

видео третьей главы:

3.1.1. Прогнозирование во множественной регрессии

<http://youtu.be/6x5FmRuTk08>

1:17 сейчас появляется надпись «Доверительный интервал ...» с формулой внизу. Они должны появиться только в 2:20

1:30 добавляем надпись «Точность прогноза определяется шириной доверительного интервала»

4:10 исправить на:

условное среднее,  $E(y_i|X)$

ошибка прогноза условного среднего,  $\hat{y}_i - E(y_i|X)$

дисперсия ошибки прогноза:

$$Var(\hat{y}_i - E(y_i|X)|X) = Var(\hat{y}_i|X) = Var(\hat{\beta}_1 + \hat{\beta}_2 x_i + \hat{\beta}_3 z_i|X)$$

5:04 исправить на:

конкретное наблюдение,  $y_i$

ошибка прогноза,  $\hat{y}_i - y_i$

дисперсия ошибки прогноза:

$$\begin{aligned} Var(\hat{y}_i - y_i|X) &= Var(\hat{y}_i - E(y_i|X) - \varepsilon_i|X) = Var(\hat{y}_i - \varepsilon_i|X) = \\ &= Var(\hat{y}_i|X) + Var(\varepsilon_i|X) = Var(\hat{\beta}_1 + \hat{\beta}_2 x_i + \hat{\beta}_3 z_i|X) + Var(\varepsilon_i|X) \end{aligned}$$

(!сейчас опечатка в начале формулы! сейчас там написано  $|X)|X)$ , кроме этого в формуле всё ок )

6:25 опечатка — пропущен пробел в «предположении о нормальности»

7:14 опечатка — пропущен пробел в «предположении о нормальности»

3.1.2 Пример построения интервалов для прогнозов

<http://youtu.be/swWNng-RBho>

ок

3.1.3. Интерпретация коэффициента при логарифмировании

<http://youtu.be/NdsfDkP0qVI>

7:50 пропущен нижний индекс 2 у коэффициента бета, должно быть:

С ростом  $x$  на единицу  $y$  растет на  $\beta_2$  единиц

7:57 пропущен нижний индекс 2 и окончание "ов"

С ростом  $x$  на один процент  $y$  растет на  $\beta_2$  процентов

8:05 пропущен нижний индекс 2 у коэффициента бета, должно быть:

С ростом  $x$  на единицу  $y$  растет на  $100\beta_2$  процентов

8:18 пропущен нижний индекс 2 и ошибка, должно быть:

С ростом  $x$  на один процент  $y$  растет на  $\beta_2/100$  единиц

3.1.4. Исправить название на «Дамми-переменные. Разные зависимости для подвыборок»

[http://youtu.be/wJj\\_oasiKgU](http://youtu.be/wJj_oasiKgU)

0:28 немного исправить на: ( $male_i$  — синим цветом, тк это формула):

Например, переменная  $male_i$ , равная 1 для мужчин и 0 — для женщин

4:02 перепутаны М и Ж, и лишний плюс в формуле, должно быть:

Для мужчин:

$$wage_i = (\beta_1 + \beta_4) + (\beta_2 + \beta_5)exper_i + \beta_3educ_i + \varepsilon_i$$

4:16 перепутаны М и Ж, должно быть:

Для женщин:

(формула верно)

5:31 перепутаны М и Ж, должно быть:

Для мужчин:

5:39 перепутаны М и Ж, должно быть:

Для женщин:

6:34 перепутаны М и Ж, должно быть:

(формула)

Для мужчин:

(формула)

Для женщин:

(формула)

7:45 сделать  $season_i$  синим цветом

8:03 в новом пункте немного переставить слова:

\* Вводим три дамми-переменных

(четыре сезона минус один базовый)

8:08-8:24 сделать  $vesna_i$ ,  $leto_i$ ,  $osen_i$  синим цветом

8:30 очистить старые надписи и вывести такую табличку

Наблюдение	Сезон	$vesna_i$	$leto_i$	$osen_i$
1	Зима	0	0	0
2	Весна	1	0	0
3	Лето	0	1	0
4	Осень	0	0	1
5	Зима	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$

12:00 оставляем прежний заголовок «Частая ошибка!» и оставляем подпись под ним

«Включить дамми-переменные на все значения факторной переменной и константу в регрессию»

ниже появляется: (сейчас лишняя буква й в линейно)

«Регрессоры линейно зависимы. Не существует единственных МНК оценок»

12:14 надпись появляется как сейчас, всё ок

12:22 надпись «Синонимы...» убираем полностью

3.1.5. Проверка гипотезы о нескольких линейных ограничениях. (6 минут)

<http://youtu.be/NqJH0VGeEhU>

1:05 — мигает pdf слайд, убираем

1:49 — виден pdf слайд, убираем

2:57 — виден pdf слайд, убираем

4:33 — виден pdf слайд, убираем

4:48 сюда вставляем кусок от 3.1.6 начинающийся с 8:32 и до конца

3.1.6. Пример проверки гипотезы о нескольких линейных ограничениях

<http://youtu.be/xludTSWApYM>

8:32—до конца отрезаем и вставляем в 3.1.5 в момент 4:48

3.1.7. добавить синюю полосу с названием фрагмента в начале «Гипотеза о незначимости регрессии»

