Notes 9 - Logistic regression

Jillian Morrison

November 18, 2019

What will we cover

- Logistic Regression
  - We want to classify observation into groups.
- For example:
  - A person arrives to the emergency room with a set of symptoms. There are 3 possible diseases Which one does the person have?
  - An online banking system wants to determine if a transaction is fraudulent or not based on things like user IP address, past transaction history, etc.
  - The doctor wants to determine if a person has cancer or not based on the DNA sequence for a number of individuals who have cancer and who does not.

Logistic Regression

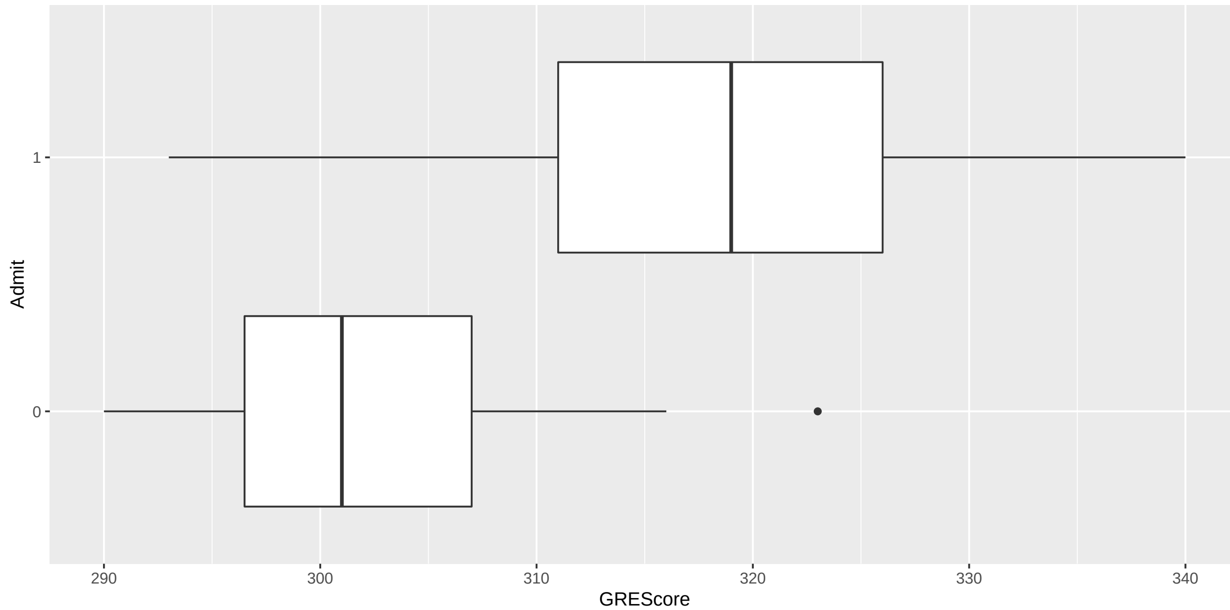
- Response variable - category (the thing you want to predict)
  - This is **UNLIKE** Linear Regression
  - We will focus on predicting 2 categories only.
- Predictor variable - category or number (the thing you are using to predict)
  - This is **LIKE** Linear Regression

Logistic Regression

Let's look at the `University_Kaggle_Grad dataset`. I added a new column for Admit (Admitted (1) if chance of Admit>0.5). We want to predict if a student will be admitted to Graduate School based on a bunch of factors.

Let's say we wanted to use GREScore to predict Admit

```
> library(readr)
> library(dplyr)
> Grad <- read_csv("University_Kaggle_Grad.csv")
> Grad<-Grad%>%mutate(Research=factor(Research), Admit=factor(Admit), UniversityRating=factor(Universit
> library(ggplot2)
> ggplot(Grad, aes(x=Admit, y=GREScore))+geom_boxplot()+coord_flip()
```

