СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$Y = \beta_1 + \beta_2 X + u$$
$$\log Y = \beta_1 + \beta_2 X + u$$

Когда альтернативные характеристики регрессионной модели имеют одну и ту же зависимую переменную, R^2 можно использовать для сравнения их пригодности.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$Y = \beta_1 + \beta_2 X + u$$

$$\log Y = \beta_1 + \beta_2 X + u$$

Однако, когда зависимая переменная отличается, этого делать нельзя.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$Y = \beta_1 + \beta_2 X + u$$

$$\log Y = \beta_1 + \beta_2 X + u$$

В случае линейной модели, *R^2* измеряет долю дисперсии в *Y*, объясненную моделью. В случае полулогарифмической модели она измеряет долю дисперсии *логарифма Y*, объясненную моделью.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$Y = \beta_1 + \beta_2 X + u$$

$$\log Y = \beta_1 + \beta_2 X + u$$

Очевидно, что они связаны, но они не совпадают, поэтому прямые сравнения бессмысленны.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$Y = \beta_1 + \beta_2 X + u$$

$$\log Y = \beta_1 + \beta_2 X + u$$

Преобразование Бокса–Кокса:

$$\frac{Y^{\lambda} - 1}{\lambda} = \beta_1 + \beta_2 X + u$$

Однако доброкачественность моделей с линейной и логарифмической версиями одной и той же зависимой переменной можно сравнить косвенно, подвергнув зависимую переменную преобразованию Бокса–Кокса и приспосабливая показанную модель.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$Y = \beta_1 + \beta_2 X + u$$

$$\log Y = \beta_1 + \beta_2 X + u$$

Преобразование Бокса–Кокса:

$$\frac{Y^{\lambda} - 1}{\lambda} = \beta_1 + \beta_2 X + u$$

Это семейство спецификаций, зависящих от параметра λ. Определение λ является эмпирическим вопросом, как и определение других параметров.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$Y = \beta_1 + \beta_2 X + u$$

$$\log Y = \beta_1 + \beta_2 X + u$$

Преобразование Бокса–Кокса:

$$\frac{Y^{\lambda} - 1}{\lambda} = \beta_1 + \beta_2 X + u$$

Модель нелинейна по параметрам, поэтому следует использовать метод нелинейной регрессии. На практике используется оценка максимального правдоподобия.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$Y = \beta_1 + \beta_2 X + u$$

$$\log Y = \beta_1 + \beta_2 X + u$$

Преобразование Бокса–Кокса:

$$\frac{Y^{\lambda} - 1}{\lambda} = \beta_1 + \beta_2 X + u$$

$$\frac{Y^{\lambda} - 1}{\lambda} = Y - 1 \quad \text{когда } \lambda = 1$$

$$\frac{Y^{\lambda} - 1}{\lambda} \to \log Y \quad \text{когда } \lambda \to 0$$

Причина, по которой это преобразование представляет интерес в данном контексте, заключается в том, что спецификации с линейными и логарифмическими зависимыми переменными являются частными случаями.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$Y = \beta_1 + \beta_2 X + u$$

$$\log Y = \beta_1 + \beta_2 X + u$$

Преобразование Бокса–Кокса:

$$\frac{Y^{\lambda} - 1}{\lambda} = \beta_1 + \beta_2 X + u$$

$$\frac{Y^{\lambda} - 1}{\lambda} = Y - 1 \quad \text{когда } \lambda = 1$$

$$\frac{Y^{\lambda} - 1}{\lambda} \to \log Y \quad \text{когда } \lambda \to 0$$

Ввод λ = 1 дает линейную модель. Зависимая переменная тогда *Y* – 1, а не *Y*, но вычитание константы из зависимой переменной не влияет на результаты регрессии, за исключением оценки перехвата.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$Y = \beta_1 + \beta_2 X + u$$

$$\log Y = \beta_1 + \beta_2 X + u$$

Преобразование Бокса–Кокса:

