STAT 135
15. Multiple linear regression and inference

Spring 2022

Lecturer: Dr Rebecca Barter (*she/her*)
Office hours: Tu 9:30-10:30 (in person), Th 4-5pm (virtual)
Office: Evans 339

Email: rebeccabarter@berkeley.edu
Twitter: @rlbarter
GitHub: rlbarter

Inference for matrix formulation of linear regression

Matrix notation

$$Y = X\beta + \epsilon$$

$$\underset{(n\times 1)}{\begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \vdots \\ y_n \end{bmatrix}} = \underset{(n\times p)}{\begin{bmatrix} 1 & x_{1,1} & x_{1,2} & \dots & x_{1,p-1} \\ 1 & x_{2,1} & x_{2,2} & \dots & x_{2,p-1} \\ 1 & x_{3,1} & x_{3,2} & \dots & x_{3,p-1} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_{n,1} & x_{n,2} & \dots & x_{n,p-1} \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}} + \underset{(n\times 1)}{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_n \end{bmatrix}}$$

The LS prediction is

$$\boxed{\widehat{\boldsymbol{\beta}} = \left(X^T X\right)^{-1} X^T Y}$$

Example

$$sale\_price_i = \beta_0 + \beta_1 area_i + \beta_2 quality_i + \beta_3 year_i + \beta_4 beds_i + \epsilon_i$$

$$Y = X\beta + \epsilon$$

$$\underset{\text{Sale price}}{\underset{(n\times 1)}{\begin{bmatrix} 177000 \\ 130000 \\ 88000 \\ 140000 \\ 125000 \\ 165000 \\ 265900 \\ 127000 \\ 155000 \\ 98000 \end{bmatrix}}} = \underset{(n\times p)}{\begin{array}{ccccc} \text{(Int)} & \text{living\_area} & \text{quality} & \text{year} & \text{beds} \end{array} \atop \begin{bmatrix} 1 & 1913 & 6 & 1928 & 3 \\ 1 & 1709 & 6 & 2004 & 3 \\ 1 & 858 & 4 & 1971 & 3 \\ 1 & 1164 & 7 & 1969 & 3 \\ 1 & 960 & 5 & 1958 & 3 \\ 1 & 1376 & 6 & 1920 & 3 \\ 1 & 1484 & 8 & 2006 & 2 \\ 1 & 864 & 5 & 1972 & 3 \\ 1 & 1200 & 7 & 2004 & 2 \\ 1 & 1436 & 4 & 1910 & 3 \end{bmatrix}} \underset{(p\times 1)}{\begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix}} + \underset{(n\times 1)}{\begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \vdots \\ \epsilon_n \end{bmatrix}}$$

The LS prediction is

$$\boxed{\widehat{\beta} = (X^T X)^{-1} X^T Y}$$

Example

The LS prediction is $\boxed{\widehat{\boldsymbol{\beta}} = \left(X^T X\right)^{-1} X^T Y}$

$$Y = \underset{(n\times 1)}{\begin{bmatrix} 177000 \\ 130000 \\ 88000 \\ 140000 \\ 125000 \\ 165000 \\ 265900 \\ 127000 \\ 155000 \\ 98000 \end{bmatrix}} \qquad X = \underset{(n\times p)}{\begin{bmatrix} 1 & 1913 & 6 & 1928 & 3 \\ 1 & 1709 & 6 & 2004 & 3 \\ 1 & 858 & 4 & 1971 & 3 \\ 1 & 1164 & 7 & 1969 & 3 \\ 1 & 960 & 5 & 1958 & 3 \\ 1 & 1376 & 6 & 1920 & 3 \\ 1 & 1484 & 8 & 2006 & 2 \\ 1 & 864 & 5 & 1972 & 3 \\ 1 & 1200 & 7 & 2004 & 2 \\ 1 & 1436 & 4 & 1910 & 3 \end{bmatrix}}$$

