Программа учебной дисциплины «Эконометрика»

Утверждена
Академическим советом ООП
Протокол № 2.3-09/ 2706-01 от «27 »июня 2018г.

Автор	Демешев Борис Борисович
Число кредитов	6
Контактная работа (час.)	56
Самостоятельная работа (час.)	172
Курс	1
Формат изучения дисциплины	без использования онлайн курса

I. ЦЕЛЬ, РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ И ПРЕРЕКВИЗИТЫ

Целями освоения дисциплины «Эконометрика» являются овладение студентами основными концепциями в области эконометрики.

В результате освоения дисциплины студент должен:

знать:

- теоретические свойства методы оценивания: метод наименьших квадратов и метод максимального правдоподобия
- границы применимости методов

уметь:

- оценивать множественную регрессию и корректно интерпретировать результаты оценивания
- корректировать результаты оценивания при наличии нарушений предпосылок теоремы Гаусса-Маркова
- оценивать модели с использованием метода максимального правдоподобия и корректно интепретировать результаты оценивания
- снижать размерность исходных данных с помощью метода главных компонент

Изучение дисциплины «Эконометрика» базируется на следующих дисциплинах:

- линейный алгебра в объёме типичного бакалаврского курса
- математический анализ в объёме типичного бакалаврского курса
- теория вероятностей и математическая статистика в объёме типичного бакалаврского курса

Для освоения учебной дисциплины студенты должны владеть навыками программирования в R, python или julia.

Основные положения дисциплины должны быть использованы в дальнейшем при изучении следующих дисциплин:

1 Макроэкономика
2 Прогнозирование временных рядов
3 Машинное обучение

II. СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Тема 1. Метод наименьших квадратов без вероятностных предпосылок.

Суть метода, матричный дифференциал, геометрический смысл условий первого порядка, геометрический смысл транспонирования, матрица-проектор, суммы TSS, ESS, RSS. Теорема Фриша-Во-Ловелла.

Тема 2. Метод главных компонент и сингулярное разложение.

Три постановки задачи метода главных компонент (минимизация расстояний, максимизация разброса, максимизация R^2). Суть сингулярного разложения. Эквивалентность SVD и PCA. SVD в методе наименьших квадратов.

Тема 3. МНК и предпосылки на ковариационную матрицу.

Работа с ковариационными матрицами. Геометрическое доказательство теоремы Гаусса-Маркова. Смысл обратной ковариационной матрицы.

Тема 4. Нормальное распределение и распределения, связанные с проекциями

Аксиомы Хершела-Максвелла, хи-квадрат, t, F распределения. Проверка гипотез в классической линейной регрессии.

Тема 5. Распределения, связанные с Пуассоновским потоком.

Работа с дифференциальными формами в теории вероятностей. Аксиоматика Пуассоновского потока. Экспоненциальное, гамма и бета распределения. Связь между гамма и хи-квадрат распределениями. Закон распределения R^2.

Тема 6. Гетероскедастичность.
Взвешенный МНК, Обобщённый МНК, робастные оценки ковариационной матрицы.

Тема 7. Эндогенность.

Предел по вероятности. Двухшаговый МНК. Метод инструментальных переменных.

Тема 8. Метод максимального правдоподобия.

Энтропия. Нормальное распределение — как распределение с максимальной энтропией. Аксиомы Гаусса для нормального распределения. Обращение блочных матриц.
ML для регрессии с единственным коэффициентом бета. ML для множественной регрессии. Три теста.

Тема 9. Логистическая регрессия.

Свойства логистического распределения. Дельта-метод. Модель логистической регрессии.

III. ОЦЕНИВАНИЕ

Все оценки выставляются по 10 балльной шкале по стандартным правилам округления.

