Подборка задач №6.
Спецификация модели. Эндогенность.

1. Покажите, что при пропуске l важных переменных в модели регрессии МНК-оценки параметров β при оставшихся регрессорах и оценка $\hat{\sigma}_\varepsilon^2 = \frac{RSS}{n-k}$ будут смещёнными.

2. Покажите, что при включении l лишних переменных в модель регрессии МНК-оценки параметров β при регрессорах и оценка $\hat{\sigma}_\varepsilon^2 = \frac{RSS}{n-k-l}$ будут несмещёнными.

3. Опишите PE тест МакКиннона, Уайта, Дэвидсона для выбора между линейной и линейной в логарифмах моделями регрессии.

4. Дайте определение эндогенного регрессора.

5. Перечислите 4 причины возникновения эндогенности. Приведите пример для каждого.

6. Перечислите предпосылки линейной модели регрессии со стохастическими регрессорами и сформулируйте теорему о состоятельности и асимптотической нормальности МНК-оценок.

7. Для выбора между линейной и полулогарифмической моделями (где EARNINGS — почасовая заработная плата в \$, S — длительность обучения, ASVABC — результаты тестов, характеризующие успеваемость) был проведен PE тест Дэвидсона, Уайта и МакКиннона и получены следующие результаты:

	Зависимая: Y	Зависимая: $\ln Y$
(Intercept)	-26.148	-1.941
	(4.17)	(3.2499)
S	2.008	0.087
	(0.276)	(0.035)
ASVABC	0.393	0.017
	(0.079)	(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

где $lin_add = \ln(\hat{Y}) - \widehat{\ln Y}$, $semilog_add = \hat{Y} - \exp(\widehat{\ln Y})$ и в скобках указаны стандартные ошибки. На уровне значимости 5% сделайте выбор между двумя оцененными моделями.

8. По данным для 23 демократических стран оценили зависимость индекса Джини от ВВП на душу населения с учетом ППС (паритета покупательной способности). Затем провели тест Рамсея.

```
      Source |       SS       df       MS              Number of obs =      23
-------------+------------------------------           F(  1,    21) =   13.05
       Model |  506.853501     1  506.853501           Prob > F      =  0.0016
    Residual |  815.572523    21  38.8367868           R-squared     =  0.3833
-------------+------------------------------           Adj R-squared =  0.3539
       Total |  1322.42602    22  60.1102738           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
```

Примечание: *ovtest* — команда в Stata для проведения теста Рамсея.

(a) Сформулируйте нулевую и альтернативную гипотезу теста Рамсея.

(b) Опишите пошагово, как проводится тест Рамсея.

(c) Прокомментируйте результаты теста Рамсея.

9. 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
Log likelihood = -2659.5656                       Prob > chi2     =      0.000

------------------------------------------------------------------------------
           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
    H0:            log likelihood       chi2       Prob > chi2
------------------------------------------------------------------------------
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
------------------------------------------------------------------------------
```

Примечание: λ — параметр для преобразования зависимой переменной, а θ — параметр для преобразования объясняющих переменных.

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

10. 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.

$$\widehat{Sales}_i = \underset{(718)}{6083} - \underset{(1.59)}{15.25}Size_i + \underset{(171)}{1452.8}Nfull_i + \underset{(423)}{420.15}Ntemp_i - \underset{(361)}{1464.1}Owner_i$$

$$\ln\widehat{Sales}_i = \underset{(0.11)}{8.59} - \underset{(0.00024)}{0.0024}Size_i + \underset{(0.026)}{0.183}Nfull_i + \underset{(0.066)}{0.102}Ntemp_i - \underset{(0.056)}{0.209}Owner_i$$

$$\ln\widehat{Sales}_i = \underset{(0.21)}{10.08} - \underset{(0.043)}{0.31}\ \ln Size_i + \underset{(0.061)}{0.22}\ \ln Nfull_i + \underset{(0.118)}{0.066}\ln Ntemp_i - \underset{(0.059)}{0.19}\ \ln Owner_i$$

В скобках приведены стандартные ошибки.

В скобках приведены стандартные ошибки.

(a) Дайте интерпретацию коэффициента при переменной Size в каждой из трёх моделей.

(b) Подробно опишите, как выбрать наилучшую из этих моделей.