高级计量经济学

理论经济学博士课程 2023-2024

Lecture 7: Hypothesis Testing

Davidson, R. & MacKinnon, J. (2009). Econometrics Theory and Methods. Oxford University Press.

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1510
E-mail huangjp@szu.edu.cn
Website https://huangjp.com

R 的典型回归结果

Kleiber & Zeileis, *Applied Econometrics with R*, Springer.: Section 3.2

```
> data("CPS1988")
> cps_lm <- lm(log(wage) ~ experience + I(experience^2) + education + ethnicity, data = CPS1988)
> summary(cps_lm)

Call:
lm(formula = log(wage) ~ experience + I(experience^2) + education + ethnicity, data = CPS1988)

Residuals:
    Min      1Q  Median      3Q     Max 
-2.9428 -0.3162  0.0580  0.3756  4.3830 

Coefficients:
                  Estimate Std. Error t value Pr(>|t|)    
(Intercept)      4.321e+00  1.917e-02  225.38   <2e-16 ***
experience       7.747e-02  8.800e-04   88.03   <2e-16 ***
I(experience^2) -1.316e-03  1.899e-05  -69.31   <2e-16 ***
education        8.567e-02  1.272e-03   67.34   <2e-16 ***
ethnicityafam   -2.434e-01  1.292e-02  -18.84   <2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.5839 on 28150 degrees of freedom
Multiple R-squared:  0.3347,Adjusted R-squared:  0.3346 
F-statistic:  3541 on 4 and 28150 DF,  p-value: < 2.2e-16
```

MATLAB 的典型回归结果

```
>> load hospital.mat;
>> mdl = fitlm(hospital,'interactions','ResponseVar','Weight',...
    'PredictorVars',{'Sex','Age','Smoker'},...
    ‘CategoricalVar',{'Sex','Smoker'});

mdl =


Linear regression model:
    Weight ~ 1 + Sex*Age + Sex*Smoker + Age*Smoker

Estimated Coefficients:
                         Estimate      SE        tStat       pValue
                         ________    _______    ________    __________

    (Intercept)             118.7     7.0718      16.785     6.821e-30
    Sex_Male               68.336     9.7153      7.0339    3.3386e-10
    Age                   0.31068    0.18531      1.6765      0.096991
    Smoker_1               3.0425     10.446     0.29127       0.77149
    Sex_Male:Age         -0.49094    0.24764     -1.9825      0.050377
    Sex_Male:Smoker_1      0.9509     3.8031     0.25003       0.80312
    Age:Smoker_1         -0.07288    0.26275    -0.27737       0.78211


Number of observations: 100, Error degrees of freedom: 93
Root Mean Squared Error: 8.75
R-squared: 0.898,  Adjusted R-Squared: 0.892
F-statistic vs. constant model: 137, p-value = 6.91e-44
```

应用研究中的典型回归结果

Busse et al, (2015). The Psychological Effect of Weather on Car Purchases. *QJE*, 130(1):371-414.
https://doi.org/10.1093/qje/qju033

TABLE III

EFFECT OF WEATHER ON FOUR-WHEEL-DRIVE PURCHASES

	Dependent variable: indicator equal to 1 if purchase was a four-wheel-drive				
	Full year	Quarter 1	Quarter 2	Quarter 3	Quarter 4
Temperature	−.032**	−.038**	−.018**	−.029**	−.038**
	(.001)	(.002)	(.003)	(.004)	(.003)
Rainfall	.084**	.119**	.081**	.054**	.132**
	(.014)	(.036)	(.026)	(.023)	(.032)
Snowfall	1.81**	1.67**	.72	125*	2.11**
	(.26)	(.33)	(.82)	(53)	(.48)
Slushfall	.504**	.540**	.27	−.029	.769**
	(.077)	(.110)	(.22)	(.167)	(.166)
Cloud cover	.461**	.337**	.512**	.383**	.598**
	(.039)	(.072)	(.077)	(.087)	(.076)
Year fixed effects	X	X	X	X	X
DMA*week-of-the-year fixed effects	X	X	X	X	X
R-squared	0.086	0.086	0.074	0.084	0.097
Observations	39,984,509	9,150,047	10,205,673	10,608,527	10,020,262

Notes. Coefficient values and clustered standard errors are presented from OLS regressions of an indicator for whether a car sold was a four-wheel-drive on weather variables: temperature (°F), rain (inches), snow (liquidized inches), slush (liquidized inches), and cloud cover (fraction of sky covered). Fixed effects are included for each year and for DMA*week-of-the-year (week 1–week 52). The first column uses all the data and the next four columns present results separately for the four quarters of the year. All coefficients and standard errors have been multiplied by 100 for ease of presentation. Thus each coefficient represents the percentage point change in probability of purchasing a covertible. Standard errors are clustered at the DMA*day. * significant at 5%; ** significant at 1%.

假设检验的基础知识

假设检验的思维方式

The Idea of Hypothesis Testing

考虑回归模型

$$y_t = \beta + u_t\ ,\quad u_t \sim \mathrm{IID}(0,\sigma^2)$$

此时 OLS 估计量满足 $\hat{\beta} = \frac{1}{n}\sum_{t=1}^{n} y_t$，$\mathrm{Var}[\hat{\beta}] = \frac{1}{n}\sigma^2$。

我们想知道总体中的 β 是否满足某种限制条件，例如

$$H_0:\ \beta = \beta_0 \quad \text{vs} \quad H_1:\ \beta \neq \beta_0$$

H_0 是零假设（null hypothesis），H_1 是备择假设（alternative hypothesis）

我们要针对 H_0 进行检验，就是要找到一个检验统计量（test statistic），并使其满足：

1. 当零假设正确时，我们知道该统计量服从分布 A；
2. 当零假设错误（即备择假设正确）时，我们知道该统计量服从分布 $B \neq A$。

如果零假设正确时很难获得样本中检验统计量的取值的话，我们就有理由相信零假设是错误的。当理由充分时，我们即可拒绝（reject）零假设，而偏向于接受（accept）备择假设。

