Из кн.: Кузнецов В.М. **Основы научных исследований в животноводстве.** Киров: Зональный НИИСХ Северо-Востока, 2006.- 568 с.

# 14. КОРРЕЛЯЦИОННЫЙ АНАЛИЗ

## 14.1. Понятие «корреляция»

У животных часто имеет место сопряженная (совместная) изменчивость признаков, например, удоя и жирномолочности, яйценоскости и веса яйца и т.д. Совместную изменчивость разных признаков называют «*корреляция*» (co-relation -связь, соотношение) и обозначают символом «r»*.

Корреляционная связь является, во-первых, вероятностной - изменение одного признака у ряда особей на *определенную* величину сопровождается изменениями другого признака на *различные* (варьирующие) значения; во-вторых, статистической - проявляется лишь в среднем для всей выборки; в отношении отдельных наблюдений она очень неполная и неточная (см. также [116]).

Корреляционную связь следует отличать от *функциональной*. При последней изменение одного показателя (аргумента) на определенную величину приводит к изменению другого показателя (функции) тоже на определенную величину (как, например, в формуле площади круга - $S = \pi R^2$, здесь R - радиус круга; $\pi = 3{,}14...$).

Корреляция не вскрывает причины связи. Она дает лишь оценку силы, или тесноты связи между переменными. Однако знать корреляции важно. Так, при селекции животных никогда не отбирают только по одному признаку. Более того, это невозможно, т.к. селекционируются особи. А особь - это десятки признаков, которые необходимо учитывать при отборе. Если бы корреляция между признаками отсутствовала, то селекция была бы проще. Отбор мог бы проводиться независимо и отдельно по каждому признаку.

Корреляцию можно рассчитать для любой пары признаков. Однако должно быть «биологическое обоснование» взаимосвязи. Например, как биологически объяснить корреляцию между инвентарными номерами быков и удоем их дочерей?

---

* Понятие «корреляция» в современном значении появилось в середине XIX века благодаря работам сэра Френсиса Гальтона (двоюродного брата Чарльза Дарвина) и Карла Пирсона. Через 20 лет после того, как Френсис Гальтон впервые приступил к решению проблемы вероятностной взаимосвязи, К.Пирсон обнаружил, что эта задача была решена 50 лет назад французским астрономом А Бравэ в статье об ошибках в определении нахождения точки в пространстве.

## 14.2. Задачи корреляционного анализа

Корреляционный анализ призван решать следующие задачи:

1. *Измерение степени связности двух и более переменных*. Наши общие знания об объективно существующих причинных связях должны дополняться научно обоснованными знаниями о *количественной* мере зависимости между переменными. Данный пункт подразумевает *верификацию* уже известных связей.
2. *Обнаружение неизвестных причинных связей*. Корреляционный анализ непосредственно не выявляет причинных связей между переменными, но устанавливает силу этих связей и их значимость. Причинный характер выясняют с помощью логических рассуждений, раскрывающих механизм связей.
3. *Отбор факторов, существенно влияющих на признак*. Самые важные те факторы, которые сильнее всего коррелируют с изучаемыми признаками.

## 14.3. Характер и сила связи

Коэффициенты корреляции могут варьировать от -1 до +1. При положительных корреляциях зависимость между признаками прямая: с увеличением одного увеличивается и другой признак. При отрицательных корреляциях зависимость обратная: увеличение одного признака приводит к уменьшению другого. Нулевая корреляция свидетельствует о независимой изменчивости двух признаков - нет *линейной* связи между признаками. Однако вполне возможно, что при этом существует *нелинейная* связь.

Коэффициент корреляции на уровне 0,5 представляется достаточно высоким. Можно даже полагать, что при такой корреляции совпадение вариации двух переменных должно быть в 50% случаев. В действительности это не так. Степень *линейной зависимости*, «связности», в вариации двух переменных более точно измеряется квадратом коэффициента корреляции - коэффициентом *детерминации* ($r^2$).

Коэффициент детерминации изменяется от 0 до 1. В случае прямолинейной связи коэффициент детерминации указывает на долю изменчивости переменной Y, которая обусловлена изменчивостью переменной X (и наоборот). Тогда $1-r^2$ - это *остаточная* доля изменчивости признака Y, обусловленная всеми

другими, не учтенными в эксперименте причинами. Так, если коэффициент корреляции между двумя признаками равен 0,5, то только 25% изменчивости одного признака объясняется изменчивостью другого признака (степень связности). По остальной же части изменчивости соотношение между признаками чисто случайное. Таким образом, корреляция ≥0,7 свидетельствует о тесной связи, порядка 0,5…0,6 – о средней и <0,5 - указывает на слабую связь.

### 14.4. Виды корреляций

Корреляции могут быть: относительно характера проявления статистической связи - *положительными* и *отрицательными*; по форме связи - *линейными* и *нелинейными*; по числу переменных - *простыми* (парными), *множественными* (между более чем двумя переменными) и *частными* - между двумя переменными при «фиксированном» влиянии остальных переменных.

Относительно природы источника совместной изменчивости различают корреляции *фенотипические*, *паратипические* (средовые) и *генетические* (см. также [101,131,137,143]).

Высокая паратипическая корреляция указывает на то, что значения признаков можно повысить, улучшая одни и те же условия среды (кормление, содержание).

Для селекционера важно, в какой степени фенотипическая связь между признаками обусловлена средой и в какой – наследственностью. При высокой генетической корреляции отбор животных можно ограничить только одним из признаков, как правило тем, который проще измерять. В этом случае можно сократить затраты на контроль других признаков. Их улучшение пойдет «само собой» при отборе по контролируемому признаку.

Высокие генетические и паратипические корреляции при высокой фенотипической свидетельствуют о том, что на совместную фенотипическую изменчивость признаков одновременно и очень заметно влияют как средовые, так и генетические факторы.

Часто бывает, что высокая фенотипическая корреляция не сопровождается столь же высокой генетической. В этом случае отбор по фенотипу одного признака приведет только к незначительному одновременному улучшению другого признака.

Генетические и паратипические корреляции могут различаться и по знаку. Различие в знаках означает, что источники изменчивости, обусловленные наследственностью и средой, влияют на признаки посредством различных физиологических механизмов.

