\hat{M} - M_0\right).$$

Если гипотеза $H_0$ справедлива, то эта статистика подчиняется $F$-распределению Фишера с $m$ и $n-m$ степенями свободы. В данном случае гипотеза $H_0$ принимается, если выполняется условие

$$T^2 < F_{m,n-m,\alpha}.$$

Величина $F_{m,n-m,\alpha}$ определяется из равенства

$$1-\alpha = P\left\{T^2 > F_{m,k,\alpha}\right\} = \left(\frac{m}{k}\right)^{m/2} \frac{\Gamma\left(\frac{m+k}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{k}{2}\right)} \int_0^{F_{m,k,\alpha}} z^{m/2-1}\left(1+\frac{m}{k}z\right)^{-(m+k)/2} dz,$$

где $k = n - m$.

В случае неизвестной матрицы по данным из файла *act.txt* имеем следующий результат:

Файл ввода act.txt

С уровнем значимости alfa = 0.0500 ОТВЕРГАЕТСЯ гипотеза: МАТЕМАТИЧЕСКОЕ ОЖИДАНИЕ равно:

| N | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 9.500000e+01 | 1.000000e+02 | 3.500000e+01 | 1.500000e+01 |

## 6.3. Комплексный анализ

Этот пункт меню может выполняться либо с проверкой предпосылок, т.е. анализом на наличие грубых ошибок измерений, на случайность и нормальность, либо без. Если задание выполняется с проверкой предпосылок и в выборке обнаруживаются аномальные измерения, то в дальнейшем анализе они не участвуют (но в исходном файле сохраняются!).

При выполнении этого задания вычисляются оценки математического ожидания, ковариационной матрицы, матрицы парных корреляций, проверяются гипотезы о значимости элементов этой матрицы, для значимых парных корреляций строятся доверительные интервалы. Далее вычисляются оценки корреляционных отношений и проверяются гипотезы о их значимости, т.е. гипотезы вида $\rho_{ij}^2 = 0$. Затем вычисляется оценка матрицы частных корреляций, элементы которой представляют собой корреляцию между соответствующими элементами, при условии, что все остальные $m-2$ компоненты $m$-мерной случайной величины фиксированы (т.е. рассматривается условное распределение 2 компонент). Все элементы этой матрицы проверяются на значи-

мость, для значимых частных коэффициентов корреляции строятся доверительные интервалы.

Вычисляется вектор множественных коэффициентов корреляции, каждый элемент которого представляет собой корреляцию (меру линейной зависимости) между соответствующей компонентой случайной величины и множеством остальных $m-1$ компонент. Относительно каждого элемента этого вектора проверяется гипотеза о значимости.

Результаты выполнения этого пункта меню без проверки предпосылок для рассматриваемого примера имеют вид:

Файл ввода act.txt

Вектор математического ожидания

| N | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 9.603333e+01 | 1.151667e+02 | 5.070000e+01 | 2.093333e+01 |

Ковариационная матрица

| N | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.0911510e+03 | 1.063689e+03 | 3.008068e+02 | -2.218081e+00 |
| 2 | 1.063689e+03 | 1.185294e+03 | 3.135423e+02 | 7.231637e+00 |
| 3 | 3.008068e+02 | 3.135423e+02 | 1.423831e+02 | -4.986442e+00 |
| 4 | -2.218081e+00 | 7.231637e+00 | -4.986442e+00 | 6.3351410e+01 |

Матрица парных коэффициентов корреляции

| N | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.000000e+00 | 9.353179e-01 | 7.631606e-01 | -8.436385e-03 |
| 2 | 9.353179e-01 | 1.000000e+00 | 7.632276e-01 | 2.639037e-02 |
| 3 | 7.631606e-01 | 7.632276e-01 | 1.000000e+00 | -5.250291e-02 |
| 4 | -8.436385e-03 | 2.639037e-02 | -5.250291e-02 | 1.000000e+00 |

Матрица значимых коэффициентов корреляции

Уровень значимости alfa = 0.0500

| N | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.000000e+00 | 9.353179e-01 | 7.631606e-01 | 0.000000e+00 |

| | | | | |
|---|---|---|---|---|
| 2 | 9.353179e-01 | 1.000000e+00 | 7.632276e-01 | 0.000000e+00 |
| 3 | 7.631606e-01 | 7.632276e-01 | 1.000000e+00 | 0.000000e+00 |
| 4 | 0.000000e+00 | 0.000000e+00 | 0.000000e+00 | 1.000000e+00 |

Доверительные интервалы для значимых парных коэф-в корреляции
Уровень значимости alfa =0.0500

| Коэффициент корреляции | Доверительный интервал |
|---|---|
| r( 1, 2) = 0.9353179 | 0.8936307 0.9610038 |
| r( 1, 3) = 0.7631606 | 0.6316382 0.8519835 |
| r( 2, 3) = 0.7632276 | 0.6317346 0.8520275 |

Матрица корреляционных отношений

| N | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.000000e+00 | 8.884497e-01 | 7.039601e-01 | 2.831308e-01 |
| 2 | 9.007756e-01 | 1.000000e+00 | 7.056496e-01 | 3.454498e-01 |
| 3 | 6.026155e-01 | 6.750155e-01 | 1.000000e+00 | 3.260075e-01 |
| 4 | 8.177995e-02 | 7.801921e-02 | 1.972346e-01 | 1.000000e+00 |

Матрица значимых корреляционных отношений
Уровень значимости alfa = 0.0500

| N | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.000000e+00 | 8.884497e-01 | 7.039601e-01 | 0.000000e+00 |
| 2 | 9.007756e-01 | 1.000000e+00 | 7.056496e-01 | 3.454498e-01 |
| 3 | 6.026155e-01 | 6.750155e-01 | 1.000000e+00 | 3.260075e-01 |
| 4 | 0.000000e+00 | 0.000000e+00 | 0.000000e+00 | 1.000000e+00 |

Матрица значимых корней кв. из корреляционных отношений

| N | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.000000e+00 | 9.425761e-01 | 8.390233e-01 | 0.000000e+00 |
| 2 | 9.490920e-01 | 1.000000e+00 | 8.400295e-01 | 5.877498e-01 |
| 3 | 7.762831e-01 | 8.215933e-01 | 1.000000e+00 | 5.709707e-01 |
| 4 | 0.000000e+00 | 0.000000e+00 | 0.000000e+00 | 1.000000e+00 |

Матрица частных коэффициентов корреляции

| N | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.000000e+00 | 8.455458e-01 | 2.073441e-01 | -7.151330e-02 |
| 2 | 8.455458e-01 | 1.000000e+00 | 2.247293e-01 | 1.153210e-01 |
| 3 | 2.073441e-01 | 2.247293e-01 | 1.000000e+00 | -9.491602e-02 |
| 4 | -7.151330e-02 | 1.153210e-01 | -9.491602e-02 | 1.000000e+00 |

Матрица значимых частных коэффициентов корреляции
Уровень значимости alfa = 0.0500

| N | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.000000e+00 | 8.455458e-01 | 0.000000e+00 | 0.000000e+00 |
| 2 | 8.455458e-01 | 1.000000e+00 | 0.000000e+00 | 0.000000e+00 |
| 3 | 0.000000e+00 | 0.000000e+00 | 1.000000e+00 | 0.000000e+00 |
| 4 | 0.000000e+00 | 0.000000e+00 | 0.000000e+00 | 1.000000e+00 |

Доверительные интервалы для
Значимых частных коэффициентов корреляции
Уровень значимости alfa = 0.0500

| Коэффициент корреляции | Доверительный интервал |
|---|---|
| r( 1, 2) =0.8455458 | 0.7513443 0.9059756 |

Вектор множественных коэффициентов корреляции

| N | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 9.387504e-01 | 9.392858e-01 | 7.781507e-01 | 1.355840e-01 |

Вектор значимых множественных коэффициентов корреляции
Уровень значимости alfa = 0.0500

| N | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 9.387504e-01 | 9.392858e-01 | 7.781507e-01 | 0.000000e+00 |

## 6.4. Парная корреляция

В данном пункте может анализироваться либо вся матрица парных корреляций (оценивание, проверка значимости и построение довери-

тельных интервалов для значимых), либо отдельный коэффициент парной корреляции. В последнем случае можно раздельно или совместно решать следующие задачи: определение оценки парного коэффициента корреляции, проверка гипотезы о его значимости (гипотезы вида: $r_{ij} = 0$), проверка при необходимости гипотезы вида $r_{ij} = r_0$ (т.е. равенства его определенному конкретному значению), а также построение для него доверительного интервала.