统计量的分布

Distribution of Test Statistic

首先我们假定零假设 H_0： $\beta = \beta_0$ 成立（这里的 β_0 就是总体中 β 的真实值）。同时我们假设 u_t 服从正态分布且 σ 已知。

一个常用的检验统计量是

$$z = \frac{\hat{\beta} - \beta_0}{\sqrt{\mathrm{Var}[\hat{\beta}]}} = \frac{\sqrt{n}}{\sigma}(\hat{\beta} - \beta_0)$$

因为 $\hat{\beta}$ 非偏，则在零假设下 $E[z] = 0$。z 的方差为

$$\mathrm{Var}[z] = E[z^2] = \frac{n}{\sigma^2}E[(\hat{\beta} - \beta_0)^2] = \frac{n}{\sigma^2}\frac{\sigma^2}{n} = 1$$

如果 u_t 服从正态分布，z 也服从正态分布，所以 $z \sim N(0,1)$。

如果备择假设 H_1： $\beta \neq \beta_0$ 成立，那针对任意 $\beta = \beta_1 \neq \beta_0$，则有 $\hat{\beta} = \beta_1 + \hat{\gamma}$。可知 $\hat{\gamma}$ 服从 $N(0, \sigma^2/n)$，因此 $z \sim N(\lambda,1)$， $\lambda = \frac{\sqrt{n}}{\sigma}(\beta_1 - \beta_0)$。

拒绝域和接受域

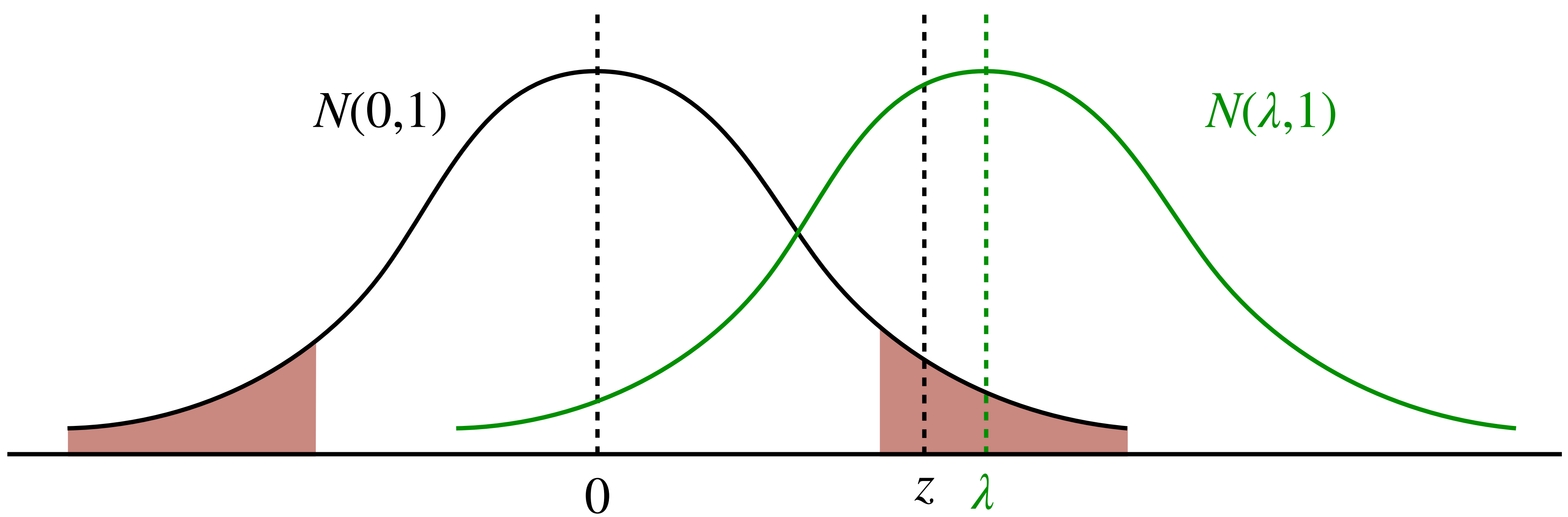
Rejection and Acceptance Regions

$$H_0: \ \beta = \beta_0 \quad \Rightarrow \quad z \sim N(0,1)$$

$$H_1: \ \beta \neq \beta_0 \quad \Rightarrow \quad \text{对任意的}\, \beta = \beta_1, \ z \sim N(\lambda, 1), \ \lambda = \frac{\sqrt{n}}{\sigma}(\beta_1 - \beta_0)$$

当 n 足够大时，则在备择假设下，大概率可以观测到 z 的取值显著不为零。如果我们确实观测到 $|z| \gg 0$，就可以拒绝零假设。

我们需要事先规定一个拒绝零假设的规则。通常我们设定一个拒绝域（rejection region），当 z 的取值在该域中时，就拒绝零假设。检验（test）= 检验统计量 + 拒绝规则。拒绝域以外的区域是接受域（acceptance region）。