$$\underset{(p\times n)}{X^T} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1913 & 1709 & 858 & 1164 & 960 & 1376 & 1484 & 864 & 1200 & 1436 \\ 6 & 6 & 4 & 7 & 5 & 6 & 8 & 5 & 7 & 4 \\ 1928 & 2004 & 1971 & 1969 & 1958 & 1920 & 2006 & 1972 & 2004 & 1910 \\ 3 & 3 & 3 & 3 & 3 & 3 & 2 & 3 & 2 & 3 \end{bmatrix}$$

$$\underset{(p\times p)}{X^T X} = \begin{array}{l} \\ \text{int} \\ \text{living\_area} \\ \text{quality\_score} \\ \text{year\_built} \\ \text{bedrooms} \end{array} \begin{array}{c} \begin{array}{ccccc} \text{int} & \text{living\_area} & \text{quality\_score} & \text{year\_built} & \text{bedrooms} \end{array} \\ \begin{bmatrix} 10 & 12964 & 58 & 19642 & 28 \\ 12964 & 17937134 & 76704 & 25446006 & 36208 \\ 58 & 76704 & 352 & 114145 & 159 \\ 19642 & 25446006 & 114145 & 38592102 & 54916 \\ 28 & 36208 & 159 & 54916 & 80 \end{bmatrix} \end{array}$$

Example

The LS prediction is $\boxed{\widehat{\boldsymbol{\beta}} = \left(X^T X\right)^{-1} X^T Y}$

$$Y = \underset{(n \times 1)}{\begin{bmatrix} 177000 \\ 130000 \\ 88000 \\ 140000 \\ 125000 \\ 165000 \\ 265900 \\ 127000 \\ 155000 \\ 98000 \end{bmatrix}} \qquad X = \underset{(n \times p)}{\begin{bmatrix} 1 & 1913 & 6 & 1928 & 3 \\ 1 & 1709 & 6 & 2004 & 3 \\ 1 & 858 & 4 & 1971 & 3 \\ 1 & 1164 & 7 & 1969 & 3 \\ 1 & 960 & 5 & 1958 & 3 \\ 1 & 1376 & 6 & 1920 & 3 \\ 1 & 1484 & 8 & 2006 & 2 \\ 1 & 864 & 5 & 1972 & 3 \\ 1 & 1200 & 7 & 2004 & 2 \\ 1 & 1436 & 4 & 1910 & 3 \end{bmatrix}}$$

$$\underset{(p \times n)}{X^T} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1913 & 1709 & 858 & 1164 & 960 & 1376 & 1484 & 864 & 1200 & 1436 \\ 6 & 6 & 4 & 7 & 5 & 6 & 8 & 5 & 7 & 4 \\ 1928 & 2004 & 1971 & 1969 & 1958 & 1920 & 2006 & 1972 & 2004 & 1910 \\ 3 & 3 & 3 & 3 & 3 & 3 & 2 & 3 & 2 & 3 \end{bmatrix}$$

$$\underset{(p \times p)}{(X^T X)^{-1}} = \begin{array}{l} \\ \text{int} \\ \text{living\_area} \\ \text{quality\_score} \\ \text{year\_built} \\ \text{bedrooms} \end{array} \begin{array}{c} \begin{array}{rrrrr} \text{int} & \text{living\_area} & \text{quality\_score} & \text{year\_built} & \text{bedrooms} \end{array} \\ \begin{bmatrix} 722.3664 & -0.0109 & 2.2965 & -0.3457 & -15.1243 \\ -0.0109 & 0.0000 & -0.0002 & 0.0000 & -0.0001 \\ 2.2965 & -0.0002 & 0.1667 & -0.0019 & 0.2462 \\ -0.3457 & 0.0000 & -0.0019 & 0.0002 & 0.0051 \\ -15.1243 & -0.0001 & 0.2462 & 0.0051 & 1.3988 \end{bmatrix} \end{array}$$

