# КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

## Случайная векторная переменная

***Слайд 2***

В задачах многомерного статистического анализа (анализ главных компонент, факторный анализ, каноническая корреляция) в матрице «объект-признак» признаки являются случайными величинами $X_j$, $j$ =1, …, $m$. Случайную величину, как это принято в теории вероятностей, будем обозначать так

$$X_j = [x_{ji}, \text{K}, x_{jn}],$$

где $x_{ij}$ –является $i$-й реализацией $j$-го признака. Тогда матрица данных «объект-признак» является случайной векторной переменной (СВП)

$$\mathbf{X} = \begin{bmatrix} X_1 \\ \text{M} \\ X_m \end{bmatrix},$$

компонентами которой являются случайные признаки $X_j$. Теперь каждый объект описывается $i$-й реализацией СВП и определяется как неслучайный $m$-мерный вектор-столбец

$$\mathbf{x}_i = [x_{1i}, \text{K}, x_{mi}]', \quad i = 1, \text{K}, n.$$

Отсюда наблюдения над $n$-объектами можно представить в виде $n$-реализаций СВП

$$\mathbf{X} = [\mathbf{x}_1, \text{K}, \mathbf{x}_n].$$

СВП является ($m \times n$)-матрицей, равной транспонированной матрице «объект-признак».

***Слайд 3***

## Ковариационная и корреляционные матрицы

Многие методы многомерного анализа данных, включая и *множественную линейную регрессию, анализ главных компонент, факторный анализ, дискриминантный и канонический корреляционный анализы* построены на преобразованиях исходной матрицы «объект-признак» в *ковариационную* или *корреляционную матрицу*. Ковариационная матрица – это квадратная матрица типа «признак-признак» размера $m \times m$, образованная из попарных ковариаций $m$ случайных признаков $X_1$, …, $X_m$ матрицы «объект-признак» **X**. Ковариационная матрица $\mathbf{K}_x$ для матрицы **X** определяется так

$$\mathbf{K}_x = \mathbf{M}[(\mathbf{X} - \mathbf{MX})'(\mathbf{X} - \mathbf{MX})],$$

где $\mathbf{MX} = (\mathbf{M}X_1, \text{K}, \mathbf{M}X_m)$ – вектор средних значений.

Компоненты ковариационной матрицы

$$k_{ij} = \mathbf{M}[(X_i - \mathbf{M}X_i)(X_j - \mathbf{M}X_j)] = \text{cov}(X_i, X_j),\ i,\ j = 1, \text{K}, m$$

при $i = j$ совпадают с дисперсией величины $X_i$. Ковариационная матрица представляет собой симметричную *неотрицательно определенную матрицу*. Пусть признаки $X_i$, $i$ =1, …, $m$ линейно независимы в том смысле, что ее все нетривиальные линейные комбинации вида $\Sigma_j a_j X_j$ ($a_j$ – компонента произвольного вектора) являются невырожденными случайными величинами, то есть не являются константами и имеют положительную дисперсию. Тогда *ковариационная матрица матрицы наблюдений* **X** *положительно определена.* Если дисперсии признаков $X_1$, …, $X_m$ равны 1, то ковариационная матрица называется *корреляционной матрицей.*

Оценка элементов ковариационной матрицы для матрицы наблюдений **X** размерности $n \times m$ находится по формуле

$$\hat{k}_{ij} = \frac{1}{n-1}\sum_{s=1}^{n}(x_{si} - \bar{x}_i)(x_{sj} - \bar{x}_j),\ i, j = 1, \text{K}, m \qquad (1)$$

где

$$\bar{x}_i = (1/n)\sum_{S=1}^{n} x_{si} \qquad (2)$$

среднее признака $X_i$.

