

1. RLMS и гетероскедастичность

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

Посмотрите описание проекта. Пролитайте 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.

Пролитайте описание переменных. Там их больше тысячи. Попадаютсь довольно прикольные. Мне нравится rs9.6.5a, «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.

Жгите! Подумайте о зависимости, которую вы хотите обнаружить. 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. Постройте прогнозы для выбранных зависимой переменной.

Придумайте вымышленного индивида или возьмите свои личные данные для построения прогнозов.

2. Временные ряды

1. Возьмите два временных ряда: один — с ежемесячными данными с сезонностью, второй — с годовыми без сезонности.

Разделите ряд на две части: обучающую и тестовую. Для месячного ряда в качестве тестовой выборки используйте последний год, для годичного — последние три. Далее оценку моделей производите по обучающей выборке.

2. Постройте графики рядов, графики корреляционной и частной корреляционной функций. Если нужно, постройте аналогичный график для разностей Δy_t .

Похожи ли ряды на стационарные?

3. Для ряда без сезонности предположите порядок ARMA модели исходя из построенных графиков. Оцените предположенную ARMA модель и аккуратно выпишите полученное уравнение.
4. Для обоих рядов самостоятельно выберите ETS модель. Оцените её и аккуратно выпишите результаты.
5. Для обоих рядов оцените автоматически подбираемую ARMA-модель с помощью функции `'auto.arima()'`. Выпишите полученное уравнение.
6. Для обоих рядов оцените автоматически подбираемую ETS модель с помощью функции `'ets()'`. Выпишите результаты оценивания.
7. Сравните ETS и ARIMA по качеству прогнозов. Для наилучшей модели по каждому ряду постройте графики прогнозов.
8. Очень важно корректно выписывать уравнения всех оцениваемых моделей. Маловероятно, что у новичка ручной подбор модели окажется лучше автоматической процедуры. Цель — понять, какую модель на выходе выдают используемые процедуры оценивания.
9. Напоите верблюдов своего любопытства в источнике знаний <https://www.otexts.org/fpp>.

В целом про работу:

1. Покажите буйство своей фантазии и аккуратность!
Смелее рискуйте что-то попровать, можно не ограничиваться списком! Но не забывайте про аккуратность в виде подписанных осей графика, указанных единиц измерения.
2. Работа должна быть написана с применением грамотного программирования R + \LaTeX или R + markdown. Работа должна представлять слитный текст, код скрывать не нужно, бессмысленные сообщения от R нужно скрыть. В конце должна быть команда `sessionInfo()`.
3. Работу можно делать в одиночку или группой из двух или трёх человек.
4. Дедлайн: второй по счёту семинар после 16 февраля.