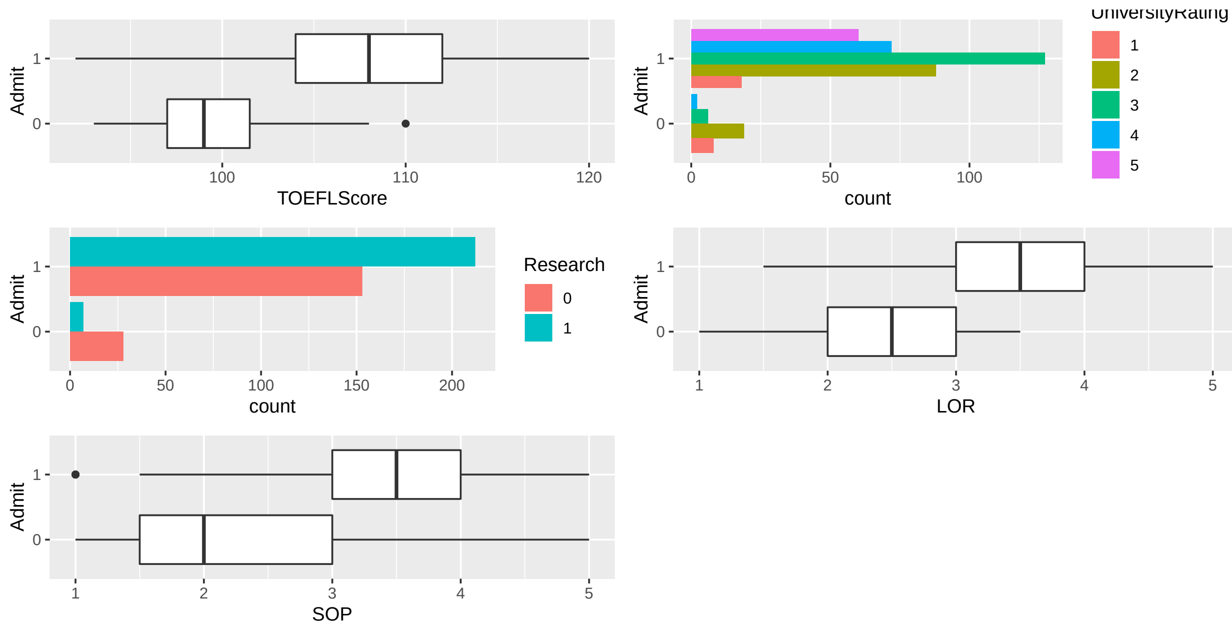


GRE May be a good predictor of Admit.

Logistic Regression

How about these?

```
> library(gridExtra)
> a=ggplot(Grad, aes(x=Admit, y=TOEFLScore))+geom_boxplot()+coord_flip()
> b=ggplot(Grad, aes(x=Admit, fill=UniversityRating))+geom_bar(position='dodge')+coord_flip()
> c=ggplot(Grad, aes(x=Admit, fill=Research))+geom_bar(position='dodge')+coord_flip()
> d=ggplot(Grad, aes(x=Admit, y=LOR))+geom_boxplot()+coord_flip()
> e=ggplot(Grad, aes(x=Admit, y=SOP))+geom_boxplot()+coord_flip()
> grid.arrange(a,b,c,d,e, nrow=3)
```



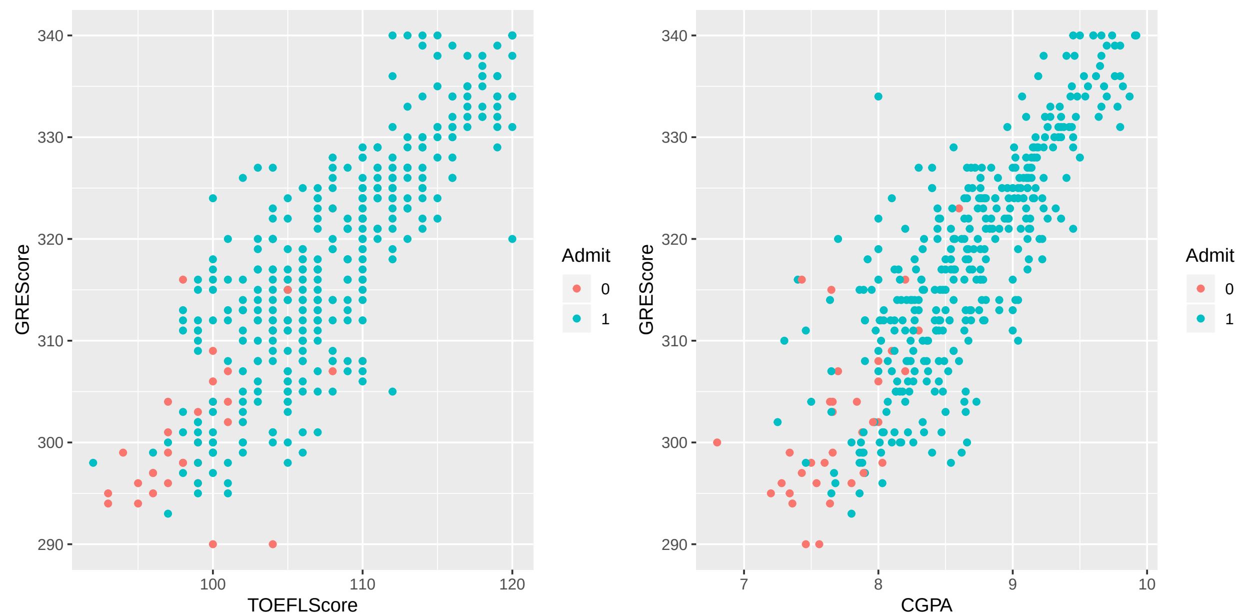
See:

