

计量经济学

课堂练习（一）：第五至七讲 gretl 练习汇总

黄嘉平

工学博士 经济学博士
深圳大学中国经济特区研究中心 讲师

办公室 粤海校区汇文楼2613
E-mail huangjp@szu.edu.cn
Website <https://huangjp.com>

第五讲

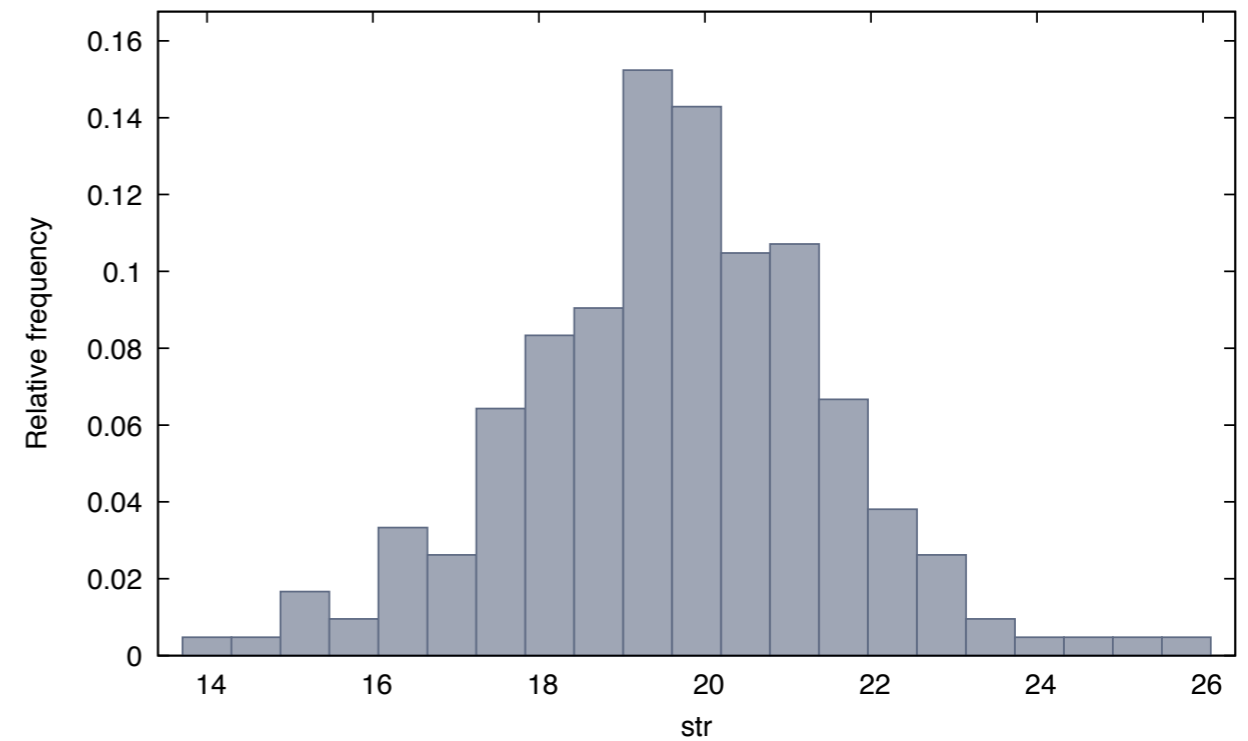
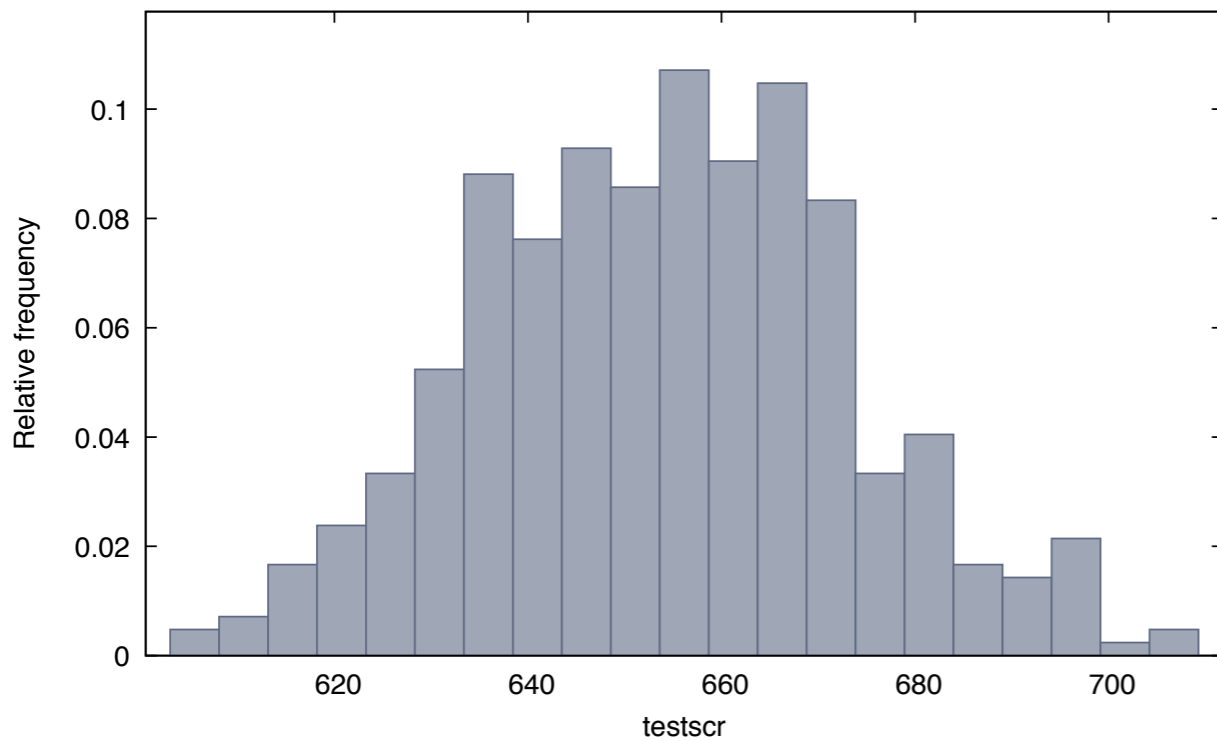
STAR 数据集

