R01b - Simple linear regression

Uncertainty and prediction intervals

STAT 5870 (Engineering)
Iowa State University

October 29, 2024

Uncertainty when explanatory variable is zero

Let

$$Y_i \stackrel{ind}{\sim} N(\beta_0 + \beta_1 X_i, \sigma^2),$$

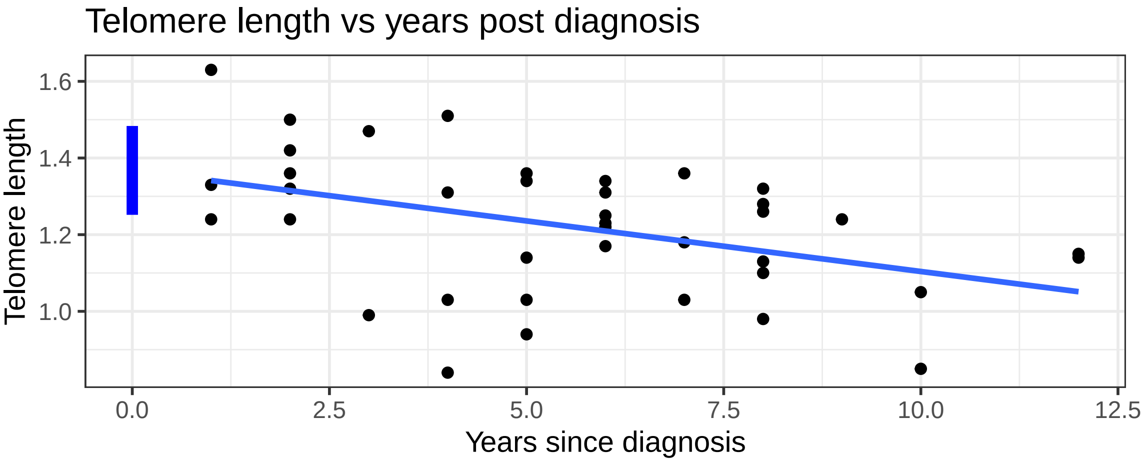
then

$$E[Y_i | X_i = 0] = \beta_0$$

and a $100(1-a)\%$ credible/confidence interval is

$$\hat{\beta}_0 \pm t_{n-2,1-a/2}\, \hat{\sigma} \sqrt{\frac{1}{n} + \frac{\overline{x}^2}{(n-1)s_x^2}}.$$