- Looks also like Toefl Score might be a good predictor of Admit
- University rating might also be a good predictor, but not as clear as Tefl
- Research Experience definitely since pattern changes
- LOR as well

Logistic Regression

Let's look at some 2 dimensional plots:

```
> a=ggplot(Grad, aes(color=Admit, x= TOEFLScore, y=GREScore))+geom_point()
> c=ggplot(Grad, aes(x=CGPA, y=GREScore, color=Admit))+geom_point()
> grid.arrange(a,c, nrow=1)
```



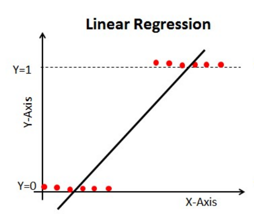
Looks like students don't get admitted if their GREScore, Toefl score, CGPA is lower.

Logistic Regression

Why can't we use linear regression?

Logistic regression?

- Why can't we use linear regression?
  - Well, we are trying to predict a category (or 0 versus 1).
  - A linear regression will predict a continuous numerical variable, so you will get things between 0 and 1, greater than 1, and less than 0 (Basically all possibilities as if it was continuous).
  - This is not what we want!



graph from: https://www.datacamp.com/community/tutorials/understanding-logistic-regression-python

Logistic Regression

In this case we want to predict a probability of being in a certain category.

So, we will want to predict $Pr(Admit = 0|ToeflScore, GREScore, etc.)$

This is read: The probability that a student is not admitted (Admit=0) given their ToeflScore, GreScore, etc..

We will also want to predict $Pr(Admit = 1|ToeflScore, GREScore, etc.)$

This is read: The probability that a student is admitted (Admit=1) given their ToeflScore, GreScore, etc..

Logistic Regression Model

If $Y = Admit$, then the model is:

$$log\Big(\frac{P(Y=1)}{P(Y=0)}\Big) = log\Big(\frac{P(Y=1)}{1-P(Y=1)}\Big) = \beta_0 + \beta_1 X$$

Logistic Regression Model

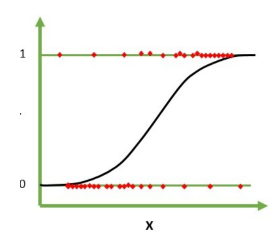
Let's talk about $log\Big(\frac{P(Y=1)}{P(Y=0)}\Big) = log\Big(\frac{P(Y=1)}{1-P(Y=1)}\Big) = log(ODDS)$

- This is called the log-odds.
- We will predict the log-odds, and then later translate it to probability
  - since the log odds is just a transformation of the probability.

Logistic Regression Model

- Why do we use this specific $log(ODDS)$ instead of just Y as in linear regression?
  - Well, if we look at $Pr(Y = 1)$ instead of considering the entire $log\Big(\frac{P(Y=1)}{1-P(Y=1)}\Big)$, this is

$$Pr(Y=1) = \frac{e^{\beta_o+\beta_1 X}}{1+e^{\beta_o+\beta_1 X}}$$



Where the steepness and specific shape depends on the values of β_0 and β_1 . Try it out at desmos.com with any value of β_0 and β_1.

