



Правительство Российской Федерации

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

Общеуниверситетская кафедра высшей математики

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 «__» _____ 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 политологических 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 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. Для указанного направления дисциплина является дисциплиной по выбору.

Изучение данной дисциплины базируется на следующих дисциплинах:

- «Алгебра и анализ» (1 курс),
- «Начала теории вероятностей и математической статистики» (1 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»,
- «Многомерные типологии политических систем современных государств».

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 кон-	Форма контроля	1 год	Параметры
----------	----------------	-------	-----------



т	роля	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 и В. 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

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_{ayd} .

Накопленная оценка учитывает результаты студента по текущему контролю следующим образом:

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

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

Если в домашней работе 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.

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:

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

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

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



Округление результирующей оценки также производится в соответствии с правилами математики.

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. Учебно-методические материалы для преподавания курсов по применению количественного инструментария к решению социально-экономических задач (<http://www.hse.ru/jesda/mathbase/>).
2. 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

1. Вакуленко Е.С., Камалова Р.У., Кисельгоф С.Г., Смирнова Ж.И., Стукал Д.К., Ха-венсон 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. 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)
3. A brief description of Stata. (http://www.stata-press.com/manuals/stata10/guide_brief.pdf).
4. Basics of Stata. (2008). (<http://www.ics.uci.edu/~dvd/Courses/Stata/stata.pdf>).
5. Long, J. Scott. 2009. The workflow of data analysis using Stata. TX: Stata Press. (Полочный индекс в библиотеке НИУ ВШЭ – 004 L87).
6. Resources to help you learn and use Stata. UCLA: Academic Technology Services, Statistical Consulting Group. (<http://www.ats.ucla.edu/stat/stata/>).
7. 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).
4. Stock J., Watson M. Introduction to Econometrics. Pearson Addison Wesley, 2008. Ch. 2.3-2.6, 3.4-3.7.

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 и категориальных переменных.

Основная литература

1. Доугерти К. Введение в эконометрику: Учебник. М.: ИНФРА-М, 1999. Гл. 2. С. 53–72.
2. Gujarati, D.N. Basic econometrics. New York McGraw-Hill, 2003. Ch. 1, 2.
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).
5. Stock J., Watson M. Introduction to Econometrics. Pearson Addison Wesley, 2008. Ch. 4.1-4.2, 5.1-5.3, 17.1-17.4.

Дополнительная литература

6. Айвазян С.А., Мхитарян В.С. Прикладная статистика и основы эконометрики. Т.2: Основы эконометрики. - М.: ЮНИТИ, 2001. – 656 с.
7. Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS : учеб. пособие для вузов. – М.: Изд. дом ГУ ВШЭ, 2006. – С. 115-146.
8. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере: учебное пособие. – М.: ИД «ФОРУМ», 2008. – С. 208-219.
9. Cook, R. D. and Weisberg, S. (1999). Applied Regression Including Computing and Graphics. New York: Wiley.
10. 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 значимость 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. Gujarati, D.N. Basic econometrics. New York McGraw-Hill, 2003. Ch. 3, 4.
3. Stock J., Watson M. Introduction to Econometrics. Pearson Addison Wesley, 2008. Ch. 4.3-4.5, 5.5-5.7, 6-7, 18.4-18.5.

Дополнительная литература

4. Доугерти К. Введение в эконометрику: Учебник. М.: ИНФРА-М, 1999. Гл. 6. 165-199.
5. Крыштановский А.О. Анализ социологических данных с помощью пакета SPSS : учеб. пособие для вузов. – М.: Изд. дом ГУ ВШЭ, 2006. – С. 146-166.
6. Тюрин Ю.Н., Макаров А.А. Анализ данных на компьютере: учебное пособие. – М.: ИД «ФОРУМ», 2008. – С. 208-219.
7. Cook, R. D. and Weisberg, S. (1999). Applied Regression Including Computing and Graphics. New York: Wiley.
8. Lewis-Beck, Michael S. and Andrew Skalaban. (1990). “When to Use R-Squared”. *The Political Methodologist* 3(2): 11–12.
9. King G. (1990). “When Not to Use R-Squared.” *The Political Methodologist*, 3(2): 9–11.
10. Goertzel, T., 2002. “Myths of Murder and Multiple Regression”, *The Sceptical Inquirer* 26, 1: 19-23 (<http://crab.rutgers.edu/%7Egoertzel/mythsofmurder.htm>).
11. Larocca, Roger. 2005. Reconciling Conflicting Gauss-Markov Conditions, *Political Analysis* 13:188–207.