Если оценка ковариационной матрицы $\hat{\Sigma}$ уже известна, то оценка парного коэффициента корреляции может быть найдена в соответствии с выражением

$$\hat{r}_{ij} = \frac{\hat{\sigma}_{ij}}{\sqrt{\hat{\sigma}_{ii}\hat{\sigma}_{jj}}},$$

где $\hat{\sigma}_{ij}$ - элемент матрицы $\hat{\Sigma}$.

При проверке гипотезы $H_0: r_{ij} = 0$ вычисляется статистика

$$t = \frac{\sqrt{n-2}\left|\hat{r}_{ij}\right|}{\sqrt{\hat{r}_{ij}^2}},$$

которая при справедливой гипотезе $H_0$ имеет $t$-распределение Стьюдента с $n-2$ степенями свободы. При конкурирующей гипотезе $H_1: r_{ij} \neq 0$ гипотеза $H_0$ принимается, если

$$t < t_{n-2,\alpha/2},$$

где $\alpha$ - уровень значимости. Величина $t_{n-2,\alpha/2}$ при $k = n-2$ определяется равенством

$$1 - \alpha = P\left\{t < t_{k,\alpha/2}\right\} = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\frac{k}{2}\right)} \int_{-t_{k,\alpha/2}}^{t_{k,\alpha/2}} \left(1 + \frac{z^2}{k}\right)^{-\frac{k+1}{2}} dz.$$

При проверке гипотезы $H_0: r_{ij} = r_0$ статистика

$$z_0 = \sqrt{n-3}\left|\frac{1}{2}\ln\frac{1+\hat{r}_{ij}}{1-\hat{r}_{ij}} - \frac{1}{2}\ln\frac{1+r_0}{1-r_0} - \frac{r_0}{2(n-1)}\right|,$$

при справедливой гипотезе $H_0$ подчиняется стандартному нормальному распределению. Гипотеза $H_0$ принимается, если

$$z < t_{\alpha/2},$$

где $t_{\alpha/2}$ - квантиль стандартного нормального распределения и

$$1-\alpha = P\{t < t_{\alpha/2}\} = \frac{1}{\sqrt{2\pi}} \int_{-t_{\alpha/2}}^{t_{\alpha/2}} e^{-t^2/2} dt.$$

Доверительный интервал для парного коэффициента корреляции определяется неравенством

$$th\left(z_0 - \frac{t_{\alpha/2}}{\sqrt{n-3}}\right) < r_{ij} < th\left(z_0 + \frac{t_{\alpha/2}}{\sqrt{n-3}}\right).$$

Ниже представлены результаты проверки гипотезы о значимости, т.е. гипотезы вида $r_{12} = 0$, и гипотезы вида $r_{12} = 0.5$ для выборки из файла *act.txt*:

Файл ввода act.txt
Парный коэффициент корреляции r( 1, 2)=0.935318
ЗНАЧИМ с уровнем значимости alfa = 0.0500
Гипотеза: r=0.50000 ОТКЛОНЯЕТСЯ при alfa = 0.0500
Гипотеза будет принята при alfa < 2.93e-05
Доверительный интервал = (0.893631,0.961004)

Зачастую хотелось бы проверить гипотезу вида $r_{12} = 1$. А сделать этого нельзя, так как в вычисляемой статистике оказывается 0 в знаменателе. Но в то же время мы можем проверить, например, гипотезу вида $r_{12} = 0.99$, т.е. сравнить с величиной, близкой к 1. Выдача результата в таком случае для нашей выборки имеет вид:

Файл ввода act.txt
Оценка парного коэф. корреляции r( 1, 2)=0.93532
Гипотеза: r=0.99000 ОТКЛОНЯЕТСЯ при alfa = 0.0500
Гипотеза будет принята при alfa < 5.36e-13

## 6.5. Корреляционное отношение

Корреляционное отношение случайной величины $x_i$ по $x_j$ определяется отношением дисперсии условного математического ожидания $x_i$ к дисперсии $x_i$:

$$\rho_{ij}^2 = D\left\{E\left[x_i \middle| x_j\right]\right\} \Big/ D\left[x_i\right].$$

Соотношение между коэффициентом корреляции $r_{ij}^2$ и корреляционным отношением $\rho_{ij}^2$ позволяет сделать следующие выводы:

а) $r_{ij}^2 = 0$, если $x_i$ и $x_j$ независимы;

б) $r_{ij}^2 = \rho_{ij}^2 = 1$, тогда и только тогда, когда имеется строгая линейная функциональная зависимость $x_i$ от $x_j$;

в) $r_{ij}^2 < \rho_{ij}^2 = 1$, тогда и только тогда, когда имеется строгая нелинейная функциональная зависимость $x_i$ от $x_j$;

г) $r_{ij}^2 = \rho_{ij}^2 < 1$, тогда и только тогда, когда регрессия $x_i$ по $x_j$ строго линейна, но нет функциональной зависимости;

д) $r_{ij}^2 < \rho_{ij}^2 < 1$, указывает на то, что не существует функциональной зависимости и некоторая нелинейная кривая регрессии "подходит" лучше, чем "наилучшая" прямая линия.

То есть равенство квадрата коэффициента корреляции корреляционному отношению указывает на то, что для регрессии нельзя найти лучшей кривой, чем прямая линия.

Оценка корреляционного отношения определяется выражением

$$\hat{\rho}_{ij}^2 = \frac{\sum_{l=1}^{k} n_l \left(\bar{x}_l^i - \bar{x}^i\right)^2}{\sum_{l=1}^{k}\sum_{s=1}^{n}\left(x_{ls}^i - \bar{x}^i\right)^2},$$

где $k$ - количество интервалов значений (сечений) для компоненты $x_i$, $\bar{x}_l^i$ - среднее значение $l$-го $x_i$-сечения, $n_l$ - число наблюдений в этом сечении, $x_{ls}^i$ - значение компоненты $x_i$ с номером $s$ в $l$-м $x_i$-сечении, $\sum_{l=1}^{k} n_l = n$.

При проверке гипотезы $H_0: \rho_{ij} = 0$ используется статистика

$$F = \frac{(n-k)\hat{\rho}_{ij}^2}{(k-1)\left(1-\hat{\rho}_{ij}^2\right)},$$

которая при справедливой гипотезе $H_0$ имеет $F$-распределение Фишера с числом степеней свободы $k-1$ и $n-k$. Гипотеза $H_0$ принимается, если

$$F < F_{k-1,n-k,\alpha},$$

где $F_{k-1,n-k,\alpha}$ - критическая точка критерия с уровнем значимости $\alpha$.

При проверке гипотезы $H_0: \rho_{ij}^2 = r_{ij}^2$ (гипотезы о линейности регрессии $x_i$ по $x_j$) статистика имеет вид

$$F = \frac{(n-k)\left(\hat{\rho}_{ij}^2 - \hat{\rho}_{ij}^2\right)}{(k-2)\left(1-\hat{\rho}_{ij}^2\right)}.$$

При справедливой гипотезе $H_0$ она подчиняется $F$-распределению Фишера с числом степеней свободы $k-2$ и $n-k$. Гипотеза $H_0$ принимается, если

$$F < F_{k-2,n-k,\alpha}.$$

Данный пункт меню предусматривает либо анализ матрицы корреляционных отношений (оценивание и проверку значимости элементов матрицы), либо анализ определенного корреляционного отношения между компонентами с номерами i и j случайной величины: вычисление оценки $\rho_{ij}^2$, проверку значимости (гипотеза: $\rho_{ij}^2 = 0$), проверку линейности регрессии $x_i$ по $x_j$ (проверяется гипотеза о равенстве корреляционного отношения квадрату парного коэффициента корреляции).

