

Домашнее задание №3.
Линейная регрессия и корреляция.
Группа 233

Дедлайн: 5 марта, 15.10.

На сайте курса скачайте базу данных `norris.sav` (также она доступна на сайте <http://www.hks.harvard.edu/fs/pnorris/Data/Data.htm>). В этой базе, собранной профессором политологии Гарвардского университета Пиппой Норрис, содержатся результаты парламентских выборов в Великобритании 2010 г. Единицей анализа является избирательный округ (`constituency`). Для анализа вам понадобятся следующие переменные:

- `Seat` – название избирательного округа (вы можете воспользоваться сайтом <http://www.election-maps.co.uk/searchByName.jsf>, чтобы посмотреть границы избирательных округов),
- `Region` – название региона, к которому принадлежит избирательный округ,
- `Con10` – процент голосов, поданных за Консервативную партию,
- `Lab10` – процент голосов, поданных за Лейбористскую партию,
- `ACORN_1p`, `ACORN_2p`, `ACORN_3p`, `ACORN_4p`, `ACORN_5p` – проценты социальных групп, выделенных согласно классификации ACORN, разработанной маркетинговой фирмой CACI (<http://www.caci.co.uk/acorn-classification.aspx>) на основе данных переписи 2001 г. Сумма этих переменных составляет 100%, т.е. население в целом. Описание групп доступно на сайте <http://www.guardian.co.uk/uk/2003/nov/23/britishidentity.socialexclusion>:
 - “Wealthy achievers” – наиболее обеспеченная прослойка населения, менеджеры и специалисты от 45 до 65 лет, живущие в отдельных домах (около 25% населения Великобритании),
 - “Urban prosperous” – от 25 до 35 лет, с университетским образованием, обеспечены, часто не женаты или не замужем (11% населения),
 - “Comfortably off” – люди со средним доходом (27% населения),
 - “Moderate means” – обеспеченный рабочий класс (15% населения),
 - “Hard pressed” – наименее обеспеченные люди (22% населения).

1. Найдите избирательные округа, в которых за Консервативную партию был подан максимальный и минимальный процент голосов. Повторите анализ для Лейбористской партии. Опишите эти округа с точки зрения их расположения и социального состава населения.
2. Рассчитайте и представьте регрессионную модель зависимости процента голосов, поданных за Консервативную партию, от региона и социального состава жителей избирательного округа. В качестве базового региона выберите Лондон.
3. Повторите тот же анализ для Лейбористской партии.
4. Выполненное задание должно представлять собой эссе, в котором обсуждалась бы зависимость голосования за Консервативную и Лейбористскую партию в Великобритании от региона, в котором находится избирательный округ, и социального состава населения избирательного округа. Вам необходимо представить эмпирические свидетельства этой зависимости (задания 1-3) и объяснить полученные результаты. Дайте ответы на следующие вопросы: почему в одних регионах люди больше поддерживают Лейбористскую, а в других Консервативную партию? Почему социальный состав избирательных округов связан с голосованием за эти две партии?

Советы по выполнению задания:

- Перед выполнением задания вкратце ознакомьтесь с партийной системой Великобритании (см., например, http://en.wikipedia.org/wiki/Politics_of_the_United_Kingdom#Political_parties).
- Статистические результаты должны быть подчинены обсуждению содержательных вопросов.
- Помните, что регион является категориальной переменной, которая вводится в регрессионное уравнение как набор фиктивных переменных.
- Вы не можете одновременно добавить в регрессию все пять переменных, характеризующих социальный состав населения. В сумме они составляют 100%, и вы не можете менять одну из этих переменных, удерживая остальные на одном уровне. Подумайте, какие переменные имеет смысл ввести в модель, чтобы решить поставленные содержательные задачи. Вы можете создать новые переменные (например, процент *wealthy achievers* и *urban prosperous* или процент *hard pressed* и *moderate means*) и добавить их в модель. Попробуйте разные варианты и выберите оптимальный.

- Не копируйте таблицы из SPSS. Таблицы с результатами регрессии должны включать следующую информацию: регрессионные коэффициенты, стандартные ошибки коэффициентов, статистическую значимость коэффициентов, количество наблюдений в модели и R-квадрат.
- Как всегда, в качестве приложения задание должно включать синтаксис SPSS.