Audrey Constant
9:30 am

February 5th Notes

Summary: We learned how to use the Gaussian (or normal) linear regression model to quantify parameter (particularly the slope and intercept of the regression model) uncertainty and to create prediction intervals that don't ignore parameter uncertainty.

1. Standard Deviation of Residuals = Average Error
   a. Average error of using systematic part of regression equation (slope and intercept) to predict Y(i) is standard deviation of residuals
   b. 2 sd from the regression line will cover ~ 95% of residual points

2. Fit Model on Transformed Scale
   a. Fit model and do all predictions on transformed scale
   b. When moving from log-log to original, the prediction interval will be two curves that fan out because error is greater for bigger values of x (Errors are additive on log scale and multiplicative on original scale)
   c. Last thing you do is undo the transformation

3. Gaussian Linear Regression Model
   a. Key Terms
      i. Sampling Distribution
         1. Distribution of the estimates of beta hat 1 (B^1)
         2. A claim about the least squares procedure
         3. A narrow sampling distribution is more precise
      ii. Standard Error – computed by taking the standard deviation of the sampling distribution (estimates of parameter)
      iii. Bootstrapping
         1. Samples act as entire populations
         2. Resampling with replacement from original sample
         3. Use same sample size as the original sample in order to mimic real samples of the population
      iv. Confidence – stability of the estimate

   b. Normality (linearity) assumption
      i. $Ei \sim N(0, \sigma^2)$
      ii. Mad Libs Example

1. Key takeaway – factors left out of the model change the predicted value of y; aggregation of all factors determine y
2. Residual is the sum total of all things that affect y variable that aren't in the x variable

iii. Schools of Minnows Example
   1. A school of minnows (Gaussian distribution) is an aggregation of each individual minnow's movements

iv. Normal Approximation to the Binomial Distribution – according to central limit theorem and the normality assumption, binomial distributions look like normal distributions

v. σ^2 = residual variance, σ = residual sd

vi. See page 111 of course packet for equations

vii. Bootstrapping and summary function in R use normality assumption to find the standard error