- STAR (California Standardized Testing and Reporting) 数据集是惯于美国加州小学教育的数据集，包含考试成绩、学校特征、学生背景等变量。本书中用到的数据涵盖1998-1999年间420个学区。
- 数据文件 `caschool.xlsx`
- 说明文件 `californiatestscores.docx`

考试成绩和学生教师比

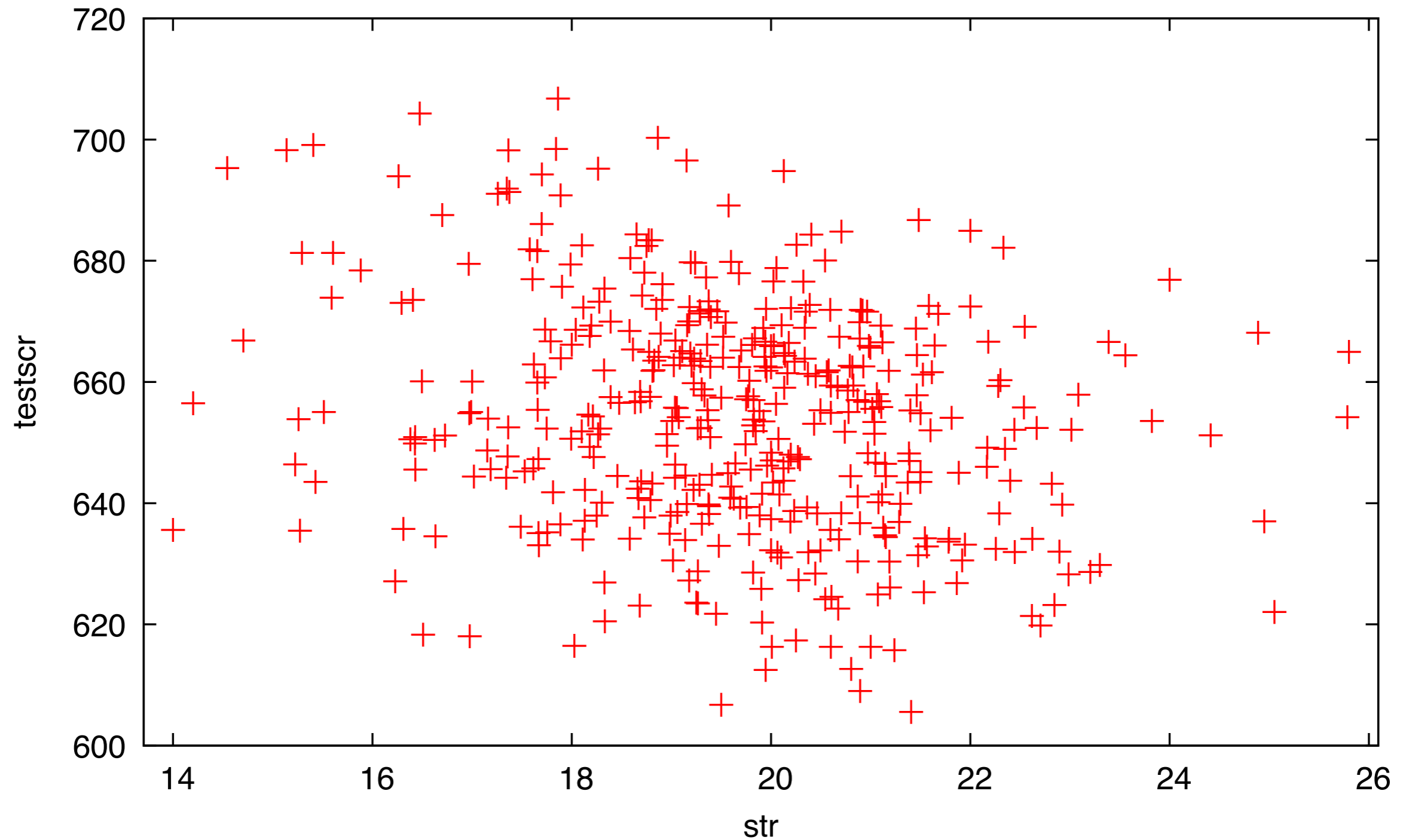
Average test score and student-teacher ratio

- “testscr”: 数学和语言考试的平均成绩
- “str”: 学生教师比 (学生人数 / 教师人数)



考试成绩和学生教师比

Average test score and student-teacher ratio



在 gretl 中进行 OLS 回归


- 在主程序窗中：

Model > Ordinary least squares >

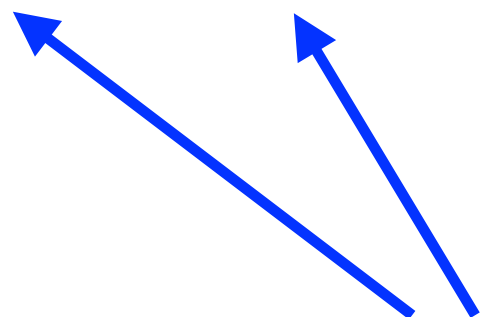
- 在编程模式中：

ols testscr **const** str

因变量



回归变量



$$\begin{aligned} Y_i &= \beta_0 + \beta_1 X_i + u_i \\ &= \beta_0 \mathbf{1} + \beta_1 X_i + u_i \end{aligned}$$

