Statistical Inference - Homework #6

Noah Johnson

April 3, 2018

[6.9.1] IVF and Birth Defects

Data Source: CDC

In November, the Centers for Disease Control and Prevention published a paper reporting that babies conceived with IVF, or with a technique in which sperm are injected directly into eggs, have a slightly increased risk of several birth defects, including a hole between the two chambers of the heart, a cleft lip or palate, an improperly developed esophagus and a malformed rectum. The study involved 9,584 babies with birth defects and 4,792 babies without. Among the mothers of babies without birth defects, 1.1 percent had used IVF or related methods, compared with 2.4 percent of mothers of babies with birth defects.

The findings are considered preliminary, and researchers say they believe IVF does not carry excessive risks. There is a 3 percent chance that any given baby will have a birth defect.

Construct a table and calculate the corresponding odds and odds ratios. Comment on the reported and calculated results.

So $P(defect) = 0.03$. We can calculate the odds of having a mother who used IVF given whether or not the baby had a birth defect. This is the opposite conditional of what we would usually expect (defect given IVF).

```
n_d <- 9584  # number of babies with birth defect
n_nd <- 4792  # number of babies with no birth defect

p_d <- 0.024  # probability of IVF given defect
p_nd <- 0.011  # probability of IVF given no defect

Y_d <- round(n_d * p_d) # Number of IVF mothers given defect
N_d <- n_d - Y_d # Number of non-IVF mothers given defect

Y_nd <- round(n_nd * p_nd) # Number of IVF mothers given no defect
N_nd <- n_nd - Y_nd # Number of non-IVF mothers given no defect
```

	IVF	not-IVF
birth defect	230	9354
no birth defect	53	4739

```
odds.y <- Y_d / (N_d)  # odds of IVF given defect
odds.n <- Y_nd / (N_nd)  # odds of IVF given no defect

print(odds.y)
```

```
## [1] 0.02458841
```

```
print(odds.n)
```

```
## [1] 0.01118379
```

```
odds.ratio <- odds.y / odds.n
print(odds.ratio)
```

```
## [1] 2.198575
```

The calculated odds are slightly higher than the reported proportions of IVF mothers across both factor levels(defect or no defect). The odds of a baby with a birth defect having a mother who used IVF is 2.2 times that of a baby without a birth defect.

[6.9.2.1] Soccer goals on target

Data from an article in Psychological Science, July 2011. [@Masters2007]

The example in the text referred to all shots, good or bad. This data relates to shots on target, 18 out of 20 shots were scored when the goalkeeper's team was behind, 71 out of 90 shots were scored when the game was tied, and when the team was not behind 55 out of 75 shots were scored. Calculate the odds of scoring for games behind, games tied, and games not behind. Construct empirical odds ratio for scoring for behind versus tied and tied versus not behind.

```
odds.behind <- 18 / (20 - 18)
odds.tied <- 71 / (90 - 71)
odds.ahead <- 55 / (75 - 55)

print(odds.behind)
```

```
## [1] 9
```

```
print(odds.tied)
```

```
## [1] 3.736842
```

```
print(odds.ahead)
```

```
## [1] 2.75
```

```
or.behindvstied <- odds.behind / odds.tied
or.tiedvsahead <- odds.tied / odds.ahead

print(or.behindvstied)
```

```
## [1] 2.408451
```

```
print(or.tiedvsahead)
```

```
## [1] 1.358852
```

- Fit a model with the categorical predictor (behind,tied,not behind) and interpret the exponeniated coefficients. How do they compare to the empirical odds ratios you calculated?

```
soccer <- tibble(game_state = factor(c("behind", "tied", "ahead")),
                 scores = c(18, 71, 55), shots = c(20, 90, 75))
soccer <- soccer %>% mutate(prop = scores / shots)
print(soccer)
```

```
## # A tibble: 3 x 4
##   game_state scores shots      prop
##       <fctr>  <dbl> <dbl>     <dbl>
## 1     behind     18    20 0.9000000
## 2       tied     71    90 0.7888889
```

```
## 3      ahead     55    75 0.7333333
```

```r
model <- glm(formula = prop ~ game_state, family = binomial(link = "logit"),
     data = soccer, weights = shots)
summary(model)
```

```
##
## Call:
## glm(formula = prop ~ game_state, family = binomial(link = "logit"),
##     data = soccer, weights = shots)
##
## Deviance Residuals:
## [1]  0  0  0
##
## Coefficients:
##                   Estimate Std. Error z value Pr(>|z|)
## (Intercept)         1.0116     0.2611   3.874 0.000107 ***
## game_statebehind    1.1856     0.7898   1.501 0.133297
## game_statetied      0.3066     0.3673   0.835 0.403783
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 2.9454e+00  on 2  degrees of freedom
## Residual deviance: 1.0880e-14  on 0  degrees of freedom
## AIC: 17.596
##
## Number of Fisher Scoring iterations: 4
```

```r
coef.exp <- exp(model$coefficients)
print(coef.exp)
```

```
##      (Intercept) game_statebehind   game_statetied
##         2.750000         3.272727         1.358852
```

3.27 is the odds ratio for scoring when behind versus ahead, while 1.36 is the odds ratio for scoring when tied versus ahead. 1.36 matches the empirical odds ratio that we calculated. The other empirical odds ratio we calculated, for scoring when behind vs tied, can be recomputed by taking the ratio of game_statebehind to game_statetied. Then the odds for scoring when ahead will cancel out, leaving the odds ratio for scoring when behind vs tied.

```r
coef.exp[[2]] / coef.exp[[3]]
```

```
## [1] 2.408451
```

```r
or.behindvstied
```

```
## [1] 2.408451
```

Success!