检验可能出错

因为样本的随机性，所有检验都可能出现错误。

通常存在两种检验错误：

	检验结果	
	接受零假设	**拒绝零假设**
零假设成立	正确	Type I error
备择假设成立	Type II error	正确

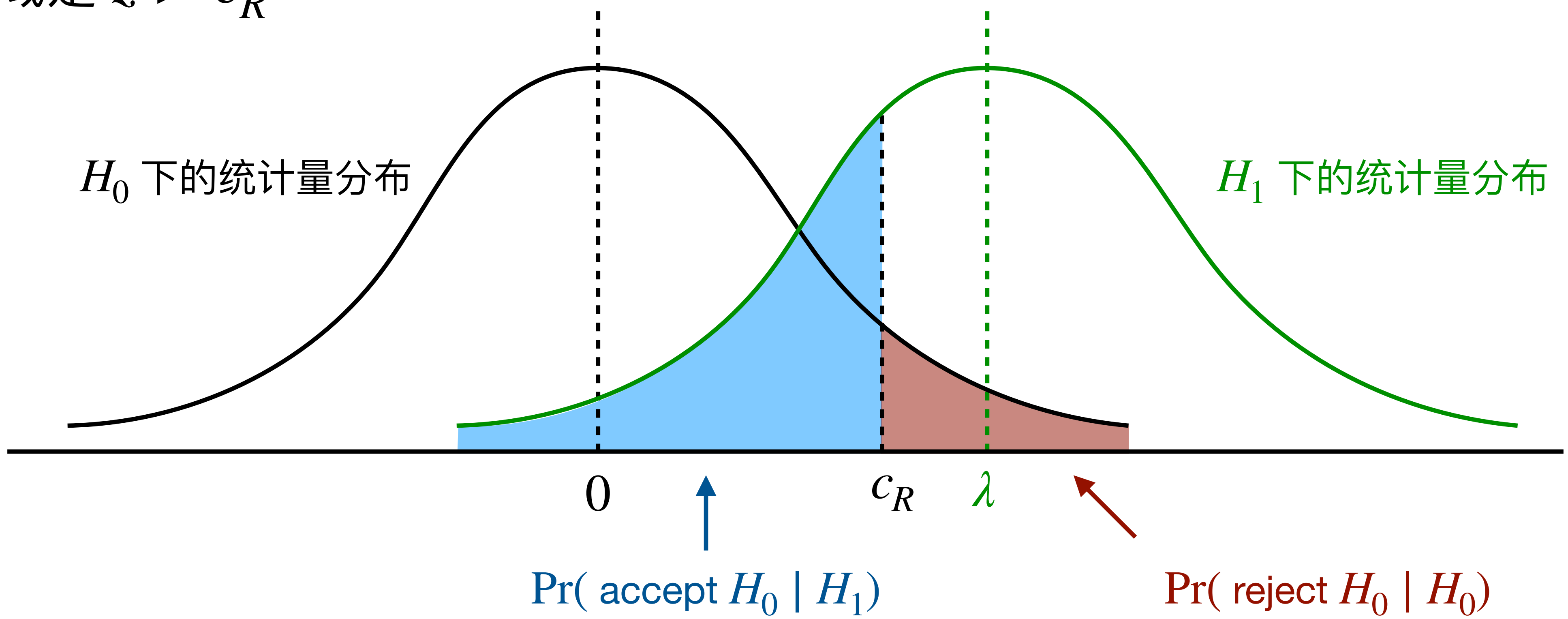
检验的 size：$\Pr($ type I error $) = \Pr($ reject $H_0 \mid H_0)$

检验的功效（power）：$1 - \Pr($ type II error$) = 1 - \Pr($ accept $H_0 \mid H_1)$

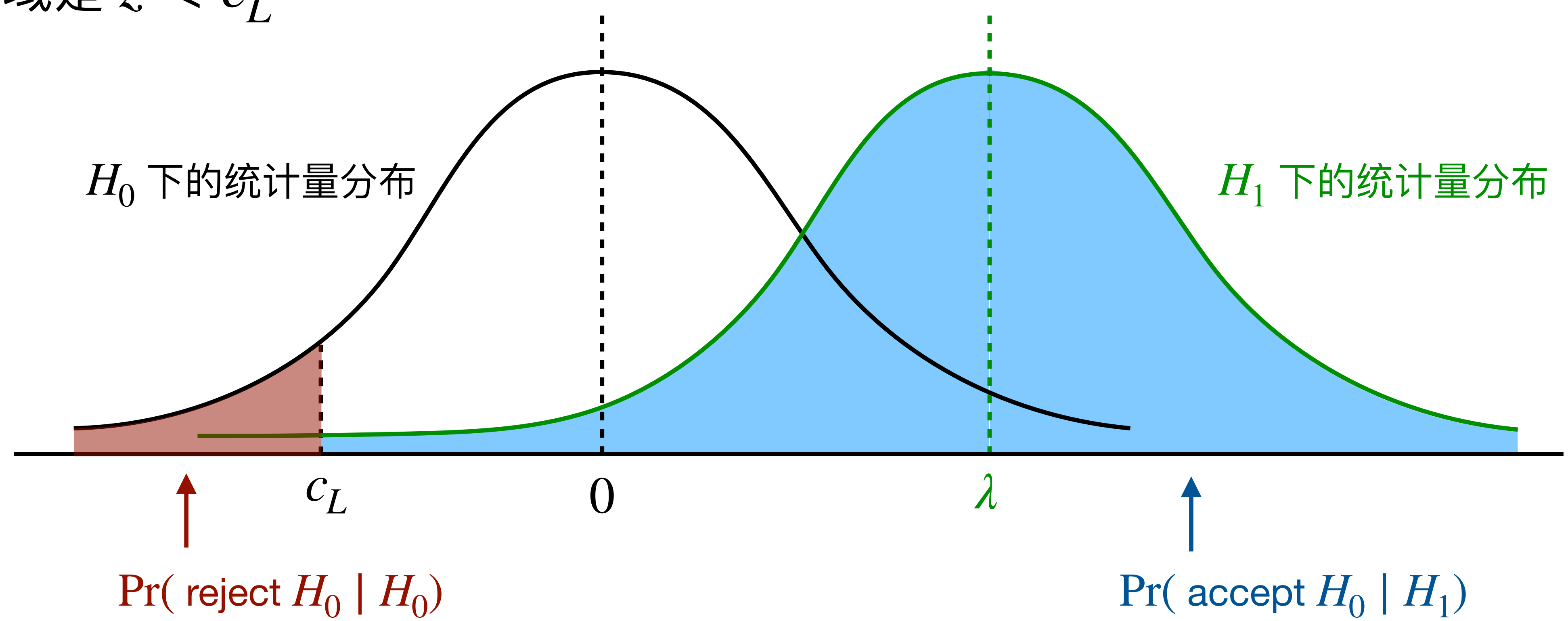
检验的拒绝域越小，size 也就越小，但同时功效也越小，因此无法同时减小两种错误的概率。我们通常选择一个可以接受的 size，然后选择功效最大的检验方法（the Neyman-Pearson approach）。

研究者可以接受的最大 size 值（允许 type I error 发生的最大概率）称为检验的显著性水平（level of significance），通常记作 α。

