Advanced Econometrics

Lecture 9: Hypothesis Testing (Hansen Chapter 9)

Instructor: Ma, Jun

Renmin University of China

December 4, 2018

Hypotheses

$Y = AK^{\alpha}L^{\beta}$

$\log Y = \log A + \alpha \log K + \beta \log L$

$H_0: \alpha + \beta = 1$ $\quad$ $\theta = \alpha + \beta$

$H_1: \alpha + \beta \neq 1$

- Hypothesis tests attempt to assess whether there is evidence to contradict a proposed parametric restriction.
- Let $\boldsymbol{\theta} = \boldsymbol{r}(\boldsymbol{\beta})$ be a $q \times 1$ parameter of interest where $\boldsymbol{r} : \mathbb{R}^k \to \Theta \subset \mathbb{R}^q$ is some transformation.
- A point hypothesis concerning $\boldsymbol{\theta}$ is a proposed restriction such as $\boldsymbol{\theta} = \boldsymbol{\theta}_0$, where $\boldsymbol{\theta}_0$ is a hypothesized (known) value.
- A hypothesis is a restriction $\boldsymbol{\beta} \in \boldsymbol{B}_0$. In the case of the hypothesis $\boldsymbol{r}(\boldsymbol{\beta}) = \boldsymbol{\theta}_0$, $\boldsymbol{B}_0 = \{\boldsymbol{\beta} : \boldsymbol{r}(\boldsymbol{\beta}) = \boldsymbol{\theta}_0\}$.

Hypotheses

Definition
The **null hypothesis**, written $\mathbb{H}_0$, is the restriction $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ or $\boldsymbol{\beta} \in \boldsymbol{B}_0$.

- We often write the null hypothesis as $\mathbb{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ or $\mathbb{H}_0 : \boldsymbol{r}(\boldsymbol{\beta}) = \boldsymbol{\theta}_0$.

Definition
The **alternative hypothesis**, written $\mathbb{H}_1$, is the set $\{\boldsymbol{\theta} \in \boldsymbol{\Theta} : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0\}$ or $\{\boldsymbol{\beta} : \boldsymbol{\beta} \notin \boldsymbol{B}_0\}$

- We often write the alternative hypothesis as $\mathbb{H}_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$ or $\mathbb{H}_1 : \boldsymbol{r}(\boldsymbol{\beta}) \neq \boldsymbol{\theta}_0$.
- The goal of hypothesis testing is to assess whether or not $\mathbb{H}_0$ is true, by asking if $\mathbb{H}_0$ is consistent with the observed data.

Acceptance and Rejection

- The decision is based on a function of the data. It is convenient to express this function as a real-valued function called a **test statistic**

$$T = T\left((Y_1, \boldsymbol{X}_1), \ldots, (Y_n, \boldsymbol{X}_n)\right).$$

这个T是我们通常T统计量取了绝对值.

$$T = \left| \frac{\hat{\theta} - \theta_0}{se(\hat{\theta})} \right|$$

- The hypothesis test then consists of the decision rule:

$$\begin{aligned} &\text{Accept } \mathbb{H}_0 \text{ if } T \leq c \\ &\text{Reject } \mathbb{H}_0 \text{ if } T > c. \end{aligned}$$

因此看起来像单侧检验.

- Small values of T are likely when $\mathbb{H}_0$ is true and large values are likely when $\mathbb{H}_1$ is true.

Acceptance and Rejection

- The most commonly used test statistic is the absolute value of the t-statistic $T = |T(\theta_0)|$ where

$$T(\theta) = \frac{\widehat{\theta} - \theta}{s\left(\widehat{\theta}\right)}.$$

  $\widehat{\theta}$ is a point estimate and $s\left(\widehat{\theta}\right)$ is its standard error.

Type I Error

- A false rejection of $\mathbb{H}_0$ (rejecting $\mathbb{H}_0$ when $\mathbb{H}_0$ is true) is called a **Type-I error**. The probability of a Type I error is

$$\Pr(\text{Reject } \mathbb{H}_0 \mid \mathbb{H}_0 \text{ true}) = \Pr(T > c \mid \mathbb{H}_0 \text{ true}).$$

- The first goal is to control the type-I error: it should not be large.
- In typical econometric models the exact sampling distributions of estimators and test statistics are unknown.

I类错误发生的概率不能太大.

