Econometrics Lecture 8

Statistical Inference on OLS Estimator

Hang Miao

Rutgers University

April 9, 2021

Overview

1. Multivariate Central Limit Theorem
2. Statistical Inference with Known σ^2 and Large n
   - Asymptotic Distribution for $\hat{\boldsymbol{\beta}}_{ols}$
   - Hypothesis Testing for $\hat{\beta}_k$
   - P-Value
   - Confidence Interval
3. Statistical Inference with Unknown σ^2 and Large n
4. Statistical Inference with Unknown σ^2 and Small n
5. Statistical Inference with Known σ^2 and Small n
6. Appendix
   - Properties of Idempotent Matrix
   - χ^2, t, F-Distribution

Multivariate Central Limit Theorem

Central Limit Theorem

Definition (Lindeberg–Levy CLT)

Suppose $\{X_n\}$ is a sequence of i.i.d. random variables with $\mathrm{E}[X_i] = \mu$ and $\mathrm{Var}[X_i] = \sigma^2 < \infty$. Then as $n \to \infty$,

$$\sqrt{n}(\overline{X}_n - \mu) \xrightarrow{d} N(0, \sigma^2)$$

Or equivalently,

$$(\overline{X}_n - \mu) \xrightarrow{d} N(0, \frac{\sigma^2}{n})$$

Or equivalently,

$$\frac{\overline{X}_n - \mu}{\frac{\sigma}{\sqrt{n}}} \xrightarrow{d} N(0, 1)$$

Multivariate Central Limit Theorem

Definition (Multivariate Central Limit Theorem)

Suppose $\{\boldsymbol{V}_n\}$ is a sequence of m-dimensional i.i.d. random variables with $\mathrm{E}[\boldsymbol{V}_i] = \boldsymbol{\mu}$ and $\mathrm{Var}[\boldsymbol{V}_i] = \boldsymbol{\Sigma_V}$, where $\boldsymbol{\Sigma_V}$ is the covariance matrix defined as $\mathrm{E}[(\boldsymbol{V}_i - \boldsymbol{\mu})(\boldsymbol{V}_i - \boldsymbol{\mu})']$. Then as $n \to \infty$,

$$\sqrt{n}(\overline{\boldsymbol{V}}_n - \boldsymbol{\mu}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Sigma_V})$$

Or equivalently,

$$(\overline{\boldsymbol{V}}_n - \boldsymbol{\mu}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{\Sigma}_{\overline{\boldsymbol{V}}_n}) = N(\boldsymbol{0}, \frac{1}{n}\boldsymbol{\Sigma_V})$$

Where $\boldsymbol{\Sigma}_{\overline{\boldsymbol{V}}_n}$ is the covariance matrix of $\overline{\boldsymbol{V}}_n - \boldsymbol{\mu}$ or $\overline{\boldsymbol{V}}_n$.

$$\boldsymbol{\Sigma}_{\overline{\boldsymbol{V}}_n} = \mathrm{Var}\left[\frac{1}{n}\sum_{i=1}^{n} \boldsymbol{V}_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} \mathrm{Var}\left[\boldsymbol{V}_i\right] = \frac{1}{n^2}\sum_{i=1}^{n} \boldsymbol{\Sigma_V} = \frac{1}{n}\boldsymbol{\Sigma_V}$$

Statistical Inference with Known σ^2 and Large n

Summary of Assumptions so Far

Assumptions For Getting the OLS Estimator

1. Underlying population space (X, Y) has a linear relationship $y = \boldsymbol{x}'\boldsymbol{\beta} + u = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_p x_p + u$ with $\mathrm{E}[u] = 0$
2. Rank of the matrix $\boldsymbol{X}$ is $p+1$ so that $(\boldsymbol{X}'\boldsymbol{X})^{-1}$ exist. This is also called No Multicollinearity

Assumptions For Getting Unbiased and Consistent OLS estimator

3. Strict Exogeneity Condition: $\mathrm{E}[\boldsymbol{u}|\boldsymbol{X}] = \boldsymbol{0}$

Assumptions For Getting the BLUE Estimator

4. Spherical Error Variance: $\mathrm{Var}[\boldsymbol{u}|\boldsymbol{X}] = \sigma^2 \boldsymbol{I}_n$
   - Homoskedasticity: $\mathrm{E}[u_i^2|\boldsymbol{X}] = \sigma^2 \quad \forall\, i = 1, 2, \cdots, n$
   - No Serial Correlation: $\mathrm{E}[u_i u_j|\boldsymbol{X}] = 0 \quad \forall\, i, j = 1, 2, \cdots, n;\ i \neq j$

$\hat{\beta}_{ols}$ as the Average of a Sequence of Vectors

$$
\begin{aligned}
\hat{\beta}_{ols} &= \beta + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{u} \\
&= \beta + \left(\frac{\mathbf{X}'\mathbf{X}}{n}\right)^{-1}\left(\frac{\mathbf{X}'\mathbf{u}}{n}\right) \\
&= \beta + \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}\left(\frac{1}{n}\sum_{i=1}^{n}u_i\mathbf{x}_i\right) \\
&= \frac{1}{n}\sum_{i=1}^{n}\left[\beta + \left(\frac{1}{n}\sum_{i=1}^{n}\mathbf{x}_i\mathbf{x}_i'\right)^{-1}(u_i\mathbf{x}_i)\right] \\
&\equiv \frac{1}{n}\sum_{i=1}^{n}\mathbf{V}_i \\
&\equiv \overline{\mathbf{V}}_n
\end{aligned}
$$

Conditional Expectation and Variance of $\hat{\beta}_{ols}$

$$
\begin{aligned}
\mathrm{E}\left[\hat{\beta}_{ols} \mid \boldsymbol{X}\right] &= \beta \equiv \mu = \mathrm{E}\left[\overline{\boldsymbol{V}}_n \mid \boldsymbol{X}\right] \\
\mathrm{Var}\left[\hat{\beta}_{ols} \mid \boldsymbol{X}\right] &= \mathrm{E}\left[\left(\hat{\beta}_{ols} - \beta\right)\left(\hat{\beta}_{ols} - \beta\right)' \mid \boldsymbol{X}\right] \\
&= \mathrm{E}\left[\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\right)\left((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\right)' \mid \boldsymbol{X}\right] \\
&= \mathrm{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\boldsymbol{u}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \mid \boldsymbol{X}\right] \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathrm{E}\left[\boldsymbol{u}\boldsymbol{u}' \mid \boldsymbol{X}\right]\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\mathrm{Var}\left[\boldsymbol{u} \mid \boldsymbol{X}\right]\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\,\sigma^2\boldsymbol{I}_n\,\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} \\
&= \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} \equiv \Sigma_{\overline{\boldsymbol{V}}_n} = \mathrm{Var}\left[\overline{\boldsymbol{V}}_n \mid \boldsymbol{X}\right]
\end{aligned}
$$