We see that this will allow us to predict only probabilities between 0 and 1! Which is what we want!

Plot from: https://techdifferences.com/difference-between-linear-and-logistic-regression.html

What is Odds?

The Odds is the ratio of the probability of an event occurring to the event not occurring.

So, if the probability it will snow today is 0.05 (i.e. 5% chance of snow), then the odds is:

$$Odds = \frac{Pr(\text{rain})}{Pr(\text{no rain})} = \frac{Pr(\text{rain})}{1 - Pr(\text{rain})} = \frac{0.05}{1 - 0.05} = \frac{0.05}{0.95} = 0.05263$$

So the Odds FOR Rain today is 0.05263.

Odds Exercise

If the probability you will win a scratch lottery if you buy one ticket is one in 1,000. What is the odds FOR winning?

Odds Exercise Solution

If the probability you will win a scratch lottery if you buy one ticket is one in 1,000. What is the odds FOR winning?

$$Pr(\text{Win}) = \frac{1}{1000} = 0.001$$

$$Odds = \frac{Pr(\text{Win})}{Pr(\text{Lose})} = \frac{Pr(\text{Win})}{1 - Pr(\text{Win})} = \frac{0.001}{1 - 0.001} = \frac{0.001}{0.999} = 0.001001001$$

The odds FOR you winning is 0.001001001

Properties of Odds

- **If $Pr(Y) = 1 - Pr(Y)$, then ODDS = 1 .**
  - This happens if $Pr(Y) = 1 - Pr(Y) = 0.5$
  - This means that the probability of the event happenning and the event nor happenning is 0.5
  - OR there is a 50/50 chance of the event.

- **If $Pr(Y) > 1 - Pr(Y)$, then ODDS > 1 .**
  - This means that the probabilty of the event happenning is larger than the event not happenning.

- **If $Pr(Y) < 1 - Pr(Y)$, then ODDS < 1 .**
  - This means that the probabilty of the event NOT happenning is larger than the event happenning.

How do we interpret log-odds?

- **If** $ODDS = 1$**, then** $log(1) = 0$**.**
  - So, if the log-odds is 0, it means that there is a 50/50 chance/probability of the event happenning
- **If** $ODDS > 1$**, then** $log(\textbf{number} > \textbf{1}) > 0$
  - So, if the log-odds is greater than 0 (or positive), it means that the probability of the event happenning is grater than the probability of the event not happenning.
  - log is increasing, so, bigger log-odss indicates the ODDs is bigger and hence the probability of the event happenning is bigger
- **If** $ODDS < 1$**, then** $log(\textbf{number} > \textbf{1}) < 0$
  - So, if the log-odds is greater than 0 (or negative), it means that the probability of the event NOT happenning is greater than the probability of the event happenning.
  - smaller log-odds indicates the ODDs is smaller and hence the probability of the event NOT happenning is bigger

SEE graph of $y = log(x)$ at link to google

How to interpret log-odds

- If $log(ODDS) = \beta_1 > 0$ or positive, then the ODDS for the event compared to the baseline will increase.
  - $ODDS = e^{\beta_1}$, which would be greater than 1
  - Increasing the ODDS for the event compared to the baseline means that the probability for that event is getting bigger compared to the baseline.
- If $log(ODDS) = \beta_1 < 0$ or negative, then the ODDS for the event compared to the baseline will decrease.
  - The ODDS will be less than 1
  - Decreasing the ODDS for the event compared to the baseline means that the probability for that event is getting smaller compared to the baseline.

Let's look at an example of Logistic Regression

Let's predict Admit (Category - 0 or 1) using GREScore (numeric)

We now use `glm()` for generalized linear model instead of `lm()` (linear model)

```
> modGRE <- glm(Admit ~ GREScore, data = Grad, family = "binomial") #Always keep family="binomial" as t
> summary(modGRE)

Call:
glm(formula = Admit ~ GREScore, family = "binomial", data = Grad)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-3.05582   0.08826   0.17668   0.34376   1.35771
```

```
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -49.97079    7.98728  -6.256 3.94e-10 ***
GREScore      0.16913    0.02625   6.443 1.17e-10 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 237.37  on 399  degrees of freedom
Residual deviance: 166.69  on 398  degrees of freedom
AIC: 170.69

Number of Fisher Scoring iterations: 7
```

