



## Правительство Российской Федерации

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

Общеуниверситетская кафедра высшей математики

### Программа дисциплины

## Анализ категориальных данных в пакете Stata

для направления 030200.62 «Политология» подготовки бакалавра

Автор программы:

Камалова Рита Ульфатовна, преподаватель ([rkamalova@hse.ru](mailto:rkamalova@hse.ru))

Гаспарян Ольга Тиграновна, преподаватель ([ogasparyan@hse.ru](mailto:ogasparyan@hse.ru))

Одобрена на заседании кафедры высшей математики «\_\_» \_\_\_\_\_ 2013 г.  
Зав. кафедрой А.А. Макаров

Рекомендована секцией УМС «\_\_\_\_\_» «\_\_» \_\_\_\_\_ 2013 г.  
Председатель \_\_\_\_\_

Утверждена УС факультета прикладной политологии «\_\_» \_\_\_\_\_ 2013 г.  
Ученый секретарь \_\_\_\_\_

Москва, 2013

*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 2013 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.

В результате освоения дисциплины студент осваивает следующие компетенции:

<i>Компетенция</i>	<i>Код по ФГОС/ НИУ</i>	<i>Дескрипторы — основные признаки освоения (показатели достижения результата)</i>	<i>Формы и методы обучения, способствующие формированию и развитию компетенции</i>
Умение логически верно, аргументировано и ясно строить устную и письменную речь	ОНК2	Демонстрирует способность написания профессиональных текстов в виде эссе/ статьи	Семинары: обсуждение статей ведущих зарубежных реферируемых журналов. Зачет: самостоятельная подготовка эссе/ статьи в соответствии со стандартами ведущих зарубежных реферируемых журналов.
Использование основных положений и методов социальных, гуманитарных и экономических наук при решении профессиональных задач, способность анализировать социальнозна-	ОНК4	Демонстрирует способность отбирать положения политической науки, а также смежных наук, позволяющие интерпретировать результаты анализа данных	Семинары: обсуждение статей ведущих зарубежных реферируемых журналов. Зачет: самостоятельная подготовка эссе/ статьи в соответствии со стандартами ведущих зарубежных реферируемых журналов.



чимые проблемы и процессы			
Способность применять математические методы анализа к социальным, политическим и экономическим явлениям	ОНК7	Оценивает и интерпретирует регрессионные модели, графики, корректно представляет результаты	Семинары: реализация методов эконометрики «руками», а также на компьютерах; обсуждение статей ведущих зарубежных реферируемых журналов.
Способность работать с информацией в глобальных компьютерных сетях	ИК2	Демонстрирует способность к нахождению статистической информации в сети Интернет	Зачет: самостоятельная подготовка эссе/ статьи в соответствии со стандартами ведущих зарубежных реферируемых журналов.
Владение иностранным языком на уровне, достаточном для разговорного общения, а также для поиска и анализа иностранных источников информации	ИК3	Демонстрирует способность к чтению профессиональной литературы на английском языке.	Семинары: обсуждение статей ведущих зарубежных реферируемых журналов. Зачет: самостоятельная подготовка эссе/ статьи в соответствии со стандартами ведущих зарубежных реферируемых журналов.

#### 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	70	14	14	42



2	Модели множественного упорядоченного выбора	42	8	8	26
3	Модели множественного неупорядоченного выбора	50	10	10	30
	<b>Итого</b>	<b>108</b>	<b>32</b>	<b>32</b>	<b>98</b>

## 6 Формы контроля знаний студентов

Тип контроля	Форма контроля	1 год				Параметры
		1	2	3	4	
Текущий	Контрольная работа	*				Письменная работа, 80 минут.
Итоговый	Зачет		*			Самостоятельная письменная работа по итогам анализа массива политологических и социально-экономических данных с использованием изученных методов. Объем: 15-25 тыс. символов с пробелами.

### 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-R<sup>2</sup>, AIC, BIC). Анализ остатков модели. Выявление нетипичных наблюдений.

### Основная литература

1. Long, J. Scott. 1997. *Regression Models for Categorical and Limited Dependent Variables*. Thousand Oaks, CA: Sage. Ch. 2, 3, 4.
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
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](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/etch/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-R<sup>2</sup> 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](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](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.
- Обобщенная модель множественного упорядоченного выбора (Generalized ordered logit) / частичная модель пропорциональных шансов (Partial proportional odds model).