在 gretl 中进行 OLS 回归

Model 1: OLS, using observations 1-420

Dependent variable: testscr

	coefficient	std. error	t-ratio	p-value	
const	698.933	9.46749	73.82	6.57e-242	***
str	-2.27981	0.479826	-4.751	2.78e-06	***
Mean dependent var	654.1565	S.D. dependent var	19.05335		
Sum squared resid	144315.5	S.E. of regression	18.58097		
R-squared	0.051240	Adjusted R-squared	0.048970		
F(1, 418)	22.57511	P-value(F)	2.78e-06		
Log-likelihood	-1822.250	Akaike criterion	3648.499		
Schwarz criterion	3656.580	Hannan-Quinn	3651.693		

第六讲

课后练习（不需提交）

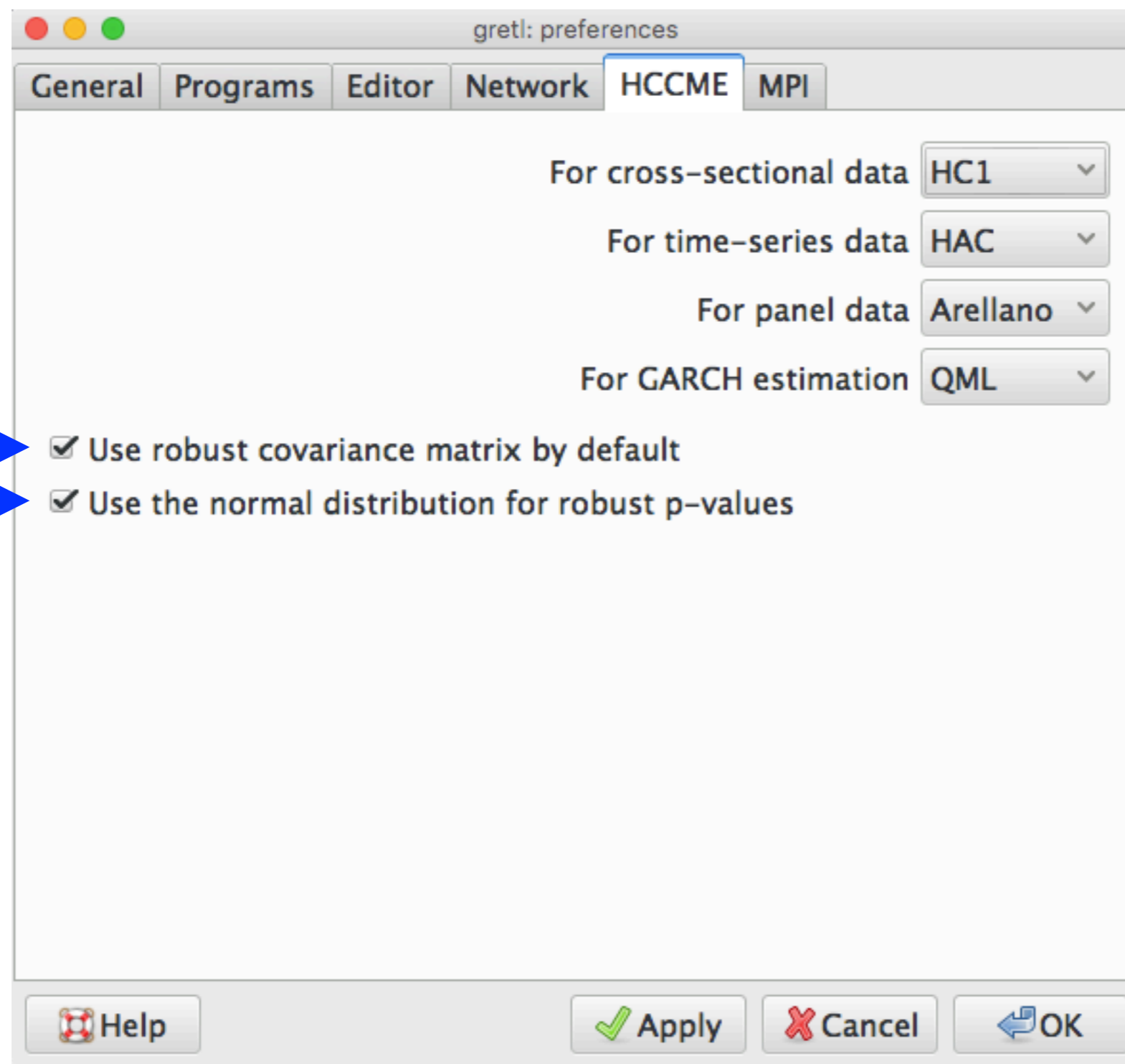
- 使用 STAR 数据集，根据 str 变量定义二值变量 D_i

$$D_i = \begin{cases} 1 & \text{if str} < 20 \\ 0 & \text{if str} \geq 20 \end{cases}$$

- 用 gretl 对模型 $Y_i = \beta_0 + \beta_1 D_i + u_i$ 进行回归分析，找到系数 β_1 的统计值和标准误。
- 令 str_L 为 $\text{str} < 20$ 的子集， str_H 为 $\text{str} \geq 20$ 的子集。针对两个子集中 testscr 的均值进行比较，并对均值差为零的原假设进行检验（参考3.4节）。比较检验结果和回归分析结果。

在 gretl 中使用异方差稳健标准误

偏好设置 Tools > Preferences > General...



HC0: White (1980)
HC1: HC0 + 调整自由度
即本书中的公式

在 gretl 中使用异方差稳健标准误

编程模式

- 在整个程序开始时运行以下命令，可适用于整个程序

```
set force_hc on  
set hc_version 1  
      # 0 (the original White's) is the default  
set robust_z on
```

- 对于单一 OLS 回归，则可以采用以下写法

```
ols yvar xvar --robust
```

(当然，你还是需要提前设置 `hc_version`)

