Lab5: Tables and OLS

Introduction to Econometrics,Spring 2023

Jiayi Cheng

Nanjing University

07/04/2023

Section 1

T-Test and Table

T-Test and Table

- Review the Theory

Hypothesis Test of of $\bar{Y}$

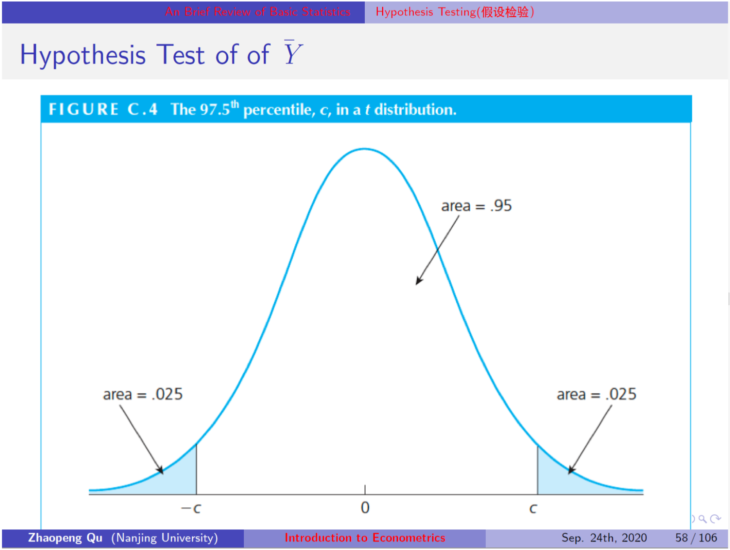
- Specify H_0 and H_1

$$H_0 : E[Y] = \mu_{Y,0} \; H_1 : E[Y] \neq \mu_{Y,0}$$

- Choose the significance level α and define a decision rule (critical region or critical value)
  - eg. if we choose $\alpha = 0.05$, then the critical value is 1.96, then the region is $(-\infty, -1.96]$ and $[1.96, +\infty)$

T-Test and Table

- Review the Theory



T-Test and Table

- Review the Theory

Hypothesis Test of of $\bar{Y}$

- Given the data compute the test statistic
  - Step1: Compute the sample average $\bar{Y}$
  - Step2: Compute the **standard error** of $\bar{Y}$

$$SE(\bar{Y}) = \frac{s_Y}{\sqrt{n}}$$

  - Step3: Compute the **t-statistic**

$$t^{act} = \frac{\bar{Y} - \mu_{Y,0}}{SE\left(\bar{Y}\right)}$$

  - Step4: Reject the null hypothesis if
    - $\mid t^{act} \mid > critical\ value$
    - or if $p - value < significance\ level$

T-Test and Table

- Review the Theory

Hypothesis Tests for the Difference Between Two Means

- To illustrate a test for the difference between two means, let μ_w be the mean hourly earning in the population of women recently graduated from college and let μ_m be the population mean for recently graduated men.
- Then the **null hypothesis** and **the two-sided alternative hypothesis** are

$$\begin{aligned} H_0 : \mu_m &= \mu_w \\ H_1 : \mu_m &\neq \mu_w \end{aligned}$$

- Consider the null hypothesis that mean earnings for these two populations differ by a certain amount, say d_0. The null hypothesis that men and women in these populations have the same mean earnings corresponds to H_0 : $H_0 : d_0 = \mu_m - \mu_w = 0$

T-Test and Table

- Review the Theory

The Difference Between Two Means

- Suppose we have samples of n_m men and n_w women drawn at random from their populations. Let the sample average annual earnings be $\overline{Y}_m$ for men and $\overline{Y}_w$ for women. Then an estimator of $\mu_m - \mu_w$ is $\overline{Y}_m - \overline{Y}_w$.
- Let us discuss the distribution of $\overline{Y}_m - \overline{Y}_w$.

$$\sim N(\mu_m - \mu_w, \frac{\sigma_m^2}{n_m} + \frac{\sigma_w^2}{n_w})$$

- if $\sigma_m^2$and σ_w^2 are known, then the this approximate normal distribution can be used to compute p-values for the test of the null hypothesis. In practice, however, these population variances are typically unknown so they must be estimated.
- Thus the *standard error* of $\overline{Y}_m - \overline{Y}_w$ is

$$SE(\overline{Y}_m - \overline{Y}_w) = \sqrt{\frac{s_m^2}{n_m} + \frac{s_w^2}{n_w}}$$

- Review the Theory

The Difference Between Two Means

- The t-statistic for testing the null hypothesis is constructed analogously to the t-statistic for testing a hypothesis about a single population mean, thus *t-statistic* for comparing two means is

$$t_{act} = \frac{\overline{Y}_m - \overline{Y}_w - d_0}{SE(\overline{Y}_m - \overline{Y}_w)}$$

- If both $n_m$and n_m are large, then this t-statistic has a standard normal distribution when the null hypothesis is true,thus $\overline{Y}_m - \overline{Y}_w = 0$.

T-Test and Table

- Review the Theory

Confidence Intervals for the Difference Between Two Means

- the 95% two-sided confidence interval for d consists of those values of d within ± 1.96 standard errors of $\overline{Y}_m - \overline{Y}_w$, thus $d = \mu_m - \mu_w$ is

$$(\overline{Y}_m - \overline{Y}_w) \pm 1.96 SE(\overline{Y}_m - \overline{Y}_w)$$

T-Test and Table

- 单样本 t 检验

```
. sysuse auto,clear
(1978 Automobile Data)
. ttest price == 6000 if foreign == 0 ,level(90)
One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [90% Conf. Interval]
---------+--------------------------------------------------------------------
   price |      52    6072.423    429.4911    3097.104    5352.903    6791.943
------------------------------------------------------------------------------
    mean = mean(price)                                            t =   0.1686
Ho: mean = 6000                                  degrees of freedom =       51
   Ha: mean < 6000             Ha: mean != 6000               Ha: mean > 6000
 Pr(T < t) = 0.5666         Pr(|T| > |t|) = 0.8668          Pr(T > t) = 0.4334

* level默认95%的水平
* 结果p值大于0.1，不能拒绝H0
```

T-Test and Table

- 独立样本 t 检验
  - 一个变量利用另一个变量来分组比较

```
. sdtest price, by(foreign)
Variance ratio test
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
Domestic |      52    6072.423    429.4911    3097.104    5210.184    6934.662
 Foreign |      22    6384.682    558.9942    2621.915     5222.19    7547.174
---------+--------------------------------------------------------------------
combined |      74    6165.257    342.8719    2949.496    5481.914      6848.6
------------------------------------------------------------------------------
    ratio = sd(Domestic) / sd(Foreign)                            f =   1.3953
Ho: ratio = 1                                  degrees of freedom =   51, 21
    Ha: ratio < 1               Ha: ratio != 1                 Ha: ratio > 1
  Pr(F < f) = 0.7963         2*Pr(F > f) = 0.4073           Pr(F > f) = 0.2037


* 方差齐性检验(F检验)
* 对两个独立样本进行比较的时候，首先要判断两总体方差是否相同，即方差齐性。
* 若两总体方差相等equal variances(方差齐)，则直接用t检验;
* 若方差不齐，选择unequal variances(方差不齐)的均值T检验去做,加unequal选项。
```

T-Test and Table

- 独立样本 t 检验
  - 一个变量利用另一个变量来分组比较

```
. ttest price, by(foreign)
Two-sample t test with equal variances
------------------------------------------------------------------------------
   Group |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
Domestic |      52    6072.423    429.4911    3097.104    5210.184    6934.662
 Foreign |      22    6384.682    558.9942    2621.915     5222.19    7547.174
---------+--------------------------------------------------------------------
combined |      74    6165.257    342.8719    2949.496    5481.914      6848.6
---------+--------------------------------------------------------------------
    diff |           -312.2587    754.4488               -1816.225    1191.708
------------------------------------------------------------------------------
    diff = mean(Domestic) - mean(Foreign)                         t =  -0.4139
Ho: diff = 0                                     degrees of freedom =       72

    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.3401         Pr(|T| > |t|) = 0.6802          Pr(T > t) = 0.6599
```

T-Test and Table

- 独立样本 t 检验
  - 在两个变量间进行比较

```
. webuse fuel,clear
```

```
. sdtest mpg1 == mpg2
Variance ratio test
```

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
mpg1	12	21	.7881701	2.730301	19.26525	22.73475
mpg2	12	22.75	.9384465	3.250874	20.68449	24.81551
combined	24	21.875	.6264476	3.068954	20.57909	23.17091

```
    ratio = sd(mpg1) / sd(mpg2)                                  f =   0.7054
Ho: ratio = 1                                   degrees of freedom =   11, 11

    Ha: ratio < 1               Ha: ratio != 1                 Ha: ratio > 1
  Pr(F < f) = 0.2862         2*Pr(F < f) = 0.5725           Pr(F > f) = 0.7138
```

T-Test and Table

- 独立样本 t 检验
  - 在两个变量间进行比较

```
. ttest mpg1 == mpg2, unpaired
Two-sample t test with equal variances
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    mpg1 |      12          21    .7881701    2.730301    19.26525    22.73475
    mpg2 |      12       22.75    .9384465    3.250874    20.68449    24.81551
---------+--------------------------------------------------------------------
combined |      24      21.875    .6264476    3.068954    20.57909    23.17091
---------+--------------------------------------------------------------------
    diff |               -1.75    1.225518               -4.291568    .7915684
------------------------------------------------------------------------------
    diff = mean(mpg1) - mean(mpg2)                                t =  -1.4280
Ho: diff = 0                                     degrees of freedom =       22
    Ha: diff < 0                 Ha: diff != 0                 Ha: diff > 0
 Pr(T < t) = 0.0837         Pr(|T| > |t|) = 0.1673          Pr(T > t) = 0.9163

* unpaired 表示对两个不同变量检验，不是配对检验
```

T-Test and Table

- 配对样本 t 检验 (单样本 t 检验的扩展)
  - 检验对象是配对样本观测值之差

```
. ttest mpg1==mpg2

Paired t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
    mpg1 |      12          21    .7881701    2.730301    19.26525    22.73475
    mpg2 |      12       22.75    .9384465    3.250874    20.68449    24.81551
---------+--------------------------------------------------------------------
    diff |      12       -1.75    .7797144     2.70101    -3.46614   -.0338602
------------------------------------------------------------------------------
     mean(diff) = mean(mpg1 - mpg2)                               t =  -2.2444
 Ho: mean(diff) = 0                              degrees of freedom =       11

 Ha: mean(diff) < 0           Ha: mean(diff) != 0           Ha: mean(diff) > 0
 Pr(T < t) = 0.0232         Pr(|T| > |t|) = 0.0463          Pr(T > t) = 0.9768


* 没有unpaired选项
* 结果p值小于0.05，拒绝H0
```

