

PART -2 Regression Assignment

Gowlapalli Rohit

10/4/2024

Contents

Significant Variables Predicting Admission	1
Inferences based on logistic model	2
Confidence Intervals	3
Wald's test	3
Odds by exponentiating coefficients	4
Predicted Probability of Admission	7
Testing Interaction effects	9
Inferences based on logistic model	10
Confidence Intervals	11
Odds by exponentiating coefficients	11
Inferences on Interaction Effects	12

Significant Variables Predicting Admission

```
library(aod)
library(ggplot2)
# Binary
mydata<-read.csv("binary.csv")
mydata$rank <- factor(mydata$rank)
fit <- glm(admit ~ gre + gpa + rank, data = mydata, family = "binomial")
```

```
summary(fit) # display results
```

```
##
## 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
```

Inferences based on logistic model

- Significant predictors ($p < 0.05$):
 - gre: $p=0.038$
 - gpa: $p=0.015$ (significant)
 - rank_2: $p=0.033$
 - rank_3: $p=0.0001$
 - rank_4: $p=0.0002$
 - Intercept: $p=0.0005$

Both GRE and GPA are statistically significant, as are the three terms for rank. The logistic regression coefficients give the change in the log odds of the outcome for a one unit increase in the predictor variable.

- For every one unit change in GRE, the log odds of admission (versus non-admission) increases by 0.002.
- For a one unit increase in GPA, the log odds of being admitted to graduate school increases by 0.804.

The indicator variables for rank have a slightly different interpretation. For example, having attended an undergraduate institution with rank of 2, versus an institution with a rank of 1, changes the log odds of admission by -0.675.

Although there is a statistically significant effect for GRE score, its impact on admission odds is relatively smaller compared to GPA and undergraduate institution rank.

GPA: The most significant odds ratios are associated with GPA. A one-unit increase in GPA is strongly linked to higher odds of admission, indicating a substantial impact on admission chances.

Rank: The ranking of the undergraduate institution also significantly influences admission odds. Candidates from institutions ranked 2, 3, or 4 have notably lower odds of admission compared to those from rank 1 institutions.

GRE: Although statistically significant, the effect of GRE score on admission odds is comparatively smaller than that of GPA and rank.

Conclusion: GPA emerges as the most critical predictor of admission in the model. With each one-unit increase in GPA, the odds of admission substantially rise, holding all other variables constant. Additionally, the rank of the applicant's undergraduate institution significantly impacts admission odds, with applicants

from institutions ranked 2, 3, or 4 having significantly lower odds of admission compared to those from rank 1 institutions. Despite its statistical significance, the impact of GRE score on admission odds is relatively smaller when compared to GPA and undergraduate institution rank.

Confidence Intervals

```
## CIs using profiled log-likelihood
confint(fit) # 95% CI for the coefficients
```

```
##              2.5 %      97.5 %
## (Intercept) -6.2716202334 -1.792547080
## gre         0.0001375921  0.004435874
## gpa         0.1602959439  1.464142727
## rank2       -1.3008888002 -0.056745722
## rank3       -2.0276713127 -0.670372346
## rank4       -2.4000265384 -0.753542605
```

Based on the 95% confidence intervals: - There is a statistically significant effect for GRE score, as the interval for GRE does not include zero. - GPA also exhibits a significant effect, as its confidence interval excludes zero. - For the rank variables, all three (rank2, rank3, and rank4) have intervals that do not encompass zero, indicating their significance.

```
## CIs using standard errors
confint.default(fit)
```

```
##              2.5 %      97.5 %
## (Intercept) -6.2242418514 -1.755716295
## gre         0.0001202298  0.004408622
## gpa         0.1536836760  1.454391423
## rank2       -1.2957512650 -0.055134591
## rank3       -2.0169920597 -0.663415773
## rank4       -2.3703986294 -0.732528724
```

We can use the `confint` function to obtain confidence intervals for the coefficient estimates. Note that for logistic models, confidence intervals are based on the profiled log-likelihood function. We can also get CIs based on just the standard errors by using the default method.

To assess the overall effect of rank, we utilize the `wald.test` function from the `aod` library. It's crucial to maintain the order of coefficients as they correspond to the terms in the model. This alignment is significant because the `wald.test` function identifies coefficients based on their model order. When using `wald.test`, the coefficients are provided as `b`, the variance-covariance matrix of error terms is supplied as `Sigma`, and `Terms` specifies the terms to be tested. In this instance, terms 4, 5, and 6 represent the three levels of rank.