Часто используется и оценка, вычисляемая по формуле

$$\hat{k}_{ij} = \frac{1}{n-1}\left[\sum_{s=1}^{n} x_{si}x_{sj} - \frac{1}{n}\bar{x}_i\bar{x}_j\right] \qquad (3)$$

Оценка по формуле (3) обладает относительно выражения (1) низкой точностью, которая объясняется тем, что она представляет часто разность двух близких друг к другу неотрицательных величин. Указанная близость чисел при вычислениях из-за округлений приводит к потере точности. Если дисперсии переменных малы, то эта оценка может дать даже отрицательные значения диагональных элементов ковариационной матрицы.

Оценки элементов $r_{ij}$ корреляционной матрицы можно получить по формуле

$$\hat{r}_{ij} = \hat{k}_{ij} / S_i S_j,$$

где $S_i = (\hat{k}_{ii})^{1/2}$ и $S_j = (\hat{k}_{jj})^{1/2}$ – среднеквадратические отклонения признаков $X_i$ и $X_j$. Если признаки $X_1$, …, $X_m$ имеют нормальное распределение с ковариационной матрицей $\mathbf{K}_x$, то оценка $\hat{\mathbf{K}}_x$ является оценкой максимума правдоподобия.

***Слайд 4***

Все приведенные выше формулы оценок ковариационной матрицы справедливы для матрицы наблюдений «объект-признак» без пропусков. Рассмотрим свойства оценок при наличии пропусков в наблюдениях. Возможны следующие варианты нахождения оценок при наличии пропусков:

1) вычисление оценки только по данным объектов без пропусков;
2) вычисление оценки по всем доступным наблюдениям;
3) вычисление оценки по данным, пропуски которых заполнены одним из существующих методов.

Для получения оценок по первому варианту используются формулы для полных данных, но при меньшем числе наблюдений. Этот вариант в пакетах прикладных программ соответствует процедуре (*Listwise*), и он позволяет получать состоятельные оценки параметров, если справедливы условия *ОПС* и име-

ются, по меньшей мере, $m + 1$ комплектных наблюдений. Если *ОПС* не соблюдается, то оценки будут смещенными.

Во втором варианте для вычисления средних, дисперсий и коэффициентов ковариаций используются все данные. Здесь возможны несколько способов нахождения оценок. Наиболее распространена процедура нахождения оценок только по всем комплектным парам (*Pairwise*). Ковариация между $X_j$ и $X_k$ по наблюдениям, для которых присутствует и $x_{ij}$ и $x_{ik}$, определяется формулой

$$\hat{k}_{jk}^{(jk)} = \frac{1}{n^{(jk)}-1}\sum_{(jk)}(x_{ij} - \bar{x}_j^{(jk)})(x_{ik} - \bar{x}_k^{(jk)}) \tag{4}$$

где $n^{(jk)}$– число наблюдений, в котором одновременно присутствуют $X_j$ и $X_k$, а средние $\bar{x}_j^{(jk)}$, $\bar{x}_k^{(jk)}$ и сумма вычисляется по этим $n^{(jk)}$ наблюдениям. Если среднеквадратические отклонения $S_j^{(jk)}$, $S_k^{(jk)}$ вычислены по $n^{(jk)}$ наблюдениям, то парные корреляции вычисляются по формуле

$$\hat{r}_{ij} = \hat{k}_{jk}^{(jk)} / S_j^{(jk)} S_k^{(jk)} \tag{5}$$

Недостатком этого варианта является то, что оцененные по формулам (4) и (5) ковариационные и корреляционные матрицы *не всегда положительно определены.* Эта проблема существенно обостряется с усилением коррелированности переменных. В случаях сильных связей между переменными более эффективным становится применение метода комплектных наблюдений. Метод парных комплектных наблюдений дает хорошие результаты при незначительных корреляциях между переменными. В целом оценки по всем доступным парам наблюдений пригодны как начальные оценки в итеративных процедурах.

