MA50260 Statistical Modelling

Lecture 4: Hypothesis Testing and Confidence Intervals for the Linear Regression Coefficients

Ilaria Bussoli

February 16, 2024

Recap and content of today's lecture

For the linear regression model

$$Y_i = \beta_1 x_{i,1} + \beta_2 x_{i,2} + \ldots + \beta_p x_{i,p} + \epsilon_i, \qquad i = 1, \ldots, n,$$

we saw that the **least square estimate** for $\underline{\beta} = (\beta_1, \ldots, \beta_p)^T$ is

$$\underline{\hat{\beta}} = \left(\mathbf{X}^T \mathbf{X}\right)^{-1} \mathbf{X}^T \mathbf{y}.$$

We further defined the **predicted value**

$$\hat{\mu}_i = \hat{\beta}_1 x_{i,1} + \hat{\beta}_2 x_{i,2} + \ldots + \hat{\beta}_p x_{i,p}, \qquad i = 1, \ldots, n,$$

which can be used (to some degree) for **prediction**.

We are left with estimating $\epsilon_1, \ldots, \epsilon_n$ and σ^2.

Estimation of σ^2

We estimate the residual variance based on the **estimated residuals**

$$\hat{\epsilon}_i = y_i - \hat{\beta}_1 x_{i,1} - \cdots - \hat{\beta}_p x_{i,p}, \qquad i = 1, \ldots, n.$$

The estimate of the residual variance is then

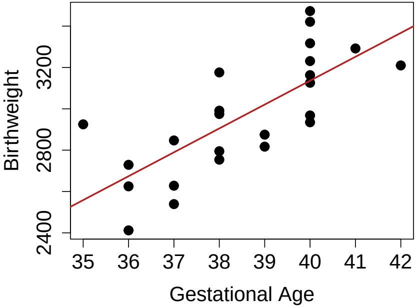
$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^{n} \hat{\epsilon}_i^2.$$

For the birth weight example with $n = 24$ observations, we

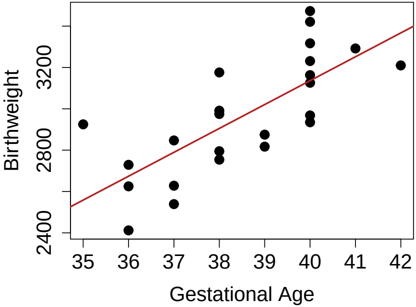
- Derive $\hat{\epsilon}_1, \ldots, \hat{\epsilon}_n$,
- Calculate

$$\hat{\sigma}^2 = \frac{1}{n-p} \sum_{i=1}^{n} \hat{\epsilon}_i^2 \approx 37094.$$

Motivating Example: Birth Weights



Motivating Example: Birth Weights



How certain are we in our estimates?

Is there evidence of a positive relationship between birth weight and gestational age?

What next?

In addition to $\underline{\hat{\beta}}$, we have to measure uncertainty using

- **Confidence intervals** or
- **Standard errors**.

We are further interested in testing significance,

$$H_0 : \beta_j = 0 \qquad \text{vs} \qquad H_1 : \beta_j \neq 0 \qquad (j = 1, \ldots, p).$$

To address these aspects, we have to derive the sampling distribution of the **least square estimator**

$$\underline{\hat{\beta}}(\mathbf{Y}) = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y}.$$

Sampling Distribution of $\underline{\hat{\beta}}(\mathbf{Y})$ (I)

We can express the least square estimator as

$$\underline{\hat{\beta}}(\mathbf{Y}) = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y} = \mathbf{A}\mathbf{Y},$$

where $\mathbf{A} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T$.

Further, the linear regression model implies

$$\mathbf{Y} \sim \mathrm{MVN}_n(\mathbf{X}\underline{\beta}, \sigma^2\mathbf{I}_n).$$

Sampling Distribution of $\underline{\hat{\beta}}(\mathbf{Y})$ (I)

We can express the least square estimator as

$$\underline{\hat{\beta}}(\mathbf{Y}) = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{Y} = \mathbf{A}\mathbf{Y},$$

where $\mathbf{A} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T$.

Further, the linear regression model implies

$$\mathbf{Y} \sim \mathrm{MVN}_n(\mathbf{X}\underline{\beta}, \sigma^2\mathbf{I}_n).$$

Since each row of $\mathbf{AY}$ is a linear combination of $Y_1, \ldots, Y_n$, we have that $\underline{\hat{\beta}}(\mathbf{Y})$ follows a multivariate normal distribution.

Sampling Distribution of $\underline{\hat{\beta}}(\mathbf{Y})$ (II)

We are left with deriving the mean and variance of $\mathbf{AY}$:

$$\mathbb{E}\left[\underline{\hat{\beta}}(\mathbf{Y})\right] = \mathbb{E}\left[\mathbf{AY}\right] = \mathbf{A}\mathbb{E}[\mathbf{Y}] = \mathbf{AX}\underline{\beta} = \mathbf{I}_{p}\underline{\beta} = \underline{\beta}$$

Sampling Distribution of $\underline{\hat{\beta}}(\mathbf{Y})$ (II)