Накопленная оценка = 0.5 * (Домашняя работа) + 0.5 * (Письменная контрольная работа)

Итоговая оценка = 0.5 * (Экзамен) + 0.5 * (Накопленная оценка)

IV. ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ

Пример домашнего задания:

Пройти курс Stastical Thinking in Python на www.datacamp.com

Пример письменной контрольной работы:

1. Найдите SVD-разложение матрицы $\begin{pmatrix} 0 & 2 & 0 \\ 1 & 1 & 1 \end{pmatrix}$.
2. В каких границах могут лежать диагональные элементы матрицы-шляпницы H? Чему равно их среднее значение?
3. Найдите дифференциал $d\frac{r^T A r}{r^T r}$, где $A^T = A$ — это матрица констант.
4. Постройте регрессию вектора $y = (4, 2, -2)^T$ на вектора $x = (1, 0, -1)^T$ и $z = (0, 1, 2)^T$ без константы. Найдите RSS, TSS, ESS. Верно ли в данной регрессии, что $TSS = ESS + RSS$ и почему?
5. Известно, что $y = x + 2z$. Храбрый исследователь Василий построил парные регрессии y на x и x на y. Нам частично известны результаты этих регрессий, $\hat{y}_i = \hat{\alpha}_1 + 16x_i$, $\hat{x}_i = \hat{\beta}_1 + 0.01y_i$.

   Найдите коэффициент $\hat{\gamma}_2$ в регрессии $\hat{x}_i = \hat{\gamma}_1 + \hat{\gamma}_2 z_i$.
6. Докажите, что в методе главных компонент с масштабированием переменных средняя величина R^2 по всем парным регрессиям исходных переменных на первую главную компоненту равна наибольшему сингулярному значению матрицы исходных переменных.
7. Поделим выборку на обучающую (X, y) и тестовую (X_{test}, y_{test}). Регрессоры X и X_{test} будем считать нестохастическими, а предпосылки теоремы Гаусса-Маркова — выполненными на всей исходной выборке.

   Найдите $\mathrm{Var}(\hat{y}_{test})$, $\mathrm{Cov}(\hat{\beta}, \hat{y}_{test})$.

Пример экзамена:

7. Величины U_1 и U_2 независимы и равномерны $U[0;1]$. Рассмотрим пару величин $Y_1 = R \cdot \cos\alpha$, $Y_2 = R \cdot \sin\alpha$, где $R = \sqrt{-2\ln U_1}$, а $\alpha = 2\pi U_2$.

   а) Выпишите дифференциальную форму для пары U_1, U_2;

   б) Выпишите дифференциальную форму для пары Y_1, Y_2;

   в) Найдите совместный закон распределения Y_1 и Y_2;

   г) Верно ли, что Y_1 и Y_2 независимы?

   д) Как распределены Y_1 и Y_2 по отдельности?

8. Эта задача посвящена доказательству неравенства Крамера-Рао. Суть его в том, что если мы возьмём любую несмещённую оценку, то её дисперсия будет не меньше некоторой границы. А именно, если $\hat{a}$ — любая несмещённая оценка вектора a, то матрица M,

$$M = \mathrm{Var}(s(a)) \cdot \mathrm{Var}(\hat{a}) - I_{k\times k}$$

   неотрицательно определена. В этой задаче $\hat{a}$ — произвольная несмещённая оценка, не обязательно равная $\hat{a}_{ML}$! Как обычно, $s(a)$ — градиент функции правдоподобия в истинной точке.

   а) Вспомните, чему равно $\mathrm{E}(s(a))$.

   б) Найдите скаляры $\mathrm{Cov}\left(\hat{a}_1, \frac{\partial \ell}{\partial a_1}\right)$, $\mathrm{Cov}\left(\hat{a}_1, \frac{\partial \ell}{\partial a_2}\right)$ и матрицу $\mathrm{Cov}\left(\hat{a}, s(a)\right)$.

   в) Рассмотрим два произвольных случайных вектора r и s и два вектора констант подходящей длины α и β. Найдите минимум функции $f(\alpha, \beta) = \mathrm{Var}(\alpha^T r + \beta^T s)$ по β. Выпишите явно $\beta^*(\alpha)$ и $f^*(\alpha)$.

   г) Докажите, что для произвольных случайных векторов положительно определена матрица

$$\mathrm{Var}(r) - \mathrm{Cov}(r, s)\,\mathrm{Var}^{-1}(s)\,\mathrm{Cov}(s, r)$$

   д) Завершите доказательство векторного неравенства Крамера-Рао.

