



Правительство Российской Федерации

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

Факультет социальных наук
Департамент политической науки

Программа дисциплины

Анализ политологических 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 «__» _____ 2014 г.
Зав. кафедрой А.А. Макаров

Рекомендована секцией УМС «_____» «__» _____ 2014 г.
Председатель _____

Утверждена УС факультета прикладной политологии «__» _____ 2014 г.
Ученый секретарь _____

Москва, 2014



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 политологических данных в пакете Stata».

Программа разработана в соответствии с рабочим учебным планом университета по направлению 030200.62 «Политология» подготовки бакалавра, утвержденным в 2014 г.

2 Цели освоения дисциплины

Целью освоения дисциплины «Анализ политологических данных в пакете Stata» является выработка компетенций по анализу данных политологических и социально-экономических данных (преимущественно методами регрессионного анализа) в пакете Stata.

3 Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате освоения дисциплины студент должен:

- Знать базовые методы анализа данных, в том числе регрессии, позволяющие решать некоторые задачи социальных наук;
- Иметь навыки их реализации в пакете Stata;
- Уметь интерпретировать и представлять полученные результаты;
- Знать ряд известных источников и баз политологических и социально-экономических данных.

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	ОНК-2	Демонстрирует способность написания профессиональных текстов в виде эссе/ статьи	Семинары: обсуждение статей ведущих зарубежных реферируемых журналов. Домашнее задание: 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	ОНК-4	Демонстрирует способность отбирать положения политической науки, а также смежных наук, позволяющие интерпретировать результаты анализа	Семинары: обсуждение статей ведущих зарубежных реферируемых журналов. Домашнее задание: самостоятельная подготовка эссе/ статьи в соответ-



анализировать социальнозначимые проблемы и процессы		данных	ствии со стандартами ведущих зарубежных реферируемых журналов.
Способность применять математические методы анализа к социальным, политическим и экономическим явлениям	ОНК-7	Оценивает и интерпретирует регрессионные модели, графики, корректно представляет результаты	Семинары: реализация методов математической статистики и эконометрики «руками», а также на компьютерах, а также на компьютерах; обсуждение статей ведущих зарубежных реферируемых журналов.
Способность работать с информацией в глобальных компьютерных сетях	ИК-2	Демонстрирует способность к нахождению статистической информации в сети Интернет	Домашнее задание: 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	ИК-3	Демонстрирует способность к чтению профессиональной литературы на английском языке.	Семинары: обсуждение статей ведущих зарубежных реферируемых журналов. Домашнее задание: самостоятельная подготовка эссе/ статьи в соответствии со стандартами ведущих зарубежных реферируемых журналов
Освоение основных теоретикометодологических подходов в сфере политического анализа и прогнозирования; владение методами сбора и обработки политической информации; методологией и методиками политического анализа и прогнозирования	ПК-9	Способен сформулировать основные допущения используемого метода анализа данных. Способен последовательно применить к одним и тем же данным разные методы анализа данных для проверки устойчивости результата к смене допущений и верно интерпретирует получаемый результат как в случае совпадения результатов, полученных разными методами, так и в случае их расхождения.	Все виды аудиторных занятий и домашние работы



4 Место дисциплины в структуре образовательной программы

Настоящая дисциплина относится к факультативам направления 030200.62 «Политология» подготовки бакалавра.

Изучение данной дисциплины базируется на следующих дисциплинах:

- Алгебра и анализ (1 курс),
- Теория вероятностей и математическая статистика (1 курс),
- Вероятностно-статистические модели в политологии (2 курс)

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

- Многомерный статистический анализ в политологии,
- Сравнительная политика,
- Политический анализ,
- Политическое поведение,
- Анализ электоральных данных,
- Прогнозирование социально-политических процессов

5 Тематический план учебной дисциплины

№	Название раздела	Всего часов	Аудиторные часы		Самостоятельная работа
			Лекции	Практические занятия	
1	Обзор источников политологических и социально-экономических данных и средств их анализа. Пакет Stata.	4	2	0	2
2	Описательные статистики, графики и рисунки в Stata.	10	2	2	6
3	Связь качественных признаков в политологии. Связь признаков в количественных и порядковых шкалах	8	2	2	4
4	Регрессионный анализ. Модель парной линейной регрессии. Разложение вариации.	34	8	8	18
5	Регрессионный анализ. Модель множественной линейной регрессии.	20	4	4	12
6	Мультиколлинеарность.	16	4	4	8
7	Гетероскедастичность.	16	4	4	8
8	Влиятельные наблюдения	16	4	4	8
9	Критерии качества регрессионных моделей. Выбор модели	16	4	4	8
10	Эффекты взаимодействия переменных и фиктивные переменные	26	6	8	12
	Итого	108	30	30	48