T-Test and Table

- -ttest-的局限
  - 每次只能对一个变量进行检验，无法批量对多个变量检验。
  - 汇报结果过于详细，有时我们只需要一个相对精简的结果，如两组各自均值，均值差异，T-Statistic 或者 P-Value。
  - 当待检验变量增加，ttest 命令费时费力。

T-Test and Table

- 多变量均值比较表格输出-ttable2-

```
ssc install ttable2

. sysuse auto,clear
(1978 Automobile Data)
. ttable2 price wei len mpg, by(foreign) f(%6.2f)
```

Variables	G1(Domestic)	Mean1	G2(Foreign)	Mean2	MeanDiff
price	52	6072.42	22	6384.68	-312.26
weight	52	3317.12	22	2315.91	1001.21***
length	52	196.13	22	168.55	27.59***
mpg	52	19.83	22	24.77	-4.95***

T-Test and Table

- 多变量均值比较表格输出-ttable2-

```
. tab rep78
     Repair
Record 1978 |      Freq.     Percent        Cum.
------------+-----------------------------------
          1 |          2        2.90        2.90
          2 |          8       11.59       14.49
          3 |         30       43.48       57.97
          4 |         18       26.09       84.06
          5 |         11       15.94      100.00
------------+-----------------------------------
      Total |         69      100.00
. ttable2 price wei len mpg if rep78==3|rep78==4, by(rep78)
--------------------------------------------------------------------------
Variables   G1(3)          Mean1    G2(4)          Mean2       MeanDiff
--------------------------------------------------------------------------
price         30        6429.233       18       6071.500     357.733
weight        30        3299.000       18       2870.000     429.000*
length        30         194.000       18        184.833       9.167
mpg           30          19.433       18         21.667      -2.233*
--------------------------------------------------------------------------
* 当组类别大于两类时，可以通过指定样本范围进行比较
```

T-Test and Table

- 结果导出-logout-

```
^^Issc install logout
^^I
^^Ilogout, save(ttable) excel replace : ttable2 price ///
^^I        wei len mpg, by(foreign) f(%6.2f)
^^Ilogout, save(ttable) word replace : ttable2 price  ///
^^I        wei len mpg, by(foreign) f(%6.2f)
^^Ilogout, save(ttable) tex replace : ttable2 price   ///
^^I        wei len mpg, by(foreign) f(%6.2f)
```

T-Test and Table

- 结果导出-t2docx-

```
ssc install t2docx

t2docx price weight length mpg  ///
    using ttable1.docx,replace     ///
    by(foreign)                 ///
    title("表1: t检验")
```

T-Test and Table

- 结果导出-esttab-

```
. sysuse auto,clear
(1978 Automobile Data)


. local var price wei len mpg
. qui estpost ttest `var´, by(foreign)
. esttab ., cell("mu_1(fmt(2)) mu_2(fmt(2)) b(star fmt(2)) t(fmt(2))")   ///
>          starlevels(* 0.10 ** 0.05 *** 0.01) replace noobs compress   ///
>          title(esttab_Table: T_test)
esttab_Table: T_test by group
------------------------------------------------------
                 (1)

                mu_1       mu_2          b            t
------------------------------------------------------
price        6072.42    6384.68    -312.26        -0.41
weight       3317.12    2315.91    1001.21***      6.25
length        196.13     168.55      27.59***      5.89
mpg            19.83      24.77      -4.95***     -3.63
------------------------------------------------------
```

T-Test and Table

- 结果导出-esttab-

```
sysuse auto,clear

local var price wei len mpg
qui estpost ttest `var´, by(foreign)
esttab using ttable2.rtf, cell("mu_1(fmt(2)) mu_2(fmt(2)) b(star fmt(2)) t(fmt(2))")  ///
        starlevels(* 0.10 ** 0.05 *** 0.01) replace noobs compress                ///
        title(esttab_Table: T_test)
```

Section 2

Descriptive Statistics Table

Descriptive Statistics Table

- 描述性统计表格导出
  - -logout-

```
logout, save(Desc1) word replace:            ///
tabstat price wei len mpg rep78,             ///
^^I  ^^Istats(mean sd min p50 max) c(s) f(%6.2f)
```

Descriptive Statistics Table

- 描述性统计表格导出
  - -sum2docx-

```
sum2docx price wei len mpg rep78 using Desc2.docx,replace                ///
    stats(N mean(%9.2f) sd(%9.3f) min(%9.2f) median(%9.2f) max(%9.2f)) ///
    title(sum2docx_Table: Descriptive statistics)


*仅sum2docx支持中文,其余命令不支持
*能设置每个统计量的小数点位数
```

Descriptive Statistics Table

- 描述性统计表格导出
  - -outreg2-

```
outreg2 using Desc3, sum(detail) replace word            ^^I     ///
    keep(price wei len mpg rep78) eqkeep(N mean sd min p50 max) ///
    fmt(f) sortvar(wage age grade)                              ///
    title(outreg2_Table: Descriptive statistics)


*若变量里有字符串变量,outreg2命令的处理最智能化:
*会在窗口说明什么变量是字符型，并在报告列表中自动剔除该变量
*支持变量排序
```

Descriptive Statistics Table

- 描述性统计表格导出
  - -esttab-

```
estpost summarize price wei len mpg rep78, detail
esttab using Desc4.rtf,                                                    ///
    cells("count mean(fmt(2)) sd(fmt(2)) min(fmt(2)) p50(fmt(2)) max(fmt(2))") ///
    noobs compress replace title(esttab_Table: Descriptive statistics)


*能设置每个统计量的小数点位数
```

Section 3

Correlation Matrix Table

Correlation Matrix Table

- 相关系数矩阵导出

```
*<方法一> -logout-
logout, save(Corr1) word replace: pwcorr price wei len mpg rep78, star(.05)


*<方法二> -esttab-
estpost correlate  price wei len mpg rep78, matrix
esttab using Corr2.rtf,                                                ///
    unstack not noobs compress nogaps replace star(* 0.1 ** 0.05 *** 0.01) ///
    b(%8.3f) p(%8.3f) title(esttab_Table: correlation coefficient matrix)


*<方法三> -corr2docx-
corr2docx price wei len mpg rep78 using Corr3.docx,                    ///
^^I^^I       replace spearman(ignore) pearson(pw) star                       ///
^^I^^I       title(corr2docx_Table: correlation coefficient matrix)
```

Section 4

OLS Regression-Estimation

Subsection 1

Data Analysis Flow

OLS Regression-Estimation

- Data Analysis Flow
  - Open the data, find the variables, and see the base case.
  - Data Cleaning.
  - Summary Statistics: Figures and Tables.
  - Model Estimation and Hypothesis Testing.
  - Report results, explain and analyze.

Subsection 2

Review the Theory

OLS Regression-Estimation

- Review the Theory

Terminology for Simple Regression Model

- The linear regression model with one regressor is denoted by

$$Y_i = \beta_0 + \beta_1 X_i + u_i$$

- Where
  - Y_i is the **dependent variable**(Test Score)
  - X_i is the **independent variable** or regressor(Class Size or Student-Teacher Ratio)
  - $\beta_0 + \beta_1 X_i$ is the **population regression line** or the **population regression function**

- Review the Theory

Review for the previous lectures

The OLS Estimator

- The estimators of the slope and intercept that *minimize the sum of the squares* of $\hat{u}_i$,thus

$$\underset{b_0,b_1}{arg\,min} \sum_{i=1}^{n} \hat{u}_i^2 = \underset{b_0,b_1}{min} \sum_{i=1}^{n} (Y_i - b_0 - b_1 X_i)^2$$

are called the **ordinary least squares (OLS) estimators** of β_0 and β_1.

OLS estimator of β_1:

$$b_1 = \hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y})}{\sum_{i=1}^{n}(X_i - \overline{X})(X_i - \overline{X})}$$

Zhaopeng Qu (Nanjing University) Lecture 4: Multiple OLS Regression 10/15/2020 5 / 79

OLS Regression-Estimation

- Review the Theory

Least Squares Assumptions

1. Assumption 1: Conditional Mean is Zero
2. Assumption 2: Random Sample
3. Assumption 3: Large outliers are unlikely

- If the 3 least squares assumptions hold the OLS estimators will be
  - **unbiased**
  - **consistent**
  - **normal sampling distribution**

OLS Regression-Estimation

- Review the Theory

Multiple regression model with k regressors

- The multiple regression model is

$$Y_i = \beta_0 + \beta_1 X_{1,i} + \beta_2 X_{2,i} + ... + \beta_k X_{k,i} + u_i, i = 1, ..., n \quad (4.1)$$

where

- Y_i is the **dependent variable**
- $X_1, X_2, ...X_k$ are the **independent variables(includes one is our of interest and some control variables)**
- $\beta_i, j = 1...k$ are slope coefficients on X_i corresponding.
- β_0 is the estimate *intercept*, the value of Y when all $X_j = 0, j = 1...k$
- u_i is the regression *error term*, still all other factors affect outcomes.

OLS Regression-Estimation

- Review the Theory

Multiple Regression: Assumption

- Assumption 1: The conditional distribution of u_i given $X_{1i}, ..., X_{ki}$ has mean zero,thus

$$E[u_i | X_{1i}, ..., X_{ki}] = 0$$

- Assumption 2: $(Y_i, X_{1i}, ..., X_{ki})$ are i.i.d.
- Assumption 3: Large outliers are unlikely.
- Assumption 4: No perfect multicollinearity.

OLS Regression-Estimation

- Review the Theory
  - The OLS estimators $\widehat{\beta}_0, \widehat{\beta}_1 ... \widehat{\beta}_k$ are *unbiased*.
  - The OLS estimators $\widehat{\beta}_0, \widehat{\beta}_1 ... \widehat{\beta}_k$ are *consistent*.
  - The OLS estimators $\widehat{\beta}_0, \widehat{\beta}_1 ... \widehat{\beta}_k$ are *normally distributed* in large samples.