STAR 数据的回归结果

异方差

Model 1: OLS, using observations 1-420

Dependent variable: testscr

Heteroskedasticity-robust standard errors, variant HC1

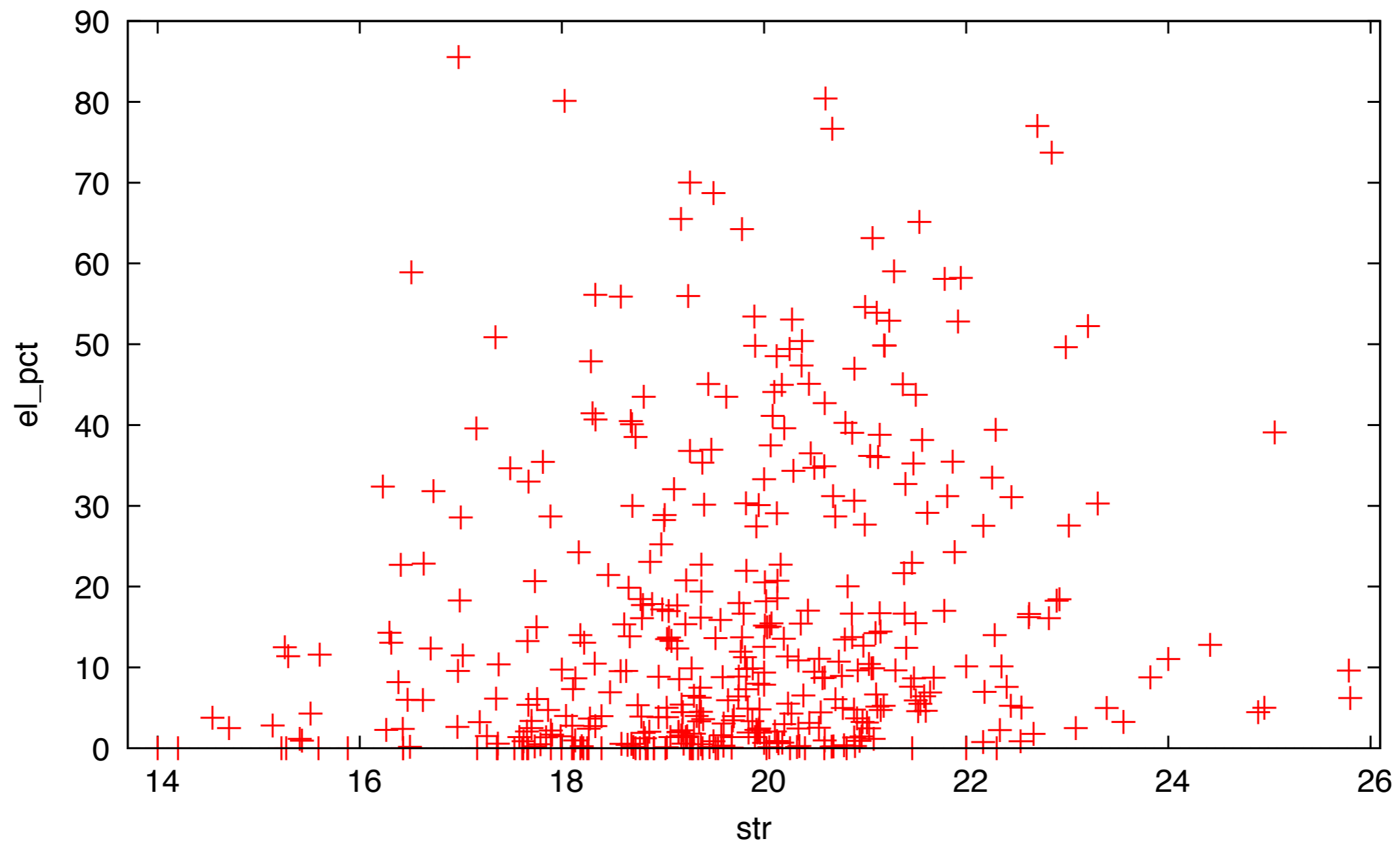
	coefficient	std. error	z	p-value	
const	698.933	10.3644	67.44	0.0000	***
str	-2.27981	0.519489	-4.389	1.14e-05	***
Mean dependent var	654.1565	S.D. dependent var	19.05335		
Sum squared resid	144315.5	S.E. of regression	18.58097		
R-squared	0.051240	Adjusted R-squared	0.048970		
F(1, 418)	19.25943	P-value(F)	0.000014		
Log-likelihood	-1822.250	Akaike criterion	3648.499		
Schwarz criterion	3656.580	Hannan-Quinn	3651.693		

第七讲

学生教师比与仍在学习英语的学生比例

Student teacher ratio and percentage of English learners

两个变量间的相关系数为 0.188



对 testscr 的多元回归分析

- 英语学习者百分比, `el_pct`, 是另一个可以影响考试成绩的变量 (`testscr` 是数学和阅读成绩的平均值, 我们有理由猜测英语非母语学生的阅读成绩低于英语母语学生)。因此, 回归模型可以扩展为:
 - 因变量: $Y = \text{testscr}$
 - 自变量: $X_1 = \text{str}, X_2 = \text{el_pct}$
 - 回归模型: $\text{testscr}_i = \beta_0 + \beta_1 \text{str}_i + \beta_2 \text{el_pct}_i + u_i$
- `grerl` 命令:
`ols testscr const str el_pct --robust`

