

Программа учебной дисциплины «Анализ категориальных данных в статистических пакетах»

Утверждена

Академическим советом ОП

Протокол № от __. __. 2019

Разработчик	Камалова Р. У., доцент, общеуниверситетская кафедра высшей математики (rkamalova@hse.ru)
Число кредитов	5
Контактная работа (час.)	56
Самостоятельная работа (час.)	134
Курс, Образовательная программа	3 курс, Политология, профиль «Политический анализ»
Формат изучения дисциплины	без использования онлайн курса

1. Цель, результаты освоения дисциплины и пререквизиты

Целью освоения дисциплины «Анализ категориальных данных в статистических пакетах» является выработка компетенций по решению политологических и социально - экономических задач, связанных с анализом данных в том случае, когда изучаемые зависимые признаки имеют дискретную природу.

В результате освоения дисциплины студент должен

знать:

- основные модели бинарного и множественного выбора;
- в каких областях применяются изученные модели

уметь:

- интерпретировать и представлять полученные результаты;
- решать задачи анализа политических и социально-экономических процессов;

владеть навыками их реализации в статистических пакетах (Rstudio).

Изучение дисциплины «Анализ категориальных данных в статистических пакетах» базируется на следующих дисциплинах:

- Математика и статистика (1 курс)
- Теория вероятностей и математическая статистика (2 курс)
- Дополнительные главы теории вероятностей и математической статистики (2 курс)

2. Содержание учебной дисциплины

Тема (раздел дисциплины)	Объем в часах	Планируемые результаты обучения (ПРО), подлежащие контролю	Формы контроля
	Семинары		
	Практические занятия		
	Сам. Работа		
Тема 1	10	Корректно интерпретирует модели бинарного выбора, знает область применения, методы диагностики этих моделей	Проверочная работа №1 (30 минут) Домашнее задание №1
	10		
	50		
Тема 2	8	Корректно применяет и интерпретирует результаты модели множественного неупорядоченного выбора	Проверочная работа №2 (30 минут) Домашнее задание №2
	8		
	40		
Тема 3	10	Корректно применяет и интерпретирует результаты модели множественного упорядоченного выбора	Проверочная работа №3 (30 минут) Домашнее задание №3
	10		
	44		
Часов по видам учебных занятий:	28		
	28		
	134		
Итого часов:	190		

Тема 1. Модели бинарного выбора

1.1 Линейная модель регрессии

Дискретные зависимые переменные. Латентный признак и наблюдаемые исходы.
Линейная вероятностная модель

1.2. Логистическая регрессия

Оценивание моделей с дискретными зависимыми переменными. Метод максимального правдоподобия (ММП). Модели бинарного выбора: логистическая регрессия (логит - модель) и пробит-модель. Условия на ошибки моделей бинарного выбора. Интерпретация коэффициентов логистической регрессии при непрерывных и категориальных переменных. Шансы и отношения шансов. Предсказанные вероятности. Доверительные интервалы. Графическое представление результатов. Сравнение коэффициентов логит - и пробит-моделей.

1.3. Проверка гипотез

Проверка гипотез о линейных ограничениях на коэффициенты: тест Вальда, тест отношения правдоподобия. Проверка гипотез об адекватности модели (критерии качества модели), сравнение моделей. (pseudo-R², AIC, BIC). Анализ остатков модели. Выявление нетипичных наблюдений.

Тема 2. Модели множественного неупорядоченного выбора

Номинальные переменные. Мультиномиальная логистическая регрессия. Интерпретация коэффициентов мультиномиальной логистической регрессии. Условная логистическая регрессия (Conditional logit). Предположение о независимости от посторонних альтернатив.

Тема 3. Модели множественного упорядоченного выбора

Порядковые переменные. Порядковая логистическая регрессия. Оценивание порядковой логистической регрессии. Интерпретация коэффициентов порядковой логистической регрессии. Предположение о параллельности регрессий: тест Бранта. Обобщенная модель множественного упорядоченного выбора (Generalized ordered logit) / частичная модель пропорциональных шансов (Partial proportional odds model).

3. Оценивание

Окончательная оценка по дисциплине рассчитывается в соответствии со следующей формулой:

$$\text{Итог.} = 0.15 \cdot \text{Одз.} + 0.15 \cdot \text{Осам.} + 0.15 \cdot \text{Ок/р} + 0.25 \cdot \text{Оэссе} + 0.3 \cdot \text{Оэкз.},$$

где *Одз.* – *оценка за домашние задания.* Оценка рассчитывается как среднее арифметическое оценок за текущие домашние задания. Округление данного компонента итоговой оценки арифметическое. Оценка за каждое домашнее задание выставляется по 10-балльной шкале. Передача домашних заданий не предусмотрена. При сдаче домашнего задания с опозданием штраф составляет 1 балл за каждый день просрочки.