## 14.5. Простая линейная корреляция

Выше отмечалось, что корреляция измеряет *совместную* изменчивость двух (и более) признаков. Однако непосредственно сравнить изменчивости признаков нельзя, т.к. они, как правило, выражаются в разных единицах измерения. Эту проблему решил Карл Пирсон. В качестве меры линейной зависимости двух признаков X и Y он предложил усредненное произведение нормированных (стандартизированных) отклонений:

$$r_{xy} = \frac{\sum_{i=1}^{n} u_{x_i} u_{y_i}}{n},$$

где $n$ - число животных; $u_i$ - отклонение продуктивности i-го животного от среднего значения, выраженное в долях сигмы:

$$u_{x_i} = \frac{x_i - \bar{x}}{\hat{\sigma}_x} \quad \text{и} \quad u_{y_i} = \frac{y_i - \bar{y}}{\hat{\sigma}_y}.$$

**Ковариансa**. Фактической мерой совместной изменчивости признаков X и Y является ковариансa, обозначают как Cov(XY) или $\sigma_{xy}$. *Ковариансa* - это отношение суммы произведений отклонений продуктивности i-го животного по каждому признаку от соответствующих средних ($SP_{xy}$) к числу степеней свободы ($df$). Для выборочной оценки используют различные формулы:

$$\hat{\sigma}_{xy} = \frac{SP_{xy}}{df} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{n-1} =$$

$$= \frac{1}{n-1}\left(\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}\right) = \frac{n\sum x_i y_i - \sum x_i \sum y_i}{n(n-1)} =$$

$$= \frac{\sum x_i y_i - n\bar{x}\,\bar{y}}{n-1}.$$

Последнее выражение наиболее удобно для расчетов.

Ковариация переменной с самой собой дает вариансу:

$$\hat{\sigma}_{xx} = \frac{\sum_{i=1}^{n} x_i x_i - n\,\bar{x}\,\bar{x}}{n-1} = \hat{\sigma}_x^2 .$$

Ковариация может варьировать от $-\infty$ до $+\infty$.

**Пример 14.1.** Пусть имеются данные по живой массе бычков при рождении (X) и последующей скорости роста (Y):

| Номер бычка (i) | $x_i$, кг | $y_i$, г/сутки | $x_i^2$ | $y_i^2$ | $x_i y_i$ |
|---|---|---|---|---|---|
| 1 | 40 | 1000 | 1600 | 1000000 | 40000 |
| 2 | 42 | 900 | 1764 | 810000 | 37800 |
| 3 | 35 | 850 | 1225 | 722500 | 29750 |
| 4 | 36 | 950 | 1296 | 902500 | 34200 |
| 5 | 45 | 920 | 2025 | 846400 | 41400 |
| 6 | 47 | 950 | 2209 | 902500 | 44650 |
| 7 | 40 | 810 | 1600 | 656100 | 32400 |
| 8 | 43 | 870 | 1849 | 756900 | 37410 |
| 9 | 41 | 930 | 1681 | 864900 | 38130 |
| 10 | 38 | 870 | 1444 | 756900 | 33060 |
| Σ | 407 | 9050 | 16693 | 8218700 | 368800 |
| Среднее | 40,7 | 905 | - | - | - |

$x_i$ и $y_i$ являются фенотипической ценностью i-го животного по живой массе при рождении и скорости роста. Сумма их произведений есть:

$$\sum_{i=1}^{10} x_i y_i = 40(1000) + 42(900) + \ldots + 38(870) = 368800,$$

и выборочная оценка ковариансы равна

$$\hat{\sigma}_{xy} = \frac{368800 - 10(40{,}7)(905)}{9} = 51{,}67 \text{ кг/(г/сутки)}.$$

**Корреляция**. Проблема с ковариансой, как меры связи между переменными, состоит в том, что ее размерность зависит от шкалы измерение. Например, если бы скорость роста измерялась в кг/сутки вместо г/сутки, то $\sigma_{xy}$ была бы 0,05167. Следовательно, размер ковариансы без единиц измерения не имеет никакого значения.

*Корреляция* - более полезная мера совместной изменчивости. Она стандартизирована, поэтому варьирует в диапазоне от -1 до +1. Коэффициент корреляции - это отношение ковариансы двух переменных к произведению их стандартных отклонений:

$$r_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}.$$

Для выборочной оценки коэффициента парной корреляции используют различные формулы:

$$\hat{r}_{xy} = \frac{\hat{\sigma}_{xy}}{\hat{\sigma}_x \hat{\sigma}_y} = \frac{\hat{\sigma}_{xy}}{\sqrt{\hat{\sigma}_x^2 \hat{\sigma}_y^2}} =$$

$$= \frac{\sum x_i y_i - \frac{\sum x_i \sum y_i}{n}}{\sqrt{\left[\sum x_i^2 - \frac{(\sum x_i)^2}{n}\right]\left[\sum y_i^2 - \frac{(\sum y_i)^2}{n}\right]}} =$$

$$= \frac{n\sum x_i y_i - \sum x_i \sum y_i}{\sqrt{[n\sum x_i^2 - (\sum x_i)^2][n\sum y_i^2 - (\sum y_i)^2]}} =$$

$$= \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)(\sum y_i^2 - n\bar{y}^2)}}.$$

Последнее выражение наиболее удобно для расчетов.

Для **примера 13.1** коэффициент парной корреляции составил

$$\hat{r}_{xy} = \frac{\hat{\sigma}_{xy}}{\sqrt{\hat{\sigma}_x^2 \hat{\sigma}_y^2}} = \frac{51{,}67}{\sqrt{(14{,}233)(3161{,}11)}} = 0{,}244 \quad \text{или}$$

$$\hat{r}_{xy} = \frac{\sum x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)(\sum y_i^2 - n\bar{y}^2)}}$$

$$= \frac{368800 - 10 \times 40{,}7 \times 905}{\sqrt{(16693 - 10 \times 40{,}7^2)(8218700 - 10 \times 905^2)}} = 0{,}244.$$

## 14.6. Проверка значимости

Оценка коэффициента корреляции является выборочной, т.к. она вычисляется на основе выборки из генеральной совокупности.

Поэтому коэффициент корреляции имеет свою ошибку. Эта ошибка является мерой расхождения между *оценкой* корреляции по выборочным данным ($\hat{r}$) и *истинной* корреляцией в генеральной совокупности - r.

Если n>100 и оценка $\hat{r}$ не очень высокая, то ошибку $m_{\hat{r}}$ рассчитывают по формуле:

$$m_{\hat{r}} = \frac{1-\hat{r}^2}{\sqrt{n}}.$$

Для малых выборок применяют формулу

$$m_{\hat{r}} = \frac{\sqrt{1-\hat{r}^2}}{\sqrt{n-2}}.$$

Проверку значимости начинают с формулировки нулевой гипотезы, которая заключается в допущении, что истинный коэффициент корреляции в генеральной совокупности равен нулю ($H_0: r=0$). Альтернативная гипотеза состоит в том, что коэффициент корреляции в генеральной совокупности отличен от нуля ($H_1: r\neq 0$). Если проверка покажет, что нулевая гипотеза не приемлема, то выборочный коэффициент корреляции ($\hat{r}$) значимо отличается от нуля и нулевую гипотезу отвергают. И наоборот, если на основе фактического критерия нулевую гипотезу принимают, т.е. $\hat{r}$ лежит в зоне *случайного* рассеяния, то нет оснований считать сомнительным предположение об отсутствии связи между переменными в генеральной совокупности.

