Regression model

\[ Y = \mu_{y|x} + \varepsilon \]

- \( Y \) represents a value of the response variable.
- \( \mu_{y|x} \) represents the population mean response for a given value of the explanatory variable, \( x \).
- \( \varepsilon \) represents the random error.

Linear Regression Model

\[ Y = \mu_{y|x} + \varepsilon = \beta_0 + \beta_1 x + \varepsilon \]

- \( \beta_0 \) is the Y-intercept parameter.
- \( \beta_1 \) is the slope parameter.
Estimated Mean

- Estimated Mean Response
  \[ \hat{\mu}_{y|x} = \hat{\beta}_0 + \hat{\beta}_1 x \]
- \( 2007: \text{CO}_2 = 382.43 \text{ ppmv} \)
  \[ \hat{\mu}_{y|x} = 9.8815 + 0.012584(382.43) \]
  \[ \hat{\mu}_{y|x} = 14.694 \text{ °C} \]

Confidence Interval

- We have estimated a population mean response for a given value of the explanatory variable. We can expand this into a confidence interval.

Confidence Interval

\[ \hat{\mu}_{y|x} \pm t^* \text{se}(\hat{\mu}_{y|x}) \]
\[ \text{se}(\hat{\mu}_{y|x}) = \sqrt{MS_{Error} \left( \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x - \bar{x})^2} \right)} \]
\[ t^* \text{ from } t \text{- table with } df = n - 2 \]
Standard Error Calculation

\[ se(\hat{\mu}_{y|x}) = \sqrt{MSE_{\text{Error}} \left( \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum (x - \bar{x})^2} \right)} \]

\[ MSE_{\text{Error}} = 0.010725, \sum (x - \bar{x})^2 = 5061.2 \]

\[ n = 20, x_0 = 382.43, \bar{x} = 341.23 \]

\[ se(\hat{\mu}_{y|x}) = 0.0643 \]

Confidence Interval

\[ \hat{\mu}_{y|x} \pm t^* \cdot se(\hat{\mu}_{y|x}) \]

\[ se(\hat{\mu}_{y|x}) = 0.0643 \]

\[ t^* = 2.101 \]

\[ 14.694 \pm 2.101(0.0643) \]

\[ 14.559 \, ^{\circ}\text{C} \text{ to } 14.829 \, ^{\circ}\text{C} \]

Interpretation – Part 1

The population mean temperature when the CO$_2$=382.43 ppmv can be any value between 14.56 $^{\circ}$C and 14.83 $^{\circ}$C.
Interpretation – Part 2

- We are 95% confident that intervals based on random samples from the population with capture the actual population mean value.
- This is confidence in the process.

\[ \mu_{y|x} = \beta_0 + \beta_1 x \]
Predicted Individual

- Predicted Individual Response
  \[ \hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 x \]
- 2007: CO₂ = 382.43 ppmv
  \[ \hat{y} = 9.8815 + 0.012584(382.43) \]
  \( \hat{y} = 14.694°C \)

Prediction Interval

- We have predicted an individual response for a given value of the explanatory variable. We can expand this into a prediction interval.

Prediction Interval

\[ \hat{y} \pm t^* se(\hat{y}) \]

\[ se(\hat{y}) = \sqrt{MS_{Error} \left( 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x - \bar{x})^2} \right)} \]

- \( t^* \) from t-table with \( df = n - 2 \)
Standard Error Calculation

\[ se(\hat{y}) = \sqrt{MS_{Error} \left( 1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{\sum(x - \bar{x})^2} \right)} \]

\[ MS_{Error} = 0.010725, \sum(x - \bar{x})^2 = 5061.2 \]

\[ n = 20, x_0 = 382.43, \bar{x} = 341.23 \]

\[ se(\hat{y}) = 0.1219 \]

For the second part:

\[ se(\hat{y}) = \sqrt{MS_{Error} + se(\hat{\mu}_{y|x})^2} \]

\[ MS_{Error} = 0.010725, se(\hat{\mu}_{y|x}) = 0.0643 \]

\[ se(\hat{y}) = \sqrt{0.010725 + (0.0643)^2} \]

\[ se(\hat{y}) = 0.1219 \]

Prediction Interval

\[ \hat{y} \pm t^* se(\hat{y}) \]

\[ se(\hat{y}) = 0.1219 \]

\[ t^* = 2.101 \]

\[ 14.694 \pm 2.101(0.1219) \]

\[ 14.438 \, ^\circ C \text{ to } 14.950 \, ^\circ C \]
Interpretation

- We are 95% confident that the annual global temperature when the CO₂ = 382.43 ppmv can be any value between 14.44 °C and 14.95 °C.

\[ \mu_{y|x} = \beta_0 + \beta_1 x \]