Результат анализа выборки из файла *act.txt* в последнем случае имеет вид:

Файл ввода act.txt
Корреляционное отношение ro( 1, 2)=0.888450
ЗНАЧИМО с уровнем значимости alfa = 0.0500
Коэффициент корреляции r ( 1, 2)=0.935318
С уровнем значимости alfa = 0.0500
ПРИНИМАЕТСЯ гипотеза: ro=r*r
(наилучшая регрессия x( 1) по x( 2) - линейная) Гипотеза: ro=r*r
будет отвергнута при alfa > 8.17e-01

## 6.6. Частная корреляция

В случае двух нормальных или почти нормальных случайных величин коэффициент корреляции между ними может быть использован в качестве меры взаимозависимости. На практике при интерпретации "взаимозависимости" приходится сталкиваться с трудностями следу-

ющего характера: если одна величина коррелирована с другой, то это может быть всего лишь отражением того факта, что они обе коррелированы с некоторой третьей величиной или совокупностью величин. Указанная возможность приводит к необходимости рассмотрения условных корреляций между двумя величинами при фиксированных значениях остальных величин, т.е. частных корреляций.

Если корреляция между двумя величинами уменьшается, когда мы фиксируем некоторую другую случайную величину (компоненту многомерной величины), то это означает, что их взаимозависимость возникает частично через воздействие этой величины; если же частная корреляция равна нулю или очень мала, то мы делаем вывод, что их взаимозависимость целиком обусловлена этим воздействием. Наоборот, когда частная корреляция больше первоначальной корреляции между двумя величинами, то мы заключаем, что другие величины ослабляли эту связь, или, можно сказать, "маскировали" корреляцию.

Представим случайный вектор $X$ в следующем виде:

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix},$$

где $X_1 = (x_1, x_2, \ldots, x_l)^{\mathrm{T}}$, $X_2 = (x_{l+1}, x_{l+2}, \ldots, x_m)^{\mathrm{T}}$, соответственно вектор математических ожиданий и ковариационную матрицу

$$M = \begin{bmatrix} M_1 \\ M_2 \end{bmatrix}, \quad \Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix}.$$

Если случайный вектор $X$ подчиняется нормальному закону распределения с вектором средних $M$ и ковариационной матрицей $\Sigma$, то условное распределение подвектора $X_1$ при известном $X_2$ является нормальным с математическим ожиданием $M_1 + B(X_2 - M_2)$ и ковариационной матрицей $\Sigma_{11\cdot 2}$, где $B = \Sigma_{12}\Sigma_{22}^{-1}$, $\Sigma_{11\cdot 2} = \Sigma_{11} - \Sigma_{12}\Sigma_{22}^{-1}\Sigma_{21}$.

ОМП для частного коэффициента корреляции определяется следующим соотношением:

$$\hat{r}_{ij\cdot l+1,\ldots,m} = \frac{\hat{\sigma}_{ij\cdot l+1,\ldots,m}}{\sqrt{\hat{\sigma}_{ii\cdot l+1,\ldots,m}\hat{\sigma}_{jj\cdot l+1,\ldots,m}}},$$

где $\hat{\sigma}_{ij\cdot l+1,\ldots,m}$ - элемент $i$-й строки и $j$-го столбца матрицы $\Sigma_{11\cdot 2}$. В данном случае при оценке взаимозависимости между компонентами $x_i$

и $x_j$ случайной величины $X$ исключается влияние компонент $x_{l+1}, x_{l+1}, \ldots, x_m$.

При проверке гипотез вида $H_0: r_{ij \cdot l+1,\ldots,m} = 0$ и $H_0: r_{ij \cdot l+1,\ldots,m} = r_0$, а также при построении доверительных интервалов для частного коэффициента корреляции используются те же самые статистики, что и для парного коэффициента корреляции. В этом случае в соответствующих соотношениях $n$ заменяется на $n-(m-l)$, где $n$ - объём выборки, $m$ - размерность случайного вектора, $l$ - число компонент в условном распределении (в частном случае $l=2$).

В этом пункте меню может анализироваться или вся матрица частных корреляций (оценивание, проверка значимости и построение доверительных интервалов для значимых коэффициентов при условии, что остальные $m-2$ компоненты $m$-мерной величины фиксированы), или отдельный коэффициент частной корреляции. В последнем случае можно, решая задачи отдельно или одновременно, найти оценку частного коэффициента корреляции, проверить гипотезу о его значимости, при необходимости проверить гипотезу вида $r_{ij} = r_0$, а также построить доверительный интервал для частного коэффициента корреляции.

Для выборки из файла *act.txt* результаты имеют следующий вид:

Файл ввода act.txt
Частный коэффициент корреляции r( 1, 2) = 0.845546
при фиксированных компонентах с индексами: 3, 4,
ЗНАЧИМ с уровнем значимости alfa = 0.0500
Гипотеза: частный коэф-т корреляции r( 1, 2)=0.700000
ОТВЕРГАЕТСЯ с уровнем значимости alfa = 0.0500
Гипотеза будет принята при alfa < 6.51e-03
Доверительный интервал =(0.751344,0.905976)

В предыдущих параграфах было получено, что парный коэффициент корреляции $r_{12} = 0.935318$, а вычисленный в данном пункте частный коэффициент корреляции при исключении влияния на корреляцию компонент $x_3$ и $x_4$ равен 0.845546. Т.е. связь $x_1$ с $x_2$

частично осуществляется через $x_3$ и $x_4$. Продолжив анализ и вычислив коэффициент корреляции при исключении влияния только $x_3$ и только $x_4$, получим соответственно 0.845086 и 0.935900. Отсюда можно сделать вывод, что связь $x_1$ с $x_2$ осуществляется частично через $x_3$, но компонента $x_4$ на эту связь практически не влияет. Кроме того, если в дальнейшем у Вас возникнет необходимость в построении регрессионной модели, то, если рассматривать только регрессию $x_1$ по $x_2$, то наилучшая модель линейная, а если включать в нее $x_3$, то, возможно, нелинейная модель окажется предпочтительней. Что касается компоненты $x_3$, то ее влияние не значимо, и, следовательно, присутствие в регрессионной модели будет излишним.

## 6.7. Частная линейная регрессия

На основе частных коэффициентов корреляции можно найти оценки параметров частных линейных регрессионных зависимостей. Каждая $i$-я строка вычисляемой матрицы представляет собой линейную регрессионную модель вида

$$E\left[x_i \middle| x_1,\ldots,x_m\right] = b_{i1}x_1+\ldots+b_{ij}x_j+\ldots+b_{im}x_m,$$

где в правой части равенства $b_{ij}$ - элементы $i$-й строки матрицы коэффициентов $(i \neq j)$, $E\left[x_i \middle| x_1,\ldots,x_m\right]$ - условное математическое ожидание случайной величины $x_i$.

Соответствующее уравнение регрессии имеет вид

$$x_i = b_{i1}x_1+\ldots+b_{ij}x_j+\ldots+b_{im}x_m.$$

Правомерность использования линейной регрессионной модели, описывающей связь между компонентами случайной величины, выявляется в результате анализа соответствующих парных и частных коэффициентов корреляции и корреляционных отношений.

Для выборки из файла *act.txt* результаты имеют следующий вид:

Файл ввода act.txt

Матрица частных коэффициентов линейной регрессии

| N | 1 | 2 | 3 | 4 |
|---|---|---|---|---|
| 1 | 1.000000e+00 | 8.147302e-01 | 3.149224e-01 | -1.032269e-01 |

| | | | | |
|---|---|---|---|---|
| 2 | 8.775273e-01 | 1.000000e+00 | 3.542393e-01 | 1.727579e-01 |
| 3 | 1.365143e-01 | 1.425688e-01 | 1.000000e+00 | -9.020554e-02 |
| 4 | -4.954284e-02 | 7.698020e-02 | -9.987264e-02 | 1.000000e+00 |

Т.е. линейная регрессия, например, для $x_1$ будет иметь вид

$$x_1 = 0.8147302x_2 + 0.3149224x_3 - 0.1032269x_4.$$

## 6.8. Множественная корреляция

Множественный коэффициент корреляции является мерой зависимости компоненты случайной величины от некоторого множества компонент.

Можно рассматривать корреляцию между одной компонентой случайного вектора и множеством всех остальных или каким-то подмножеством.

Следует отметить, что множественный коэффициент корреляции $r_i$ случайной величины $x_i$ относительно некоторого множества других случайных величин всегда не меньше, чем абсолютная величина любого парного коэффициента корреляции $r_{ij}$ с таким же первичным индексом. Более того, множественный коэффициент корреляции никогда нельзя уменьшить путем расширения множества величин, относительно которых измеряется зависимость $x_i$.

Если коэффициент корреляции между $x_i$ и множеством всех остальных компонент многомерной случайной величины равен нулю ($r_i = 0$), то все коэффициенты корреляции этой величины относительно любого подмножества также равны 0, т.е. величина $x_i$ полностью некоррелирована со всеми остальными величинами.

