Lesson_20_Boardsheet

Kevin Cummiskey

3/5/2020

Review

Last lesson, we discussed the difference between *fixed effects* and *random effects* in modeling catcher framing ability.

Briefly explain the difference between a fixed effect and a random effect.

If you have 50 catchers, how many parameters are there in a fixed effects model?

If you have 50 catchers, how many parameteris are there in a random effects model?

If you wanted to know if Sandy Leon of the Boston Red Sox had a higher called strike percentage than other catchers, should you use a fixed or random effect?

If you wanted to know if the variability in catcher called strike percentage is "interesting", should you use a fixed or random effect?

Models with multiple random effects

Last class, we discussed a model for called strike percentage with a catcher random effect that adjusts for pitch location.

$$Strike_i \sim \text{Bernoulli}(p_i)$$

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + f(px, pz) + catcher_j$$

$$catcher_j \sim \text{Normal}(0, \sigma^2)$$

Read in the data

Follow the steps in Lesson 19 to read in Statcast data.

Fit the model adjusting for pitch location

```
library(mgcv)
library(broom)
library(lme4)
```

1

```r
library(ggrepel)
#fit pitch location model (plate_x = px, plate_z = pz)
model.location <- gam(type == "S" ~ s(plate_x, plate_z),
                      family = "binomial",
                      data = taken)
#get predictions from location model
taken %>%
  mutate(strike_prob = predict(model.location, newdata = .,
                               type = "response")) -> taken


#fit random effects model adjusting for pitch location
model.random.adj <- glmer(type == "S" ~ strike_prob + (1|catcher),
                          family = "binomial",
                          data = taken)
# random effects
model.random.adj %>% tidy(effects = "ran_pars")
```

```
## # A tibble: 1 x 3
##   term                group   estimate
##   <chr>               <chr>      <dbl>
## 1 sd_(Intercept).catcher catcher   0.203
```

```r
# fixed effects
model.random.adj %>% tidy(effects = "fixed")
```

```
## # A tibble: 2 x 5
##   term        estimate std.error statistic p.value
##   <chr>          <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)    -2.92    0.0411     -70.9       0
## 2 strike_prob     6.12    0.0594     103.        0
```

```r
#get random effects
model.random.adj %>%
  ranef() %>%
  as_tibble() %>%
  transmute(id = levels(grp),
            effect = condval) %>%
  arrange(desc(effect)) -> catcher_effects.adj

catcher_effects.adj %>% head(4)
```

```
## # A tibble: 4 x 2
##   id            effect
##   <chr>          <dbl>
## 1 Chris Stewart  0.339
## 2 Mike Zunino    0.285
## 3 Jason Castro   0.257
## 4 Jeff Mathis    0.249
```

```r
catcher_effects.adj %>% tail(4)
```
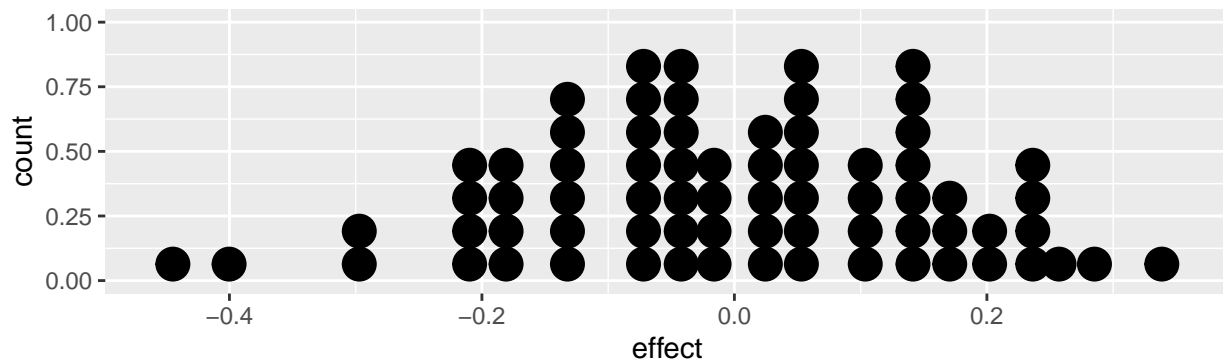
```
## # A tibble: 4 x 2
##   id            effect
##   <chr>          <dbl>
## 1 Drew Butera   -0.285
## 2 Andrew Knapp  -0.309
```

```
## 3 Tyler Flowers    -0.400
## 4 Tuffy Gosewisch -0.445
```

```r
catcher_effects.adj %>%
  ggplot(aes(x = effect, label = id)) +
  geom_dotplot()
```


Models with multiple random effects

What other variables should we adjust for?

Write a model adjusting pitch location (fixed effect) and for pitcher using a random effect.

```r
#fit random effects model adjusting for pitch location and pitcher
model.random.adj2 <- glmer(type == "S" ~ strike_prob + (1|catcher) + (1|pitcher),
                           family = "binomial",
                           data = taken)
# random effects
model.random.adj2 %>% tidy(effects = "ran_pars")
```

```
## # A tibble: 2 x 3
##   term                   group    estimate
##   <chr>                  <chr>       <dbl>
## 1 sd_(Intercept).pitcher pitcher     0.220
## 2 sd_(Intercept).catcher catcher     0.175
```

```r
# fixed effects
model.random.adj2 %>% tidy(effects = "fixed")
```

```
## # A tibble: 2 x 5
##   term         estimate std.error statistic p.value
##   <chr>           <dbl>     <dbl>     <dbl>   <dbl>
## 1 (Intercept)     -2.94    0.0415     -70.8       0
## 2 strike_prob      6.17    0.0608     102.        0
```

```r
#get random effects
model.random.adj2 %>%
  ranef() %>%
  as_tibble() %>%
  filter(grpvar == "catcher") %>%
  transmute(id = grp,
```

```
        effect = condval) %>%
  arrange(desc(effect)) -> catcher_effects.adj2

catcher_effects.adj2 %>% head(4)
```

```
## # A tibble: 4 x 2
##   id             effect
##   <fct>           <dbl>
## 1 Austin Barnes   0.201
## 2 Jeff Mathis     0.197
## 3 Yasmani Grandal 0.193
## 4 Evan Gattis     0.192
```

```
catcher_effects.adj2 %>% tail(4)
```

```
## # A tibble: 4 x 2
##   id             effect
##   <fct>           <dbl>
## 1 Mike Zunino    -0.222
## 2 Andrew Knapp   -0.250
## 3 Ryan Hanigan   -0.321
## 4 Devin Mesoraco -0.384
```

```
catcher_effects.adj2 %>%
  ggplot(aes(x = effect, label = id)) +
  geom_dotplot()
```

```
## `stat_bindot()` using `bins = 30`. Pick better value with `binwidth`.
```