We are left with deriving the mean and variance of $\mathbf{AY}$:

$$\mathbb{E}\left[\underline{\hat{\beta}}(\mathbf{Y})\right] = \mathbb{E}\left[\mathbf{AY}\right] = \mathbf{A}\mathbb{E}[\mathbf{Y}] = \mathbf{AX}\underline{\beta} = \mathbf{I}_p\underline{\beta} = \underline{\beta}$$

and

$$\mathrm{Var}\left[\underline{\hat{\beta}}(\mathbf{Y})\right] = \mathrm{Var}(\mathbf{AY}) = \mathbf{A}\mathrm{Var}(\mathbf{Y})\mathbf{A}^T = \sigma^2\mathbf{AA}^T,$$

with

$$\begin{aligned} \mathbf{AA}^T &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T]^T \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\ &= (\mathbf{X}^T\mathbf{X})^{-1}. \end{aligned}$$

Sampling Distribution of $\underline{\hat{\beta}}(\mathbf{Y})$ (II)

We are left with deriving the mean and variance of $\mathbf{AY}$:

$$\mathbb{E}\left[\underline{\hat{\beta}}(\mathbf{Y})\right] = \mathbb{E}\left[\mathbf{AY}\right] = \mathbf{A}\mathbb{E}[\mathbf{Y}] = \mathbf{AX}\underline{\beta} = \mathbf{I}_p\underline{\beta} = \underline{\beta}$$

and

$$\mathrm{Var}\left[\underline{\hat{\beta}}(\mathbf{Y})\right] = \mathrm{Var}(\mathbf{AY}) = \mathbf{A}\mathrm{Var}(\mathbf{Y})\mathbf{A}^T = \sigma^2\mathbf{AA}^T,$$

with

$$\begin{aligned} \mathbf{AA}^T &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T[(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T]^T \\ &= (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1} \\ &= (\mathbf{X}^T\mathbf{X})^{-1}. \end{aligned}$$

$$\Rightarrow\ \underline{\hat{\beta}}(\mathbf{Y}) \sim \mathrm{MVN}_p\left(\underline{\beta},\, \sigma^2(\mathbf{X}^T\mathbf{X})^{-1}\right)$$

Hypothesis Tests for β_j

We want to test

$$H_0 : \beta_j = b \qquad \text{vs} \qquad H_1 : \beta_j \neq b.$$

The test statistic is

$$t = \frac{\hat{\beta}_j - b}{\sqrt{\hat{\sigma}^2 (\mathbf{X}^T \mathbf{X})^{-1}_{j,j}}}$$

Since

- $\hat{\beta}_j(\mathbf{Y})$ follows a normal distribution,
- $\hat{\sigma}^2(\mathbf{Y})$ follows a χ^2_{n-p} distribution and
- $\hat{\beta}_j(\mathbf{Y})$ is independent of $\hat{\sigma}^2(\mathbf{Y})$,

the test statistic is t-distributed with $n - p$ degrees of freedom under the null hypothesis.

Example: Birth Weights

Recall the model

$$\mathbb{E}(Y_i) = \beta_1 + \beta_2 x_i,$$

and we want to test

$$H_0 : \beta_2 = 0 \qquad \text{vs.} \qquad H_1 : \beta_2 > 0.$$

We have

$$(\mathbf{X}^T\mathbf{X})^{-1} = \begin{bmatrix} 19.6 & -0.507 \\ -0.507 & 0.0132 \end{bmatrix}.$$

and

$$t = \frac{\hat{\beta}_2 - b}{\sqrt{\hat{\sigma}^2(\mathbf{X}^T\mathbf{X})^{-1}_{2,2}}} = \frac{116}{22.2} = 5.22.$$

Compare $t = 5.22$ to the t_{22} distribution. Since $5.22 > 1.72$, there is evidence to **reject** H_0 at the 5% level.

Confidence Interval for β_j

A $100(1-\alpha)\%$ confidence interval for β_j is given by

$$\hat{\beta}_j \pm t_{n-p}(1-\alpha/2) \times \sqrt{\hat{\sigma}^2(\mathbf{X}^T\mathbf{X})^{-1}_{j,j}}.$$

We can use this for testing at the $\alpha\%$ significance level:

- Testing $H_0 : \beta_j = b$ against $H_1 : \beta_j \neq b$
  Calculate the $100(1-\alpha)\%$ confidence interval and reject H_0 if b does **not** lie within it.
- Testing against $H_1 : \beta_j > b$.
  Calculate the $100(1-2\alpha)\%$ confidence interval and reject H_0 if b lies **below** it.

Example: Birth Weights

What is the 95% confidence interval for β_2 in the linear regression model

$$\mathbb{E}(Y_i) = \beta_1 + \beta_2 x_i?$$

We have $\hat{\beta}_2 = 115.5$, $\sqrt{\hat{\sigma}^2(\mathbf{X}^T\mathbf{X})^{-1}_{2,2}} = 22.2$ and $t_{22}(0.975) = 2.074$.

Then the 95% confidence interval for β_2 is

$$\hat{\beta}_2 \pm t_{22}(0.975) \times \sqrt{\hat{\sigma}^2(\mathbf{X}^T\mathbf{X})^{-1}_{2,2}} = 115.5 \pm 2.074 \times 22.2 = (69.5, 161.5).$$

So we reject $H_0 : \beta_2 = 0$ at the 5% significance level for $H_1 : \beta_2 \neq 0$.

Consequently, there is evidence of a positive relationship between birth weight and gestational age.