Simple Linear Regression

Question

- Is annual carbon dioxide concentration related to annual global temperature?

---

Simple Linear Regression

- Response variable, \( Y \).
  - Annual global temperature (\(^\circ\) C).
- Explanatory (predictor) variable, \( x \).
  - Annual atmospheric CO\(_2\) concentration.

---

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 Model

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

- \( \beta_0 \) The Y-intercept parameter.
- \( \beta_1 \) The slope parameter.

Conditions

- The relationship is linear.
- The random error term, \( \varepsilon \), is
  - Independent
  - Identically distributed
  - Normally distributed with standard deviation, \( \sigma \).

\[ \mu_{y|x} = \beta_0 + \beta_1 x \]
Describe the plot.
- Direction – positive/negative.
- Form – linear/non-linear.
- Strength.
- Unusual points?

**CO₂ and Temperature.**
- There is a fairly strong, positive linear relationship between CO₂ and temperature.
- Larger (smaller) values of CO₂ are associated with larger (smaller) values of temperature.
Method of Least Squares

- Find estimates of $\beta_0$ and $\beta_1$ such that the sum of squared vertical deviations from the estimated straight line is the smallest possible.

Least Squares Estimates

Slope: \[
\hat{\beta}_1 = \frac{\sum (x - \bar{x})(y - \bar{y})}{\sum (x - \bar{x})^2}
\]

Intercept: \[
\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}
\]

Line of Best fit: \[
\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x
\]
Stat 301– Lecture 6

### Linear Fit

\[ \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x \]

- Predicted Temperature = 9.8815 + 0.012584\*CO₂

### Interpretation

- Estimated Y-intercept.
  - This does not have an interpretation within the context of the problem.
  - Having no CO₂ in the atmosphere is not reasonable given the data.
Interpretation

- Estimated slope.
- For each additional 1 ppmv of CO₂, the annual global temperature goes up 0.012584 °C, on average.

How Strong?

- The strength of a linear relationship can be measured by $R^2$, the coefficient of determination.
- RSquare in JMP output.
How Strong?

\[ R^2 = \frac{SS_{Model}}{SS_{Total}} \]

\[ R^2 = \frac{0.80145}{0.99450} = 0.806 \]

Interpretation

- 80.6% of the variation in the global temperature can be explained by the linear relationship with carbon dioxide concentration.
- 19.4% is unexplained.

Interpretation

- There is a fairly strong positive linear relationship between carbon dioxide concentration and global temperature.
- Cause and effect?
Cause and Effect?

- There is a strong positive linear relationship between the number of 2nd graders in communities and the number of crimes committed in those communities.

Connection to Correlation

- If you square the correlation coefficient, $r$, relating carbon dioxide to global temperature you get $R^2$, the coefficient of determination.

$$ r = \pm \sqrt{R^2} = \pm \sqrt{0.806} = \pm 0.898 $$

Connection to Correlation

$$ \hat{\beta}_1 = r \left( \frac{s_y}{s_x} \right) $$

$s_y$ is the standard deviation of the $y$ values

$s_x$ is the standard deviation of the $x$ values