С другой стороны, если $r_i$ относительно множества всех остальных компонент равен единице ($r_i = 1$), то, по крайней мере, один из коэффициентов корреляции относительно некоторого подмножества компонент должен быть равен 1.

Надо отметить, что коэффициент корреляции, например, между $x_1$ и множеством всех остальных компонент является обычным

коэффициентом корреляции между $x_1$ и условным математическим ожиданием $E[x_1|x_2, x_3, \ldots, x_n]$.

Если представим случайный вектор $X$ в виде

$$X = \begin{bmatrix} X_1 \\ X_2 \end{bmatrix},$$

где $X_1 = (x_1, x_2, \ldots, x_l)^{\mathrm{T}}$, $X_2 = (x_{l+1}, x_{l+2}, \ldots, x_m)^{\mathrm{T}}$, и соответственно ковариационную матрицу

$$\Sigma = \begin{bmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{bmatrix},$$

тогда ОМП множественного коэффициента корреляции между $x_i$, $i \le l$ и множеством компонент $x_{l+1}, x_{l+2}, \ldots, x_m$ определится соотношением

$$\hat{r}_{i\cdot l+1,\ldots,m} = \sqrt{\frac{\hat{\sigma}_{(i)}\hat{\Sigma}_{22}^{-1}\hat{\sigma}_{(i)}^{\mathrm{T}}}{\hat{\sigma}_{ii}}},$$

где $\hat{\sigma}_{ii}$ - элемент матрицы $\hat{\Sigma}$, а $\hat{\sigma}_{(i)}$ - $i$-я строка матрицы $\hat{\Sigma}_{12}$.

При проверке гипотезы $H_0: r_{i\cdot l+1,\ldots,m} = 0$ используется статистика

$$F = \frac{n-m+l-1}{m-l}\frac{\hat{r}_{ij}^2}{1-\hat{r}_{ij}^2},$$

которая при справедливой гипотезе $H_0$ имеет $F$-распределение Фишера с числом степеней свободы $m-l$ и $n-m+l-1$. Гипотеза $H_0$ принимается, если

$$F < F_{m-l,n-m+l-1,\alpha},$$

где $\alpha$ - уровень значимости, а $F_{m-l,n-m+l-1,\alpha}$ - критическая точка критерия с уровнем значимости $\alpha$.

В данном пункте меню можно анализировать или вектор множественных коэффициентов корреляции (относительно множества всех остальных величин, образующих многомерную величину), находить оценки и проверять значимость компонент этого вектора, или анализировать множественный коэффициент корреляции между определенной компонентой и задаваемым подмножеством остальных компонент.

В последнем случае для выборки из файла *act.txt* получается результат:

Файл ввода act.txt

Множественный коэффициент корреляции между 1 компонентой и множеством компонент с индексами: 3, 4,

r = 0.763818

ЗНАЧИМ с уровнем значимости alfa = 0.0500

Для сравнения: парный коэффициент корреляции $r_{13} = 0.7631606$, а $r_{14} = -0.0843685$, и по абсолютной величине меньше вычисленного множественного коэффициента корреляции.

## 6.9. Покомпонентный анализ

Этот пункт предполагает решение задач проверки гипотез двух видов для каждой из компонент многомерной случайной величины:

- проверку гипотез о равенстве математических ожиданий компонент определенным значениям, причем стандартные отклонения для этих компонент могут быть либо известны, либо неизвестны (гипотезы: $m_i = m_{0i}$);
- проверку гипотез о равенстве стандартных отклонений известным величинам, и, в свою очередь, соответствующие математические ожидания могут быть известны или нет (гипотезы: $\sigma_i = \sigma_{0i}$).

Следует подчеркнуть, что, строго говоря, проведение покомпонентного анализа оправдано при некоррелированности величин, входящих в систему.

Проверка гипотезы $H_0: m_i = m_{0i}$ при *известной* дисперсии компоненты $\sigma_i^2$. Вычисляемая статистика

$$T = \frac{\sqrt{n}\left(\bar{x}_i - m_0\right)}{\sigma_i},$$

где $\bar{x}_i$ - среднее значение компоненты $x_i$, при справедливой гипотезе $H_0$ подчиняется стандартному нормальному распределению. Гипотеза $H_0$ принимается, если

$$|T| \le t_{\alpha/2},$$

где критическое значение $t_{\alpha/2}$ при заданном $\alpha$ определяется равенством

$$1-\alpha = P\{t < t_{\alpha/2}\} = \frac{1}{\sqrt{2\pi}} \int_{-t_{\alpha/2}}^{t_{\alpha/2}} e^{-t^2/2} dt.$$

При проверке аналогичной гипотезы при *неизвестной* дисперсии компоненты $\sigma_i^2$ вычисляемая статистика

$$T = \frac{\sqrt{n}(\overline{x}_i - m_0)}{s_i},$$

где $s = \sqrt{\sum_{j=1}^{n} (x_{ij} - \overline{x}_i)^2 / (n-1)}$ и $x_{ij}$ - $j$-е наблюдение компоненты $x_i$, при справедливой гипотезе $H_0$ имеет $t$-распределение Стьюдента с $n$ степенью свободы. Гипотеза $H_0$ принимается, если

$$|T| \le t_{n-1,\alpha/2},$$

где критическое значение $t_{n-1,\alpha/2}$ при заданном $\alpha$ и $k = n-1$ определяется равенством

$$1-\alpha = P\{t < t_{k,\alpha/2}\} = \frac{\Gamma\left(\frac{k+1}{2}\right)}{\sqrt{k\pi}\,\Gamma\left(\frac{k}{2}\right)} \int_{-t_{k,\alpha/2}}^{t_{k,\alpha/2}} \left(1 + \frac{z^2}{k}\right)^{-\frac{k+1}{2}} dz.$$

Проверка гипотезы $H_0: \sigma_i^2 = \sigma_0^2$ при *известном математическом ожидании компоненты* $m_i$. В этом случае статистика

$$\chi^2 = \sum_{j=1}^{n} \frac{(x_{ij} - m_0)^2}{\sigma_i^2}$$

имеет $\chi^2$-распределение с $n$ степенями свободы, если гипотеза $H_0$ справедлива. Гипотеза $H_0$ принимается, если

$$\chi^2_{n,\alpha/2} \le \chi^2 \le \chi^2_{n,1-\alpha/2},$$

где значения $\chi^2_{n,\alpha/2}$ и $\chi^2_{n,1-\alpha/2}$ определяются равенством

$$\alpha/2 = \int_0^{\chi^2_{n,\alpha/2}} f(y)dy = \int_{\chi^2_{n,1-\alpha/2}}^{\infty} f(y)dy,$$

и $f(y) = \frac{1}{2^{n/2}\Gamma(n/2)} y^{n/2-1} e^{-y/2}$ - функция плотности $\chi^2$-распределения с $n$ степенями свободы.

Если *математическое ожидание компоненты неизвестно*, то при проверке данной гипотезы вычисляемая статистика имеет вид

$$\chi^2 = \sum_{j=1}^{n} \frac{\left(x_{ij} - \overline{x}_i\right)^2}{\sigma_i^2}$$

и при справедливой гипотезе $H_0$ имеет $\chi^2$-распределение с $n-1$ степенями свободы. Решение о принятии или отклонении гипотезы $H_0$ осуществляется, как и в предыдущем случае.

В случае проверки гипотез относительно математических ожиданий при неизвестных стандартных отклонениях для выборки из *act.txt* результаты имеют вид:

Файл ввода act.txt
"Покомпонентный анализ"
Проверка гипотезы о значимости отклонения математического ожидания от заданного значения при известном векторе стандартных отклонений.
Уровень значимости alfa = 0.0500

| N компоненты | Заданное значение мат.ожидания | Вычисленное значение мат.ожидания | Значимость отклонения |
|---|---|---|---|
| 1 | 9.600000e+01 | 9.603333e+01 | Незначимо |
| 2 | 1.150000e+02 | 1.151667e+02 | Незначимо |
| 3 | 5.000000e+01 | 5.070000e+01 | Незначимо |
| 4 | 2.000000e+01 | 2.093333e+01 | Незначимо |

## 6.10. Статистические распределения

В данном пункте меню предоставляется возможность вычисления вероятностей или квантилей для распределений, наиболее часто используемых в задачах статистического анализа:

- стандартного нормального распределения;
- $\chi^2$-распределения;
- $t$-распределения Стьюдента;
- $F$-распределения Фишера;
- $B$-распределения (бета).