当拒绝域是 $z > c_R$



当拒绝域是 $z < c_L$



临界值和 *P* 值

Critical Value and the *P* Value

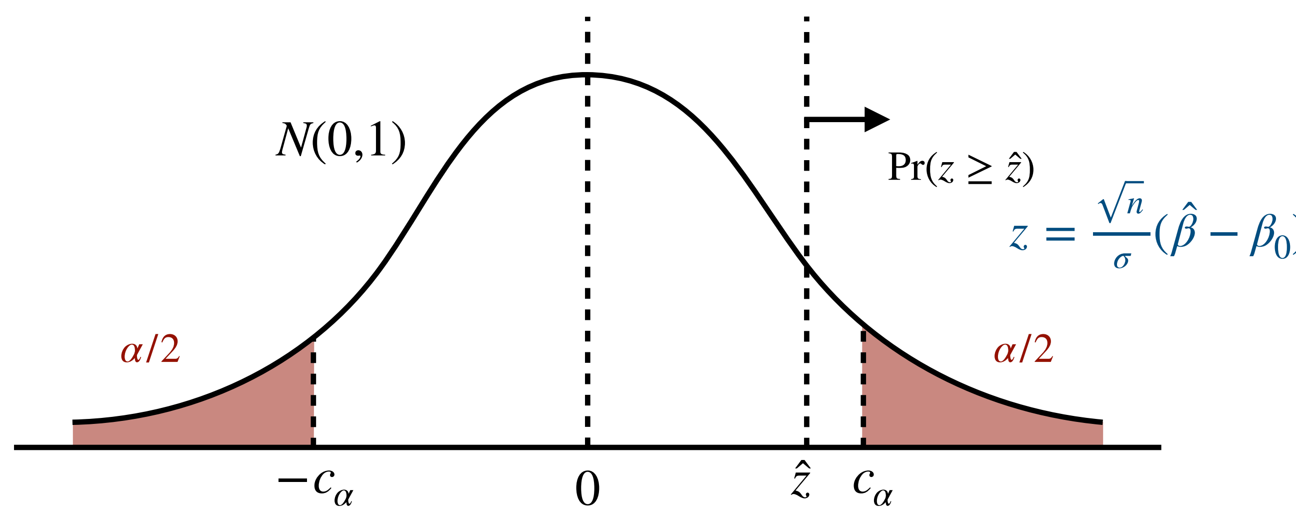
当给出检验的显著性水平 α 时，我们可以通过确定临界值的方式给出该检验的拒绝域。

z 统计量服从标准正态分布，则双尾检验的临界值 c_α 可以通过下面的隐函数定义

$$\Phi(c_\alpha) = 1-\frac{\alpha}{2}$$

$$\Leftrightarrow \quad c_\alpha = \Phi^{-1}(1-\frac{\alpha}{2})$$

Φ 是标准正态分布的累积分布函数



当我们观测到估计值 $\hat{z}$ 时，我们将 $2\mathrm{Pr}(z > |\hat{z}|)$ 定义为该检验的 *P* 值（marginal/observed significance level）。*P* 值是临界值为 $\hat{z}$ 的检验所对应的显著性水平。

通过比较 *P* 值和 α 的大小，可以直观地判断检验结果（拒绝或接受零假设）。

P 值受样本量的影响。样本量 n 越大，$\hat{z}$ 的取值就越大，*P* 值就越小。因此，在大样本下，我们应当审视 5% 或 1% 常用显著性水平的合理性（应当选择更小的 α），并关注估计值的经济学含义（如何解释回归系数）。

常用的概率分布

正态分布

The Normal Distribution

正态分布也称高斯分布（Gaussian distribution），由期望值 μ 和方差 σ^2 两个参数决定。

$N(\mu, \sigma^2)$ 的密度函数是

$$f(x \mid \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$$

如果 $x \sim N(\mu, \sigma^2)$，则 $z = \frac{x-\mu}{\sigma} \sim N(0,1)$。

- 服从正态分布的独立变量的线性结合也服从正态分布。（pp.130-132）

多变量正态分布

The Multivariate Normal Distribution

随机向量 $\boldsymbol{x} = (x_1, \ldots, x_m)$ 服从期望值为 $\boldsymbol{\mu}$，协方差矩阵为 $\boldsymbol{\Omega}$ 的多变量正态分布可以表达为 $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Omega})$。其密度函数为

$$f(\boldsymbol{x} \mid \boldsymbol{\mu}, \boldsymbol{\Omega}) = \frac{1}{(2\pi)^{m/2} \lvert \boldsymbol{\Omega} \rvert^{1/2}} \exp\Big(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{\mu})^\top \boldsymbol{\Omega}^{-1}(\boldsymbol{x} - \boldsymbol{\mu})\Big)$$

因为 $\boldsymbol{\Omega}$ 是协方差矩阵，因此是半正定。
又因 $\boldsymbol{\Omega}$ 可逆，因此是正定矩阵。

- 如果 $\boldsymbol{x} \sim N(\boldsymbol{\mu}, \boldsymbol{\Omega})$，则 $\boldsymbol{a}^\top \boldsymbol{x} \sim N(\boldsymbol{a}^\top \boldsymbol{\mu}, \boldsymbol{a}^\top \boldsymbol{\Omega} \boldsymbol{a})$。
- 如果 $\boldsymbol{x}$ 服从多变量正态分布，且协方差都为零（要素间相互不相关），则 $x_1, \ldots, x_m$ 相互独立。

相互独立（mutual independence）：
$f(x_1, x_2, \ldots, x_m) = f(x_1)f(x_2)\cdots f(x_m)$