1. Про законы распределения:

   а) Василий спроецировал n-мерный стандартный нормальный вектор u на линейную оболочку независимых векторов a, b и c. Квадрат длины проекции назовём буквой Z. Как распределена величина Z?

   б) Далее Василий спроецировал u на линейную оболочку векторов a и b. Квадрат длины проекции назовём буквой W. Как распределена величина W?

   в) Неугомонный Василий спроецировал u на линейную оболочку векторов d и e. Вектора d и e независимы и ортогональны a, b и c. Квадрат длины проекции назовём буквой Q. Как распределена величина Q?

   г) Какое известное распределение можно получить из Q/Z? Как конкретно его получить?

   д) Какое известное распределение можно получить из W/Z? Как конкретно его получить?

2. Рассмотрим модель множественной регрессии, $y = X\beta + u$, где регрессоры детерминистические, а $u \sim \mathcal{N}(0; \sigma^2 \cdot I)$. Величина σ^2 известна. Мы хотим проверить гипотезу H_0: $\beta = 0$.

   а) Выведите формулы для статистик W, LR, LM.

   б) Сравните эти статистики между собой.

3. Рассмотрим модель множественной регрессии, $y = X\beta + u$, где регрессоры детерминистические, а $u \sim \mathcal{N}(0; \sigma^2 \cdot I)$. Величина σ^2 неизвестна и тоже оценивается. Мы хотим проверить гипотезу H_0: $\beta = 0$.

   а) Выведите формулы для статистик W, LR, LM.

   б) Сравните эти статистики между собой.

4. Сэр Томас Байес в 18 веке решил задачу, которая на современном языке формулируется так: Величина R имеет равномерное распределение на отрезке $[0; 1]$. Мы изготавливаем монетку, выпадающую орлом с вероятностью R. Затем подбрасываем её n раз. Из этих n раз оказывается X орлов и Y решек.

   а) Как выглядит условная плотность величины R при известных X и Y с точностью до константы?

   б) Какова условная вероятность того, что монетка выпадет орлом, при известных X и Y?

   Хинт: какое там есть распределение-то на отрезке $[0; 1]$? А тут ещё две известных величины, X и Y завалялись :)

5. Вспомнив Матрицу-Мать-Всех-Регрессий, докажите, что в регрессии

$$\hat{y}_i = \hat{\beta}_1 + \hat{\beta}_x x_i + \hat{\beta}_z z_i + \hat{\beta}_w w_i$$

   величину R^2 можно разложить в сумму:

$$R^2 = \hat{\beta}_x \frac{\text{sCov}(y, x)}{\text{sVar}(y)} + \hat{\beta}_z \frac{\text{sCov}(y, z)}{\text{sVar}(y)} + \hat{\beta}_w \frac{\text{sCov}(y, w)}{\text{sVar}(y)}$$

V. РЕСУРСЫ

V.1Основная литература

1. Michael Creel, Econometrics [Электронный ресурс], режим доступа: https://econpapers.repec.org/paper/aubautbar/575.03.htm
2. Bruce Hansen, Econometrics [Электронный ресурс], режим доступа: https://www.ssc.wisc.edu/~bhansen/econometrics/

V.2 Дополнительная литература

1. John Stachursky, First course in Econometric theory [Электронный ресурс], режим доступа: http://lectures.quantecon.org/_downloads/course_notes.pdf

2. Kevin Sheppard, Financial Econometrics [Электронный ресурс], режим доступа: https://www.kevinsheppard.com/Category:MFE

V.3 Программное обеспечение

Студенты при прохождении курса могут использовать открытое программное обеспечение:

1. Дистрибутив языка программирования python "Anaconda", условие доступа: https://www.anaconda.com/

2. Дистрибутив языка программирования R, условие доступа: https://www.r-project.org/

3. IDE Rstudio, условие доступа: https://www.rstudio.com/

V5. Профессиональные базы данных, информационные справочные системы, интернет-ресурсы (электронные образовательные ресурсы)

1. Платформа "Открытое образование", условие доступа: httpts://openedu.ru

2. Платформа "Data Camp", условие доступа: https://www.datacamp.com

V6. Материально-техническое обеспечение дисциплины

Для проведения лекций нужен мультимедийный проектор с дистанционным управлением.