Если в формуле (1) среднеквадратические значения $S_j^{(j)}$, $S_k^{(k)}$, вычислим по доступным наблюдениям $X_j$ и $X_k$, то получим следующие элементы корреляционной матрицы

$$\hat{r}_{ik} = \hat{k}_{jk}^{(jk)} / S_j^{(j)} S_k^{(k)} \tag{6}$$

В формуле (6) более полно используются данные, но полученная корреляционная матрица не является положительно определенной и, следовательно, она может привести к трудностям на первой итерации в итеративных процедурах. Однако оценки величины $r_{jk}$ могут оказаться вне отрезка [–1, 1], что противоречит смыслу истинной корреляции. Можно получить еще несколько вариантов оценок, заменяя средние в формулах (1) – (3) их средними по всем присутствующим наблюдениям. Например, применяя такой способ к выражению (4), получаем оценку

$$\hat{k}_{jk}^{(jk)} = \frac{1}{n^{(jk)}-1}\sum_{(jk)}(x_{ij} - \bar{x}_j^{(j)})(x_{ik} - \bar{x}_k^{(k)}) \tag{7}$$

которая используется в системе BMDP8D под именем *ALLVALU.*

Рассмотренные оценки по доступным наблюдениям типа (4) – (7) при дают возможность получать состоятельные ковариации и корреляции по отдельности. Если их рассматривать в совокупности, то выясняется, что они обладают недостатками, существенно снижающими их практическую применимость.

Лучших результатов можно достичь, если находить оценки с применением оптимальных методов с восстановлением пропусков. Эти методы требуют больших вычислительных затрат, и в них искажение оценок увеличивается с

ростом доли пропусков, поэтому по возможности целесообразно пользоваться другими методами, не связанными с заполнением пропусков.

***Слайд 5***

**Матрицы близостей**

Ряд процедур статистической обработки, такие как Q-факторный и кластерный анализы, многомерное шкалирование, анализ соответствий использует в качестве исходных данных квадратную матрицу Δ «объект-объект» размерности $n \times n$, элементами которых являются меры сходства или различий $\delta_{ij}$ между объектами $X_j$ и $X_j$, $i, j = 1, \ldots, n$. Каждая строка и каждый столбец матрицы Δ соответствует одному объекту. Элементом $\delta_{ij}$ в $i$-й строке и $j$-м столбце матрицы Δ является мера сходства между объектами $i$ и $j$.

Количественное оценивание сходства тесно связано с понятием *метрика*. Пусть объекты представлены точками координатного пространства и меру сходства или различия оценивают в соответствии с *метрическим расстоянием* между точками. В метрическом пространстве каждой паре элементов $X_i$ и $X_j$ поставлено в соответствие расстояние, обозначаемое $d_{ij} = d(X_i, X_j)$. Для того, чтобы мера сходства была метрикой любых объектов $X_i$, $X_j$ и $X_k$, она должна удовлетворять аксиомам:

1) $d_{ij} = 0 \Leftrightarrow X_i = X_j$ для всех $i, j = 1, \text{K}, n$ (аксиома тождества);

2. $d_{ij} \le d_{ik} + d_{jk}$ для всех $i, j = 1, \text{K}, n$ (аксиома треугольника);

3. $d_{ij} = d_{ji}$ для всех $i, j = 1, \text{K}, n$ (аксиома симметрии).

Важность метрики как меры сходства неоспорима, однако, она не является единственным способом описания близости объектов. Существует несколько видов сходства, в которых меры не являются метриками. Например, оценивание близости объектов может основываться на процессе сопоставления признаков. Это понятие не связано с размерностью пространства. Более того, в социальных исследованиях близости объектов оценивают непосредственно. Например, за основу берут степень взаимосвязи объектов. В этих случаях часто нарушается аксиома симметрии: *A* соответствует *B*, но *B* соответствует *A* не в той же степени. Нарушение симметрии вызывает определенные трудности при исследованиях, но применение таких мер может оказаться целесообразным и необходимым при решении конкретных задач. Рассмотрим некоторые меры сходства.

