

Причинность, регрессия, корреляция

Исследование объективно существующих связей между явлениями - важнейшая задача общей теории статистики. В процессе статистического исследования зависимостей вскрываются причинно-следственные отношения между явлениями, что позволяет выявлять факторы (признаки), оказывающие существенное влияние на вариацию изучаемых явлений и процессов. Причинно-следственные отношения - это связь явлений и процессов, когда изменение одного из них - причины - ведет к изменению другого - следствия.

Причина - это совокупность условий, обстоятельств, действие которых приводит к появлению следствия. Если между явлениями действительно существуют причинно-следственные отношения, то эти условия должны обязательно реализовываться вместе с действием причин. Причинные связи носят всеобщий и многообразный характер, и для обнаружения причинно-следственных связей необходимо отбирать отдельные явления и изучать их изолированно.

Особое значение при исследовании причинно-следственных связей имеет выявление временной последовательности: причина всегда должна предшествовать следствию, однако не каждое предшествующее событие следует считать причиной, а последующее следствием.

В реальной социально-экономической действительности причину и следствие необходимо рассматривать как смежные явления, появление которых обусловлено комплексом сопутствующих более простых причин и следствий. Между сложными группами причин и следствий возможны многозначные связи, когда за одной причиной будет следовать то одно, то другое действие или одно действие имеет несколько различных причин. Чтобы установить однозначную причинную связь между явлениями или предсказать возможные следствия конкретной причины, необходима полная абстракция от всех прочих явлений в исследуемой временной или пространственной среде. Теоретически такая абстракция воспроизводится. Приемы абстракции часто применяются при изучении взаимосвязей между двумя признаками (парной корреляции). Но чем сложнее изучаемые явления, тем труднее выявить причинно-следственные связи между ними. Взаимное переплетение различных внутренних и внешних факторов неизбежно приводит к некоторым ошибкам в определении причины и следствия.

Особенностью причинно-следственных связей в социально-экономических явлениях является их транзитивность, т. е. причина X и следствие Y связаны соотношением $X \rightarrow X' \rightarrow X'' \rightarrow Y$, а не непосредственно $X \rightarrow Y$. Однако промежуточные факторы, как правило, при анализе опускаются.

Так, например, используя показатели международной методологии расчетов, фактором валовой прибыли (Y) является валовое накопление основных и оборотных фондов (X), но при этом допускаются такие факторы, как валовой выпуск (X'), оплата труда (X'') и т. д. Правильно вскрытые причинно-

следственные связи позволяют установить силу воздействия отдельных факторов на результаты хозяйственной деятельности.

Социально-экономические явления представляют собой результат одновременного воздействия большого числа причин. Следовательно, при изучении этих явлений необходимо выявлять главные, основные причины, абстрагируясь от второстепенных.

В основе первого этапа статистического изучения связи лежит качественный анализ изучаемого явления, связанный с анализом природы социального или экономического явления методами экономической теории, социологии, конкретной экономики. Второй этап - построение модели связи. Он базируется на методах статистики: группировках, средних величинах, таблицах и т. д. Третий, последний этап - интерпретация результатов - вновь связан с качественными особенностями изучаемого явления.

Статистика разработала множество методов изучения связей, выбор которых зависит от целей исследования и от поставленных задач. Связи между признаками и явлениями, ввиду их большого разнообразия, классифицируются по ряду оснований. Признаки по их значению для изучения взаимосвязи делятся на два класса. Признаки, обуславливающие изменения других, связанных с ними признаков, называются факторными, или просто факторами. Признаки, изменяющиеся под действием факторных признаков, являются результативными. Связи между явлениями и их признаками классифицируются по степени тесноты связи, направлению и аналитическому выражению.

В статистике различают функциональную связь и стохастическую зависимость. Функциональной называют такую связь, при которой определенному значению факторного признака соответствует одно и только одно значение результативного признака. Функциональная связь проявляется во всех случаях наблюдения и для каждой конкретной единицы исследуемой совокупности.

Если причинная зависимость проявляется не в каждом отдельном случае, а в общем, среднем при большом числе наблюдений, то такая зависимость называется стохастической. Частным случаем стохастической связи является корреляционная связь, при которой изменение среднего значения результативного признака обусловлено изменением факторных признаков.

По степени тесноты связи различают количественные критерии оценки тесноты связи (таблица 25).

Таблица 25 - Количественные критерии оценки тесноты связи

Величина коэффициента корреляции			Характер связи
До	±	0,3	Практически отсутствует
0,3 -		0,5	Слабая
0,5		0,7	Умеренная
0,7		1,0	Сильная

По направлению выделяют связь прямую и обратную. При прямой связи с увеличением или уменьшением значений факторного признака происходит увеличение или уменьшение значений результативного. Так, например, рост производительности труда способствует увеличению уровня рентабельности производства. В случае обратной связи значения результативного признака изменяются под воздействием факторного, но в противоположном направлении по сравнению с изменением факторного признака. Так, с увеличением уровня фондоотдачи снижается себестоимость единицы производимой продукции.