Example

The LS prediction is $\boxed{\widehat{\boldsymbol{\beta}} = \left(X^T X\right)^{-1} X^T Y}$

$$\underset{(n\times 1)}{Y} = \begin{bmatrix} 177000 \\ 130000 \\ 88000 \\ 140000 \\ 125000 \\ 165000 \\ 265900 \\ 127000 \\ 155000 \\ 98000 \end{bmatrix}$$

$$\underset{(p\times n)}{X^T} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1913 & 1709 & 858 & 1164 & 960 & 1376 & 1484 & 864 & 1200 & 1436 \\ 6 & 6 & 4 & 7 & 5 & 6 & 8 & 5 & 7 & 4 \\ 1928 & 2004 & 1971 & 1969 & 1958 & 1920 & 2006 & 1972 & 2004 & 1910 \\ 3 & 3 & 3 & 3 & 3 & 3 & 2 & 3 & 2 & 3 \end{bmatrix}$$

$$\underset{(p\times p)}{(X^T X)^{-1}} = \begin{array}{l} \\ \text{int} \\ \text{living\_area} \\ \text{quality\_score} \\ \text{year\_built} \\ \text{bedrooms} \end{array} \begin{array}{c} \begin{array}{rrrrr} \text{int} & \text{living\_area} & \text{quality\_score} & \text{year\_built} & \text{bedrooms} \end{array} \\ \begin{bmatrix} 722.3664 & -0.0109 & 2.2965 & -0.3457 & -15.1243 \\ -0.0109 & 0.0000 & -0.0002 & 0.0000 & -0.0001 \\ 2.2965 & -0.0002 & 0.1667 & -0.0019 & 0.2462 \\ -0.3457 & 0.0000 & -0.0019 & 0.0002 & 0.0051 \\ -15.1243 & -0.0001 & 0.2462 & 0.0051 & 1.3988 \end{bmatrix} \end{array}$$

$$\underset{(p\times 1)}{\hat{\beta}} = (X^T X)^{-1} X^T Y = \begin{array}{l} \text{int} \\ \text{living\_area} \\ \text{quality\_score} \\ \text{year\_built} \\ \text{bedrooms} \end{array} \begin{bmatrix} 731418.02374 \\ 18.97778 \\ 26094.62901 \\ -330.71265 \\ -39533.52530 \end{bmatrix}$$

$\hat{\beta}$ is unbiased

$\hat{\beta}$ is unbiased

Assume that the errors, ϵ_i, are IID with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$

This means that $Cov(\epsilon) = \sigma^2 I_{n \times n}$ and $E(\epsilon) = 0$

The LS estimator is:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

The LS estimate, $\hat{\beta}$ is unbiased:

$$\boxed{E[\hat{\beta}] = \beta}$$

Proof:

$$\begin{aligned}
E\left(\widehat{\boldsymbol{\beta}}\right) &= E\left(\left(X^T X\right)^{-1} X^T Y\right) \\
&= \left(X^T X\right)^{-1} X^T E\left(Y\right) \\
&= \left(X^T X\right)^{-1} X^T E\left(X\boldsymbol{\beta} + \boldsymbol{\epsilon}\right) \\
&= \left(X^T X\right)^{-1} X^T X\boldsymbol{\beta} \\
&= \boldsymbol{\beta}
\end{aligned}$$

$\mathrm{Cov}(\hat{\beta})$

$\hat{\beta}$ variance-covariance matrix

Note if A is a constant matrix and Y is a random vector, then:

$$Cov(AY) = ACov(Y)A^T$$
$$Cov(A + Y) = Cov(Y)$$

Assume that the errors, ϵ_i, are IID with $E(\epsilon_i) = 0$, $Var(\epsilon_i) = \sigma^2$
This means that $Cov(\epsilon) = \sigma^2 I_{n\times n}$ and $E(\epsilon) = 0$