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. Gujarati, D.N. Basic econometrics. New York McGraw-Hill, 2003. Ch. 10, 12.
3. Stock J., Watson M. Introduction to Econometrics. Pearson Addison Wesley, 2008. Ch. 6.7.

Дополнительная литература

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. Gujarati, D.N. Basic econometrics. New York McGraw-Hill, 2003. Ch. 11.
3. Stock J., Watson M. Introduction to Econometrics. Pearson Addison Wesley, 2008. Ch. 5.4-5.7, 18.2, 18.5

Дополнительная литература

4. Braumoeller B. (2006). “Explaining Variance.” *Political Analysis*, Vol.14, No.3, pp. 268–290.
5. Cook, R. D. and Weisberg, S. (1999). Applied Regression Including Computing and Graphics. New York: Wiley.



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.
3. Bluman, Allan G. Elementary Statistics: a Step by Step Approach. 8th Ed. Larson & Farber, 2009. Ch. 3, pp. 142-171

Дополнительная литература

4. Cook, R. D. and Weisberg, S. (1999). Applied Regression Including Computing and Graphics. New York: Wiley.
5. 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. Gujarati, D.N. Basic econometrics. New York McGraw-Hill, 2003. Ch. 13.
3. 21–23.
4. Stock J., Watson M. Introduction to Econometrics. Pearson Addison Wesley, 2008. Ch. 6.4, 7.5, 9.

Дополнительная литература

5. Bartels L.M. (1990). “Five Approaches to Model Specification.” The Political Methodologist, Vol. 3, No. 2, pp. 2–6.
6. Cook, R. D. and Weisberg, S. (1999). Applied Regression Including Computing and Graphics. New York: Wiley.
7. Fox, John. Applied Regression Analysis and Generalized Linear Models, 2nd Edition. Sage, 2008. Ch. 9.
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.

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. Gujarati, D.N. Basic econometrics. New York McGraw-Hill, 2003. Ch. 9.
3. Stock J., Watson M. Introduction to Econometrics. Pearson Addison Wesley, 2008. Ch. 5.3, 8.3-8.4.

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. Brambor Th., Clark W., Golder M. (2006). “Understanding Interaction Models: Improving Empirical Analyses.” *Political Analysis*, Vol. 14, pp. 63–82.
6. Cook, R. D. and Weisberg, S. (1999). *Applied Regression Including Computing and Graphics*. New York: Wiley.

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)

Домашнее задание – самостоятельная (индивидуальная) письменная работа по результатам анализа массива политологических и социально-экономических данных с использованием изученных методов. Рекомендуемый объем: 8-20 тыс. символов с пробелами.

В электронном виде сдается до 16 июня, 23:59 МСК, на адрес rkamalova@hse.ru. В печатном виде сдается 17 июня перед зачетной работой. Отсутствие печатного текста автоматически влечет неудовлетворительный балл за домашнее задание.

Структура домашнего задания в обязательном порядке должна включать в себя следующие разделы:

1. Аннотация (от 500 до 800 знаков с пробелами)
2. Введение (постановка задачи, обзор литературы и краткая «приманка» для читателя в виде формулировки основных полученных в работе результатов)
3. Гипотезы (формулируются и обосновываются гипотезы работы)
4. Данные и используемые методы (описываются имеющиеся данные и используемые методы; выбор методов обосновывается)
5. Результаты (в табличной форме, соответствующей зарубежным публикационным стандартам, представляются полученные результаты, а также предлагается их интерпретация)
6. Обсуждение полученных результатов (описывается, в какой мере полученные результаты соответствуют существующей литературе; обсуждаются возможные причины расхождений; обсуждается устойчивость полученных результатов к изменению массива данных и изменению методов анализа)
7. Заключение (формулируются выводы работы, обсуждаются теоретические следствия из полученных выводов, указываются возможные направления дальнейшего уточнения результатов)
8. Список использованной литературы

Объем домашнего задания – 15-25 тыс. символов с пробелами.

