Данные для разных вариантов содержатся в указанных в условии файлах, номер варианта отражается в окончаниях названия переменных. Так, данные для варианта 3 хранятся в переменных y\_3, x2\_3, price\_3, brick\_3 и т. д.

К решению обеих задач напишите пояснительный текст: постарайтесь ясно и грамотно описать свои действия и полученные результаты.

При проверке гипотез формулируйте основную и альтернативную гипотезы, ваше решение относительно отвержения или принятия основной гипотезы, приводите основание для этого решения (включая ваши расчёты или результаты применения статистических программ).

В задаче №5 требуется выбрать функциональную форму. О том, как это делается, можете почитать тут:

<https://github.com/bdemeshev/epsilon/raw/master/e_001/functional-form/functional-form.pdf>

**Задача №4.**

Используйте данные, соответствующие вашему варианту, из файла «Данные к задаче 4 дз (мультиколлинеарность).xls». Вариант, как и раньше, отмечается окончанием в названии переменных: \_1, \_2, …, \_10.

*Часть I.* Оцените линейную зависимость **y** от **x2**, **x3** и **x4** методом наименьших квадратов. Проверьте значимость регрессии в целом и значимость коэффициентов при объясняющих переменных по отдельности. Проверьте гипотезу о совместной значимости коэффициентов при переменных **x3** и **x4**.

Рассчитайте корреляционную матрицу для объясняющих переменных.

*Часть II.*

**Кое-какие сведения.** В случае, когда между объясняющими переменными возникает нестрогая мультиколлинеарность (зависимость, очень близкая к линейной), оценки МНК при тесно коррелирующих переменных имеют низкую точность (высокую дисперсию) и чувствительны к небольшим изменениям в данных. В такой ситуации вместо обычной оценки может быть использована оценка , где D — некая матрица (обычно диагональная с неотрицательными элементами на главной диагонали). Оценка называется оценкой ридж-регрессии, она смещённая, но имеют низкую дисперсию по сравнению с оценкой МНК, поэтому может иметь и меньшее значение MSE, если правильно подобрать матрицу D.

Рассмотрим частный случай ридж-регрессии — оценку , то есть на главной диагонали матрицы D ставится одно и то же число λ. Попробуйте оценить зависимость **y** от **x2**, **x3** и **x4** с помощью оценки , подставляя значения λ от 0 до 2 с шагом 0.1.

Постройте график зависимости оценок коэффициентов регрессии от λ. Что происходит с оценками, если дальше увеличивать λ?

Полезные команды в Stata: **regress**, **test**, **correlate** (или **pwcorr**), **matrix**.

Не знаю, есть ли в Stata готовая команда для ридж-регрессии, но её легко сделать самому. Также не знаю, есть ли красивый и удобный способ перенести наблюдения в матрицу, но можно делать в лоб:

matrix X=J(40,4,1)

matrix Y=J(40,1,0)

forval i=1/40 {

matrix Y[`i',1]=y[`i']

forval j=2/4 {

matrix X[`i',`j']=x`j'[`i']

}

}

Здесь J(n,m,x) – функция, возвращающая матрицу n на m, в которой каждый элемент равен x.

**Задача №5.**

Файл «Данные к задаче 5 дз.xls» содержит данные о стоимости квартир в Москве, собранные студентами РЭШ осенью 1997 г.

|  |  |
| --- | --- |
| **totsp** | общая площадь квартиры (в кв.м); |
| **price** | цена квартиры в долларах США; |
| **d2** | =1, если квартира двухкомнатная, 0 иначе; |
| **d3** | =1, если квартира трёхкомнатная, 0 иначе; |
| **d4** | =1, если квартира четырёхкомнатная, 0 иначе; |
| **dist** | расстояние от центра Москвы (в км); |
| **walk** | =1, если до метро можно быстро дойти пешком, 0 иначе; |
| **brick** | =1, если дом кирпичный, 0 иначе; |
| **bal** | =1, если есть балкон, 0 иначе; |
| **floor** | =0, если этаж первый или последний, 1 иначе. |

Комнат в каждой квартире было не больше четырёх, так что d2=d3=d4=0 для однокомнатных квартир.

Постройте регрессионную модель, связывающую цену квартиры с её харакеристиками, выберите подходящую функциональную форму. Возможно, вам не удастся построить совершенную модель, проходящую тесты на правильность функциональной формы, не надо пытаться получить идеальный результат любой ценой — пусть итоговая регрессия не точно описывает данные, но не имеет бросающихся в глаза недостатков и позволяет осмыслить результаты оценивания.

Опишите полученные результаты — попробуйте дать интерпретацию каждому значимому коэффициенту.

Рассчитайте прогноз цены однокомнатной квартиры с указанными характеристиками:

* площадь 40 кв.м.,
* есть балконом,
* кирпичный девятиэтажный дом,
* третий этаж,
* в 10 км от центра Москвы,
* рядом с метро.