# 第六讲 受约束的回归模型

- $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots \beta_k x_k + u$

# 一、对多个线性约束的检验：F检验

- 目前我们仅涉及检验一个单一的线性约束，如$\beta_1 = 0$或$\beta_1 = \beta_2$。
- 然而，我们希望对参数的多个假设进行联合检验。
- 经典例子是检验“排除性约束”：**一组参数**是否等于零。

# 检验排除性约束

- 虚拟假设**$H_0$: $\beta_{k-q+1} = 0, \ldots, \beta_k = 0$;**
- 对立假设是**H1: H0 不正确（即至少有一个参数不为零）.**
- 可否单独检验每一个t统计量? 由于我们希望了解q个参数的**联合**显著性，单独检验t无法做到这一点。

# 排除性约束检验（续）

- 分别估计**受约束模型**和**不受约束模型**。

- 直观的，我们希望了解两个模型残差平方和的变化是否足够大以确定是否应该包括被排除掉的变量$x_{k-q+1}, \cdots, x_k$。

$$F \equiv \frac{\left(SSR_r - SSR_{ur}\right)/q}{SSR_{ur}/\left(n-k-1\right)}$$

其中 $r$ 是受约束模型，$ur$代表不受约束模型。

# $F$ 统计量

- $F$ 统计量总是为正，既然受限制模型的残差平方SSR和不可能小于不受限制的残差平方和。
- 事实上，F统计量衡量的是残差平方和SSR从不受限制模型到受限制模型的相对增加。
- $q$ 限制条件个数， $df_r - df_{ur}$
- $n - k - 1 = df_{ur}$

# $F$ 统计量（续）

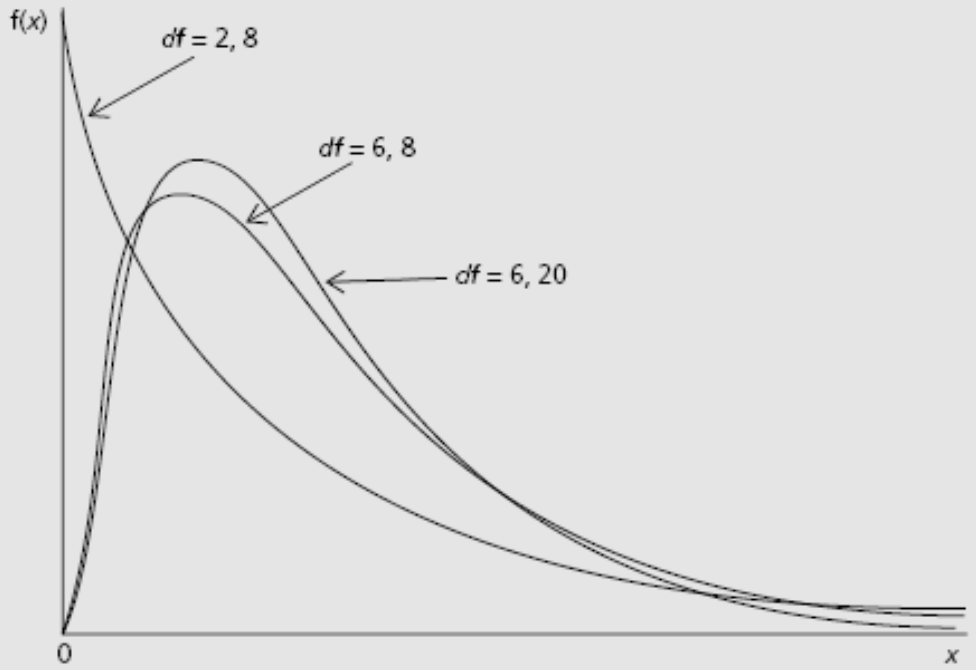
- 为了决定残差平方和的这一增加是否**足够大**以拒绝这一限制性条件，我们需要了解F统计量的样本分布。
- $F \sim F_{q,n-k-1}$，其中$q$ 指F统计量分子的自由度，$n-k-1$指分母的自由度。