Осам. – *оценка за самостоятельные работы.* Оценка рассчитывается как среднее арифметическое оценок за проверочные работы, выполняемые студентами на занятиях. Выполнение проверочных работ осуществляется в режиме closed-book. Оценка за каждую проверочную работу выставляется по 10-балльной шкале. Проверочные работы не переписываются.

Оэссе – *оценка за эссе по курсу.* Требования к написанию эссе представлены в следующем разделе. Оценка выставляется по 10-балльной шкале.

Ок/р. – *оценка за промежуточную контрольную работу.* Выполнение проверочных работ осуществляется в режиме closed-book. Оценка выставляется по 10-балльной шкале. Передача и переписывание контрольной работы не предусмотрено.

Оэкз. – *оценка за экзаменационную работу по курсу.* Оценка выставляется по 10-балльной шкале. Экзаменационная работа выполняется в режиме closed-book. Первая и вторая передачи экзамена предусматривают выполнение письменной работы, по заданиям схожим с экзаменационной работой по курсу. На передаче окончательная оценка по дисциплине выставляется по той же самой формуле, что и до передач.

4. Примеры оценочных средств

Блокирующие элементы по курсу не предусмотрены.

Эссе - самостоятельная (индивидуальная) письменная работа по результатам анализа массива опросных данных с использованием изученных методов на тему «Социально - экономические и психологические факторы электорального поведения россиян».

Вам даны три массива данных исследования «Курьер» Аналитического Центра Юрия Левады, относящиеся к трем разным периодам времени. Задача состоит в том, чтобы

построить три одинаковые и разумные модели для трех <<временных точек>> и сравнить полученные результаты. Сравнить в смысле неизменности порождающего данные процесса: выяснить, насколько существенно меняются коэффициенты и содержательные результаты, получаемые по регрессионным моделям.

Для анализа предлагаются следующие зависимые переменные:

1. Электоральный выбор.
2. Поддержка В. Путина и Д. Медведева.
3. Готовность участвовать в протестах.

Предикторы должны быть выбраны самостоятельно в соответствии с теоретическими соображениями (научной литературой).

5. Ресурсы

5.1. Рекомендуемая основная литература

1. Rudas T. Lectures on Categorical Data Analysis. Springer, New York, 2018. Available at: <https://proxylibrary.hse.ru:2120/10.1007/978-1-4939-7693-5>
2. Прикладная статистика и основы эконометрики: Учебник для вузов / С. А. Айвазян, В. С. Мхитарян. – М.: ЮНИТИ, 1998. – 1023 с.

5.2 Рекомендуемая дополнительная литература

1. Analysis of multivariate social science data / D. J. Bartholomew [et al.]. – 2nd ed. – Boca Raton; London; New York: CRC Press, 2008. – 371 с. – (Statistics in the social and behavioral sciences series) . – На англ. яз.
2. Burnham, Kenneth P. and David R. Anderson (2004). Multimodel Inference: Understanding AIC and BIC in Model Selection. Sociological Methods & Research, 33, pp. 261-304.
3. Путеводитель по современной эконометрике: учеб. пособие для вузов / М. Вербик; Пер. с англ. В. А. Банникова; Науч. ред., предисл. С. А. Айвазяна. – М.: Научная книга, 2008. – 615 с. – (Б-ка Солев) . – НП.

5.3 Программное обеспечение

№п/п	Наименование	Условия доступа/скачивания
1	RStudio	Свободное лицензионное соглашение

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

№п/п	Наименование	Условия доступа/скачивания
1	Открытое образование	URL: https://openedu.ru/

5.5 Материально-техническое обеспечение дисциплины

Учебные аудитории для семинарских занятий по дисциплине обеспечивают использование и демонстрацию тематических иллюстраций, соответствующих программе дисциплины в составе:

- ПЭВМ с доступом в Интернет (операционная система, офисные программы, антивирусные программы);
- мультимедийный проектор с дистанционным управлением.

Учебные аудитории для практических занятий по дисциплине оснащены ПЭВМ, с возможностью подключения к сети Интернет и доступом к электронной информационно-образовательной среде НИУ ВШЭ.

6 Особенности организации обучения для лиц с ограниченными возможностями здоровья и инвалидов

В случае необходимости, обучающимся из числа лиц с ограниченными возможностями здоровья (по заявлению обучающегося) предлагаются следующие варианты восприятия учебной информации, в том числе с применением электронного обучения и дистанционных технологий:

- 6.1 для лиц с нарушениями зрения:* в печатной форме увеличенным шрифтом; в форме электронного документа; индивидуальные задания и консультации.
- 6.2 для лиц с нарушениями слуха:* в печатной форме; в форме электронного документа; индивидуальные задания и консультации.
- 6.3 для лиц с нарушениями опорно-двигательного аппарата:* в печатной форме; в форме электронного документа; индивидуальные задания и консультации.