Wald's test

```
wald.test(b = coef(fit), Sigma = vcov(fit), Terms = 4:6)
```

```
## Wald test:
## -----
##
## Chi-squared test:
## X2 = 20.9, df = 3, P(> X2) = 0.00011
```

The chi-squared test statistic of 20.9, with three degrees of freedom, yields a p-value of 0.00011, indicating a statistically significant overall effect of rank.

We can further investigate differences in coefficients for different levels of rank. For instance, we test whether the coefficient for rank=2 is equal to the coefficient for rank=3. The first line of code creates a vector 'l' defining the test, specifying the difference between the terms for rank=2 and rank=3 (the 4th and 5th terms in the model). By multiplying one term by 1 and the other by -1, we contrast these two terms. Other terms in the model, not involved in the test, are multiplied by 0.

```
l <- cbind(0, 0, 0, 1, -1, 0)
wald.test(b = coef(fit), Sigma = vcov(fit), L = l)
```

```
## Wald test:
## -----
##
## Chi-squared test:
## X2 = 5.5, df = 1, P(> X2) = 0.019
```

With a chi-squared test statistic of 5.5 and 1 degree of freedom, the associated p-value is 0.019. This suggests that the discrepancy between the coefficient for rank=2 and the coefficient for rank=3 is statistically significant.

Odds by exponentiating coefficients

```
exp(coef(fit)) # exponentiated coefficients
```

```
## (Intercept)      gre      gpa      rank2      rank3      rank4
##  0.0185001    1.0022670  2.2345448  0.5089310  0.2617923  0.2119375
```

Exponentiated Coefficients (Odds Ratios): - Intercept: 0.018 (Baseline odds of admission) - gre: 1.002 (A 1-unit increase in GRE increases odds by 0.2%) - gpa: 2.235 (A 1-unit increase in GPA increases odds by 123.5%) - rank2: 0.509 (Being in Rank 2 decreases odds by 49.1% compared to Rank 1) - rank3: 0.262 (Being in Rank 3 decreases odds by 73.8% compared to Rank 1) - rank4: 0.212 (Being in Rank 4 decreases odds by 78.8% compared to Rank 1)

Based on the exponentiated coefficients: - GPA has the highest impact on admission odds, with each one-unit increase associated with approximately a 2.23-fold increase in the odds of admission. - GRE has a minimal impact, with each one-unit increase in GRE score resulting in only a slight increase in admission odds. - Students from institutions ranked 2, 3, or 4 have significantly lower odds of admission compared to those from rank 1 institutions, as indicated by the odds ratios less than 1.

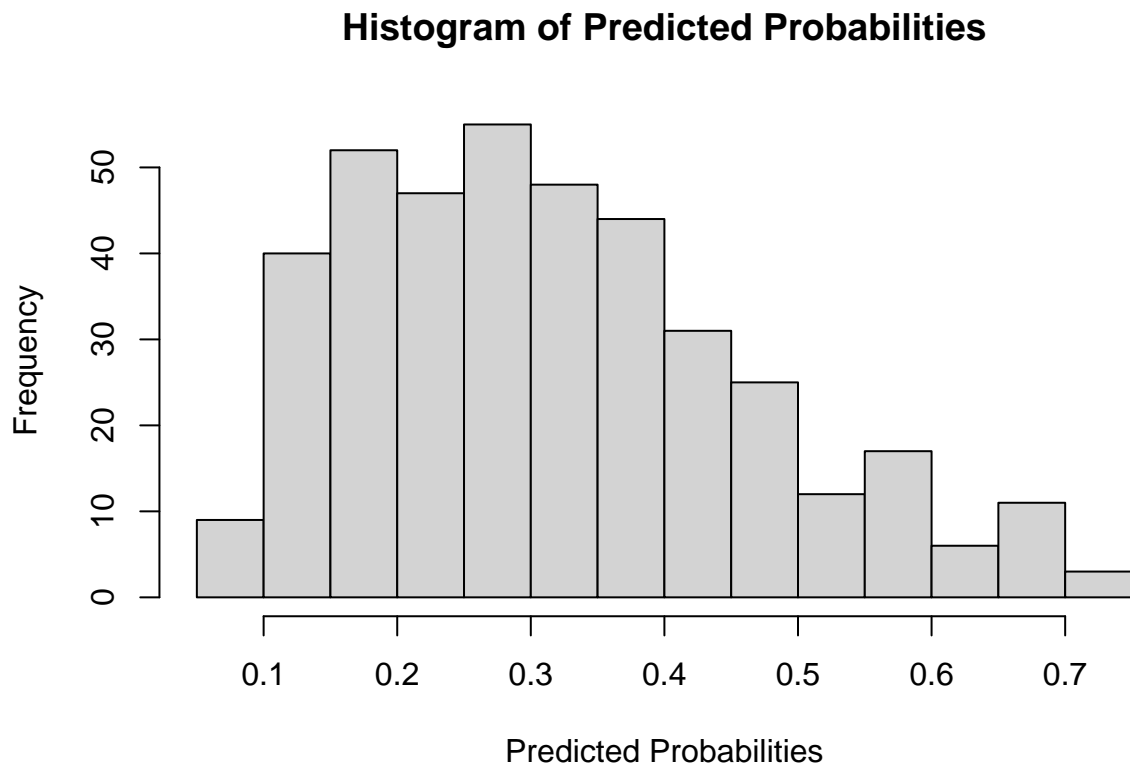
```
exp(confint(fit)) # 95% CI for exponentiated coefficients
```

```
##           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
```

