**Вопросы для защиты индивидуальных домашних заданий 6-8.**

1. Системы случайных величин. Их законы распределения.

Системы-наборы случайных величин, рассматриваемых совместно. Могут быть дискретными и непрерывными.

Случайная величина – функция, ставит каждому элементарному исходу вещественное число. Дискретная – конечное число значений. Непрерывная – принимает любое значение из интервала.

Закон распределения для дискретной СВ – функция . X – случайная величина, x – ее значение, p – вероятности. В сумме вероятности равны 1.

Закон распределения для непрерывной СВ – функция плотности вероятности.

Совместное распределение – вероятность одновременного принятия ими различных значений. Для дискретных: . Для непрерывных с помощью плотности вероятности.

Ковариация – степень линейной зависимости между двумя СВ. . Корреляция – измеряет степень линейной зависимости в пределах [-1;1]. .

1. Условные законы распределения. Независимость случайных величин.

Условный закон распределения СВ описывает распределение одной величины при условии, что значения другой известны.

Для двух дискретных СВ X и Y условное распределение X при условии определяется, как , где 𝑃(𝑋=𝑥\_𝑖,𝑌=𝑦\_𝑗) — совместное распределение, а 𝑃(𝑌=𝑦\_𝑗)— маргинальное распределение *Y*.

Для двух непрерывных СВ X и Y условное распределение X при условии Y=y определяется, как , где – совместная плотность распределния, а – маргинальная плотность Y.

Независимость СВ – значение одной СВ не дает информации о значении другой. Для дискретных:

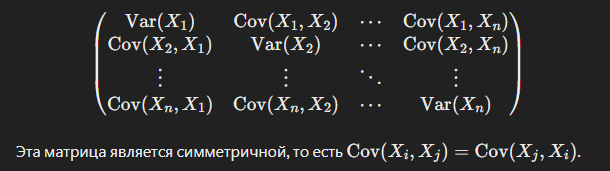
1. Числовые характеристики системы случайных величин. Ковариация и коэффициент корреляции, их свойства. Матрица корреляционных моментов. Зависимость и коррелированность. Условное математическое ожидание, функция регрессии.

Матожидание СВ (среднее значение) – E[X], представляет среднее значение, которое принимает X. .

Дисперсия СВ – D[X], показывает разброс значений СВ относительно мат. ожидания. .

Ковариация – см выше. Если она положительна, величины имеют тенденцию изменяться вместе. Если она отрицательна, то одна СВ увеличивается, когда другая уменьшается.

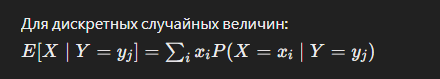
Коэффициент корреляции – см выше. Если он равен 1, то полная положительная линейная зависимость, если он равен -1, то полная отрицательная линейная зависимость. Если равен 0, зависмомть отсутсвует.

Матрица корреляционных моментов – ковариации между всеми парами случайных величин в системе. 

Зависимость между случайными величинами означает, что значения одной случайной величины оказывают влияние на значения другой. Зависимость может быть любой, не обязательно линейной.

Коррелированность — это мера линейной зависимости между случайными величинами. Две случайные величины могут быть зависимыми, но некоррелированными, если их связь нелинейная.

Условное матожидание E[X|Y=y] СВ X при условии, что Y принимает значение y - среднее значение X при известном Y.



Функция регрессии описывает зависимость матожидания X от Y. Для линейной регрессии E[X|Y] = a\_0 + a\_1\*Y.

