

Парная регрессия и корреляция

1. Статистическая зависимость (независимость) случайных переменных.
2. Анализ линейной статистической связи экономических данных.
3. Нелинейные модели и их линеаризация.

Статистическая зависимость случайных переменных

Экономические явления:

- Обладают большим разнообразием;
- Характеризуются множеством признаков, которые отражают те или иные их свойства.

Признаки экономических явлений:

- Изменяются (варьируются) во времени и пространстве.
- Изменения признаков взаимозависимы и взаимообусловлены.

Связь между признаками экономических явлений оказывается:

- очень тесной (например, часовая выработка и заработная плата);
- вовсе не обнаруживается или выражается слабо (например, пол студента и их успеваемость).



ВАЖНО!

Чем теснее связь между признаками, тем точнее принимаемые решения и легче управление системами.

Среди многих форм связей явлений важнейшую роль играет причинная, определяющая все другие формы.

Сущность причинности состоит в порождении одного явления другим.

В любой конкретной связи одни признаки выступают в качестве факторов, воздействующих на другие и обуславливающие их изменения, другие – в качестве результатов действия этих факторов.

Функциональная зависимость

ЭТО

связь, при которой каждому значению независимой переменной x соответствует точно *определенное значение* зависимой переменной y

Статистическая зависимость

ЭТО

связь, при которой каждому значению независимой переменной x соответствует *множество значений* зависимой переменной y , причем неизвестно заранее, какое именно значение примет y

**Частным случаем статистической зависимости
является**

**Корреляционная
зависимость**

ЭТО

связь, при которой каждому значению независимой переменной x соответствует *определенное математическое ожидание (среднее значение)* зависимой переменной y

ВАЖНО!

Корреляционная связь является «неполной» зависимостью, которая проявляется только в средних величинах при достаточно большом числе случаев

Пример

Повышение квалификации работника ведет к росту производительности труда. Это положение подтверждается в массе явлений и не означает, что у двух или более рабочих одного разряда, занятых аналогичным процессом, будет одинаковая производительность труда. Уровни их выработки будут различаться, т.к. у этих рабочих могут быть различными стаж работы, состояние здоровья и т.д.

Особенности зависимости

Функциональная

- Всегда выражается формулами, и присуще точным наукам (математика, физика).
- С одинаковой силой проявляется у всех единиц совокупности.
- Является полной и точной, т.к. известен перечень всех факторов и механизм их воздействия на переменную в виде уравнения

Корреляционная

- Разнообразие факторов, их взаимосвязи и противоречивые действия вызывают широкое варьирование переменной y
- Обнаруживается в массе и требует для своего исследования массовых наблюдений.
- Связь между переменными неполная и проявляется лишь в средних величинах

Виды функциональной и корреляционной зависимости

Прямая

- С увеличением (уменьшением) значение факторного признака происходит увеличение (уменьшение) результативного признака

Обратная

- С увеличением (уменьшением) значений факторного признака происходит уменьшение (увеличение) результативного признака

Функциональная и корреляционная зависимость

Прямолинейной

- С возрастанием величины факторного признака происходит равномерное возрастание (или убывание) величин результативного признака
- Выражается уравнением прямой линии

Криволинейной

- С возрастанием величины факторного признака возрастание (или убывание) результативного признака происходит неравномерно
- Выражаются уравнениями кривых линий

Корреляционные связи

```
graph TD; A[Корреляционные связи] --> B[Однофакторные (парные)]; A --> C[Многофакторные (множественные)]; B --> D[Связь между одним признаком-фактором и результативным признаком]; C --> E[Связь между несколькими факторными признаками и результативным признаком];
```

Однофакторные (парные)

- Связь между одним признаком-фактором и результативным признаком

Многофакторные (множественные)

- Связь между несколькими факторными признаками и результативным признаком

Понятие корреляционного анализа

Корреляционный
анализ

ЭТО

*Раздел математической статистики,
посвященный изучению взаимосвязей между
случайными величинами*

ВАЖНО!

Применяется анализ тогда, когда данные наблюдения можно считать случайными и выбранными из генеральной совокупности, распределенной по многомерному нормальному закону

Пример

Корреляционный анализ заключается в *количественном определении тесноты* связи между двумя признаками (при парной связи) и между результативным и множеством факторных признаков (при многофакторной связи)

Понятие корреляции

Корреляция

ЭТО

Статистическая зависимость между случайными величинами, при которой изменение одной из случайных величин приводит к изменению математического ожидания другой

Варианты корреляции

Парная

Связь между двумя признаками (результативным или факторным или двумя факторными)

Частная

Зависимость между результативным и одним факторным признаками или двумя факторными признаками при фиксированном значении других факторных признаков

Множественная

Зависимость между результативным признаком и двумя и более факторными признаками, включенными в исследование

ВАЖНО!

**Теснота связи количественно
выражается величиной
коэффициентов корреляции.**

Ковариация

**Характеризует сопряженность
вариации двух признаков и
представляет собой
статистическую меру
взаимодействия двух случайных
переменных**

Формула определения ковариации

$$\text{Cov}(y, x) = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{n}$$

n – объем исследуемой совокупности,

x_i - i -е значение независимой переменной ($i = 1, 2, \dots, n$),

y_i - i -е значение зависимой переменной ($i = 1, 2, \dots, n$),

\bar{x} - среднее значение независимой переменной, рассчитываемая по формуле:

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i$$

\bar{y} - среднее значение зависимой переменной, рассчитываемая по формуле:

$$\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$$



Ковариация

Виды связи в результате расчета ковариации:

- Если ковариация будет положительной, то между случайными величинами существует **прямая связь**.
- Если ковариация будет отрицательной, то между случайными величинами существует **обратная связь**.
- Если ковариация будет близка к нулю, то между случайными величинами **отсутствует связь**.

Формула линейного коэффициента корреляции

$$r_{yx} = \frac{\text{cov}(y,x)}{\sigma_x \sigma_y} = \frac{\sum_{i=1}^n (x - \bar{x})(y - \bar{y})}{n \sigma_x \sigma_y}$$

$\sigma_x \sigma_y$ - средние квадратические отклонения случайных величин x и y ,

Определяются по формуле:

$$\sigma_x = \sqrt{\frac{\sum_{i=1}^n (x - \bar{x})^2}{n}}$$

$$\sigma_y = \sqrt{\frac{\sum_{i=1}^n (y - \bar{y})^2}{n}}$$



Ковариация



ВАЖНО

- Коэффициент корреляции принимает значение от -1 до $+1$.
- Положительное значение коэффициента свидетельствует о наличии прямой связи.
- Отрицательное значение коэффициента свидетельствует о наличии обратной связи.
- Если $r_{yx} = \pm 1$, то корреляционная связь представляется линейной функциональной зависимостью.
- Если $r_{yx} = 0$, то линейная корреляционная связь отсутствует

Качественная характеристика связи

```
graph TD; A[Качественная характеристика связи] --> B[значение]; A --> C[Характер связи]; B --> D[ ]; C --> E[Практически отсутствует<br/>Слабая<br/>Умеренная<br/>Сильная];
```

значение

Характер связи

Практически отсутствует
Слабая
Умеренная
Сильная

Спасибо за внимание