Now we can say that for a one unit increase in GPA, the odds of being admitted to graduate school (versus not being admitted) increase by a factor of 2.23.

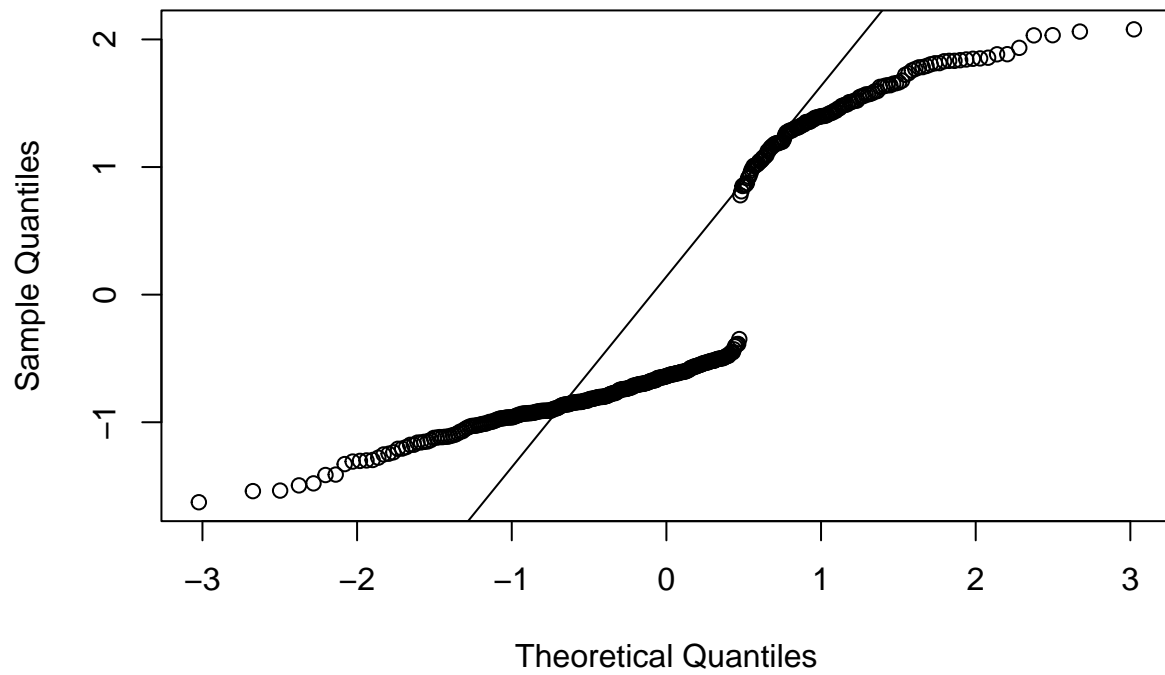
```
predictions <- predict(fit, type="response") # predicted values
residuals <- residuals(fit, type="deviance") # residuals
```

```
hist(predictions, breaks=20, main="Histogram of Predicted Probabilities", xlab="Predicted Probabilities")
```