$H_0: \beta = \beta_0$

$H_1: \beta \neq \beta_0$

$T = \left| \frac{\hat\theta - \theta_0}{se(\hat\theta)} \right| > c$

$\Rightarrow \frac{\sqrt{n}(\hat\beta - \beta_0)}{\sqrt{n}\, se(\hat\beta)}$

H_0 成立时 $\beta = \beta_0 \Rightarrow$

$\frac{\sqrt{n}(\hat\beta - \beta) \to_d N(0, V_\beta)}{\sqrt{n}\, se(\hat\beta) \to_p \sqrt{V_\beta}} \to N(0,1)$

$\hat\beta \overset{a}{\sim} N(\beta, \frac{V_\beta}{n})$. $se(\hat\beta) = \sqrt{\frac{\hat V_\beta}{n}}$

Type I Error

- Suppose that when $\mathbb{H}_0$ is true,

$$T \xrightarrow{d} \xi.$$

  Let $G(u) = \Pr(\xi \le u)$ be the distribution of ξ. We call G the asymptotic null distribution. In simple cases, G is known and does not depend on unknown parameters.

- We define the **asymptotic size** of the test as the asymptotic probability of a Type I error:

$$\begin{aligned}\lim_{n\to\infty} \Pr(T > c \mid \mathbb{H}_0 \text{ true}) &= \Pr(\xi > c) \\ &= 1 - G(c).\end{aligned}$$

零假设成立下T的渐近分布,
卡方或标准正态的绝对值.

找一个C 使 $1-G(c)=\alpha$

- In the dominant approach to hypothesis testing, the researcher pre-selects a significance level $\alpha \in (0,1)$ and then selects c so that the asymptotic size is no larger than α.

Type I Error

$$\theta = r(\beta)$$

$$T = \frac{\hat{\theta} - \theta_0}{se(\hat{\theta})}$$

$$\sqrt{n}(\hat{\theta} - \theta) = \sqrt{n}\left(r(\hat{\beta}) - r(\beta)\right)$$

$$\approx \sqrt{n}\,\frac{\partial r(\beta)}{\partial \beta'}(\hat{\beta} - \beta)$$

$$\xrightarrow{d} N\left(0, \frac{\partial r(\beta)}{\partial \beta'}(\hat{\beta} - \beta)\frac{\partial r(\beta)'}{\partial \beta}\right)$$

$$\Rightarrow \hat{\theta} \overset{a}{\sim} N\left(\theta, \frac{\frac{\partial r(\beta)}{\partial \beta'} V_\beta \frac{\partial r(\beta)'}{\partial \beta}}{n}\right) \rightarrow \frac{\frac{\partial r(\hat{\beta})}{\partial \hat{\beta}'} \hat{V}_\beta \frac{\partial r(\hat{\beta})'}{\partial \hat{\beta}}}{n}$$

$$se(\hat{\theta}) = \sqrt{\frac{\frac{\partial r(\hat{\beta})}{\partial \hat{\beta}'} \hat{V}_\beta \frac{\partial r(\hat{\beta})'}{\partial \hat{\beta}}}{n}}$$

$$\frac{\partial r(\beta)}{\beta'} \longrightarrow \left.\frac{\partial r(b)}{\partial b'}\right|_{b=\beta}$$

β ← 参数真值

β' ← 哑变量

$$T = \frac{\hat{\theta} - \theta_0}{se(\hat{\theta})} = \frac{\sqrt{n}(\hat{\theta} - \theta_0)}{\sqrt{n}\, se(\hat{\theta})} = \frac{\sqrt{n}(\hat{\theta} - \theta_0)}{\sqrt{\frac{\partial r(\hat{\beta})}{\partial \hat{\beta}'} \hat{V}_\beta \frac{\partial r(\hat{\beta})'}{\partial \hat{\beta}}}}$$

$$\underset{\theta}{\underset{\|}{\sqrt{n}(\hat{\theta} - \theta_0)}} \xrightarrow{d} N\left(0, \frac{\partial r(\beta)}{\partial \beta'} V_\beta \frac{\partial r(\beta)'}{\partial \beta}\right)$$

$$\sqrt{\frac{\partial r(\hat{\beta})}{\partial \hat{\beta}'} \hat{V}_\beta \frac{\partial r(\hat{\beta})'}{\partial \hat{\beta}}} \xrightarrow{p} \sqrt{\frac{\partial r(\beta)}{\partial \beta'} V_\beta \frac{\partial r(\beta)'}{\partial \beta}}$$

t tests

- The most common test of "scalar" hypothesis: $\mathbb{H}_0 : \theta = \theta_0$ against $\mathbb{H}_1 : \theta \neq \theta_0$.

