**Making Predictions with Linear Models: A Murder Mystery Case Study**

**Author:** Miranda Chen Musgrove, University of Colorado, Boulder, [mich6131@colorado.edu](mailto:mich6131@colorado.edu)

**Course Information**

Department: **Ecology and Evolutionary Biology (EBIO)**

Level: **Lower Undergraduate**

Course type: **Both**

Students: **Majors**

Number of Students: **100 students**

**Focus:**

This Swirl lesson will introduce students to the idea of confidence intervals (CIs; or uncertainty) around linear slopes versus individual values. Often predicting means are easier than predicting individual values (e.g. estimating what movie genre is popular in the U.S. versus estimating what movie genre YOU as an individual enjoy), and the uncertainty surrounding such individual values are much greater than around the means. For example, we are confident from surveys that 18-year-old white men often prefer action movies and buddy comedies over rom-coms generally, but we are much less confident predicting the tastes as a single individual 18-year-old male. Students will be presented with a crime scene investigation/murder mystery challenge where they need to create a linear model and use it to determine a y-value (time of death) of a particular x-value (hypoxanthine concentration in a posthumous eye). Data for hours after death and hypoxanthine concentration will be provided. They will be asked to calculate with both mean CIs and individual CIs for the data around the slope. By the end of the lesson, students will know how to visualize linear regressions, confidence intervals, and be able to predict individual values for the murder mystery.

**Overview:**

This Swirl lesson will introduce two main concepts to students: 1) the idea of uncertainty around linear slopes versus individual values and 2) predicting values using linear models. Students will be presented with a crime scene investigation/murder mystery challenge where they need to create a linear model and use it to determine a y-value (time of death) of a particular x-value (hypoxanthine concentration in a posthumous eye). Students will be encouraged to develop hypotheses for what happened to the dead victim--if the individual was murdered or not. Data for hours after death and hypoxanthine concentration will be provided. They will be asked to calculate the time of death with both mean CIs and individual CIs for the data around the slope. Lastly, students will interpret results to then evaluate their hypotheses and determine whether they have enough evidence to conclude that a murder took place or not.

**Learning objectives:**

By the end of the lesson, students will know how to:

1. Generate multiple hypotheses for the case study
2. Construct a linear model based on data for the amount of hypoxanthine and time of death
3. Estimate the prediction uncertainty for the time of death of a single individual based on a measurement of the amount of hypoxanthine in the eye
4. Analyze and interpret data for evaluating alternative hypotheses
5. Explain why predicting the value of an individual is more uncertain than predicting the mean for a population using a statistical linear model

**Lesson sequence:** Provide a numbered, ordered list of the activities within your swirl lesson. This list can be taken from step 4 in your initial lesson design, with any modifications that were introduced.

1. Present crime scene investigation/murder mystery case study challenge to students. Students can develop multiple hypotheses for what occurred.
2. Students will construct a linear model based on case study data (amount of hypoxanthine and time of death)
3. Students will predict the y value (the time of death) of a single individual and the confidence intervals based on a given x (a measurement of the amount of hypoxanthine in the eye)
4. Students will determine if the data support the hypotheses of murder or not.

**Pre-lesson activities:** Before asking students to conduct the Swirl Lesson, it may be good to present the case study ahead of time and actually go through different hypotheses for if a murder took place or not. This could be in small groups where students outline what the time of dealth would look like if murder did and did NOT occur.

**Implementation notes:** There are many lines in which Swirl simply asks students to type in given code. Perhaps consider prompting students to type it out to familiarize themselves with coding syntax.

**Crime Scene Scenario:**

A United States (U.S.) Senator was traveling to Venezuela for top secret and controversial negotiations between the two countries. The airplane carrying the senator goes down in a remote jungle location. There is some intelligence evidence suggesting the plane may have been deliberately crashed because the U.S. Senator was killed prior to the crash, soon after the plane left Miami. Crime scene investigators get to the crash site about 2 days (25 hours) after the crash. The flight time from Miami to Venezuela is about 4 hours.

Develop a model to assess whether it is possible to determine whether the senator was murdered soon after take-off and the plane crashed to cover up the murder or whether the Senator died as a consequence of the plane crash. The time of death can be estimated using the amount of hypoxanthine in the vitreous humour of the eye (the fluid in the eye). The amount of hypoxanthine in the eye of the dead senator received 25 hours after the plane crashed was 103.