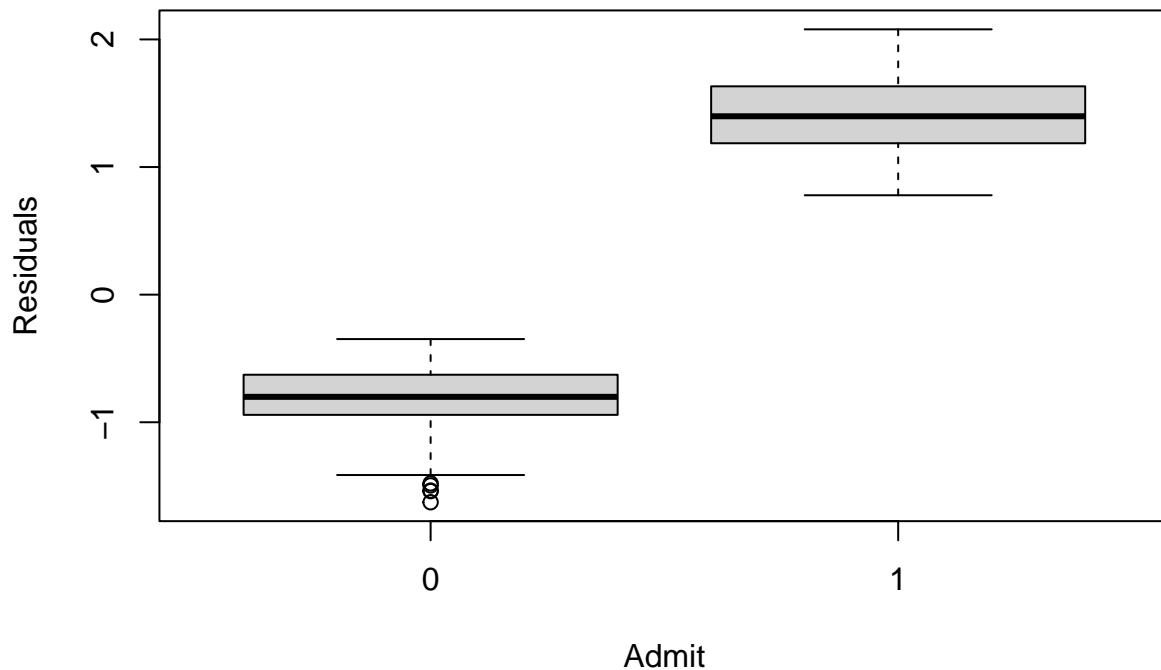
```
qqnorm(residuals)
qqline(residuals)
```

Normal Q-Q Plot



```
boxplot(residuals ~ mydata$admit, xlab="Admit", ylab="Residuals", main="Residuals by Admit")
```

Residuals by Admit



```
with(fit, null.deviance - deviance)
```

```
## [1] 41.45903
```

```
with(fit, df.null - df.residual)
```

```
## [1] 5
```

```
with(fit, pchisq(null.deviance - deviance, df.null - df.residual, lower.tail = FALSE))
```

```
## [1] 7.578194e-08
```

```
logLik(fit)
```

```
## 'log Lik.' -229.2587 (df=6)
```

With a chi-square value of 41.46 and 5 degrees of freedom, along with a p-value of less than 0.001, we can conclude that our model significantly outperforms an empty model. We calculate the predicted probability of admission at each value of rank, holding gre and gpa at their means

Predicted Probability of Admission

```
newdata1 <- with(mydata, data.frame(gre = mean(gre), gpa = mean(gpa), rank = factor(1:4)))
```

```
newdata1$rankP <- predict(fit, newdata = newdata1, type = "response")
newdata1
```

```
##      gre      gpa rank      rankP
## 1 587.7 3.3899     1 0.5166016
## 2 587.7 3.3899     2 0.3522846
## 3 587.7 3.3899     3 0.2186120
## 4 587.7 3.3899     4 0.1846684
```

