**Корреляционный анализ**.

**Уравнение парной регрессии**.

**Использование графического метода**.

Этот метод применяют для наглядного изображения формы связи между изучаемыми экономическими показателями. Для этого в прямоугольной системе координат строят график, по оси ординат откладывают индивидуальные значения результативного признака Y, а по оси абсцисс - индивидуальные значения факторного признака X.

Совокупность точек результативного и факторного признаков называется **полем корреляции**.

На основании поля корреляции можно выдвинуть гипотезу (для генеральной совокупности) о том, что связь между всеми возможными значениями X и Y носит линейный характер.

Линейное уравнение регрессии имеет вид y = bx + a

Оценочное уравнение регрессии (построенное по выборочным данным) будет иметь вид y = bx + a + ε, где ei – наблюдаемые значения (оценки) ошибок εi, a и b соответственно оценки параметров α и β регрессионной модели, которые следует найти.

Здесь ε - случайная ошибка (отклонение, возмущение).

Причины существования случайной ошибки:

1. Невключение в регрессионную модель значимых объясняющих переменных;

2. Агрегирование переменных. Например, функция суммарного потребления – это попытка общего выражения совокупности решений отдельных индивидов о расходах. Это лишь аппроксимация отдельных соотношений, которые имеют разные параметры.

3. Неправильное описание структуры модели;

4. Неправильная функциональная спецификация;

5. Ошибки измерения.

Так как отклонения εi для каждого конкретного наблюдения i – случайны и их значения в выборке неизвестны, то:

1) по наблюдениям xi и yi можно получить только оценки параметров α и β

2) Оценками параметров α и β регрессионной модели являются соответственно величины а и b, которые носят случайный характер, т.к. соответствуют случайной выборке;

Для оценки параметров α и β - используют МНК (метод наименьших квадратов).

Метод наименьших квадратов дает наилучшие (состоятельные, эффективные и несмещенные) оценки параметров уравнения регрессии. Но только в том случае, если выполняются определенные предпосылки относительно случайного члена (ε) и независимой переменной (x).

Формально критерий МНК можно записать так:

S = ∑(yi - y∙i)2 → min

Система нормальных уравнений.

a·n + b·∑x = ∑y

a·∑x + b·∑x2 = ∑y·x

Для расчета параметров регрессии построим расчетную таблицу (табл. 1)

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| x | y | x2 | y2 | x∙y |
| 0.9937 | 3.0204 | 0.9874 | 9.123 | 3.0014 |
| 0.9912 | 3.6147 | 0.9825 | 13.066 | 3.5829 |
| 0.9885 | 3.2347 | 0.9771 | 10.4636 | 3.1975 |
| 0.9791 | 2.3076 | 0.9586 | 5.3249 | 2.2593 |
| 0.9698 | 2.2094 | 0.9405 | 4.8813 | 2.1426 |
| 0.9396 | 2.2513 | 0.8828 | 5.0683 | 2.1153 |
| 0.9423 | 2.0844 | 0.8879 | 4.3448 | 1.9642 |
| 0.9624 | 2.0149 | 0.9262 | 4.0598 | 1.9391 |
| 0.9504 | 1.7579 | 0.9033 | 3.0901 | 1.6707 |
| 0.9356 | 2.0281 | 0.8753 | 4.1134 | 1.8975 |
| 0.9188 | 1.3481 | 0.8442 | 1.8173 | 1.2386 |
| 0.9118 | 0.3075 | 0.8314 | 0.09455 | 0.2804 |
| 0.7374 | 1.8437 | 0.5438 | 3.3993 | 1.3596 |
| 0.8451 | -0.3285 | 0.7142 | 0.1079 | -0.2776 |
| 0.783 | 1.209 | 0.6131 | 1.4616 | 0.9466 |
| 0.7838 | 1.4861 | 0.6143 | 2.2086 | 1.1648 |
| 0.6979 | -0.2614 | 0.4871 | 0.06831 | -0.1824 |
| 0.5486 | -2.3026 | 0.301 | 5.3019 | -1.2632 |
| 15.879 | 27.8254 | 14.2708 | 77.9947 | 27.0374 |

Для наших данных система уравнений имеет вид

18a + 15.879·b = 27.825

15.879·a + 14.271·b = 27.037

Домножим уравнение (1) системы на (-0.882), получим систему, которую решим методом алгебраического сложения.

-15.879a -14.005 b = -24.542

15.879∙a + 14.271∙b = 27.037

Получаем:

0.266∙b = 2.495

Откуда b = 9.4759

Теперь найдем коэффициент «a» из уравнения (1):

18a + 15.879∙b = 27.825

18a + 15.879∙9.4759 = 27.825

18a = -122.642

a = -6.8135

Получаем эмпирические коэффициенты регрессии: b = 9.4759, a = -6.8135

Уравнение регрессии (эмпирическое уравнение регрессии):

y = 9.4759 x -6.8135

Эмпирические коэффициенты регрессии a и b являются лишь оценками теоретических коэффициентов βi, а само уравнение отражает лишь общую тенденцию в поведении рассматриваемых переменных.

**1. Параметры уравнения регрессии**.

Выборочные средние.