Distribution of $\hat{\beta}_{ols}$ with Known σ^2 and Large Sample

Taking the sample data matrix $\boldsymbol{X}$ as given, according to the Multivariate Central Limit Theorem,

$$(\overline{\boldsymbol{V}}_n - \boldsymbol{\mu}) \xrightarrow{d}, N\Big(\mathbf{0}, \, \Sigma_{\overline{\boldsymbol{V}}_n}\Big)$$

the distribution for OLS estimator $\hat{\beta}_{ols}$ follows

$$(\hat{\beta}_{ols} - \beta) \xrightarrow{d} N\Big(\mathbf{0}, \, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\Big)$$

Or equivalently

$$\hat{\beta}_{ols} \xrightarrow{d} N\Big(\beta, \, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\Big)$$

Matrix Expression for $\hat{\boldsymbol{\beta}}_{ols} \sim N\Big(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\Big)$

$$\hat{\boldsymbol{\beta}}_{ols} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_p \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \begin{bmatrix} m_{(00)} & m_{(01)} & m_{(02)} & \cdots & m_{(0p)} \\ m_{(10)} & m_{(11)} & m_{(12)} & \cdots & m_{(1p)} \\ m_{(20)} & m_{(21)} & m_{(22)} & \cdots & m_{(2p)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{(p0)} & m_{p1)} & m_{(p2)} & \cdots & m_{(pp)} \end{bmatrix} \right)$$

where

$$\sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1} = \begin{bmatrix} m_{(00)} & m_{(01)} & m_{(02)} & \cdots & m_{(0p)} \\ m_{(10)} & m_{(11)} & m_{(12)} & \cdots & m_{(1p)} \\ m_{(20)} & m_{(21)} & m_{(22)} & \cdots & m_{(2p)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{(p0)} & m_{p1)} & m_{(p2)} & \cdots & m_{(pp)} \end{bmatrix}$$

Distribution of Each Individual Entry $\hat{\beta}_k$ of $\hat{\boldsymbol{\beta}}_{ols}$ with Known σ^2 and Large Sample

Since we have $\hat{\boldsymbol{\beta}}_{ols} \sim N\Big(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\Big)$, according to the property of Multivariate Normal Distribution, the k-th entry of vector $\hat{\boldsymbol{\beta}}_{ols}$ denoted as $\hat{\beta}_k$ follows regular Univariate Normal distribution $N\Big(\beta_k, m_{(kk)}\Big)$. For example $\hat{\beta}_2 \sim N(\beta_2, m_{(22)})$, $\hat{\beta}_p \sim N(\beta_p, m_{(pp)})$

$$\hat{\boldsymbol{\beta}}_{ols} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_p \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \begin{bmatrix} m_{(00)} & m_{(01)} & m_{(02)} & \cdots & m_{(0p)} \\ m_{(10)} & m_{(11)} & m_{(12)} & \cdots & m_{(1p)} \\ m_{(20)} & m_{(21)} & m_{(22)} & \cdots & m_{(2p)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{(p0)} & m_{p1)} & m_{(p2)} & \cdots & m_{(pp)} \end{bmatrix} \right)$$

Hypothesis Testing for β_k

Null Hypothesis H_0

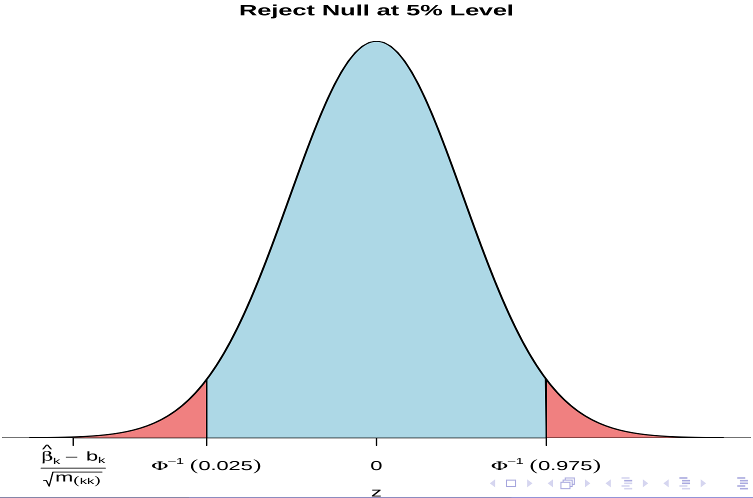
- Suppose we propose the the Null Hypothesis as $H_0 : \beta_k = b_k$
- Since $\hat{\beta}_k \sim N(\beta_k, m_{(kk)})$, under Null Hypothesis H_0, $\hat{\beta}_k \sim N(b_k, m_{(kk)})$, thus we could construct the test statistic

$$z = \frac{\hat{\beta}_k - b_k}{\sqrt{m_{(kk)}}} \sim N(0,1)$$

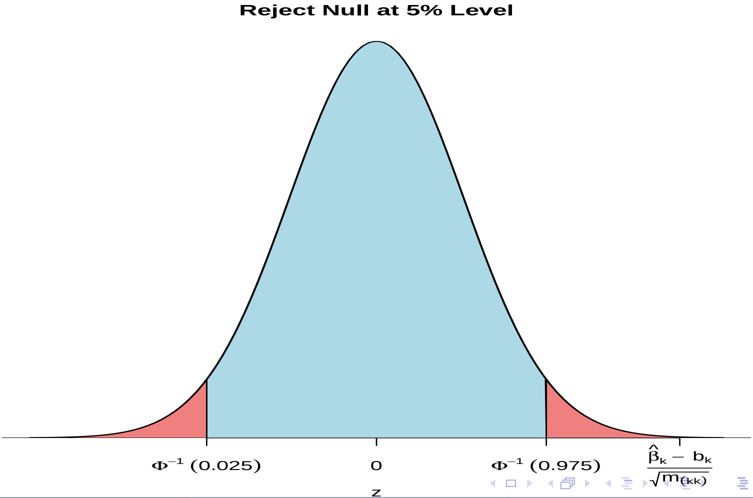
Alternative Hypothesis H_1

- If $H_1 : \beta_k \neq b_k$, we have two-sided reject region for significant level and P-Value
- If $H_1 : \beta_k > b_k$ or $H_1 : \beta_k < b_k$, we have one-sided reject region for significant level and P-Value.

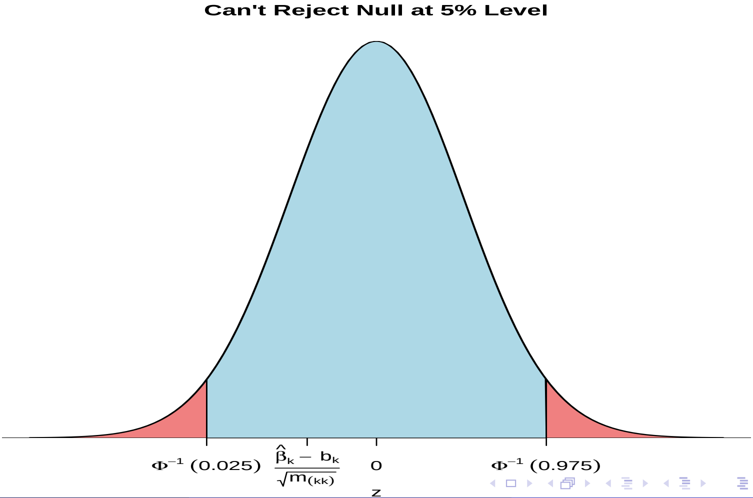
Reject the H_0 at 5% Significant Level under $H_1 : \beta_k \neq b_k$: Case I