6 Формы контроля знаний студентов

Тип контроля	Форма контроля	1 год				Параметры
		1	2	3	4	



Текущий	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: 8-15 тыс. символов с пробелами.
Итоговый	Зачет			*	Письменный зачет, 160 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.

Письменный зачет состоит из двух частей: А (обязательная) и В (основная). В части А будут представлены задачи, проверяющие базовые навыки и элементарные компетенции. В части В будут представлены задачи, предполагающие более творческий подход к решению, проверку способностей переводить содержательные задачи на формальный язык и применять к их решению методы эконометрики. Задачи части В не проверяются, если не набран заданный минимум баллов из части А.

6.2 Порядок формирования оценок по дисциплине

Итоговая (результатирующая) оценка по дисциплине формируется из накопленной оценки и оценки за экзамен.

Накопленная оценка ($O_{накопл}$) учитывает результаты студента по текущему контролю следующим образом:

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

где $k_{к/р} = 0.2$, $k_{д/з} = 0.4$, $k_{ауд} = 0.4$.

Оценка за аудиторную работу ($O_{ауд}$) формируется из оценок за текущие домашние работы¹, проверочные работы и семинарскую активность. Преподаватель оценивает активность студента в дискуссиях, правильность решения задач на семинаре, правильность и своевременность решения задач в текущих домашних заданиях и прочих заданиях, которые выдаются на семинарских занятиях.

Если в домашней работе ($O_{д/з}$) допущены грубые ошибки (например, использование зависимой переменной, измеренной на качественном уровне; примерный список таких ошибок будет сообщен студентам), то работа оценивается не выше 4 баллов.

Накопленная оценка по 10-ти балльной шкале за работу на семинарских и практических занятиях определяется перед промежуточным или итоговым контролем.

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.

Результатирующая оценка за дисциплину рассчитывается следующим образом:

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



$$O_{результ} = k_{накопл} \cdot O_{накопл} + k_{экзамен} \cdot O_{экзамен},$$

где $k_{накопл} = 0.6$, $k_{экзамен} = 0.4$.

Округление результирующей оценки также производится в соответствии с правилами математики.

На пересдаче студенту не предоставляется возможность получить дополнительный балл для компенсации оценки за текущий контроль. Накопленная оценка не обнуляется.

В диплом выставляется результирующая оценка по учебной дисциплине.

7 Содержание дисциплины

Раздел 1. Обзор источников политологических и социально-экономических данных и средств их анализа. Пакет Stata.

Основная литература

1. Вакуленко Е.С., Камалова Р.У., Кисельгоф С.Г., Смирнова Ж.И., Стукал Д.К., Хавенсон Т.Е. Разработка учебно-методических материалов для преподавания курсов по применению количественного инструментария к решению социально-экономических задач. // Экономическая социология (электронный журнал), 2010. Т. 11. № 4. С. 190—194. (http://ecsoc.hse.ru/data/2010/09/29/1234590525/ecsoc_t11_n4.pdf#page=190)
2. Учебно-методические материалы для преподавания курсов по применению количественного инструментария к решению социально-экономических задач (<http://www.hse.ru/jesda/mathbase/>).
3. Long, J. Scott and Jeremy Freese. 2005. *Regression Models for Categorical Dependent Variables Using Stata*. 2nd Edition. College Station, TX: Stata Press. Ch. 1, 2.
4. A brief description of Stata. (http://www.stata-press.com/manuals/stata10/guide_brief.pdf).
5. Basics of Stata. (2008). (<http://www.ics.uci.edu/~dvd/Courses/Stata/stata.pdf>).