По аналитическому выражению выделяют связи прямолинейные (или просто линейные) и нелинейные. Если статистическая связь между явлениями может быть приближенно выражена уравнением прямой линии, то ее называют линейной связью; если же она выражается уравнением какой-либо кривой линии (параболы, гиперболы, степенной, показательной, экспоненциальной и т. д.), то такую связь называют нелинейной, или криволинейной.

В статистике не всегда требуются количественные оценки связи, часто важно определить лишь ее направление и характер, выявить форму воздействия одних факторов на другие. Для выявления наличия связи, ее характера и направления в статистике используются методы: приведения параллельных данных; аналитических группировок; графический; корреляции.

Метод приведения параллельных данных основан на сопоставлении двух или нескольких рядов статистических величин. Такое сопоставление позволяет установить наличие связи и получить представление о ее характере. Сравним изменения двух величин X и Y :

С увеличением величины X величина Y также возрастает. Поэтому связь между ними прямая, и описать ее можно или уравнением прямой, или уравнением параболы второго порядка.

Графически взаимосвязь двух признаков изображается с помощью поля корреляции. В системе координат на оси абсцисс откладываются значения факторного признака, а на оси ординат - результативного. Каждое пересечение линий, проводимых через эти оси, обозначается точкой. При отсутствии тесных связей имеет место беспорядочное расположение точек на графике. Чем сильнее связь между признаками, тем теснее будут группироваться точки вокруг определенной линии, выражающей форму связи.

11.2 Корреляционные связи

Для социально-экономических явлений характерно, что наряду с существенными факторами, формирующими уровень результативного признака на него оказывают воздействие многие факторы.

Это свидетельствует о том, что взаимосвязи явлений, которые изучает статистика, носят корреляционный характер и аналитически выражаются функцией вида $y_x = f(x)$.

Корреляция - это статистическая зависимость между случайными величинами, не имеющими строго функционального характера, при которой изменение одной из случайных величин приводит к изменению математического ожидания другой.

В статистике принято различать следующие варианты зависимостей.

1. Парная корреляция - связь между двумя признаками (результативным и факторным или двумя факторными).

2. Частная корреляция - зависимость между результативным и одним факторным признаками при фиксированном значении других факторных признаков.

3. Множественная корреляция - зависимость результативного и двух или более факторных признаков, включенных в исследование.

Корреляционный анализ имеет своей задачей количественное определение тесноты связи между двумя признаками (при парной связи) и между результативным и множеством факторных признаков (при многофакторной связи).

Теснота связи количественно выражается величиной коэффициентов корреляции. Коэффициенты корреляции, представляя количественную характеристику тесноты связи между признаками, дают возможность определять «полезность» факторных признаков при построении уравнений множественной регрессии. Величина коэффициента корреляции служит также оценкой соответствия уравнения регрессии выявленным причинно-следственным связям.

Первоначально исследования корреляции проводились в биологии, а позднее распространились и на другие области, в том числе на социально-экономическую. Одновременно с корреляцией начала использоваться и регрессия. Корреляция и регрессия тесно связаны между собой: первая оценивает силу (тесноту) статистической связи, вторая исследует ее форму. Та и другая служат для установления соотношения между явлениями, для определения наличия или отсутствия связи.

Корреляционно-регрессионный анализ как общее понятие включает в себя измерение тесноты, направления связи и установление аналитического выражения (формы) связи (регрессионный анализ).

Регрессионный анализ заключается в определении аналитического выражения связи, в котором изменение одной величины (называемой зависимой или результативным признаком) обусловлено влиянием одной или нескольких независимых величин (факторов), а множество всех прочих факторов, также оказывающих влияние на зависимую величину, принимается за постоянные и средние значения. Регрессия может быть однофакторной (парной) и многофакторной (множественной).

По форме зависимости различают:

а) линейную регрессию, которая выражается уравнением прямой (линейной функцией) вида: $Y_x = a_0 + a_1 x$;

б) нелинейную регрессию, которая выражается уравнениями вида:

парабола - $Y_x = a_0 + a_1x + a_2x^2$.

По направлению связи различают:

а) прямую регрессию (положительную), возникающую при условии, если с увеличением или уменьшением независимой величины значения зависимой также соответственно увеличиваются или уменьшаются;

б) обратную (отрицательную) регрессию, появляющуюся при условии, что с увеличением или уменьшением независимой величины зависимая соответственно уменьшается или увеличивается.