R Data Analysis Examples: Logit Regression

Idre at UCLA: [http://www.ats.ucla.edu/stat/r/dae/logit.htm](http://www.ats.ucla.edu/stat/r/dae/logit.htm)

```r
1. mydata <- read.csv("http://www.ats.ucla.edu/stat/data/binary.csv")
2. head(mydata)
3. mydata$rank <- factor(mydata$rank)
4. mylogit <- glm(admit ~ gre + gpa + rank, data = mydata, family = "binomial")
5. summary(mylogit)
6. newdata1 <- with(mydata, data.frame(gre = mean(gre), gpa = mean(gpa), rank = factor(1:4)))
7. newdata1
8. newdata1.rank <- predict(mylogit, newdata = newdata1, type = "response")
9. newdata1
10. newdata2 <- with(mydata, data.frame(gre = rep(seq(from = 200, to = 800, length.out = 100), 4),
11. gpa = mean(gpa), rank = factor(rep(1:4, each = 100))))
12. newdata3 <- cbind(newdata2, predict(mylogit, newdata = newdata2, type = "link",
13. se = TRUE))
14. newdata3 <- within(newdata3, {
15. PredictedProb <- plogis(fit - (1.96 * se.fit))
16. UL <- plogis(fit + (1.96 * se.fit))
17. })
18. head(newdata3)
19. tail(newdata3)
```

Line #19 is the key. (Lines 19-20). The 2\textsuperscript{nd} argument to the with() function is the data.frame() function. Note that GRE is repeated 4 times (FOUR times). What will occur FOUR times? A sequence of 100 values going from 200 to 800. So the 400 observations in mydata are 100 rank #1 school students, 100 rank #2 school students, 100 rank #3 school students, and finally 100 rank #4 students. So we will have a range of 100 GRE scores going from 200 to 800 for each and every rank category. That is why it is repeated 4 times: rep(seq(...), 4).

What will the GPA be? We will give all 400 observations (rows, cases, graduate school applicants from our sample) the MEAN or AVERAGE GPA. The average GPA was 3.39 for the sample. Line #20 shows the argument gpa = mean(gpa).

So if we want to see the statistics and the graphics for GPA as to its effect on being admitted to graduate school by rank of the undergraduate institution the applicant is getting a degree from, what should we change in lines #19 and #20?

GRE should be the MEAN(GRE), right?

GPA should get the rep() function with the nested seq( ...) sequence, as shown above, right?

GPA has a minimum of 2.26 and a maximum of 4.00, right?

So you could substitute 2.2 or 2.0 for 200 in the from = 200 above.

And for sure, you just substitute 4.0 for the 800 above and to = 800 becomes to = 4.0 or to = 4.00, right?