Дополнительная литература

6. Единый архив социологических данных (рук. — Л. Б. Косова) // Экономическая социология (электронный журнал), 2010, т. 11, № 1, с. 133–140. (http://www.hse.ru/data/2011/12/08/1208204985/ecsoc_t11_n1.pdf#page=133)
7. Long, J. Scott. 2009. *The workflow of data analysis using Stata*. TX: Stata Press. (Полочный индекс в библиотеке НИУ ВШЭ – 004 L87).
8. Resources to help you learn and use Stata. UCLA: Academic Technology Services, Statistical Consulting Group. (<http://www.ats.ucla.edu/stat/stata/>).
9. Rodríguez, Germán. (<http://data.princeton.edu/stata/>).

Раздел 2. Описательные статистики в Stata. Графики и рисунки в Stata.

Основная литература

1. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере: учебное пособие. – М.: ИД «ФОРУМ», 2008. – С. 24 – 49, 131 – 138.
2. Descriptive information and statistics. UCLA: Academic Technology Services, Statistical Consulting Group. (<http://www.ats.ucla.edu/stat/stata/modules/descript.htm>).
3. Graphics. UCLA: Academic Technology Services, Statistical Consulting Group. (<http://www.ats.ucla.edu/stat/stata/topics/graphics.htm>).
4. Long, J. Scott and Jeremy Freese. 2005. *Regression Models for Categorical Dependent Variables Using Stata*. 2nd Edition. College Station, TX: Stata Press. Ch. 1, 2.

Дополнительная литература



5. Long, J. Scott. 2009. The workflow of data analysis using Stata. TX: Stata Press. (Полочный индекс в библиотеке НИУ ВШЭ – 004 L87).
6. Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS : учеб. пособие для вузов. – М.: Изд. дом ГУ ВШЭ, 2006. – С. 10-39.
7. Тюрин Ю.Н., Макаров А.А., Высоцкий И.Р., Яценко И.В. Теория вероятностей и статистика. — М.: МЦНМО, 2004. – С. 27-44, 44-63, 189-205.
8. Data Management. UCLA: Academic Technology Services, Statistical Consulting Group. (http://www.ats.ucla.edu/stat/stata/topics/data_management.htm).

Раздел 3. Связь качественных признаков в политологии. Связь признаков в количественных и порядковых шкалах.

Основная литература

1. Доугерти К. Введение в эконометрику: Учебник. М.: ИНФРА-М, 1999. Гл. 1. С. 34–52.
2. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере: учебное пособие. – М.: ИД «ФОРУМ», 2008. – С. 244-263.
3. Stata Annotated Output. Correlation. UCLA: Academic Technology Services, Statistical Consulting Group. (http://www.ats.ucla.edu/stat/stata/output/stata_corr_output.htm).

Дополнительная литература

1. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. Т.2: Теория вероятностей и прикладная статистика. — М.: Юнити, 2001. Гл. 11. с. 396-457.
2. Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS : учеб. пособие для вузов. — М.: Изд. дом ГУ ВШЭ, 2006. – С. 47-81.

Раздел 4. Регрессионный анализ. Модель парной линейной регрессии. Парная линейная регрессия как линейная аппроксимация условного математического ожидания. Оценивание коэффициентов парной линейной регрессии: метод наименьших квадратов. Интерпретация коэффициентов регрессии при непрерывных переменных.

Основная литература

1. Доугерти К. Введение в эконометрику: Учебник. М.: ИНФРА-М, 1999. Гл. 2. С. 53–72.
2. Stock J, Watson M. Introduction to Econometrics (2rd edition). Addison Wesley Longman; 2008. pp. 109-143, 143-147, 149-158.
3. Regression Analysis. UCLA: Academic Technology Services, Statistical Consulting Group. (<http://www.ats.ucla.edu/stat/stata/topics/regression.htm>).
4. Stata Annotated Output. Regression Analysis. UCLA: Academic Technology Services, Statistical Consulting Group. (http://www.ats.ucla.edu/stat/stata/output/reg_output.htm).

Дополнительная литература

5. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. Т.2: Основы эконометрики. - М.: ЮНИТИ, 2001. – 656 с.
6. Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS : учеб. пособие для вузов. – М.: Изд. дом ГУ ВШЭ, 2006. – С. 115-146.
7. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере: учебное пособие. – М.: ИД «ФОРУМ», 2008. – С. 208-219.
8. Gujarati, D.N. Basic econometrics. New York McGraw-Hill, 2003. Ch. 1, 2.
9. Kritzer, H.M., 1996. The Data Puzzle: The Nature of Interpretation in Quantitative Research, *American Journal of Political Science*, 40(1): 1-32.