卡方分布

The Chi-squared Distribution

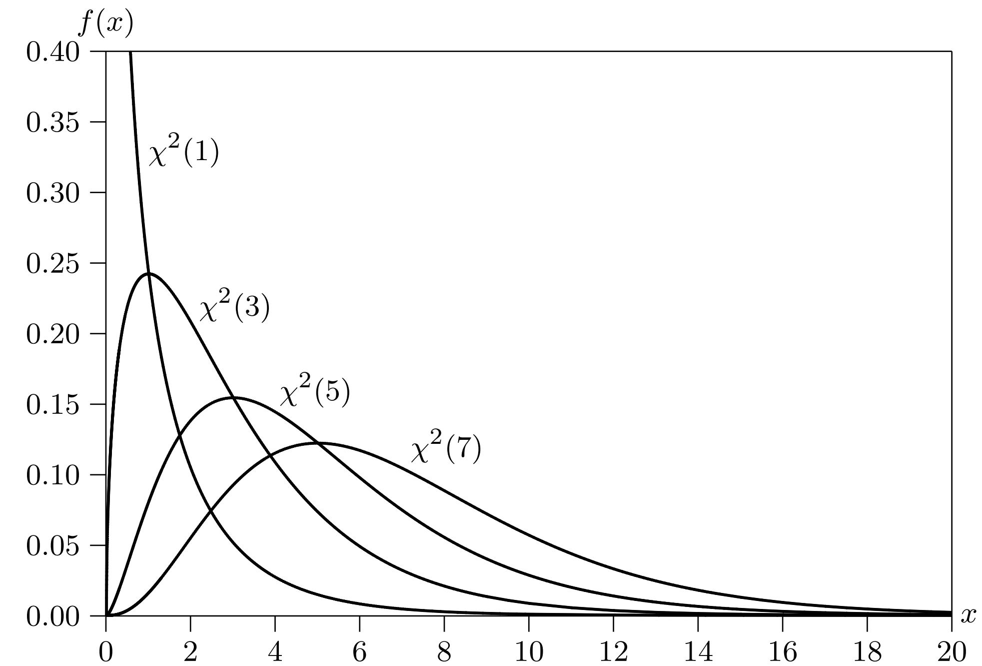


Figure 4.4 Various chi-squared PDFs

当随机向量 $\boldsymbol{z} \sim N(\boldsymbol{0}, \boldsymbol{I})$ 时，随机变量

$$y \equiv \|\boldsymbol{z}\|^2 = \boldsymbol{z}^\top \boldsymbol{z} = \sum_{t=1}^{m} z_t^2$$

服从自由度为 m 的卡方分布（chi-squared distribution with m degrees of freedom），写成 $y \sim \chi^2(m)$。

- $E[y] = \sum_{t=1}^{m} E[z_t^2] = \sum_{t=1}^{m} 1 = m$
- $\mathrm{Var}[y] = \sum_{t=1}^{m} \mathrm{Var}[z_t^2] = mE[(z_t^2 - 1)^2]$
  $= mE[z_t^4 - 2z_t^2 + 1] = m(3 - 2 + 1) = 2m$
- $y_1 \sim \chi^2(m_1),\ y_2 \sim \chi^2(m_2) \quad \Rightarrow \quad (y_1 + y_2) \sim \chi^2(m_1 + m_2)$

服从卡方分布的随机变量

定理 4.1

1. 如果长度为 m 的随机向量 $\boldsymbol{x} \sim N(\mathbf{0}, \boldsymbol{\Omega})$，则 $\boldsymbol{x}^\top \boldsymbol{\Omega}^{-1} \boldsymbol{x} \sim \chi^2(m)$；
2. 如果 $\boldsymbol{P}$ 是投影矩阵且 $\operatorname{rank}(\boldsymbol{P}) = r$，$\boldsymbol{z} \sim N(\mathbf{0}, \boldsymbol{I})$ 是长度为 n 的随机向量，则 $\boldsymbol{z}^\top \boldsymbol{P} \boldsymbol{z} \sim \chi^2(r)$。

证明：

1. $\boldsymbol{\Omega}$ 是对称正定矩阵，因此存在非奇异矩阵 $\boldsymbol{A}$ 满足 $\boldsymbol{\Omega} = \boldsymbol{A}\boldsymbol{A}^\top$（Section 3.4）。此时考虑 $\boldsymbol{z} = \boldsymbol{A}^{-1}\boldsymbol{x}$：$E[\boldsymbol{z}] = E[\boldsymbol{A}^{-1}\boldsymbol{x}] = \boldsymbol{A}^{-1}E[\boldsymbol{x}] = \mathbf{0}$，且

$$\begin{aligned}\operatorname{Var}[\boldsymbol{z}] &= \operatorname{Var}[\boldsymbol{A}^{-1}\boldsymbol{x}] = E[\boldsymbol{A}^{-1}\boldsymbol{x}\boldsymbol{x}^\top(\boldsymbol{A}^\top)^{-1}] \\ &= \boldsymbol{A}^{-1}E[\boldsymbol{x}\boldsymbol{x}^\top](\boldsymbol{A}^\top)^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A}\boldsymbol{A}^\top(\boldsymbol{A}^\top)^{-1} = \boldsymbol{I}_m\end{aligned}$$

   因此 $\boldsymbol{z} \sim N(\mathbf{0}, \boldsymbol{I})$。而 $\boldsymbol{x}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{x} = \boldsymbol{x}^\top(\boldsymbol{A}^\top)^{-1}\boldsymbol{A}^{-1}\boldsymbol{x} = \boldsymbol{z}^\top\boldsymbol{z}$，所以 $\boldsymbol{x}^\top\boldsymbol{\Omega}^{-1}\boldsymbol{x} \sim \chi^2(m)$。

2. 假设 $\boldsymbol{P}$ 投影到 $n \times r$ 矩阵 $\boldsymbol{Z}$ 的列空间，因此 $\boldsymbol{P} = \boldsymbol{Z}(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top$，

$$\boldsymbol{z}^\top\boldsymbol{P}\boldsymbol{z} = \boldsymbol{z}^\top\boldsymbol{Z}(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{Z}^\top\boldsymbol{z}$$