***Слайд 6***

***Корреляционная матрица «объект-объект».*** Пусть признаки в матрице «объект-признак» измерены в количественной шкале. За меру сходства двух объектов в этой матрице возьмем квадратный корень из коэффициента корреляции

$$\hat{r}_{ik} = \frac{\sum_{j=1}^{m}(x_{ij} - \bar{x}_i)(x_{kj} - \bar{x}_k)}{\{\sum_{j=1}^{m}(x_{ij} - \bar{x}_i)^2 \sum_{j=1}^{m}(x_{kj} - \bar{x}_k)^2\}^{1/2}}, \quad i,\ k = 1, \text{K}, n, \tag{8}$$

где $x_{ij}$– значение $j$-ой переменной для $i$-го объекта, $\bar{x}_i$ – среднее для всех переменных $i$-го объекта, $n$ – число объектов, $m$ – число признаков. Вычисленные коэффициенты корреляции образуют корреляционную матрицу типа «объект-

объект» размерности $n \times n$. Если признаки измерены в номинальной или в порядковой шкале, то вместо $r_{ik}$ используются соответственно коэффициенты сопряженности и ранговой корреляции.

Коэффициенты корреляционной матрицы «объект-объект» в отличие от корреляционной матрицы «признак-признак» не имеют ясного физического смысла, так как средние $\bar{x}_j$ в формуле (8) определяются по всем признакам одного объекта. Поэтому полученный коэффициент отражает форму, но не чувствителен к различиям в величине переменных, используемых при вычислении коэффициента. Чувствительность к форме особенно важна в таких науках как социология, психология и антропология, в которых используются профили. Профиль – это вектор значений признаков объекта, изображенный в виде ломаной линии. Два профиля могут иметь корреляцию, равную 1, но не будут идентичными (рис. 1).

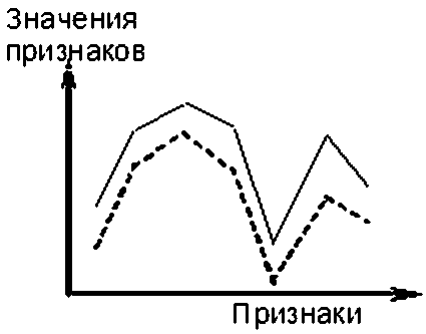

Рис. 1. Профили двух объектов

Недостатком корреляционной матрицы «объект-объект» как меры сходства является то, что элементы этой матрицы не удовлетворяют аксиоме треугольника. Эта мера является полезной в приложениях кластерного анализа и многомерного шкалирования, где важна форма, а не сдвиг и масштаб. Если необходимо анализировать не меру сходства, а меру различий, то корреляции нужно преобразовать в различия:

$$\delta_{ij} = (1 - r_{ij})^{1/2}.$$

***Слайд 7***

***Меры расстояния***. Для этой меры большому расстоянию соответствует меньшее сходство. Поэтому эту меру лучше бы называть мерой *несходства*, *различия*. Меры расстояния не ограничены сверху и зависят от выбора масштаба измерений. Если исходные данные заданы матрицей «объект-признак» **X**, то возможно вычисление различных типов расстояний между объектами $X_i$ и $X_j$.

Одним из наиболее часто применяемых расстояний является *расстояние Минковского*[1].

$$d_{ij} = d(X_i, X_j) = \left( \sum_{k=1}^{m} w_k \mid x_{ik} - x_{jk} \mid^q \right)^{1/q},$$

где $w_k > 0$ – весовые коэффициенты при признаках, $q \geq 1$ –параметр степени. Веса выражают степень важности измерения при различении объектов. При сопоставлении двух объектов измерениям признака с меньшей значимостью нужно присваивать меньшие веса. Назначение весов признаков можно осуществить посредством линейного преобразования. В тех случаях, когда линейное преобразование сводится к изменению масштабов, матрица весов является диагональной $W = diag(w_1, \ldots, w_m)$. Параметр степени, как установили психологи, представляет собой монотонную функцию от степени внимательности.

