



Правительство Российской Федерации

Федеральное государственное автономное образовательное учреждение
высшего профессионального образования
«Национальный исследовательский университет
«Высшая школа экономики»

Факультет прикладной политологии

Программа дисциплины

Анализ категориальных данных в пакете Stata

для направления 030200.62 «Политология» подготовки бакалавра

Автор программы:

Камалова Рита Ульфатовна, преподаватель (rkamalova@hse.ru)

Одобрена на заседании кафедры высшей математики «__» _____ 2012 г.
Зав. кафедрой А.А. Макаров

Рекомендована секцией УМС «_____» «__» _____ 2012 г.
Председатель _____

Утверждена УС факультета прикладной политологии «__» _____ 2012 г.
Ученый секретарь _____

Москва, 2012

Настоящая программа не может быть использована другими подразделениями университета и другими вузами без разрешения кафедры-разработчика программы.



1 Область применения и нормативные ссылки

Настоящая программа учебной дисциплины устанавливает минимальные требования к знаниям и умениям студента и определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих данную дисциплину, учебных ассистентов и студентов направления 030200.62 «Политология» подготовки бакалавра, изучающих дисциплину «Анализ категориальных данных в пакете Stata».

Программа разработана в соответствии с рабочим учебным планом университета по направлению 030200.62 «Политология» подготовки бакалавра, утвержденным в 2012 г.

2 Цели освоения дисциплины

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

3 Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате освоения дисциплины студент должен:

- Знать основные модели бинарного и множественного выбора, позволяющие решать задачи анализа политических и социально-экономических процессов;
- Иметь навыки их реализации в пакете Stata;
- Уметь интерпретировать и представлять полученные результаты;
- Знать в каких областях применяются изученные модели.

4 Место дисциплины в структуре образовательной программы

Настоящая дисциплина относится к факультативам направления 030200.62 «Политология» подготовки бакалавра. Для указанного направления дисциплина является дисциплиной по выбору.

Изучение данной дисциплины базируется на следующих дисциплинах:

- «Алгебра и анализ» (1 курс),
- «Теория вероятностей, начала статистики и обработки данных» (2 курс),
- «Эконометрика для политологов» (3 курс).

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

- «Политический анализ»,
- «Политическое поведение»,
- «Анализ электоральных данных»,
- «Прогнозирование социально-политических процессов»,
- «Многомерные типологии политических систем современных государств».



5 Тематический план учебной дисциплины

№	Название раздела	Всего часов	Аудиторные часы		Самостоятельная работа
			Лекции	Практические занятия	
1	Модели бинарного выбора	48	14	14	20
2	Модели множественного упорядоченного выбора	30	8	8	14
3	Модели множественного неупорядоченного выбора	30	8	8	14
	Итого	108	30	30	48

6 Формы контроля знаний студентов

Тип контроля	Форма контроля	1 год				Параметры
		1	2	3	4	
Текущий (неделя)	Контрольная работа			*		Письменная работа, 80 минут.
	Домашнее задание				*	Самостоятельная письменная работа по итогам анализа массива политологических и социально-экономических данных с использованием изученных методов. Объем: 10-20 тыс. символов с пробелами.
Итоговый	Зачет				*	Письменный зачет, 100 минут.

6.1 Критерии оценки знаний, навыков

Оценки за домашние задания и зачет выставляются, исходя из следующих критериев:

- правильность решения задачи,
- полнота решения задачи,
- наличие и корректность интерпретации полученных результатов.

Оценки по всем формам текущего и итогового контроля выставляются по 10-ти балльной шкале.

7 Содержание дисциплины

Раздел 1. Модели бинарного выбора

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

Основная литература



1. Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage. Ch. 2, 3, 4. (Полочный индекс в библиотеке НИУ ВШЭ – 519 L87).
2. Long, J. Scott and Jeremy Freese. 2005. *Regression Models for Categorical Dependent Variables Using Stata*. 2nd Edition. College Station, TX: Stata Press. Ch. 3, 4. (Полочный индекс в библиотеке НИУ ВШЭ – 519 L87).
3. Stata Topics. Logistic (and Categorical) Regression UCLA: Academic Technology Services, Statistical Consulting Group. (http://www.ats.ucla.edu/stat/stata/topics/logistic_regression.htm).