Фактический критерий значимости (К) для коэффициента корреляции ($t_{\hat{r}}$) рассчитывают из отношения:

$$t_{\hat{r}} = \frac{|\hat{r}|}{m_{\hat{r}}}.$$

$t_{\hat{r}}$-статистику сравнивают с критическим значением, $t_{\alpha;df}$, при уровне значимости $\alpha$ и степени свободы $df = n-2$ (находят по табл. А.8 Приложения А; двусторонняя область). Если $t_{\hat{r}} \geq t_{\alpha;df}$, то нулевую гипотезу на уровне значимости $\alpha$ *отвергают*, т.е. связь между переменными считают значимой, допуская ошибку в $\alpha$ % случаев (см. также табл. А.12 Приложения А). При $t_{\hat{r}} < t_{\alpha;df}$ нулевую гипотезу *принимают* и говорят, что связь между переменными не подтверждается.

Для **примера 14.1** получим:

$$m_{\hat{r}} = \frac{\sqrt{1-0{,}244^2}}{\sqrt{10-2}} = 0{,}343 \quad \text{и} \quad t_{\hat{r}} = \frac{0{,}244}{0{,}343} = 0{,}71.$$

Число степеней свободы: $df = n - 2 = 10 - 2 = 8$.

Критическое значение (табл. А.8): $t_{0{,}05;8} = 2{,}31$.

**Вывод**. Корреляционный анализ выявил слабую взаимосвязь между живой массой телят при рождении и среднесуточными привесами до годовалого возраста. Значение $t_{\hat{r}}$-статистики свидетельствовало о том, что отклонение выборочной оценки корреляции ($\hat{r}$) от аналогичного параметра в генеральной совокупности ($r = 0$) можно приписать случайной вариации. Данные выборки характеризуют нулевую гипотезу как весьма возможную и правдоподобную. Другими словами, гипотеза об отсутствии связи между живой массой телят при рождении и среднесуточным привесом до годовалого возраста не вызывает возражения.

Для значимого коэффициента корреляции ($\hat{r}$) определяют доверительный интервал (интервальную оценку), который с заданной надежностью ($P = 1 - \alpha$) «накроет» неизвестный генеральный коэффициент корреляции ($r$). Для построения такого интервала необходимо знать выборочное распределение коэффициента корреляции $\hat{r}$, которое при $r \neq 0$ несимметрично и очень медленно (с ростом n) сходится к нормальному распределению. Поэтому Р.А. Фишер в 1921 г. предложил z-преобразование случайной величины $\hat{r}$:

$$\hat{z} = \frac{1}{2} \ln \frac{1+\hat{r}}{1-\hat{r}},$$

которое не зависит ни от r, ни от n (ln - натуральный логарифм с основанием e=2,71828…). Если n>50, то распределение $\hat{z}$ близко к нормальному с математическим ожиданием и вариансой:

$$\mu_z \approx \frac{1}{2} \times \ln \frac{1+r}{1-r} + \frac{r}{2(n-1)},$$

$$\sigma_z^2 = \frac{1}{n-3}.$$

Поэтому вначале строят доверительный интервал для $\mu_z$:

$$(\hat{z} - t_{\alpha;df}\,\sigma_z) < \mu_z < (\hat{z} + t_{\alpha;df}\,\sigma_z).$$

При определении доверительных границ для r, т.е. для перехода от $\hat{z}$ к $\hat{r}$ используют формулу:

$$\hat{r}_i = \text{th}\,\hat{z}_i = \frac{e^{\hat{z}_i} - e^{-\hat{z}_i}}{e^{\hat{z}_i} + e^{-\hat{z}_i}},$$

где $\text{th}\,\hat{z}_i$ - гиперболический тангес $\hat{z}_i$ ($\hat{z}_i$ - минимальная (максимальная) граница для $\mu_z$).

Z-преобразование Фишера используют также для проверки существенности (значимости) различия двух коэффициентов корреляции ($\hat{r}_1$ и $\hat{r}_2$), полученных по выборкам объемов $n_1$ и $n_2$, т.е. для проверки гипотезы $H_0 : r_1 = r_2$; альтернатива - $H_1 : r_1 \neq r_2$. Статистический критерий имеет вид:

$$t_{z_1 - z_2} = \frac{|\hat{z}_1 - \hat{z}_2|}{\sqrt{1/(n_1 - 3) + 1/(n_2 - 3)}}.$$

Гипотезу $H_0$ отвергают на уровне значимости $\alpha$, если $t_{z_1 - z_2} \geq t_{\alpha,df}$, и принимают, если $t_{z_1 - z_2} < t_{\alpha,df}$ ($t_{\alpha,df}$ находят по табл. А.8 Приложения А; двусторонний критерий, $df = n_1 + n_2 - 4$).

Чем меньше выборка, тем сильнее оценка коэффициента корреляции подвержена случайному влиянию. Коэффициенты корреляции, вычисленные по различным выборкам одной совокупности, могут различаться даже по знаку. Поэтому (1) следует с осторожностью подходить к обобщению результатов анализа, выполненного на небольшой выборке, и (2) не всегда правомерно распространять результаты и выводы эксперимента на более крупные совокупности животных (например, породу).

Коэффициенты корреляции не являются «аддитивными». Например, коэффициент корреляции, вычисленный по нескольким объединенным выборкам, не совпадает с «усредненной корреляцией» по этим выборкам. Коэффициенты корреляции не могут быть просто усреднены. Если интерес представляет обобщенный коэффициент корреляции, то следует преобразовать коэффициенты корреляции по выборкам в такую меру зависимости, которая будет аддитивной. Например, до того, как усреднить коэффициенты корреляции, их можно возвести в квадрат, получить коэффициенты детерминации, которые уже будут аддитивными.

## 14.7. Частная корреляция

На величину коэффициента парной корреляции могут оказывать влияние другие переменные. Интенсивность связи в «чистой» форме определяют с помощью коэффициента *частной* корреляции. В этом случае сопряженную вариацию между двумя переменными оценивают при фиксировании (исключении) влияния остальных переменных (см. также [116]).

Расчет коэффициента частной корреляции базируется на оценках коэффициентов парных корреляций. Так, для трех признаков выборочный коэффициент *частной* корреляции рассчитывают из отношения:

$$\hat{r}_{12.3} = \frac{\hat{r}_{12} - \hat{r}_{13}\ \hat{r}_{23}}{\sqrt{(1-\hat{r}_{13}^2)(1-\hat{r}_{23}^2)}},$$

где $\hat{r}_{12.3}$ - корреляция между признаками 1 и 2 при элиминации влияния на эту связь признака 3 (если есть основание полагать, что связь между признаками 1 и 2 возникает за счет связи с признаком 3).