### **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](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](http://www.ats.ucla.edu/stat/stata/seminars/stata_BeyondBinaryLogistic/Stata_ordinal_seminar.htm)).
3. Peterson, Bercedis and Frank E. Harrell. (1990). Partial Proportional Odds Models for Ordinal Response Variables, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 39, No. 2, pp. 205-217.



4. Fu, V. 1998. sg88: Estimating generalized ordered logit models. *Stata Technical Bulletin* 44: 27–30.

### Раздел 3. Модели множественного неупорядоченного выбора

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

#### **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](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. McFadden, D., 'Conditional Logit Analysis of Qualitative Choice Behaviour', in Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, 1973. (<http://elsa.berkeley.edu/reprints/mcfadden/zarembka.pdf>)
3. R Data Analysis Examples. Multinomial Logistic Regression. UCLA: Academic Technology Services, Statistical Consulting Group. (<http://www.ats.ucla.edu/stat/r/dae/mlogit.htm>).
4. Stata Data Analysis Examples. Multinomial Logistic Regression. UCLA: Academic Technology Services, Statistical Consulting Group. (<http://www.ats.ucla.edu/stat/stata/dae/mlogit.htm>).
5. 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](http://www.ats.ucla.edu/stat/stata/library/odds_ratio_mlogit.htm)).
6. Steckel J, Vanhonacker W. 1988. A heterogeneous conditional logit model of choice. *Journal of Business and Economic Statistics* 6: 391-398.
7. Thurner P. (2000) The empirical application of the spatial theory of voting in multiparty systems with random utility models. *Electoral Studies*, No. 19, pp. 493–517.

## **8 Образовательные технологии**

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

## **9 Оценочные средства для текущего контроля и аттестации студента**

### **9.1 Тематика заданий текущего контроля**

#### *Домашняя работа (пример)*

На предложенном массиве *LEVADA\_CUR11\_17.dta* постройте регрессию, моделирующую голосование в выборах 4 декабря 2011 года.

Зависимая переменная q29 – ПРИНИМАЛИ ЛИ ВЫ УЧАСТИЕ В ВЫБОРАХ В ГОСУДАРСТВЕННУЮ ДУМУ РОССИИ 4 ДЕКАБРЯ ЭТОГО ГОДА?

Объясняющие переменные. Непрерывные:



- qS2 – Сколько лет вам исполнилось?
- qD9 – Каков был общий доход вашей семьи за прошлый месяц из всех источников, включая доходы от основной и дополнительной работы, пенсии, стипендии, другие денежные доходы, как выплаченные, так и невыплаченные, которые получили все члены вашей семьи в прошлом месяце?

Объясняющие переменные. Категориальные:

- qS1 – Пол респондента
- qS3 – Образование
- D14\_1 – Пользуетесь ли вы интернетом, и если да, то для каких целей? Для того, чтобы следить за последними новостями.
- q4B – Вы в целом одобряете или не одобряете деятельность Владимира Путина на посту председателя правительства России?

Используйте как минимум одну непрерывную и одну категориальную переменную в качестве объясняющей.

1. Сохраните в массиве только те переменные, которые Вам нужны. Удалите наблюдения с пропусками. Приведите описательные статистики переменных (*по каждой переменной должно быть одинаковое число наблюдений!*). Следите за тем, чтобы наблюдений в массиве было больше 100, а лучше 500.
2. При необходимости перекодируйте переменные или превратите категориальную в набор бинарных переменных. Перекодируйте зависимую в 0 и 1!
3. Постройте модель регрессии, приведите оценки коэффициентов  $\beta$  при всех переменных, отношение шансов (factor change coefficients --- odds ratio), а также стандартизованные коэффициенты (выберите, какие Вам больше нравятся) при переменных, для которых это релевантно.
4. Проверьте значимость коэффициентов z-критерием (сформулировав  $H_0$  и  $H_1$ , приведя значения статистики критерия и p-value).
5. Проинтерпретируйте полученные оценки коэффициентов  $\beta$  при одной непрерывной и одной категориальной переменной, в том числе стандартизованные коэффициенты.
6. Рассчитайте предельные эффекты (marginal effect) или другие соответствующие изменения в вероятностях принадлежать группе голосовавших в выборах  $y = 1$  и проинтерпретируйте их.
7. Рассчитайте предсказанные вероятности для двух индивидов с заданными Вами характеристиками; рассчитайте разность вероятностей, проверьте статистическую значимость и проинтерпретируйте ее.
8. Постройте график, в котором по оси OY отложены предсказанные вероятности  $P(y=1|X)$ , а по оси OX --- объясняющая переменная. Проинтерпретируйте график.
9. Сравните результаты (коэффициенты и значимость) для моделей логит- и пробит- на противоречия.
10. Проверьте гипотезу о линейных ограничениях критерием Вальда (сформулировав  $H_0$  и  $H_1$ , приведя значения статистики критерия и p-value).
11. Проверьте значимость коэффициентов критерием отношения правдоподобия, т.е. сравните Вашу модель и модель на константу, т.е. без объясняющих переменных (сформулировав  $H_0$  и  $H_1$ , приведя значения статистики критерия и p-value). Сделайте вывод.
12. Предложите модель, вложенную в Вашу. Проверьте гипотезу об ограничениях вида  $\beta_k = \beta_p = 0$  критерием отношения правдоподобия (LR test) (сформулировав  $H_0$  и  $H_1$ , приведя значения статистики критерия и p-value).
13. Сравните Вашу модель и вложенную с помощью двух информационных критериев. Сделайте выбор в пользу одной из них.