Функция плотности нормального распределения имеет вид

$$f(x)=\frac{2}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}},$$

$\chi^2$-распределения с числом степеней свободы $n$ -

$$f(x)=\frac{1}{2^{n/2}\Gamma(n/2)}x^{n/2-1}e^{-x/2},$$

$t$-распределения Стьюдента с числом степеней свободы $n$ -

$$f(x)=\frac{\Gamma\left(\frac{n+1}{2}\right)}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}\left(1+\frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

$F$-распределения Фишера с числом степеней свободы $m$ и $n$

$$f(x)=\left(\frac{m}{n}\right)^{m/2}\frac{\Gamma\left(\frac{m+n}{2}\right)}{\Gamma\left(\frac{m}{2}\right)\Gamma\left(\frac{n}{2}\right)}x^{m/2-1}\left(1+\frac{m}{n}x\right)^{-(m+n)/2},$$

$B$-распределения -

$$f(x)=\frac{1}{B(a,b)}x^{a-1}(1-x)^{b-1},$$

где $B(a,b)=\Gamma(a)\Gamma(b)/\Gamma(a+b)$ - бета-функция.

При определении вероятности того, что случайная величина $X$ больше $x$, вычисляется интеграл вида

$$P\{X>x\}=\int_x^\infty f(x)dx,$$

а при вычислении квантили решается уравнение

$$P\{X>x\}=\int_x^\infty f(x)dx=\alpha.$$

Например, при вычислении квантили для $F$-распределения Фишера с числом степеней свободы 3 и 8 результат имеет вид

F-распределение Фишера (числа ст.св.- 3,8)
P{ X >=2.92380e+00} = 1.00000e-01

Здесь символ P{.} означает вероятность события. А для квантили $B$-распределения с параметрами a=5 и b=4:

B(5.000000,4.000000)-распределение
P{ X >=3.44623e-01} = 9.00000e-01

## 6.11. Пример анализа

Проанализируем следующую выборку объёмом в 200 наблюдений 5-мерной случайной величины.

| x1 | x2 | x3 | x4 | x5 |
|---|---|---|---|---|
| 1.24401 | 1.2256 | 2.48063 | 5.09715 | 5.67616 |
| 1.71241 | 0.626988 | 3.43145 | 5.28711 | 5.04412 |
| 1.29376 | 1.20933 | 2.58338 | 5.17358 | 5.73493 |
| 1.32286 | 1.0218 | 2.6555 | 4.93999 | 5.32052 |
| 1.41837 | 1.00409 | 2.84452 | 5.15056 | 5.47015 |
| 1.24692 | 1.46749 | 2.50888 | 5.46174 | 6.27485 |
| 1.40566 | 1.27103 | 2.81701 | 5.5629 | 6.0952 |
| 1.00601 | 1.13827 | 2.04045 | 4.34976 | 4.97117 |
| 1.17192 | 0.997179 | 2.33227 | 4.52311 | 4.94684 |
| 1.05463 | 1.36699 | 2.10472 | 4.82985 | 5.6535 |
| 1.65896 | 1.18183 | 3.3072 | 6.03614 | 6.40737 |
| 1.6825 | 1.28108 | 3.35661 | 6.24083 | 6.70489 |
| 1.30867 | 1.37217 | 2.6172 | 5.4758 | 6.17937 |
| 1.36076 | 1.51697 | 2.74501 | 5.82998 | 6.67746 |
| 1.479 | 1.05567 | 2.975 | 5.38695 | 5.7242 |
| 1.28845 | 0.766223 | 2.58322 | 4.4392 | 4.57574 |
| 1.66204 | 1.16597 | 3.31735 | 6.00566 | 6.32761 |
| 1.32855 | 1.59532 | 2.64287 | 5.88024 | 6.79189 |
| 1.30138 | 1.46439 | 2.60434 | 5.59894 | 6.40702 |
| 1.57241 | 0.764188 | 3.15051 | 5.13944 | 5.15057 |
| 1.60426 | 0.865616 | 3.20698 | 5.39251 | 5.46464 |
| 1.33294 | 0.867808 | 2.65662 | 4.71216 | 4.91132 |
| 1.84472 | 1.2165 | 3.68943 | 6.5491 | 6.90022 |
| 1.67768 | 1.30356 | 3.35608 | 6.2458 | 6.7743 |
| 2.09582 | 0.836802 | 4.20317 | 6.58409 | 6.38023 |
| 1.58499 | 1.25133 | 3.17203 | 5.94918 | 6.4411 |
| 1.79649 | 1.48236 | 3.59709 | 6.85967 | 7.49063 |
| 1.20041 | 1.41042 | 2.38901 | 5.2748 | 6.04652 |
| 1.1727 | 1.00727 | 2.34265 | 4.55354 | 4.98338 |
| 1.44199 | 1.36453 | 2.88975 | 5.7743 | 6.45937 |
| 1.78469 | 1.07384 | 3.56705 | 6.17195 | 6.36525 |
| 1.43938 | 1.18638 | 2.89787 | 5.5189 | 5.9697 |
| 1.5918 | 0.748616 | 3.17617 | 5.19534 | 5.14832 |
| 1.00155 | 1.4443 | 1.99476 | 4.79865 | 5.82385 |

| | | | | |
|---|---|---|---|---|
| 1.21119 | 1.09715 | 2.40933 | 4.78141 | 5.26095 |
| 1.89164 | 1.66817 | 3.79332 | 7.39909 | 8.08679 |
| 1.92501 | 1.31477 | 3.86136 | 6.9249 | 7.2562 |
| 1.3071 | 0.902101 | 2.59565 | 4.70909 | 4.95217 |
| 1.30553 | 1.49053 | 2.61575 | 5.62943 | 6.46633 |
| 1.38003 | 1.10771 | 2.76548 | 5.21542 | 5.64391 |
| 1.62849 | 1.58657 | 3.24878 | 6.61229 | 7.35888 |
| 1.38386 | 1.22136 | 2.77493 | 5.42297 | 5.9469 |
| 1.61045 | 0.644472 | 3.21246 | 5.038 | 4.89067 |
| 1.34355 | 1.43703 | 2.68659 | 5.65824 | 6.39915 |
| 1.47766 | 0.780442 | 2.95208 | 4.94146 | 5.00136 |
| 1.67379 | 0.953223 | 3.33968 | 5.71707 | 5.86704 |
| 1.24026 | 1.00932 | 2.46313 | 4.71767 | 5.10629 |
| 1.54236 | 1.20769 | 3.07441 | 5.79177 | 6.21144 |
| 1.38848 | 1.23728 | 2.78381 | 5.45152 | 5.98313 |
| 1.59789 | 1.07262 | 3.19644 | 5.71943 | 5.97903 |
| 1.46522 | 1.05868 | 2.928 | 5.36248 | 5.65149 |
| 1.34187 | 0.970568 | 2.69884 | 4.87945 | 5.20928 |
| 1.1542 | 1.10764 | 2.29957 | 4.67858 | 5.20517 |
| 1.78304 | 1.42447 | 3.58213 | 6.75278 | 7.26703 |
| 1.55925 | 1.3524 | 3.13699 | 6.05541 | 6.64916 |
| 1.21769 | 0.984217 | 2.44419 | 4.62245 | 5.01171 |
| 1.51113 | 1.34183 | 3.03326 | 5.93869 | 6.50574 |
| 1.26026 | 1.16637 | 2.51831 | 5.0282 | 5.54764 |
| 1.59664 | 1.3566 | 3.19053 | 6.1497 | 6.70063 |
| 1.42146 | 1.75347 | 2.84532 | 6.36413 | 7.39606 |
| 1.41708 | 1.84192 | 2.82756 | 6.49711 | 7.64694 |
| 1.66885 | 0.807161 | 3.33482 | 5.46284 | 5.43288 |
| 1.4798 | 1.16967 | 2.95222 | 5.58812 | 5.97844 |
| 1.50818 | 1.20522 | 3.00328 | 5.69629 | 6.12389 |
| 1.57933 | 1.58461 | 3.16432 | 6.48487 | 7.27747 |
| 1.42522 | 2.08873 | 2.85696 | 6.91472 | 8.26546 |
| 1.66549 | 1.02867 | 3.34 | 5.81373 | 6.01504 |
| 1.47943 | 1.31809 | 2.95446 | 5.82316 | 6.38761 |
| 1.45707 | 1.36877 | 2.91037 | 5.83317 | 6.47863 |
| 1.36451 | 1.1248 | 2.71099 | 5.19753 | 5.66211 |
| 1.57157 | 0.766671 | 3.14808 | 5.14317 | 5.14139 |
| 1.60351 | 1.26107 | 3.1994 | 6.02818 | 6.4721 |
| 1.66603 | 1.08977 | 3.30302 | 5.91026 | 6.15316 |
| 1.3873 | 1.32632 | 2.77961 | 5.58333 | 6.23729 |
| 1.54472 | 0.854598 | 3.11136 | 5.23656 | 5.33406 |
| 1.68483 | 1.16151 | 3.38362 | 6.06918 | 6.42687 |
| 1.81629 | 1.07399 | 3.62502 | 6.25261 | 6.42965 |
| 1.6459 | 0.459585 | 3.27606 | 4.85133 | 4.50704 |
| 1.33775 | 1.02766 | 2.68228 | 5.00576 | 5.35163 |

