Outliers and Influential Points

READING:	1.5
EXERCISES:	CH.1: 33, 43, 47
ASSIGNED:	HW 4
PRODUCER:	DR. MARIO

IMG CREDIT: ALEX RIEGERT-WATERS

Types of "Unusual Points"

- **Outlier:** Observation with a Residual Far Away From Zero
  - Recall: Mathematical Definition from Boxplots
  - Far Away Depends on the Standard Error of the Regression and the Distribution of the Residuals

- **Influential Point:** Observation Heavily Effects the Regression Fit
  - Determined by How the y-Intercept and Slope Change When the Observation is Removed
  - Removal will Drastically Impact Predictions/ Fitted Values

Detecting Outliers

- **Standardized Residual:** Adjust by Standard Error of the Regression

$$\frac{y - \hat{y}}{\hat{\sigma}_{\epsilon}}$$

- **Studentized Residual:** Use Standard Error of Regression After Removing the Point from the Regression

- Under Conditions, 95% of Residuals Within -2 and +2

Detecting Influential Points

- Remove Each Observation and See How the Line Changes
- Points Farther Away from the Mean $\bar{x}$ Have More Impact
- Fewer Data Points Leads to Each Point Having More Impact
- **Leverage:** Formula Based Off Distance From $\bar{x}$ and sample size n
- Future Information in Chapter 4

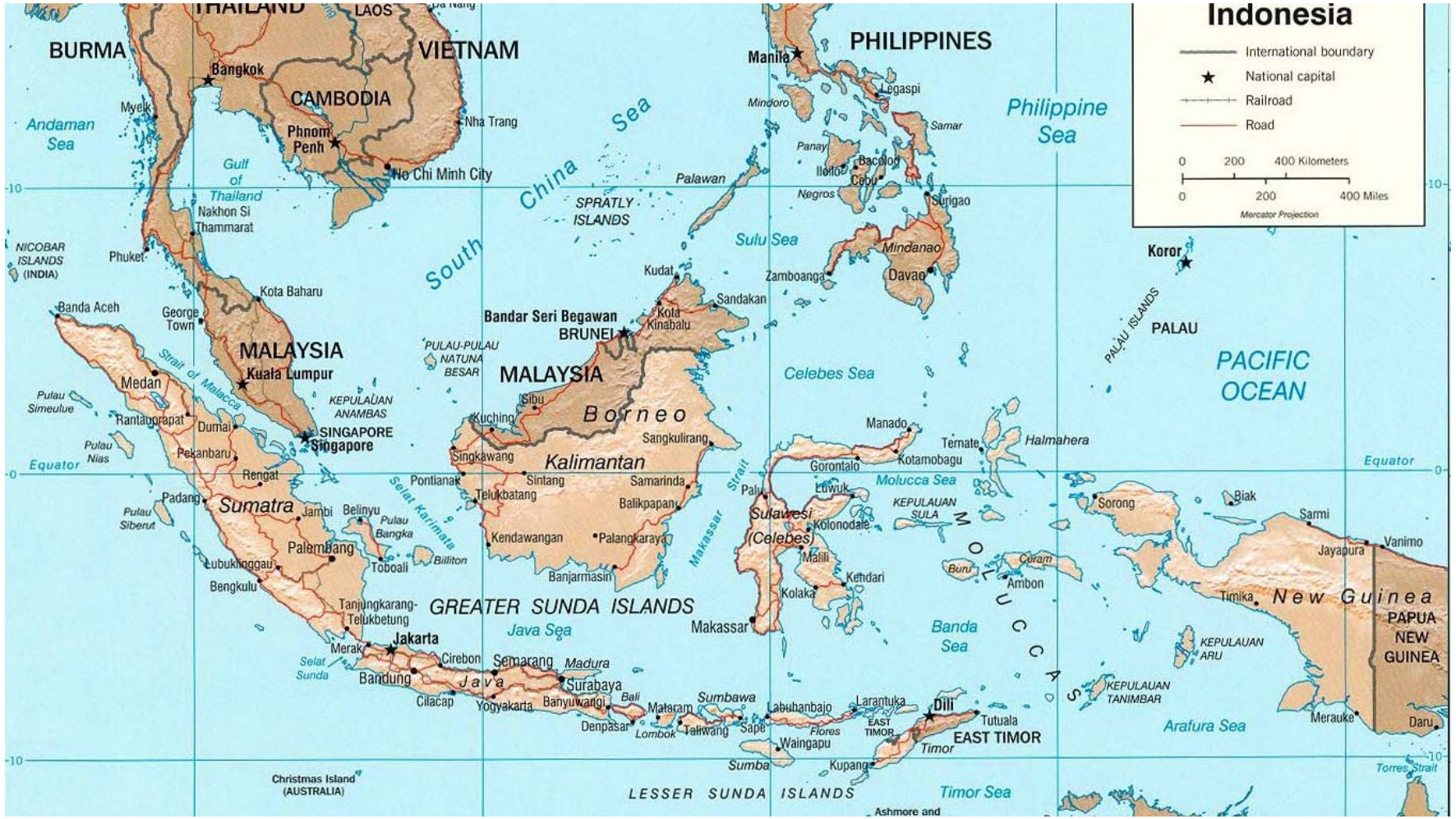
Extrapolation

- **Extrapolation:** Using Model to Make Prediction for an Unusual Value of the Predictor Variable *X*
- Highly Discouraged
- No Guarantee Model Pattern Continues Outside Range of *X*

Example: Mammal Species

- Y = Number of Mammal Species on an Island
- X = Area of the Island
- Question: *Do bigger islands tend to have a larger variety of mammals?*
- Data from 14 Islands in Southeast Asia

Example: Mammal Species



Supplement for Lecture 7

- Inspect for Outliers Using Default Plots from the **lm()** Function
- Perform Transformations to Meet Linearity Condition
- Obtain Fitted Values and Residuals from Models
- Extrapolate Versus Interpolate

Methamatics

We fit our transformed model and get

$$\widehat{\log(y)} = 3 + 5\log(x)$$

However, we want to Predict *y* and not *log(y)*, therefore

$$\hat{y} = e^{\widehat{\log(y)}} = e^{3+5\log(x)} = e^3 e^{5\log(x)} = e^3 e^{\log(x^5)} = e^3 x^5$$

In general,

$$\hat{y} = e^{\widehat{\log(y)}} = e^{\widehat{\beta_0}} x^{\widehat{\beta_1}}$$

Thank You

Make Reasonable Decisions