Требования к оформлению: шрифт Times New Roman, 12 кегль, полуторный интервал; нумерация страниц в левом нижнем углу с отсутствием номера на титульной странице.

Оцениваются адекватность формулировки задачи/проблемы, соответствие заявленных для проверки гипотез сформулированной задаче, корректность подбора методов и интерпретации результатов, обоснованность выводов. Обращайте внимание на грамотность русского языка и корректность использования терминов, аккуратность оформления, в т.ч. библиографии.



Задание выполняется в пакете для анализа данных Stata. Возможно самостоятельное использование других программ: Excel, SPSS, R.

Если объем и характер заимствования, оформленного в виде ссылки, ставят под сомнение самостоятельность выполнения работы, преподаватель действует в соответствии с п. 2.5 Положения о плагиате НИУ ВШЭ (http://www.hse.ru/org/hse/antiplagiat_info/plagiat).

Темы домашних заданий:

- Социально-экономические факторы электорального поведения россиян: анализ результатов выборов 4 декабря 2011 г. на уровне субъектов федерации.
- Этническая гетерогенность населения как фактор гражданских конфликтов.
- Зависимость экономического роста от качества государственного управления (на данных проекта Всемирного банка "Worldwide Governance Indicators").
- Социально-экономические показатели регионов Российской Федерации и эффективность деятельности губернаторов (по методике оценки эффективности, утвержденной Указом Президента Российской Федерации от 21 августа 2012 года № 1199 «Об оценке эффективности деятельности органов исполнительной власти субъектов Российской Федерации»).

9.2 Примерные вопросы для оценки качества освоения дисциплины

1. Перед Вами следующий результат из пакета Stata. Запишите медиану проеобразованного индекса Гастила и дисперсию.

```
. summarize gastil , detail
```

Индекс Гастила				
	Percentiles	Smallest		
1%	1	1		
5%	1	1		
10%	1	1	obs	85
25%	1.333333	1	Sum of wgt.	85
50%	2.277778		Mean	2.44469
		Largest	Std. Dev.	1.228502
75%	3.444444	4.666667	Variance	1.509218
90%	4.277778	4.875	Skewness	.3933883
95%	4.555555	4.888889	Kurtosis	1.881263
99%	4.888889	4.888889		

2. Перед Вами следующий результат из пакета Stata.

	gin~8090	cp19500	gastil
gini_8090	1.0000		
	72		
cp19500	0.3223*	1.0000	
	0.0100		
	63	71	
gastil	0.3800*	0.6819*	1.0000
	0.0010	0.0000	
	72	71	85

Формально запишите H_0 и H_1 относительно связи индекса демократии (gastil) и индекса восприятия коррупции (cp1). Запишите формулу расчета коэффициента корреляции Пирсона, его величину, соответствующее p-value и число наблюдений. Сделайте вывод относительно H_0 .

3. Докажите, что

$$Corr(\hat{y}, y) = \sqrt{R^2}$$

4. Докажите, что

$$Cov(\hat{y}, \hat{\varepsilon}) = 0$$



5. Рассчитайте t-статистику и соответствующее p-value для коэффициента $b_1 = 0.093$ со стандартной ошибкой $s.e(b_1) = 0.003$.
6. Сформулируйте условия Гаусса-Маркова, укажите условия «<<ядра>>» и «<<периферии>>».
7. Представлены данные об эффективности борьбы с коррупцией и качестве гос. управления в 5 странах мира. Оба показателя измерены в шкале от 0 до 15. Постройте модель парной линейной регрессии качества гос. управления на эффективность борьбы с коррупцией, зная, что оценка коэффициента корреляции Пирсона между двумя переменными равна 0.957, а также известно, что стандартное отклонение показателя борьбы с коррупцией равно 3.899, а эффективности борьбы с коррупцией – 3.578. Выпишите уравнение регрессии.

<i>n</i>	<i>Качество гос.управления</i>	<i>Эф-ть борьбы с коррупцией</i>
1.	1	2
2.	2	5
3.	4	4
4.	6	8
5.	10	12

8. Покажите, что

$$RSS = \sum_{i=1}^n (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2 = S_{yy} - \hat{\beta}_1 S_{xy},$$

где $S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2$, а $S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})$.