[1] Минковский (Minkowski) Герман (22.6.1864 – 12.1.1909) – немецкий математик и физик.

Это расстояние нужно применять тогда, когда все признаки матрицы **X** взаимно независимы и однородны, например, измерены в одних и тех же единицах и одинаково важны для решаемой задачи. Частными случаями расстояния Минковского являются:

*а*) *евклидово*[2] *расстояние*

$d_{ij} = \left(\sum_{k=1}^{m} (x_{ik} - x_{jk})^2\right)^{1/2}, \quad w_k = 1,\ q = 2,$

*б*) *манхеттингское расстояние* (*расстояние* «*Сити-блок*»

$d_{ij} = \sum_{k=1}^{m} |x_{ik} - x_{jk}|, \quad w_k = 1,\ q = 1,$

*в*) *доминантное расстояние*

$d_{ij} = \max_k |x_{ik} - x_{jk}|, \quad w_k = 1,\ q = \infty.$

*Расстояние Махаланобиса* [3]

$d_{ij} = (\mathbf{x}_i - \mathbf{x}_j)' \mathbf{K}_x^{-1} (\mathbf{x}_i - \mathbf{x}_j),$

где $\mathbf{x}_i$, $\mathbf{x}_j$ – векторы переменных объектов *i* и *j*, $\mathbf{K}_x^{-1}$ – обратная корреляционная матрица признаков. Это расстояние применяется при сильной зависимости и физической неоднородности исследуемых признаков. В отличие от евклидовой метрики и метрики Минковского, эта метрика через ковариационную матрицу связана с корреляциями признаков. Если признаки некоррелированы, то расстояние Махаланобиса эквивалентно евклидовому расстоянию.

Несмотря на важность рассмотренных метрик, они имеют серьезные недостатки. Наиболее существенным недостатком является то, что оценка сходства сильно зависит от различий в сдвигах признаков. Если одновременно велики абсолютные значения и стандартные отклонения признаков, то эти признаки могут подавить влияние переменных с меньшими абсолютными значениями и стандартными отклонениями. Для уменьшения указанных влияний можно использовать нормированные признаки с нулевыми средними и единичными дисперсиями. Однако нормировка также может сильно повлиять на меру сходства.

*Хеммингово расстояние.* В теории кодирования количество разрядов, в которых кодовая комбинация *A* отличается от кодовой комбинации *B*, называется *хемминговым расстоянием* $d_{AB}$ между этими комбинациями. Нетрудно убедиться, что хеммингово расстояние удовлетворяет всем трем аксиомам метрики, а именно: 1) $d_{AB} = 0$ тогда и только тогда, когда комбинации A и B тождественно равны, 2) $d_{AB} = d_{BA}$ и 3) для любых трех комбинаций выполняется аксиома треугольника $d_{AC} \le d_{AB} + d_{BC}$.

***Слайд 8***

**Определение статистической значимости коэффициентов корреляции**

Исследование структуры статистических связей на основе корреляционного анализа достаточно широко используется в статистических исследованиях

---

[2] Евклид (Eukléidês) (ок. 340 – ок.287 до н. э.) – древнегреческий математик

[3] Махаланобис(Mahalanobis) Прасанта Чандра (29.6.1893 – 27.6.1972) – индийский статистик и экономист, член АНСССР (1958).

данных, как самостоятельно, так и как часть общего статистического анализа данных. В частности, этот вид анализа целесообразно применять для установления наличия статистических значимых связей между входными переменными (проверка выполнимости одного из допущений при построении регрессионной модели по экспериментальным данным), а также между входными и выходной(ыми) переменной(ыми). Для осуществления требуемых вычислений рекомендуется использовать пакет Microsoft Excel, или Statistica.