Theorem

Under $\mathbb{H}_0 : \theta = \theta_0$,

$$T(\theta_0) \xrightarrow{d} Z.$$

For c satisfying $\alpha = 2(1 - \Phi(c))$,

$$\Pr(|T(\theta_0)| > c \mid \mathbb{H}_0) \to \alpha,$$

and the test "Reject $\mathbb{H}_0$ if $|T(\theta_0)| > c$" has asymptotic size α.

- The alternative $\theta \neq \theta_0$ is called a two-sided alternative.

$T \to_d Z$

$|T| \to_d |Z|$

$P(|Z| > c) = 2(1 - \Phi(c))$

$Z \sim N(0,1)$

$\Rightarrow \Pr(|T| > c) \to \Pr(|Z| > c)$

$= \underbrace{2(1 - \Phi(c))}_{= \alpha}$

t tests

- One-sided alternative could be $\mathbb{H}_1 : \theta > \theta_0$.
- Tests of $\theta = \theta_0$ against $\theta > \theta_0$ are based on the signed t-statistic $T = T(\theta_0)$.
- We reject $\mathbb{H}_0$ if $T > c$ where c satisfies $\alpha = 1 - \Phi(c)$. Negative values of are not taken as evidence against $\mathbb{H}_0$.
- We should use one-sided tests and critical values only when the parameter space is known to satisfy a one-sided restriction such as $\theta \geq \theta_0$.

单侧假设检验的符号与H_1有关.
H_1是">", 拒绝域就在右边.
H_1是"<", 拒绝域就在左边.

Type II Error and Power 取伪错误

II类错误的概率 Pr(Accept H0 | H1 is true)

- A false acceptance of the null hypothesis $\mathbb{H}_0$ (accepting $\mathbb{H}_0$ when $\mathbb{H}_1$ is true) is called a **Type II error**.
- The rejection probability under the alternative hypothesis is called the power of the test.
- Power = 1 - the probability of a Type II error:

$$\pi(\boldsymbol{\theta}) = \Pr(\text{Reject } \mathbb{H}_0 \mid \mathbb{H}_1 \text{ true}) = \Pr(T > c \mid \mathbb{H}_1 \text{ true})$$

  $\pi(\boldsymbol{\theta})$ is called **power function**. The power depends on the true value of the parameter $\boldsymbol{\theta}$.
- A well behaved test the power is increasing both as $\boldsymbol{\theta}$ moves away from $\boldsymbol{\theta}_0$ and as the sample size n increases.

检验功效 power of the test.

power = 1 − II类错误的概率.

做出正确判断的可能性.

功效函数与参数真值有关.

Type II Error and Power

- Four possibilities:

		Truth	
		H_0	H_1
Decision	H_0	✓	Type II error
	H_1	Type I error	✓

- When $T \leq c$, we accept H_0 (and risk making a Type II error).
- When $T > c$, we reject H_0 (and risk making a Type I error).

如果c很大，那么 T>c 更不容易，降低I类错误

但同时会提高II类错误.

Type II Error and Power

- Unfortunately, the probabilities of Type I and II errors are inversely related.
- By decreasing the probability of Type I error, one makes c larger, which increases the probability of the Type II error. Thus it is impossible to make both errors arbitrary small.
- We want the probability of a type-II error to be as small as possible for a given probability of a type-I error.

希望在控制I类错误的同时，尽可能减小II类错误

选择功效更大的.

P-Values

P值是随机变量.

G是T的渐近分布

- p-value is a measure of the strength of information against the null hypothesis:

$$p = 1 - G\left(T\right).$$

T特别大说明越违背H0. G(·)严格递增 ⇔ p特别小.

  G is the (asymptotic) distribution of T under $\mathbb{H}_0$.

一种理解: P值是边际意义上的显著水平.

- p-value is the marginal significant level: the largest value of α for which the test rejects $\mathbb{H}_0$.
- $T \to_d \xi$ under $\mathbb{H}_0$, then $p = 1 - G\left(T\right) \to_d 1 - G\left(\xi\right)$:

$$\begin{aligned}
\Pr\left(1 - G\left(\xi\right) \leq u\right) &= \Pr\left(1 - u \leq G\left(\xi\right)\right) \\
&= 1 - \Pr\left(\xi \leq G^{-1}\left(1 - u\right)\right) \\
&= 1 - G\left(G^{-1}\left(1 - u\right)\right) \\
&= 1 - \left(1 - u\right) \\
&= u.
\end{aligned}$$

