

# Эконометрика, 2020-2021, 2 модуль

## Семинар 6

30.11.20

для

Группы Э\_Б2018\_Э\_3

Семинарист О.А.Демидова

### Мультиколлинеарность данных. Метод главных компонент

1) Признаком мультиколлинеарности служит:

1. маленькие  $t$ -статистики при  $R^2$ , близком к 1
2. близкое к 0 значение коэффициента множественной детерминации
3. значительные изменения в оценках коэффициентов регрессии при небольших изменениях в данных
4. близкие к 0 значения коэффициентов корреляции регрессоров
5. все ответы верны

2) Оцененная с помощью МНК зависимость заработной платы индивида EARNINGS от его возраста AGE, опыта EXP, пола MALE, длительности обучения S, длительности обучения матери SM имеет вид (в скобках стандартные отклонения коэффициентов):

$$\widehat{EARN} = -24 - 0.099 AGE + 2.49 S + 0.26 SM + 0.46 EXP + 6.23 MALE, R^2 = 0.247$$

(10.6)      (0.25)      (0.249)      (0.24)      (0.14)      (1.11)

Были оценены также вспомогательные регрессии:

$$\widehat{AGE} = -.007 + 0.53 AGE - 0.6S + 0.23SM + 1.23MALE, R^2 = 0.2,$$

$$\widehat{S} = 8.47 + 0.095 AGE + 0.4SM - 0.2EXP + 0.12MALE, R^2 = 0.25,$$

$$\widehat{SM} = 6.16 - 0.045 AGE + 0.42S + 0.08EXP + 0.42MALE, R^2 = 0.18,$$

$$\widehat{EXP} = -.07 + 0.53 AGE - 0.6S + 0.23SM + 1.23MALE, R^2 = 0.2,$$

VIF для переменной EXP равен \_\_\_\_.

Ответ. 1.25

3) При применении к модели, результаты оценки которой приведены ниже,

EARNINGS	Coef.	Std. Err.	t	P> t	[95% Conf. Interval	
S	2.578227	.2288185	11.27	0.000	2.128729	3.027726
AGE	-10.70493	9.211662	-1.16	0.246	-28.80062	7.390769
Agesq	.1300605	.1125515	1.16	0.248	-.0910395	.3511605
EXP	.4429137	.1442633	3.07	0.002	.159518	.7263094
ETHHISP	-1.078255	2.268688	-0.48	0.635	-5.534941	3.378432
ETHBLACK	-4.014172	2.152185	-1.87	0.063	-8.241996	.2136528
MALE	6.364055	1.111968	5.72	0.000	4.179668	8.548442
_cons	193.7202	187.6859	1.03	0.302	-174.9761	562.4165

. vif

Variable	VIF	1/VIF
AGE	1411.96	0.000708
Agesq	1411.13	0.000709
EXP	1.29	0.778114
S	1.14	0.875122
ETHBLACK	1.04	0.962602
MALE	1.03	0.966488
ETHHISP	1.02	0.983851
Mean VIF	404.09	

метода последовательного исключения, на ближайшем шаге из уравнения регрессии будет удалена переменная

- 1) S 2) AGE 3) EXPSQ 4) EXP 5) ETHWHITE 6) ETHHISP 7) FEMALE  
8) ни одна из перечисленных

4) Первой главной компонентой системы показателей  $X_1, \dots, X_k$  называется такая линейная комбинация этих показателей

1. в которой коэффициент при  $X_1$  равен 1 2. которая обладает наименьшей дисперсией 3. которая обладает наибольшей дисперсией 4. которая ортогональна всем  $X_j, j = 1, \dots, k$

5) (Д.А.Борзых, Б.Б.Демешев, задача 7.4)

Пионеры, Крокодил Гена и Чебурашка собирали металлолом несколько дней подряд. В распоряжение иностранной шпионки, гражданки Шапокляк, попали ежедневные данные по количеству собранного металлолома: вектор  $g$  – для Крокодила Гены, вектор  $h$  – для Чебурашки и вектор  $x$  – для пионеров. Гена и Чебурашка собирали вместе, поэтому выборочная корреляция  $\hat{c}or(g, h) = -0.9$ . Гена и Чебурашка собирали независимо от пионеров, поэтому  $\hat{c}or(g, x) = 0$ ,  $\hat{c}or(h, x) = 0$ . Если регрессоры  $g, h, x$  центрировать и нормировать, то получится матрица  $\tilde{X}$ .

1) Найдите параметр обусловленности матрицы  $\tilde{X}\tilde{X}$ .