Рассмотрим следующую задачу определения наличия статистической связи между случайными величинами на основе корреляционного анализа и оценка коэффициента корреляции генеральной совокупности на основе значения коэффициента корреляции выборки.

Напомним, что корреляция или корреляционная зависимость — это статистическая взаимосвязь двух или более случайных величин. Корреляционный анализ – метод, позволяющий обнаружить зависимость между несколькими случайными величинами. Взаимосвязь между переменными описывается количественно, чтобы выявить уровень связи между переменными. Для этого вводится коэффициент корреляции.

Одна из наиболее распространенных задач статистического исследования состоит в изучении связи между выборками. Обычно связь между выборками носит не функциональный, а вероятностный (или стохастический) характер. В этом случае отсутствует строгая, однозначная зависимость между величинами.

Корреляционный анализ заключается в определении степени связи между двумя случайными величинами *X* и *Y*. В качестве меры тесноты такой связи используется коэффициент корреляции. Коэффициент корреляции оценивается по выборке объема *n* связанных пар наблюдений (*x, y*) из совместной генеральной совокупности *X* и *Y*. Для оценки степени взаимосвязи величин *X* и *Y*, измеренных в количественных шкалах, используется коэффициент линейной корреляции $r_в$, предполагающий, что выборки *X* и *Y* распределены по нормальному закону.

1. Для определения связи переменных используется *коэффициент корреляции,* который определяется по формуле (9):

$$r = \frac{\sum \left(x_{1,i} - \overline{x_1}\right) \cdot \left(x_{2,i} - \overline{x_2}\right)}{\sqrt{\sum \left(x_{1,i} - \overline{x_1}\right)^2} \cdot \sqrt{\sum \left(x_{2,i} - \overline{x_2}\right)^2}}. \qquad (9)$$

где $x_1$ и $x_2$ – заданные переменные, а $\overline{x_1}$ и $\overline{x_2}$ – их средние значения. Коэффициент является линейным и изменяется в пределах [-1; 1]. Он определяет степень, тесноту линейной связи между величинами: -1 – строгая обратная линейная зависимость; +1 – строгая прямая линейная зависимость.

При большом числе наблюдений, когда коэффициенты корреляции необходимо последовательно вычислять для нескольких выборок, для удобства получаемые коэффициенты сводят в таблицы, называемые корреляционными матрицами. *Корреляционная матрица* — это квадратная таблица, в которой на

пересечении соответствующих строки и столбца находится коэффициент корреляции между соответствующими параметрами.

2. Чтобы доказать гипотезу, что коэффициент корреляции значимо отличается от 0 вычисляется значение тестовой статистики по формуле (10):

$$\xi = \left( 0.5 \cdot \ln\left(\frac{1+r}{1-r}\right) - \frac{|r|}{2\cdot(n-1)} \right) \cdot \sqrt{n-3}\,. \qquad (10)$$

которая сравнивается с табличным значением критерия распределения Стьюдента при бесконечном числе опытов и вероятностью 0.95. Он равен 1.96.

3.В случае, если тестовая статистика больше критерия Стьюдента, то коэффициент корреляции значимо отличается от 0, в противном случае – не значимо.

***Слайд 9***

**Пример выполнения работы**

Исследуются 3 характеристики автомобиля: количество лет, пробег (тыс. км) и количество поломок. Обозначим их, соответственно, $x_1$, $x_2$ и $y$. Проводятся 18 экспериментов и получившиеся данные вводятся в таблицу 1 экспериментальных данных.

| Параметр | | |
|---|---|---|
| $x_1$ | $x_2$ | $y$ |
| 1 | 8 | 1 |
| 2 | 11 | 1 |
| 3 | 18 | 2 |
| 4 | 24 | 2 |
| 5 | 29 | 2 |
| 6 | 37 | 3 |
| 7 | 41 | 3 |
| 8 | 43 | 3 |
| 9 | 45 | 3 |
| 10 | 55 | 4 |
| 11 | 65 | 5 |
| 12 | 66 | 6 |
| 13 | 69 | 9 |
| 14 | 77 | 12 |
| 15 | 99 | 15 |
| 16 | 125 | 17 |
| 17 | 128 | 17 |
| 18 | 140 | 20 |