多元回归结果

model1: OLS, using observations 1-420

Dependent variable: testscr

Heteroskedasticity-robust standard errors, variant HC1

	coefficient	std. error	z	p-value	
const	686.032	8.72822	78.60	0.0000	***
str	-1.10130	0.432847	-2.544	0.0109	**
el_pct	-0.649777	0.0310318	-20.94	2.36e-97	***

Mean dependent var	654.1565	S.D. dependent var	19.05335
Sum squared resid	87245.29	S.E. of regression	14.46448
R-squared	0.426431	Adjusted R-squared	0.423680
F(2, 417)	223.8229	P-value(F)	9.28e-67
Log-likelihood	-1716.561	Akaike criterion	3439.123
Schwarz criterion	3451.243	Hannan-Quinn	3443.913

多重共线性的检测

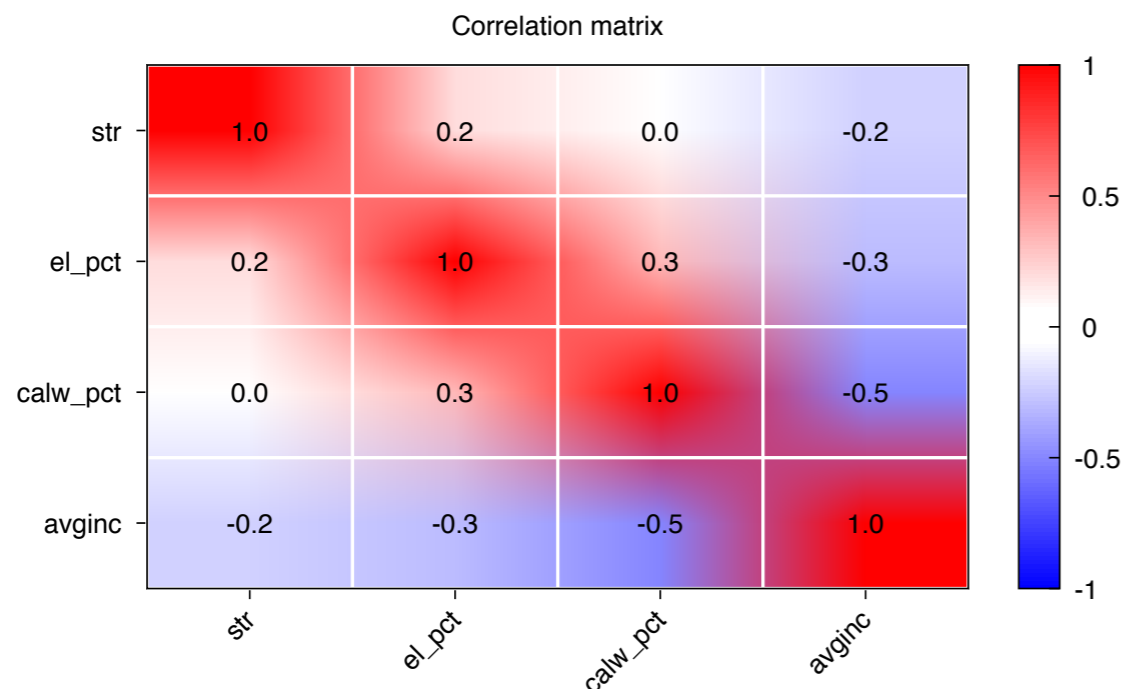
回归变量的样本相关矩阵

- 在 gretl 中可以用下列命令计算变量间的相关系数矩阵

```
corr str el_pct calw_pct avginc --plot=display
```

```
Correlation Coefficients, using the observations 1 - 420  
5% critical value (two-tailed) = 0.0957 for n = 420
```

```
      str      el_pct      calw_pct      avginc      str  
1.0000      0.1876      0.0183      -0.2322      el_pct  
              1.0000      0.3196      -0.3074      calw_pct  
              0.0183      1.0000      -0.5127      avginc  
              -0.2322      -0.3074      -0.5127      1.0000
```



