

# Logistic Regression in UCLA Data

Hunter Grigsby

## Data Preperation

Resources: “<https://stats.idre.ucla.edu/stat/data/binary.csv>”

```
# packages required
library(aod)
library(ggplot2)

path <- "https://stats.idre.ucla.edu/stat/data/binary.csv"
mydata <- read.csv(path)
```

- The dataset contains information on applicants, specifically whether they were admitted (admit), their GRE scores (gre), GPAs (gpa), and the rank of their undergraduate institution (rank).

## Data Checkup

```
str(mydata)

## 'data.frame':    400 obs. of  4 variables:
## $ admit: int  0 1 1 1 0 1 1 0 1 0 ...
## $ gre : int  380 660 800 640 520 760 560 400 540 700 ...
## $ gpa : num  3.61 3.67 4 3.19 2.93 3 2.98 3.08 3.39 3.92 ...
## $ rank : int  3 3 1 4 4 2 1 2 3 2 ...

mydata$rank = as.factor(mydata$rank)
str(mydata)

## 'data.frame':    400 obs. of  4 variables:
## $ admit: int  0 1 1 1 0 1 1 0 1 0 ...
## $ gre : int  380 660 800 640 520 760 560 400 540 700 ...
## $ gpa : num  3.61 3.67 4 3.19 2.93 3 2.98 3.08 3.39 3.92 ...
## $ rank : Factor w/ 4 levels "1","2","3","4": 3 3 1 4 4 2 1 2 3 2 ...

summary(mydata)

##      admit      gre      gpa      rank
## Min.   :0.0000  Min.   :220.0  Min.   :2.260  1: 61
## 1st Qu.:0.0000  1st Qu.:520.0  1st Qu.:3.130  2:151
## Median :0.0000  Median :580.0  Median :3.395  3:121
## Mean   :0.3175  Mean   :587.7  Mean   :3.390  4: 67
## 3rd Qu.:1.0000  3rd Qu.:660.0  3rd Qu.:3.670
## Max.   :1.0000  Max.   :800.0  Max.   :4.000
```

- The summary(mydata) output provides key descriptive statistics for each variable in the dataset, such as minimum, maximum, median, mean, and the quartiles, which help in understanding the distribution and central tendencies of GRE scores, GPAs, and undergraduate institution ranks. Additionally, the

conversion of the rank variable to a factor type in R, as shown in the `str(mydata)` output, facilitates categorical analysis, ensuring that statistical methods appropriate for nominal data are applied in subsequent analyses.

## Regression = Estimation

```
mylogit = glm(admit ~ gre + gpa + rank, data = mydata, family = "binomial")
summary(mylogit)
```

```
##
## Call:
## glm(formula = admit ~ gre + gpa + rank, family = "binomial",
##      data = mydata)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.989979   1.139951  -3.500 0.000465 ***
## gre          0.002264   0.001094   2.070 0.038465 *
## gpa          0.804038   0.331819   2.423 0.015388 *
## rank2       -0.675443   0.316490  -2.134 0.032829 *
## rank3       -1.340204   0.345306  -3.881 0.000104 ***
## rank4       -1.551464   0.417832  -3.713 0.000205 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##    Null deviance: 499.98  on 399  degrees of freedom
## Residual deviance: 458.52  on 394  degrees of freedom
## AIC: 470.52
##
## Number of Fisher Scoring iterations: 4
```

- The output from the logistic regression model shows that GRE scores ( $p = 0.038465$ ) and GPA ( $p = 0.015388$ ) are statistically significant predictors of admission, indicating that higher values in these variables increase the likelihood of admission. The negative coefficients for rank2, rank3, and rank4 (all  $p < 0.05$ ) suggest that applicants from lower-ranked institutions have a reduced chance of admission compared to those from top-ranked institutions, controlling for other factors.

## Calculate Odds Ratios

```
coefs = coefficients(mylogit)
ORs = exp(coefs)
CIs = exp(confint(mylogit))
```

```
## Waiting for profiling to be done...
```

```
CIs
```

```
##              2.5 %    97.5 %
## (Intercept) 0.001889165 0.1665354
## gre         1.000137602 1.0044457
## gpa         1.173858216 4.3238349
## rank2       0.272289674 0.9448343
```

```
## rank3      0.131641717 0.5115181
## rank4      0.090715546 0.4706961
```

```
cbind(round(ORs,2), round(CIs,2))
```

```
##           2.5 % 97.5 %
## (Intercept) 0.02 0.00 0.17
## gre         1.00 1.00 1.00
## gpa         2.23 1.17 4.32
## rank2       0.51 0.27 0.94
## rank3       0.26 0.13 0.51
## rank4       0.21 0.09 0.47
```

- GRE Scores (gre): The odds ratio of 1.00 indicates negligible impact on admission odds per unit increase in GRE scores.
- GPA (gpa): With an odds ratio of 2.23, each unit increase in GPA significantly boosts admission odds, suggesting GPA is a strong predictor.
- Rank 2 (rank2): Applicants from institutions ranked '2' have roughly half the admission odds compared to top-ranked schools.
- Rank 3 (rank3) and Rank 4 (rank4): These ranks significantly lower the admission odds to 0.26 and 0.21, respectively, indicating much lower chances of admission for applicants from these institutions.