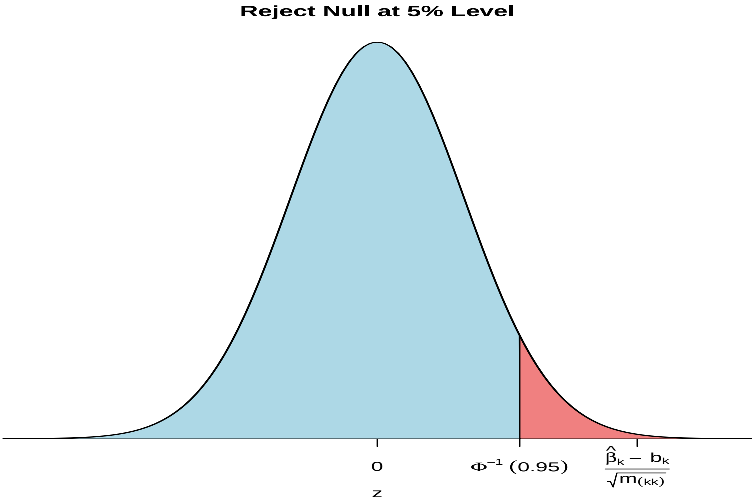
Reject the H_0 at 5% Significant Level under $H_1 : \beta_k \neq b_k$: Case II



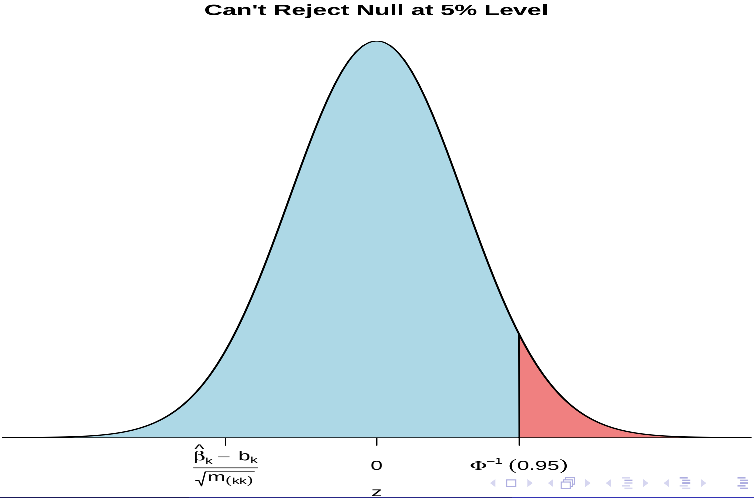
Can't Reject the H_0 at 5% Significant Level under $H_1 : \beta_k \neq b_k$



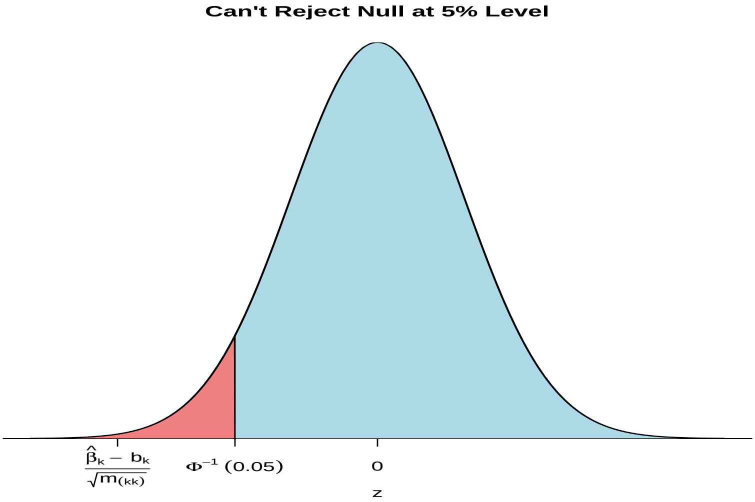
Reject the H_0 at 5% Significant Level under $H_1: \beta_k > b_k$



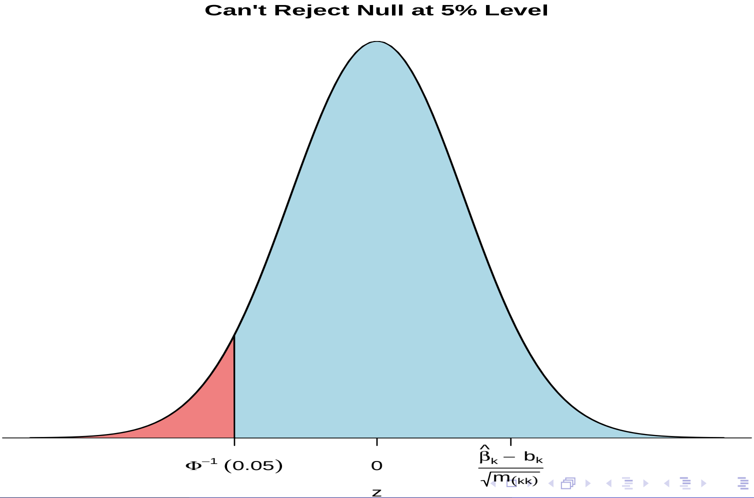
Can't Reject the H_0 at 5% Significant Level under $H_1 : \beta_k > b_k$



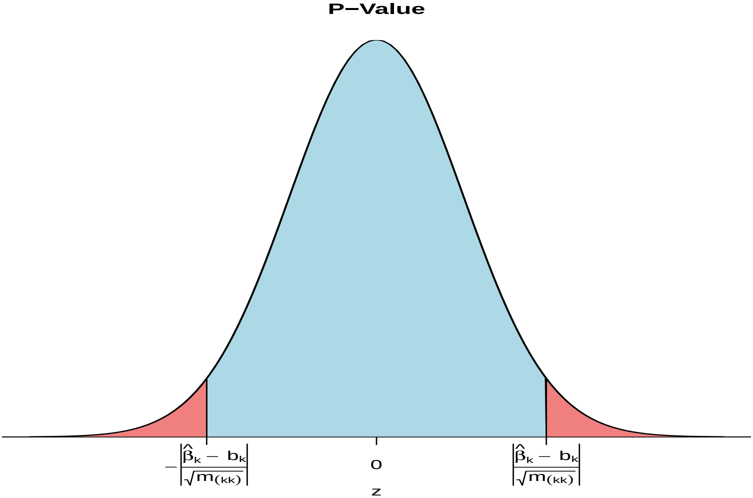
Reject the H_0 at 5% Significant Level under $H_1 : \beta_k < b_k$