   如果令 $\boldsymbol{x} = \boldsymbol{Z}^\top\boldsymbol{z}$，则 $E[\boldsymbol{x}] = \mathbf{0}$, $\operatorname{Var}[\boldsymbol{x}] = E[\boldsymbol{Z}^\top\boldsymbol{z}\boldsymbol{z}^\top\boldsymbol{Z}] = \boldsymbol{Z}^\top\boldsymbol{Z}$，因此 $\boldsymbol{x} \sim N(\mathbf{0}, \boldsymbol{Z}^\top\boldsymbol{Z})$。可见 $\boldsymbol{z}^\top\boldsymbol{P}\boldsymbol{z} = \boldsymbol{x}^\top(\boldsymbol{Z}^\top\boldsymbol{Z})^{-1}\boldsymbol{x} \sim \chi^2(r)$。

Student's *t* 分布

The Student's *t* Distribution

如果 $z \sim N(0,1)$，$y \sim \chi^2(m)$，且 z 和 y 相互独立，则

$$t_m \equiv \frac{z}{\sqrt{\frac{y}{m}}}$$

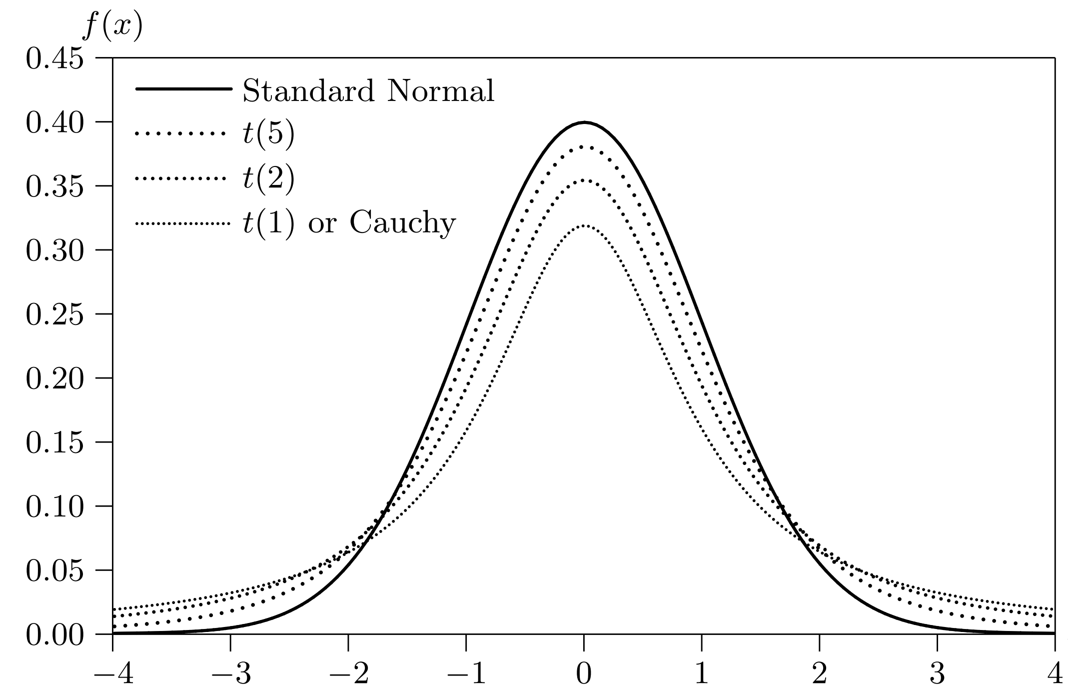


Figure 4.5 PDFs of the Student's *t* distribution

服从自由度为 m 的 Student's *t* 分布，写成 $t_m \sim t(m)$。

- 自由度为 m 的 *t* 分布存在前 $m-1$ 个矩。因此 $t(1)$ 没有矩，$t(2)$ 仅有期望值。
- 若存在，则 $E[t_m]=0$，$\mathrm{Var}[t_m]=m\big/(m-2)$。
- 当 $m \to \infty$ 时，*t* 分布的方差趋近于 1，分布自身趋近于正态分布。

F 分布

The *F* Distribution

当 $y_1 \sim \chi^2(m_1)$，$y_2 \sim \chi^2(m_2)$，且 y_1 与 y_2 相互独立时，

$$F_{m_1,m_2} = \frac{y_1 \big/ m_1}{y_2 \big/ m_2}$$

服从自由度为 m_1 和 m_2 的 *F* 分布，写成 $F_{m_1,m_2} \sim F(m_1, m_2)$。

- $E[F_{m_1,m_2}] = (m_2 - 1) \big/ (m_2 - 3)$。
- 当 $m_2 \to \infty$ 时，$m_1 F_{m_1,m_2} \to y_1 \sim \chi^2(m_1)$。
- 由 *t* 分布的定义可知，当 $x \sim t(m_2)$ 时，$x^2 \sim F(1, m_2)$。

古典正态模型下的精确检验

精确检验与古典正态模型

Exact Tests and the Classical Normal Linear Model

当检验统计量在 H_0 下的分布为已知时，该检验被称为精确检验（exact test）。

假设古典正态线性模型：

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u} \ , \quad \boldsymbol{u} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$$

其中 $\boldsymbol{u}$ 独立于 $\boldsymbol{X}$（$\boldsymbol{X}$ 固定或在 $\boldsymbol{y}$ 之前生成）。

“古典” 一般指 $\boldsymbol{X}$ 为固定变量或 $\boldsymbol{X}$ 与 $\boldsymbol{u}$ 独立，在此假设下我们只需要考虑普通的期望值。“正态” 指 $\boldsymbol{u}$ 服从正态分布。只有在针对 $\boldsymbol{\beta}$ 进行假设检验时才需要关于分布的完整信息。

单一约束的检验

Tests for a Single Restriction

首先我们考虑针对单个系数的假设检验。我们可以把系数向量写成 $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^\top, \beta_2]^\top$，并针对 $H_0 : \beta_2 = 0$ 进行检验。

这里的 $\beta_2 = 0$ 可以看作系数的线性约束条件，而零假设和备择假设分别是在有约束和没有约束下的回归。

原回归模型可以写成

$$\boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{x}_2\beta_2 + \boldsymbol{u} \ , \quad \boldsymbol{u} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$$

$\boldsymbol{X}_1$ 是 $n \times (k-1)$ 矩阵，$\boldsymbol{x}_2$ 是 $n \times 1$ 向量，$\boldsymbol{X} = [\boldsymbol{X}_1, \boldsymbol{x}_2]$。