| | | | | |
|---|---|---|---|---|
| 1.83904 | 1.34001 | 3.6822 | 6.75383 | 7.16409 |
| 1.65125 | 1.21312 | 3.29912 | 6.08326 | 6.43743 |
| 1.49721 | 1.71816 | 2.99484 | 6.49994 | 7.45546 |
| 1.45079 | 0.904319 | 2.90132 | 5.0743 | 5.2421 |
| 1.82627 | 1.31303 | 3.64481 | 6.6691 | 7.06011 |
| 1.57454 | 0.882094 | 3.15175 | 5.34891 | 5.44047 |
| 1.52467 | 1.2328 | 3.04518 | 5.78761 | 6.30932 |
| 1.51681 | 1.47898 | 3.02587 | 6.16238 | 6.89092 |
| 1.42522 | 0.458566 | 2.85277 | 4.29468 | 4.04382 |
| 1.59861 | 1.29664 | 3.17657 | 6.07419 | 6.58023 |
| 1.46306 | 1.13572 | 2.91546 | 5.46568 | 5.88235 |
| 1.94733 | 1.13304 | 3.90023 | 6.66939 | 6.81622 |
| 1.54466 | 1.49117 | 3.08551 | 6.25612 | 6.95903 |
| 1.33182 | 1.35337 | 2.67797 | 5.48987 | 6.16495 |
| 1.61215 | 1.61973 | 3.21276 | 6.60808 | 7.45825 |
| 1.37886 | 1.43233 | 2.75926 | 5.7398 | 6.45111 |
| 1.51058 | 1.03671 | 3.02052 | 5.44175 | 5.71259 |
| 1.4434 | 1.56007 | 2.87869 | 6.10642 | 6.93417 |
| 1.42947 | 1.52384 | 2.87464 | 6.01334 | 6.85903 |
| 1.69337 | 1.39789 | 3.40394 | 6.47819 | 7.0393 |
| 1.54311 | 0.84544 | 3.0923 | 5.21515 | 5.28432 |
| 1.61035 | 0.647367 | 3.2091 | 5.08203 | 4.94336 |
| 1.46701 | 1.53372 | 2.93444 | 6.11735 | 6.91365 |
| 1.76831 | 1.42462 | 3.53699 | 6.69545 | 7.24432 |
| 1.46439 | 0.806636 | 2.91132 | 4.94761 | 5.01033 |
| 1.30939 | 0.682523 | 2.60087 | 4.35689 | 4.39265 |
| 1.53599 | 0.919842 | 3.06458 | 5.30057 | 5.47006 |
| 1.79477 | 1.5265 | 3.58569 | 6.93832 | 7.54557 |
| 1.81682 | 0.815109 | 3.63838 | 5.85174 | 5.74856 |
| 1.45599 | 1.34709 | 2.91643 | 5.7917 | 6.41529 |
| 1.54587 | 1.15713 | 3.08407 | 5.70687 | 6.12244 |
| 1.61082 | 1.70442 | 3.20537 | 6.76595 | 7.65464 |
| 1.294 | 1.20338 | 2.59069 | 5.16254 | 5.72614 |
| 1.34539 | 1.4255 | 2.68907 | 5.63966 | 6.39542 |
| 1.4849 | 1.2618 | 2.97161 | 5.73564 | 6.26433 |
| 1.43582 | 1.12897 | 2.87087 | 5.37222 | 5.80084 |
| 1.42438 | 1.27147 | 2.82977 | 5.6094 | 6.17364 |
| 1.72713 | 1.09831 | 3.46446 | 6.06621 | 6.32003 |
| 1.27251 | 1.24247 | 2.5481 | 5.16907 | 5.80226 |
| 1.66012 | 1.25353 | 3.33083 | 6.15325 | 6.58147 |
| 1.34708 | 1.37957 | 2.69794 | 5.56976 | 6.29585 |
| 1.42908 | 1.16076 | 2.85946 | 5.42588 | 5.85552 |
| 1.30776 | 1.35186 | 2.61509 | 5.43799 | 6.13265 |
| 1.32715 | 0.990423 | 2.66229 | 4.90526 | 5.21798 |
| 1.30768 | 0.882293 | 2.60497 | 4.68246 | 4.90686 |

| | | | | |
|---|---|---|---|---|
| 1.52897 | 1.1069 | 3.04494 | 5.59731 | 5.9343 |
| 1.18732 | 0.834908 | 2.37299 | 4.31443 | 4.51995 |
| 1.23307 | 0.853202 | 2.46297 | 4.44875 | 4.69166 |
| 1.70182 | 0.728847 | 3.38282 | 5.40226 | 5.2832 |
| 1.4798 | 1.26712 | 2.95122 | 5.72378 | 6.23038 |
| 1.26761 | 1.20091 | 2.54592 | 5.10169 | 5.65925 |
| 1.57715 | 1.20654 | 3.16325 | 5.87628 | 6.27827 |
| 1.50879 | 1.89595 | 3.00988 | 6.80127 | 7.97374 |
| 1.70961 | 1.30586 | 3.42639 | 6.36296 | 6.80934 |
| 1.23087 | 1.35963 | 2.46502 | 5.23965 | 5.98938 |
| 1.83816 | 1.63492 | 3.67741 | 7.21999 | 7.93946 |
| 1.72783 | 0.608112 | 3.44183 | 5.30432 | 5.05766 |
| 1.64165 | 0.929653 | 3.29318 | 5.59732 | 5.67981 |
| 1.54849 | 1.20997 | 3.10534 | 5.81219 | 6.22168 |
| 1.27493 | 1.24931 | 2.54631 | 5.18626 | 5.81362 |
| 1.53751 | 0.782996 | 3.08761 | 5.09509 | 5.10766 |
| 1.54239 | 1.43903 | 3.09205 | 6.15761 | 6.82049 |
| 1.22154 | 0.935143 | 2.44227 | 4.5601 | 4.88344 |
| 1.72933 | 1.02503 | 3.46821 | 5.97082 | 6.13656 |
| 1.52819 | 1.32725 | 3.06647 | 5.93795 | 6.5051 |
| 1.11506 | 1.1638 | 2.22453 | 4.64913 | 5.25791 |
| 1.57551 | 1.26498 | 3.13515 | 5.97837 | 6.45655 |
| 1.40097 | 1.01784 | 2.7942 | 5.12801 | 5.44336 |
| 1.83623 | 1.01268 | 3.6809 | 6.20774 | 6.3132 |
| 1.55502 | 1.0843 | 3.11781 | 5.62535 | 5.91954 |
| 1.56438 | 0.911979 | 3.12657 | 5.3526 | 5.49822 |
| 1.34502 | 0.829792 | 2.69847 | 4.67901 | 4.84952 |
| 1.78971 | 0.90419 | 3.57937 | 5.91267 | 5.90741 |
| 1.68758 | 1.42639 | 3.38736 | 6.523 | 7.08313 |
| 1.21643 | 0.670999 | 2.42476 | 4.10568 | 4.16007 |
| 1.73636 | 0.756009 | 3.46621 | 5.53229 | 5.44644 |
| 1.66967 | 0.85911 | 3.31936 | 5.56384 | 5.57753 |
| 1.3223 | 1.24236 | 2.65339 | 5.30137 | 5.8766 |
| 1.36747 | 1.26422 | 2.72754 | 5.43657 | 6.01695 |
| 1.96619 | 1.70476 | 3.94627 | 7.64195 | 8.39332 |
| 1.27467 | 1.27734 | 2.56813 | 5.2109 | 5.87214 |
| 1.14296 | 0.900382 | 2.29276 | 4.29719 | 4.61606 |
| 1.38622 | 1.63 | 2.77733 | 6.08048 | 6.97478 |
| 1.36675 | 1.09948 | 2.72947 | 5.17543 | 5.58449 |
| 1.59466 | 1.46363 | 3.18913 | 6.3378 | 6.97631 |
| 1.5542 | 1.02977 | 3.09078 | 5.52445 | 5.81798 |
| 1.28773 | 1.87586 | 2.56876 | 6.22385 | 7.46697 |
| 1.82525 | 1.3904 | 3.65212 | 6.78801 | 7.27904 |
| 1.498 | 0.884795 | 3.00755 | 5.14921 | 5.3173 |
| 1.71347 | 1.14478 | 3.42325 | 6.11658 | 6.39527 |