$$\frac{Y^\lambda - 1}{\lambda} = \beta_1 + \beta_2 X + u$$

$$\frac{Y^\lambda - 1}{\lambda} = Y - 1 \quad \text{когда } \lambda = 1$$

$$\frac{Y^\lambda - 1}{\lambda} \to \log Y \quad \text{когда } \lambda \to 0$$

Ввод λ = 0 дает (полу)логарифмическую модель. Конечно, нельзя говорить о том, чтобы ввести *l* ровно равным 0, потому что тогда зависимая переменная становится нулевой, деленной на ноль. Речь идет о предельном виде, когда λ стремится к нулю, и мы воспользовались правилом Лопиталя.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$Y = \beta_1 + \beta_2 X + u$$

$$\log Y = \beta_1 + \beta_2 X + u$$

Преобразование Бокса–Кокса:

$$\frac{Y^\lambda - 1}{\lambda} = \beta_1 + \beta_2 X + u$$

$$\frac{Y^\lambda - 1}{\lambda} = Y - 1 \quad \text{когда } \lambda = 1$$

$$\frac{Y^\lambda - 1}{\lambda} \to \log Y \quad \text{когда } \lambda \to 0$$

Таким образом, можно подогнать общую модель и посмотреть, близок ли λ к 0 или близок к 1. Конечно, "близкий" не имеет значения в эконометрике. Чтобы подойти к этому вопросу технически, нужно проверить гипотезы: λ = 0 и λ = 1.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$Y = \beta_1 + \beta_2 X + u$$

$$\log Y = \beta_1 + \beta_2 X + u$$

Преобразование Бокса–Кокса:

$$\frac{Y^\lambda - 1}{\lambda} = \beta_1 + \beta_2 X + u$$

$$\frac{Y^\lambda - 1}{\lambda} = Y - 1 \quad \text{когда } \lambda = 1$$

$$\frac{Y^\lambda - 1}{\lambda} \to \log Y \quad \text{когда } \lambda \to 0$$

Результатом может быть то, что один отклоняется, а другой не отклоняется, но, конечно, возможно, что ни один из них не отклоняется или оба отклоняются, всё зависит от выбранного вами уровня значимости.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$Y = \beta_1 + \beta_2 X + u$$

$$\log Y = \beta_1 + \beta_2 X + u$$

Преобразование Бокса–Кокса:

$$\frac{Y^\lambda - 1}{\lambda} = \beta_1 + \beta_2 X + u$$

$$\frac{Y^\lambda - 1}{\lambda} = Y - 1 \quad \text{когда } \lambda = 1$$

$$\frac{Y^\lambda - 1}{\lambda} \to \log Y \quad \text{когда } \lambda \to 0$$

Если вы заинтересованы только в сравнении соответствий линейных и логарифмических спецификаций, существует короткая процедура, которая включает только стандартные регрессии наименьших квадратов.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$Y = \beta_1 + \beta_2 X + u$$

$$\log Y = \beta_1 + \beta_2 X + u$$

$Y^* = Y$ / среднее геометрическое Y

Первый шаг - разделить наблюдения по зависимой переменной на их среднее геометрическое. Будем называть преобразованную переменную *Y*.*

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$Y = \beta_1 + \beta_2 X + u$$

$$\log Y = \beta_1 + \beta_2 X + u$$

$$Y^* = Y \, / \, \text{среднее геометрическое } Y$$

$$Y^* = \beta_1' + \beta_2' X + u$$

$$\log Y^* = \beta_1' + \beta_2' X + u$$

Теперь вы регрессируете Y^* и $\log_e Y^*$, оставляя правую часть уравнения без изменений. (Параметрам были даны простые оценки, чтобы подчеркнуть, что коэффициенты не будут оценками исходных β_1 и β_2.).