<http://youtu.be/fxnJk6bb7Yo>

0:24 появляется заголовок «Гипотеза о незначимости регрессии»

0:30 появляется формула

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \beta_4 w_i + \beta_5 d_i + \beta_6 s_i + \varepsilon_i$$

0:40 появляется еще одна формула с подписью ниже:

$$H_0 : \begin{cases} \beta_2 = 0 \\ \beta_3 = 0 \\ \beta_4 = 0 \\ \vdots \end{cases}$$

Всего  $(k - 1)$  ограничение

1:12 очищаем экран (кроме заголовка) и пишем:

Общая формула:

$$F = \frac{(RSS_R - RSS_{UR})/r}{RSS_{UR}/(n - k_{UR})} \sim F_{r, n - k_{UR}}$$

Частный случай:

$$F = \frac{ESS/(k - 1)}{RSS/(n - k)} \sim F_{k-1, n-k}$$

3.1.8. Пример проверки гипотезы о незначимости регрессии

<http://youtu.be/xbpA6Dji5mY>

3.1.9. Лишние и пропущенные регрессоры

[http://youtu.be/80\\_36Dfeba8](http://youtu.be/80_36Dfeba8)

0:24 изменяем заголовок на «Предпосылки:» (убираем «снова БСХС»)

1:01 изменяем заголовок на «Предположения на  $\varepsilon_i$ :» (убираем БСХС)

1:27 перед формулой добавляем «Условная некоррелированность», т.е. должно быть:

\* Условная некоррелированность

$Cov(\varepsilon_i, \varepsilon_j | X) = 0$  при  $i \neq j$

1:34 изменяем заголовок на «Предпосылки на регрессоры:» (убираем БСХС)

2:05 изменяем заголовок на «Асимптотические свойства (плюс новое):» (убираем БСХС)

2:40 формулу  $\hat{\sigma}^2 = \frac{RSS}{n-k}$  удаляем полностью

3:03 изменяем заголовок на «Свойства при нормальности (плюс новое):» (убираем БСХС)

3:11 добавить двоеточие после слова «Новое»

3.1.10. Тест Рамсея

<http://youtu.be/51qvueLMChw>

2:20 исправить формулу, убрать крышку над первым  $y_i$ , должно быть:

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \gamma_1 \hat{y}_i^2 + \gamma_2 \hat{y}_i^3 + \dots + \gamma_p \hat{y}_i^{p+1} + \varepsilon_i$$

### 3.1.11. Простые показатели качества модели

<http://youtu.be/ZG10ncL0H4w>

1:32 немного поменять, чтобы вышло:

$$* R^2 = \frac{ESS}{TSS} = 1 - \frac{RSS}{TSS}$$

растёт с добавлением регрессоров,  $R_{UR}^2 > R_R^2$ ,

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

3:52 Добавить строчку:

Штрафуем модель за большое  $k$  и большую  $RSS$

5:35 добавить в список пункт

\* простые критерии качества

### 3.2.1. Графики и переход к логарифмам

Диаграмма рассеяния для большого количества переменных, мозаичный график

<http://youtu.be/eZssYMz7Fgs>

5:36 - 6:36 (примерно) — удалить из видео. По смыслу: там график выделился синим, и я пытался снять выделение, но никак не получалось, пока я его не перестроил. Помоему, хороший момент склейки — когда я кликал на кнопку zoom. От момента клика до момента клика можно вырезать.

### 3.2.2. Графики для качественных и количественных переменных

(гистограммы разным цветом, фасетки) <http://youtu.be/Ng-W5BqlpzE>

0:16 заменить название лекции на «Проверка сложных гипотез и прогнозирование».

Текущее название «мультиколлинеарность» относится к четвертой лекции.

### 3.2.3. Изменить название фрагмента на «Оценивание моделей с дамми-переменными.

Интерпретация»

<http://youtu.be/Syab2Kc9Cs0>

0:16 заменить название лекции на «Проверка сложных гипотез и прогнозирование».

Текущее название «мультиколлинеарность» относится к четвертой лекции.

### 3.2.4. Изменить название фрагмента на «Построение прогнозов. Доверительный и пре-

диктивный интервал».

<http://youtu.be/c0wFwacaAV8>

0:16 заменить название лекции на «Проверка сложных гипотез и прогнозирование».

Текущее название «мультиколлинеарность» относится к четвертой лекции.

### 3.2.5. Изменить название фрагмента на «Проверка гипотезы о линейных ограничениях.

Графическое представление результатов»

<http://youtu.be/Du0GNgkmyR8>

0:16 заменить название лекции на «Проверка сложных гипотез и прогнозирование».

Текущее название «мультиколлинеарность» относится к четвертой лекции.

### 3.2.6. Ловушка дамми-переменных. Информационные критерии. Тест Рамсея.

<http://youtu.be/FW3v22BDY98>

0:16 заменить название лекции на «Проверка сложных гипотез и прогнозирование».

Текущее название «мультиколлинеарность» относится к четвертой лекции.

### 3.2.7 Нано-исследование

<http://youtu.be/ySVk5QgHGfQ>

0:16 заменить название лекции на «Проверка сложных гипотез и прогнозирование».

Текущее название «мультиколлинеарность» относится к четвертой лекции.