| | | | | |
|---|---|---|---|---|
| 1.18288 | 1.10028 | 2.37754 | 4.73173 | 5.22001 |
| 1.76138 | 1.0378 | 3.53594 | 6.07323 | 6.21645 |
| 1.64432 | 1.19307 | 3.26918 | 6.00692 | 6.39149 |
| 1.66202 | 1.15409 | 3.32677 | 6.0008 | 6.3344 |
| 1.60212 | 1.6845 | 3.20938 | 6.69505 | 7.5438 |
| 1.37164 | 1.52984 | 2.75856 | 5.88636 | 6.69324 |
| 1.61281 | 1.56986 | 3.21502 | 6.54989 | 7.30232 |
| 1.67274 | 1.50002 | 3.35116 | 6.57673 | 7.2391 |
| 1.63481 | 1.15343 | 3.28419 | 5.93019 | 6.25926 |
| 1.45753 | 1.04317 | 2.91135 | 5.32522 | 5.61975 |
| 1.36059 | 1.46122 | 2.71267 | 5.74994 | 6.52728 |
| 1.34388 | 1.30683 | 2.69123 | 5.46972 | 6.08964 |
| 1.77562 | 1.18359 | 3.5399 | 6.3311 | 6.65325 |
| 1.64576 | 1.86375 | 3.29662 | 7.10264 | 8.15092 |
| 1.33979 | 1.44296 | 2.67476 | 5.66298 | 6.43814 |
| 1.7063 | 1.53408 | 3.42768 | 6.71992 | 7.39294 |
| 1.42237 | 0.951325 | 2.8432 | 5.0781 | 5.32959 |
| 1.46821 | 0.80818 | 2.95352 | 4.96957 | 5.03536 |
| 1.13663 | 1.5671 | 2.25583 | 5.36095 | 6.32884 |
| 1.43349 | 0.551934 | 2.85593 | 4.45601 | 4.28982 |
| 1.60298 | 1.40917 | 3.18974 | 6.26216 | 6.87401 |
| 1.40617 | 1.52629 | 2.8148 | 5.95934 | 6.79871 |
| 1.75501 | 0.354564 | 3.50081 | 4.95047 | 4.43598 |
| 1.58236 | 0.850557 | 3.16951 | 5.31713 | 5.39102 |
| 1.6998 | 1.55122 | 3.40582 | 6.74306 | 7.43312 |
| 1.45459 | 1.09787 | 2.91639 | 5.40352 | 5.76469 |
| 1.45371 | 1.09149 | 2.90583 | 5.38438 | 5.70413 |
| 1.19975 | 1.36895 | 2.39739 | 5.18489 | 5.95598 |
| 1.8629 | 1.45191 | 3.73232 | 6.9738 | 7.51626 |
| 1.37538 | 1.08964 | 2.76518 | 5.16681 | 5.5616 |
| 1.30086 | 1.15456 | 2.59957 | 5.09375 | 5.61028 |

Проверка предпосылок показала, что выборка нормальна, случайна и не содержит аномальных измерений. Ниже приведены некоторые оценки и результаты анализа.

Вектор математического ожидания

| N | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 1.499448 | 1.191171 | 2.998918 | 5.655132 | 6.097118 |

Матрица парных коэффициентов корреляции

| N | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 1.000000e+00 | -9.399218e-03 | 9.996689e-01 | 7.198058e-01 | 4.547873e-01 |

| | | | | | |
|---|---|---|---|---|---|
| 2 | -9.399218e-03 | 1.000000e+00 | -6.477115e-03 | 6.872305e-01 | 8.860626e-01 |
| 3 | 9.996689e-01 | -6.477115e-03 | 1.000000e+00 | 7.216361e-01 | 4.572405e-01 |
| 4 | 7.198058e-01 | 6.872305e-01 | 7.216361e-01 | 1.000000e+00 | 9.452698e-01 |
| 5 | 4.547873e-01 | 8.860626e-01 | 4.572405e-01 | 9.452698e-01 | 1.000000e+00 |

Матрица значимых коэффициентов корреляции
Уровень значимости alfa = 0.0500

| N | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 1.000000+00 | 0.000000e+00 | 9.996689e-01 | 7.198058e-01 | 4.547873e-01 |
| 2 | 0.000000e+00 | 1.000000e+00 | 0.000000e+00 | 6.872305e-01 | 8.860626e-01 |
| 3 | 9.996689e-01 | 0.000000e+00 | 1.000000e+00 | 7.216361e-01 | 4.572405e-01 |
| 4 | 7.198058e-01 | 6.872305e-01 | 7.216361e-01 | 1.000000e+00 | 9.452698e-01 |
| 5 | 4.547873e-01 | 8.860626e-01 | 4.572405e-01 | 9.452698e-01 | 1.000000e+00 |

Вектор множественных коэффициентов корреляции

| N | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 9.998986e-01 | 9.998923e-01 | 9.998984e-01 | 9.998467e-01 | 9.997837e-01 |

Каждый элемент этого вектора представляет собой коэффициент корреляции между соответствующей компонентой и множеством всех остальных компонент. Анализ значений вектора показывает, что взаимозависимость очень сильная. И больше информации из этого вектора извлечь нельзя. В то же время из матрицы парных корреляций видно, что, по-видимому, взаимозависимость между компонентами первой и второй, а также второй и третьей практически отсутствует.

Матрица частных коэффициентов корреляции

| N | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 1.000000+00 | -6.892923e-01 | 5.129580e-01 | 6.657743e-01 | 4.063153e-01 |
| 2 | -6.892923e-01 | 1.000000e+00 | -2.152277e-01 | 7.434679e-01 | 6.992099e-01 |
| 3 | 5.129580e-01 | -2.152277e-01 | 1.000000e+00 | 1.126719e-01 | 4.598697e-01 |
| 4 | 6.657743e-01 | 7.434679e-01 | 1.126719e-01 | 1.000000e+00 | 1.347956e-02 |
| 5 | 4.063153e-01 | 6.992099e-01 | 4.598697e-01 | 1.347956e-02 | 1.000000e+00 |

Матрица частных коэффициентов линейной регрессии

| N | 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|
| 1 | 1.000000+00 | -5.162699e-01 | 1.600408e-01 | 2.205248e-01 | 6.227712e-02 |
| 2 | -1.183319e+00 | 1.000000e+00 | -4.479462e-03 | 3.414649e-01 | 1.740037e-01 |
| 3 | 1.960046e+00 | -2.393514e-02 | 1.000000e+00 | 2.396893e-02 | -4.080884e-03 |
| 4 | 2.499815e+00 | 1.688773e+00 | 2.218521e-02 | 1.000000e+00 | -3.231518e-02 |
| 5 | 2.266627e+00 | 2.763025e+00 | -1.212747e-02 | -1.037547e-01 | 1.000000e+00 |

В матрице частных коэффициентов корреляции приведены значения коэффициентов при исключении влияния всех компонент кроме тех, корреляция между которыми вычисляется. Посмотрим, что получится, если будем исключать влияние только некоторых компонент. Исключая влияние второй компоненты, найдём $r_{13\cdot 2} = 0.999688$, $r_{14\cdot 2} = 0.999813$, $r_{15\cdot 2} = 0.999074$. Видно, что вторая компонента существенно маскировала корреляции между первой и четвертой и первой и пятой компонентами, которые на самом деле существенно сильнее. Исключая влияние первой компоненты, найдем $r_{23\cdot 1} = 0.116107$, $r_{24\cdot 1} = 0.999795$, $r_{25\cdot 1} = 0.999749$, $r_{34\cdot 1} = 0.116556$, $r_{35\cdot 1} = 0.115891$. Как видим, здесь маскировалась корреляция между второй и четвёртой и второй и пятой компонентами, а вот взаимозависимость третьей компоненты с четвёртой и пятой, по-видимому, в основном определяется их общей зависимостью от первой компоненты. В то же время $r_{45\cdot 1} = 0.999518$, $r_{45\cdot 2} = 0.998839$. А при исключении влияния первой и второй компонент - $r_{45\cdot 1,2} = -0.058195$. Это даёт возможность предполагать, что взаимозависимость четвёртой и пятой компонент определяется их зависимостью от первой и второй.