The LS estimator is:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

The LS estimate, $\hat{\beta}$ is has variance-covariance matrix:

$$\underset{(p\times p)}{Cov(\hat{\beta}) = \sigma^2 (X^T X)^{-1}}$$

Proof:

$$\begin{aligned}
Cov\left(\widehat{\boldsymbol{\beta}}\right) &= Cov\left(\left(X^T X\right)^{-1} X^T Y\right) \\
&= \left(X^T X\right)^{-1} X^T Cov(Y) X \left(X^T X\right)^{-1} \\
&= \left(X^T X\right)^{-1} X^T Cov(X\boldsymbol{\beta} + \boldsymbol{\epsilon}) X \left(X^T X\right)^{-1} \\
&= \left(X^T X\right)^{-1} X^T Cov(\boldsymbol{\epsilon}) X \left(X^T X\right)^{-1} \\
&= \left(X^T X\right)^{-1} X^T \sigma^2 I_{n\times n} X \left(X^T X\right)^{-1} \\
&= \sigma^2 \left(X^T X\right)^{-1}
\end{aligned}$$

The distribution of $\hat{\beta}$

The theoretical distribution of the $\hat{\beta}$ vector

Assume that the errors, $\epsilon_i \overset{IID}{\sim} N(0, \sigma^2)$ i.e., $\epsilon \overset{IID}{\sim} N(0, \sigma^2 I_{n \times n})$

The LS estimator is:

$$\hat{\beta} = (X^T X)^{-1} X^T Y$$

And we have shown that:

$$E[\hat{\beta}] = \beta \qquad Cov(\hat{\beta}) = \sigma^2 (X^T X)^{-1}$$

Because the components of $\hat{\beta}$ are linear combinations of independent normal RVs, they are themselves normal, and so

The LS estimate, $\hat{\beta}$ is has distribution:

$$\hat{\beta} \sim N\left(\beta,\ \sigma^2 (X^T X)^{-1}\right)$$

This means that each element

$$\hat{\beta}_j \sim N(\beta_j, \sigma^2 c_{jj})$$

(Where $C = (X^T X)^{-1}$)

But this assumes we know $\hat{\sigma}^2$

Distribution of $\hat{\beta}_j$ when estimating variance

Assume that the errors, $\epsilon_i \overset{IID}{\sim} N(0, \sigma^2)$ i.e., $\epsilon \overset{IID}{\sim} N(0, \sigma^2 I_{n \times n})$

We have that $\hat{\beta}_j \sim N(\beta_j, \sigma^2 c_{jj})$, where $C = (X^T X)^{-1}$ (c_{jj} is the jth diagonal entry)

If we estimate σ^2 with some $\hat{\sigma}^2$, then we can estimate the variance of $\hat{\beta}_j$ as:

$$\hat{\sigma}^2_{\hat{\beta}_j} = \hat{\sigma}^2 c_{jj}$$

Then, we have

$$\frac{\hat{\beta}_j - \beta_j}{\hat{\sigma}_{\hat{\beta}_j}} \sim t_{n-p}$$

But what should $\hat{\sigma}^2$ be?

Estimating σ^2

Assume that the errors, $\epsilon_i \overset{IID}{\sim} N(0,\sigma^2)$ i.e., $\epsilon \overset{IID}{\sim} N(0,\sigma^2 I_{n\times n})$

The variance-covariance matrix is: $Cov(\hat{\beta}) = \sigma^2(X^TX)^{-1}$

Similarly to the 1-dim case, we can estimate σ^2 using the RSS

Let $r = Y - X\hat{\beta}$ be the vector of residuals

Then, an unbiased estimator for σ^2 is:

$$\hat{\sigma}^2 = \frac{RSS}{n-p} = \frac{\|r\|_2^2}{n-p} = \frac{\sum_i r_i^2}{n-p}$$

The proof of unbiasedness requires some understanding of matrix properties such as the *trace.* If you're curious, check out Section 14.4.3 of Rice.