Путем соответствующих перестановок цифр в субиндексах можно записать формулы для $\hat{r}_{13.2}$ и $\hat{r}_{23.1}$. Точки между цифрами отделяют признаки, корреляции с которыми элиминируются.

Частная корреляция при четырех переменных:

$$\hat{r}_{12.34} = \frac{\hat{r}_{12.4} - \hat{r}_{13.4}\ \hat{r}_{23.4}}{\sqrt{(1-\hat{r}_{13.4}^2)(1-\hat{r}_{23.4}^2)}}.$$

Обобщение на любое число переменных:

$$\hat{r}_{12.3...m} = \frac{\hat{r}_{12.4...m} - \hat{r}_{13.4..m}\ \hat{r}_{23.4...m}}{\sqrt{(1-\hat{r}_{13.4...m}^2)(1-\hat{r}_{23.4...m}^2)}}.$$

Расчет коэффициента частной корреляции порядка m сводится к оценке коэффициентов частной корреляции порядка m-1. Сначала рассчитывают коэффициенты парной корреляции, а затем приступают к вычислению коэффициентов корреляций более высокого порядка. Частные коэффициенты корреляции также варьируют от –1 до +1.

**Пример 14.2**. Пусть у телят измеряют три признака: 1 - возраст; 2 - среднесуточный привес; 3 - уровень общего белка в крови. Связь общего белка сыворотки крови с возрастом составила $\hat{r}_{31}=0{,}25$, с привесом -

$\hat{r}_{23}=0{,}40$ и привеса с возрастом - $\hat{r}_{21}=0{,}50$. Привес и белок крови зависят от возраста. Поэтому требуется вычислить частный коэффициент корреляции привеса и белка при исключении влияния возраста:

$$\hat{r}_{23.1}=\frac{\hat{r}_{23}-\hat{r}_{12}\hat{r}_{13}}{\sqrt{(1-\hat{r}_{12}^2)(1-\hat{r}_{13}^2)}}=$$

$$=\frac{0{,}40-0{,}50\times 0{,}25}{\sqrt{(1-0{,}50^2)(1-0{,}25^2)}}\approx 0{,}32\,.$$

$\hat{r}_{23.1}^2\approx 0{,}10$, что свидетельствует о низкой связи привеса с процентом общего белка крови в пределах отдельных возрастных групп животных.

## 14.8. Множественная корреляция

В биозоотехнических исследованиях чаще всего встречаются сложные взаимосвязи между переменными. Для определения интенсивности или тесноты связи одной из переменных с совокупностью остальных переменных используют коэффициент *множественной* корреляции. Например, коэффициент корреляции $r_{y.12}$ показывает интенсивность связи при условии, что переменная Y одновременно зависит от переменных 1 и 2:

$$\hat{r}_{y.12}=\sqrt{\frac{\hat{r}_{y1}^2+\hat{r}_{y2}^2-2\hat{r}_{y1}\hat{r}_{y2}\hat{r}_{12}}{1-\hat{r}_{12}^2}}\,.$$

Коэффициенты множественной корреляции варьируют от 0 до 1. По их значениям *нельзя* сделать вывод о характере взаимосвязи, т.е. «+» или «-» корреляции между переменными. Только если все коэффициенты парной корреляции имеют одинаковый знак, то этот знак можно отнести также к коэффициенту множественной корреляции и утверждать о соответствующем характере множественной связи.

Используя матричную форму записи, выражение коэффициента множественной корреляции для любого числа объясняющих переменных можно получить из уравнения:

$$r_{y.12...n}=\sqrt{r'\,R^{-1}\,r}\,,$$

где

$$r = \begin{bmatrix} r_{y1} \\ r_{y2} \\ \vdots \\ r_{yn} \end{bmatrix}$$ - вектор корреляций признака Y с переменными 1,2,…,n;

$r'$ - трансформированный вектор $r$;

$$R = \begin{bmatrix} r_{11} & r_{12} & \cdots & r_{1n} \\ r_{21} & r_{22} & \cdots & r_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ r_{n1} & r_{n2} & \cdots & r_{nn} \end{bmatrix}$$ - корреляционная матрица для 1, 2,…, n переменных.

**Пример 14.3.**

Пусть: $r' = [0{,}9687 \;\; 0{,}4257 \;\; -0{,}5189]$;

$$R = \begin{bmatrix} 1 & 0{,}3620 & -0{,}5038 \\ 0{,}3620 & 1 & -0{,}3778 \\ -0{,}5038 & -0{,}3778 & 1 \end{bmatrix};$$

$$R^{-1} = \begin{bmatrix} 1{,}4049 & -0{,}2813 & 0{,}6015 \\ -0{,}2813 & 1{,}2228 & -0{,}3203 \\ 0{,}6015 & 0{,}3203 & 1{,}4240 \end{bmatrix}.$$

Тогда:

$$\hat{r}^2_{y.123} = [0{,}9687 \;\; 0{,}4257 \;\; -0{,}5189] \begin{bmatrix} 1{,}4049 & -0{,}2813 & 0{,}6015 \\ -0{,}2813 & 1{,}2228 & 0{,}3203 \\ 0{,}6015 & 0{,}3203 & 1{,}4240 \end{bmatrix} \begin{bmatrix} 0{,}9687 \\ 0{,}4257 \\ -0{,}5189 \end{bmatrix} = 0{,}9451;$$

$$\hat{r}_{y.123} = \sqrt{0{,}9451} = 0{,}9722$$

Высокое значение $\hat{r}_{y.123}$ свидетельствует о тесной связи признака Y с переменными 1, 2 и 3.

В заключение следует отметить, что корреляционный анализ можно применять тогда, когда данные наблюдений или эксперимента можно считать случайными и выбранными из совокупности, распределенной по многомерному нормальному закону. Если эти предпосылки нарушаются, то коэффициент корреляции не следует рассматривать как строгую меру взаимосвязей переменных.

## 14.9. Ранговая корреляция

Для вычисления парных корреляций необходимо, чтобы исходные данные были выражены достаточно точно и имели

*нормальное* распределение. Это не всегда возможно. Существуют признаки, которые с трудом поддаются точной оценке, например, балл за экстерьер. Кроме того, распределение одного или обоих признаков может быть очень неравномерным и неправильным. В таких случаях для количественной оценки связи между признаками используют метод *ранговой* корреляции Спирмена (этот и последующий метод относятся к *непараметрической статистике*; см.главу 12).