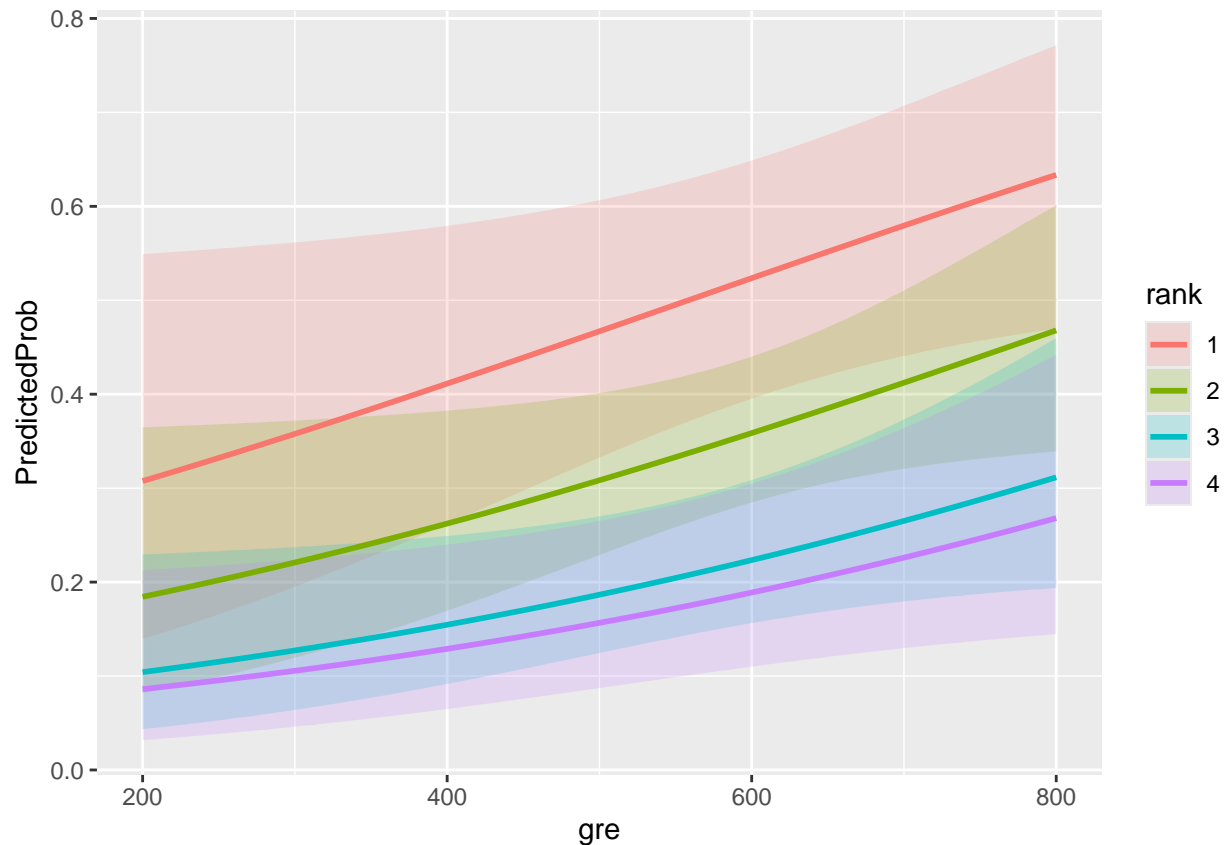
From the above results, we observe that the predicted probability of acceptance into a graduate program is 0.52 for students from the highest-ranked undergraduate institutions (rank=1), and 0.18 for students from the lowest-ranked institutions (rank=4), while holding GRE and GPA at their means. To create a table of predicted probabilities varying GRE and rank, we can follow a similar approach. We'll generate 100 values of GRE between 200 and 800 for each rank value (1, 2, 3, and 4) to plot the results.

```
newdata2 <- with(mydata, data.frame(gre = rep(seq(from = 200, to = 800, length.out = 100),
4), gpa = mean(gpa), rank = factor(rep(1:4, each = 100))))
newdata3 <- cbind(newdata2, predict(fit, newdata = newdata2, type = "link",
se = TRUE))
newdata3 <- within(newdata3, {
  PredictedProb <- plogis(fit)
  LL <- plogis(fit - (1.96 * se.fit))
  UL <- plogis(fit + (1.96 * se.fit))
})

## view first few rows of final dataset
head(newdata3)
```

```
##      gre      gpa rank      fit      se.fit residual.scale      UL      LL
## 1 200.0000 3.3899     1 -0.8114870 0.5147714             1 0.5492064 0.1393812
## 2 206.0606 3.3899     1 -0.7977632 0.5090986             1 0.5498513 0.1423880
## 3 212.1212 3.3899     1 -0.7840394 0.5034491             1 0.5505074 0.1454429
## 4 218.1818 3.3899     1 -0.7703156 0.4978239             1 0.5511750 0.1485460
## 5 224.2424 3.3899     1 -0.7565919 0.4922237             1 0.5518545 0.1516973
## 6 230.3030 3.3899     1 -0.7428681 0.4866494             1 0.5525464 0.1548966
##      PredictedProb
## 1      0.3075737
## 2      0.3105042
## 3      0.3134499
## 4      0.3164108
## 5      0.3193867
## 6      0.3223773
```

```
ggplot(newdata3, aes(x = gre, y = PredictedProb)) + geom_ribbon(aes(ymin = LL,
ymax = UL, fill = rank), alpha = 0.2) + geom_line(aes(colour = rank),
size = 1)
```