Выборочные дисперсии:

Среднеквадратическое отклонение

Коэффициент корреляции b можно находить по формуле, не решая систему непосредственно:

**1.1. Коэффициент корреляции**.

*Ковариация*.

Рассчитываем показатель тесноты связи. Таким показателем является выборочный линейный коэффициент корреляции, который рассчитывается по формуле:

Линейный коэффициент корреляции принимает значения от –1 до +1.

Связи между признаками могут быть слабыми и сильными (тесными). Их критерии оцениваются по шкале Чеддока:

0.1 < rxy < 0.3: слабая;

0.3 < rxy < 0.5: умеренная;

0.5 < rxy < 0.7: заметная;

0.7 < rxy < 0.9: высокая;

0.9 < rxy < 1: весьма высокая;

В нашем примере связь между признаком Y и фактором X высокая и прямая.

Кроме того, коэффициент линейной парной корреляции может быть определен через коэффициент регрессии b:

**2.1. Значимость коэффициента корреляции**.

Выдвигаем гипотезы:

H0: rxy = 0, нет линейной взаимосвязи между переменными;

H1: rxy ≠ 0, есть линейная взаимосвязь между переменными;

Для того чтобы при уровне значимости α проверить нулевую гипотезу о равенстве нулю генерального коэффициента корреляции нормальной двумерной случайной величины при конкурирующей гипотезе H1 ≠ 0, надо вычислить наблюдаемое значение критерия (величина случайной ошибки)

и по таблице критических точек распределения Стьюдента, по заданному уровню значимости α и числу степеней свободы k = n - 2 найти критическую точку tкрит двусторонней критической области. Если tнабл < tкрит оснований отвергнуть нулевую гипотезу. Если |tнабл| > tкрит — нулевую гипотезу отвергают.

По таблице Стьюдента с уровнем значимости α=0.05 и степенями свободы k=16 находим tкрит:

tкрит(n-m-1;α/2) = tкрит(16;0.025) = 2.473

где m = 1 - количество объясняющих переменных.

Если |tнабл| > tкритич, то полученное значение коэффициента корреляции признается значимым (нулевая гипотеза, утверждающая равенство нулю коэффициента корреляции, отвергается).

Поскольку |tнабл| > tкрит, то отклоняем гипотезу о равенстве 0 коэффициента корреляции. Другими словами, коэффициент корреляции статистически - значим.

В парной линейной регрессии t2r = t2b и тогда проверка гипотез о значимости коэффициентов регрессии и корреляции равносильна проверке гипотезы о существенности линейного уравнения регрессии.

**2.2. Интервальная оценка для коэффициента корреляции (доверительный интервал)**.

Доверительный интервал для коэффициента корреляции.

r∈(0.469;1)

**1.2. Уравнение регрессии** (оценка уравнения регрессии).

Линейное уравнение регрессии имеет вид y = 9.476∙x -6.813

Коэффициентам уравнения линейной регрессии можно придать экономический смысл.

Коэффициент регрессии b = 9.476 показывает среднее изменение результативного показателя (в единицах измерения у) с повышением или понижением величины фактора х на единицу его измерения. В данном примере с увеличением на 1 единицу y повышается в среднем на 9.476.

Коэффициент a = -6.813 формально показывает прогнозируемый уровень у, но только в том случае, если х=0 находится близко с выборочными значениями.

Но если х=0 находится далеко от выборочных значений х, то буквальная интерпретация может привести к неверным результатам, и даже если линия регрессии довольно точно описывает значения наблюдаемой выборки, нет гарантий, что также будет при экстраполяции влево или вправо.

Подставив в уравнение регрессии соответствующие значения х, можно определить выровненные (предсказанные) значения результативного показателя y(x) для каждого наблюдения.

Связь между у и х определяет знак коэффициента регрессии b (если > 0 – прямая связь, иначе - обратная). В нашем примере связь прямая.

**2. Оценка параметров уравнения регрессии**.

**Выводы**.

Изучена зависимость Y от X. На этапе спецификации была выбрана парная линейная регрессия. Оценены её параметры методом наименьших квадратов: y = 9.476∙x -6.813 Возможна экономическая интерпретация параметров модели - увеличение X на 1 ед.изм. приводит к увеличению Y в среднем на 9.476 ед.изм.

Линейный коэффициент корреляции равен 0.821, следовательно, связь между признаком Y и фактором X высокая и прямая. Также подтверждается его значимость.

Решение было получено и оформлено с помощью сервиса:

[Уравнение регрессии](https://math.semestr.ru/corel/corel.php)

Вместе с этой задачей решают также:

[Уравнение множественной регрессии](https://math.semestr.ru/regress/corel.php)

[Коэффициент корреляции Спирмена](https://math.semestr.ru/corel/spirmen.php)

[Выявление тренда методом аналитического выравнивания](https://math.semestr.ru/trend/analis.php)

[Уравнение нелинейной регрессии](https://math.semestr.ru/corel/noncorel.php)

[Показатели динамики: цепные и базисные](https://axd.semestr.ru/dinam/group.php)

[Коэффициент корреляции Пирсона](https://math.semestr.ru/corel/prim.php)