- Multiple OLS estimator

$$\widehat{\beta}_j = \frac{\sum_{i=1}^{n} \tilde{X}_{j,i} Y_i}{\sum_{i=1}^{n} \tilde{X}_{j,i}^2} \; for \; j = 1, 2, .., k$$

Subsection 3

OLS in stata

OLS Regression-Estimation

- 普通最小二乘法 (OLS)

```
. *help  reg
. *regress depvar [indepvars] [if] [in] [weight] [, options]  //因变量, 自变量


. sysuse auto, clear
(1978 Automobile Data)

. reg price weight mpg turn foreign
      Source |       SS           df       MS      Number of obs   =        74
-------------+----------------------------------   F(4, 69)        =     19.23
       Model |   334771309         4  83692827.3   Prob > F        =    0.0000
    Residual |   300294087        69  4352088.22   R-squared       =    0.5271
-------------+----------------------------------   Adj R-squared   =    0.4997
       Total |   635065396        73  8699525.97   Root MSE        =    2086.2

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      weight |   4.284532   .7404967     5.79   0.000     2.807282    5.761783
         mpg |  -.4660076   73.51407    -0.01   0.995    -147.1226    146.1905
        turn |  -229.2059   114.2423    -2.01   0.049    -457.1131   -1.298676
     foreign |   3221.415   706.4847     4.56   0.000     1812.017    4630.813
       _cons |   1368.197   4887.597     0.28   0.780    -8382.292    11118.69
------------------------------------------------------------------------------
```

OLS Regression-Estimation

- 普通最小二乘法 (OLS)

```
. regress weight length, noconstant  //不包括截距项（constant）

      Source |       SS           df       MS      Number of obs   =        74
-------------+----------------------------------   F(1, 73)        =   3450.13
       Model |   703869302         1   703869302   Prob > F        =    0.0000
    Residual |  14892897.8        73  204012.299   R-squared       =    0.9793
-------------+----------------------------------   Adj R-squared   =    0.9790
       Total |   718762200        74   9713002.7   Root MSE        =    451.68

------------------------------------------------------------------------------
      weight |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      length |   16.29829   .2774752    58.74   0.000     15.74528     16.8513
------------------------------------------------------------------------------
```

OLS Regression-Estimation

- 普通最小二乘法 (OLS)

```
. reg price weight mpg turn foreign, robust  //稳健标准误 (robust)
Linear regression                                Number of obs     =         74
                                                 F(4, 69)          =      12.46
                                                 Prob > F          =     0.0000
                                                 R-squared         =     0.5271
                                                 Root MSE          =     2086.2

------------------------------------------------------------------------------
             |               Robust
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      weight |   4.284532   .9164881     4.67   0.000     2.456188    6.112876
         mpg |  -.4660076   84.34373    -0.01   0.996    -168.7271    167.7951
        turn |  -229.2059   136.4962    -1.68   0.098    -501.5084    43.09658
     foreign |   3221.415   690.7001     4.66   0.000     1843.506    4599.324
       _cons |   1368.197   6008.419     0.23   0.821    -10618.27    13354.66
------------------------------------------------------------------------------
```

- 回归结果

```
. regress price mpg weight foreign
      Source |       SS           df       MS      Number of obs   =        74
-------------+----------------------------------   F(3, 70)        =     23.29
       Model |   317252881         3   105750960   Prob > F        =    0.0000
    Residual |   317812515        70  4540178.78   R-squared       =    0.4996
-------------+----------------------------------   Adj R-squared   =    0.4781
       Total |   635065396        73  8699525.97   Root MSE        =    2130.8

------------------------------------------------------------------------------
       price |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         mpg |    21.8536   74.22114     0.29   0.769    -126.1758     169.883
      weight |   3.464706    .630749     5.49   0.000     2.206717    4.722695
     foreign |    3673.06   683.9783     5.37   0.000     2308.909    5037.212
       _cons |  -5853.696   3376.987    -1.73   0.087    -12588.88    881.4934
------------------------------------------------------------------------------
```

OLS Regression-Estimation

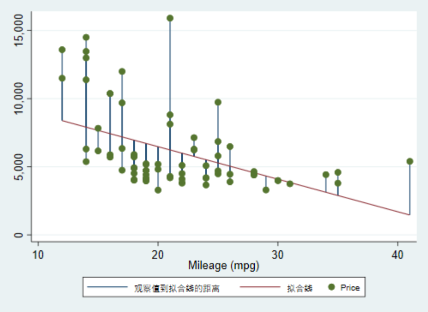
- 回归结果

```
  predict yhat, xb        //price的拟合值
  predict e, residual     //残差
  vce                     //获取变量的方差—协方差矩阵

. test mpg = 20               //单变量检验
 ( 1)  mpg = 20
       F(  1,    70) =    0.00
            Prob > F =    0.9801
. test mpg weight foreign  //联合检验
 ( 1)  mpg = 0
 ( 2)  weight = 0
 ( 3)  foreign = 0
       F(  3,    70) =   23.29
            Prob > F =    0.0000
```

OLS Regression-Estimation

- 回归结果

```
. qui reg price mpg
. predict yhat_p, xb
(option xb assumed; fitted values)
. twoway (rspike price yhat_p mpg )    ///
>        (lfit price mpg)              ///
>        (scatter price mpg ),         ///
>        legend(label(1 "观察值到拟合线的距离") label(2 "拟合线") row(1) size(small))

. graph export olsf.png, width(500) replace
(note: file olsf.png not found)
(file olsf.png written in PNG format)
```



Subsection 4

OLS Result Table

OLS Regression-Estimation

- 回归结果输出-esttab-
  - word 文档

```
sysuse nlsw88, clear

reg wage age married occupation
est store m1
reg wage age married collgrad occupation
est store m2
xi: reg wage age married collgrad occupation i.race
est store m3

esttab m1 m2 m3 using ols.rtf, scalar(r2 r2_a N F) compress  ///
    star(* 0.1 ** 0.05 *** 0.01)                              ///
    b(%6.3f) t(%6.3f) r2(%9.3f) ar2                           ///
    mtitles("OLS-1" "OLS-2" "OLS-3")                          ///
    title(esttab_Table: regression result)
```

OLS Regression-Estimation

- 回归结果输出-esttab-
  - Tex 文档

```
    esttab m1 m2 m3 using ols.tex, replace          ///
        star( * 0.10 ** 0.05 *** 0.01 ) compress    ///
        b(%6.3f) t(%6.3f) r2(%9.3f) ar2             ///
        mtitles("OLS-1" "OLS-2" "OLS-3")            ///
        title(esttab_Table: regression result)      ///
        booktabs page width(\hsize)

/*
esttab 的 LaTeX 输出的专有选项:
1. booktabs: 用 booktabs 宏包输出表格(三线表格)。
2. page[(packages)]: 创建完成的 LaTeX 文档以及添加括号里的宏包
3. 如果写了 booktabs 选项, 则 page[(packages)] 将自动添加\usepackagebooktabs。
4. alignment(cccccc): 定义从第二列开始的列对齐方式(默认居中)。
5. width(\hsize): 可以使得表格宽度为延伸至页面宽度
6. fragment: 不输出表头表尾, 只输出表格本身内容, 其不能与 page[(packages)] 选项共存。
*/
```

OLS Regression-Estimation

表 1: esttab_Table: regression result

	(1) OLS-1	(2) OLS-2	(3) OLS-3
age	-0.064	-0.059	-0.067*
	(-1.637)	(-1.579)	(-1.796)
married	-0.469*	-0.472**	-0.629**
	(-1.873)	(-1.983)	(-2.578)
occupation	-0.284***	-0.384***	-0.370***
	(-8.055)	(-11.251)	(-10.756)
collgrad		4.220***	4.133***
		(15.444)	(15.051)
_Irace_2			-0.784***
			(-2.897)
_Irace_3			-0.224
			(-0.210)
_cons	11.910***	11.168***	11.753***
	(7.654)	(7.545)	(7.878)
N	2237	2237	2237
R^2	0.031	0.125	0.128
adj. R^2	0.030	0.123	0.126

t statistics in parentheses
* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

Section 5

Nonlinear Regression

Subsection 1

Review the Theory

Nonlinear Regression

- Review the Theory

Nonlinear Regression Regression Functions

- So far our regression model is

$$Y_i = \beta_0 + \beta_1 X_{1,i} + ... + \beta_k X_{k,i} + u_i$$

- The effect of Y on a change in X_j by 1 (unit) is constant and equals β_j:

$$\beta_j = \frac{\partial Y_i}{\partial X_{ji}}$$

- But if a relation between Y and X is nonlinear:
  - The effect on Y of a change in X depends on the value of X – that is, the *marginal effect of X is not constant*.
  - A linear regression is misspecified – the functional form is wrong.
  - The estimator of the effect on Y of X is biased(a special case of OVB).
- The solution to this is to estimate a regression function that is nonlinear in X.

Nonlinear Regression

- Review the Theory

Nonlinear in Xs

Two Cases:

1. The effect of change in X_1 on Y depends on X_1 itself.
   - eg. the effect of a change in class size on test scores is bigger when initial class size is small.
2. The effect of change in X_1 on Y depends on another variable, like X_2
   - eg.the effect of class size on test scores depends on the percentage of disadvantaged pupils in the class.

Zhaopeng Qu (Nanjing University) Lecture 6: Nonlinear Regression Functions 10/29/2020 12 / 76

- Review the Theory

Two Complementary Approaches:

1. **Polynomials in X**
   - The population regression function is approximated by a quadratic, cubic, or higher-degree polynomial.
2. **Logarithmic transformations**
   - Y and/or X is transformed by taking its logarithm
   - this gives a *percentages* interpretation that makes sense in many applications

Nonlinear Regression

- Review the Theory

Introduction

- The product of two variables is called an **interaction term**.
- Try to answer *how the effect on Y of a change in an independent variable depends on the value of another independent variable.*
- Consider three cases:

1. Interactions between two binary variables.
2. Interactions between a binary and a continuous variable.
3. Interactions between two continuous variables.