2) Вычислите одну или две главные компоненты (выразите их через вектор-столбцы матрицы  $\tilde{X}$ ), объясняющие не менее 70% общей выборочной дисперсии регрессоров.

3. Шпионка Шапокляк пытается смоделировать ежедневный выпуск танков,  $y$ . Выразите оценки коэффициентов регрессии  $y = \beta_1 + \beta_2 g + \beta_3 h + \beta_4 x + \varepsilon$  через оценки коэффициентов регрессии на главные компоненты, объясняющие не менее 70% общей выборочной дисперсии.

Ответ.  $(\tilde{X}_1 - \tilde{X}_2)/\sqrt{2}; \tilde{X}_3$

- 6) (Демешев, Борзых, 7.13)

Известно, что выборочная корреляция между переменными  $x$  и  $z$  равна 0.9.

1. Найдите коэффициенты  $VIF$  для  $x$  и  $z$  в регрессии  $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \varepsilon_i$ .
2. В каких пределах могут лежать коэффициенты  $VIF$  для  $x$  и  $z$  в регрессии  $y_i = \beta_1 + \beta_2 x_i + \beta_3 z_i + \beta_4 w_i + \varepsilon_i$ ?

- 7) (Демешев, Борзых, 7.11)

Эконометресса Алевтина перешла от исходных регрессоров к трём главным компонентам,  $z_1$ ,  $z_2$  и  $z_3$ . И далее посчитала коэффициенты вздутия дисперсии,  $VIF_j$ , для главных компонент. Чему они оказались равны?

### Проблемы мультиколлинеарности при моделировании продаж одежды

В файле clothing (STATA) содержатся данные о продажах одежды в 400 немецких магазинах одежды.

#### Переменные:

tsales – среднегодовые продажи в гульденах,  
sales - продажи в расчете на квадратный метр,  
margin – маржинальная валовая прибыль,  
pown – количество собственников (менеджеров),  
nfull – количество полностью занятых,  
part - количество частично занятых,  
paux – количество временно работающих,  
hoursw – общее число отработанных часов,  
hourspw – количество отработанных часов в расчете на одного работающего,  
inv1 – капиталовложения в помещения,  
inv2 - капиталовложения в автоматизацию,  
ssize – размер магазина в  $m^2$ ,  
start – год открытия магазина.

- 1) Оцените зависимость среднегодовых продаж (переменная tsales) или продаж в расчете на квадратный метр (sales) или маржинальной валовой прибыли (margin) от всех остальных переменных.
- 2) Проверьте адекватность регрессии. Если регрессия адекватна, то переходите к следующим пунктам.
- 3) Рассчитайте  $VIF$ -ы. Существует ли для построенной регрессии проблема мультиколлинеарности?

- 4) Выберите факторы, которые должны быть исключены из уравнения регрессии, используя метод пошагового исключения незначимых переменных.
- 5) Выберите факторы, которые должны быть включены в уравнение регрессии, используя метод пошагового включения переменных.
- 6) Сравните результаты, полученные в пунктах 1, 4, 5. В качестве показателя качества подгонки регрессии используйте коэффициент множественной детерминации, скорректированный на число степеней свободы. Дайте экономическую интерпретацию полученным результатам.

**Методические рекомендации по выполнению упражнения в пакете STATA**

1) Для оценки, например, параметров уравнения регрессии

$$\text{sales} = \beta_1 + \beta_2 \text{nown} + \beta_3 \text{nfull} + \beta_4 \text{npart} + \beta_5 \text{naux} + \beta_6 \text{hoursw} + \beta_7 \text{hourspw} + \beta_8 \text{inv1} + \beta_9 \text{inv2} + \beta_{10} \text{ssize} + \beta_{11} \text{start} + \varepsilon$$

наберите в командном окне

`reg sales nown nfull npart naux hoursw hourspw inv1 inv2 ssize start`

2) Для нахождения VIFов после оценки уравнения регрессии необходимо набрать команду `vif`

и воспользоваться выданной таблицей..

3) Для применения метода пошагового исключения незначимых переменных из уравнения регрессии, наберите в командном окне

`stepwise, pr(0.1): reg имя зависимой переменной имена независимых переменных`

Выбранный в скобках уровень значимости 0.1 можно изменить.

4) Для применения метода пошагового включения переменных в уравнение регрессии, наберите в командном окне

`stepwise, pr(0.1): reg имя зависимой переменной имена независимых переменных`

Выбранный в скобках уровень значимости 0.1 можно изменить.