根据 FWL 定理，$\hat{\beta}_2$ 和 $\hat{\boldsymbol{u}}$ 可以从下面的短模型获得

$$\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{M}_1\boldsymbol{x}_2\beta_2 + \boldsymbol{v} \ , \quad \boldsymbol{M}_1 = \boldsymbol{I} - \boldsymbol{X}_1(\boldsymbol{X}_1^\top\boldsymbol{X}_1)^{-1}\boldsymbol{X}_1^\top$$

此时可得 $\hat{\beta}_2 = \dfrac{\boldsymbol{x}_2^\top\boldsymbol{M}_1\boldsymbol{y}}{\boldsymbol{x}_2^\top\boldsymbol{M}_1\boldsymbol{x}_2}, \qquad \mathrm{Var}[\hat{\beta}_2] = \dfrac{\sigma^2}{\boldsymbol{x}_2^\top\boldsymbol{M}_1\boldsymbol{x}_2}$

单一约束的检验：σ^2 已知

当 σ^2 已知时，我们可以用下面的统计量检验 $H_0 : \beta_2 = 0$

$$z_{\beta_2} \equiv \frac{\hat{\beta}_2 - 0}{\sqrt{\mathrm{Var}[\hat{\beta}_2]}} = \frac{\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{y}}{\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2} \cdot \frac{\sqrt{\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2}}{\sigma} = \frac{\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{y}}{\sigma\sqrt{\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2}}$$

当 $H_0 : \beta_2 = 0$ 成立时，$\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{M}_1\boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{M}_1\boldsymbol{u} = \boldsymbol{M}_1\boldsymbol{u}$，因此

$$z_{\beta_2} = \frac{\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{u}}{\sigma\sqrt{\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2}}$$

z_{β_2} 是 $u_1, \ldots, u_n$ 的线性结合，因此服从正态分布。$E[z_{\beta_2}] = 0$,

$$\mathrm{Var}[z_{\beta_2}] = \frac{E[\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{u}\boldsymbol{u}^\top \boldsymbol{M}_1 \boldsymbol{x}_2]}{\sigma^2 \boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2} = \frac{\boldsymbol{x}_2^\top \boldsymbol{M}_1 E[\boldsymbol{u}\boldsymbol{u}^\top] \boldsymbol{M}_1 \boldsymbol{x}_2}{\sigma^2 \boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2} = \frac{\sigma^2 \boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2}{\sigma^2 \boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2} = 1$$

所以 $z_{\beta_2} \sim N(0,1)$。

单一约束的检验：σ^2 未知

当 σ^2 未知时，我们可以用回归标准误 s 替代 σ。因为 $s^2 = \frac{1}{n-k}\sum_{t=1}^{n} \hat{u}_t^2 = \frac{\boldsymbol{y}^\top \boldsymbol{M}_X \boldsymbol{y}}{n-k}$，我们可以定义下面的统计量

$$t_{\beta_2} \equiv \frac{\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{y}}{s\sqrt{\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2}} = \sqrt{\frac{n-k}{\boldsymbol{y}^\top \boldsymbol{M}_X \boldsymbol{y}}} \cdot \frac{\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{y}}{\sqrt{\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{x}_2}} = \frac{z_{\beta_2}}{\sqrt{\frac{\boldsymbol{y}^\top \boldsymbol{M}_X \boldsymbol{y}}{(n-k)\sigma^2}}}$$

已知 $z_{\beta_2} \sim N(0,1)$，如果我们能证明 $\boldsymbol{y}^\top \boldsymbol{M}_X \boldsymbol{y} \big/ \sigma^2 \sim \chi^2(n-k)$，且 z_{β_2} 和 $\boldsymbol{y}^\top \boldsymbol{M}_X \boldsymbol{y} \big/ \sigma^2$ 独立，即可得到 $t_{\beta_2} \sim t(n-k)$。

$\boldsymbol{y}^\top \boldsymbol{M}_X \boldsymbol{y} \big/ \sigma^2 \sim \chi^2(n-k)$：因为 $\boldsymbol{M}_X$ 是投影到 $\mathcal{S}^\perp(\boldsymbol{X})$ 的投影矩阵，所以

$$\frac{\boldsymbol{y}^\top \boldsymbol{M}_X \boldsymbol{y}}{\sigma^2} = \frac{\boldsymbol{u}^\top \boldsymbol{M}_X \boldsymbol{u}}{\sigma^2} = \boldsymbol{\varepsilon}^\top \boldsymbol{M}_X \boldsymbol{\varepsilon}, \qquad \boldsymbol{\varepsilon} = \boldsymbol{u}/\sigma \sim N(\boldsymbol{0}, \boldsymbol{I})$$

$\dim(\mathcal{S}^\perp(\boldsymbol{X})) = n-k \;\Rightarrow\; \operatorname{rank}(\boldsymbol{M}_X) = n-k$，因此 $\boldsymbol{y}^\top \boldsymbol{M}_X \boldsymbol{y} \big/ \sigma^2 \sim \chi^2(n-k)$。

单一约束的检验：σ^2 未知

z_{β_2} 和 $\boldsymbol{y}^\top \boldsymbol{M}_X \boldsymbol{y} \big/ \sigma^2$ 独立：两个变量中随机要素全都来自 $\boldsymbol{u}$。$\boldsymbol{y}^\top \boldsymbol{M}_X \boldsymbol{y} \big/ \sigma^2$ 只通过 $\boldsymbol{M}_X \boldsymbol{u}$ 受 $\boldsymbol{u}$ 的影响，而由

$$\boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{u} = \boldsymbol{x}_2^\top \boldsymbol{P}_X \boldsymbol{M}_1 \boldsymbol{u} = \boldsymbol{x}_2^\top (\boldsymbol{P}_X - \boldsymbol{P}_X \boldsymbol{P}_1) \boldsymbol{u} \overset{\boldsymbol{P}_X \boldsymbol{P}_1 = \boldsymbol{P}_1 \boldsymbol{P}_X}{=} \boldsymbol{x}_2^\top \boldsymbol{M}_1 \boldsymbol{P}_X \boldsymbol{u}$$