Subsection 2

In practice

Nonlinear Regression

- In practice

```
. use caschool.dta, clear
```

```
. ***** 数据转化 ************

. gen avginc2 = avginc^2    //avginc平方项
. gen avginc3 = avginc^3    //avginc立方项

. gen loginc = ln(avginc)   //avginc对数
. gen logtest = ln(testscr) //testscr对数

. gen loginc2 = loginc^2    //avginc对数平方项
. gen loginc3 = loginc^3    //avginc对数立方项

. gen histr = (str>=20)     //histr = 1 if the str > 20
. gen hiel = (el_pct >= 10) //hiel = 1 if the el_pct > 10%
```

Nonlinear Regression

- In practice

```
. gen hisxhie = histr*hiel  //D1i × D2i
. gen strxhiel = str*hiel      //Xi × Di
. gen strxelpc = str*el_pct //X1i × X2i

. gen sttr2 = str^2         //str平方项
. gen sttr3 = str^3           //str立方项

. gen str2hiel = sttr2*hiel //Xi × Di
. gen str3hiel = sttr3*hiel //Xi × Di
```

Nonlinear Regression

- In practice

$$\text{Test Score } = \beta_0 + \beta_1 STR + \beta_2 \text{ English } + \beta_3 \text{ Lunch } + u$$

```
. reg testscr str el_pct meal_pct, r    //多元线性回归模型
Linear regression                               Number of obs     =        420
                                                F(3, 416)         =     453.48
                                                Prob > F          =     0.0000
                                                R-squared         =     0.7745
                                                Root MSE          =     9.0801
```

testscr	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
str	-.9983092	.2700799	-3.70	0.000	-1.529201	-.4674178
el_pct	-.1215733	.0328317	-3.70	0.000	-.18611	-.0570366
meal_pct	-.5473456	.0241072	-22.70	0.000	-.5947328	-.4999583
_cons	700.15	5.56845	125.74	0.000	689.2042	711.0958

```
. est store m1
```

Nonlinear Regression

- In practice

$$\text{Test Score } = \beta_0 + \beta_1 STR + \beta_2 \text{ English } + \beta_3 \text{ Lunch } + \beta_4 \ln(\text{Income}) + u$$

```
. reg testscr str el_pct meal_pct loginc, r //控制对数收入的影响
Linear regression                               Number of obs     =        420
                                                F(4, 415)         =     417.20
                                                Prob > F          =     0.0000
                                                R-squared         =     0.7962
                                                Root MSE          =     8.6426

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |   -.734326   .2567803    -2.86   0.004    -1.239078   -.2295738
      el_pct |  -.1755344   .0336606    -5.21   0.000     -.241701   -.1093678
    meal_pct |  -.3982342   .0331741   -12.00   0.000    -.4634443    -.333024
      loginc |   11.56897   1.818811     6.36   0.000     7.993736     15.1442
       _cons |    658.552   8.641528    76.21   0.000     641.5653    675.5386
------------------------------------------------------------------------------

. est store m2
```

Nonlinear Regression

- In practice

$$\text{Test Score} = \beta_0 + \beta_1 \text{ STR } + \beta_2 \text{ HiEL} + \beta_3 (STR \times HiEL) + u$$

```
. reg testscr str hiel strxhiel, r //考虑生师比与英语学习者百分比的交互影响
Linear regression                               Number of obs     =        420
                                                F(3, 416)         =      63.67
                                                Prob > F          =     0.0000
                                                R-squared         =     0.3103
                                                Root MSE          =      15.88

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -.9684601   .5891016    -1.64   0.101    -2.126447    .1895268
        hiel |   5.639141   19.51456     0.29   0.773    -32.72029    43.99857
    strxhiel |  -1.276613   .9669194    -1.32   0.187     -3.17727    .6240436
       _cons |   682.2458   11.86781    57.49   0.000     658.9175    705.5742
------------------------------------------------------------------------------

. est store m3
```

Nonlinear Regression

- In practice

```
. test str strxhiel //联合假设的F 统计量和P 值
 ( 1)  str = 0
 ( 2)  strxhiel = 0
       F(  2,   416) =    5.64
            Prob > F =    0.0038
```

Nonlinear Regression

- In practice

$$\text{Test Score} = \beta_0 + \beta_1 \text{ STR } + \beta_2 \text{HiEL} + \beta_3 (STR \times HiEL)$$
$$+ \beta_4 \text{ Lunch } + \beta_5 \ln(\text{Income}) + u$$

```
. reg testscr str hiel strxhiel meal_pct loginc, r //控制学生特征的基础上考虑交互影响
Linear regression                               Number of obs     =        420
                                                F(5, 414)         =     335.85
                                                Prob > F          =     0.0000
                                                R-squared         =     0.7974
                                                Root MSE          =     8.6286

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |  -.5310319   .3418468    -1.55   0.121    -1.203004      .14094
        hiel |   5.498208      9.795     0.56   0.575    -13.75593    24.75234
    strxhiel |  -.5776664   .4957779    -1.17   0.245    -1.552222    .3968896
    meal_pct |  -.4113776   .0288276   -14.27   0.000    -.4680443   -.3547109
      loginc |   12.12447   1.797513     6.75   0.000     8.591078    15.65786
       _cons |   653.6661   9.869378    66.23   0.000     634.2658    673.0665
------------------------------------------------------------------------------

. est store m4
```

Nonlinear Regression

- In practice

```
. test str strxhiel //联合假设的F统计量和P值
 ( 1)  str = 0
 ( 2)  strxhiel = 0
       F(  2,   414) =    5.92
            Prob > F =    0.0029
```

Nonlinear Regression

- In practice

$$\text{TestScore} = \beta_0 + \beta_1 \times \text{STR} + \beta_2 \times \text{STR}^2 + \beta_3 \times \text{STR}^3 \\ + \beta_4\ \text{HiEL} + \beta_5\ \text{Lunch}\ + \beta_6 \ln(\text{Income}) + u$$

```
. reg testscr str sttr2 sttr3 hiel meal_pct loginc, r //考虑生师比的非线性影响
Linear regression                               Number of obs     =        420
                                                F(6, 413)         =     281.14
                                                Prob > F          =     0.0000
                                                R-squared         =     0.8011
                                                Root MSE          =     8.5593

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |   64.33964   24.86053     2.59   0.010     15.47069    113.2086
       sttr2 |  -3.423925   1.249901    -2.74   0.006    -5.880886   -.9669636
       sttr3 |   .0592895    .020763     2.86   0.005     .0184752    .1001037
        hiel |  -5.473991   1.033762    -5.30   0.000    -7.506081     -3.4419
    meal_pct |  -.4200571   .0285258   -14.73   0.000    -.4761309   -.3639833
      loginc |   11.74819   1.771437     6.63   0.000     8.266032    15.23035
       _cons |   252.0458   163.6341     1.54   0.124    -69.61384    573.7055
------------------------------------------------------------------------------

. est store m5
```

Nonlinear Regression

- In practice

```
. test str sttr2 sttr3   //联合假设的F统计量和P值
 ( 1)  str = 0
 ( 2)  sttr2 = 0
 ( 3)  sttr3 = 0
       F(  3,   413) =    6.31
            Prob > F =    0.0003
. test sttr2 sttr3       //联合假设的F统计量和P值
 ( 1)  sttr2 = 0
 ( 2)  sttr3 = 0
       F(  2,   413) =    6.17
            Prob > F =    0.0023
```

- In practice

$$\text{TestScore} = \beta_0 + \beta_1 \times \text{STR} + \beta_2 \times \text{STR}^2 + \beta_3 \times \text{STR}^3$$
$$+\beta_4 \text{ HiEL} + \beta_5 (STR \times HiEL) + \beta_6 \left(STR^2 \times HiEL\right) +$$
$$\beta_7 \left(STR^3 \times HiEL\right) + \beta_8 \text{ Lunch } + \beta_9 \ln(\text{Income}) + u$$

```
. reg testscr str sttr2 sttr3 hiel strxhiel str2hiel str3hiel meal_pct loginc, r
//考虑所有因素
Linear regression                               Number of obs     =        420
                                                F(9, 410)         =     199.88
                                                Prob > F          =     0.0000
                                                R-squared         =     0.8031
                                                Root MSE          =     8.5475

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |   83.70272    28.4967     2.94   0.003     27.68485    139.7206
       sttr2 |  -4.380841   1.441017    -3.04   0.003    -7.213544   -1.548139
       sttr3 |   .0749236    .024008     3.12   0.002     .0277295    .1221178
        hiel |   816.0904   327.6745     2.49   0.013     171.9587    1460.222
    strxhiel |  -123.2842   50.21281    -2.46   0.014    -221.9909   -24.57757
    str2hiel |    6.12124   2.541978     2.41   0.016     1.124303    11.11818
    str3hiel |    -.1006    .0425094    -2.37   0.018    -.1841636   -.0170364
    meal_pct |  -.4177876   .0287011   -14.56   0.000    -.4742073   -.3613679
      loginc |   11.80035   1.778008     6.64   0.000     8.305197     15.2955
       _cons |   122.3464    185.518     0.66   0.510    -242.3388    487.0315
------------------------------------------------------------------------------

. est store m6
```

Nonlinear Regression

- In practice

```
. test str sttr2 sttr3 strxhiel str2hiel str3hiel //联合假设的F 统计量和P 值
 ( 1)  str = 0
 ( 2)  sttr2 = 0
 ( 3)  sttr3 = 0
 ( 4)  strxhiel = 0
 ( 5)  str2hiel = 0
 ( 6)  str3hiel = 0
       F(  6,   410) =    4.96
            Prob > F =    0.0001
```

Nonlinear Regression

- In practice

```
. test sttr2 sttr3 //联合假设的F 统计量和P 值
 ( 1)  sttr2 = 0
 ( 2)  sttr3 = 0
       F(  2,   410) =    5.81
            Prob > F =    0.0033
```

Nonlinear Regression

- In practice

```
. test strxhiel str2hiel str3hiel //联合假设的F 统计量和P 值
 ( 1)  strxhiel = 0
 ( 2)  str2hiel = 0
 ( 3)  str3hiel = 0
       F(  3,   410) =    2.69
            Prob > F =    0.0460
```

Nonlinear Regression

- In practice

$$\text{TestScore} = \beta_0 + \beta_1 \times \text{STR} + \beta_2 \times \text{STR}^2 + \beta_3 \times \text{STR}^3$$
$$+\beta_4 \text{ English } + \beta_5 \text{ Lunch } + \beta_6 \ln(\text{Income}) + u$$

```
. reg testscr str sttr2 sttr3 el_pct meal_pct loginc, r //除交互影响外所有因素
Linear regression                               Number of obs     =        420
                                                F(6, 413)         =     280.81
                                                Prob > F          =     0.0000
                                                R-squared         =     0.8007
                                                Root MSE          =     8.5679

------------------------------------------------------------------------------
             |               Robust
     testscr |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         str |   65.28595   25.25864     2.58   0.010     15.63443    114.9375
       sttr2 |  -3.465567   1.270893    -2.73   0.007    -5.963793   -.9673414
       sttr3 |    .059909   .0211205     2.84   0.005     .0183919    .1014262
      el_pct |   -.165687   .0343657    -4.82   0.000    -.2332405   -.0981334
    meal_pct |  -.4024177   .0332667   -12.10   0.000    -.4678108   -.3370246
      loginc |   11.50894   1.806403     6.37   0.000     7.958044    15.05983
       _cons |   244.8025   165.7221     1.48   0.140     -80.9614    570.5665
------------------------------------------------------------------------------

. est store m7
```

Nonlinear Regression

- In practice

```
. test str sttr2 sttr3 //联合假设的F 统计量和P 值
 ( 1)  str = 0
 ( 2)  sttr2 = 0
 ( 3)  sttr3 = 0
       F(  3,   413) =    5.91
            Prob > F =    0.0006
```