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$Y = \beta_1 + \beta_2 X + u$$

$$\log Y = \beta_1 + \beta_2 X + u$$

$Y^* = Y$ / **среднее геометрическое** Y

$$Y^* = \beta_1' + \beta_2' X + u$$

$$\log Y^* = \beta_1' + \beta_2' X + u$$

Остаточные суммы квадратов теперь прямо сопоставимы. Таким образом, спецификация с меньшим *RSS* обеспечивает лучшую подгонку.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$Y = \beta_1 + \beta_2 X + u$$

$$\log Y = \beta_1 + \beta_2 X + u$$

$Y^* = Y$ / среднее геометрическое Y

$$Y^* = \beta_1' + \beta_2' X + u$$

$$\log Y^* = \beta_1' + \beta_2' X + u$$

Мы будем использовать преобразование для сравнения приступов линейной и полулогарифмической версий простого уравнения заработной платы, используя набор данных *EAWE* Data Set 21.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$e^{\frac{1}{n}\sum \log Y_i}$$

Первый шаг - вычислить среднее геометрическое зависимой переменной. Самый простой способ сделать это - взять экспоненту среднего значения логарифма зависимой переменной.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$e^{\frac{1}{n}\sum \log Y_i} = e^{\frac{1}{n}\log\left(Y_1 Y_2 \ldots Y_n\right)}$$

Сумма логарифмов *Y* равна логарифму произведений *Y*.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$e^{\frac{1}{n}\sum \log Y_i} = e^{\frac{1}{n}\log\left(Y_1 Y_2 \ldots Y_n\right)}$$

$$= e^{\log\left(Y_1 Y_2 \ldots Y_n\right)^{\frac{1}{n}}}$$

Теперь мы используем правило, согласно которому *a*log *X* совпадает с выражением log *X*[*a*]

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$e^{\frac{1}{n}\sum \log Y_i} = e^{\frac{1}{n}\log(Y_1 Y_2 \ldots Y_n)}$$

$$= e^{\log(Y_1 Y_2 \ldots Y_n)^{\frac{1}{n}}} = (Y_1 Y_2 \ldots Y_n)^{\frac{1}{n}}$$

И, наконец, мы используем тот факт, что экспонента логарифма *X* сводится к *X*.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$e^{\frac{1}{n}\sum \log Y_i} = e^{\frac{1}{n}\log(Y_1 Y_2 \ldots Y_n)}$$

$$= e^{\log(Y_1 Y_2 \ldots Y_n)^{\frac{1}{n}}} = (Y_1 Y_2 \ldots Y_n)^{\frac{1}{n}}$$

```
. sum LGEARN

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      LGEARN |       500    2.824265     .553001   .6931472   4.642948
```

LGEARN уже определен как логарифм *EARNINGS*. Мы находим его среднее. В программе Stata это делается с помощью команды «sum».

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$e^{\frac{1}{n}\sum \log Y_i} = e^{\frac{1}{n}\log(Y_1 Y_2 \ldots Y_n)}$$

$$= e^{\log(Y_1 Y_2 \ldots Y_n)^{\frac{1}{n}}} = (Y_1 Y_2 \ldots Y_n)^{\frac{1}{n}}$$

```
. sum LGEARN

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      LGEARN |       500    2.824265     .553001   .6931472   4.642948

. gen EARNSTAR = EARNINGS/exp(2.8243)
```

Затем мы определяем *EARNSTAR*, разделив *EARNINGS* на экспоненту среднего значения *LGEARN.*

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

$$e^{\frac{1}{n}\sum \log Y_i} = e^{\frac{1}{n}\log(Y_1 Y_2 \ldots Y_n)}$$

$$= e^{\log(Y_1 Y_2 \ldots Y_n)^{\frac{1}{n}}} = (Y_1 Y_2 \ldots Y_n)^{\frac{1}{n}}$$

```
. sum LGEARN

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
      LGEARN |       500    2.824265     .553001   .6931472   4.642948

. gen EARNSTAR = EARNINGS/exp(2.8243)

. gen LGEARNST = ln(EARNSTAR)
```