Таким образом, достаточно беглый анализ позволяет выделить 3 группы взаимозависимых компонент: 1 и 3; 1, 2 и 4; 1, 2 и 5. Это значит, что можно попробовать провести отдельно корреляционный анализ выделенных групп компонент.

Анализируя отдельно только компоненты 1 и 3, получаем коэффициенты корреляции $r_{13} = r_{31} = 0.9996689$, корреляционные отношения $\rho_{13}^2 = 0.998123$ и $\rho_{31}^2 = 0.998236$. В результате проверки гипотезы $H_0: \rho_{13}^2 = r_{13}^2$ убеждаемся, что гипотеза $H_0$ не отклоняется при заданном уровне значимости $\alpha = 0.05$. Это означает, что наилучшая регрессия $x_1$ по $x_3$ - линейная. Аналогично, при проверке соответствующей гипотезы убеждаемся, что наилучшая регрессия $x_3$ по $x_1$ также линейная.

Матрица частных коэффициентов линейной регрессии имеет вид:

| N | 1 | 3 |
|---|---|---|
| 1 | 1.000000+00 | 4.989953e-01 |
| 3 | 2.002758e+00 | 1.000000e+00 |

Следовательно, соотношения для линейной регрессии могут быть представлены зависимостями

$$x_1 = 0.498995x_3 + \varepsilon_1,$$
$$x_3 = 2.002758x_1 + \varepsilon_3,$$

где $\varepsilon_1$ и $\varepsilon_3$ - ошибки наблюдений.

Более того, значения парных коэффициентов корреляции и корреляционных отношений позволяют предположить, что справедлива, например, и гипотеза вида $H_0: r_{31}^2 = \rho_{31}^2 = 1$. А это возможно лишь тогда, когда имеется *строгая линейная функциональная* зависимость $x_3$ от $x_1$.

Анализируя отдельно только компоненты 1, 2 и 4, замечаем, что корреляция между первой и второй компонентой незначима, а компонента $x_4$ коррелирована с $x_1$ и $x_2$. Коэффициенты корреляции $r_{41} = 0.7198058$, $r_{42} = 0.6872305$, корреляционные отношения $\rho_{41}^2 = 0.6031383$ и $\rho_{42}^2 = 0.5660371$. В результате проверки гипотез $H_0: \rho_{41}^2 = r_{41}^2$ и $H_0: \rho_{42}^2 = r_{42}^2$ убеждаемся, что гипотезы $H_0$ при заданном уровне значимости $\alpha = 0.05$ не отклоняются.

случаях. Это означает, что наилучшая регрессия $x_4$ по $x_1$ и $x_4$ по $x_2$ - линейная.

Вычисленная матрица частных коэффициентов линейной регрессии в данном случае имеет вид:

| N | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 1.000000+00 | -6.426641e-01 | 4.004236e-01 |
| 2 | -1.554817e+00 | 1.000000e+00 | 6.228157e-01 |
| 3 | 2.496423e+00 | 1.604955e+00 | 1.000000e+00 |

Следовательно, соотношение для линейной регрессии $x_4$ по $x_1$ и $x_2$ может быть представлено зависимостью

$$x_4 = 2.496423x_1 + 1.604955x_2 + \varepsilon_4,$$

где $\varepsilon_4$ - случайная ошибка наблюдений.

Аналогично, анализируя только компоненты 1, 2 и 5, отмечаем, что корреляция между первой и второй компонентой незначима, а компонента $x_5$ коррелирована с $x_1$ и $x_2$. Коэффициенты корреляции $r_{51} = 0.4547873$, $r_{52} = 0.8860626$, корреляционные отношения $\rho_{51}^2 = 0.3465583$ и $\rho_{52}^2 = 0.8205661$. Результаты проверки гипотез $H_0: \rho_{51}^2 = r_{51}^2$ и $H_0: \rho_{52}^2 = r_{52}^2$ показывают, что гипотезы $H_0$ при заданном уровне значимости $\alpha = 0.05$ не отклоняются в обоих случаях. Это означает, что наилучшая регрессия $x_5$ по $x_1$ и $x_5$ по $x_2$ - линейная.

Матрица частных коэффициентов линейной регрессии в этом случае имеет вид:

| N | 1 | 2 | 3 |
|---|---|---|---|
| 1 | 1.000000+00 | -1.291135e+00 | 4.972641e-01 |
| 2 | -7.727050e-01 | 1.000000e+00 | 3.849475e-01 |
| 3 | 2.007282e+00 | 2.596454e+00 | 1.000000e+00 |

Следовательно, соотношение для линейной регрессии $x_5$ по $x_1$ и $x_2$ может быть представлено зависимостью

$$x_5 = 2.007282x_1 + 2.596454x_2 + \varepsilon_5,$$

где $\varepsilon_5$ - случайная ошибка наблюдений.

В заключение отметим, что данная выборка была смоделирована следующим образом. Компоненты $x_1$ и $x_2$ моделировались как независимые нормальные случайные величины, причём $x_1$ - с математическим ожиданием 1.5 и среднеквадратическим отклонением 0.2, $x_2$ - с математическим ожиданием 1.2 и среднеквадратическим отклонением 0.3, $x_3$ моделировалась как $x_3 = 2x_1 + \varepsilon_3$, где случайная ошибка наблюдения $\varepsilon_3$ моделировалась как нормальная величина с нулевым математическим ожиданием и среднеквадратическим отклонением 0.01, $x_4$ - как $x_4 = 2.5x_1 + 1.6x_2 + \varepsilon_4$ и случайная ошибка наблюдения $\varepsilon_4$ имела тот же закон распределения, что и $\varepsilon_3$, $x_5$ - как $x_5 = 2x_1 + 2.6x_2 + \varepsilon_5$, а ошибка наблюдения $\varepsilon_5$ распределена как нормальная величина с нулевым математическим ожиданием и среднеквадратическим отклонением 0.02.

Таким образом, используя по существу только процедуры корреляционного анализа, нам удалось выделить зависимые и независимые компоненты и восстановить функциональные соотношения. Очевидно, что при обработке реальных наблюдений всё оказывается более сложным и не таким очевидным. И шаги, связанные с построением регрессионных зависимостей или восстановлением функциональных, лежат уже за пределами корреляционного анализа: в рамках регрессионного анализа или соответствующей предметной области. В то же время данный пример достаточно наглядно иллюстрирует возможности методов корреляционного анализа.

## Рекомендуемая литература


1. Андерсон Т. Введение в многомерный статистический анализ. - М.: Физматгиз, 1963.
2. Кендалл М., Стьюарт А.. Статистические выводы и связи. - М.: Наука, 1973.
3. Кендалл М., Стьюарт А.. Многомерный статистический анализ и временные ряды. - М.: Наука, 1976.

## Содержание

Борис Юрьевич Лемешко

# КОРРЕЛЯЦИОННЫЙ АНАЛИЗ МНОГОМЕРНЫХ СЛУЧАЙНЫХ ВЕЛИЧИН

## Программная система

Редактор *И. Л. Кескевич*
Технический редактор *Г. Е. Телятникова*

Лицензия № 020356 от 27.12.91. Подписано в печать 30.06.95. Формат 60×84 1/16. Бумага оберточная. Тираж 170 экз. Уч.-изд. л. 2,5. Печ. л. 2,75. Изд. № 119. Заказ № 308 Цена договорная.

Отпечатано в типографии
Новосибирского государственного технического университета
630092, г. Новосибирск, пр. К. Маркса, 20.