

Семинар 13.

Выбор функциональной зависимости.

1. Пусть x — цена мороженого, а y — дневная выручка от продаж мороженого.
Оценивание регрессии вида

$$y_i = \beta_1 + \beta_2 x_i + \beta_3 x_i^2 + \varepsilon_i$$

дало следующий результат:

Dependent Variable: Y				
Method: Least Squares				
Sample: 1 50				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	100.2079	1.967171	50.94010	0.0000
X	10.03677	0.903745	11.10576	0.0000
X2	-0.817382	0.084591	-9.662803	0.0000
R-squared	0.765563	Mean dependent var	123.4150	
Adjusted R-squared	0.755587	S.D. dependent var	8.089480	
S.E. of regression	3.999287	Akaike info criterion	5.668234	
Sum squared resid	751.7320	Schwarz criterion	5.782955	
Log likelihood	-138.7058	Hannan–Quinn criter.	5.711920	
F-statistic	76.74024	Durbin–Watson stat	1.738403	
Prob(F-statistic)	0.000000			

Coefficients covariance matrix			
	C	X	X2
C	3.869764	-1.598561	0.134292
X	-1.598561	0.816755	-0.074654
X2	0.134292	-0.074654	0.007156

Найдите 95%-ый доверительный интервал для значения цены $x = x_0$, при котором выручка максимальна.

2. По данным для 23 демократических стран оценили зависимость индекса Джини от ВВП на душу населения с учетом ППС (паритета покупательной способности). Затем провели тест Рамселя.
- (а) Сформулируйте нулевую и альтернативную гипотезу теста Рамселя.
 - (б) Опишите пошагово, как проводится тест Рамселя.
 - (в) Прокомментируйте результаты теста Рамселя.
3. Для выбора между линейной и полулогарифмической моделями (где $EARNINGS$ — почасовая заработка в \$, S — длительность обучения, $ASVABC$ — результаты тестов, характеризующие успеваемость) был проведен тест Дэвидсона, Уайта и МакКиннона и получены следующие результаты:

```
. reg gini gdp if democ==1
Source          SS           df           MS
Model          506.853501      1   506.853501
Residual       815.572523     21  38.8367868
Total          1322.42602    22  60.1102738
Number of obs = 23
F( 1, 21) = 13.05
Prob > F = 0.0016
R-squared = 0.3833
Adj R-squared = 0.3539
Root MSE = 6.2319

gini          Coef.        Std. Err.         t      P>|t|      [95% Conf. Interval]
gdp          -.0006307    .0001746     -3.61    0.002    -.0009937   -.0002676
_cons        44.30983    3.572733    12.40    0.000    36.87993    51.73974

. ovtest
Ramsey RESET test using powers of the fitted values of gini
Ho: model has no omitted variables
F(3, 18) = 5.16
Prob > F = 0.0095
```

	Зависимая: Y	Зависимая: $\ln Y$
(Intercept)	-26.148 (4.17)	-1.941 (3.2499)
S	2.008 (0.276)	0.087 (0.035)
ASVABC	0.393 (0.079)	0.017 (0.007)
lin_add	-15.373 (5.984)	
semilog_add		-0.029 (0.065)
R^2	0.2071	0.2212
F	46.59	50.74
Adj. R^2	0.2027	0.2168
Num. obs.	540	540
RSS	90975.57	148.1
$\hat{\sigma}$	13.04	0.5256

Здесь $\text{lin_add} = \ln Y - \bar{\ln} Y$, $\text{semilog_add} = Y - \exp(\bar{\ln} Y)$ и в скобках указаны стандартные ошибки. С помощью РЕ-теста на уровне значимости 5% сделайте выбор между двумя оцененными моделями.

4. Для 400 голландских магазинов модной одежды с помощью трёх моделей оценили зависимость продаж в расчете на квадратный метр в гульденах, $Sales$, от:

- общей площади магазина, $Size$, в м^2 ;
- количества сотрудников, работающих целый день, $Nfull$;
- количества временных рабочих, $Ntemp$;
- дамми-переменной $Owner$, равной единице, если собственник один, и нулю иначе.

$$\widehat{Sales}_i = 6083 - 15.25Size_i + 1452.8Nfull_i + 420.15Ntemp_i - 1464.1Owner_i$$

$$\ln \widehat{Sales}_i = 8.59 - 0.0024Size_i + 0.183Nfull_i + 0.102Ntemp_i - 0.209Owner_i$$

$$\ln \widehat{Sales}_i = 10.08 - 0.31 \ln Size_i + 0.22 \ln Nfull_i + 0.066 \ln Ntemp_i - 0.19 \ln Owner_i$$

(0.21) (0.043) (0.061) (0.118) (0.059)

В скобках приведены стандартные ошибки.

- (а) Дайте интерпретацию коэффициента при переменной *Size* в каждой из трёх моделей;
- (б) Подробно опишите, как выбрать наилучшую из этих моделей.
5. С помощью теста Бокса–Кокса оценили зависимость веса индивида (в килограммах) от его роста (в сантиметрах):

							Number of obs	=	540
							LR chi2(2)	=	230.68
							Prob > chi2	=	0.000
Log likelihood = -2659.5656									
W	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]				
/lambda	1.055498	1.892654	0.56	0.577	-2.654035	4.76503			
/theta	-.0263371	.1471576	-0.18	0.858	-.3147607	.2620865			
Estimates of scale-variant parameters									
	Coef.								
Notrans									
_cons	2.936809								
Trans									
H	.0237224								
/sigma	.1660251								
Test	Restricted	chi2	Prob > chi2						
H0:	log likelihood								
theta=lambda = -1	-2680.8693	42.61	0.000						
theta=lambda = 0	-2659.7618	0.39	0.531						
theta=lambda = 1	-2685.5201	51.91	0.000						

Какую спецификацию модели (линейную, линейную в логарифмах, полулогарифмическую) следует предпочесть и почему?