

Методический семинар  
12.04.2012, часть 3

**Мультиколлинеарность**

Демидова О.А.

# Мультиколлинеарность

**Теоретическая мультиколлинеарность данных – явление, наблюдаемое при нарушении условий теоремы Гаусса – Маркова об отсутствии точной линейной связи между регрессорами. При наличии теоретической мультиколлинеарности однозначное нахождение оценок МНК коэффициентов регрессии невозможно.**

# Мультиколлинеарность

**При работе с реальными данными часто имеет место квазимультиколлинеарность, когда между регрессорами существует почти линейная зависимость.**

# Мультиколлинеарность

## Признаки мультиколлинеарности:

- Небольшие изменения в данных приводят к значительным изменениям в оценках коэффициентов регрессии.
- Многие коэффициенты по-отдельности не значимы, хотя в целом регрессия адекватная,  $R^2$  может быть достаточно высоким.
- Оценки коэффициентов регрессии (обычно незначимых) могут иметь “неправильный” знак (с экономической точки зрения).

# Мультиколлинеарность

**Индикаторы мультиколлинеарности:**

- В корреляционной матрице факторов встречаются элементы, по модулю близкие к 1.
- CN – conditional number (параметр обусловленности матрицы  $X'X$ )

$$CN = \sqrt{\frac{\lambda_{\max}}{\lambda_{\min}}},$$

**Если  $CN > 30$ , то это признак мультиколлинеарности**

# Мультиколлинеарность

**Индикаторы мультиколлинеарности:**

**VIF – variance inflation factor**

$$VIF(X_j) = \frac{1}{1 - R_j^2},$$

где  $R_j^2$  – коэффициент множественной детерминации регрессора  $X_j$  на все остальные регрессоры.

**Если  $VIF > 8$ , то это м.б. признаком мультиколлинеарности**

# Мультиколлинеарность

```
gen SMALE = S* MALE
reg EARNINGS      MALE ETHBLACK ETHHISP S SM ASVAB02 ASVAB03 ASVAB04
SMALE
```

Source	SS	df	MS	Number of obs = 540	F( 9, 530) = 26.15
Model	37064.5295	9	4118.28105	Prob > F = 0.0000	
Residual	83453.8984	530	157.460186	R-squared = 0.3075	
				Adj R-squared = 0.2958	
Total	120518.428	539	223.596341	Root MSE = 12.548	

EARNINGS	Coef.	Std. Err.	t	P>t	[95% Conf. Interval]	
MALE	-17.25304	6.096429	-2.83	0.005	-29.22917	-5.276907
ETHBLACK	-1.710566	1.907612	-0.90	0.370	-5.457974	2.036843
ETHHISP	-3.599198	2.619646	-1.37	0.170	-8.745361	1.546965
S	1.225908	.3390308	3.62	0.000	.5598985	1.891917
SM	.0264917	.2452239	0.11	0.914	-.4552383	.5082217
ASVAB02	.2801644	.0862475	3.25	0.001	.1107355	.4495934
ASVAB03	-.1264856	.0977774	-1.29	0.196	-.3185645	.0655933
ASVAB04	.0889209	.0944876	0.94	0.347	-.0966953	.2745371
SMALE	1.752398	.4381509	4.00	0.000	.8916722	2.613123
_cons	-12.43911	5.105659	-2.44	0.015	-22.46892	-2.409297

Пример оценки зависимости почасовой з/п от длительности обучения, расы и способностей индивида. Коэффициенты при многих переменных оказались незначимыми.

## Мультиколлинеарность

**vif**

<b>Variable</b>	<b>VIF</b>	<b>1/VIF</b>
<b>SMALE</b>	<b>32.59</b>	<b>0.030688</b>
<b>MALE</b>	<b>31.87</b>	<b>0.031382</b>
<b>ASVAB03</b>	<b>2.93</b>	<b>0.341043</b>
<b>ASVAB04</b>	<b>2.84</b>	<b>0.351965</b>
<b>ASVAB02</b>	<b>2.54</b>	<b>0.394438</b>
<b>S</b>	<b>2.43</b>	<b>0.411756</b>
<b>SM</b>	<b>1.41</b>	<b>0.707806</b>
<b>ETHBLACK</b>	<b>1.27</b>	<b>0.788433</b>
<b>ETHHISP</b>	<b>1.16</b>	<b>0.864274</b>
<b>Mean VIF</b>	<b>8.78</b>	

**Vif для MALE, SMALE больше 8, следовательно, может существовать проблема мультиколлинеарности.**



# Мультиколлинеарность

cor (MALE ETHBLACK ETHHISP S SM ASVAB02 ASVAB03 ASVAB04 SMALE)  
(obs=540)

	MALE	ETHBLACK	ETHHISP	S	SM	ASVAB02	ASVAB03	ASVAB04	SMALE
MALE	1.0000								
ETHBLACK	-0.0581	1.0000							
ETHHISP	0.0167	-0.0842	1.0000						
S	-0.0246	-0.1486	-0.0560	1.0000					
SM	0.0318	-0.1094	-0.2995	0.4218	1.0000				
ASVAB02	0.2043	-0.3902	-0.1243	0.4959	0.3318	1.0000			
ASVAB03	0.0783	-0.3839	-0.1548	0.4945	0.3871	0.6869	1.0000		
ASVAB04	-0.0104	-0.3480	-0.1265	0.4560	0.3703	0.6836	0.7626	1.0000	
SMALE	0.9694	-0.0899	0.0178	0.1467	0.1111	0.2897	0.1684	0.0749	1.0000

В корреляционной матрицы есть достаточно близкий к 1 элемент, что свидетельствует о мультиколлинеарности.