# F分布

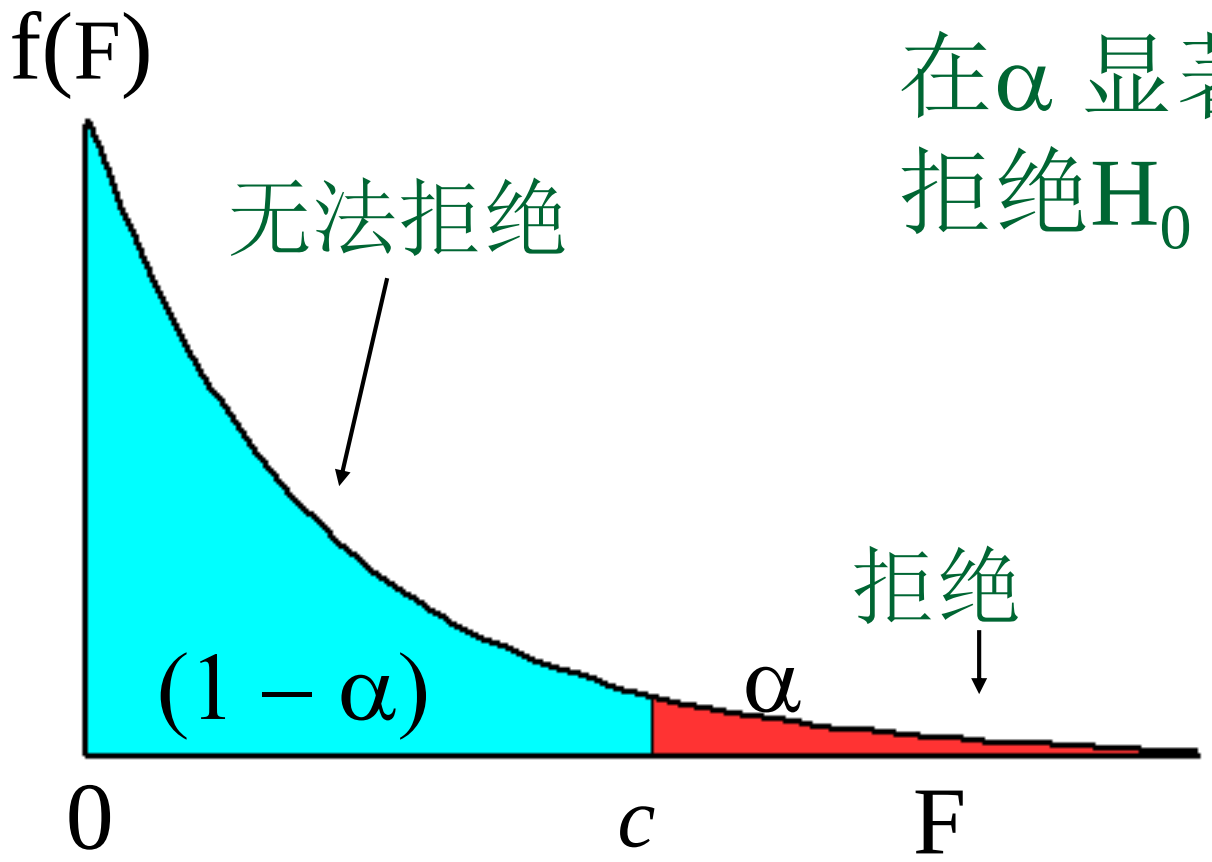
令$X_1 \sim \chi^2_{k_1}$和$X_2 \sim \chi^2_{k_2}$，并假定$X_1$和$X_2$独立，则随机变量$F = \dfrac{(X_1/k_1)}{(X_2/k_2)}$有一个自由度为$(k_1, k_2)$的F分布。记为：$F \sim F_{k_1 k_2}$

**Figure B.11**

The $F_{k_1,k_2}$ distribution for various degrees of freedom, $k_1$ and $k_2$.



# F分布（续）




在α 显著性水平下
拒绝$H_0$ ,如果$F > c$

# 例子：运动员表现及其薪水

$$\log(salary) = \beta_0 + \beta_1 years + \beta_2 gamesyr + \beta_3 bavg + \beta_4 hrunsyr + \beta_5 rbisyr + u, \quad (4.28)$$

$$H_0 : \beta_3 = 0, \beta_4 = 0, \beta_5 = 0.$$

$$H_1 : H_0 \text{不正确。}$$

$$\widehat{\log(salary)} = \underset{(0.29)}{11.10} + \underset{(.0121)}{.0689}\, years + \underset{(.0026)}{.0126}\, gamesyr + \underset{(.00110)}{.00098}\, bavg + \underset{(.0161)}{.0144}\, hrunsyr + \underset{(.0072)}{.0108}\, rbisyr \qquad (4.31)$$

$$n = 353,\ SSR = 183.186,\ R^2 = .6278,$$

$$\widehat{\log(salary)} = \underset{(0.11)}{11.22} + \underset{(.0125)}{.0713}\, years + \underset{(.0013)}{.0202}\, gamesyr \qquad (4.33)$$

$$n = 353,\ SSR = 198.311,\ R^2 = .5971.$$

$$F = \frac{(198.311 - 183.186)/3}{183.186/347} \approx 9.55 > 2.60_{(3,347,5\%)}$$