Дополнительная литература

1. Вербик М. Путеводитель по современной эконометрике. Пер. с англ. В. А. Банникова. Под науч. ред. и предисл. С.А. Айвазяна. – М.: Научная книга, 2008. - 616 с. (Гл. 7 с. 296-316).
2. Cook, R. D. and Weisberg, S. (1999). Applied Regression Including Computing and Graphics. New York: Wiley. Ch. 21, 22.
3. Allison, P. D. (1999). Comparing logit and probit coefficients across groups. *Sociological Methods and Research* 28: 186–208.
4. Gujarati, D.N. Basic econometrics. New York McGraw-Hill, 2003. Ch. 15.1-15.10.
5. Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, 96(1):3-14. (<http://sta559s11.pbworks.com/w/file/fetch/37766848/IntroLogisticRegressionPengEducResearch.pdf>).
6. Peng, C.-Y. J., & So, T.-S. H. (2002). Logistic regression analysis and reporting: A primer. *Understanding Statistics*, 1(1), 31-70. (<http://www.indiana.edu/~jopeng51/teaching-logistic.pdf>).
7. Bo Hu, Jun Shao and Mari Palta (2006). Pseudo-R² in Logistic Regression Model, *Statistica Sinica* 16:847-860. (<http://www3.stat.sinica.edu.tw/statistica/oldpdf/A16n39.pdf>).
8. Logistic Regression with Stata. UCLA: Academic Technology Services, Statistical Consulting Group. (http://www.ats.ucla.edu/stat/stata/seminars/stata_logistic/default.htm).
9. Logistic Regression with Stata by Xiao Chen, Phil Ender, Michael Mitchell & Christine Wells. Stata Web Books. UCLA: Academic Technology Services, Statistical Consulting Group. (<http://www.ats.ucla.edu/stat/stata/webbooks/logistic/>).
10. What are pseudo R-squareds? UCLA: Academic Technology Services, Statistical Consulting Group. (http://www.ats.ucla.edu/stat/mult_pkg/faq/general/psuedo_rsquareds.htm).

Раздел 2. Модели множественного упорядоченного выбора

- Порядковые переменные. Порядковая логистическая регрессия. Оценивание порядковой логистической регрессии. Интерпретация коэффициентов порядковой логистической регрессии. Предположение о параллельности регрессий: тест Бранта.

Основная литература

1. Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage. Ch. 5. (Полочный индекс в библиотеке НИУ ВШЭ – 519 L87).
2. Long, J. Scott and Jeremy Freese. 2005. *Regression Models for Categorical Dependent Variables Using Stata*. 2nd Edition. College Station, TX: Stata Press. Ch. 3, 5. (Полочный индекс в библиотеке НИУ ВШЭ – 519 L87).
3. Stata Topics. Logistic (and Categorical) Regression UCLA: Academic Technology Services, Statistical Consulting Group. (http://www.ats.ucla.edu/stat/stata/topics/logistic_regression.htm).

Дополнительная литература



1. Вербик М. Путеводитель по современной эконометрике. Пер. с англ. В. А. Банникова. Под науч. ред. и предисл. С.А. Айвазяна. – М.: Научная книга, 2008. - 616 с. (Гл. 7 с. 316-324).
2. Beyond Binary: Ordinal Logistic Regression in Stata. UCLA: Academic Technology Services, Statistical Consulting Group. (http://www.ats.ucla.edu/stat/stata/seminars/stata_BeyondBinaryLogistic/Stata_ordinal_seminar.htm).

Раздел 3. Модели множественного неупорядоченного выбора

- Номинальные переменные. Мультиномиальная логистическая регрессия. Интерпретация коэффициентов мультиномиальной логистической регрессии. Предположение о независимости от посторонних альтернатив.

Основная литература

1. Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage. Ch. 6. (Полочный индекс в библиотеке НИУ ВШЭ – 519 L87).
2. Long, J. Scott and Jeremy Freese. 2005. *Regression Models for Categorical Dependent Variables Using Stata*. 2nd Edition. College Station, TX: Stata Press. Ch 3, 6. (Полочный индекс в библиотеке НИУ ВШЭ – 519 L87).
3. Stata Topics. Logistic (and Categorical) Regression UCLA: Academic Technology Services, Statistical Consulting Group. (http://www.ats.ucla.edu/stat/stata/topics/logistic_regression.htm).