Nonlinear Regression

- In practice

```
. test sttr2 sttr3 //联合假设的F 统计量和P 值
 ( 1)  sttr2 = 0
 ( 2)  sttr3 = 0
       F(  2,   413) =    5.96
            Prob > F =    0.0028
```

Nonlinear Regression

- In practice

```
*导出到Latex（或Word）
. esttab m* using nl_ols.tex, replace                          ///
      star( * 0.10 ** 0.05 *** 0.01 ) compress                 ///
      b(%6.3f) t(%6.3f) r2(%9.3f) ar2                        ///
      booktabs page width(\\hsize)
(output written to nl_ols.tex)
```

Nonlinear Regression

- In practice

	(1) testscr	(2) testscr	(3) testscr	(4) testscr	(5) testscr	(6) testscr	(7) testscr
str	-0.998***	-0.734***	-0.968	-0.531	64.340***	83.703***	65.286**
	(-3.696)	(-2.860)	(-1.644)	(-1.553)	(2.588)	(2.937)	(2.585)
el_pct	-0.122***	-0.176***					-0.166***
	(-3.703)	(-5.215)					(-4.821)
meal_pct	-0.547***	-0.398***		-0.411***	-0.420***	-0.418***	-0.402***
	(-22.705)	(-12.004)		(-14.270)	(-14.726)	(-14.556)	(-12.097)
loginc		11.569***		12.124***	11.748***	11.800***	11.509***
		(6.361)		(6.745)	(6.632)	(6.637)	(6.371)
hiel			5.639	5.498	-5.474***	816.090**	
			(0.289)	(0.561)	(-5.295)	(2.491)	
strxhiel			-1.277	-0.578		-123.284**	
			(-1.320)	(-1.165)		(-2.455)	
sttr2					-3.424***	-4.381***	-3.466***
					(-2.739)	(-3.040)	(-2.727)
sttr3					0.059***	0.075***	0.060***
					(2.856)	(3.121)	(2.837)
str2hiel						6.121**	
						(2.408)	
str3hiel						-0.101**	
						(-2.367)	
_cons	700.150***	658.552***	682.246***	653.666***	252.046	122.346	244.803
	(125.735)	(76.208)	(57.487)	(66.232)	(1.540)	(0.659)	(1.477)
N	420	420	420	420	420	420	420
R^2	0.775	0.796	0.310	0.797	0.801	0.803	0.801
adj. R^2	0.773	0.794	0.305	0.795	0.798	0.799	0.798

Section 6

Nonlinear Regression (continued)

Subsection 1

Dummy Variables

- Dummy Variables

```
. use Nations2.dta,clear
(UN Human Development Indicators)
. des
Contains data from Nations2.dta
  obs:           194                          UN Human Development Indicators
 vars:            13                          2 Jul 2012 06:11
------------------------------------------------------------------------------------------
              storage   display    value
variable name   type    format     label      variable label
------------------------------------------------------------------------------------------
country         str21   %21s                  Country
region          byte    %8.0g      region     Region
gdp             float   %9.0g                 Gross domestic product per cap 2005$, 200
school          float   %9.0g                 Mean years schooling (adults) 2005/2010
adfert          float   %8.0g                 Adolescent fertility: births/1000 fem 15-
                                                2010
chldmort        float   %9.0g                 Prob dying before age 5/1000 live births
                                                2005/2009
life            float   %9.0g                 Life expectancy at birth 2005/2010
pop             float   %9.0g                 Population 2005/2010
urban           float   %9.0g                 Percent population urban 2005/2010
femlab          float   %9.0g                 Female/male ratio in labor force 2005/200
literacy        float   %9.0g                 Adult literacy rate 2005/2009
co2             float   %9.0g                 Tons of CO2 emitted per cap 2005/2006
gini            float   %9.0g                 Gini coef income inequality 2005/2009
------------------------------------------------------------------------------------------
Sorted by: region  country
```

Nonlinear Regression (continued)

- Dummy Variables

```
. tab region,gen(reg)  //为每一类创建虚拟变量
     Region |      Freq.     Percent        Cum.
------------+-----------------------------------
     Africa |         52       26.80       26.80
   Americas |         35       18.04       44.85
       Asia |         49       25.26       70.10
     Europe |         43       22.16       92.27
    Oceania |         15        7.73      100.00
------------+-----------------------------------
      Total |        194      100.00
. gen loggdp=ln(gdp)
(15 missing values generated)


. label values reg1 reg1
. label define reg1 0 "others" 1 "Africa" //reg1 = 1 (Africa); = 0 (elsewhere)
```

Nonlinear Regression (continued)

- Dummy Variables

```
. eststo : qui reg life reg1
. eststo : qui reg life loggdp chldmort
. eststo : qui reg life reg1 reg2 reg3 reg4 loggdp chldmort
. eststo : qui reg life reg1 reg2 reg4 loggdp chldmort
. eststo : qui reg life reg1 loggdp chldmort


. esttab using e2.tex, replace nonumbers            ///
      title(Dummy_Var_Example Table)                ///
      mtitles(m1 m2 m3 m4)                          ///
      star( * 0.10 ** 0.05 *** 0.01 ) compress      ///
      b(%6.3f) t(%6.3f) r2(%9.3f)                   ///
      booktabs page
(output written to e2.tex)
```

Nonlinear Regression (continued)

- Dummy Variables

Table 1: Dummy_Var_Example Table

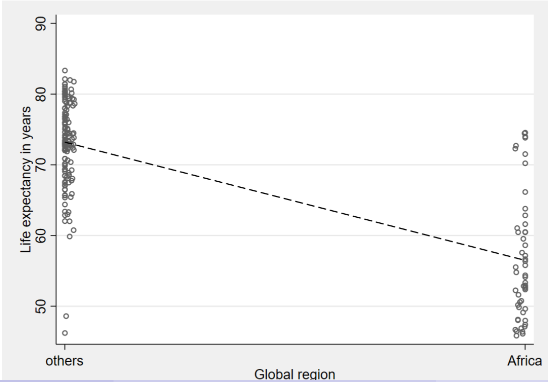
	m1	m2	m3	m4	est5
reg1	-16.721***		-2.151*	-2.927***	-3.144***
	(-15.175)		(-1.775)	(-3.570)	(-3.979)
loggdp		1.525***	1.407***	1.422***	1.483***
		(4.834)	(4.412)	(4.468)	(4.894)
chldmort		-0.146***	-0.127***	-0.127***	-0.128***
		(-19.041)	(-14.605)	(-14.610)	(-14.781)
reg2			1.487	0.692	
			(1.264)	(0.933)	
reg3			0.984		
			(0.871)		
reg4			1.455	0.649	
			(1.213)	(0.852)	
_cons	73.211***	62.286***	62.162***	62.821***	62.657***
	(128.333)	(20.464)	(20.043)	(20.901)	(21.430)
N	194	178	178	178	178
R^2	0.545	0.880	0.891	0.891	0.890

t statistics in parentheses

* $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$

- Dummy Variables

```
. predict lifehat
. graph twoway scatter life  reg1, msymbol(oh S) jitter(5)           ///
      || lfit life reg1                                              ///
      || , legend(off) xlabel(0 "others" 1 "Africa")  scheme(s2mono)  ///
      xtitle("Global region") ytitle("Life expectancy in years")
. cap graph export nat1.png, replace
```



Nonlinear Regression (continued)

- Dummy Variables
  - Explanation:

m1 : The difference between the two means equals –16.72 years.

m2 : **loggpd** and **chldmort**(child mortality rate) together explain about 88% of the variance in life expectancy(m2). Including four dummy variables for regions 1–4 raises this only to about 89% (m3).

m3 : It is not possible to include all five in one regression because of **multicollinearity**. None of the regional dummy variables have significant effects.

m4 : Dropping reg3, the weakest of these predictors.The coefficient on reg1 now appears significant.

m5 : A reduced model.

Conclusion : The differences in life expectancy among **other regions** of the world are largely accounted for by variations in **wealth** and **child mortality**, but in **Africa** there are circumstances at work (such as **wars**) that further depress life expectancy.

Subsection 2

Interaction Effects

Nonlinear Regression (continued)

- Interaction Effects
  - consider some different variables:

    per capita carbon dioxide emissions (co2)

    percent of the population living in urban areas (urban)

    dummy variable reg4 defined as 1 for European countries and 0 for all others

    form an interaction term named urb_reg4 by multiplying the dummy variable reg4 times the measurement variable urban

Nonlinear Regression (continued)

- Interaction Effects

```
. label val reg4 reg4
. label define reg4 0 "others" 1 "Europe"
. gen logco2 = log10(co2)
(9 missing values generated)
. label var logco2 "log10(per cap CO2)"


. gen urb_reg4 = urban * reg4
. label variable urb_reg4 "interaction urban*reg4 (Europe)"
```

Nonlinear Regression (continued)

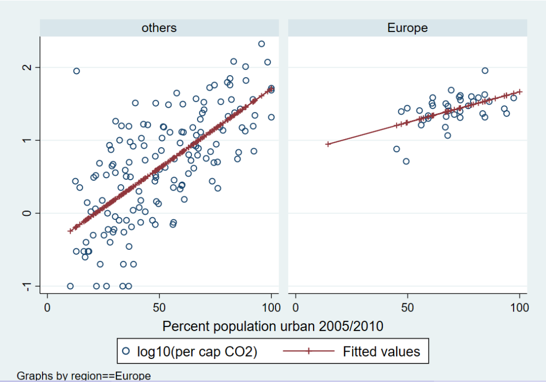
- Interaction Effects

```
. qui reg logco2 urban reg4 urb_reg4
. qui reg logco2 c.urban i.reg4 c.urban#i.reg4
. qui reg logco2 c.urban##i.reg4


 * factor-variable :
 * i. indicator variables
 * c. continuous variables
 * #  an interaction between two variables
 * ## factorial interaction which automatically includes all the lower-level
interactions involving those variables
```

Nonlinear Regression (continued)

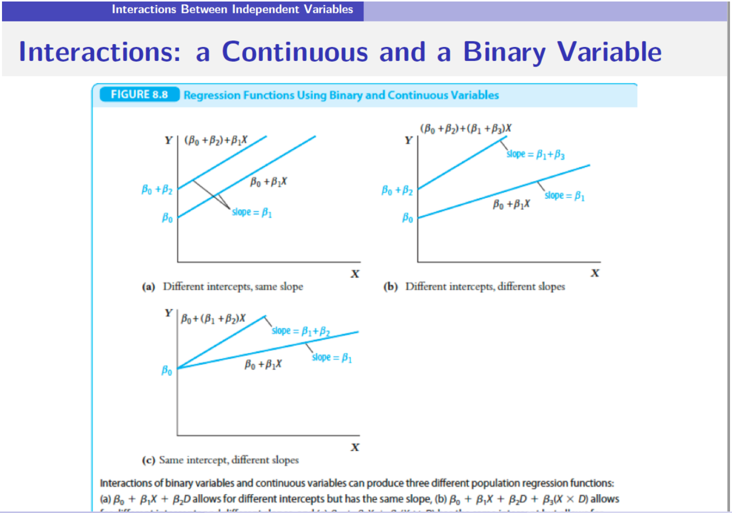
- Interaction Effects

```
. predict co2hat
(option xb assumed; fitted values)
. graph twoway scatter logco2 urban, msymbol(Oh)     ///
      || connect co2hat urban, msymbol(+)            ///
      || , by(reg4)
. cap graph export nat2.png, replace
```



Nonlinear Regression (continued)

- Interaction Effects
  - Explanation:



Nonlinear Regression (continued)

- Interaction Effects
  - Explanation:

    The line in the left-hand (reg4 $= 0$) panel has a slope of .0217 and y-intercept $-.4682$.