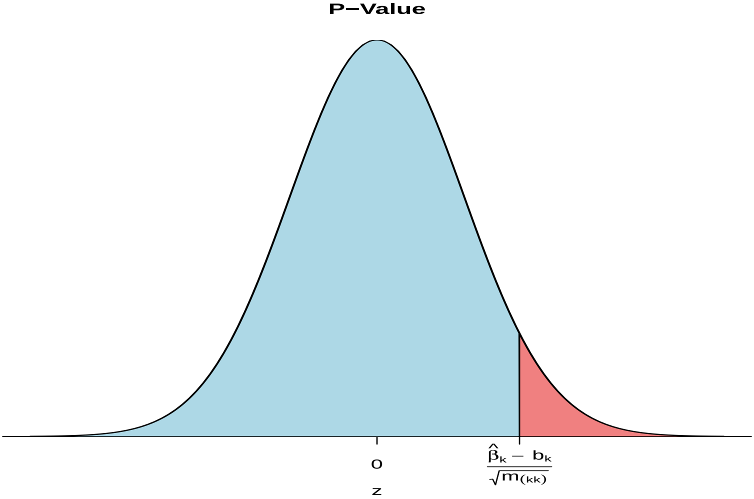
Can't Reject the H_0 at 5% Significant Level under $H_1 : \beta_k < b_k$



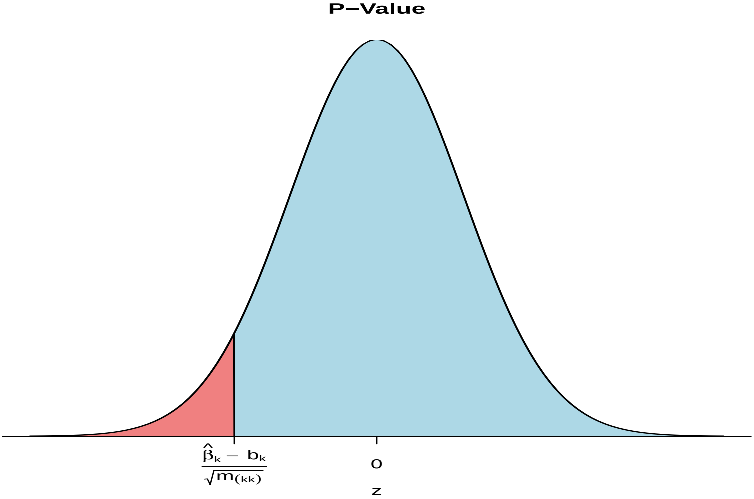
P-Value under $H_1 : \beta_k \neq b_k$



P-Value under $H_1 : \beta_k > b_k$



P-Value under $H_1 : \beta_k < b_k$



Confidence Intervals

Definition (Confidence intervals)

Confidence Interval (CI) is a type of interval estimate, computed from the statistics of the observed data, that might contain the true value of an unknown population parameter.

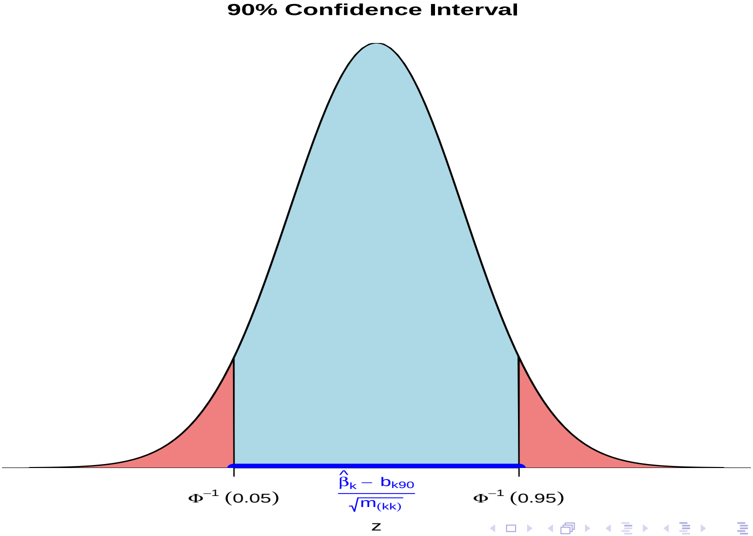
Confidence Intervals for the Population Mean

$$\text{99\% confidence interval for } \beta_k = \left[\hat{\beta}_k \pm 2.58 \times \sqrt{m_{(kk)}}\right]$$

$$\text{95\% confidence interval for } \beta_k = \left[\hat{\beta}_k \pm 1.96 \times \sqrt{m_{(kk)}}\right]$$

$$\text{90\% confidence interval for } \beta_k = \left[\hat{\beta}_k \pm 1.64 \times \sqrt{m_{(kk)}}\right]$$

90% Confidence Interval



90% Confidence Interval

We want the population mean b_{k90} takes the value such that statistic Z_n lie in the region where the area (probability) under the PDF is 90%.

$$\Phi^{-1}(0.05) \leq \frac{\hat{\beta}_k - b_{k90}}{\sqrt{m_{(kk)}}} \leq \Phi^{-1}(0.95)$$

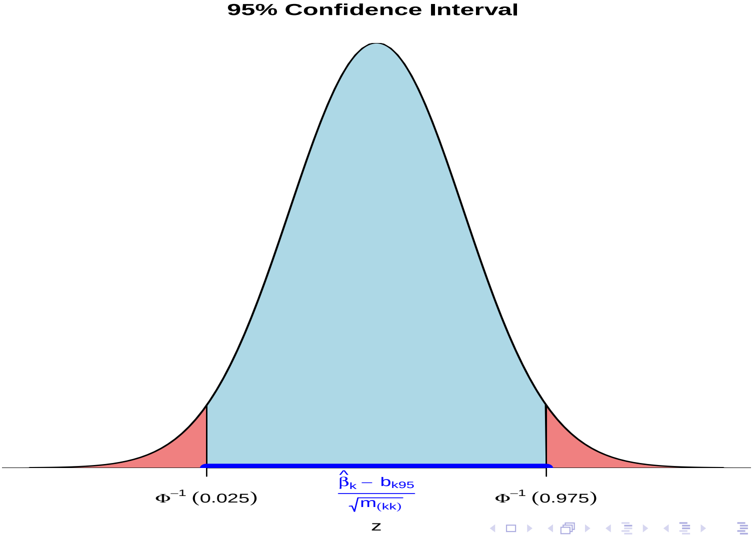
$$\sqrt{m_{(kk)}}\Phi^{-1}(0.05) \leq \hat{\beta}_k - b_{k90} \leq \sqrt{m_{(kk)}}\Phi^{-1}(0.95)$$

$$\hat{\beta}_k - \sqrt{m_{(kk)}}\Phi^{-1}(0.95) \leq b_{k90} \leq \hat{\beta}_k - \sqrt{m_{(kk)}}\Phi^{-1}(0.05)$$

Since normal distribution is symmetrical,
$\Phi^{-1}(0.05) = -\Phi^{-1}(0.95) = -1.64$. Therefore,

$$b_{k90} \in \left[\hat{\beta}_k - 1.64\sqrt{m_{(kk)}}, \; \hat{\beta}_k + 1.64\sqrt{m_{(kk)}}\right]$$

95% Confidence Interval



95% Confidence Interval

We want the population mean b_{k95} takes the value such that statistic Z_n lie in the region where the area (probability) under the PDF is 95%.

$$\Phi^{-1}(0.025) \leq \frac{\hat{\beta}_k - b_{k95}}{\sqrt{m_{(kk)}}} \leq \Phi^{-1}(0.975)$$