- 因为F统计量大于临界值2.6，因此，我们拒绝bavg，hrunsyr和rbisyr对薪水没有影响的假设。
- 为何bavg，hrunsyr和rbisyr三变量的参数估计值未通过t检验，而其F检验却是显著的？

  当自变量存在多重共线性时，模型结果难以发现每个变量的偏效应，但却可能发现联合显著性。

$$t = \frac{\hat{\beta}_j}{se\left(\hat{\beta}_j\right)} = \frac{\hat{\beta}_j}{\hat{\sigma}\Big/\left[SST_j\left(1-R_j^2\right)\right]^{1/2}}$$

# F统计量与t统计量的关系

- 当F统计量检验单个变量的排除性时，等于对应的t统计量的平方。
- 给定对立假设为双侧，$t^2_{n-k-1}$与$F_{1,n-k-1}$拥有同样的分布，两种方法的结果一致。

# F统计量的$R^2$ 型

- SSR很大程度上依赖于度量单位， 可以用R2计算F统计量。
- 依据**SSR = SST(1 – R²)** ,F统计量的R2型为：

$$F \equiv \frac{\left(R_{ur}^2 - R_r^2\right)/q}{\left(1-R_{ur}^2\right)/\left(n-k-1\right)},$$

其中，$r$ 代表受限制模型，$ur$代表不受限制模型。

# 续上例，F统计量的R2型为

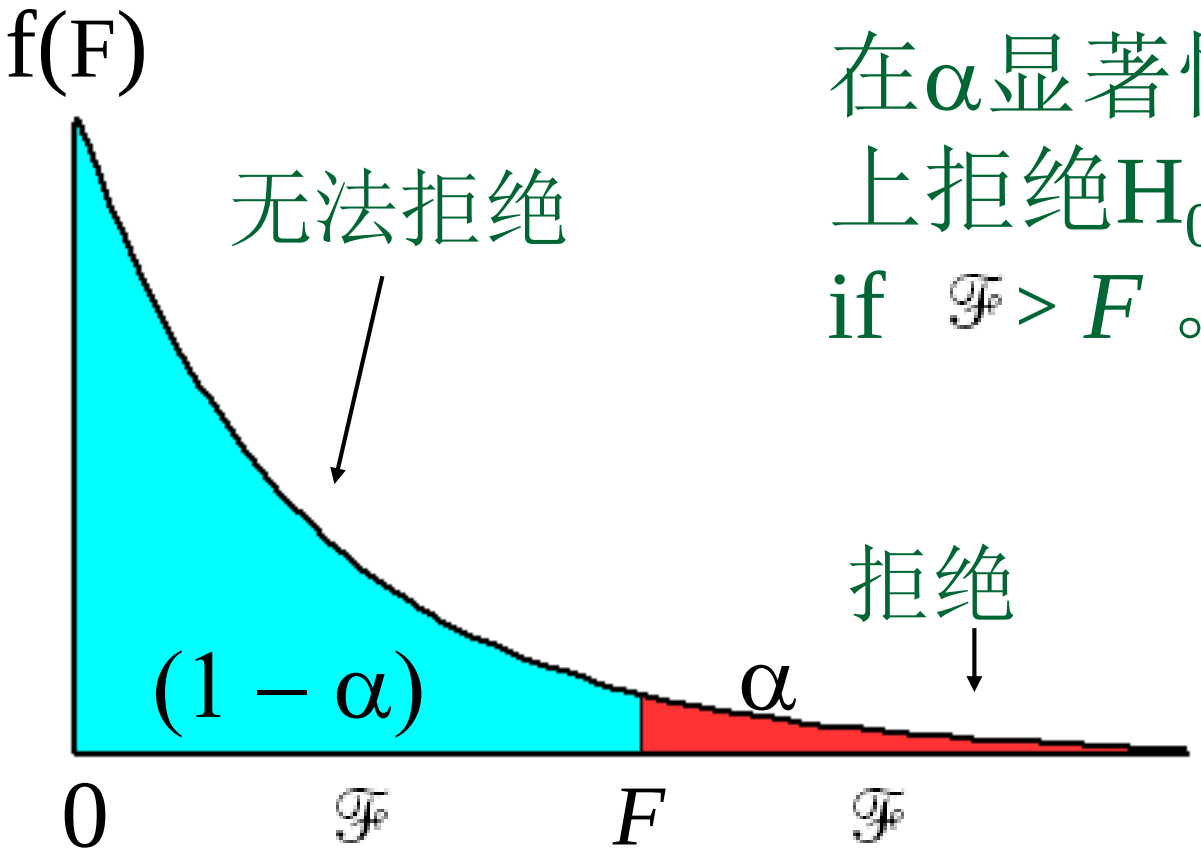
$$F = \frac{(0.6278 - 0.5971)}{1 - 0.6278} \frac{347}{3} \approx 9.54$$