    The line in the right panel (reg4 $= 1$) has a less-steep slope (.0084) and a higher y-intercept (.826).

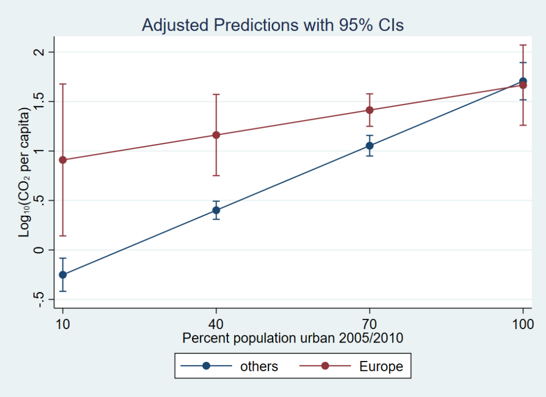
    No European countries exhibit the low-urbanization, low-CO2 profile seen in other parts of the world.

    Even European nations with middling urbanization have relatively high CO2 emissions.

Nonlinear Regression (continued)

- Interaction Effects

```
. qui margins, at(urban = (10(30)100) reg4 = (0 1)) vsquish
. marginsplot, ytitle("Log{subscript:10}(CO{subscript:2} per capita)") xlabel(10(30)10(

. cap graph export nat3.png, replace
```



Nonlinear Regression (continued)

- Interaction Effects
  - Interactions Between Two Continuous Variables
  - **centering** makes their main effects easier to interpret
  - Interacting variables have been centered, can be interpreted as the effect of each variable when the other is at its mean.

```
. summarize urban loggdp
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       urban |        194    55.43488     23.4391      10.25        100
      loggdp |        179    8.693936    1.297024   5.634075   11.22399
```

Nonlinear Regression (continued)

- Interaction Effects

```
. sum urban
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       urban |        194    55.43488     23.4391      10.25        100
. gen urban0 = urban - r(mean)


. sum loggdp
    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
      loggdp |        179    8.693936    1.297024   5.634075   11.22399
. gen loggdp0 = loggdp - r(mean)
(15 missing values generated)


. gen urb_gdp = urban0 * loggdp0
(15 missing values generated)
```

Nonlinear Regression (continued)

- Interaction Effects

```
. reg logco2 c.loggdp0 c.urban0 c.loggdp0#c.urban0
      Source |       SS           df       MS      Number of obs   =       175
-------------+----------------------------------   F(3, 171)       =    371.66
       Model |  83.4990751         3  27.833025    Prob > F        =    0.0000
    Residual |  12.8060512       171  .074889188   R-squared       =    0.8670
-------------+----------------------------------   Adj R-squared   =    0.8647
       Total |  96.3051263       174  .553477737   Root MSE        =    .27366

------------------------------------------------------------------------------------
            logco2 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------------+----------------------------------------------------------------
           loggdp0 |   .4848767   .0242429    20.00   0.000     .4370228    .5327306
            urban0 |   .0025141     .00137     1.84   0.068    -.0001903    .0052185
                   |
c.loggdp0#c.urban0 |  -.0035963   .0007565    -4.75   0.000    -.0050895   -.0021031
                   |
             _cons |   .8903587   .0267759    33.25   0.000     .8375049    .9432124
------------------------------------------------------------------------------------
```

Nonlinear Regression (continued)

- Interaction Effects
  - Explanation :

Interactions Between Independent Variables

Interactions Between Two Continuous Variables

- Thus the effect on Y of a change in X_1, holding X_2 constant, is

$$\frac{\Delta Y}{\Delta X_1} = \beta_1 + \beta_3 X_2$$

- A similar calculation shows that the effect on Y of a change ΔX_1 in X_2, holding X_1 constant, is

$$\frac{\Delta Y}{\Delta X_2} = \beta_2 + \beta_3 X_1$$

- That is, if X_1 changes by ΔX_1 and X_2 changes by ΔX_2, then the expected change in Y

$$\Delta Y = (\beta_1 + \beta_3 X_2)\Delta X_1 + (\beta_2 + \beta_3 X_1)\Delta X_2 + \beta_3 \Delta X_1 \Delta X_2$$

Nonlinear Regression (continued)

- Interaction Effects
  - Explanation :

    predicted logco2 rises by 0.65 with each 1-unit increase in loggdp, when urban is at its mean

    predicted logco2 rises by only a small amount, .0025, with each 1-unit increase in urban when loggdp is at its mean.

    each 1-unit increase in urbanization, the effect of loggdp on logco2 becomes weaker, decreasing by −.004.

    二氧化碳排放量随着财富的增加而增加，但在城市化程度较高的国家，二氧化碳排放量的增加幅度较小。

Section 7

Making Regression Tables in Stata : estout package

Subsection 1

Description

Making Regression Tables in Stata : estout package

- Description
  - The estout package provides tools for making regression tables in Stata. The package currently contains the following commands.

    esttab : A command for **publication-style regression tables** that display nicely in Stata's results window, and can be exported to various formats such as **CSV**, **RTF**, **HTML**, **LaTeX**.

    estout : A **generic program for making a table** from one or more sets of estimation results. estout is **the engine behind esttab**.

    eststo : A utility command to **store estimation results** for later tabulation.

    estadd : A utility command to **add additional results** to an existing estimation set.

    estpost : A utility command to **post results** from various **non-eclass commands** as estimation results (so that they can be tabulated).

Making Regression Tables in Stata : estout package

- Installation

```
ssc install estout, replace
help esttab
```

- Reference
  Making regression tables in Stata

Subsection 2

-esttab-

Making Regression Tables in Stata : estout package

- -esttab-
  - *-esttab-* is a wrapper for *-estout-*

```
esttab [ namelist ] [ using filename ] [, options estout_opt
```

- The procedure :
  - Store a number of models
  - Apply esttab to these stored estimation sets to compose a regression
  - Table produces a fully formatted right away

Making Regression Tables in Stata : estout package

- -esttab-
  - Example

```
. sysuse auto, clear
(1978 Automobile Data)
. qui eststo: reg price weight mpg
. qui eststo: reg price weight mpg foreign
. esttab          //Models stored are automatically picked up by esttab
----------------------------------------------------------------------------
                      (1)             (2)             (3)             (4)
                    price           price           price           price
----------------------------------------------------------------------------
weight              1.747**         3.465***        1.747**         3.465***
                   (2.72)          (5.49)          (2.72)          (5.49)
mpg                -49.51           21.85          -49.51           21.85
                  (-0.57)          (0.29)         (-0.57)          (0.29)
foreign                            3673.1***                       3673.1***
                                   (5.37)                          (5.37)
_cons              1946.1         -5853.7          1946.1         -5853.7
                   (0.54)         (-1.73)          (0.54)         (-1.73)
----------------------------------------------------------------------------
N                      74              74              74              74
----------------------------------------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
. eststo clear    //removes the models from memory

* eststo : Store the regression models.
```

Making Regression Tables in Stata : estout package

- -esttab-
  - Or –>

```
. sysuse auto, clear
(1978 Automobile Data)
. qui reg price weight mpg
. est store model1
. qui reg price weight mpg foreign
. est store model2
. esttab model1 model2
--------------------------------------------
                      (1)             (2)
                    price           price
--------------------------------------------
weight              1.747**         3.465***
                   (2.72)          (5.49)
mpg                -49.51           21.85
                  (-0.57)          (0.29)
foreign                            3673.1***
                                   (5.37)
_cons              1946.1         -5853.7
                   (0.54)         (-1.73)
--------------------------------------------
N                      74              74
--------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
. est clear
```

Making Regression Tables in Stata : estout package

- -esttab-
  - Standard errors, p-values, and summary statistics

```
* Default : (t-statistics) and the number of observations in the table footer


. sysuse auto, clear
(1978 Automobile Data)
. qui eststo: reg price weight mpg
. qui eststo: reg price weight mpg foreign
. esttab, se ar2 //replace by standard errors and add the adjusted R-squared
--------------------------------------------
                      (1)             (2)
                    price           price
--------------------------------------------
weight              1.747**         3.465***
                  (0.641)         (0.631)
mpg                -49.51           21.85
                  (86.16)         (74.22)
foreign                            3673.1***
                                  (684.0)
_cons              1946.1         -5853.7
                 (3597.0)        (3377.0)
--------------------------------------------
N                      74              74
adj. R-sq           0.273           0.478
--------------------------------------------
Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001
* t-statistics can also be replaced by : p, ci, aux
```

Making Regression Tables in Stata : estout package

- -esttab-
  - Standard errors, p-values, and summary statistics

```
* Further summary statistics options : pr2, bic, scalars()

. esttab, p scalars(F df_m df_r)
--------------------------------------------
                      (1)             (2)   
                    price           price   
--------------------------------------------
weight              1.747**         3.465***
                  (0.008)         (0.000)   
mpg                -49.51           21.85   
                  (0.567)         (0.769)   
foreign                            3673.1***
                                  (0.000)   
_cons              1946.1         -5853.7   
                  (0.590)         (0.087)   
--------------------------------------------
N                      74              74   
F                   14.74           23.29   
df_m                    2               3   
df_r                   71              70   
--------------------------------------------
p-values in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

Making Regression Tables in Stata : estout package

- -esttab-
  - Numerical formats

```
* Default :
* t-statistics are printed using 2 decimal places.
* R-squared measures are printed using 3 decimal places.
* point estimates, standard errors using an adaptive display format.(a3)

. esttab, b(a6) p(4) r2(4) nostar
----------------------------------------------------
                     (1)             (2)
                   price           price
----------------------------------------------------
weight          1.746559        3.464706
                (0.0081)        (0.0000)
mpg            -49.51222        21.85360
                (0.5673)        (0.7693)
foreign                         3673.060
                                (0.0000)
_cons           1946.069       -5853.696
                (0.5902)        (0.0874)
----------------------------------------------------
N                     74              74
R-sq              0.2934          0.4996
----------------------------------------------------
p-values in parentheses