Wald Tests

$\theta = r(\beta)$ $\quad r: \mathbb{R}^K \to \mathbb{R}^q$

$\beta \in \mathbb{R}^K$

$\theta \in \mathbb{R}$. $\quad H_0: \theta = \theta_0$

$H_1: \theta \neq \theta_0$

$|T| = \left| \frac{\hat{\theta} - \theta_0}{se(\hat{\theta})} \right|$

$\hat{\theta} = r(\hat{\beta})$ $\quad H_0: \theta = \theta_0$. $\theta_0 \in \mathbb{R}^q$

$H_1: \theta \neq \theta_0$

$\|\hat{\theta} - \theta_0\|^2 = (\hat{\theta} - \theta_0)' W (\hat{\theta} - \theta_0)$

$\hat{R} = \frac{\partial r(\hat{\beta})}{\partial \beta'}$, $\quad R = \frac{\partial r(\beta)}{\partial \beta'}$

$\sqrt{n}(\hat{\beta} - \beta) \to_d N(0, V_\beta)$

$\sqrt{n}\left(r(\hat{\beta}) - r(\beta)\right) \to_d N(0, R V_\beta R')$

- The parameter of interest is $\boldsymbol{\theta} = \boldsymbol{r}(\boldsymbol{\beta})$. Estimator: $\widehat{\boldsymbol{\theta}} = \boldsymbol{r}\left(\widehat{\boldsymbol{\beta}}\right)$. To test $\mathbb{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$ against $\mathbb{H}_1 : \boldsymbol{\theta} \neq \boldsymbol{\theta}_0$, one approach is to measure the discrepancy $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$:

$$W = n\left(\boldsymbol{r}\left(\widehat{\boldsymbol{\beta}}\right) - \boldsymbol{\theta}_0\right)' \left(\widehat{\boldsymbol{R}}' \widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}} \widehat{\boldsymbol{R}}\right)^{-1} \left(\boldsymbol{r}\left(\widehat{\boldsymbol{\beta}}\right) - \boldsymbol{\theta}_0\right).$$

- When $\boldsymbol{r}(\boldsymbol{\beta}) = \boldsymbol{R}'\boldsymbol{\beta}$, r是线性函数的例子

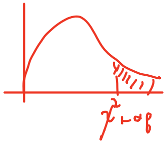
$$W = \left(\boldsymbol{R}'\widehat{\boldsymbol{\beta}} - \boldsymbol{\theta}_0\right)' \left(\boldsymbol{R}' \widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}} \boldsymbol{R}\right)^{-1} \left(\boldsymbol{R}'\widehat{\boldsymbol{\beta}} - \boldsymbol{\theta}_0\right).$$

$\Rightarrow \hat{R} \hat{V}_\beta \hat{R}' \to_p R V_\beta R'$

$\Rightarrow \sqrt{n}\,(r(\hat{\beta}) - \theta_0)' \hat{R} \hat{V}_\beta \hat{R}' \sqrt{n}\,(r(\hat{\beta}) - \theta_0)$

$\parallel$

Wald $\to_d \chi^2_q$ $\quad (H_0: r(\beta) = \theta_0)$



reject H_0 if Wald $> \chi^2_{1-\alpha, q}$

$\Pr(\text{Wald} > \chi^2_{1-\alpha, q}) \to \alpha$

Wald Tests

① Wald 计算无约束最小二乘

② 拉格朗日乘数法计算有约束最小二乘

$\min (Y-Xb)'(Y-Xb)$

s.t. $r(b)=\theta_0$

看乘数大小.

③ 第三种方法. 有约束无约束都算

⇑

三位一体的检验方法

trinity

Theorem

Under $\mathbb{H}_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$,

then

$$W \xrightarrow{d} \chi^2_q,$$

and for c satisfying $\alpha = 1 - G_q(c)$,

$$\Pr(W > c \mid \mathbb{H}_0) \to \alpha$$

so the test "Reject $\mathbb{H}_0$ *if* $W > c$*" has asymptotic size* α.