Testing Interaction effects

```
interaction_fit <- glm(admit ~ gre + gpa + rank + gpa*rank, data = mydata, family = "binomial")
summary(interaction_fit) # display results
```

```
##
## Call:
## glm(formula = admit ~ gre + gpa + rank + gpa * rank, family = "binomial",
##      data = mydata)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.985768   2.480668  -2.010   0.0444 *
## gre          0.002287   0.001102   2.075   0.0380 *
## gpa          1.089088   0.726130   1.500   0.1337
## rank2        0.503294   2.961800   0.170   0.8651
## rank3        0.450796   3.243486   0.139   0.8895
## rank4       -1.508472   4.172185  -0.362   0.7177
## gpa:rank2    -0.342951   0.858440  -0.400   0.6895
## gpa:rank3    -0.515245   0.929281  -0.554   0.5793
## gpa:rank4    -0.009246   1.212095  -0.008   0.9939
## ---
## 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: 458.11 on 391 degrees of freedom
## AIC: 476.11
##
## Number of Fisher Scoring iterations: 4
```

Inferences based on logistic model

- Significant predictors ($p < 0.05$):
 - gre: $p=0.038$ (significant)
 - Intercept: $p=0.044$ (significant)
- gre:** The coefficient estimate for 'gre' is 0.002287 with a standard error of 0.001102. It has a z-value of 2.075 and a p-value of 0.0380, indicating statistical significance at the 0.05 level. This suggests that for every one-unit increase in GRE score, the log odds of admission increase by approximately 0.23%.
- gpa:** With a coefficient estimate of 1.089088 and a standard error of 0.726130, 'gpa' has a z-value of 1.500 and a p-value of 0.1337, which is not statistically significant at the 0.05 level. This implies that although GPA has a positive impact on admission odds, the effect is not statistically supported in this model.
- rank2, rank3, rank4:** These are the indicator variables representing different levels of the 'rank' feature. 'Rank2' to 'rank4' have coefficient estimates of 0.503294, 0.450796, and -1.508472, respectively. None of these coefficients are statistically significant based on their p-values, indicating no clear evidence of a significant effect of the rank levels on admission odds compared to the reference level.
- Interaction terms (gpa:rank2, gpa:rank3, gpa:rank4):** These interaction terms represent the combined effect of GPA and each rank level. None of these interaction terms are statistically significant, as indicated by their p-values (> 0.05). This suggests that the interaction between GPA and rank does not significantly influence admission odds beyond the individual effects of GPA and rank.

The logistic regression model included the GPA*rank interaction term to assess its impact on admission odds. However, this interaction was found to be statistically non-significant ($p > 0.05$), indicating that the relationship between GPA and admission odds remains consistent across different ranks of undergraduate institutions. Despite GPA's lack of statistical significance alone, it still holds considerable influence on admission odds, alongside gre, which remains significant. Therefore, while both GPA and rank independently affect admission odds, their interaction does not significantly contribute to predicting admission outcomes in this analysis.

```
## CIs using profiled log-likelihood
confint(interaction_fit) # 95% CI for the coefficients
```

```
##                2.5 %      97.5 %
## (Intercept) -1.013852e+01 -0.2855434
## gre          1.441291e-04  0.0044754
## gpa          -2.899299e-01  2.5947186
## rank2        -5.220220e+00  6.4899913
## rank3        -5.901029e+00  6.9159969
## rank4        -9.990756e+00  6.5227383
## gpa:rank2    -2.078618e+00  1.3141072
## gpa:rank3    -2.374097e+00  1.2960931
## gpa:rank4    -2.361837e+00  2.4314703
```