Example

$$\underset{(n\times 1)}{Y} = \begin{bmatrix} 177000 \\ 130000 \\ 88000 \\ 140000 \\ 125000 \\ 165000 \\ 265900 \\ 127000 \\ 155000 \\ 98000 \end{bmatrix} \quad \underset{(n\times p)}{X} = \begin{bmatrix} 1 & 1913 & 6 & 1928 & 3 \\ 1 & 1709 & 6 & 2004 & 3 \\ 1 & 858 & 4 & 1971 & 3 \\ 1 & 1164 & 7 & 1969 & 3 \\ 1 & 960 & 5 & 1958 & 3 \\ 1 & 1376 & 6 & 1920 & 3 \\ 1 & 1484 & 8 & 2006 & 2 \\ 1 & 864 & 5 & 1972 & 3 \\ 1 & 1200 & 7 & 2004 & 2 \\ 1 & 1436 & 4 & 1910 & 3 \end{bmatrix} \quad \underset{(p\times 1)}{\hat{\beta}} = \begin{array}{l} \texttt{int} \\ \texttt{living\_area} \\ \texttt{quality\_score} \\ \texttt{year\_built} \\ \texttt{bedrooms} \end{array} \begin{bmatrix} 731418.02374 \\ 18.97778 \\ 26094.62901 \\ -330.71265 \\ -39533.52530 \end{bmatrix}$$

$$\underset{(n\times 1)}{\hat{Y}} = X\hat{\beta} = \begin{bmatrix} 1 & 1913 & 6 & 1928 & 3 \\ 1 & 1709 & 6 & 2004 & 3 \\ 1 & 858 & 4 & 1971 & 3 \\ 1 & 1164 & 7 & 1969 & 3 \\ 1 & 960 & 5 & 1958 & 3 \\ 1 & 1376 & 6 & 1920 & 3 \\ 1 & 1484 & 8 & 2006 & 2 \\ 1 & 864 & 5 & 1972 & 3 \\ 1 & 1200 & 7 & 2004 & 2 \\ 1 & 1436 & 4 & 1910 & 3 \end{bmatrix} \begin{bmatrix} 731418.02374 \\ 18.97778 \\ 26094.62901 \\ -330.71265 \\ -39533.52530 \end{bmatrix} = \begin{bmatrix} 168075.72 \\ 139070.09 \\ 81644.26 \\ 166396.77 \\ 113973.89 \\ 160530.35 \\ 225861.45 \\ 107522.04 \\ 195038.55 \\ 112786.89 \end{bmatrix}$$

The residual vector is:

$$\underset{(n\times 1)}{r} = Y - \hat{Y} = \begin{bmatrix} 177000 \\ 130000 \\ 88000 \\ 140000 \\ 125000 \\ 165000 \\ 265900 \\ 127000 \\ 155000 \\ 98000 \end{bmatrix} - \begin{bmatrix} 168075.72 \\ 139070.09 \\ 81644.26 \\ 166396.77 \\ 113973.89 \\ 160530.35 \\ 225861.45 \\ 107522.04 \\ 195038.55 \\ 112786.89 \end{bmatrix} = \begin{bmatrix} 8924.283 \\ -9070.088 \\ 6355.741 \\ -26396.771 \\ 11026.114 \\ 4469.649 \\ 40038.554 \\ 19477.958 \\ -40038.554 \\ -14786.886 \end{bmatrix}$$

Example

$$\underset{(n \times 1)}{r} = Y - \hat{Y} = \begin{bmatrix} 177000 \\ 130000 \\ 88000 \\ 140000 \\ 125000 \\ 165000 \\ 265900 \\ 127000 \\ 155000 \\ 98000 \end{bmatrix} - \begin{bmatrix} 168075.72 \\ 139070.09 \\ 81644.26 \\ 166396.77 \\ 113973.89 \\ 160530.35 \\ 225861.45 \\ 107522.04 \\ 195038.55 \\ 112786.89 \end{bmatrix} = \begin{bmatrix} 8924.283 \\ -9070.088 \\ 6355.741 \\ -26396.771 \\ 11026.114 \\ 4469.649 \\ 40038.554 \\ 19477.958 \\ -40038.554 \\ -14786.886 \end{bmatrix}$$