9. Вам известно, что в уравнении регрессии качества гос. управления на эффективность борьбы с коррупцией, известно, что $S_{yy} = 12.8$, $S_{xy} = 9.66$. Рассчитайте остаточную сумму квадратов.
10. Рассчитайте коэффициент детерминации в модели регрессии качества гос. управления на эффективность борьбы с коррупцией. Считайте, что объясненная сумма квадратов равна 46.9.
11. В регрессии качества гос. управления на эффективность борьбы с коррупцией проверьте гипотезу о том, что модель на константу не хуже модели с предиктором. $n = 5$. Не забудьте сформулировать гипотезу на языке математической статистики!
12. Вам предложена некоторая выдача пакета Stata для вышеописанной модели регрессии. Проверьте гипотезу о значимости коэффициента при предикторе x – эффективность борьбы с коррупцией.

```
. matrix list e(v)
      symmetric e(v) [2,2]
              x      _cons
      x      .02357095
      _cons  -.14613992  1.1926903
```

13. 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.
14. Сформулируйте как минимум 3 способа выявления мультиколлинеарности в модели множественной регрессии. Каковы последствия мультиколлинеарности? Сформулируйте как минимум 2 причины возникновения гетероскедастичности в модели регрессионной модели. Каковы последствия гетероскедастичности?
15. Укажите два случая, когда можно считать, что оценки коэффициентов регрессии описываются нормальным распределением.
16. Опишите идею теста Уайта (White test).



17. Проверьте модель на мультиколлинеарность, используя предложенные выдачи (сделайте вывод, страдает ли модель от мультиколлинеарности, и объясните, на каком основании сделан вывод --- сформулируйте гипотезы, приведите рассчитанные статистики и пр.). Если да, то как можно улучшить качество модели?
18. Перед Вами выдача Stata для регрессии процента голосов за партию «Яблоко» (*Yabloko*) в субъектах РФ на явку, %, (*yavka*), количество персональных компьютеров на 100 работников с доступом в Интернет (*comp_internet100*) и индекс Джини, 0-100, (*gini100*) в субъектах РФ. Сформулируйте исследовательский вопрос, для ответа на который, на Ваш взгляд, можно оценить приведенную модель. Выпишите в явном виде модель регрессии, проверьте гипотезы о статистической значимости, содержательно проинтерпретируйте оценки коэффициентов при предикторах, охарактеризуйте качество модели, используя доступные вам способы.

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

Source	SS	df	MS	Number of obs =	79
Model	167.343137	3	55.7810455	F(3, 75) =	34.02
Residual	122.957063	75	1.63942751	Prob > F =	0.0000
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	0.000	-.0897032 -.0398626
comp_int~100	.1984785	.0390788	5.08	0.000	.1206296 .2763274
gini100	.056672	.0677672	0.84	0.406	-.0783271 .1916711
_cons	1.587854	2.433101	0.65	0.516	-3.259133 6.434841

19. Как в регрессионной модели можно добиться R^2 , равного в точности единице, не зная процесса, порождающего данные?
20. В чем недостатки скорректированного R^2 ?
21. Предложите 2 критерия для сравнения моделей

$$\text{ВВП} = 5 - 3 \times \text{безработица} + 10 \times \text{производительность труда}, \\ R^2 = 0.54, n=200.$$

$$\text{ВВП} = 2 - 4 \times \text{безработица} + 9 \times \text{производительность труда} + 2 \times \text{демократия}, \\ R^2 = 0.62, n=200.$$

если известно, что в обоих случаях остатки не описываются нормальным распределением.

22. Каковы последствия включения нерелевантного регрессора?
23. Перечислите известные Вам шаговые алгоритмы отбора. В чем заключаются их недостатки?
24. В чем разница между двумя информационными критериями --- Акаике и Шварца?
25. Вам предложены две модели регрессии. Можно ли делать обоснованный выбор в пользу одной из них и на основании чего?



```
. reg ER gorod rus dohod_pc
==== omitted output ====
. estat ic
      Model |      Obs   ll(null)   ll(model)    df         AIC         BIC
-----+-----
      model 1 |      82   -343.6012   -312.0067     4       632.0135       641.6404

. reg ER gorod rus dohod_pc gini
==== omitted output ====
. estat ic
      Model |      Obs   ll(null)   ll(model)    df         AIC         BIC
-----+-----
      model 2 |      82   -343.6012   -310.4186     5       630.8373       640.8709
```

ER — % голосов за «Единую Россию» в регионе, gorod — % городского населения, rus — % русского населения, dohod_pc — среднедушевой доход, gini — коэффициент Джини.