В данном методе необходимы не точные значения количественных признаков, а их ранги (порядковые номера животных по соответствующему признаку). Коэффициент ранговой корреляции является парным. Поэтому оценивается *соответствие* между двумя рядами порядковых номеров.

Ранги присваивают по нисходящей: от бóльшего значения к меньшему. Если встречаются два или более животных с одинаковыми (связными) значениями, то используют метод «средних рангов». Например, присвоили по живой массе ранги пяти лучшим животным. У следующих трех животных живая масса была по 420 кг. Необходимо усреднить ранги, *которые имели бы эти животные, если бы их значения различались*: (6+7+8)/3=7. Таким образом, всем трем животным присваивают ранг 7. Последующему, с живой массой ниже 420 кг, но выше, чем у остальных не ранжированных, присваивают ранг 9*.

Если обозначить ранги, соответствующие значениям переменной X, через v, а ранги, соответствующие значениям переменной Y, - через w, то коэффициент ранговой корреляции Спирмена ($r_s$) вычисляют по формуле:

$$\hat{r}_s = 1 - \frac{6\sum_{i=1}^{n}(v_i - w_i)^2}{n(n^2-1)};$$

где n - размер выборки.

Процедура проверки значимости коэффициента ранговой корреляции аналогична соответствующей процедуре для коэффициента парной корреляции.

---

* При наличии связных рангов в коэффициент ранговой корреляции вводят относительно сложную поправку, расчет которой дан в книге: Э. Фёрстера, Б. Рёнца «Методы корреляционного и регрессионного анализа» [116].

**Пример 14.4.** Сравнение результатов оценки племенной ценности 10 быков по качеству потомства методами CC и BLUP.

| Ранг по BLUP, v | Ранг по CC, w | D=v-w | $D^2$ |
|---|---|---|---|
| 1 | 6 | -5 | 25 |
| 2 | 5 | -3 | 9 |
| 3 | 1 | +2 | 4 |
| 4 | 4 | 0 | 0 |
| 5 | 2 | +3 | 9 |
| 6 | 7 | -1 | 1 |
| 7 | 8 | -1 | 1 |
| 8 | 10 | -2 | 4 |
| 9 | 3 | +6 | 36 |
| 10 | 9 | +1 | 1 |
| | Σ | 0 | 90 |

$$\hat{r}_s = 1 - \frac{6 \times 90}{10(10^2 - 1)} \approx 0{,}45;$$

$$m_{\hat{r}} = \frac{\sqrt{1 - 0{,}45^2}}{\sqrt{10 - 2}} = 0{,}316;$$

$$t_{\hat{r}} = \frac{0{,}45}{0{,}316} = 1{,}42.$$

Число степеней свободы: $df = n - 2 = 10 - 2 = 8.$

Критическое значение (табл. А.8): $t_{0{,}05;8} = 2{,}31.$

**Вывод**. Корреляция Спирмена указывала на значительное расхождение рангов быков, оцененных методами CC и BLUP. Это могло бы свидетельствовать о том, что методы по-разному классифицируют быков по племенной ценности. Однако $t_{\hat{r}} < t_{0{,}05;8}$. Поэтому нулевая гипотеза не может быть отвергнута. Различия в рангах племенной ценности одних и тех же быков, рассчитанной разными методами, данным экспериментом не доказаны. Для получения объективных результатов необходимо повторить исследование на бóльшем числе быков.

## 14.10. Коэффициент конкордации

В животноводстве существуют признаки, которые не поддаются точной количественной оценке. Это т.н. атрибутивные признаки. Например, ранжируют животных изучаемой выборки, приписывая каждому из них порядковый номер. Если число переменных больше двух, то в результате n животных имеют m рангов. Для проверки, согласованности этих m ранжировок друг с

другом, используют коэффициент *конкордации* Кендалла, W:

$$W = \frac{12\sum_i D_i^2}{m^2(n^3 - n)}.$$

При наличии связных рангов коэффициент конкордации вычисляют по формуле:

$$W = \frac{12\sum_i D_i^2}{m^2(n^3 - n) - mB},$$

где $D_i = \sum_{j=1}^{m} R_{ij} - \frac{\sum_j \sum_i R_{ij}}{n}$, при $i=1,2,...,n;\ j=1,2,...,m;$ - есть сумма рангов, приписанных i-ому животному выборки, минус среднее значение этой суммы рангов; m - число признаков-переменных, связь между которыми оценивается;

$B = \sum_{k=1}^{z} (B_k^3 - B_k)$, где $B_k$ - число неразличимых рангов в k-ой группе признаков.

Коэффициент W принимает значение в интервале от 0 до 1.

**Пример 14.5.** Пусть 3 специалиста оценивают (ранжируют) 6 одних и тех же животных. Результаты представлены в столбцах 2, 3, 4 табл. 32.

**32. Ранжирование 6 животных тремя специалистами**

| № жив-го ( i ) | Эксперт ( j) | | | Сумма рангов | | |
|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | $\sum_{j=1}^{3} R_{ij}$ | $D_i$ | $D_i^2$ |
| *1* | *2* | *3* | *4* | *5* | *6* | *7* |
| 1 | 1 | 2 | 1 | 4 | -6,5 | 42,25 |
| 2 | 2 | 1 | 3 | 6 | -4,5 | 20,25 |
| 3 | 4 | 4,5 | 3 | 11,5 | +1,0 | 1,00 |
| 4 | 5 | 4,5 | 6 | 15,5 | +5,0 | 25,00 |
| 5 | 3 | 3 | 3 | 9 | -1,5 | 2,25 |
| 6 | 6 | 6 | 5 | 17 | +6,5 | 42,25 |
| Сумма | 21 | 21 | 21 | 63 | - | 133,00 |

Сумма рангов для каждого i-го животного указана в столбце 5. Для определения D вначале вычисляют среднее значение по суммам рангов:

$$\frac{\sum_{j=1}^{3}\sum_{i=1}^{6} R_{ij}}{6} = \frac{63}{6} = 10{,}5.$$

Полученное среднее (10,5) вычитают из каждой i-ой суммы рангов, и разность записывают в столбец 6. Сумма квадратов разностей есть элемент числителя для W. Поправка на связность, B:

$$B = (2^3 - 2) + (3^3 - 3) = 30 .$$

Число стад n=6, число экспертов m=3. Тогда

$$W = \frac{12 \times 133}{3^2 (6^3 - 6) - 3 \times 30} = 0{,}8867.$$

Значимость коэффициента W проверяют критерием $\chi^2$:

$$\chi^2 = m(n-1)W = \\ = 3(6-1)0{,}8867 = 13{,}3 ,$$

с df=n-1 степенями свободы.

