

Программа по эконометрике

Борис Демешев

2019-04-04

Темы

Тема 1. Множественная регрессия без статистических предпосылок

Задача оптимизации. Правила работы с матричным дифференциалом. Геометрическая интерпретация. Показатели RSS, ESS, TSS, R^2 .

Тема 2. Множественная регрессия с предпосылками на дисперсию

Свойства ковариационных матриц. Теорема Гаусса-Маркова с доказательством.

Тема 3. Множественная регрессия и нормальные остатки

Тестирование гипотезы об отдельном коэффициенте. Ограниченная и неограниченная модель. F-тест для выбора вложенных моделей.

Тема 4. Гетероскедастичность

Нахождение эффективных оценок при известной форме гетероскедастичности. Корректировка стандартных ошибок при неизвестной форме гетероскедастичности. Тест Уайта на гетероскедастичность

Тема 5. Метод максимального правдоподобия

Идея метода максимального правдоподобия. Информация Фишера. Тест отношения правдоподобия. Тест Вальда

Тема 6. Модель логистической регрессии

Предпосылки модели. Оценка коэффициентов и их стандартных ошибок. Интерпретация коэффициентов. Нахождение предельных эффектов.

Тема 7. Модели одномерных временных рядов

ARIMA-модель. ETS-модель. Уравнение модели. Алгоритм подбора гиперпараметров. Прогнозирование в рамках модели.

Литература

- Никита Артамонов, Введение в эконометрику. Курс лекций.
- Kurt Schmidheiny, Short guides on econometrics, schmidheiny.name/teaching/shortguides.htm
- Michael Creel, Econometric Lecture notes, econpapers.repec.org/paper/aubautbar/575.03.htm
- Материалы курса эконометрики ВШЭ, bdemeshev.github.io/em301/

Формула оценивания

Итоговая оценка = 0.4 * Домашнее задание + 0.6 * Письменный экзамен

Пример домашнего задания

Пройдите курсы на datacamp:

<https://www.datacamp.com/courses/introduction-to-the-tidyverse>,

<https://www.datacamp.com/courses/multiple-and-logistic-regression>,

<https://www.datacamp.com/courses/communicating-with-data-in-the-tidyverse>,

<https://www.datacamp.com/courses/forecasting-using-r>

Пример экзамена

Задача 1.

Майор прониин наблюдает хочет оценить модель $y_t = \beta x_t + u_t$, величины u_t независимы с нулевым ожиданием и дисперсией пропорциональной $1/x_t$. Известны наблюдения:

y_t	1	2	3
x_t	1	1	2

1. Оцените параметр β с помощью МНК. Является ли оценка $\hat{\beta}_{ols}$ несмещённой?
2. Найдите самую эффективную линейную оценку параметра β , $\hat{\beta}_{eff}$?
3. Найдите стандартную ошибку $se(\hat{\beta}_{eff})$.

Задача 2.

Билл Гейтс оценил регрессию $\hat{Y}_i = 4 + 0.4X_i + 0.9W_i$, $RSS = 520$, $R^2 = 2/15$.

Про матрицу регрессоров X известно, что

$$X'X = \begin{pmatrix} 29 & 0 & 0 \\ 0 & 50 & 10 \\ 0 & 10 & 80 \end{pmatrix}$$

1. Сколько наблюдений было у Билла Гейтса?
2. Найдите выборочное среднее переменных X , W и Y .

3. Постройте 95%-й доверительный интервал для фактического значения зависимой переменной при $X = 1$ и $W = 3$.

Задача 3.

По 200 наблюдениям исследователь Иннокентий строит модель зависимости финального балла за метрику от числа часов подготовки для разных студентов.

Он оценил одну и ту же модель на трёх выборках

выборка	уравнение	RSS	наблюдений
все студенты	$\hat{y}_i = 40 + 2x_i$	7000	200
любители пиццы	$\hat{y}_i = 50 + 3x_i$	3000	100
нелюбители пиццы	$\hat{y}_i = 37 + 1.8x_i$	2000	100

1. Переменная d_i равна 1 для любителей пиццы и 0 для остальных студентов. Какие оценки коэффициентов получит Иннокентий при оценке модели по всей выборке?

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_2 x_i + \hat{\beta}_3 d_i + \hat{\beta}_4 d_i x_i$$

2. Протестируйте гипотезу о том, что симпатия к пицце не влияет на ожидаемый балл по эконометрике.

Задача 4.

Джеймс Бонд наблюдает значения независимых случайных величин y_1, y_2, \dots, y_{400} , закон распределения которых задан табличкой:

y_i	1	2	3
вероятность	$2a$	$3a$	$1 - 5a$

Из 400 наблюдений: 100 наблюдений равны 1, 100 наблюдений равны 2, и 200 наблюдений равны 3.

- Оцените неизвестный параметр a методом моментов и методом максимального правдоподобия.
- Постройте 95%-й доверительный интервал для параметра a .

Задача 5.

По 200 наблюдениям исследователь Иннокентий оценил модель логистической регрессии для вероятности сдать экзамен по метрике:

$$\hat{P}(Y_i = 1) = \Lambda(1.5 + 0.3X_i - 0.4D_i),$$

где Y_i — бинарная переменная равная 1, если студент сдал экзамен; X_i — количество часов подготовки студента; D_i — бинарная переменная равная 1, если студент любит пиццу.

Оценка ковариационной матрицы оценок коэффициентов имеет вид:

$$\begin{pmatrix} 0.04 & -0.01 & 0 \\ -0.01 & 0.01 & 0 \\ 0 & 0 & 0.09 \end{pmatrix}$$

1. Проверьте гипотезу о том, что количество часов подготовки не влияет на вероятность сдать экзамен.

2. Посчитайте предельный эффект увеличения каждого регрессора на вероятность сдать экзамен для студента не любящего пиццу и готовившегося 24 часа. Кратко, одной-двумя фразами, прокомментируйте смысл полученных цифр.
3. Постройте 95%-й доверительный интервал для разницы вероятностей сдать экзамен двумя студентами, если оба студента готовились 20 часов, однако один любит пиццу, а второй — нет.
4. При каком значении D_i предельный эффект увеличения X_i на вероятность сдать экзамен максимален, если $X_i = 20$?