Дополнительная литература

1. Вербик М. Путеводитель по современной эконометрике. Пер. с англ. В. А. Банникова. Под науч. ред. и предисл. С.А. Айвазяна. – М.: Научная книга, 2008. - 616 с. (Гл. 7 с. 324-329).
2. Stata Library. Understanding RR Ratios in Multinomial Logistic Regression. UCLA: Academic Technology Services, Statistical Consulting Group. (http://www.ats.ucla.edu/stat/stata/library/odds_ratio_mlogit.htm).
3. Stata Data Analysis Examples. Multinomial Logistic Regression. UCLA: Academic Technology Services, Statistical Consulting Group. (<http://www.ats.ucla.edu/stat/stata/dae/mlogit.htm>).
4. R Data Analysis Examples. Multinomial Logistic Regression. UCLA: Academic Technology Services, Statistical Consulting Group. (<http://www.ats.ucla.edu/stat/r/dae/mlogit.htm>).

8 Образовательные технологии

Занятия проводятся в форме лекций и практических занятий в компьютерном классе.

9 Порядок формирования оценок по дисциплине

Преподаватель оценивает работу студентов на практических занятиях: активность в дискуссиях, правильность решения задач на семинаре, правильность и своевременность решения задач в текущих домашних заданиях¹ и прочих заданиях, которые выдаются на семинарских занятиях. Оценки за работу на практических занятиях преподаватель выставляет в рабочую ведомость. Накопленная оценка по 10-ти балльной шкале за работу на семинарских и практических занятиях определяется перед промежуточным или итоговым контролем – $O_{\text{итд}}$.

Накопленная оценка учитывает результаты студента по текущему контролю следующим образом:

¹ Имеются ввиду текущие домашние работы, которые не включаются в РУП, это не форма текущего контроля «Домашнее задание».



$$O_{\text{накопл.}} = k_{\kappa/p} \cdot O_{\kappa/p} + k_{\delta/3} \cdot O_{\delta/3} + k_{\text{ауд.}} \cdot O_{\text{ауд.}},$$

где $k_{\kappa/p} = 0.4$, $k_{\delta/3} = 0.4$, $k_{\text{ауд.}} = 0.2$.

Округление каждого компонента накопленной оценки производится в соответствии с правилами математики и происходит до расчета накопленной оценки. Округление накопленной оценки также производится в соответствии с правилами математики.

Результирующая оценка за дисциплину рассчитывается следующим образом:

$$O_{\text{результ}} = k_{\text{накопл.}} \cdot O_{\text{накопл.}} + k_{\text{зачет.}} \cdot O_{\text{зачет.}},$$

где $k_{\text{накопл.}} = 0.5$, $k_{\text{зачет.}} = 0.5$.

Округление результирующей оценки также производится в соответствии с правилами математики.

В случае если $O_{\delta/3} < 4$, до зачета студент может переписать домашнее задание в соответствии с полученными замечаниями. Тогда выполнение домашнего задания в соответствии с требованиями оценивается в 4 балла.

В диплом выставляется результирующая оценка по учебной дисциплине.

10 Учебно-методическое и информационное обеспечение дисциплины

10.1 Базовые учебники

1. Long, J. Scott. 1997. Regression Models for Categorical and Limited Dependent Variables. Thousand Oaks, CA: Sage. (Полочный индекс в библиотеке НИУ ВШЭ – 519 L87).
2. Long, J. Scott and Jeremy Freese. 2005. Regression Models for Categorical Dependent Variables Using Stata. 2nd Edition. College Station, TX: Stata Press. (Полочный индекс в библиотеке НИУ ВШЭ – 519 L87).