По табл. А.9 для α=0,05 и df=5 находим $\chi^2_{0,05;5} = 11{,}07$. Так как

$$\chi^2 = 13{,}3 > \chi^2_{0,05;5} = 11{,}07 ,$$

то с вероятностью 1-α нулевая гипотеза отвергается.

**Вывод** Оценку животных тремя экспертами на уровне значимости α=5% можно считать вполне согласованной.

Если вместо экспертов рассматривать признаки, то коэффициент W будет единой выборочной мерой связи.

## 14.11. Причины смещенных оценок

**Выбросы**. По определению, выбросы являются нетипичными, резко выделяющимися наблюдениями. Выбросы могут не только искусственно увеличить значение коэффициента корреляции, но также реально уменьшить существующую корреляцию.

Обычно считается, что выбросы представляют собой случайную ошибку, которую следует контролировать. К сожалению, не существует общепринятого метода автоматического удаления выбросов.

Некоторые исследователи применяют численные методы удаления выбросов. Например, исключаются значения, которые выходят за границы ±2 стандартных отклонений (и даже ±1,5 стандартных отклонений) вокруг выборочного среднего. В ряде случаев такая «чистка» данных абсолютно необходима. Однако определение выбросов субъективно, поэтому решение должно приниматься индивидуально в каждом эксперименте (с учетом особенностей эксперимента или «сложившейся практики»).

Следует заметить, что в некоторых случаях относительная частота выбросов к численности групп может быть исследована и разумно проинтерпретирована с точки зрения самой организации эксперимента.

**Неоднородность групп**. Отсутствие однородности в выборке также является фактором, смещающим (в ту или иную сторону) выборочную корреляцию. Допустим, что коэффициент корреляции вычислен по данным, которые поступили из двух различных экспериментальных групп, но это было проигнорировано при вычислениях. Далее, пусть действия экспериментатора в одной из групп увеличивают значения обоих коррелированных признаков, и, таким образом, данные каждой группы сильно различаются. В подобных ситуациях *высокая корреляция может быть следствием разбиения данных на две группы*, а вовсе не отражать «истинную» зависимость между двумя переменными (которая может практически отсутствовать).

Если такое явление имеет место, то необходимо разделить данные на «подмножества» и вычислить корреляции отдельно для каждого множества. Если неясно, как определить подмножества, то следует применить многомерные методы разведочного анализа (например, *кластерный анализ*).

**Нелинейная зависимость**. Другим возможным источником трудностей, связанным с линейной корреляцией, является форма зависимости. Корреляция Пирсона хорошо подходит для описания линейной зависимости. Отклонения от линейности приводит к смещенной оценке коэффициента корреляции, даже если имеют место очень тесные связи между переменными.

Что делать, если корреляция сильная, однако зависимость явно нелинейная? К сожалению, не существует простого ответа на данный вопрос, так как не имеется естественного обобщения коэффициента корреляции Пирсона на случай нелинейных зависимостей. Однако, если кривая зависимости монотонна (монотонно возрастает или, напротив, монотонно убывает), то можно преобразовать одну или обе переменные, чтобы сделать зависимость линейной, а затем уже вычислить корреляцию между преобразованными величинами. Для этого часто используется логарифмическое преобразование.

Другой подход состоит в использовании непараметрической корреляции (например, корреляции Спирмена). Иногда этот метод

приводит к успеху, хотя непараметрические корреляции чувствительны только к упорядоченным значениям переменных, например, по определению, они пренебрегают монотонными преобразованиями данных. К сожалению, два самых точных метода исследования нелинейных зависимостей непросты и требуют хорошего навыка «экспериментирования» с данными. Эти методы состоят в следующем:

1. Нужно попытаться найти функцию, которая наилучшим способом описывает данные и проверить ее «степень согласия» с данными (используя хи-квадрат).
2. Можно некоторой «группирующей переменной» дифференцировать данные, а затем применить дисперсионный анализ.

**Построчное и попарное удаление пропущенных данных**. При *построчном удалении* наблюдений с пропусками удаляется вся строка, в которой имеется хотя бы одно пропущенное значение. Этот способ приводит к «правильной» корреляционной матрице в том смысле, что все коэффициенты вычисляются по одному и тому же множеству наблюдений. Однако, если пропущенные значения распределены случайным образом в переменных, то данный метод может привести к тому, что в рассматриваемом множестве данных не останется ни одного неисключенного наблюдения (в каждой строке наблюдений встретится, по крайней мере, одно пропущенное значение).

Чтобы избежать подобной ситуации, используют способ, называемый *попарным удалением*. В этом способе учитывают только пропуски в каждой выбранной паре переменных и игнорируют пропуски в других переменных. Корреляцию между парой переменных вычисляют по наблюдениям, где нет пропусков. Во многих ситуациях, особенно, когда число пропусков относительно мало, скажем 10%, и пропуски распределены достаточно хаотично, этот метод не приводит к серьезным ошибкам.

Например, в систематическом смещении (сдвиге) оценки может «скрываться» систематическое расположение пропусков, являющееся причиной различия коэффициентов корреляции, построенных по разным подмножествам.

Другая проблема, связанная с корреляционной матрицей, вычисленной при попарном удалении пропусков, возникает при использовании этой матрицы в других видах анализа (например, множественная регрессия, факторный или кластерный анализы). В них предполагают, что корреляционная матрица «правильная» с определенным уровнем состоятельности и «соответствия» различных коэффициентов. Использование матрицы с «плохими» (смещенными) оценками приводит к тому, что программа либо не в состоянии анализировать такую матрицу, либо результаты будут ошибочными. Поэтому, если применяется попарный метод исключения пропущенных данных, то необходимо проверить, имеются или нет систематические закономерности в распределении пропусков (отсутствующих значений).

Если попарное исключение пропущенных данных не приводит к какому-либо систематическому сдвигу в оценках, то все эти статистики будут похожи на аналогичные статистики, вычисленные при построчном способе удаления пропусков. Если наблюдается значительное различие, то есть основание предполагать наличие сдвига в оценках. Например, если среднее (или стандартное отклонение) значение переменной A, которое использовалось при вычислении ее корреляции с переменной B, много меньше среднего (или стандартного отклонения) тех же значений переменной A, которые использовались при вычислении ее корреляции с переменной C, то имеются все основания ожидать, что эти две корреляции (A-B и A-C) основаны на разных подмножествах данных, и, таким образом, в оценках корреляций имеется сдвиг, вызванный неслучайным расположением пропусков в значениях переменных.