Раздел 5. Модель множественной линейной регрессии. Теорема Гаусса-Маркова и теорема Рао о свойствах МНК-оценок. Статистический вывод в регрессии: статистическая значи-



мость коэффициентов. Критерии качества моделей, сравнение моделей. Разложение вариации.

Основная литература

1. Доугерти К. Введение в эконометрику: Учебник. М.: ИНФРА-М, 1999. Гл. 3. С. 73–114, Гл. 5. С. 134–140, 146–155, Обзор. С. 15-20, 26-28.
2. Stock J, Watson M. Introduction to Econometrics (2nd edition). Addison Wesley Longman; 2008. pp.166-169, 171-185, 186-206, 211-219, 220-239.

Дополнительная литература

3. Доугерти К. Введение в эконометрику: Учебник. М.: ИНФРА-М, 1999. Гл. 6. 165-199.
4. Крышгановский А.О. Анализ социологических данных с помощью пакета SPSS : учеб. пособие для вузов. – М.: Изд. дом ГУ ВШЭ, 2006. – С. 146-166.
5. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере: учебное пособие. – М.: ИД «ФОРУМ», 2008. – С. 208-219.
6. Gujarati, D.N. Basic econometrics. New York McGraw-Hill, 2003. Ch. 3, 4.
7. Lewis-Beck, Michael S. and Andrew Skalaban. (1990). “When to Use R-Squared”. *The Political Methodologist* 3(2): 11–12.
8. King G. (1990). “When Not to Use R-Squared.” *The Political Methodologist*, 3(2): 9–11.
9. Goertzel, T., 2002. “Myths of Murder and Multiple Regression”, *The Sceptical Inquirer* 26, 1: 19-23 (<http://crab.rutgers.edu/%7Egoertzel/mythsofmurder.htm>).
10. Larocca, Roger. 2005. Reconciling Conflicting Gauss-Markov Conditions, *Political Analysis* 13:188–207.
11. Stock J, Watson M. Introduction to Econometrics (2nd edition). Addison Wesley Longman; 2008. pp. 704-727, 743-752.

Раздел 6. Мультиколлинеарность. Источники мультиколлинеарности. Последствия мультиколлинеарности для статистического вывода. Способы диагностики. Меры борьбы с мультиколлинеарностью. Автокорреляция. Источники и последствия автокорреляции для статистического вывода.

Основная литература

1. Доугерти К. Введение в эконометрику: Учебник. М.: ИНФРА-М, 1999. С. Гл. 5. 155–164, Гл. 7. 217–234.
2. Stock J, Watson M. Introduction to Econometrics (2nd edition). Addison Wesley Longman; 2008. pp. 206-210.

Дополнительная литература

3. Gujarati, D.N. Basic econometrics. New York McGraw-Hill, 2003. Ch. 10, 12.
4. Cook, R. D. and Weisberg, S. (1999). Applied Regression Including Computing and Graphics. New York: Wiley.

Раздел 7. Гетероскедастичность. Источники гетероскедастичности. Последствия гетероскедастичности для статистического вывода. Способы диагностики. Меры борьбы с гетероскедастичностью.

Основная литература

1. Доугерти К. Введение в эконометрику: Учебник. М.: ИНФРА-М, 1999. Гл. 7. С. 200–217.
2. Stock J, Watson M. Introduction to Econometrics (2nd edition). Addison Wesley Longman; 2008. pp. 160-165, 604-608.

Дополнительная литература

3. Braumoeller B. (2006). “Explaining Variance.” *Political Analysis*, Vol.14, No.3, pp. 268–290.



4. Cook, R. D. and Weisberg, S. (1999). *Applied Regression Including Computing and Graphics*. New York: Wiley.
5. Gujarati, D.N. *Basic econometrics*. New York McGraw-Hill, 2003. Ch. 11.

Раздел 8. Нетипичные и влиятельные наблюдения. Последствия и способы диагностики (межквартильный размах, коробчатая диаграмма Тьюки и статистические выбросы, мера потенциального влияния, мера Кука, графики остатков).

Основная литература

1. Айвазян С. А., Мхитарян В. С. *Прикладная статистика и основы эконометрики*. В 2 частях. — М.: Юнити, 2001. С. 735-737.
2. Fox, John. *Applied Regression Analysis and Generalized Linear Models*, 2nd Edition. Sage, 2008. Ch. 11. (будет выслана преподавателем).