Notice that the results look very similar to what we are used to with `lm()`

Example of Logistic Regression

Questions might be:

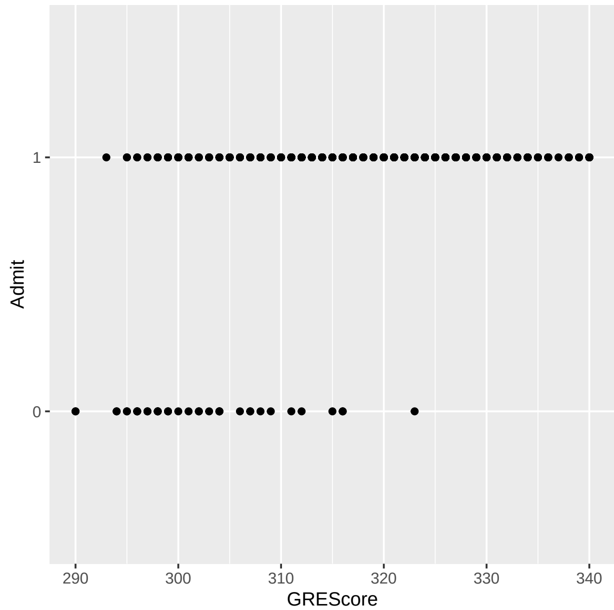
1. Is there a relationship between `GREScore` and `Admit`?
2. How strong is the relationship between `GREScore` and `Admit`?
3. What is the effect of `GREScore` and `Admit`?
4. Is `GREScore` a good predictor of `Admit`?
5. How good are the predictions based on your model?

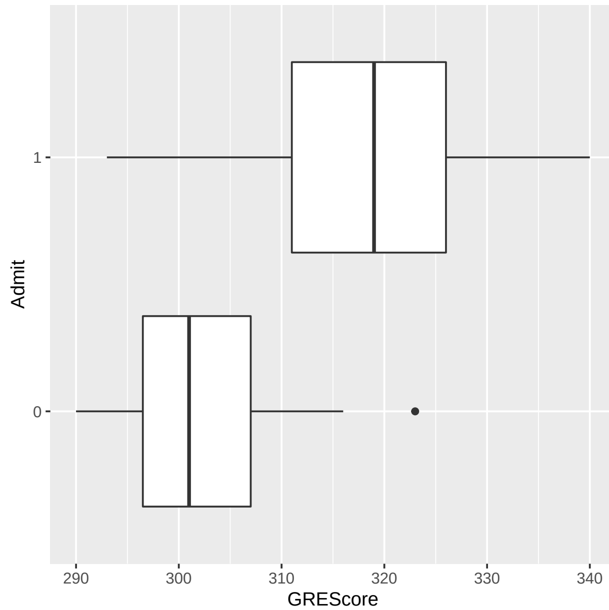
Example - Is there a relationship?

1. Is there a relationship between `GREScore` and `Admit`?

Again:

```
> a=ggplot(Grad, aes(x=Admit, y=GREScore))+geom_point()+coord_flip()
> b=ggplot(Grad, aes(x=Admit, y=GREScore))+geom_boxplot()+coord_flip()
> grid.arrange(a,b,nrow=1)
```





- See that there might be a difference, but not sure how big.
- Same plot, but 2 types.

Example - Strength of Relationship

2. How strong is the relationship between `GREScore` and `Admit`?
   - We will answer this by asking is the logistic model a good fit for the data?
   - We answer this using Deviance

- The **null deviance** shows how well the response variable is predicted by a model that includes only the intercept (overall mean) (i.e. the rpredictor is not included)
- The **residual deviance** shows how well the response variable is predicted by a model that includes the predictor.

In our results: **Null deviance: 237.37 on 399 df and Residual deviance: 166.69 on 398**

Above, you can see that addition of 1 (399-398 = 1) independent variable, the deviance decreased from 237.37 to 166.69. This is some reduction of the deviance, but there is still 166.69. How big is this? Have to compare to another model.

NOTE: If your Null Deviance is really small (close to 0), it means that the Null Model explains the data pretty well. Likewise with your Residual Deviance.