26. 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. (Полочный индекс в библиотеке НИУ ВШЭ — 330.4 Д71).
2. Gujarati, D.N. “Basic econometrics”. New York McGraw-Hill, 2003. - 1002 с. (Полочный индекс в библиотеке НИУ ВШЭ — 330.4 G94).

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. 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.
8. King G. (1990). “When Not to Use R-Squared.” *The Political Methodologist*, Vol. 3, No. 2, pp. 9–11.
9. Lewis-Beck, Michael S. and Andrew Skalaban. 1990. When to Use R-Squared. *The Political Methodologist* 3 (2): 11–12.
10. Luskin R. (1991). “R-Squared Encore.” *The Political Methodologist*, Vol. 4, No. 1, pp. 21–23.
11. A brief description of Stata. (http://www.stata-press.com/manuals/stata10/guide_brief.pdf).
12. Basics of Stata. (2008). (<http://www.ics.uci.edu/~dvd/Courses/Stata/stata.pdf>).



13. Учебно-методические материалы для преподавания курсов по применению количественного инструментария к решению социально-экономических задач (<http://www.hse.ru/jesda/mathbase/>).
14. 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).

10.3 Дополнительная литература

15. Айвазян С. А., Мхитарян В. С. Прикладная статистика и основы эконометрики. В 2 частях. — М.: Юнити, 2001.
16. Айвазян С.А. Методы эконометрики. Учебник. — М.: Инфра-М, 2010. — 512 с.
17. 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)
18. 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.
19. 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)
20. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс. 6-е изд., перераб. и доп. — М.: Дело, 2004. — 576 с.
21. Тюрин Ю.Н., Макаров А.А., Высоцкий И.Р., Ященко И.В. Теория вероятностей и статистика. — М.: МЦНМО, 2004.
22. Braumoeller V. (2006). “Explaining Variance.” *Political Analysis*, Vol.14, No.3, pp. 268–290.
23. Cook, R. D. and Weisberg, S. (1999). *Applied Regression Including Computing and Graphics*. New York: Wiley.
24. Data Management. UCLA: Academic Technology Services, Statistical Consulting Group. (http://www.ats.ucla.edu/stat/stata/topics/data_management.htm).
25. Descriptive information and statistics. UCLA: Academic Technology Services, Statistical Consulting Group. (<http://www.ats.ucla.edu/stat/stata/modules/descript.htm>).
26. Fox, J., 2008. *Applied Regression Analysis and Generalized Linear Models*, London: Sage.
27. Fox, John. *Applied Regression Analysis and Generalized Linear Models*, 2nd Edition. Sage, 2008.
28. Fox, John. *Regression Diagnostics: An Introduction*. Sage, 1991.
29. Goertzel, T., 2002. “Myths of Murder and Multiple Regression”, *The Sceptical Inquirer* 26, 1: 19-23 (<http://crab.rutgers.edu/%7Egoertzel/mythsofmurder.htm>).
30. Graphics. UCLA: Academic Technology Services, Statistical Consulting Group. (<http://www.ats.ucla.edu/stat/stata/topics/graphics.htm>).
31. Kritzer, H.M., 1996. The Data Puzzle: The Nature of Interpretation in Quantitative Research, *American Journal of Political Science*, 40(1): 1-32.
32. Larocca, Roger. 2005. Reconciling Conflicting Gauss-Markov Conditions, *Political Analysis* 13:188–207.
33. Long, J. Scott. 2009. *The workflow of data analysis using Stata*. TX: Stata Press. (Полочный индекс в библиотеке НИУ ВШЭ – 004 L87).
34. Regression Analysis. UCLA: Academic Technology Services, Statistical Consulting Group. (<http://www.ats.ucla.edu/stat/stata/topics/regression.htm>).



35. Resources to help you learn and use Stata. UCLA: Academic Technology Services, Statistical Consulting Group. (<http://www.ats.ucla.edu/stat/stata/>).
36. Stata Annotated Output. Regression Analysis. UCLA: Academic Technology Services, Statistical Consulting Group. (http://www.ats.ucla.edu/stat/stata/output/reg_output.htm).

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.