Дополнительная литература

3. Cook, R. D. and Weisberg, S. (1999). *Applied Regression Including Computing and Graphics*. New York: Wiley.
4. Fox, John. *Regression Diagnostics: An Introduction*. Sage, 1991.

Раздел 9. Критерии качества регрессионных моделей. Выбор модели.

Основная литература

1. Доугерти К. *Введение в эконометрику: Учебник*. М.: ИНФРА-М, 1999. Гл. 7. С. 159-164.
2. Bartels L.M. (1990). "Five Approaches to Model Specification." *The Political Methodologist*, Vol. 3, No. 2, pp. 2–6.
3. King G. (1986). "How Not to Lie With Statistics: Avoiding Common Mistakes in Quantitative Political Science." *American Journal of Political Science*, Vol. 30, pp. 666–687.
4. King G. (1990). "When Not to Use R-Squared." *The Political Methodologist*, Vol. 3, No. 2, pp. 9–11.
5. Lewis-Beck, Michael S. and Andrew Skalaban. 1990. When to Use R-Squared. *The Political Methodologist* 3 (2): 11–12.
6. Luskin R. (1991). "R-Squared Encore." *The Political Methodologist*, Vol. 4, No. 1, pp. 21–23.
7. Stock J, Watson M. *Introduction to Econometrics* (2nd edition). Addison Wesley Longman; 2008. pp. 235-239, 312–347.

Дополнительная литература

8. Cook, R. D. and Weisberg, S. (1999). *Applied Regression Including Computing and Graphics*. New York: Wiley.
9. Fox, John. *Applied Regression Analysis and Generalized Linear Models*, 2nd Edition. Sage, 2008. Ch. 9.
10. Gujarati, D.N. *Basic econometrics*. New York McGraw-Hill, 2003. Ch. 13.
11. Stock J, Watson M. *Introduction to Econometrics* (2nd edition). Addison Wesley Longman; 2008. pp. 312-344.

Раздел 10. Эффекты взаимодействия переменных и фиктивные переменные. Фиктивные переменные как способ включения в анализ категориальных переменных. Фиктивные переменные для константы и для коэффициента наклона. Взаимодействия категориальных переменных. Взаимодействия непрерывных переменных.

Основная литература

1. Доугерти К. *Введение в эконометрику: Учебник*. М.: ИНФРА-М, 1999. Гл. 9. С. 262–285.
2. Brambor Th., Clark W., Golder M. (2006). "Understanding Interaction Models: Improving Empirical Analyses." *Political Analysis*, Vol. 14, pp. 63–82.



3. Stock J, Watson M. Introduction to Econometrics (2nd edition). Addison Wesley Longman; 2008. pp. 158-160, 277-290.

Дополнительная литература

4. Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS: учеб. пособие для вузов. – М.: Изд. дом ГУ ВШЭ, 2006. – С. 166-182.
5. Cook, R. D. and Weisberg, S. (1999). Applied Regression Including Computing and Graphics. New York: Wiley.
6. Gujarati, D.N. Basic econometrics. New York McGraw-Hill, 2003. Ch. 9.
7. Stock J, Watson M. Introduction to Econometrics (2nd edition). Addison Wesley Longman; 2008. pp. 347-378.

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 Примеры заданий промежуточного /итогового контроля

Часть А. Обязательная. Вам нужно набрать как минимум 12 баллов

1. (2 балла) Сформулируйте условия Гаусса–Маркова (в расширенном варианте т.е., пользуясь терминологией Дениса, укажите условия «ядра» и «периферии»).

2. (2 балла) Подчеркните статистически значимые оценки коэффициентов в модели регрессии процента голосов за партию «Яблоко» на социально-экономические показатели, приняв 90%-ный уровень доверия (1 балл). Проинтерпретируйте оценку коэффициента при предикторе «Явка» (1 балл). (*yavka* – Явка в %; *compinternet* – Количество компьютеров с доступом в Интернет, нормированное на численность населения; *gini* – Коэффициент Джини).

```
. regress Yabloko yavka comp_internet100 gini100
```

Source	SS	df	MS	Number of obs =	79
Model	167.343137			F(3, 75) =	
Residual		75		Prob > F =	
Total	290.300199	78	3.72179743	R-squared =	0.5764
				Adj R-squared =	0.5595
				Root MSE =	1.2804