多重共线性的检测

Variance Inflation Factors (VIF)

- Variance Inflation Factors (VIF)

$$\text{VIF}_k = \frac{1}{1 - R_k^2},$$

此处 R_k^2 是用其他自变量回归 X_k 时得到的 R^2 。

- 在运行 **ols** 命令后运行 **vif** 命令可得

```
Variance Inflation Factors  
Minimum possible value = 1.0  
Values > 10.0 may indicate a collinearity problem
```

```
      str      1.036  
el_pct      1.036
```

$\text{VIF}(j) = 1/(1 - R(j)^2)$, where $R(j)$ is the multiple correlation coefficient between variable j and the other independent variables

涉及多个系数的单个约束检验

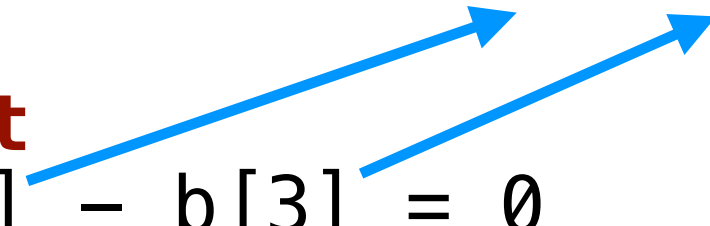
- 有时我们需要检验涉及两个或多个回归系数的单个约束。例如，如果理论指出 $\beta_1 = \beta_2$ ，则应针对下列假设进行检验：

$$H_0 : \beta_1 = \beta_2 ; \quad H_1 : \beta_1 \neq \beta_2$$

- 因为只有一个约束条件， $q = 1$ ， F 统计量在大样本下服从 $F_{1,\infty}$ 分布。
- 在 gretl 中，我们可以先运行 OLS 回归，之后进行检验

```
ols testscr const str expn el_pct --robust
```

```
restrict  
    b[2] - b[3] = 0  
end restrict
```



回归结果

Model 2: OLS, using observations 1-420

Dependent variable: testscr

Heteroskedasticity-robust standard errors, variant HC1

	coefficient	std. error	z	p-value	
const	649.578	15.4583	42.02	0.0000	***
str	-0.286399	0.482073	-0.5941	0.5524	
expn	3.86790	1.58072	2.447	0.0144	**
el_pct	-0.656023	0.0317844	-20.64	1.21e-94	***

Mean dependent var	654.1565	S.D. dependent var	19.05335
Sum squared resid	85699.71	S.E. of regression	14.35301
R-squared	0.436592	Adjusted R-squared	0.432529
F(3, 416)	147.2037	P-value(F)	5.20e-65
Log-likelihood	-1712.808	Akaike criterion	3433.615
Schwarz criterion	3449.776	Hannan-Quinn	3440.003

Excluding the constant, p-value was highest for variable 14 (str)

检验结果

Restriction:

$$b[\text{str}] - b[\text{expn}] = 0$$

Test statistic: Robust F(1, 416) = 8.9403, with p-value = 0.00295511

Restricted estimates:

	coefficient	std. error	t-ratio	p-value	
const	685.822	11.3696	60.32	4.31e-208	***
str	-0.854052	0.459004	-1.861	0.0635	*
expn	-0.854052	0.459004	-1.861	0.0635	*
el_pct	-0.656690	0.0396393	-16.57	9.96e-48	***

Standard error of the regression = 14.5489

课后练习（不需提交）

- 阅读第 7.6 节并尝试复制其中的分析结果。
- 在 gretl 中，**modeltab** 命令可以帮助我们制作和书中类似的表格。具体用法参考 command reference。
(注：gretl 无法替代你制作表格，最终表格的格式还需要手动编辑。)

例：

```
modeltab free  
ols testscr const str --robust --quiet  
modeltab add  
ols testscr const str el_pct --robust --quiet  
modeltab add  
modeltab show
```