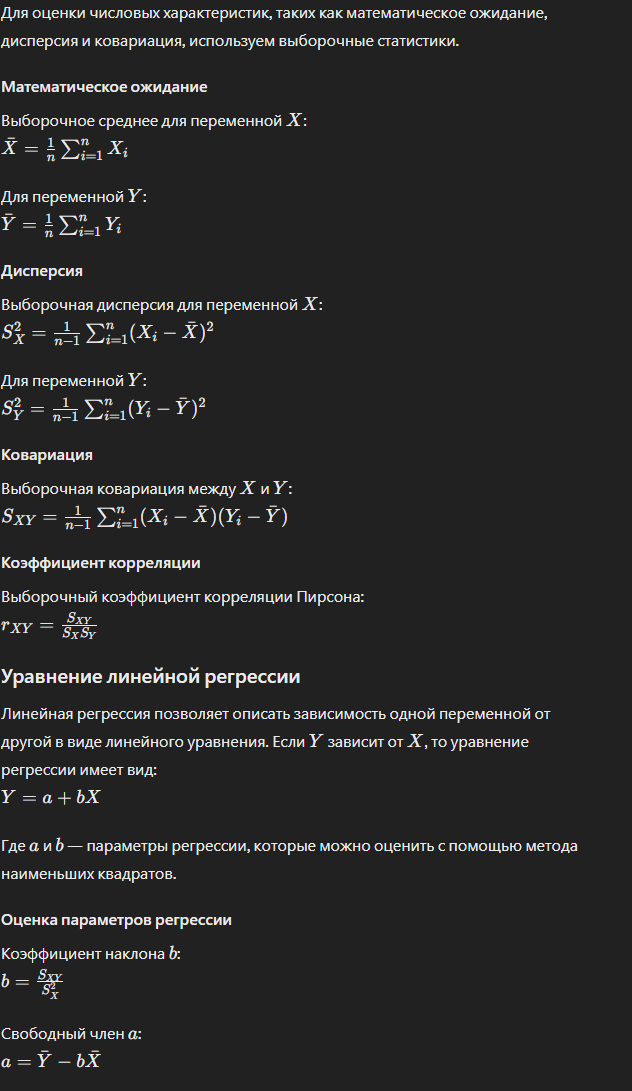
1. Как составляется корреляционная таблица для двумерной выборки?

Корреляционная таблица (таблица сопряжённости) используется для анализа связи между двумя переменными в двумерной выборке. Она показывает, как частоты одной переменной распределены в зависимости от значений другой переменной. В этой таблице содержатся значения корреляции между всеми парами переменных.

|  |  |  |
| --- | --- | --- |
|  | X | Y |
| X | 1 | pX,Y |
| Y | pX,Y | 1 |

1. Построение условного и безусловного вариационного ряда по корреляционной таблице. Нахождение оценок числовых характеристик системы. Уравнение линейной регрессии.

Вариационный ряд — это упорядоченный по возрастанию или убыванию ряд значений выборки. Безусловный вариационный ряд строится для всей выборки, а условный вариационный ряд строится для подвыборок при фиксированных значениях другой переменной.

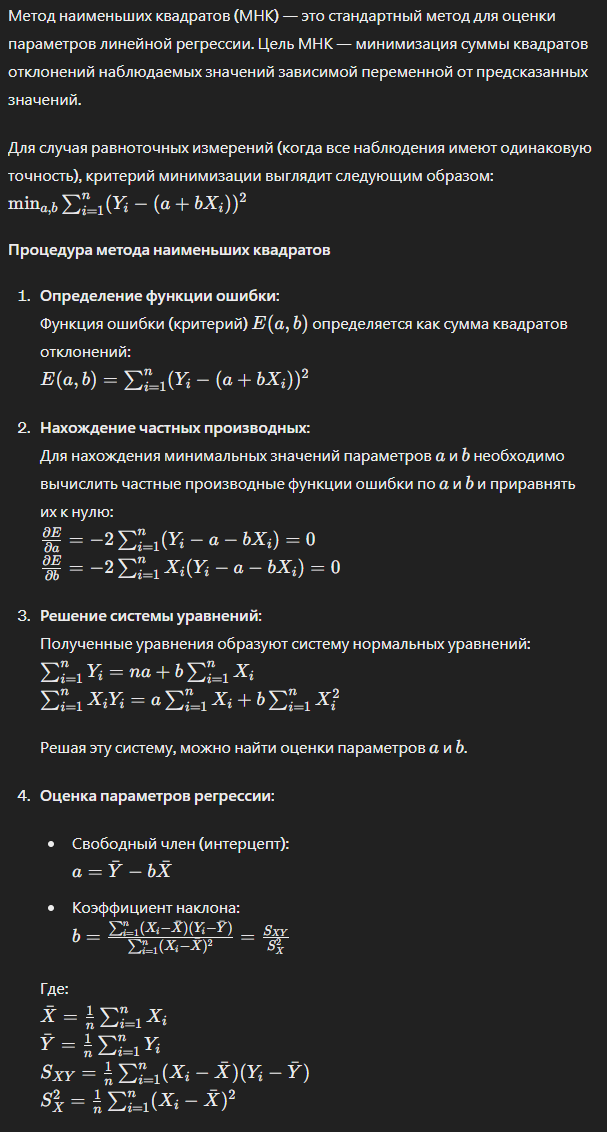
Оценки числовых характеристик и линейная регрессия.  