Yabloko	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]
yavka	-.0647829	.0125096	-5.18		-.0897032 -.0398626
compinternet	.1984785	.0390788	5.08		.1206296 .2763274
gini	5.66720	.0677672	0.84		-.0783271 .1916711
_cons	1.587854	2.433101	0.65		-3.259133 6.434841

3. (3 балла) Используя выдачу Stata из задания 2, сформулируйте нулевую и альтернативную гипотезы для F-критерия (отображаемого в правой верхней части выдачи Stata) (1 балл). Рассчитайте значение F-статистики, опираясь на представленные в задании 2 данные (1 балл). Сформулируйте содержательный вывод (1 балл).



4. (3 балла) Построена регрессия индекса потребительских цен (ИПЦ) на уровень безработицы в странах мира. Среднее значение ИПЦ равно 1.5. Выборочная оценка ковариации ИПЦ и индекса безработицы равна -15 , а выборочная оценка дисперсии уровня безработицы равна 30. Найдите МНК-оценки коэффициентов регрессии и запишите полученную модель.

5. (5 баллов) У Вас есть следующие данные для трех стран мира по качеству государственного управления (x) и уровню младенческой смертности (y): ($x_i = 5$; $y_i = 2$), (3; 4), (2; 5).

С помощью метода наименьших квадратов (МНК), оцените регрессию без константы, т.е. получите МНК-оценку коэффициента β_1 из модели $y_i = \beta_1 \times x_i + e_i$.

6. (2 балла) Сравните предложенные ниже модели и сделайте обоснованный выбор в пользу одной из них, если это возможно.

```
. regress Yabloko compinternet unempl gorod rus
-----+-----
Source |      SS      df      MS                Number of obs =      83
-----+-----                F( 4,   78) =    35.54
Model | 194.671569      4 48.6678922          Prob > F      = 0.0000
Residual | 106.818466     78 1.36946751          R-squared     = 0.6457
-----+-----                Adj R-squared = 0.6275
Total | 301.490035     82 3.67670774          Root MSE     = 1.1702

-----+-----
Yabloko |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----
compinternet |   .191554   .0321178     5.96  0.000     .1276123   .2554957
unempl |   .0579075  .0266021     2.18  0.033     .0049469   .1108682
gorod |   .0584879  .0130776     4.47  0.000     .0324524   .0845235
rus |   2.052794  .7022181     2.92  0.005     .6547851   3.450803
_cons |  -6.112517  1.031041    -5.93  0.000    -8.165163  -4.059872

-----+-----
. estat ic
-----+-----
Model |      Obs   ll(null)   ll(model)   df       AIC       BIC
-----+-----
. |      83   -171.3026  -128.2419    5       266.4839  278.5781
-----+-----
```



```
. regress LDPR compinternet unempl gorod rus cpi
-----+-----
Source |      SS      df      MS                Number of obs =      83
-----+-----                F( 5, 77) =      20.52
Model | 1413.76065      5 282.752129            Prob > F      = 0.0000
Residual | 1060.81245     77  13.776785            R-squared     = 0.5713
-----+-----                Adj R-squared = 0.5435
Total | 2474.57309     82  30.177206            Root MSE     = 3.7117
```

```
-----+-----
LDPR |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----
compinternet |   .0531108   .104591     0.51   0.613    -.1551566   .2613781
unempl |   .0655002   .0845976    0.77   0.441    -.1029551   .2339555
gorod |   .1145754   .0415262    2.76   0.007    .0318861   .1972647
rus |   12.84184   2.230467    5.76   0.000    8.400414   17.28327
cpi |   -.2659963   .2559907   -1.04   0.302    -.7757389   .2437463
_cons |   22.97214   28.79047    0.80   0.427   -34.35702   80.3013
```

```
. estat ic
```

```
-----+-----
Model |      Obs   ll(null)   ll(model)      df          AIC          BIC
-----+-----
. |      83   -258.6637   -223.5118        6       459.0236       473.5367
```

(*Yabloko* — процент голосов за «Яблоко»; *LDPR* — процент голосов за ЛДПР; *compinternet* — Количество компьютеров с доступом в Интернет, нормированное на численность населения; *unempl* — уровень безработицы; *gorod* — процент городского населения; *rus* — процент русского населения; *cpi* — индекс потребительских цен).