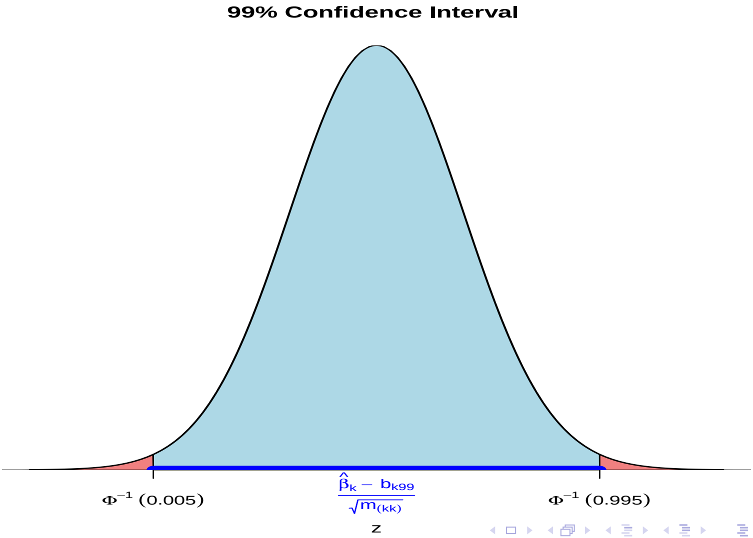
$$\sqrt{m_{(kk)}}\Phi^{-1}(0.025) \leq \hat{\beta}_k - b_{k95} \leq \sqrt{m_{(kk)}}\Phi^{-1}(0.975)$$

$$\hat{\beta}_k - \sqrt{m_{(kk)}}\Phi^{-1}(0.975) \leq b_{k95} \leq \hat{\beta}_k - \sqrt{m_{(kk)}}\Phi^{-1}(0.025)$$

Since normal distribution is symmetrical,
$\Phi^{-1}(0.025) = -\Phi^{-1}(0.975) = -1.96$. Therefore,

$$b_{k95} \in \left[\hat{\beta}_k - 1.96\sqrt{m_{(kk)}},\ \hat{\beta}_k + 1.96\sqrt{m_{(kk)}}\right]$$

99% Confidence Interval



99% Confidence Interval

We want the population mean b_{k99} takes the value such that statistic Z_n lie in the region where the area (probability) under the PDF is 99%.

$$\Phi^{-1}(0.005) \leq \frac{\hat{\beta}_k - b_{k99}}{\sqrt{m_{(kk)}}} \leq \Phi^{-1}(0.995)$$

$$\sqrt{m_{(kk)}}\Phi^{-1}(0.005) \leq \hat{\beta}_k - b_{k99} \leq \sqrt{m_{(kk)}}\Phi^{-1}(0.995)$$

$$\hat{\beta}_k - \sqrt{m_{(kk)}}\Phi^{-1}(0.995) \leq b_{k99} \leq \hat{\beta}_k - \sqrt{m_{(kk)}}\Phi^{-1}(0.005)$$

Since normal distribution is symmetrical,
$\Phi^{-1}(0.005) = -\Phi^{-1}(0.995) = -2.58.$ Therefore,

$$b_{k99} \in \left[\hat{\beta}_k - 2.58\sqrt{m_{(kk)}},\, \hat{\beta}_k + 2.58\sqrt{m_{(kk)}}\right]$$

Statistical Inference with Unknown σ^2 and Large n

Estimate σ^2

If σ^2 is unkown, we could still estimate it. From Homoskedasticity of assumption 4: $\mathrm{E}[u_i^2|\boldsymbol{X}] = \sigma^2$, We could estimate this σ^2 by taking the sample mean of $\overline{u^2} = \frac{1}{n}\sum_{i=1}^{n} u_i^2$. Since u_i is unobservable, it is replaced by $\hat{u}_i = y_i - \hat{y}_i$. So we could estimate σ^2 by the following estimator:

$$\hat{\sigma}^2 = \frac{1}{n-p-1}\sum_{i=1}^{n}\hat{u}_i^2 = \frac{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}}{n-p-1} = \frac{\mathrm{SSR}}{n-p-1}$$

We will show $\hat{\sigma}^2$ is unbiased that is:

$$\mathrm{E}[\hat{\sigma}^2|\boldsymbol{X}] = \mathrm{E}[u_i^2|\boldsymbol{X}] = \sigma^2$$

Unbiasedness of Estimator $\hat{\sigma}^2$

$$
\begin{aligned}
\mathrm{E}[\,\hat{\sigma}^2\,|\boldsymbol{X}] &= \mathrm{E}\left[\frac{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}}{n-p-1}|\boldsymbol{X}\right] = \frac{1}{n-p-1}\mathrm{E}\left[\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}|\boldsymbol{X}\right] \\
\mathrm{E}\left[\,\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}\,|\boldsymbol{X}\right] &= \mathrm{E}[\,(\boldsymbol{y}-\hat{\boldsymbol{y}})'\,(\boldsymbol{y}-\hat{\boldsymbol{y}})\,|\boldsymbol{X}] = \mathrm{E}[\,(\boldsymbol{y}-\boldsymbol{X}\hat{\beta}_{ols})'\,(\boldsymbol{y}-\boldsymbol{X}\hat{\beta}_{ols})\,|\boldsymbol{X}] \\
&= \mathrm{E}[\,(\boldsymbol{y}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y})'\,(\boldsymbol{y}-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y})\,|\boldsymbol{X}] \\
&= \mathrm{E}[\,\boldsymbol{y}'(\boldsymbol{I}_n-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')'\,(\boldsymbol{I}_n-\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}')\boldsymbol{y}\,|\boldsymbol{X}] \\
&= \mathrm{E}[\,\boldsymbol{y}'\,\boldsymbol{M}'\boldsymbol{M}\,\boldsymbol{y}\,|\boldsymbol{X}] = \mathrm{E}[\,\boldsymbol{y}'\boldsymbol{M}\,\boldsymbol{y}\,|\boldsymbol{X}] \\
&= \mathrm{E}[\,(\boldsymbol{X}\beta+\boldsymbol{u})'\boldsymbol{M}\,(\boldsymbol{X}\beta+\boldsymbol{u})\,|\boldsymbol{X}] \\
&= \mathrm{E}[\,\boldsymbol{u}'\boldsymbol{M}\,\boldsymbol{u}\,|\boldsymbol{X}] = \sum_{i,j}^{n}\boldsymbol{M}_{ij}\,\mathrm{E}[\,u_i\,u_j\,|\boldsymbol{X}] \\
&= \sum_{i=1}^{n}\boldsymbol{M}_{ii}\,\mathrm{E}[\,u_i\,u_i\,|\boldsymbol{X}] + \sum_{i\neq j}^{n}\boldsymbol{M}_{ij}\,\mathrm{E}[\,u_i\,u_j\,|\boldsymbol{X}] = \sigma^2\cdot\mathrm{trace}(\boldsymbol{M}) \\
&= \sigma^2(n-p-1)
\end{aligned}
$$

