Name			
Email			

Reading Quiz for Sheather Chapter 8

[Based on Sheather, Ch 8, #1]

• •

Chapter 6 of Bradbury (2007), a book on baseball, uses regression analysis to compare the success of the 30 Major League Baseball teams. For example, the author considers the relationship between x_i , market size (i.e., the population in millions of the city associated with each team) and Y_i , the number of times team *i* made the post-season playoffs in the $m_i = 10$ seasons between 1995 and 2004. The author found that "it is hard to find much correlation between market size (population of the host city) and... success in making the playoffs. The relationship... is quite weak." Bradbury's analysis was similar to this:

```
> with (baseballdata, plot (PlayoffAppearances ~ Population))
> summary(lm.1 <- lm(PlayoffAppearances ~ Population,
+ data=baseballdata))
. . .
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)
              1.7547
                                    2.319
                                            0.0279 *
                          0.7566
Population
               0.1684
                          0.1083
                                    1.555
                                            0.1311
. . .
Multiple R-squared:
                      0.07952,
                                 Adjusted R-squared:
                                                        0.04664
> abline(lm.1)
```



Another researcher did an analysis like that on the back of this page.

```
> ProbOfPlayoffAppearance <- baseballdata$PlayoffAppearances/10
> with(baseballdata,plot(ProbOfPlayoffAppearance Population))
> summary(glm.1 <- glm(cbind(PlayoffAppearances, 10-PlayoffAppearances)</pre>
+ ~ Population, data=baseballdata, family=binomial))
. . .
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
                        0.21102 -6.911
                                          4.8e-12 ***
(Intercept) -1.45843
Population
             0.07807
                         0.02751
                                   2.838 0.00455 **
. . .
> curve(invlogit(cbind(1,x)%*%coef(glm.1)), add=T)
```



1. Which analysis (the one on the front of the quiz, or on the back) is likely to produce a model whose assumptions are better satisfied (and therefore has more believable results)?

2. Why?

- 3. Can we use AIC and BIC to compare the models from the two analyses?
- 4. Why or why not?