Часть В. Продвинутая. Задачи части В проверяются, только если за решения задач части А набрано 10 и более баллов

7. (2 балла) Сформулируйте как минимум 3 способа выявления мультиколлинеарности в модели множественной регрессии.

8. (2 балла) Каковы последствия гетероскедастичности? Сформулируйте как минимум два.

9. (3 балла) Проинтерпретируйте график из модели регрессии процента голосов за партию «Яблоко»: объясните, что показывают *Leverage* и *Normalized Residual Squared*, о каких проблемах (если они есть) свидетельствует график и что Вы предложили бы сделать.

Table 1 The impact of presidential elections on the effective number of electoral parties. Dependent variable: Effective number of electoral parties

<i>Regressor</i>	<i>Model</i>
Proximity	−3.44** (0.49)
PresidentialCandidates	0.29* (0.07)
Proximity*PresidentialCandidates	0.82** (0.22)
Controls	—
Constant	3.01** (0.33)
R^2	0.34
N	522

* $p < 0.05$; ** $p < 0.01$ (two-tailed). Control variables not shown here.
 Robust standard errors clustered by country in parentheses.

- Проинтерпретируйте оценку коэффициента при переменной «PresidentialCandidate» (1 балл).
- Проинтерпретируйте оценку коэффициента при переменной «Proximity*PresidentialCandidates» (2 балла).

14. (4 балла) Для модели регрессии процента голосов за ЛДПР на социально-экономические показатели были получены МНК-оценки, и уравнение регрессии выглядит следующим образом:

$$Y_{abloko_i} = 13.8 \times railways_i + 3.76 \times grp_pc_i - 12.2 \times temperature_i + 9.2 \times gorod_i + e_i,$$

$n = 79, R^2 = 0.93,$

где *railways* — плотность железнодорожной сети, *grp_pc* — ВВП/чел., *temperature* — средняя июньская температура, *gorod* — процент городского населения.

На уровне доверия 95% статистически значимы ВВП/чел. и процент городского населения. Была также получена следующая выдача:

```
. estat vif
```

Variable	VIF	1/VIF
gorod	1.86	0.536349
temperature	1.56	0.639024
grp_pc	1.51	0.661598
railways	1.50	0.666150
Mean VIF	1.61	

В чем на Ваш взгляд причина такого высокого значения коэффициента детерминации R^2 ? Проинтерпретируйте R^2 .

15. (5 баллов) Некоторый неопытный исследователь решил проверить, не страдает ли построенная им модель от эндогенности, рассчитав коэффициент корреляции между модельными значениями зависимой переменной и регрессионными остатками. Объясните, почему эта идея бессмысленна? Каков будет полученный неопытным исследователем ответ?



10 Учебно-методическое и информационное обеспечение дисциплины

10.1 Базовые учебники

1. Доугерти К. Введение в эконометрику: Учебник. 2-е изд. – М.: ИНФРА-М, 1999. – 402 с.
2. Stock J, Watson M. Introduction to Econometrics (2nd edition). Addison Wesley Longman; 2008.

10.2 Основная литература

3. Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS: учеб. пособие для вузов. — М.: ГУ–ВШЭ, 2006. – 281 с.
4. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере: учебное пособие. – М.: ИД «ФОРУМ», 2008. – 368 с.
5. Bartels L.M. (1990). “Five Approaches to Model Specification.” *The Political Methodologist*, Vol. 3, No. 2, pp. 2–6.
6. Brambor, T., W.R. Clark, M. Golder, 2006. Understanding Interaction Models: Improving Empirical Analysis, *Political Analysis* 14: 63-82.
7. Gujarati, D.N. “Basic econometrics”. New York McGraw-Hill, 2003. - 1002 с.
8. King G. (1986). “How Not to Lie With Statistics: Avoiding Common Mistakes in Quantitative Political Science.” *American Journal of Political Science*, Vol. 30, pp. 666–687.
9. King G. (1990). “When Not to Use R-Squared.” *The Political Methodologist*, Vol. 3, No. 2, pp. 9–11.
10. Lewis-Beck, Michael S. and Andrew Skalaban. 1990. When to Use R-Squared. *The Political Methodologist* 3 (2): 11–12.
11. Luskin R. (1991). “R-Squared Encore.” *The Political Methodologist*, Vol. 4, No. 1, pp. 21–23.
12. A brief description of Stata. (http://www.stata-press.com/manuals/stata10/guide_brief.pdf).
13. Basics of Stata. (2008). (<http://www.ics.uci.edu/~dvd/Courses/Stata/stata.pdf>).
14. Учебно-методические материалы для преподавания курсов по применению количественного инструментария к решению социально-экономических задач (<http://www.hse.ru/jesda/mathbase/>).
15. Long, J. Scott and Jeremy Freese. 2005. *Regression Models for Categorical Dependent Variables Using Stata*. 2nd Edition. College Station, TX: Stata Press.