Confidence Intervals

- **Intercept:** The 95% confidence interval for the intercept ranges from -10.14 to -0.29, indicating a relatively wide range of uncertainty regarding the baseline log odds of admission.
- **gre:** The narrow confidence interval for 'gre' suggests relatively precise estimation, with values ranging from 0.000144 to 0.004475.
- **gpa:** Similarly, 'gpa' also has a narrow confidence interval, ranging from -0.290 to 2.595, implying precise estimation.
- **rank2, rank3, rank4:** The confidence intervals for the rank-related variables ('rank2' to 'rank4') are wider compared to 'gre' and 'gpa'. For example, the confidence interval for 'rank2' ranges from -5.22 to 6.49, indicating a higher level of uncertainty in the estimated coefficients.
- **Interaction terms (gpa:rank2, gpa:rank3, gpa:rank4):** These interaction terms also have wider confidence intervals compared to 'gre' and 'gpa', indicating increased uncertainty in their estimated effects on admission odds.

```
## CIs using standard errors
confint.default(interaction_fit)
```

```
##              2.5 %      97.5 %
## (Intercept) -9.8477868302 -0.123748236
## gre          0.0001267495  0.004447873
## gpa          -0.3341009991  2.512276953
## rank2        -5.3017285640  6.308315977
## rank3        -5.9063202983  6.807912189
## rank4        -9.6858046572  6.668860334
## gpa:rank2    -2.0254627529  1.339561306
## gpa:rank3    -2.3366021920  1.306112623
## gpa:rank4    -2.3849079114  2.366416000
```

Odds by exponentiating coefficients

```
exp(coef(interaction_fit)) # exponentiated coefficients
```

```
## (Intercept)      gre      gpa      rank2      rank3      rank4
## 0.00683453  1.00228993  2.97156270  1.65416063  1.56956097  0.22124775
##   gpa:rank2   gpa:rank3   gpa:rank4
## 0.70967318  0.59735435  0.99079666
```

Exponentiated Coefficients (Odds Ratios): - gre: 1.002 (A 1-unit increase in GRE score increases odds by 0.2%) - gpa: 2.972 (A 1-unit increase in GPA increases odds by 197.2%) - rank2: 1.654 (Being in Rank 2 increases odds by 65.4% compared to Rank 1) - rank3: 1.570 (Being in Rank 3 increases odds by 57% compared to Rank 1) - rank4: 0.221 (Being in Rank 4 decreases odds by 77.9% compared to Rank 1) - gpa:rank2: 0.710 (Interaction between GPA and Rank 2 decreases odds by 29%) - gpa:rank3: 0.597 (Interaction between GPA and Rank 3 decreases odds by 40.3%) - gpa:rank4: 0.991 (Interaction between GPA and Rank 4 decreases odds by 0.9%)

```
exp(confint(interaction_fit)) # 95% CI for exponentiated coefficients
```

```
##              2.5 %      97.5 %
## (Intercept) 3.952710e-05  0.7516057
```

## gre	1.000144e+00	1.0044854
## gpa	7.483160e-01	13.3928177
## rank2	5.406141e-03	658.5176210
## rank3	2.736627e-03	1008.2756690
## rank4	4.582155e-05	680.4391030
## gpa:rank2	1.251030e-01	3.7214272
## gpa:rank3	9.309848e-02	3.6549890
## gpa:rank4	9.424696e-02	11.3755954

Inferences on Interaction Effects

The coefficients gpa:rank2, gpa:rank3, and gpa:rank4 denote the interaction between GPA and each rank level, with rank 1 serving as the reference level. None of these interaction terms exhibit significant p-values (all p-values > 0.05). Therefore, there is insufficient evidence to assert that the impact of GPA on admission probability varies based on the rank of the undergraduate institution. The lack of statistical significance in the interaction terms implies that although GPA remains a significant predictor of admission, its effect does not seem to vary significantly depending on whether a student originates from a higher or lower-ranked institution.