Мы также определяем LGEARNST, логарифм EARNSTAR.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

```
. reg EARNSTAR S EXP

------------------------------------------------------------------------------
      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  2,   497) =   35.24
       Model |  30.7698527     2  15.3849264           Prob > F      =  0.0000
    Residual |  216.958472   497  .436536161           R-squared     =  0.1242
-------------+------------------------------           Adj R-squared =  0.1207
       Total |  247.728325   499  .496449549           Root MSE      =  .66071
------------------------------------------------------------------------------
    EARNSTAR |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
           S |   .1114334   .0132792     8.39   0.000     .0853432    .1375236
         EXP |   .0583614   .0124543     4.69   0.000     .0338918    .0828311
       _cons |  -.8705654    .254515    -3.42   0.001    -1.370623   -.3705073
------------------------------------------------------------------------------
```

Вот регрессия *EARNSTAR* на *S* и *EXP*. Остаточная сумма квадратов равна 217,0.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

```
. reg LGEARNST S EXP

----------------------------------------------------------------------------
      Source |       SS       df       MS              Number of obs =     500
-------------+------------------------------           F(  2,    497) =   40.12
       Model |  21.2104061     2  10.6052031           Prob > F      =  0.0000
    Residual |  131.388814   497  .264363811           R-squared     =  0.1390
-------------+------------------------------           Adj R-squared =  0.1355
       Total |   152.59922   499   .30581006           Root MSE      =  .51416
----------------------------------------------------------------------------
    LGEARNST |      Coef.   Std. Err.      t    P>|t|      [95% Conf. Interval]
-------------+--------------------------------------------------------------
           S |   .0916942   .0103338     8.87   0.000      .0713908    .1119976
         EXP |   .0405521    .009692     4.18   0.000      .0215098    .0595944
       _cons |  -1.624505   .1980634    -8.20   0.000     -2.013649    -1.23536
----------------------------------------------------------------------------
```

Мы запускаем параллельную регрессию для *LGEARNST*. Остаточная сумма квадратов равна 131,4. Таким образом, мы заключаем, что полулогарифмическая версия лучше подходит.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

```
. boxcox EARNINGS S EXP

                                                     Number of obs   =        500
                                                     LR chi2(2)      =      76.08
Log likelihood = -1785.403                           Prob > chi2     =      0.000

------------------------------------------------------------------------------
    EARNINGS |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      /theta |   .1088657     .05362     2.03   0.042      .0037726    .2139589
------------------------------------------------------------------------------

---------------------------------------------------------
   Test         Restricted      LR statistic      P-value
    H0:       log likelihood        chi2       Prob > chi2
---------------------------------------------------------
theta = -1      -2025.7902        480.77           0.000
theta =  0      -1787.4901          4.17           0.041
theta =  1      -1912.8953        254.98           0.000
---------------------------------------------------------

.
```

Вот результат для полной регрессии Бокса-Кокса. Параметр, который мы обозначили λ (lambda), называется «theta» в программе Stata. Он оценивается в 0.11. Поскольку он ближе к 0, чем к 1, он указывает, что зависимая переменная должна быть логарифмической, а не линейной.

СРАВНЕНИЕ ЛИНЕЙНЫХ И ЛОГАРИФМИЧЕСКИХ ХАРАКТЕРИСТИК

```
. boxcox EARNINGS S EXP

                                                  Number of obs   =        500
                                                  LR chi2(2)      =      76.08
Log likelihood = -1785.403                        Prob > chi2     =      0.000

------------------------------------------------------------------------------
    EARNINGS |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      /theta |   .1088657     .05362     2.03   0.042     .0037726    .2139589
------------------------------------------------------------------------------


---------------------------------------------------------
   Test         Restricted      LR statistic      P-value
    H0:       log likelihood        chi2       Prob > chi2
---------------------------------------------------------
theta = -1      -2025.7902       480.77           0.000
theta =  0      -1787.4901         4.17           0.041
theta =  1      -1912.8953       254.98           0.000
---------------------------------------------------------
```

.

Однако даже значение 0 не (полностью) лежит в 95-процентном доверительном интервале. (Тест вероятностей логарифма объясняется в главе 10.).