1. Задача регрессии. Метод наименьших квадратов для случая равноточных измерений. Линейная регрессия.

Задача регрессии заключается в нахождении зависимости между зависимой переменной Y и одной или несколькими независимыми переменными X. Основная цель регрессионного анализа — построение модели, которая позволяет предсказывать значения зависимой переменной на основе значений независимых переменных.

Линейная регрессия

Линейная регрессия — это метод для моделирования линейной зависимости между зависимой переменной Y и одной или несколькими независимыми переменными X. Уравнение линейной регрессии имеет вид: 𝑌 = 𝑎+𝑏𝑋



Либо через матрицы



Потом уравнение y = a\_0+a\_1\*x

Свойства оценок МНК

Несмещённость: Оценки a и b, полученные методом наименьших квадратов, являются несмещёнными, что означает, что в среднем они равны истинным значениям параметров.

Эффективность: Эти оценки имеют наименьшую возможную дисперсию среди всех несмещённых оценок.

Состоятельность: При увеличении объема выборки оценки, полученные методом наименьших квадратов, сходятся к истинным значениям

1. Статистическая гипотеза об адекватности модели в задаче регрессии.

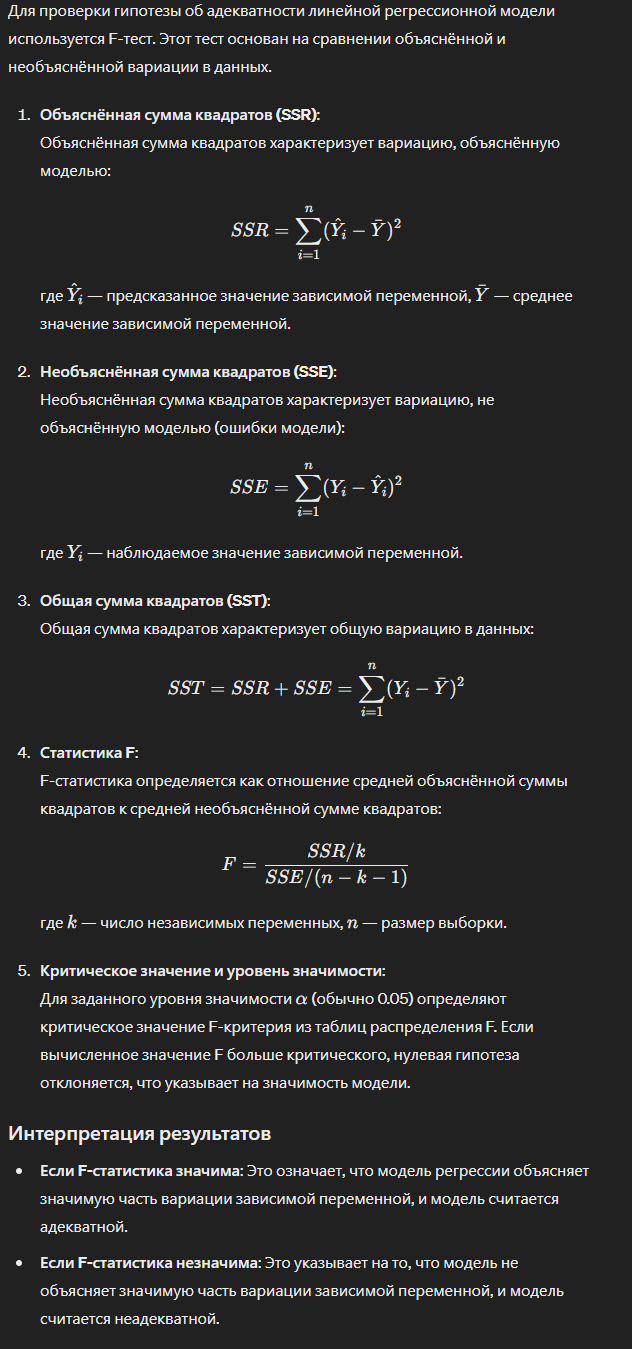
При оценке модели регрессии важно проверить, насколько хорошо модель объясняет данные. Для этого используется тестирование гипотез об адекватности модели. Основная цель — определить, объясняет ли модель значимую часть вариации зависимой переменной или нет.

Гипотеза об адекватности модели

Нулевая гипотеза (H₀): Регрессионная модель неадекватна, то есть независимая переменная(ые) не оказывает значимого влияния на зависимую переменную.