Consistency of Estimator $\hat{\sigma}^2$

$$
\begin{aligned}
\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}} &= (\boldsymbol{y} - \boldsymbol{X}\hat{\beta}_{ols})'(\boldsymbol{y} - \boldsymbol{X}\hat{\beta}_{ols}) \\
&= (\boldsymbol{X}\beta + \boldsymbol{u} - \boldsymbol{X}\hat{\beta}_{ols})'(\boldsymbol{X}\beta + \boldsymbol{u} - \boldsymbol{X}\hat{\beta}_{ols}) \\
&= ((\beta - \hat{\beta}_{ols})'\boldsymbol{X}' + \boldsymbol{u}')(\boldsymbol{X}(\beta - \hat{\beta}_{ols}) + \boldsymbol{u}) \\
&\xrightarrow{p} \boldsymbol{u}'\boldsymbol{u} = \sum_{i=1}^{n} u_i^2 \\
\hat{\sigma}^2 \,|\, \boldsymbol{X} &= \frac{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}}{n-p-1} \,|\, \boldsymbol{X} \\
&= \frac{n}{n-p-1}\frac{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}}{n} \,|\, \boldsymbol{X} \\
&\xrightarrow{p} 1 \cdot \frac{\sum_{i=1}^{n} u_i^2}{n} \,|\, \boldsymbol{X} \\
&\xrightarrow{p} \mathrm{E}[u_i^2 \,|\, \boldsymbol{X}] = \sigma^2
\end{aligned}
$$

Distribution of $\hat{\beta}_{ols}$ with Unknown σ^2 and Large Sample $\hat{\beta}_{ols} \sim N\left(\beta, \hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}\right)$

$$\hat{\beta}_{ols} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_p \end{pmatrix} \sim N\left(\begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{pmatrix}, \begin{bmatrix} m_{(00)} & m_{(01)} & m_{(02)} & \cdots & m_{(0p)} \\ m_{(10)} & m_{(11)} & m_{(12)} & \cdots & m_{(1p)} \\ m_{(20)} & m_{(21)} & m_{(22)} & \cdots & m_{(2p)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{(p0)} & m_{p1)} & m_{(p2)} & \cdots & m_{(pp)} \end{bmatrix} \right)$$

where

$$\color{red}{\hat{\sigma}^2}(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} m_{(00)} & m_{(01)} & m_{(02)} & \cdots & m_{(0p)} \\ m_{(10)} & m_{(11)} & m_{(12)} & \cdots & m_{(1p)} \\ m_{(20)} & m_{(21)} & m_{(22)} & \cdots & m_{(2p)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ m_{(p0)} & m_{p1)} & m_{(p2)} & \cdots & m_{(pp)} \end{bmatrix}$$

Statistical Inference with Unknown σ^2 and Small n

t-statistic

If the sample size n is small, Central Limit Theory can not be applied, so the distribution of $\hat{\beta}_k$ does not follow Normal distribution $N(\beta_k, m_{kk})$. But if we have the following assumption satisfied:

Assumptions For Getting OLS estimator follows t-distribution

5. uncertainty u follows Normal Distribution: $\boldsymbol{u}|\boldsymbol{X} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$
i.e $u_i|\boldsymbol{X} \sim N(0, \sigma^2)$

Then the normalized t-statistic with the same format as the previous z-statistic follows standard t-distribution with degree $n - p - 1$:

$$t = \frac{\hat{\beta}_k - \beta_k}{\sqrt{m_{(kk)}}} \sim t(n - p - 1)$$

where $m_{(kk)}$ is the estimated variance of $\hat{\beta}_k$ i.e the k-th diagonal entry of the estimated covariance matrix $\hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ defined above.

Proof t-statistic follows t-distribution

$$t = \frac{\hat{\beta}_k - \beta_k}{\sqrt{m_{(kk)}}} = \frac{\hat{\beta}_k - \beta_k}{\sqrt{\hat{\sigma}^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}_{(kk)}}} = \frac{\hat{\beta}_k - \beta_k}{\sqrt{\sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}_{(kk)}}} \sqrt{\frac{\sigma^2}{\hat{\sigma}^2}}$$

We want to show:

- $\frac{\hat{\beta}_k - \beta_k}{\sqrt{\sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}_{(kk)}}} \sim N(0,1)$
- $\frac{(n-p-1)\hat{\sigma}^2}{\sigma^2} \sim \chi^2(n-p-1)$
- $\frac{\hat{\beta}_k - \beta_k}{\sqrt{\sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}_{(kk)}}}$ and $\frac{\hat{\sigma}^2}{\sigma^2}$ are independent

$\frac{\hat{\beta}_k - \beta_k}{\sqrt{\sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}_{(kk)}}} \sim N(0,1)$

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta} &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} \\
\because\ \boldsymbol{u} &\sim N(0, \sigma^2 \boldsymbol{I}_n) \quad \textit{from new assumption 5} \\
\therefore\ (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} &\sim N\Big(0, (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\sigma^2 \boldsymbol{I}_n \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\Big) \\
\therefore\ \hat{\boldsymbol{\beta}}_{ols} - \boldsymbol{\beta} &\sim N\Big(0, \sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}\Big) \\
\therefore\ \frac{\hat{\beta}_k - \beta_k}{\sqrt{\sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}_{(kk)}}} &\sim N(0,1) \quad \forall k = 1, 2, \cdots, p+1
\end{aligned}
$$

$$
\begin{aligned}
\frac{(n-p-1)\hat{\sigma}^2}{\sigma^2} &= \frac{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}}{\sigma^2} \\
&= \frac{\boldsymbol{u}'\boldsymbol{M}\,\boldsymbol{u}}{\sigma^2} \\
&= \frac{\boldsymbol{u}'\boldsymbol{Q}'\boldsymbol{\Lambda}\boldsymbol{Q}\,\boldsymbol{u}}{\sigma^2} \quad \textit{(by eigen-decomposition)} \\
&= \frac{\boldsymbol{u}'\boldsymbol{Q}'\begin{bmatrix} \boldsymbol{I}_{(n-p-1)} & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix}\boldsymbol{Q}\,\boldsymbol{u}}{\sigma^2} \quad \textit{(eigenvalues are either 1 or 0)} \\
&= \frac{\sum_{i=1}^{n-p-1} u_i^2}{\sigma^2} = \sum_{i=1}^{n-p-1}\left(\frac{u_i}{\sigma}\right)^2 \sim \chi(n-p-1)
\end{aligned}
$$

$$
\left(\because u_i \sim N(0,\sigma^2) \implies \frac{u_i}{\sigma} \sim N(0,1)\right)
$$

$\frac{\hat{\beta}_k - \beta_k}{\sqrt{\sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}_{(kk)}}}$ and $\frac{\hat{\sigma}^2}{\sigma^2}$ are independent