$= \Pr(\text{type-I}) \to \alpha$

$\forall\, \theta \le \theta_0$

$\lim_{n\to\infty} \Pr(T > z_{1-\alpha}) = \alpha$

↑

渐近 size α 检验

$\boxed{\text{size } \alpha \text{ 检验一定是 level } \alpha \text{ 检验.}}$

一种情况: $H_0: \theta \le \theta_0$

$H_1: \theta > \theta_0$

$T = \dfrac{\hat\theta - \theta_0}{se(\hat\theta)}$

$\lim_{n\to\infty} \Pr(T > z_{1-\alpha}) = \alpha$

$T = \dfrac{\hat\theta - \theta + \theta - \theta_0}{se(\hat\theta)}$

$= \underbrace{\dfrac{\hat\theta - \theta}{se(\hat\theta)}}_{\to_d N(0,1)} + \underbrace{\dfrac{\theta - \theta_0}{se(\hat\theta)}}_{\le 0\ (H_0\text{为真})}$

$\Pr(T > z_{1-\alpha}) = \Pr\left(\dfrac{\hat\theta - \theta}{s.e.} + \dfrac{\theta - \theta_0}{s.e.} > z_{1-\alpha}\right)$

$\le \Pr\left(\dfrac{\hat\theta - \theta}{s.e} > z_{1-\alpha}\right) \to \alpha$

$\forall\, \theta \le \theta_0$

$\limsup_{n\to\infty} \Pr(T > z_{1-\alpha}) \le \alpha$

↑

level α 检验

Homoskedastic Wald Tests

$V_\beta = \sigma^2 (E X X')^{-1}$

$= V_\beta^0$

$\hat{V}_\beta^0 = s^2 \left(\frac{1}{n} \sum x_i x_i'\right)^{-1}$

- If the error is known to be homoskedastic,

$$W^0 = \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)' \left(\widehat{\boldsymbol{V}}^0_{\widehat{\boldsymbol{\theta}}}\right)^{-1} \left(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\right)$$

$$= \left(\boldsymbol{r}\left(\widehat{\boldsymbol{\beta}}\right) - \boldsymbol{\theta}_0\right)' \left(\widehat{\boldsymbol{R}}' \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \widehat{\boldsymbol{R}}\right)^{-1} \left(\boldsymbol{r}\left(\widehat{\boldsymbol{\beta}}\right) - \boldsymbol{\theta}_0\right) / s^2.$$

- In the case of linear hypotheses $\mathbb{H}_0 : \boldsymbol{R}'\boldsymbol{\beta} = \boldsymbol{\theta}_0$,

$$W^0 = \left(\boldsymbol{R}'\widehat{\boldsymbol{\beta}} - \boldsymbol{\theta}_0\right)' \left(\boldsymbol{R}' \left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} \boldsymbol{R}\right)^{-1} \left(\boldsymbol{R}'\widehat{\boldsymbol{\beta}} - \boldsymbol{\theta}_0\right) / s^2.$$

- In this case, the F testing statistic: $F = W^0/q$ and $F \to_d \chi^2_q / q$.

$F \sim F_{q, n-k}$ 对∀n都成立

Power and Test Consistency

- The power of a test is the probability of rejecting $\mathbb{H}_0$ when $\mathbb{H}_1$ is true.
- Random sample from $N\left(\theta, \sigma^2\right)$, with σ^2 known: $\{Y_1, ..., Y_n\}$. For testing $\mathbb{H}_0 : \theta = 0$ against $\mathbb{H}_1 : \theta > 0$,

$$T = \frac{\sqrt{n}\overline{Y}}{\sigma}.$$

We reject $\mathbb{H}_0$ if $T > c$.

- Note $T = \frac{\sqrt{n}\left(\overline{Y} - \theta\right)}{\sigma} + \frac{\sqrt{n}\theta}{\sigma}$. The power of the test is

$$\Pr\left(T > c\right) = \Pr\left(Z + \sqrt{n}\theta/\sigma > c\right) = 1 - \Phi\left(c - \sqrt{n}\theta/\sigma\right).$$

- This power function is monotonically increasing in θ and n.
- If $\theta > 0$, the power increases to 1 as $n \to \infty$. This means whenever $\mathbb{H}_1$ is true, the test will reject $\mathbb{H}_0$ with a high probability if n is sufficiently large.

Power and Test Consistency

△ 记住结论

功率→1时检验是一致的.