Альтернативная гипотеза (H₁): Регрессионная модель адекватна, то есть независимая переменная(ые) оказывает значимое влияние на зависимую переменную.

Критерий F-теста



Основные предположения F-теста

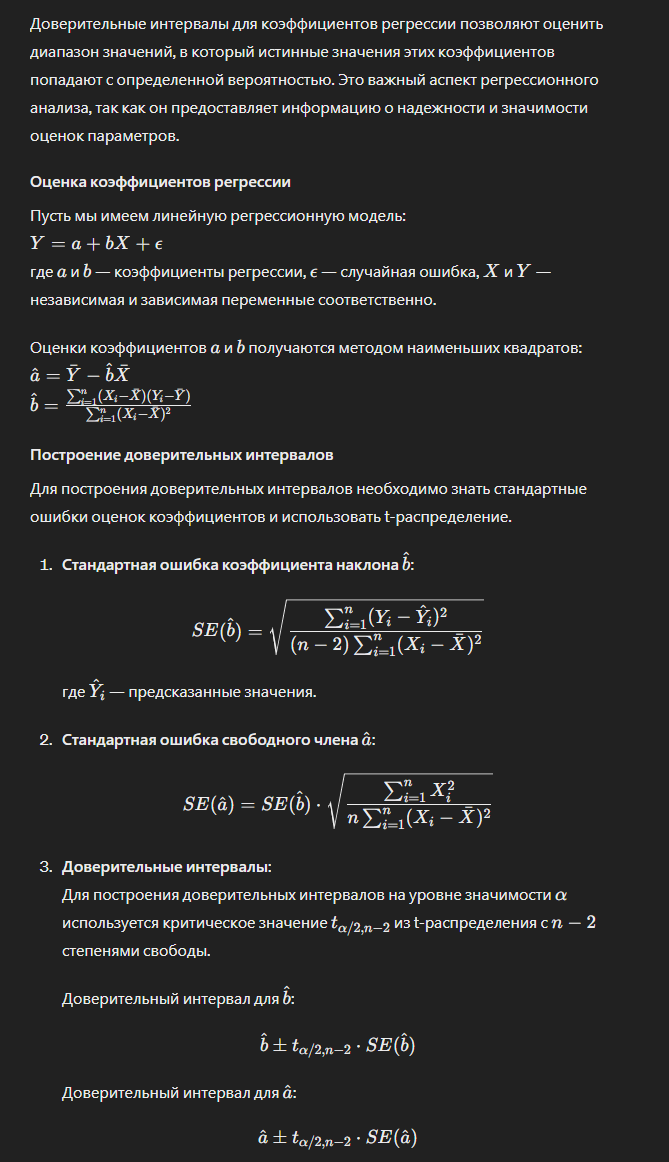
Линейность: Связь между зависимой и независимыми переменными должна быть линейной.

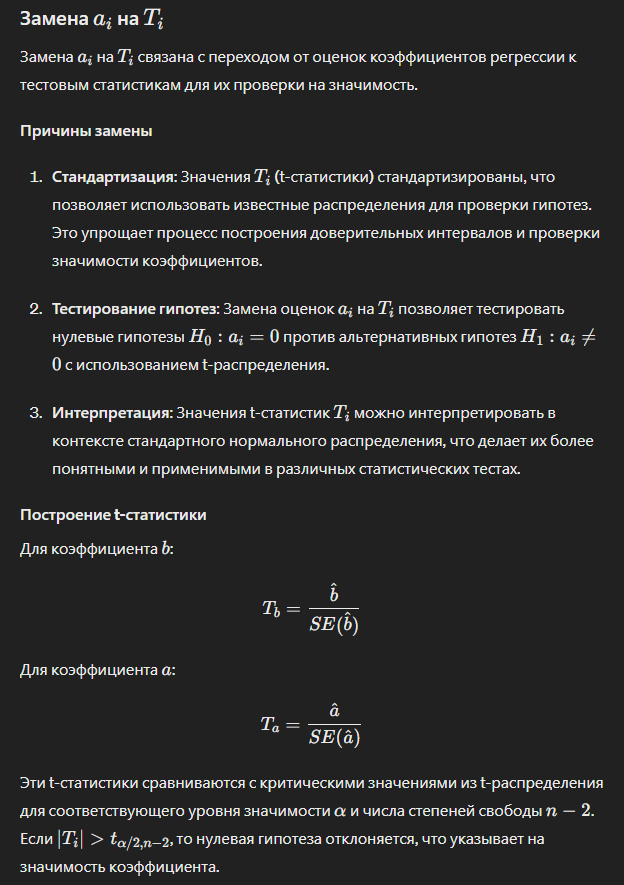
Независимость ошибок: Ошибки модели должны быть независимы.