* increase precision for the point estimates
* display p-values and the R-squared using 2 decimal places
```

Making Regression Tables in Stata : estout package

- -esttab-
  - Labels, titles, and notes

```
. esttab, label                                   ///
      title(This is a regression table)           ///
      nonumbers mtitles("Model A" "Model B")      ///
      addnote("Source: auto.dta")
This is a regression table
----------------------------------------------------
                          Model A         Model B
----------------------------------------------------
Weight (lbs.)              1.747**         3.465***
                          (2.72)          (5.49)
Mileage (mpg)             -49.51           21.85
                         (-0.57)          (0.29)
Car type                                  3673.1***
                                          (5.37)
Constant                  1946.1         -5853.7
                          (0.54)         (-1.73)
----------------------------------------------------
Observations                  74              74
----------------------------------------------------
t statistics in parentheses
Source: auto.dta
* p<0.05, ** p<0.01, *** p<0.001

. eststo clear
```

Making Regression Tables in Stata : estout package

- -esttab-
  - Labels, titles, and notes

```
* About factor variables and interactions

. sysuse auto, clear
(1978 Automobile Data)
. qui eststo: reg price mpg i.foreign
. qui eststo: reg price c.mpg##i.foreign
. esttab, varwidth(25) label nobaselevels interaction(" X ")
--------------------------------------------------------------
                                   (1)             (2)   
                                 Price           Price   
--------------------------------------------------------------
Mileage (mpg)                   -294.2***       -329.3***
                               (-5.28)         (-4.39)   
Foreign                         1767.3*         -13.59   
                                (2.52)         (-0.01)   
Foreign X Mileage (mpg)                          78.89   
                                                (0.70)   
Constant                       11905.4***      12600.5***
                               (10.28)          (8.25)   
--------------------------------------------------------------
Observations                        74              74   
--------------------------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

. eststo clear
```

- -esttab-
  - Compressed table

```
. sysuse auto, clear
(1978 Automobile Data)
. qui eststo: reg price weight
. qui eststo: reg price weight mpg
. qui eststo: reg price weight mpg foreign
. qui eststo: reg price weight mpg foreign displacement
. esttab, compress
------------------------------------------------------------
                 (1)          (2)          (3)          (4)
               price        price        price        price
------------------------------------------------------------
weight         2.044***     1.747**      3.465***     2.458**
              (5.42)       (2.72)       (5.49)       (2.82)
mpg                        -49.51        21.85        19.08
                          (-0.57)       (0.29)       (0.26)
foreign                                 3673.1***    3930.2***
                                        (5.37)       (5.67)
displace~t                                            10.22
                                                     (1.65)
_cons         -6.707       1946.1      -5853.7      -4846.8
             (-0.01)       (0.54)      (-1.73)      (-1.43)
------------------------------------------------------------
N                 74           74           74           74
------------------------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
* reduces horizontal spacing to fit more models on screen without line breaking

. eststo clear
```

Making Regression Tables in Stata : estout package

- -esttab-
  - Significance stars

```
. sysuse auto, clear
(1978 Automobile Data)
. qui eststo: reg price weight mpg
. qui eststo: reg price weight mpg foreign
. esttab, star(+ 0.10 * 0.05)
----------------------------------------
                      (1)          (2)
                    price        price
----------------------------------------
weight              1.747*       3.465*
                   (2.72)       (5.49)
mpg                -49.51        21.85
                  (-0.57)       (0.29)
foreign                         3673.1*
                                (5.37)
_cons              1946.1      -5853.7+
                   (0.54)      (-1.73)
----------------------------------------
N                      74           74
----------------------------------------
t statistics in parentheses
+ p<0.10, * p<0.05
* default symbols and thresholds are:  * for p<.05, ** for p<.01, and *** for p<.001.
```

Making Regression Tables in Stata : estout package

- -esttab-
  - With Excel

```
esttab using example.csv
esttab using example.csv,replace wide plain
```

  - With Word

```
esttab using example.rtf
esttab using example.rtf, append wide label modelwidth(8)

* varwidth() and modelwidth() change the column widths


lab var mpg "The mgp variable has a really long label"
esttab using example.rtf, replace label nogap onecell

* onecell : placed beneath one another in the same table cell
```

Making Regression Tables in Stata : estout package

- -esttab-
  - With LaTeX

```
. esttab using example.tex, label nostar replace page

* page[(packages)] adds opening and closing code to define a whole LaTeX document
```

- -esttab-
  - With LaTeX

	(1) Price	(2) Price
Weight (lbs.)	1.747	3.465
	(2.72)	(5.49)
Mileage (mpg)	-49.51	21.85
	(-0.57)	(0.29)
Car type		3673.1
		(5.37)
Constant	1946.1	-5853.7
	(0.54)	(-1.73)
Observations	74	74

t statistics in parentheses

- -esttab-
  - With LaTeX

```
. esttab using example.tex, label nostar replace page booktabs

* produces a LaTeX formatted table for use with LaTeX´s booktabs package
```

Making Regression Tables in Stata : estout package

- -esttab-
  - With LaTeX

	(1) Price	(2) Price
Weight (lbs.)	1.747	3.465
	(2.72)	(5.49)
Mileage (mpg)	-49.51	21.85
	(-0.57)	(0.29)
Car type		3673.1
		(5.37)
Constant	1946.1	-5853.7
	(0.54)	(-1.73)
Observations	74	74

t statistics in parentheses

Making Regression Tables in Stata : estout package

- -esttab-
  - With LaTeX

```
. esttab using example.tex, label nostar replace page booktabs ///
    width(0.8\hsize)


* width(\hsize) in LaTeX or width(100%) in HTML to span the whole page
```

- -esttab-
  - With LaTeX

	(1) Price	(2) Price
Weight (lbs.)	1.747	3.465
	(2.72)	(5.49)
Mileage (mpg)	-49.51	21.85
	(-0.57)	(0.29)
Car type		3673.1
		(5.37)
Constant	1946.1	-5853.7
	(0.54)	(-1.73)
Observations	74	74

t statistics in parentheses

Making Regression Tables in Stata : estout package

- -esttab-
  - With LaTeX

```
. esttab using example.tex, label nostar replace page booktabs ///
    width(0.8\hsize) alignment(ll)


* specify the alignment of the models´ columns in LaTeX
```

Making Regression Tables in Stata : estout package

- -esttab-
  - With LaTeX

	(1) Price	(2) Price
Weight (lbs.)	1.747	3.465
	(2.72)	(5.49)
Mileage (mpg)	-49.51	21.85
	(-0.57)	(0.29)
Car type		3673.1
		(5.37)
Constant	1946.1	-5853.7
	(0.54)	(-1.73)
Observations	74	74

t statistics in parentheses

Making Regression Tables in Stata : estout package

- -esttab-
  - With LaTeX

```
. esttab using example.tex, label nostar replace page booktabs ///
    width(0.8\hsize) alignment(ll) title(Regression table)
```

Making Regression Tables in Stata : estout package

- -esttab-
  - With LaTeX

表 2: Regression table

	(1) Price	(2) Price
Weight (lbs.)	1.747 (2.72)	3.465 (5.49)
Mileage (mpg)	-49.51 (-0.57)	21.85 (0.29)
Car type		3673.1 (5.37)
Constant	1946.1 (0.54)	-5853.7 (-1.73)
Observations	74	74

t statistics in parentheses

Making Regression Tables in Stata : estout package

- -esttab-
  - With LaTeX–分组样本回归 mgroup() 选项

```
sysuse auto, clear
eststo clear
eststo: qui reg weight mpg
eststo: qui reg weight mpg foreign
eststo: qui reg price weight mpg
eststo: qui reg price weight mpg foreign
esttab using mgroups.tex, replace                          ///
         star(* 0.1 ** 0.05 *** 0.01)                      ///
         compress nogaps                                   ///
         title(An Illustration of mgroup() in esttab)     ///
         mgroups("Group A" "Group B",                      ///
             pattern(1 0 1 0) span                         ///
             prefix(\multicolumn{@span}{c}{) suffix(})   ///
             erepeat(\cmidrule(lr){@span}) )               ///
         booktabs page(dcolumn) alignment(D{.}{.}{-1})
```

- -esttab-
  - With LaTeX–分组样本回归 mgroup() 选项

```
* "Group A" "Group B" : 组别名称
* mgroup()子选项 :
* pattern(1 0 1 0) : 1--该组别的第一个模型, 0--该组别的其他模型
* span : 定组别名能在表格中跨列
* prefix(\multicolumn{@span}{c}{)和suffix(}):组别名在LaTeX代码中跨行
* erepeat(\cmidrule(lr){@span}) : 设定跨行代码,下面加底部表格线
* page(dcolumn) : 添加加载宏包dcolumn
* alignment(D{.}{.}{-1} : 调整单元格对齐方式,小数点对齐
```

Making Regression Tables in Stata : estout package

- -esttab-
  - With LaTeX–分组样本回归 mgroup() 选项

Table 1: An Illustration of mgroup() in esttab

	Group A		Group B	
	(1) weight	(2) weight	(3) price	(4) price
mpg	-108.4^{***}	-91.22^{***}	-49.51	21.85
	(-11.60)	(-10.34)	(-0.57)	(0.29)
foreign		-550.1^{***}		3673.1^{***}
		(-4.96)		(5.37)
weight			1.747^{***}	3.465^{***}
			(2.72)	(5.49)
_cons	5328.8^{***}	5125.7^{***}	1946.1	-5853.7^{*}
	(25.85)	(27.93)	(0.54)	(-1.73)
N	74	74	74	74

t statistics in parentheses

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Making Regression Tables in Stata : estout package

- -esttab-
  - Non-standard contents

    Sometimes it is necessary to include parameter statistics in a table

    main() option : replacing the point-estimates

    aux() option : replacing the t-statistics

Making Regression Tables in Stata : estout package

- -esttab-
  - Non-standard contents

    For example, to include VIF(variance inflation factors) instead of t-statistics after reg

```
. sysuse auto, clear
. eststo clear
. reg price weight mpg foreign
. estadd vif

    Variable |       VIF       1/VIF
-------------+----------------------
      weight |      3.86    0.258809
         mpg |      2.96    0.337297
     foreign |      1.59    0.627761
-------------+----------------------
    Mean VIF |      2.81
added matrix:
                e(vif) :  1 x 4
```

- -esttab-
  - Non-standard contents