Since β_k, $\sqrt{\sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}_{(kk)}}$, σ^2 are constant, and $\hat{\sigma}^2 = \frac{\hat{\boldsymbol{u}}'\hat{\boldsymbol{u}}}{n-p-1}$ it is equivalent to show that $\hat{\boldsymbol{\beta}}_{ols}$ and $\hat{\boldsymbol{u}}$ are independent

$$
\begin{aligned}
\hat{\boldsymbol{\beta}}_{ols} &= \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} \sim N\left(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}\right) \\
\hat{\boldsymbol{u}} &= \boldsymbol{M}\boldsymbol{u} \sim N\left(\boldsymbol{0}, \sigma^2\boldsymbol{M}\right) \\
\mathrm{Cov}(\hat{\boldsymbol{\beta}}_{ols}, \hat{\boldsymbol{u}} \mid \boldsymbol{X}) &= \mathrm{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\,(\boldsymbol{M}\boldsymbol{u})' \mid \boldsymbol{X}\right] \\
&= \mathrm{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\,\boldsymbol{u}'\boldsymbol{M} \mid \boldsymbol{X}\right] \\
&= \mathrm{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\sigma^2\boldsymbol{I}_n\boldsymbol{M} \mid \boldsymbol{X}\right] \\
&= \sigma^2\mathrm{E}\left[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{M} \mid \boldsymbol{X}\right] \\
&= \sigma^2\mathrm{E}\left[\boldsymbol{0} \mid \boldsymbol{X}\right] \\
&= \boldsymbol{0}
\end{aligned}
$$

Since both $\hat{\boldsymbol{\beta}}_{ols}$ and $\hat{\boldsymbol{u}}$ follows joint Normal distribution, therefore uncorrelation implies independence due to the properties of normal

Statistical Inference with Known σ^2 and Small n

z Test

Because the sample size n is small, Central Limit Theory can not be applied, the same additional assumption has to be made.

Normality Assumption

5. uncertainty u follows Normal Distribution: $\boldsymbol{u}|\boldsymbol{X} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I}_n)$
   i.e $u_i|\boldsymbol{X} \sim N(0, \sigma^2)$

With σ^2 known, the Z test statistic could be constructed same as the Large n Case. Previously, we have proved that under assumption 1-5 $\frac{\hat{\beta}_k - \beta_k}{\sqrt{\sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}_{(kk)}}} \sim N(0,1)$. So in case of known σ^2 and small n, the test statistic and the associated distribution is summarized as follow :

z Test Statistic and Distribution

$$z = \frac{\hat{\beta}_k - \beta_k}{\sqrt{\sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1}_{(kk)}}} = \frac{\hat{\beta}_k - \beta_k}{\sqrt{m_{(kk)}}} \sim N(0,1)$$

Appendix

Properties of $\boldsymbol{M}$

Let $\boldsymbol{M} = \boldsymbol{I}_n - \boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'$

1. $\boldsymbol{M}$ is symmetric: $\boldsymbol{M}' = \boldsymbol{M}$
2. $\boldsymbol{M}$ is idempotent: $\boldsymbol{M}\boldsymbol{M} = \boldsymbol{M}$
3. $\boldsymbol{M}\boldsymbol{X} = \boldsymbol{0}$
4. $\text{trace}(\boldsymbol{M}) = n - p - 1$
5. The eigenvalues of $\boldsymbol{M}$ are either 0 or 1

The eigenvalues of $\boldsymbol{M}$ are either 0 or 1

Let λ be the eigenvalue and $\boldsymbol{v}$ be the associated eigenvector of $\boldsymbol{M}$

$$\begin{aligned}
\boldsymbol{M}\boldsymbol{v} &= \lambda\boldsymbol{v} \\
\boldsymbol{M}\boldsymbol{v} &= \boldsymbol{M}\boldsymbol{M}\boldsymbol{v} = \boldsymbol{M}(\lambda\boldsymbol{v}) = \lambda\boldsymbol{M}\boldsymbol{v} = \lambda^2\boldsymbol{v} \\
\therefore \lambda^2\boldsymbol{v} &= \lambda\boldsymbol{v} \\
&\Longrightarrow (\lambda^2 - \lambda)\boldsymbol{v} = \boldsymbol{0} \Longrightarrow (\lambda^2 - \lambda) = \lambda(\lambda - 1) = 0 \\
\therefore \quad &\lambda \quad \textit{should be either 0 or 1}
\end{aligned}$$

trace($\boldsymbol{M}$) $= n - p - 1$

We know $\boldsymbol{X}$ is $n \times (p+1)$ matrix, $\boldsymbol{B}$ is $n \times m$ matrix

$$\begin{aligned}
\text{trace}(\boldsymbol{M}) &= \text{trace}(\boldsymbol{I}_n) - \text{trace}(\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}') \\
&= n - \text{trace}(\boldsymbol{X}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}) \\
&= n - \text{trace}(\boldsymbol{I}_{p+1}) \\
&= n - p - 1
\end{aligned}$$

trace($\boldsymbol{AB}$) = trace($\boldsymbol{BA}$)

Suppose $\boldsymbol{A}$ is $m \times n$ matrix, $\boldsymbol{B}$ is $n \times m$ matrix

$$\text{trace}(\boldsymbol{AB}) = \sum_{i=1}^{m}\sum_{k=1}^{n} a_{ik} b_{ki} = \sum_{k=1}^{n}\sum_{i=1}^{m} b_{ki} a_{ik} = \text{trace}(\boldsymbol{BA})$$

χ^2-Distribution

Definition

- Let $\{Z_n\}$ be a sequence of i.i.d. random variables.
- For each $Z_i \sim N(0, 1)$
- Let $W = \sum_{i=1}^{n} Z_i^2$

Then the distribution for the random variable W follows χ^2-Distribution with n degree of freedom, denoted as $\chi^2(n)$

t-Distribution

Definition

- Let $Z \sim N(0,1)$
- Let $W \sim \chi^2(n)$
- Z and W are independent
- Let $T = \frac{Z}{\sqrt{W/n}}$

Then the distribution for the random variable T follows t-Distribution with n degree of freedom, denoted as $t(n)$

F-Distribution

Definition

- Let $W_1 \sim \chi^2(n_1)$
- Let $W_2 \sim \chi^2(n_2)$
- W_1 and W_2 are independent
- Let $F = \frac{W_1/n_1}{W_2/n_2}$

Then the distribution for the random variable F follows F-Distribution with degree n_1 and n_2, denoted as $F(n_1, n_2)$

$\boldsymbol{X}'\boldsymbol{X} = \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'$

$$
\begin{aligned}
\boldsymbol{X}'\boldsymbol{X} &= \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & x_{31} & \dots & x_{n1} \\ x_{12} & x_{22} & x_{32} & \dots & x_{n2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & x_{3p} & \dots & x_{np} \end{bmatrix} \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ 1 & x_{31} & x_{32} & \dots & x_{3p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{n1} & x_{n2} & \dots & x_{np} & \end{bmatrix} \\
&= \begin{bmatrix} \boldsymbol{x}_1 & \boldsymbol{x}_2 & \boldsymbol{x}_3 & \dots & \boldsymbol{x}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{x}_1' \\ \boldsymbol{x}_2' \\ \boldsymbol{x}_3' \\ \vdots \\ \boldsymbol{x}_n' \end{bmatrix} \\
&= \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i'
\end{aligned}
$$

Matrix Expression for $x_i x_i'$

$$
\begin{aligned}
\boldsymbol{x}_i \boldsymbol{x}_i' &= \begin{pmatrix} 1 \\ x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix} \begin{pmatrix} 1 & x_{i1} & x_{i2} & \dots & x_{ip} \end{pmatrix} \\
&= \begin{bmatrix} 1 & x_{i1} & x_{i2} & \dots & x_{ip} \\ x_{i1} & x_{i1}^2 & x_{i1}x_{i2} & \dots & x_{i1}x_{ip} \\ x_{i2} & x_{i2}x_{i1} & x_{i2}^2 & \dots & x_{i2}x_{ip} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{ip} & x_{ip}x_{i1} & x_{ip}x_{i2} & \dots & x_{ip}^2 \end{bmatrix}
\end{aligned}
$$

