

1. Прочитайте про RLMS, <http://www.hse.ru/rlms/>

Посмотрите описание проекта. Проллистайте вестник RLMS, чтобы иметь представление о том, какие исследования можно строить на основе RLMS.

2. Скачайте любую волну RLMS по своему выбору. Скачайте описание переменных.

Проллистайте описание переменных. Там их больше тысячи. Попадаютсь довольно прикольные. Мне нравится rc9.6.5a, «У Вас есть GPRS навигатор?»

3. Загрузите данные в R.

Данные RLMS выложены на сайте в формате SPSS. SPSS — это потихоньку погибающий статистический пакет для домохозяек. Для удобства можно воспользоваться готовой функцией для чтения данных RLMS в пакете `rlms`.

```
library("rlms")  
h <- rlms_read("/home/boris/downloads/r20hall23c.sav")
```

Про установку пакета `rlms` можно прочитать на страничке <https://github.com/bdemeshev/rlms>

Описания переменных при этом также загружаются в таблицу данных. Можно их посмотреть:

```
var_meta <- rlms_show_variable_labels(df)  
var_meta
```

4. Выберите любую количественную переменную в качестве зависимой и несколько переменных в качестве объясняющих.

Цель этой домашки скорее ознакомится с наличием мониторинга RLMS, поэтому можно не сильно заморачиваться с этим этапом. Хотя в реальности тут-то всё самое интересное и начинается. За оригинальные гипотезы будут плюшки. Кстати, неплохо бы дать выбранным переменным понятные названия.

5. Опишите выбранные переменные.

Постройте симпатичные графики. Посчитайте описательные статистики. Много ли пропущенных наблюдений? Есть ли что-нибудь интересненькое?

6. Постройте регрессию зависимой переменной на объясняющие.

Проверьте гипотезу о значимости каждого полученного коэффициента. Проверьте гипотезу о значимости регрессии в целом. Для нескольких коэффициентов (двух достаточно) постройте 95%-ый доверительный интервал.

7. Разберитесь с возможным наличием гетероскедастичности в данных.

С какой переменной может быть связана дисперсия $\text{Var}(\varepsilon_i|X)$? Проведите визуальный анализ на гетероскедастичность. Проведите формальные тесты на гетероскедастичность. Примените оценки дисперсии $\hat{\beta}$ устойчивые к гетероскедастичности. Прокомментируйте. Может помочь http://bdemeshev.github.io/r_cycle/cycle_files/12_hetero.html

8. Покажите буйство своей фантазии и аккуратность!

Не стоит думать, что побуквенное выполнение этих инструкций гарантирует оценку в десять баллов. Эконометрика — это не ремесло, а искусство! Фантазируйте! Убедите меня в работе, что вы были на лекциях, даже если это не так :) Аккуратность в виде подписанных осей на графиках, указанных единицах измерения также не повредит.

9. Срок сдачи — 22 декабря 2016 года.

10. Работа должна быть написана с применением грамотного программирования R + \LaTeX или markdown. Каждый день более поздней сдачи умножает оценку за работу на 0.8. Работа должна представлять слитный текст, код скрывать не нужно. В конце должна быть команда `sessionInfo()`.
11. Работа сдаётся обязательно в двух вариантах: печатном и электронном. Электронная версия должна быть прислана на boris.demeshev@gmail.com строго под темой «em301: ht 01».