```
. esttab, aux(vif 2) wide nopar
----------------------------------------
                      (1)
                    price
----------------------------------------
weight              3.465***        3.86
mpg                 21.85           2.96
foreign            3673.1***        1.59
_cons             -5853.7
----------------------------------------
N                      74
----------------------------------------
vif in second column
* p<0.05, ** p<0.01, *** p<0.001
```

- -esttab-
  - Non-standard contents

    More than two kinds of parameter statistics

    Switch to **estout syntax** and make use of the cells() option

    cells() disables b(), beta(), main(), t(), abs, se(), p(), ci()...

Making Regression Tables in Stata : estout package

- -esttab-
  - Non-standard contents

    For example,to print point estimates, t-statistics, and variance inflation factors in one table

```
. esttab, cells("b(fmt(a3) star) vif(fmt(2))" t(par fmt(2)))
----------------------------------------------
                      (1)
                    price
                      b/t                  vif
----------------------------------------------
weight              3.465***              3.86
                   (5.49)
mpg                 21.85                 2.96
                   (0.29)
foreign            3673.1***              1.59
                   (5.37)
_cons             -5853.7
                  (-1.73)
----------------------------------------------
N                      74
----------------------------------------------
```

- -esttab-
  - Non-standard contents

    complicated summary statistics section in the table footer

    r2, ar2, pr2, aic, bic, scalars()...

    estout's stats() option, equivalently

Making Regression Tables in Stata : estout package

- -esttab-
  - Non-standard contents

```
return list
dis r(cmdline)
```

Subsection 3

-estout-

Making Regression Tables in Stata : estout package

- -estout-
  - The full syntax of estout is rather complex
  - The most important options:

```
estout [ namelist ] [ using filename ] [, cells(array)
                     stats(scalarlist) style(style) optio
```

  - cells() and stats() options : determine the primary contents of the table
  - style() option : determines the basic formatting of the table.

Making Regression Tables in Stata : estout package

- -estout-
  - Choosing an output format

```
. sysuse auto, clear
. eststo: qui reg price weight mpg
. eststo: qui reg price weight mpg foreign
. estout, style(tex)
            &        est1&        est2&        est3&        est4\\
            &           b&           b&           b&           b\\
weight      &    1.746559&    3.464706&    1.746559&    3.464706\\
mpg         &   -49.51222&     21.8536&   -49.51222&     21.8536\\
foreign     &            &     3673.06&            &     3673.06\\
_cons       &    1946.069&   -5853.696&    1946.069&   -5853.696\\
```

Making Regression Tables in Stata : estout package

- -estout-
  - The cells option
  - The stats option

```
. eststo clear
. estout, stats(r2 bic N)
----------------------------------
                                 .
                                 b
----------------------------------
weight                    3.464706
mpg                        21.8536
foreign                    3673.06
_cons                    -5853.696
----------------------------------
r2                        .4995594
bic                       1357.414
N                               74
----------------------------------
```

Making Regression Tables in Stata : estout package

- -estout-
  - Using labels

```
. sysuse auto, clear
. eststo clear
. eststo, title("Model 1"): qui reg price weight mpg
. eststo, title("Model 2"): qui reg price weight mpg foreign
. label variable foreign "Car type (1=foreign)"
. estout, cells("b(star label(Coef.)) se(label(Std. err.))")  ///
      stats(r2 N, labels(R-squared "N. of cases"))            ///
      label legend varlabels(_cons Constant)
```

	Model 1		Model 2	
	Coef.	Std. err.	Coef.	Std. err.
Weight (lbs.)	1.746559**	.6413538	3.464706***	.630749
Mileage (mpg)	-49.51222	86.15604	21.8536	74.22114
Car type (1=foreign)			3673.06***	683.9783
Constant	1946.069	3597.05	-5853.696	3376.987
R-squared	.2933891		.4995594	
N. of cases	74		74	

* p<0.05, ** p<0.01, *** p<0.001

Subsection 4

-eststo-

Making Regression Tables in Stata : estout package

- -eststo-
  - Stores a copy of the active estimation results for later tabulation.
  - Analogous to official Stata's **estimates store**.
  - Does not require the user to specify a name for the stored estimates.

```
sysuse auto, clear
reg price weight mpg
eststo
reg price weight mpg foreign
eststo
esttab
```

  - As a prefix

```
sysuse auto, clear
eststo: qui reg price weight mpg
eststo: qui reg price weight mpg foreign
esttab
```

Making Regression Tables in Stata : estout package

- -eststo-
  - Using by

```
. sysuse auto, clear
. eststo clear
. by foreign : eststo: qui reg price weight mpg
------------------------------------------------------------------------------------
-> Domestic
(est1 stored)
------------------------------------------------------------------------------------
-> Foreign
(est2 stored)
. esttab, label nodepvar nonumber
----------------------------------------------------
                          Domestic         Foreign
----------------------------------------------------
Weight (lbs.)                4.415***        5.156***
                            (4.66)          (5.85)
Mileage (mpg)                237.7          -19.78
                            (1.71)         (-0.34)
Constant                  -13285.4*        -5065.8
                           (-2.32)         (-1.58)
----------------------------------------------------
Observations                    52              22
----------------------------------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

Making Regression Tables in Stata : estout package

- -eststo-
  - Adding additional statistics

```
. sysuse auto, clear
. eststo clear
. qui reg price weight mpg
. test weight = mpg
 ( 1)  weight - mpg = 0
       F(  1,    71) =    0.36
            Prob > F =    0.5514
. eststo, addscalars(p_diff r(p))
(e(p_diff) = .55138216 added)
(est1 stored)
. esttab, scalars(p_diff) obslast
----------------------------
                      (1)
                    price
----------------------------
weight              1.747**
                   (2.72)
mpg                -49.51
                  (-0.57)
_cons              1946.1
                   (0.54)
----------------------------
p_diff              0.551
N                      74
----------------------------
t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001
```

Subsection 5

-estadd-

- -estadd-
  - Add results to (stored) estimates
  - Results that are included in the e()-returns for the models can be tabulated by estout or esttab
  - 举例：向内存中添加两个统计量：

    一个是文字类型的返回值 Industry，采用暂元（local）存储

    一个是数值类型的返回值 Mean_Wage，采用单值（scalar）来存储

Making Regression Tables in Stata : estout package

- -estadd-

```
. eststo clear
. sysuse nlsw88.dta, clear
(NLSW, 1988 extract)
. qui reg wage ttl_exp married
. estadd  local Industry "Yes"
added macro:
           e(Industry) : "Yes"
```

```
. qui sum wage
. estadd  scalar Mean_Wage = r(mean)
added scalar:
          e(Mean_Wage) =  7.766949


. qui ereturn list
```

Making Regression Tables in Stata : estout package

- -estadd-

```
^^Isysuse nlsw88.dta, clear

^^Iglobal xx "ttl_exp married south hours tenure age i.industry"
^^I
^^Iqui reg wage $xx if race==1
^^Iestadd local Industry   "Yes"
^^Iestadd local Occupation "No"
^^Iest store m1
^^I
^^Iqui reg wage $xx if race==2
^^Iestadd local Industry   "Yes"
^^Iestadd local Occupation "No"
^^Iest store m2
^^I
^^Iqui reg wage $xx i.occupation if race==1
^^Iestadd local Industry   "Yes"
^^Iestadd local Occupation "Yes"
^^Iest store m3
^^I
^^Iqui reg wage $xx i.occupation if race==2
^^Iestadd local Industry   "Yes"
^^Iestadd local Occupation "Yes"
^^Iest store m4
```

Making Regression Tables in Stata : estout package

- -estadd-

```
local m "m1 m2 m3 m4"
esttab `m´, mtitle(White Black White Black) b(%6.3f) nogap compress  ///
            star(* 0.1 ** 0.05 *** 0.01)                              ///
            drop(*.industry *.occupation)                             ///
            ar2 scalar(N Industry Occupation)
```

	(1) White	(2) Black	(3) White	(4) Black
ttl_exp	0.251***	0.271***	0.176***	0.193***
	(6.49)	(4.82)	(4.61)	(3.61)
married	-0.737**	0.082	-0.657**	0.099
	(-2.31)	(0.21)	(-2.12)	(0.27)
south	-0.814***	-2.038***	-0.758***	-1.791***
	(-2.72)	(-4.92)	(-2.61)	(-4.56)
hours	0.051***	0.036	0.021	0.007
	(3.81)	(1.35)	(1.56)	(0.25)
tenure	0.025	-0.003	0.039	-0.014
	(0.76)	(-0.08)	(1.25)	(-0.34)
age	-0.073	-0.216***	-0.058	-0.148**
	(-1.58)	(-3.12)	(-1.31)	(-2.23)
_cons	5.576**	9.490**	9.607***	8.420**
	(2.22)	(2.53)	(3.82)	(2.39)
N	1615	572	1612	571
adj. R-sq	0.112	0.166	0.176	0.281
Industry	Yes	Yes	Yes	Yes
Occupation	No	No	Yes	Yes

t statistics in parentheses
* p<0.1, ** p<0.05, *** p<0.01

Subsection 6

-estpost-

- -estpost-
  - esttab and estout tabulate the e()-returns of a command, but not all commands return their results in e()
  - Posting results from **non-eclass** commands
  - Review

- 结果导出-esttab-

```
sysuse auto,clear

local var price wei len mpg
qui estpost ttest `var´, by(foreign)
esttab using ttable2.rtf, cell("mu_1(fmt(2)) mu_2(fmt(2)) b(star fmt(2)) t(fmt(2))")  ///
        starlevels(* 0.10 ** 0.05 *** 0.01) replace noobs compress                    ///
        title(esttab_Table: T_test)
```

- 描述性统计表格导出
  - -esttab-

```
estpost summarize price wei len mpg rep78, detail
esttab using Desc4.rtf,                                                        ///
    cells("count mean(fmt(2)) sd(fmt(2)) min(fmt(2)) p50(fmt(2)) max(fmt(2))") ///
    noobs compress replace title(esttab_Table: Descriptive statistics)
```

Making Regression Tables in Stata : estout package

- -estpost-
  - Review
  - 相关系数矩阵导出

```
*<方法一> -logout-
logout, save(Corr1) word replace: pwcorr price wei len mpg rep78, star(.05)


*<方法二> -esttab-
estpost correlate  price wei len mpg rep78, matrix
esttab using Corr2.rtf,                                                   ///
    unstack not noobs compress nogaps replace star(* 0.1 ** 0.05 *** 0.01) ///
    b(%8.3f) p(%8.3f) title(esttab_Table: correlation coefficient matrix)


*<方法三> -corr2docx-
corr2docx price wei len mpg rep78 using Corr3.docx,                       ///
        replace spearman(ignore) pearson(pw) star                         ///
        title(corr2docx_Table: correlation coefficient matrix)
```