1. Необходимо рассчитать средние значения для каждой из переменных:

$$\overline{x_1} = \frac{\sum x_{1,i}}{n} = 9.5\,;\ \overline{x_2} = \frac{\sum x_{2,i}}{n} = 59.5 \text{ и } \overline{y} = \frac{\sum y_i}{n} = 6.9\,.$$

Образуем все возможные пары значений результатов измерений и для каждой пары найдём коэффициент корреляции, используя формулу (9).

Между $x_1$ и $y$: $r = \dfrac{\sum\left(x_{1,\mathrm{i}} - \overline{x_1}\right)\cdot\left(y_{\mathrm{i}} - \overline{y}\right)}{\sqrt{\sum\left(x_{1,\mathrm{i}} - \overline{x_1}\right)^2}\cdot\sqrt{\sum\left(y_{\mathrm{i}} - \overline{y}\right)^2}} = 0.874\,.$

Между $x_2$ и $y$: $r = \dfrac{\sum\left(x_{2,\mathrm{i}} - \overline{x_2}\right)\cdot\left(y_{\mathrm{i}} - \overline{y}\right)}{\sqrt{\sum\left(x_{2,\mathrm{i}} - \overline{x_2}\right)^2}\cdot\sqrt{\sum\left(y_{\mathrm{i}} - \overline{y}\right)^2}} = 0.739\,.$

Между $x_1$ и $x_2$: $r = \dfrac{\sum\left(x_{1,\mathrm{i}} - \overline{x_1}\right)\cdot\left(x_{2,\mathrm{i}} - \overline{x_2}\right)}{\sqrt{\sum\left(x_{1,\mathrm{i}} - \overline{x_1}\right)^2}\cdot\sqrt{\sum\left(x_{2,\mathrm{i}} - \overline{x_2}\right)^2}} = 0.815\,.$

Из формулы (9) видно, что при перестановке переменных местами, результат остается прежним. Об этом свидетельствует произведение в числителе и квадраты в знаменателе. Следовательно, для пар $\{x_2; x_1\}$, $\{y; x_1\}$ и $\{y; x_2\}$ значения вычислены.

Получившиеся результаты заносятся в таблицу 2.

Таблица 2 – Корреляционная матрица

| | $x_1$ | $x_2$ | $y$ |
|---|---|---|---|
| $x_1$ | 1,000 | 0.815 | 0.874 |
| $x_2$ | 0.815 | 1,000 | 0.739 |
| $y$ | 0.874 | 0.739 | 1,000 |

2. Проверим гипотезу, что коэффициент корреляции значимо отличается от 0. Вычислим тестовые статистики по формуле (10):

Пара $\{x_1; y\}$ и $\{y; x_1\}$:

$$\xi = \left(0.5 \cdot \ln\left(\frac{1+r}{1-r}\right) - \frac{|r|}{2 \cdot (n-1)}\right) \cdot \sqrt{n-3} = 5.128$$

Пара $\{x_2; y\}$ и $\{y; x_2\}$:

$$\xi = \left(0.5 \cdot \ln\left(\frac{1+r}{1-r}\right) - \frac{|r|}{2 \cdot (n-1)}\right) \cdot \sqrt{n-3} = 3.58$$

Пара $\{x_1; x_2\}$ и $\{x_2; x_1\}$:

$$\xi = \left(0.5 \cdot \ln\left(\frac{1+r}{1-r}\right) - \frac{|r|}{2 \cdot (n-1)}\right) \cdot \sqrt{n-3} = 4.3$$

3. Сравниваем каждое из этих значений с критерием Стьюдента, получаем, что все коэффициенты корреляции значимо отличаются от 0.