Luke Watson

10-8-24

**Exercise 5**

*Part One*

# plot a normal density curve, create data in-line

plot(function(x) dnorm(x), -3, 3)

# compute likelihood ratio of getting either 1 or 2 using this distribution

prob\_one <- dnorm(1)

prob\_two <- dnorm(2)

lr\_one\_to\_two <- prob\_one/prob\_two

# LR is 4.48 in favor of probability of getting one given normal dist.

# now how much more likely is 0 to 1 2 and 3

prob\_zero <- dnorm(0)

prob\_three <- dnorm(3)

lr\_zero\_to\_one <- prob\_zero/prob\_one

lr\_zero\_to\_two <- prob\_zero/prob\_two

lr\_zero\_to\_three <- prob\_zero/prob\_three

# make into df

lr\_table <- data.frame(

comparison\_to\_zero = c(1, 2, 3),

lr = c(lr\_zero\_to\_one, lr\_zero\_to\_two, lr\_zero\_to\_three)

)

lr\_table

# comparison\_to\_zero LR in favor of zero

#1 1 1.648721

#2 2 7.389056

#3 3 90.017131

# repeat the above except now using a mean of 50 (no longer standard, but same)

plot(function(x) dnorm(x, mean = 50), 47, 53)

prob\_51 <- dnorm(51, mean = 50)

prob\_52 <- dnorm(52, mean = 50)

prob\_53 <- dnorm(53, mean = 50)

prob\_50 <- dnorm(50, mean = 50)

lr\_50\_to\_51 <- prob\_50/prob\_51

lr\_50\_to\_52 <- prob\_50/prob\_52

lr\_50\_to\_53 <- prob\_50/prob\_53

lr\_50\_table <- data.frame(

comparison\_to\_zero = c(1, 2, 3),

lr = c(lr\_50\_to\_51, lr\_50\_to\_52, lr\_50\_to\_53)

)

# combine to see there is no difference

full\_lr\_table <- rbind(

lr\_table, lr\_50\_table

)

full\_lr\_table

# comparison\_to\_zero lr

# 1 1 1.648721

# 2 2 7.389056

# 3 3 90.017131

# comparison\_to\_fifty

# 4 1 1.648721

# 5 2 7.389056

# 6 3 90.017131

# now use lr to compare binomial outcomes

seventy\_prob <- dbinom(20, 25, .7)

fifty\_prob <- dbinom(20, 25, .5)

# now compare

lr\_70\_to\_fifty <- seventy\_prob / fifty\_prob

# LR is 65.1 in favor of the 70% to press left hypothesis

*Part Two*

Currently, major league baseball teams are duking it out in a winner takes all playoff for a chance to be crowned the world series champion. Every year, new storylines emerge with unexpected teams and players leading their teams to glory or an early exit from the playoff race. What sorts of factors from the regular season predict playoff performance? This is a question that has perplexed those who report and chronicle baseball throughout the years, and with my analysis, I was hoping to give a little more insight into how a particular class of player performance metrics, pitch dynamics like velocity and break, influence subsequent playoff performance in terms of runs allowed (earned run average or ERA). To start, I conducted a linear mixed effects model with fastball (4-seam, 2-seam) average velocity, offspeed pitch (changeup, splitter, forkball) average break, and breaking pitch (slider, curveball, slurve) average break as fixed effects with a random effect of player predicting the log-transformed ERA of that player in the subsequent postseason. The adjustment of log-transforming ERA is due to frequent outliers of players with poor performances (several runs in an inning or less) having massively inflated ERAs of 50, 100, and even 150. In this model, no fixed effects were considered significant. I opted to run another model that included these three same predictors of fastball average velocity, offspeed average break, and breaking pitch average break as well as all of their interaction. To preemptively adjust for multicollinearity, I centered each of the predictors. As with the previous model, no predictors reached significance. The estimates of these two models are compared in Table 1.

**Table 1**

|  |  |  |  |
| --- | --- | --- | --- |
| Model/Term | Estimate | SE | t |
| Model 1 |  |  |  |
| Intercept | .32 | 2.00 | .16 |
| Fastball velocity | .01 | .02 | .52 |
| Offspeed break | .01 | .01 | .32 |
| Breaking break | .02 | .02 | .76 |
| Model 2 |  |  |  |
| Intercept | 1.64 | .06 | 29.00 |
| Fastball velocity | .01 | .02 | .21 |
| Offspeed break | .02 | .02 | .84 |
| Breaking break | .01 | .01 | .24 |
| Fastball velocity x offspeed break | .01 | .01 | 1.28 |
| Fastball velocity x breaking break | .00 | .01 | .159 |
| Offspeed break x breaking break | -.00 | .01 | -.04 |
| Fastball velocity x offspeed break x breaking break | .01 | .02 | .56 |

*Note*. Parameter estimates for each of the predictors for model 1 and model 2. Note that both models are predicting log transformed estimates and giving standard errors of those estimates. Also, the predictors for the second model were mean-centered prior to analysis.

*Model comparison*

To compare the two models, I used a few methods. First, I looked at the models with the log-likelihood F-test (or in this instance with the mixed models, a Chi-squared test), and found there were no differences reported between the models (*X*2 = 2.21, *p* = .698). The likelihood ratio for the two models was .95, indicating a slight favor toward the second model over the first. Lastly, I conducted an analysis of the Akaike weights and found an overwhelming favor for the first model, giving it a 100% probability of generating the data when compared to model 2 (weight = 1.1 e -9). AIC has been my favorite method to compare models given all the different ways these values can be manipulated to ease the interpretation of model comparison like the Akaike weight. According to AIC and Akaike weights, I prefer the simpler first model with no interactions.

*Three new models*

To wrap up, I will conduct three new multi-level models and compare them with Akaike weights. The first model included regular season walk percentage of a pitcher along with expected weighted on base average (xwoba) to predict postseason log era. The second model included the same walk percentage along with strikeout percentage. The last model included weighted on base percentage (woba), fastball average velocity, and fastball average induced vertical break (higher spins make fastballs fall less, resulting in “rising” fastball phenomenon that is hard to hit). The first model was the preferred model according to Akaike weights, given a 98% chance to have generated the data compared to .3% for the second model and 1.2% chance for the 3rd model.