Telomere data: uncertainty



Uncertainty when explanatory variable is x

Let

$$Y_i \stackrel{ind}{\sim} N(\beta_0 + \beta_1 X_i, \sigma^2),$$

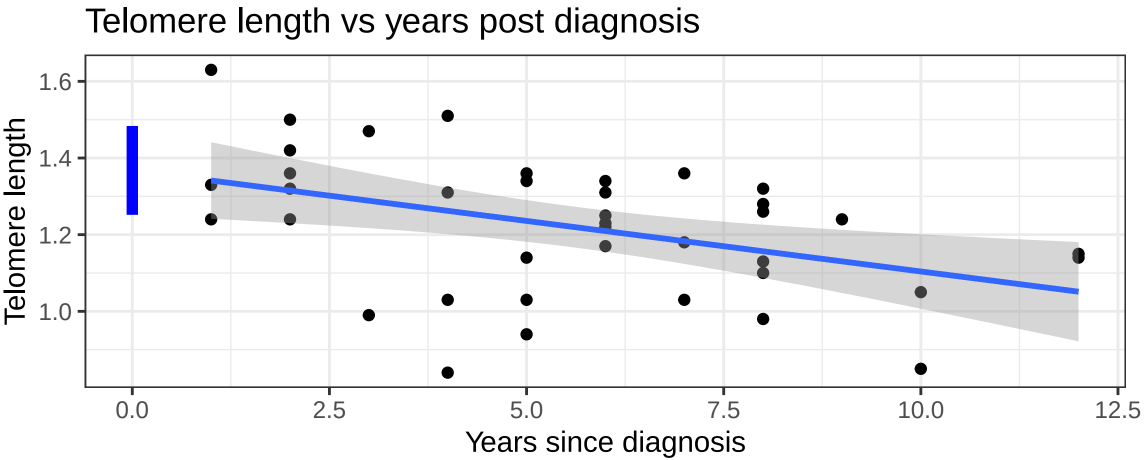
then

$$E[Y_i | X_i = x] = \beta_0 + \beta_1 x$$

and a $100(1-a)\%$ credible/confidence interval is

$$\hat\beta_0 + \hat\beta_1 x \pm t_{n-2,1-a/2}\, \hat\sigma \sqrt{\frac{1}{n} + \frac{(\overline{x} - x)^2}{(n-1)s_x^2}}.$$

Telomere data: uncertainty



Prediction intervals

Let

$$Y_i \stackrel{ind}{\sim} N(\beta_0 + \beta_1 X_i, \sigma^2),$$

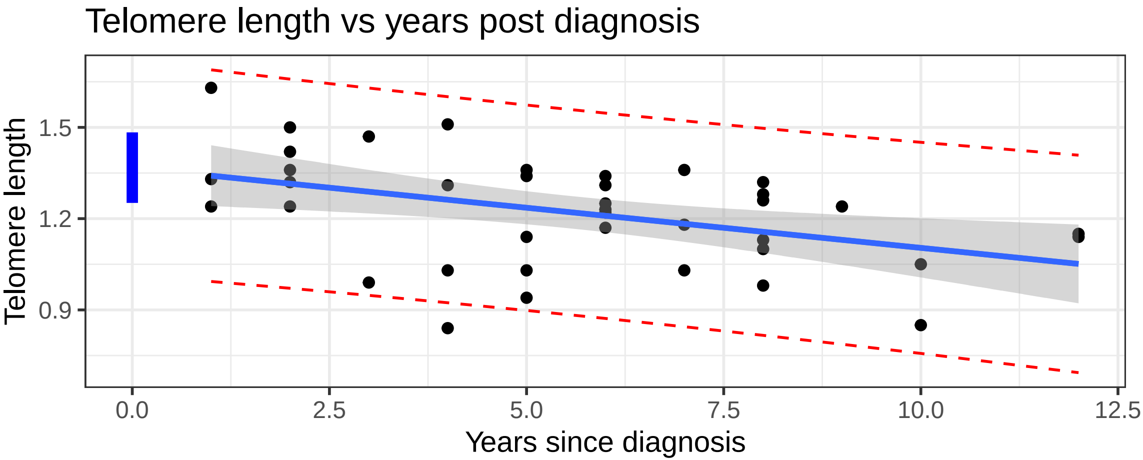
then

$$E[Y_i | X_i = x] = \beta_0 + \beta_1 x$$

and a $100(1-a)\%$ prediction interval is

$$\hat\beta_0 + \hat\beta_1 x \pm t_{n-2,1-a/2}\, \hat\sigma \sqrt{1 + \frac{1}{n} + \frac{(\overline{x} - x)^2}{(n-1)s_x^2}}.$$

Telomere data: prediction intervals



Summary

Two main types of uncertainty intervals:

- where is the line?

$$\hat{\beta}_0 + \hat{\beta}_1 x \pm t_{n-2,1-a/2}\, \hat{\sigma} \sqrt{\quad \frac{1}{n} + \frac{(\overline{x} - x)^2}{(n-1)s_x^2}}$$

- where will a new data point fall?

$$\hat{\beta}_0 + \hat{\beta}_1 x \pm t_{n-2,1-a/2}\, \hat{\sigma} \sqrt{1 + \frac{1}{n} + \frac{(\overline{x} - x)^2}{(n-1)s_x^2}}$$

Both intervals are confidence and credible intervals.