We can estimate σ^2 using:

$$\hat{\sigma}^2 = \frac{\|r\|_2^2}{n-p} = \frac{\sum_i r_i^2}{n-p} = \frac{8924.3^2 + 9070.1^2 + \ldots + 14786.9^2}{10-5} = 968{,}972{,}338$$

Which we can plug into:

$$\hat{\sigma}^2_{\hat{\beta}_j} = \hat{\sigma}^2 c_{jj} \cdots\cdots\blacktriangleright \underset{(p \times p)}{C} = (X^T X)^{-1} = \begin{array}{l} \\ \text{int} \\ \text{living\_area} \\ \text{quality\_score} \\ \text{year\_built} \\ \text{bedrooms} \end{array} \begin{array}{c} \begin{array}{rrrrr} \text{int} & \text{living\_area} & \text{quality\_score} & \text{year\_built} & \text{bedrooms} \end{array} \\ \begin{bmatrix} 722.3664 & -0.0109 & 2.2965 & -0.3457 & -15.1243 \\ -0.0109 & 0.0000 & -0.0002 & 0.0000 & -0.0001 \\ 2.2965 & -0.0002 & 0.1667 & -0.0019 & 0.2462 \\ -0.3457 & 0.0000 & -0.0019 & 0.0002 & 0.0051 \\ -15.1243 & -0.0001 & 0.2462 & 0.0051 & 1.3988 \end{bmatrix} \end{array}$$

So the SD estimate for $\hat{\beta}_{quality}$ is:

$$\hat{\sigma}^2_{\hat{\beta}_{quality}} = \hat{\sigma}^2 c_{quality,quality} = 968{,}972{,}338 \times 0.1667 = 161{,}527{,}689$$

$$\implies \hat{\sigma}_{\hat{\beta}_{quality}} = 12{,}709$$

Hypothesis testing and confidence intervals for $\hat{\beta}$

Hypothesis testing for β_j

Assume that the errors, $\epsilon_i \overset{IID}{\sim} N(0,\sigma^2)$ i.e., $\epsilon \overset{IID}{\sim} N(0,\sigma^2 I_{n\times n})$

$$H_0 : \beta_j = 0 \text{ against } H_1 : \beta_j \neq 0$$

Test statistic: $t = \dfrac{\hat{\beta}_j}{\hat{\sigma}_{\hat{\beta}_j}}$ **P-value:** $P(|T| \geq |t|)$

Where $T \sim t_{n-p}$
(p is the number of parameters in the model)

These standardized coefficients are comparable to one another

Where

$$\hat{\sigma}^2_{\hat{\beta}_j} = \hat{\sigma}^2 c_{jj}$$

And $$\hat{\sigma}^2 = \frac{RSS}{n-p} = \frac{\|r\|_2^2}{n-p} = \frac{\|Y - X\hat{\beta}\|_2^2}{n-p}$$

Confidence interval for β_j

Assume that the errors, $\epsilon_i \overset{IID}{\sim} N(0,\sigma^2)$ i.e., $\epsilon \overset{IID}{\sim} N(0,\sigma^2 I_{n\times n})$

A $(1-\alpha)\,\%$ **confidence interval** is:

$$[\hat{\beta}_j - t_{n-p,\alpha/2}\hat{\sigma}_{\hat{\beta}_j},\ \hat{\beta}_j + t_{n-p,\alpha/2}\hat{\sigma}_{\hat{\beta}_j}]$$

Where

$$\hat{\sigma}^2_{\hat{\beta}_j} = \hat{\sigma}^2 c_{jj}$$

And

$$\hat{\sigma}^2 = \frac{RSS}{n-p} = \frac{\|r\|_2^2}{n-p} = \frac{\|Y - X\hat{\beta}\|_2^2}{n-p}$$

