Stat 414 – Day 3
Parameter Estimation

Last Time

- Adjusted vs. Unadjusted associations, Transformations vs. Polynomials, Advantages of Centering, Interpreting interactions
- (Day 2b) Can *model* heterogeneity
    - GLS is a variation of weighted least squares that iterates between estimating regression slopes and variance terms

Example 1: Airfares from San Luis Obispo to a "random" sample of 12 major U.S. cities as found March 31, 2014 on Travelocity.com for travel on May 8-May 12, 2017.

631.8	338.6	627.9	352.6	699.8	470.7	557.8	547.6	569.83	321.1	344.7	427.6

(a) Identify the observational units

In the population of American cities, μ represents the population mean airfare and σ the city-to-city variation in airfares.
(b) Suggest an estimator for μ. Why do you think this is a good estimator?

Recall: Least Squares estimators minimize the sum of the squared "errors" $\sum_{i=1}^{n}(y_i - k)^2$

(c) How can we find the value of *k* that minimizes this sum?

(d) In R, fit the "null model" (intercept only). What is the estimate of the intercept? What is the value of *SSE = sum of squared errors*? How is this related to *Residual standard error*?

(e) How do we estimate σ? (the "random noise" variance, verify in R, interpret)

(f) If we know distances to the cities, x_i, how can we find the values for β_0 and β_1 that minimize $\sum_{i=1}^{n}(y_i - \beta_0 + \beta_1 x_i)^2$

(g) But what about σ^2, how do we estimate the "random noise" variance? (Verify in R, interpret)

Measures of model fit

- Want MSE = $\hat{\sigma}^2$ to be small
- R^2 = proportion of variation in *y* explained by the model, 1 – SSError/SSTotal
- R^2_{adj} = 1-(MSE/MSTotal) = 1 – SSE/(n – p – 1)/(SST/(n – 1)) = R^2 – p/(n – (p + 1))(1 – R^2)
    - "MSTotal" estimates the unexplained variability in the response variable
    - MSE measures the unexplained variation in the response after adjusting for model

(h) **Inference:** Verify F = (drop in SSE)/(difference in df) / MSE_{full}, = $\frac{R^2/1}{(1-R^2)/10}$, df = 1, 10

There are alternatives to least squares estimation. One method we will see is *maximum likelihood estimation*. The *likelihood* is the pdf of a random variable, but viewed as a function of the parameter(s). We want to choose parameter values that maximize the likelihood of observing the data we have. Again, we are trying to match what we observe with what we expect to see. One advantage of MLEs is they work for non-normal distributions.

Suppose our data follow a normal distribution $L(\mu, \sigma; y) = 1/(\sigma\sqrt{2\pi})\, e^{-(y-\mu)^2/2\sigma^2}$.
With independent observations, the *joint likelihood* will be the product $1/(\sigma\sqrt{2\pi})^n \; e^{-\Sigma(y_i-\mu)^2/2\sigma^2}$.

(h) So how do we find the values of μ that maximize this function for our observed data? What about the linear model?

For our simple model (assuming independence, normality) the ML slope and intercept estimates will be the same as with LS estimation. But we will judge models by the value of the likelihood function at those parameter estimates, $L = -\frac{n}{2}\ln(2\pi) - \frac{n}{2}\ln(\frac{SSE}{n}) - \frac{n}{2}$.

(i) In R, verify the value of the log likelihood for the null model

(j) In R, determine the value of the log likelihood for the linear model. Which is better?

More measures of model fit (aka information criteria)
Want maximize the log likelihood but can also penalize you for the number of parameters
Here: p = number of parameters (intercept, slopes, σ)

- Want (2)(log)likelihood values to be large
- Want small BIC = -2 x log-likelihood + p x ln(n)
- Want small AIC = -2 x log-likelihood + 2p

(k) Verify the AIC value in R.