10.3 Дополнительная литература

16. Айвазян С. А., Мхитарян В. С. Прикладная статистика и основы эконометрики. В 2 частях. — М.: Юнити, 2001.
17. Айвазян С.А. Методы эконометрики. Учебник. — М.: Инфра-М, 2010. — 512 с.
18. Вакуленко Е.С., Камалова Р.У., Кисельгоф С.Г., Смирнова Ж.И., Стукал Д.К., Хавенсон Т.Е. Разработка учебно-методических материалов для преподавания курсов по применению количественного инструментария к решению социально-экономических задач. // *Экономическая социология (электронный журнал)*, 2010. Т. 11. № 4. С. 190—194. (http://ecsoc.hse.ru/data/2010/09/29/1234590525/ecsoc_t11_n4.pdf#page=190)
19. 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.
20. Единый архив социологических данных (рук. — Л. Б. Косова) // *Экономическая социология (электронный журнал)*, 2010, т. 11, № 1, с. 133–140. (http://www.hse.ru/data/2011/12/08/1208204985/ecsoc_t11_n1.pdf#page=133)



21. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс. 6-е изд., перераб. и доп. — М.: Дело, 2004. — 576 с.
22. Тюрин Ю.Н., Макаров А.А., Высоцкий И.Р., Ященко И.В. Теория вероятностей и статистика. — М.: МЦНМО, 2004.
23. Baum C.F. An Introduction to Modern Econometrics Using Stata, 2006, Stata Press.
24. Braumoeller B. (2006). “Explaining Variance.” *Political Analysis*, Vol.14, No.3, pp. 268–290.
25. Cameron, A. Colin and Pravin K. Trivedi. Microeconometrics using Stata, Revised Edition, 2010, Stata Press, 706 p.
26. Cook, R. D. and Weisberg, S. (1999). Applied Regression Including Computing and Graphics. New York: Wiley.
27. Fox, J., 2008. Applied Regression Analysis and Generalized Linear Models, London: Sage.
28. Fox, John. Applied Regression Analysis and Generalized Linear Models, 2nd Edition. Sage, 2008.
29. Fox, John. Regression Diagnostics: An Introduction. Sage, 1991.
30. Goertzel, T., 2002. “Myths of Murder and Multiple Regression”, *The Sceptical Inquirer* 26, 1: 19-23 (<http://crab.rutgers.edu/%7Egoertzel/mythsofmurder.htm>).
31. Graphics. UCLA: Academic Technology Services, Statistical Consulting Group. (<http://www.ats.ucla.edu/stat/stata/topics/graphics.htm>).
32. Kritzer, H.M., 1996. The Data Puzzle: The Nature of Interpretation in Quantitative Research, *American Journal of Political Science*, 40(1): 1-32.
33. Larocca, Roger. 2005. Reconciling Conflicting Gauss-Markov Conditions, *Political Analysis* 13:188–207.
34. Long, J. Scott. And Jeremy Freese. Regression Models for Categorical Dependent Variables Using Stata, 2001, Stata Press.
35. Regression Analysis. UCLA: Academic Technology Services, Statistical Consulting Group. (<http://www.ats.ucla.edu/stat/stata/topics/regression.htm>).
36. Resources to help you learn and use Stata. UCLA: Academic Technology Services, Statistical Consulting Group. (<http://www.ats.ucla.edu/stat/stata/>).
37. Stock J, Watson M. Introduction to Econometrics (2nd edition). Addison Wesley Longman; 2008.

10.4 Программные средства

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

10.5 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, Python.



Национальный исследовательский университет «Высшая школа экономики»
Программа дисциплины «Анализ политологических данных в пакете Stata» для направления
030200.62 «Политология» подготовки бакалавра

Для проведения семинарских занятий требуется компьютерный класс.