14. Классифицируйте наблюдения на 0 и 1, самостоятельно выбрав порог отсечения. Представьте результаты классификации в таблице.
15. Сколько наблюдений было верно классифицировано? Какой % от общего числа они составляют? Рассчитайте чувствительность и специфичность.
16. Постройте ROC-кривую и график «чувствительность VERSUS (1 --- специфичность)». Какова площадь под ROC-кривой? Что Вы можете сказать о качестве Вашей модели?
17. Проверьте массив на атипичные наблюдения. При их обнаружении, удалите и вновь оцените модель регрессии. Произошли ли значимые изменения в оценках коэффициентов  $\beta$ ? (Возможно, Вам понадобится создать идентификатор наблюдений. Это можно сделать командой `gen n=(\_n)`).

Результаты необходимо представить в виде тезисов --- ответов на вопросы --- и сопроводить выдачами из пакета Stata там, где это необходимо.

Сдается в электронном виде до 7 марта, 23:59 на адрес [rkamalova@hse.ru](mailto:rkamalova@hse.ru). К работе также необходимо приложить log-файл и do-файл.

## 9.2 Вопросы для оценки качества освоения дисциплины

### Проверочная работа (пример)

1. Покажите, что  $P(y_i = 1|X) = F(\beta_0 + \beta_1 x_i)$ , приняв, что  $y_i = 1$ , если латентный  $y_i^* > 0$ , и  $y_i = 0$  в противном случае.
2. Вы оценили модель логистической регрессии на данных GSS (*General Social Survey*) по США

$$P(y_i = 1|X_i) = 2.5 - 0.11education_i + 0.53male_i + 0.70bush92_i - 0.07children_i$$

```
. logit cappun educ male bush92 chldrn, nolog
```

```
Logistic regression                               Number of obs   =       5570
                                                    LR chi2(4)      =       248.17
                                                    Prob > chi2     =       0.0000
Log likelihood = -2831.6609                       Pseudo R2      =       0.0420
```

	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
cappun						
educ	-.1133811	.0127096	-8.92	0.000	-.1382916	-.0884707
male	.5295012	.0667557	7.93	0.000	.3986624	.66034
bush92	.7005197	.0736915	9.51	0.000	.556087	.8449524
chldrn	-.0748111	.0221608	-3.38	0.001	-.1182455	-.0313767
_cons	2.500112	.1964794	12.72	0.000	2.115019	2.885204

Зависимая переменная --- отношение к смертной казни (0 --- отрицательное), *education* --- количество лет обучения, *male* --- мужчина (дамми), *bush92* --- голосовал за Буша в 1992 г. (дамми), *children* --- число детей у респондента. Проинтерпретируйте коэффициенты при количестве лет обучения и половой принадлежности в терминах отношения шансов.

3. Во сколько раз шансы мужчины без детей на то, чтобы поддерживать смертную казнь, больше/меньше, чем у женщины с 2 детьми (при прочих равных (зафиксированных на среднем значении)).
4. Какими критериями можно проверить значимость коэффициентов при предикторах в модели бинарного выбора:





1. Критерий Фишера
  2. Критерий Вальда
  3. Критерий отношения правдоподобия
  4. Критерий Хаусмана
5. Рассчитайте для модели поддержки смертной казни AIC и BIC.
6. У альтернативной модели без переменной *bush92* AIC = 8174.8, BIC = 8202.7. Какую Вы предпочтете? Обоснуйте свой выбор.
7. Рассчитайте скорректированный  $R^2_{count}$ , специфичность и чувствительность модели.

```
. estat classification, cutoff(.6)
```

```
Logistic model for cappun
```

Classified	True		Total
	D	~D	
+	4224	1163	5387
-	104	79	183
Total	4328	1242	5570

8. Нарисуйте ROC-кривую. Подпишите оси.