可知 z_{β_2} 只通过 $\boldsymbol{P}_X \boldsymbol{u}$ 受 $\boldsymbol{u}$ 的影响。

$\boldsymbol{M}_X \boldsymbol{u}$ 与 $\boldsymbol{P}_X \boldsymbol{u}$ 的期望值都为 $\boldsymbol{0}$，因此它们的协方差是

$$E[\boldsymbol{M}_X \boldsymbol{u} \boldsymbol{u}^\top \boldsymbol{P}_X] = \boldsymbol{M}_X E[\boldsymbol{u} \boldsymbol{u}^\top] \boldsymbol{P}_X = \sigma^2 \boldsymbol{M}_X \boldsymbol{P}_X \overset{\boldsymbol{P}_X \boldsymbol{M}_X = \boldsymbol{O}}{=} \boldsymbol{O}$$

因此 $\boldsymbol{M}_X \boldsymbol{u}$ 与 $\boldsymbol{P}_X \boldsymbol{u}$ 不相关。因为 $\boldsymbol{M}_X \boldsymbol{u}$ 与 $\boldsymbol{P}_X \boldsymbol{u}$ 都服从正态分布，所以 $\boldsymbol{M}_X \boldsymbol{u}$ 独立于 $\boldsymbol{P}_X \boldsymbol{u}$，所以。$\boldsymbol{y}^\top \boldsymbol{M}_X \boldsymbol{y} \big/ \sigma^2$ 独立于 z_{β_2}。

综上，我们可得 $t_{\beta_2} \sim t(n-k)$。

多重约束的联合检验

Joint Tests for Multiple Restrictions

当存在 r 个线性约束条件时（$r \leq k$），我们可以将其写成 $H_0 : \boldsymbol{\beta}_2 = \mathbf{0}$。此时，在不同的假设下我们可以得到下面的回归模型：

$$
\begin{aligned}
H_0 &: \boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{u} \ , \quad \boldsymbol{u} \sim N(\mathbf{0}, \sigma^2\boldsymbol{I}) \\
H_1 &: \boldsymbol{y} = \boldsymbol{X}_1\boldsymbol{\beta}_1 + \boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{u} \ , \quad \boldsymbol{u} \sim N(\mathbf{0}, \sigma^2\boldsymbol{I})
\end{aligned}
$$

因为需要检验的参数大于 1，我们不能继续使用 z 检验或 t 检验，而需要找到另一个合适的检验统计量。这里我们用有约束（restricted）回归和无约束（unrestricted）回归的拟合（fit）结果来建立检验统计量，并用两个回归的残差平方和作为拟合的指标。

若将有约束回归的残差平方和写为 RSSR，将无约束回归的写为 USSR，我们可以定义下面的统计量

$$
F_{\boldsymbol{\beta}_2} \equiv \frac{(\mathrm{RSSR} - \mathrm{USSR})\big/r}{\mathrm{USSR}\big/(n-k)}
$$

如果 H_0 不成立，RSSR 将大于 USSR。
因此当 $F_{\boldsymbol{\beta}_2}$ 取值很大时，我们倾向于拒绝 H_0。

多重约束的联合检验

Joint Tests for Multiple Restrictions

我们将要证明 $F_{\beta_2} \equiv \dfrac{(\mathrm{RSSR}-\mathrm{USSR})/r}{\mathrm{USSR}/(n-k)}$ 在 H_0 下服从 $F(r, n-k)$。

$\mathrm{RSSR} = \boldsymbol{y}^\top \boldsymbol{M}_1 \boldsymbol{y}$，$\mathrm{USSR} = \boldsymbol{y}^\top \boldsymbol{M}_X \boldsymbol{y}$。根据 FWL 定理，USSR 等于短模型 $\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{M}_1\boldsymbol{X}_2\boldsymbol{\beta}_2 + \boldsymbol{v}$ 的残差平方和

$$\hat{\boldsymbol{v}}^\top \hat{\boldsymbol{v}} = \boldsymbol{y}^\top \boldsymbol{M}_1 \boldsymbol{y} - \boldsymbol{y}^\top \boldsymbol{M}_1 \boldsymbol{X}_2 (\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{y}$$

勾股定理

因此，$\mathrm{RSSR} - \mathrm{USSR} = \boldsymbol{y}^\top \boldsymbol{M}_1 \boldsymbol{X}_2 (\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{y}$。一般情况下 $\boldsymbol{M}_X\boldsymbol{y} = \boldsymbol{M}_X\boldsymbol{u}$，在 H_0 成立时 $\boldsymbol{M}_1\boldsymbol{y} = \boldsymbol{M}_1\boldsymbol{u}$，则

$$F_{\beta_2} = \frac{\boldsymbol{\varepsilon}^\top \boldsymbol{M}_1 \boldsymbol{X}_2 (\boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{X}_2)^{-1} \boldsymbol{X}_2^\top \boldsymbol{M}_1 \boldsymbol{\varepsilon} \big/ r}{\boldsymbol{\varepsilon}^\top \boldsymbol{M}_X \boldsymbol{\varepsilon} \big/ (n-k)}, \qquad \boldsymbol{\varepsilon} = \boldsymbol{u}/\sigma \sim N(\boldsymbol{0}, \boldsymbol{I})$$

$\mathrm{rank}(\boldsymbol{X}_2) = r$

已知 $\boldsymbol{\varepsilon}^\top \boldsymbol{M}_X \boldsymbol{\varepsilon} \sim \chi^2(n-k)$，而分子可以写成 $\boldsymbol{\varepsilon}^\top \boldsymbol{P}_{M_1X_2} \boldsymbol{\varepsilon}$ 并服从 $\chi^2(r)$。又因为 $\boldsymbol{M}_X$ 和 $\boldsymbol{P}_{M_1X_2}$ 正交，加上 $\boldsymbol{u}$ 服从正态分布可得分子与分母相互独立。因此 $F_{\beta_2} \sim F(r, n-k)$。