Нормальность ошибок: Ошибки должны быть нормально распределены.

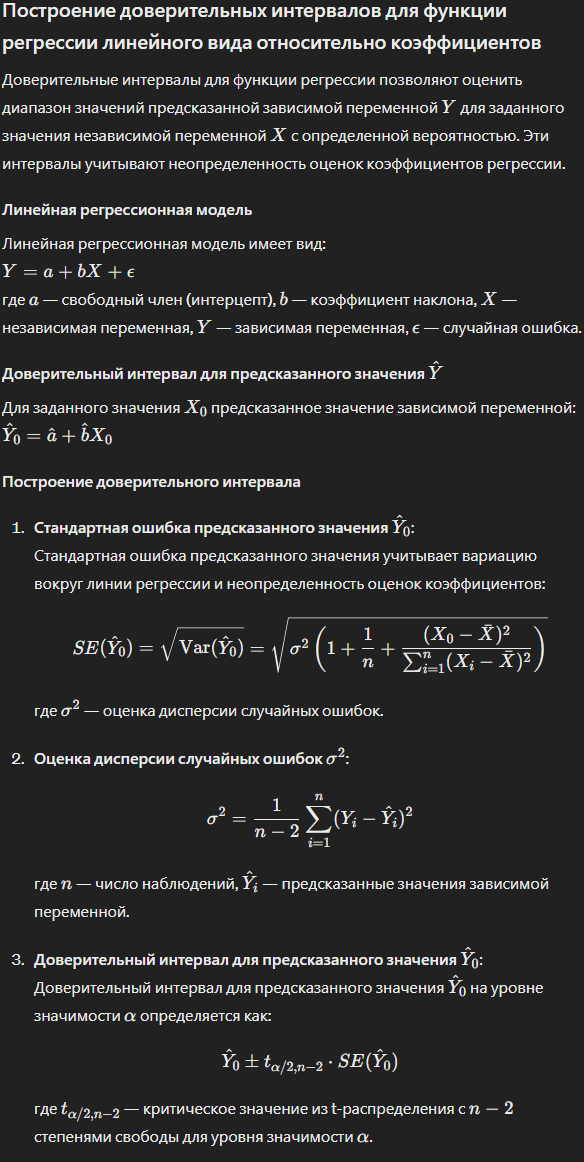
Гомоскедастичность: Дисперсия ошибок должна быть постоянной (гомоскедастичность).

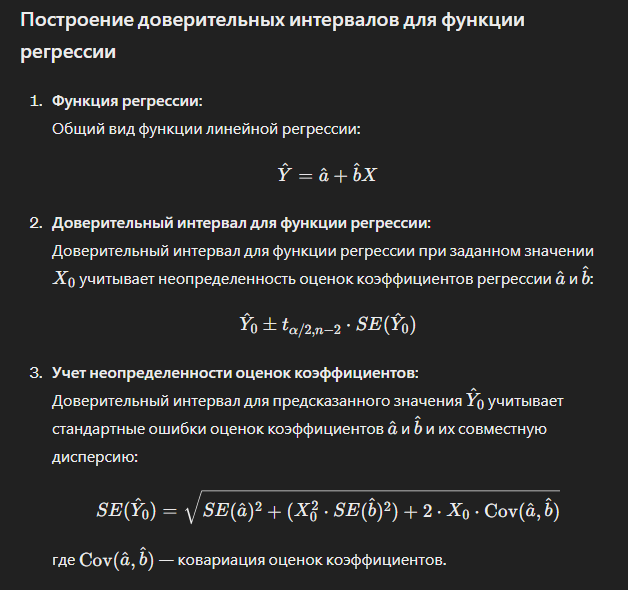
1. Построение доверительных интервалов для коэффициентов функции регрессии. Зачем делалась замена  на ?





1. Построение доверительных интервалов для функции регрессии линейного вида относительно коэффициентов.

****

****

1. Нелинейная регрессия. Уметь составлять систему для нахождения оценок коэффициентов функции регрессии произвольного вида.

См 8 лабу