**Подстановка среднего значения**. Другим общим методом, позволяющим избежать потери наблюдений при построчном способе удаления наблюдений с пропусками, является замена средним (для каждой переменной пропущенные значения заменяются средним значением этой переменной). Подстановка среднего вместо пропусков имеет свои преимущества и недостатки в сравнении с попарным способом удаления

пропусков. Основное преимущество в том, что он дает состоятельные оценки, однако имеет следующие недостатки:

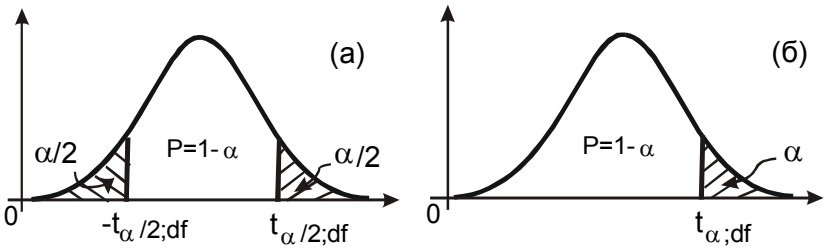
- искусственно уменьшается разброс данных - чем больше пропусков, тем больше данных, совпадающих со средним значением;
- так как пропущенные данные заменяются искусственно созданными «средними», то корреляции могут сильно уменьшиться.

**Ложные корреляции**. По коэффициентам корреляции нельзя строго доказать причинной зависимости между переменными. Однако можно определить ложные корреляции, т.е. корреляции, которые обусловлены влияниями «других», остающихся вне поля зрения исследователя, переменных (которые влияют на коррелируемые переменные). При «контролировании» (исключении) этих переменных исходная корреляция либо исчезнет, либо, возможно, даже изменит свой знак. Основная проблема ложной корреляции состоит в том, что исследователь не знает, кто является ее «агентом». Тем не менее, исследователь может воспользоваться частными корреляциями, чтобы контролировать (частично исключая) влияние определенных переменных.

Приложение

## А.8. Критические значения t-распределения Стьюдента

(здесь и далее P - доверительная вероятность)

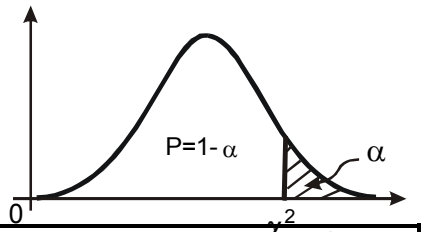



| df | Уровень значимости (ошибка, α) | | | | |
|---|---|---|---|---|---|
| | Двусторонняя критическая область (а) | | | | |
| | **0,100** | **0,050** | **0,020** | **0,010** | **0,001** |
| | Односторонняя критическая область (б) | | | | |
| | **0,050** | **0,025** | **0,010** | **0,005** | **0,0005** |
| **1** | 6,314 | 12,706 | 31,821 | 63,657 | 637 |
| **2** | 2,920 | 4,303 | 6,965 | 9,925 | 31,598 |
| **3** | 2,353 | 3,182 | 4,541 | 5,841 | 12,941 |
| **4** | 2,132 | 2,776 | 3,747 | 4,604 | 8,610 |
| **5** | 2,015 | 2,571 | 3,365 | 4,032 | 6,859 |
| **6** | 1,943 | 2,447 | 3,143 | 3,707 | 5,959 |
| **7** | 1,895 | 2,365 | 2,998 | 3,499 | 5,405 |
| **8** | 1,860 | 2,306 | 2,896 | 3,355 | 5,041 |
| **9** | 1,833 | 2,262 | 2,821 | 3,250 | 4,781 |
| **10** | 1,812 | 2,228 | 2,764 | 3,169 | 4,587 |
| **11** | 1,796 | 2,201 | 2,718 | 3,106 | 4,437 |
| **12** | 1,782 | 2,179 | 2,681 | 3,055 | 4,318 |
| **13** | 1,771 | 2,160 | 2,650 | 3,012 | 4,221 |
| **14** | 1,761 | 2,145 | 2,624 | 2,977 | 4,140 |
| **15** | 1,753 | 2,131 | 2,602 | 2,947 | 4,073 |
| **20** | 1,725 | 2,086 | 2,528 | 2,845 | 3,850 |
| **25** | 1,708 | 2,060 | 2,485 | 2,787 | 3,725 |
| **30** | 1,697 | 2,042 | 2,457 | 2,750 | 3,646 |
| **35** | 1,690 | 2,030 | 2,432 | 2,724 | 3,591 |
| **40** | 1,684 | 2,021 | 2,408 | 2,704 | 3,551 |
| **50** | 1,676 | 2,008 | 2,384 | 2,678 | 3,496 |
| **100** | 1,661 | 1,982 | 2,360 | 2,625 | 3,390 |
| **∞** | 1,645 | 1,960 | 2,326 | 2,576 | 3,291 |

**Примечание**. В последней строке даны значения нормированной случайной величины $t = u \sim N(0;1)$.

Из кн.: Кузнецов В.М. **Основы научных исследований в животноводстве.** Киров: Зональный НИИСХ Северо-Востока, 2006.- 568 с.