# 计算F检验的p值

在F检验的背景下，p值被定义为

p值=P$(f > F)$:给定虚拟假设正确，观察到的F值至少和我们所得到的F值一样大的概率。

其中，$f$表示一个自由度为$(q, n-k-1)$的F堆积变量，F是检验统计量的实际值。




在α显著性水平上拒绝$H_0$，如果 if $\mathcal{F} > F$ 。

# 回归整体显著性的F统计量

- 排除性约束的一个特例是检验$H_0$: $\beta_1 = \beta_2 = \ldots = \beta_k = 0$，即假设模型中没有任何一个解释变量对y有作用。
- 既然只有截距项的模型$R^2$ 等于零，则整体显著性的F统计量为：

$$F = \frac{R^2 / k}{\left(1 - R^2\right) / \left(n - k - 1\right)}$$

- 如果$H_0$被拒绝，则我们得到结论认为模型中的变量的确对y有解释力，意味着回归是总体显著的。
- 相反，如果我们无法拒绝$H_0$,则没有证据表明模型中的任何一个变量有助于解释y，我们必须需找其他变量来解释y。
- 因此，我们必须计算F统计量来检验联合显著性，而非仅仅看$R^2$的大小。

# 检验一般的线性约束

- F统计量的基本形式可适用于任何的线性约束，而非仅仅是排除性约束。
- 先估计受约束模型，再估计不受约束模型，然后记录两个模型的残差平方和
- 施加约束可以变得很有技巧，类似于重新定义变量。

# 二、邹氏参数稳定性检验

在含有$k$个解释变量和一个截距项的一般模型中，假设有两组，称为g=1和g=2。我们想检验这两组的截距和所有的斜率都相同。

$$y = \beta_{g,0} + \beta_{g,1}x_1 + \beta_{g,2}x_2 + \ldots + \beta_{g,k}x_k + u,$$

不受约束模型除了截距项和变量本身外，还有一组虚拟变量和交互项，其自由度为$n-2(k+1)$。不受约束模型的残差平方和可通过两个分离的回归得到。令$SSR_1$、$SSR_2$表示第一组、第二组估计所得到的残差平方和，SSR为受约束模型的残差平方和（将两组混合并估计一个方程所得到）。

$$F = \frac{\left[SSR - \left(SSR_1 + SSR_2\right)\right] / \left(k+1\right)}{\left(SSR_1 + SSR_2\right) / \left(\left(n - 2(k+1)\right)\right)} \longrightarrow Chow\ statistic$$

# 邹至庄检验的步骤：

- 确立回归的一般模型，明确受约束模型和不受约束模型的自由度。（约束个数：k+1）
- 分别将两组进行回归得到$SSR_1$和$SSR_2$。
- 将两组数据混合并重新估计模型得到受约束模型的残差平方和SSR.
- 运用公式计算邹统计量，检验显著性水平。

# 三、案例分析

- 本实验中，我们将利用数据：“usaauto.dta”，来研究回归系数存在约束的情况下，价格、汽车重量等因素对每加仑汽油所行驶的路程的影响。我们将介绍如何定义约束、列出已定义的约束、取消已定义的约束、以及在定义好约束后如何进行约束回归。

例如，我们想拟合以下的模型：

$$mpg = \beta_0 + \beta_1\ price + \beta_2 weight + \beta_3 displ + \beta_4 gear\_ratio + \beta_5 foreign + \beta_6 length + u$$

且该模型有这样的约束：$\beta_1 = \beta_2 = \beta_3 = \beta_6$，$\beta_4 = \beta_5 = \beta_0/20$。那么，我们可以定义约束如下：

```
constraint 1 price=weight
constraint 2 displ=weight
constraint 3 displ=length
constraint 4 gear_ratio=foreign
constraint 5 foreign=_cons/20
```

拟合前面的约束回归：

```
cnsreg mpg price weight displ gear_ratio foreign length, c(1-5)
```

命令中，cnreg代表进行约束回归，mpg是被解释变量的名称，price weight displ gear_ratio foreign length为各个解释变量的名称，选项c(1-5)表示在1到5个约束之下进行回归。