ZOO 800

Homework Week 9

Submission instructions

Submit a single URL to a public GitHub repository on Canvas. Please make sure it works – i.e., that you can clone the repo as a project yourself. Be sure to indicate in the submission who is in your group. Submit a single URL for each group, but if you're not the one submitting the URL, submit a comment mentioning the name of the person submitting for your group.

Problem

One important use of probability distributions is to test the ability of statistical tests to provide an accurate and unbiased estimate of effects of known size and determine whether they are significantly different from some reference value – usually zero (power analysis). More broadly, the combination of simulating data from a known process and then attempting to estimate parameters of the data generation process is called a simulation-estimation framework. Here we will begin to explore this approach using a few different probability distributions: the uniform, normal, binomial, and beta.

Objective 1

A. Use the equation for a simple linear regression to generate 100 observations of *x* and *y*. Consistent with the assumptions of linear regression, *x* should be measured without error, but the observed values of *y* should include normally distributed error with mean zero.

$$y_i = \alpha + \beta x_i + \varepsilon_i.$$

where y_i is an individual observation of the response (dependent) variable, α is the intercept, β is the slope, x_i is an individual observation of the predictor (independent) variable, and ε_i is a random variable drawn from a normal distribution with mean = 0 and standard deviation = σ.

You can make up values for the constants α and β such that the y_i fall between anywhere betwen -100 and +100 and draw random values of x_i from a uniform distribution between 0 and 10.

B. Using facet wrapping in ggplot, create a multipanel figure comparing plots of y vs. x in one row for three different values of σ: 1, 10, and 25.

C. How does the ability to visually detect a relationship between y and x change as observation error increases (no answer or analysis needed – just think about this).

Objective 2

Recalling Don Corleone's loaded coin, we're interested in knowing how many coin flips are required to be able to consistently determine whether the coin is unfair ($p(heads)>0.5$) for different degrees of unfairness.

A. Using simulations of coin flips (Bernoulli trials), plot the probability (number of times out of 100) that you determine the coin is significantly unfair ($alpha < 0.05$) for 1 to 20 coin flips when $p = 0.55$.

B. Repeat this analysis for $p = 0.6$ and $p = 0.65$ and add these lines to the plot from A.