> **Definition**
> A test of $\mathbb{H}_0 : \boldsymbol{\theta} \in \boldsymbol{\Theta}_0$ is **consistent against fixed alternatives** if for all $\boldsymbol{\theta} \in \boldsymbol{\Theta}_1$, $\Pr(\text{Reject } \mathbb{H}_0 \mid \boldsymbol{\theta}) \to 1$ as $n \to \infty$.

eg. $X_1, \cdots, X_n$

$H_0: \mu = 0$. $H_1: \mu \neq 0$

$$T = \frac{\bar{X}_n}{\sqrt{\frac{1}{n}\underbrace{\left(\frac{1}{n}\sum (X_i - \bar{X}_n)^2\right)}_{\hat{V}_\theta}}}$$

- In general, t test and Wald test are consistent. Take a t statistic for testing $\mathbb{H}_0 : \theta = \theta_0$,

$$|T| = \left|\frac{\widehat{\theta} - \theta_0}{s\left(\widehat{\theta}\right)}\right| = \left|\underbrace{\frac{\widehat{\theta} - \theta}{s\left(\widehat{\theta}\right)}}_{\to_d N(0,1)} + \underbrace{\frac{\sqrt{n}\,(\theta - \theta_0)}{\sqrt{\widehat{V}_\theta}}}_{\rightarrow H_1\text{为真时不为0.}}\right|.$$

$\sqrt{\widehat{V}_\theta} \to_p C > 0$

reject H_0 if $|T| > z_{1-\frac{\alpha}{2}}$

- $\frac{\widehat{\theta}-\theta}{s(\widehat{\theta})}$ converges in distribution to $N(0,1)$ but $\frac{\sqrt{n}(\theta-\theta_0)}{\sqrt{\widehat{V}_\theta}}$ tends to be large if n is large, since $\sqrt{\widehat{V}_\theta}$ converges in probability to a positive constant.

Bonferroni Corrections

① $E(Y|X) = X'\beta$

$(u = Y - E(Y|X))$

② $E(u^2|X) = \text{Constant}$

H_0: ① 且 ②

- Under the joint hypothesis that a set of k hypotheses are all true, what is the probability that the smallest p-value is smaller than α?
- Suppose our null hypothesis $\mathbb{H}_0$ is a joint hypothesis: "$\mathbb{H}_0^1$ is true, $\mathbb{H}_0^2$ is true, ..., and $\mathbb{H}_0^k$ is true" and for each hypothesis we have a test (a testing statistic with an asymptotic p-value p_j).
- Consider the following rule: reject $\mathbb{H}_0$ if any of the hypotheses is rejected, or the smallest p-value is smaller than α.

让最小的p值小于 α

$H_0: \beta_1 = 0$ 且 $\beta_2 = 0$

$H_0^1: \beta_1 = 0 \quad T_1 = \frac{\hat\beta_1 - 0}{se(\hat\beta_1)}$

$H_0^2: \beta_2 = 0 \quad T_2 = \frac{\hat\beta_2 - 0}{se(\hat\beta_2)}$

reject H_0 if $|T_1| > z_{1-\frac{\alpha}{2}}$

或 $|T_2| > z_{1-\frac{\alpha}{2}}$

$Pr(\text{type-I}) = Pr(|T_1| > z_{1-\frac{\alpha}{2}}$ or $|T_2| > z_{1-\frac{\alpha}{2}} \mid H_0 \text{ is true})$

$= \underbrace{Pr(|T_1| > z_{1-\frac{\alpha}{2}})}_{\to\alpha} + \boxed{Pr(|T_2| > z_{1-\frac{\alpha}{2}}) - Pr(|T_1| > z_{1-\frac{\alpha}{2}} \text{ 且 } |T_2| > z_{1-\frac{\alpha}{2}})}$

> 0

$\Rightarrow Pr(\text{type-I}) \neq \alpha$

Bonferroni Corrections

- But the test may not have "correct size" (the type-I error could be very large):

$$\Pr\left(\min_{1\le j\le k} p_j < \alpha\right) \le \sum_{j=1}^{k} \Pr\left(p_j < \alpha\right) \to k\alpha.$$

$\Pr(\min_{1\le j\le k} P_j < \alpha) = \Pr(P_1<\alpha$ 或 $P_2<\alpha$ 或 $\cdots) = \Pr(P_1<\alpha \cup P_2<\alpha \cup \cdots \cup P_k<\alpha)$

$\le \sum_{j=1}^{k} \Pr(P_j<\alpha) \to k\alpha$

- Bonferroni correction: use the adjusted significance level α/k,

$$\underbrace{\Pr\left(\min_{1\le j\le k} p_j < \frac{\alpha}{k}\right)}_{\text{上极限肯定} \le \alpha} \le \sum_{j=1}^{k} \Pr\left(p_j < \frac{\alpha}{k}\right) \to \alpha.$$

$\Rightarrow$ type-I error 的概率是 $\le \alpha$ 的.

So the type-I error associated with the decision rule should not be much larger than α.