Example

$$\underset{(p \times 1)}{\hat{\beta}} = \begin{array}{l} \text{int} \\ \text{living\_area} \\ \text{quality\_score} \\ \text{year\_built} \\ \text{bedrooms} \end{array} \begin{bmatrix} 731418.02374 \\ 18.97778 \\ 26094.62901 \\ -330.71265 \\ -39533.52530 \end{bmatrix}$$

$$C = (X^T X)^{-1} = \begin{array}{ccccc} \text{int} & \text{living\_area} & \text{quality\_score} & \text{year\_built} & \text{bedrooms} \end{array}$$
$$\begin{bmatrix} 722.3664 & -0.0109 & 2.2965 & -0.3457 & -15.1243 \\ -0.0109 & 0.0000 & -0.0002 & 0.0000 & -0.0001 \\ 2.2965 & -0.0002 & 0.1667 & -0.0019 & 0.2462 \\ -0.3457 & 0.0000 & -0.0019 & 0.0002 & 0.0051 \\ -15.1243 & -0.0001 & 0.2462 & 0.0051 & 1.3988 \end{bmatrix}$$

$$\hat{\sigma}^2 = \frac{\|r\|_2^2}{n-p} = 968{,}972{,}338$$

$\hat{\beta}_j$	$(\hat{\sigma}^2_{\hat{\beta}_j} = \hat{\sigma}^2 c_{jj})$	$(t_{\hat{\beta}_j} = \hat{\beta}_j / \hat{\sigma}_{\hat{\beta}_j})$	$P(\lvert T \rvert \geq \lvert t \rvert)$
$\hat{\beta}_{int} = 731{,}418.02$	$\hat{\sigma}_{\hat{\beta}_{int}} = 836{,}632.00$	$t_{\hat{\beta}_{int}} = 0.874$	$p_{\hat{\beta}_{int}} = 0.4220$
$\hat{\beta}_{area} = 18.98$	$\hat{\sigma}_{\hat{\beta}_{area}} = 35.12$	$t_{\hat{\beta}_{area}} = 0.540$	$p_{\hat{\beta}_{area}} = 0.6122$
$\hat{\beta}_{quality} = 26{,}094.63$	$\hat{\sigma}_{\hat{\beta}_{quality}} = 12{,}708.79$	$t_{\hat{\beta}_{quality}} = 2.053$	$p_{\hat{\beta}_{quality}} = 0.0952$
$\hat{\beta}_{year} = -330.71$	$\hat{\sigma}_{\hat{\beta}_{year}} = 406.57$	$t_{\hat{\beta}_{year}} = -0.813$	$p_{\hat{\beta}_{year}} = 0.4530$
$\hat{\beta}_{beds} = -39{,}533.53$	$\hat{\sigma}_{\hat{\beta}_{beds}} = 36{,}816.08$	$t_{\hat{\beta}_{beds}} = -1.074$	$p_{\hat{\beta}_{beds}} = 0.3320$

Example

```
> ls_fit <- lm(sale_price ~ ., ames_train)
> summary(ls_fit)

Call:
lm(formula = sale_price ~ ., data = ames_train)

Residuals:
     1      2      3      4      5      6      7      8      9     10 
  8924  -9070   6356 -26397  11026   4470  40039  19478 -40039 -14787 

Coefficients:
                   Estimate Std. Error t value Pr(>|t|)  
(Intercept)       731418.02  836632.00   0.874   0.4220  
total_living_area     18.98      35.12   0.540   0.6122  
quality_score       26094.63   12708.79   2.053   0.0952 .
year_built           -330.71     406.57  -0.813   0.4530  
bedrooms           -39533.53   36816.08  -1.074   0.3320  
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 31130 on 5 degrees of freedom
Multiple R-squared:  0.785,	Adjusted R-squared:  0.6129 
F-statistic: 4.563 on 4 and 5 DF,  p-value: 0.06353
```

Example

lm_inference_multi.R