## 10 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 заданиях 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. Накопленная оценка учитывает результаты студента по текущему контролю следующим образом:

$$O_{\text{накопл.}} = k_{\text{к/р}} \cdot O_{\text{к/р}} + k_{\text{д/з}} \cdot O_{\text{д/з}} + k_{\text{ауд}} \cdot O_{\text{ауд}},$$

где  $k_{\text{к/р}} = 0.3$ ,  $k_{\text{д/з}} = 0.4$ ,  $k_{\text{ауд}} = 0.3$ .

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:

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

где  $k_{\text{накопл.}} = 0.6$ ,  $k_{\text{зачет}} = 0.4$ .

Округление результирующей оценки также производится в соответствии с правилами математики.

Если в зачетной работе допущены грубые ошибки (например, интерпретация регрессионного коэффициента без учета нелинейности модели регрессии. Примерный список таких ошибок будет сообщен студентам), то работа оценивается не выше 4 баллов.



В случае  $O_{\partial/3} < 4$ , 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.

## 11 Учебно-методическое и информационное обеспечение дисциплины

### 11.1 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).

### 11.2 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, A. (1996) An Introduction to Categorical Data Analysis. New York: John Wiley & Sons, Inc.
9. Agresti, A. 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.
10. Allison, P. D. 1999. Comparing logit and probit coefficients across groups. Sociological Methods and Research 28: 186–208.
11. 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](http://www.ats.ucla.edu/stat/stata/seminars/stata_BeyondBinaryLogistic/Stata_ordinal_seminar.htm)).
12. Bo Hu, Jun Shao and Mari Palta (2006). Pseudo- $R^2$  in Logistic Regression Model, *Statistica Sinica* 16:847-860. (<http://www3.stat.sinica.edu.tw/statistica/oldpdf/A16n39.pdf>).
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. Logistic Regression with Stata. UCLA: Academic Technology Services, Statistical Consulting Group. ([http://www.ats.ucla.edu/stat/stata/seminars/stata\\_logistic/default.htm](http://www.ats.ucla.edu/stat/stata/seminars/stata_logistic/default.htm)).
18. McFadden, D., 'Conditional Logit Analysis of Qualitative Choice Behaviour', in Zarembka (ed.), *Frontiers in Econometrics*, Academic Press, 1973. (<http://elsa.berkeley.edu/reprints/mcfadden/zarembka.pdf> )
19. Menard, S. (1995). Applied logistic regression analysis. Thousand Oaks, CA: Sage.



20. 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>).
21. 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/etch/37766848/IntroLogisticRegressionPengEducReseach.pdf>).
22. Peterson, Bercedis and Frank E. Harrell. (1990). Partial Proportional Odds Models for Ordinal Response Variables, *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 39, No. 2, pp. 205-217.
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](http://www.ats.ucla.edu/stat/stata/topics/logistic_regression.htm)).
25. Steckel J, Vanhonacker W. 1988. A heterogeneous conditional logit model of choice. *Journal of Business and Economic Statistics* 6: 391-398.
26. Thurner P. (2000) The empirical application of the spatial theory of voting in multiparty systems with random utility models. *Electoral Studies*, No. 19, pp. 493–517
27. Williams, R. 2006. Generalized ordered logit/partial proportional odds models for ordinal dependent variables. *Stata Journal* 6: 58–82.

### 11.3 Программные средства

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

### 11.4 Дистанционная поддержка дисциплины

При выполнении домашних работ студентам рекомендуется пользоваться материалами

- Единого архива экономических и социологических данных НИУ ВШЭ (<http://sophist.hse.ru/>) (преимущественно результаты опросов населения Аналитического Центра Юрия Левады, Фонда Общественное Мнение, Всероссийского центра изучения общественного мнения),
- Межуниверситетского консорциума по политическим и социальным исследованиям (ICPSR) (<http://www.icpsr.umich.edu/icpsrweb/ICPSR/index.jsp>),
- Проекта «Разработка учебно-методических материалов для преподавания курсов по применению количественного инструментария к решению социально-экономических задач» (<http://www.hse.ru/jesda/mathbase/>).

## 12 Материально-техническое обеспечение дисциплины

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