Example - Effect of Predictor

3. What is the effect of `GREScore` and `Admit`?

Since `GREScore` is numeric, we interpret it as a slope. However, as `X` changes, the `log(ODDS)` is changing.

```
(Intercept)    GREScore
-49.9707924   0.1691343
```

- as GREScore increases by 1 unit, the $log(ODDS)$ for 1 (Admitted) versus 0 (Not Admitted) increases by 0.1691343 units
  - For the response, the baseline is set alphabetically/numerically. So, we know that Admit = 0 is the baseline (i.e. the denominator of our odds). We are comparing the other category to this baseline.
$$ODDS = \frac{Pr(Admit = 1)}{Pr(Admit = 0)} = \frac{Pr(Admit = 1)}{1 - Pr(Admit = 1)}$$
  - Because this $log(ODDS) > 0$, it means that the $ODDS > 1$
  - Which then means that the odds for the Event (Admit=1) compared to the BASELINE is increasing as GREScores increases
  - This means that it is more likely for students to be in the Admit=1 Category as the GREScore increases.

Example - Effect of Predictor - In terms of ODDS

Recall from the previous slide: as GREScore increases by 1 unit, the $log(ODDS)$ for 1 (Admitted) versus 0 (Not Admitted) increases by 0.1691343 units

- So we can say that as GREScore increases, the odds for being Admitted (Admit=1) versus not (Admit=0) also increases.
  - Remember, if the log odds is increasing, so is the odds because the log function is an increasing function
- Specifically, we can say the odds for being Admitted (Admitted = 1) versus not is increasing by $e^{0.1691343} = 1.184279$ for every unit increase in GREScore
  - Recall that the inverse of the natural log is e^x

Example - Is it a good predictor?

4. Is `GREScore` a good predictor of `Admit`?

We do a hypothesis test here on the coefficients.

Justification of Hypotheses:

- No difference would imply that there is a 50/50 chance to be in either group. This would mean that $ODDS = 1$ and $log(ODDS) = 0$
  - **Research Hypothesis:** $log(ODDS) \neq 0$ or there is some difference between groups in terms of probability
  - **Null hypothesis:** $log(ODDS) = 0$ or there is no difference between groups in terms of probability

Solution:

- Here, `Pr(>|z|)` = 1.17 x 10^{-10} < 0.05, so we can conclude that we have evidence in favor of the Research Hypothesis that there is some difference.
- So, GREScore is a good predictor of Admit

Example - Accuracy of Predictions

5. How good are the predictions based on your model?
    - We will use AIC to select the model with the better predictions.
    - AIC is an estimator of out-of-sample prediction error
    - Lower AIC is better.
    - Again, we will need another model to compare this to in order to determine whch model is better for predictions.

In this case, AIC = 170.69

Exercise:

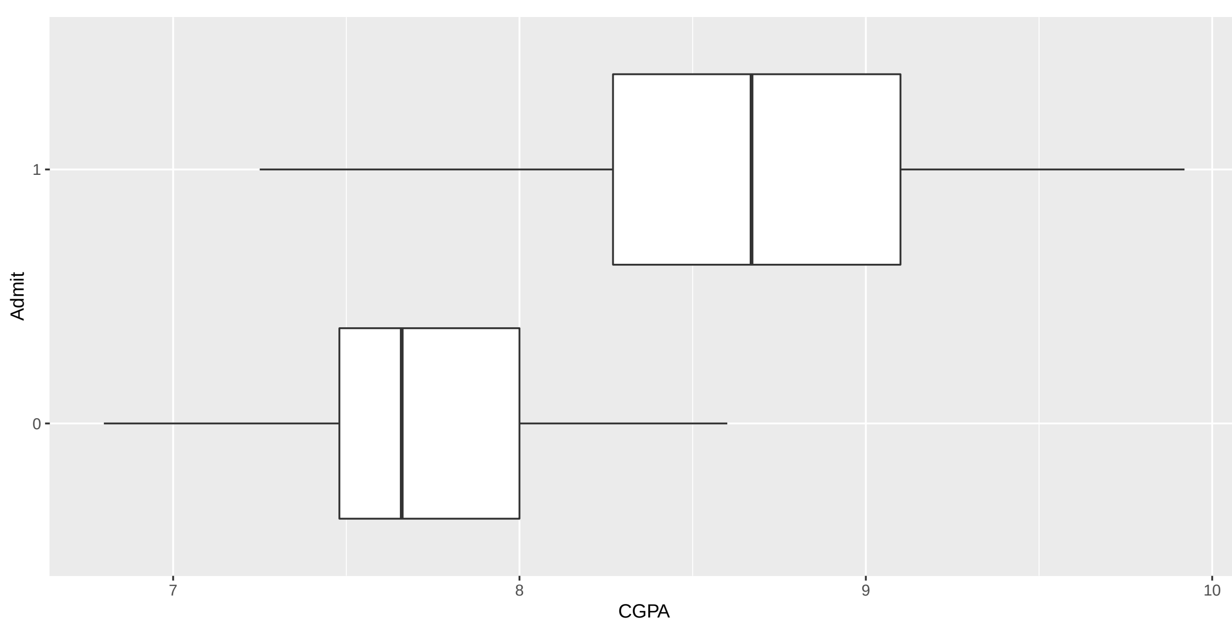
Use CGPA to predict Admit, then answe the following:

1. Is there a relationship between `CGPA` and `Admit`?
2. How strong is the relationship between `CGPA` and `Admit`?
3. What is the effect of `CGPA` and `Admit`?
4. Is `CGPA` a good predictor of `Admit`?
5. How good are the predictions based on your model?

Exercise Solutions

1. Is there a relationship between `CGPA` and `Admit`?

```
> ggplot(Grad, aes(x=Admit, y=CGPA))+geom_boxplot()+coord_flip()
```



There probably is a relationship.

Exercise Solutions

```
> modGPA <- glm(Admit ~ CGPA, data = Grad, family = "binomial") #Always keep family="binomial" as this
> summary(modGPA)

Call:
glm(formula = Admit ~ CGPA, family = "binomial", data = Grad)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-2.96569   0.04213   0.11535   0.28681   1.90345

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -33.9543     5.1407  -6.605 3.98e-11 ***
CGPA          4.4581     0.6498   6.861 6.84e-12 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 237.37  on 399  degrees of freedom
Residual deviance: 135.59  on 398  degrees of freedom
AIC: 139.59

Number of Fisher Scoring iterations: 7
```

Exercise Solutions

2. How strong is the relationship between `CGPA` and `Admit`?

Residual Deviance is 135.59. This model is a better fit than when GREScore was used to predict (Residual Deviance = 166.69)

3. What is the effect of `CGPA` and `Admit`?

```
(Intercept)        CGPA
 -33.954291    4.458088
```

- Effect
  - As CGPA increases by 1 unit, the $log(ODDS)$ for being Admitted (Admit=1) versus not (Admit=0) increases by 4.4581.
  - This suggests that the $ODDS$ for Admit=1 compared to the BASELINE, Admit=0, is increasing by $e^{4.4581} = 86.32334$. Notice that GPA has a bigger effect on Admit than GRE Score (where the odds only increased by 1.18)
  - Since the $ODDS$ for admitted versus not increases as CGPA increases, it is more likely for a student to be admitted (Admit=1) versus not (Admit=0) as CGPA increases.

Exercise Solutions

4. Is `CGPA` a good predictor of `Admit`?

Here, `Pr(>|z|)` $= 6.84 \text{ x } 10^{-12} < 0.05$, so we can conclude that we have evidence in favor of the Research Hypothesis that there is some difference. So, GGPA does contribute to predicting Admit.

5. How good are the predictions based on your model?

- In this case, AIC = 139.59.
- This is lower than when GREScore was used (170.69).
- So CGPA offers better predictions than GREScore.