## А.9. Критические значения $\chi^2$-распределения Пирсона




| df | Уровень значимости (α) | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | **0,99** | **0,95** | **0,90** | **0,75** | **0,50** | **0,25** | **0,10** | **0,05** | **0,025** | **0,010** |
| **1** | … | … | 0,02 | 0,10 | 0,45 | 1,32 | 2,71 | 3,84 | 5,02 | 6,63 |
| **2** | 0,02 | 0,10 | 0,21 | 0,58 | 1,39 | 2,77 | 4,61 | 5,99 | 7,38 | 9,21 |
| **3** | 0,11 | 0,35 | 0,58 | 1,21 | 2,37 | 4,11 | 6,25 | 7,81 | 9,35 | 11,34 |
| **4** | 0,30 | 0,71 | 1,06 | 1,92 | 3,36 | 5,39 | 7,78 | 9,49 | 11,14 | 13,28 |
| **5** | 0,55 | 1,15 | 1,61 | 2,67 | 4,35 | 6,63 | 9,24 | 11,07 | 12,83 | 15,09 |
| **6** | 0,87 | 1,64 | 2,20 | 3,45 | 5,35 | 7,84 | 10,64 | 12,59 | 14,45 | 16,81 |
| **7** | 1,24 | 2,17 | 2,83 | 4,25 | 6,35 | 9,04 | 12,02 | 14,07 | 16,01 | 18,48 |
| **8** | 1,65 | 2,73 | 3,49 | 5,07 | 7,34 | 10,22 | 13,36 | 15,51 | 17,53 | 20,09 |
| **9** | 2,09 | 3,33 | 4,17 | 5,90 | 8,34 | 11,39 | 14,68 | 16,92 | 19,02 | 21,67 |
| **10** | 2,56 | 3,94 | 4,87 | 6,74 | 9,34 | 12,55 | 15,99 | ***18,31*** | 20,48 | 23,21 |
| **11** | 3,05 | 4,57 | 5,58 | 7,58 | 10,34 | 13,70 | 17,28 | 19,68 | 21,92 | 24,72 |
| **12** | 3,57 | 5,23 | 6,30 | 8,44 | 11,34 | 14,85 | 18,55 | 21,03 | 23,34 | 26,22 |
| **13** | 4,11 | 5,89 | 7,04 | 9,30 | 12,34 | 15,98 | 19,81 | 22,36 | 24,74 | 27,69 |
| **14** | 4,66 | 6,57 | 7,79 | 10,17 | 13,34 | 17,12 | 21,06 | 23,68 | 26,12 | 29,14 |
| **15** | 5,23 | 7,26 | 8,55 | 11,04 | 14,34 | 18,25 | 22,31 | 25,00 | 27,49 | 30,58 |
| **16** | 5,81 | 7,96 | 9,31 | 11,91 | 15,34 | 19,37 | 23,54 | 26,30 | 28,85 | 32,00 |
| **17** | 6,41 | 8,67 | 10,09 | 12,79 | 16,34 | 20,49 | 24,77 | 27,59 | 30,19 | 33,41 |
| **18** | 7,01 | 9,39 | 10,86 | 13,68 | 17,34 | 21,60 | 25,99 | 28,87 | 31,53 | 34,81 |
| **19** | 7,63 | 10,12 | 11,65 | 14,56 | 18,34 | 22,72 | 27,20 | 30,14 | 32,85 | 36,19 |
| **20** | 8,26 | 10,85 | 12,44 | 15,45 | 19,34 | 23,83 | 28,41 | 31,41 | 34,17 | 37,57 |
| **21** | 8,90 | 11,59 | 13,24 | 16,34 | 20,34 | 24,93 | 29,62 | 32,67 | 35,48 | 38,93 |
| **22** | 9,54 | 12,34 | 14,04 | 17,24 | 21,34 | 26,04 | 30,81 | 33,92 | 36,78 | 40,29 |
| **23** | 10,20 | 13,09 | 14,85 | 18,14 | 22,34 | 27,14 | 32,01 | 35,17 | 38,08 | 41,64 |
| **24** | 10,86 | 13,85 | 15,66 | 19,04 | 23,34 | 28,24 | 33,20 | 36,42 | 39,36 | 42,98 |
| **25** | 11,52 | 14,61 | 16,47 | 19,94 | 24,34 | 29,34 | 34,38 | 37,65 | 40,65 | 44,31 |
| **26** | 12,20 | 15,38 | 17,29 | 20,84 | 25,34 | 30,43 | 35,56 | 38,89 | 41,92 | 45,64 |
| **27** | 12,88 | 16,15 | 18,11 | 21,75 | 26,34 | 31,53 | 36,74 | 40,11 | 43,19 | 46,96 |
| **28** | 13,56 | 16,93 | 18,94 | 22,66 | 27,34 | 32,62 | 37,92 | 41,34 | 44,46 | 48,28 |
| **29** | 14,26 | 17,71 | 19,77 | 23,57 | 28,34 | 33,71 | 39,09 | 42,56 | 45,72 | 49,59 |
| **30** | 14,95 | 18,49 | 20,60 | 24,48 | 29,34 | 34,80 | 40,26 | 43,77 | 46,98 | 50,89 |
| **40** | 22,16 | 26,51 | 29,05 | 33,66 | 39,34 | 45,62 | 51,80 | 55,76 | 59,34 | 63,69 |
| **50** | 29,71 | 34,76 | 37,69 | 42,94 | 49,33 | 56,33 | 63,17 | 67,50 | 71,42 | 76,15 |
| **60** | 37,48 | 43,19 | 46,46 | 52,29 | 59,33 | 66,98 | 74,40 | 79,08 | 83,30 | 88,38 |
| **70** | 45,44 | 51,74 | 55,33 | 61,70 | 69,33 | 77,58 | 85,53 | 90,53 | 95,02 | 100,42 |
| **80** | 53,54 | 60,39 | 64,28 | 71,14 | 79,33 | 88,13 | 96,58 | 01,88 | 06,63 | 112,33 |
| **90** | 61,75 | 69,13 | 73,29 | 80,62 | 89,33 | 98,64 | 07,56 | 13,14 | 18,14 | 124,12 |
| **100** | 70,06 | 77,93 | 82,36 | 90,13 | 99,33 | 109,14 | 18,50 | 24,34 | 29,56 | 135,81 |

## А.12. Значения коэффициента корреляции (r) при различных уровнях значимости (α) и числе степеней свободы (df)

| df | α | | df | α | |
|---|---|---|---|---|---|
| | 0,05 | 0,01 | | 0,05 | 0,01 |
| **1** | 0,997 | 1,000 | **24** | 0,388 | 0,496 |
| **2** | 0,950 | 0,990 | **25** | 0,381 | 0,487 |
| **3** | 0,878 | 0,959 | **26** | 0,374 | 0,478 |
| **4** | 0,811 | 0,917 | **27** | 0,367 | 0,470 |
| **5** | 0,754 | 0,874 | **28** | 0,361 | 0,463 |
| **6** | 0,707 | 0,834 | **29** | 0,355 | 0,456 |
| **7** | 0,666 | 0,798 | **30** | 0,349 | 0,449 |
| **8** | 0,632 | 0,765 | **35** | 0,325 | 0,418 |
| **9** | 0,602 | 0,735 | **40** | 0,304 | 0,393 |
| **10** | 0,576 | 0,708 | **45** | 0,288 | 0,372 |
| **11** | 0,553 | 0,684 | **50** | 0,273 | 0,354 |
| **12** | 0,532 | 0,661 | **60** | 0,250 | 0,325 |
| **13** | 0,514 | 0,641 | **70** | 0,232 | 0,302 |
| **14** | 0,497 | 0,623 | **80** | 0,217 | 0,283 |
| **15** | 0,482 | 0,606 | **90** | 0,205 | 0,267 |
| **16** | 0,468 | 0,590 | **100** | 0,195 | 0,254 |
| **17** | 0,456 | 0,575 | **125** | 0,174 | 0,228 |
| **18** | 0,444 | 0,561 | **150** | 0,159 | 0,208 |
| **19** | 0,433 | 0,549 | **200** | 0,138 | 0,181 |
| **20** | 0,423 | 0,537 | **300** | 0,113 | 0,148 |
| **21** | 0,413 | 0.526 | **400** | 0,098 | 0,128 |
| **22** | 0,404 | 0,515 | **500** | 0,088 | 0,115 |
| **23** | 0,396 | 0,505 | **1000** | 0,062 | 0,081 |