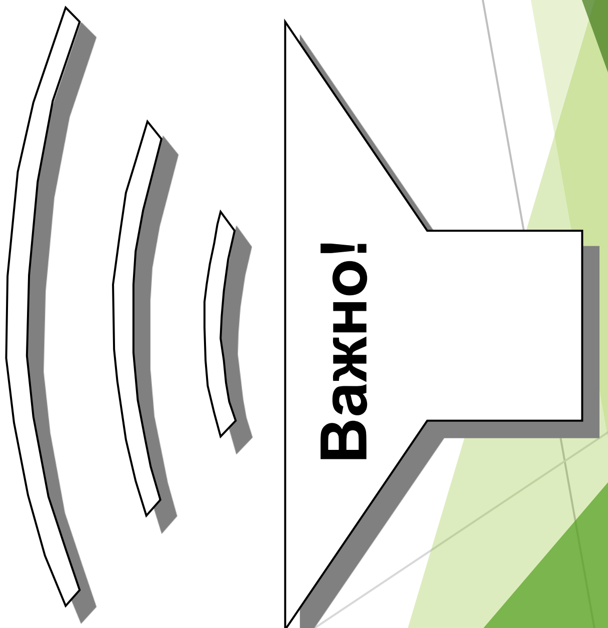
Парная регрессия и корреляция

Если выполняется неравенство $t > t_{\ell}$, то значение коэффициента корреляции признается значимым, т.е нулевая гипотеза, утверждающая равенство нулю коэффициента Корреляции, отвергается и делается вывод о том, что между исследуемыми переменными есть тесная статистическая связь



Сущность регрессионного анализа

Регрессионный анализ заключается в определении аналитической формы связи, в которой изменение результативного признака обусловлено влиянием одного или нескольких факторных признаков, а множество всех прочих факторов, также оказывающих влияние на результативный признак, принимается за постоянные и средние значения

Цель

Оценка функциональной зависимости условного среднего значения результативного признака от факторных признаков.

Основной предпосылкой регрессионного анализа

Является то, что только результативный признак подчиняется нормальному закону распределения, а факторные признаки – произвольному закону распределения.

Уравнение регрессии

Уравнение регрессии, или модель связи социально-экономических явлений, выражается функцией

$\widehat{y_x} = f(x)$

Парная регрессия (характеризует связь между двумя признаками: результативным и факторным)

$\widehat{y_x} = (x_1, x_2, \ldots x_k)$

Где k - число факторных признаков

Множественная регрессия (характеризует связь между результативным признаком и двумя и более факторными признаками)

Уравнение адекватно реальному моделируемому явлению или процессу в случае соблюдению требований его построения.

Требования к построению уравнения регрессии

- Совокупность исходных данных должна быть однородной и математически описываться непрерывными функциями
- Наличие достаточно большого объема исследуемой выборочной совокупности
- Возможность описания моделируемого явления одним или несколькими уравнениями причинно-следственных связей
- Причинно-следственные связи между явлениями и процессами, по возможности, следует описывать линейной (или приводимой к линейной) формой зависимости
- Отсутствие количественных ограничений на параметры модели
- Количественное выражение факторных признаков
- Постоянство территориальной и временной структуры изучаемой совокупности

Теоретическая обоснованность моделей

Условия Теоретической Обоснованности моделей

Все признаки и их совместные распределения должны починяться нормальному закону распределения

Дисперсия моделируемого признака должна все время оставаться постоянной при изменении величины и значений факторных признаков

Отдельные наблюдения должны быть независимыми, т.е. результаты, полученные в i-м наблюдении, не должны быть связаны с предыдущими о последующих наблюдениях, а также влиять на них

Уравнение линейной парной регрессии

Уравнение линейной парной регрессии

Где a_0, a_1 - параметры модели;
ϵ_i - случайная величина (величина остатка)

Параметры модели и их содержание

Параметр	Содержание параметра
	Свободный коэффициент (член) регрессионного уравнения. Не имеет экономического смысла и показывает значение результативного признака у, если факторный признак х=0
	Независимая, нормально распределенная случайная величина, остаток с нулевым математическим ожиданием и постоянной дисперсией. Отражает тот факт, что изменение у будет неточно описываться изменением х, т.к. присутствуют другие факторы, не учтенные в данной модели.

Система нормальных уравнений для нахождения параметра линейной парной регрессии методом наименьших квадратов

Система нормальных уравнений

$$na_0 + a_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i\,;\ a_0 \sum_{i=1}^{n} x_i + a_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i$$

Формулы для определения значения параметров a_0 и a_1

Расчет параметров a_0 и a_1

$$a_1 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})(y_i-\bar{y})}{\sum_{i=1}^{n}(x_i-\bar{x})^2};\ a_0 = \bar{y} - a_1\bar{x}$$

Формулы определения коэффициента эластичности

Коэффициент эластичности

$$Э_{yx} = a_1 \frac{\bar{x}}{\bar{y}}$$

Формула определения бета-коэффициента

Бета-коэффициент

$\beta_{yx} = a_1 \frac{\sigma_x}{\sigma_y}$; где σ_x и σ_y - средние квадратические отклонения случайных величин x и y

Важно!

Бета-коэффициент показывает, на какую часть своего среднего квадратического отклонения изменится в среднем значение результативного признака при изменении факторного признака на величину своего среднего квадратического отклонения

Проверка адекватности и точности уравнения регресии

Проверка адекватности модели

Проверка значимости модели

Установление наличия или отсутствия систематической ошибки

Формулы для определения t-критерия Стьюдента

t – критерий Стьюдента для оценки коэффициентов регрессии

$$t_{\text{расч}a_0} = \frac{a_0}{S_{a_0}};\ t_{\text{расч}a_1} = \frac{a_1}{S_{a_1}};\ S_{a_0} = S_\varepsilon\sqrt{\frac{\sum_{i=1}^{n} x_i^2}{n\sum_{i=1}^{n}(x_i-\bar{x})^2}};\ S_{a_1} = \frac{S_\varepsilon}{\sqrt{\sum_{i=1}^{n}(x_i-\bar{x})^2}};$$

$$S_\varepsilon\sqrt{\frac{\sum_{i=1}^{n}\varepsilon^2}{n-k-1}}$$

Формула определения F-критерия Фишера

$$F_{\text{расч}} = \frac{r_{yx}^2}{1 - r_{yx}^2} * (n - k - 1)$$


Требования, при которых модель считается адекватной
Уровни ряда остатков имеют случайный характер
Математическое ожидание уровней ряда остатков равно нулю
Дисперсия каждого отклонения E: одинакова для всех значений x_1
Уровни ряда остатков распределены по нормальному закону
Значения уровней ряда остатков независимы друг от друга

Соблюдение требований, которым должен удовлетворять ряд остатков

Требование	Метод проверки требований
Первое	
Второе	

Требование	Метод проверки требований
Третье	
Четвертое	

Требование	Метод проверки требований
Пятое	

Спасибо за внимание!