10.2 Дополнительная литература

3. Айвазян С.А. Методы эконометрики. Учебник. – М.: Инфра-М, 2010. 512 с.
4. Айвазян С. А., Мхитарян В. С. Теория вероятностей и прикладная статистика, в 2 ч. М.: Юнити, 2001.
5. Вербик М. Путеводитель по современной эконометрике. Пер. с англ. В. А. Банникова. Под науч. ред. и предисл. С.А. Айвазяна. – М.: Научная книга, 2008. - 616 с.
6. Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS. М.: ГУ–ВШЭ, 2006.
7. Шандор З. Мультиномиальные модели дискретного выбора // Квантиль, №7, с. 9-19.
8. Agresti and I. Liu. (2005). The analysis of ordered categorical data: An overview and a survey of recent developments, invited discussion paper for the Spanish Statistical Journal, TEST.
9. Agresti, A. (1996) An Introduction to Categorical Data Analysis. New York: John Wiley & Sons, Inc.
10. Allison, P. D. 1999. Comparing logit and probit coefficients across groups. Sociological Methods and Research 28: 186–208.
11. Bo Hu, Jun Shao and Mari Palta (2006). Pseudo-R² in Logistic Regression Model, *Statistica Sinica* 16:847-860. (<http://www3.stat.sinica.edu.tw/statistica/oldpdf/A16n39.pdf>).
12. Cook, R. D. and Weisberg, S. (1999). Applied Regression Including Computing and Graphics. New York: Wiley.
13. Fu, V. 1998. sg88: Estimating generalized ordered logit models. Stata Technical Bulletin 44: 27–30.
14. Gujarati, D.N. Basic econometrics. New York McGraw-Hill, 2003.
15. Hosmer, D. and Lemeshow, S. (2000) Applied Logistic Regression (Second Edition). New York: John Wiley & Sons, Inc..



16. Jann, B. 2005. Making regression tables from stored estimates. *Stata Journal* 5: 288–308.
17. Menard, S. (1995). *Applied logistic regression analysis*. Thousand Oaks, CA: Sage.
18. Peng, C.-Y. J., & So, T.-S. H. (2002). Logistic regression analysis and reporting: A primer. *Understanding Statistics*, 1(1), 31-70. (<http://www.indiana.edu/~jopeng51/teaching-logistic.pdf>).
19. Peng, C.-Y. J., Lee, K. L., & Ingersoll, G. M. (2002). An introduction to logistic regression analysis and reporting. *The Journal of Educational Research*, 96(1):3-14. (<http://sta559s11.pbworks.com/w/file/fetch/37766848/IntroLogisticRegressionPengEducResearch.pdf>).
20. Williams, R. 2006. Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata Journal* 6: 58–82.
21. Beyond Binary: Ordinal Logistic Regression in Stata. UCLA: Academic Technology Services, Statistical Consulting Group. (http://www.ats.ucla.edu/stat/stata/seminars/stata_BeyondBinaryLogistic/Stata_ordinal_seminar.htm).
22. Logistic Regression with Stata. UCLA: Academic Technology Services, Statistical Consulting Group. (http://www.ats.ucla.edu/stat/stata/seminars/stata_logistic/default.htm).
23. Stata Data Analysis Examples. Multinomial Logistic Regression. UCLA: Academic Technology Services, Statistical Consulting Group. (<http://www.ats.ucla.edu/stat/stata/dae/mlogit.htm>).
24. Stata Topics. Logistic (and Categorical) Regression UCLA: Academic Technology Services, Statistical Consulting Group. (http://www.ats.ucla.edu/stat/stata/topics/logistic_regression.htm).

Используются электронные справочники, издания и базы статистических данных Федеральной службы государственной статистики – www.gks.ru, в том числе:

- Ежегодный бюллетень «Численность населения Российской Федерации по городам, поселкам городского типа и районам» на 1 января текущего года;
- Материалы всероссийской переписи населения 2010 г.
- Центральная База Статистических Данных.

10.3 Программные средства

Для успешного освоения дисциплины, студент использует пакет для анализа данных Stata.

10.4 Дистанционная поддержка дисциплины

При выполнении домашних работ студентам рекомендуется пользоваться материалами

- Единого архива экономических и социологических данных НИУ ВШЭ (<http://sophist.hse.ru/>),
- Проекта «Разработка учебно-методических материалов для преподавания курсов по применению количественного инструментария к решению социально-экономических задач» (<http://www.hse.ru/jesda/mathbase/>).

11 Материально-техническое обеспечение дисциплины

Пакет для анализа данных Stata. Возможно самостоятельное использование других программ: Excel, SPSS, R.