# Часть 1. Выборочный коэффициент корреляции

Цель: научиться рассчитывать линейные коэффициенты корреляции, проверять их значимость и строить интервальные оценки, интерпретировать параметры связи.

Постановка задачи: в 12 анализах крови здоровых пациентов определяли содержание гемоглобина (в %) и оседание эритроцитов крови за 24 часа (в мм).

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Гемоглобин % | 24 | 44 | 62 | 65 | 73 | 82 | 74 | 83 | 79 | 83 | 82 | 85 |
| Оседание Эритроцитов, мм | 9 | 19 | 25 | 27 | 28 | 29 | 30 | 30 | 31 | 33 | 34 | 36 |

Требуется найти выборочный коэффициент корреляции и проверить его значимость, т.е. ответить на вопрос: является зависимость оседания эритроцитов от гемоглобина статистически значимой? Распределение случайных величин X и Y предполагается нормальным. Уровень значимости α составляет 0,05.

# Решение

Заполним таблицу:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| № | x | y | x\*y | x^2 | y^2 |
| 1 | 24 | 9 | 216 | 576 | 81 |
| 2 | 44 | 19 | 836 | 1936 | 361 |
| 3 | 62 | 25 | 1550 | 3844 | 625 |
| 4 | 65 | 27 | 1755 | 4225 | 729 |
| 5 | 73 | 28 | 2044 | 5329 | 784 |
| 6 | 82 | 29 | 2378 | 6724 | 841 |
| 7 | 74 | 30 | 2220 | 5476 | 900 |
| 8 | 83 | 30 | 2490 | 6889 | 900 |
| 9 | 79 | 31 | 2449 | 6241 | 961 |
| 10 | 83 | 33 | 2739 | 6889 | 1089 |
| 11 | 82 | 34 | 2788 | 6724 | 1156 |
| 12 | 85 | 36 | 3060 | 7225 | 1296 |
| Среднее | 69,67 | 27,58 | 2043,75 | 5173,17 | 810,25 |
| Сумма | 836 | 331 | 24525 | 62078 | 9723 |

Найдем выборочный коэффициент корреляции по формуле:

Проверим гипотезу с помощью критерия Стьюдента, для этого найдем наблюдаемое значение статистики:

Затем, по заданному уровню значимости и числу степеней свободы находим критическое значение статистики :

Так как , делаем вывод, что коэффициент корреляции является значимым, то есть нулевая гипотеза отвергается с вероятностью ошибки .

# Часть 2. Частный и множественный коэффициенты корреляции

По данным 12 машиностроительных предприятий методами корреляционного анализа исследуется взаимосвязь между следующими показателями:

* x1 – рентабельность, %;
* x2 – премии и вознаграждения на одного работника, млн. руб.;
* x3 – фонтдоотдача, представленными в таблице:

|  |  |  |  |
| --- | --- | --- | --- |
| №п/п | *x1* | *х2* | *x3* |
| 1 | 12,06 | 1,25 | 1,35 |
| 2 | 11,26 | 1,14 | 1,25 |
| 3 | 13,26 | 1,23 | 1,45 |
| 4 | 10,16 | 1,04 | 1,30 |
| 5 | 13,72 | 1,80 | 1,37 |
| 6 | 12,82 | 0,43 | 1,65 |
| 7 | 10,63 | 0,88 | 1,91 |
| 8 | 9,12 | 0,57 | 1,68 |
| 9 | 25,83 | 1,72 | 1,94 |
| 10 | 23,39 | 1,70 | 1,89 |
| 11 | 14,68 | 0,84 | 1,94 |
| 12 | 10,05 | 0,60 | 2,06 |

Требуется:

а) рассчитать векторы средних и средних квадратических отклонений, матрицу парных коэффициентов корреляции ( х, S, R);

б) проверить при α = 0,05 значимость парного коэффициента корреляции ρ12 и найти его интервальную оценку с доверительной вероятностью γ = 0,95;

в) по корреляционной матрице R рассчитать частный коэффициент корреляции r12/3;

г) проверить при α = 0,05 значимость частного коэффициента корреляции ρ12/3 и определить его интервальную оценку при γ = 0,95;

д) по корреляционной матрице R вычислить оценку множественного коэффициента корреляции r1(2,3) и при α = 0,05 проверить гипотезу H0: r1(2,3)=0.

Решение

Построим таблицу:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| №п/п | x1 | х2 | x3 | x1\*x2 | x1\*x3 | x2\*x3 |
| 1 | 12,06 | 1,25 | 1,35 | 15,08 | 16,28 | 1,69 |
| 2 | 11,26 | 1,14 | 1,25 | 12,84 | 14,08 | 1,43 |
| 3 | 13,26 | 1,23 | 1,45 | 16,31 | 19,23 | 1,78 |
| 4 | 10,16 | 1,04 | 1,30 | 10,57 | 13,21 | 1,35 |
| 5 | 13,72 | 1,80 | 1,37 | 24,70 | 18,80 | 2,47 |
| 6 | 12,82 | 0,43 | 1,65 | 5,51 | 21,15 | 0,71 |
| 7 | 10,63 | 0,88 | 1,91 | 9,35 | 20,30 | 1,68 |
| 8 | 9,12 | 0,57 | 1,68 | 5,20 | 15,32 | 0,96 |
| 9 | 25,83 | 1,72 | 1,94 | 44,43 | 50,11 | 3,34 |
| 10 | 23,39 | 1,70 | 1,89 | 39,76 | 44,21 | 3,21 |
| 11 | 14,68 | 0,84 | 1,94 | 12,33 | 28,48 | 1,63 |
| 12 | 10,05 | 0,60 | 2,06 | 6,03 | 20,70 | 1,24 |
| Среднее | 13,92 | 1,10 | 1,65 | 16,84 | 23,49 | 1,79 |
| Ср кв откл | 5,06 | 0,44 | 0,28 |  |  |  |

|  |  |
| --- | --- |
| R12 | 0,6820 |
| R13 | 0,3790 |
| R23 | -0,1939 |

а) Основываясь на таблицах, получим значения векторов:

б) Проверим при α = 0,05 значимость парного коэффициента корреляции:

Так как |R12|>rкр, значит, удалось установить значимую взаимосвязь между признаками, нулевая гипотеза отвергается. Далее определим интервальную оценку. Для этого используем Z-преобразование Фишера:

По таблице Z-преобразования Фишера для R12=0,682 будем иметь Z’(0,682)=0,830. По таблице нормального распределения . Тогда

Обратный переход по таблице Z-преобразования:

Полученная интервальная оценка подтверждает вывод о значимости парного коэффициента корреляции , т.к. ноль находится вне доверительного интервала.

в) Найдем точечную оценку множественного коэффициента корреляции . Вычислим алгебраические дополнения:

г) Для проверки значимости частных коэффициентов корреляции по таблице Фишера-Йейтса найдем:

Так как , нулевая гипотеза отвергается. Далее определим интервальную оценку, где, (с помощью функции ФИШЕР):

Полученная интервальная оценка подтверждает вывод о значимости парного коэффициента корреляции , т.к. ноль находится вне доверительного интервала.

г) Найдем оценку множественного коэффициента корреляции :

Проверим гипотезу . Имеем

Критическое значение по таблице F-распределения:

# Заключение

Так как , то гипотеза отвергается, то есть множественный коэффициент не равен нулю (). Зависимость между переменными присутствует.