Matrix Expression for $X'X$

$$
\begin{aligned}
\boldsymbol{X}'\boldsymbol{X} &= \sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i' \\
&= \sum_{i=1}^{n} \begin{bmatrix} 1 & x_{i1} & x_{i2} & \cdots & x_{ip} \\ x_{i1} & x_{i1}^2 & x_{i1}x_{i2} & \cdots & x_{i1}x_{ip} \\ x_{i2} & x_{i2}x_{i1} & x_{i2}^2 & \cdots & x_{i2}x_{ip} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{ip} & x_{ip}x_{i1} & x_{ip}x_{i2} & \cdots & x_{ip}^2 \end{bmatrix} \\
&= \begin{bmatrix} \sum_{i=1}^{n} 1 & \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i2} & \cdots & \sum_{i=1}^{n} x_{ip} \\ \sum_{i=1}^{n} x_{i1} & \sum_{i=1}^{n} x_{i1}^2 & \sum_{i=1}^{n} x_{i1}x_{i2} & \cdots & \sum_{i=1}^{n} x_{i1}x_{ip} \\ \sum_{i=1}^{n} x_{i2} & \sum_{i=1}^{n} x_{i2}x_{i1} & \sum_{i=1}^{n} x_{i2}^2 & \cdots & \sum_{i=1}^{n} x_{i2}x_{ip} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum_{i=1}^{n} x_{ip} & \sum_{i=1}^{n} x_{ip}x_{i1} & \sum_{i=1}^{n} x_{ip}x_{i2} & \cdots & \sum_{i=1}^{n} x_{ip}^2 \end{bmatrix}
\end{aligned}
$$

$\frac{\mathbf{X}'\mathbf{X}}{n} \xrightarrow{p} \mathrm{E}[\mathbf{x}\mathbf{x}']$

$$
\begin{aligned}
\frac{\mathbf{X}'\mathbf{X}}{n} &= \frac{1}{n}\sum_{i=1}^{n} \mathbf{x}_i \mathbf{x}_i' \\
&= \begin{bmatrix}
\frac{\sum_{i=1}^n 1}{n} & \frac{\sum_{i=1}^n x_{i1}}{n} & \frac{\sum_{i=1}^n x_{i2}}{n} & \cdots & \frac{\sum_{i=1}^n x_{ip}}{n} \\
\frac{\sum_{i=1}^n x_{i1}}{n} & \frac{\sum_{i=1}^n x_{i1}^2}{n} & \frac{\sum_{i=1}^n x_{i1}x_{i2}}{n} & \cdots & \frac{\sum_{i=1}^n x_{i1}x_{ip}}{n} \\
\frac{\sum_{i=1}^n x_{i2}}{n} & \frac{\sum_{i=1}^n x_{i2}x_{i1}}{n} & \frac{\sum_{i=1}^n x_{i2}^2}{n} & \cdots & \frac{\sum_{i=1}^n x_{i2}x_{ip}}{n} \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\frac{\sum_{i=1}^n x_{ip}}{n} & \frac{\sum_{i=1}^n x_{ip}x_{i1}}{n} & \frac{\sum_{i=1}^n x_{ip}x_{i2}}{n} & \cdots & \frac{\sum_{i=1}^n x_{ip}^2}{n}
\end{bmatrix} \\
&\xrightarrow{p} \begin{bmatrix}
1 & \mathrm{E}[x_1] & \mathrm{E}[x_2] & \cdots & \mathrm{E}[x_p] \\
\mathrm{E}[x_1] & \mathrm{E}[x_1^2] & \mathrm{E}[x_1x_2] & \cdots & \mathrm{E}[x_1x_p] \\
\mathrm{E}[x_2] & \mathrm{E}[x_1^2] & \mathrm{E}[x_2^2] & \cdots & \mathrm{E}[x_2x_p] \\
\vdots & \vdots & \vdots & \ddots & \vdots \\
\mathrm{E}[x_p] & \mathrm{E}[x_1x_p] & \mathrm{E}[x_2x_p] & \cdots & \mathrm{E}[x_p^2]
\end{bmatrix} \equiv \mathrm{E}[\mathbf{x}\mathbf{x}']
\end{aligned}
$$

Matrix Expression for $\boldsymbol{X}'\boldsymbol{u}$

$$
\begin{aligned}
\boldsymbol{X}'\boldsymbol{u} &= \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & x_{31} & \dots & x_{n1} \\ x_{12} & x_{22} & x_{32} & \dots & x_{n2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ x_{1p} & x_{2p} & x_{3p} & \dots & x_{np} \end{bmatrix} \begin{pmatrix} u_1 \\ u_2 \\ u_3 \\ \vdots \\ u_n \end{pmatrix} \\
&= \begin{pmatrix} \sum_{i=1}^n u_i \\ \sum_{i=1}^n x_{i1} u_i \\ \sum_{i=1}^n x_{i2} u_i \\ \vdots \\ \sum_{i=1}^n x_{ip} u_i \end{pmatrix} = \sum_{i=1}^n \begin{pmatrix} u_i \\ x_{i1} u_i \\ x_{i2} u_i \\ \vdots \\ x_{ip} u_i \end{pmatrix} = \sum_{i=1}^n u_i \begin{pmatrix} 1 \\ x_{i1} \\ x_{i2} \\ \vdots \\ x_{ip} \end{pmatrix} \\
&= \sum_{i=1}^n u_i \boldsymbol{x}_i
\end{aligned}
$$

$$
\begin{aligned}
\frac{\boldsymbol{X}'\boldsymbol{u}}{n} &= \begin{pmatrix} \frac{\sum_{i=1}^{n} u_i}{n} \\ \frac{\sum_{i=1}^{n} x_{i1} u_i}{n} \\ \frac{\sum_{i=1}^{n} x_{i2} u_i}{n} \\ \vdots \\ \frac{\sum_{i=1}^{n} x_{ip} u_i}{n} \end{pmatrix} = \frac{1}{n}\sum_{i=1}^{n} u_i \boldsymbol{x}_i \\
&\xrightarrow{p} \begin{pmatrix} \mathrm{E}[u] \\ \mathrm{E}[x_1 u] \\ \mathrm{E}[x_2 u] \\ \vdots \\ \mathrm{E}[x_p u] \end{pmatrix} \equiv \mathrm{E}[u\boldsymbol{x}] \\
\mathrm{E}[u\boldsymbol{x}] &= \mathrm{E}_{\boldsymbol{x}}\Big[\mathrm{E}_u[u\boldsymbol{x}|\boldsymbol{x}]\Big] = \mathrm{E}_{\boldsymbol{x}}\Big[\boldsymbol{x}\mathrm{E}_u[u|\boldsymbol{x}]\Big] = \mathrm{E}_{\boldsymbol{x}}\Big[\boldsymbol{x}0\Big] = 0
\end{aligned}
$$