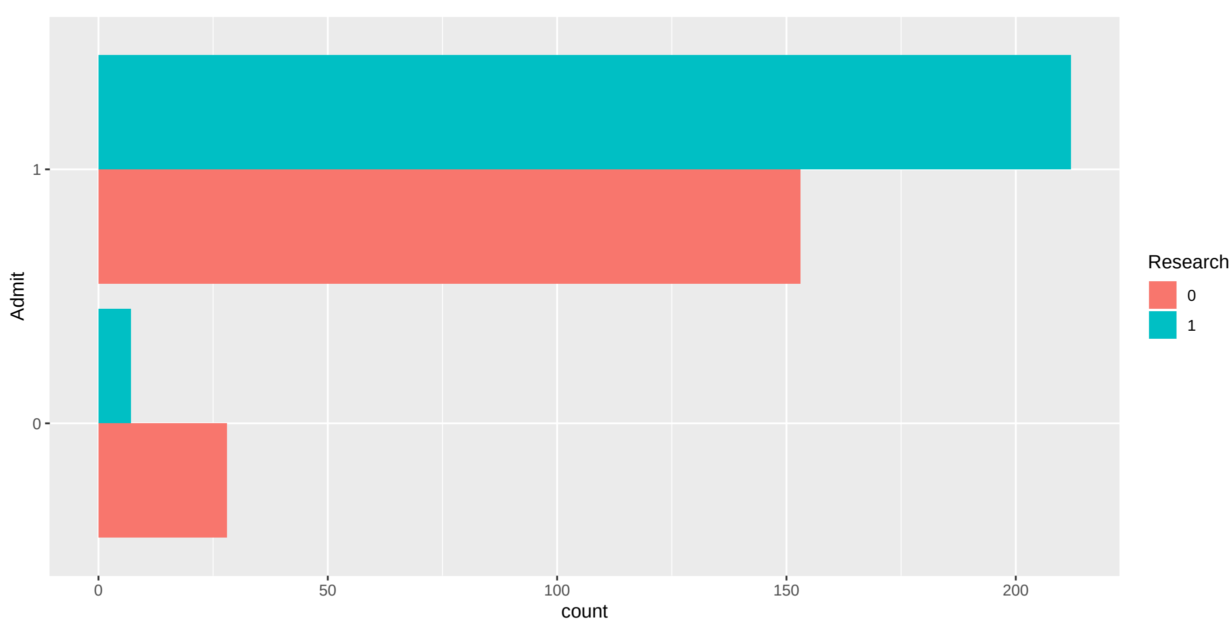
Qualitative Predictors

Let's use Research Experience to predict `Admit`

1. Is there a relationship between `Research` and `Admit`?
2. How strong is the relationship between `Research` and `Admit`?
3. What is the effect of `Research` and `Admit`?
4. Is `Research` a good predictor of `Admit`?
5. How good are the predictions based on your model?

Qualitative Predictors - Relationship

```
> ggplot(Grad, aes(x=Admit, fill=Research))+geom_bar(position="dodge")+coord_flip()
```



There probably is a relationship. The patterns for research experience flips for Admitted students versus those who were not admitted.

Qualitative Predictors - Fit Model

```
> modR <- glm(Admit ~ Research, data = Grad, family = "binomial") #Always keep family="binomial" as thi
> summary(modR)

Call:
glm(formula = Admit ~ Research, family = "binomial", data = Grad)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6242   0.2549   0.2549   0.5798   0.5798

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.6982     0.2055   8.262  < 2e-16 ***
Research1     1.7124     0.4357   3.930 8.48e-05 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 237.37  on 399  degrees of freedom
Residual deviance: 217.92  on 398  degrees of freedom
AIC: 221.92

Number of Fisher Scoring iterations: 6
```

Qualitative Predictors - Strength of Relationship

2. How strong is the relationship between `Research` and `Admit`?

Residual Deviance is 217.92 compared to the Null Deviance of 237.37.

This Residual Deviance is more than when GGPA was used (135.59) and also more than when GREScore was used (166.69). This model does not have as good a fit as when CGPA or GREScore was used. CGPA sill has the best fit.

Qualitative Predictios - Effect of Predictors

3. What is the effect of `Research` on `Admit`?

- **Recall:** we are predicting the log-odds for a student being admitted to grad school versus not based on Research Experience

```
(Intercept)   Research1
   1.698233    1.712443
```

- **BASELINE**
  - Predictor - Since the coefficient is for Research = 1 (Student has Research Experience), the baseline is Research = 0 (No Research Experience).

    - Response - We already know that Admit=0 is the baseline, so the log-odds compare Admit=1 to Admit=0

- **INTERCEPT INTERPRETATION**
    - The intercept refers to the log-odds for Admit=1 compared to the BASELINE, Admit=0, for students who do not have research experience (BASELINE)
        * the intercept is 1.698, which is > 0 and so the ODDS for ADMIT=1 versus the BASELINE , Admit=0, is > 1 and is increasing
        * this means that students are more likely to be Admitted (1), versus not (0 - BASELINE), if they do not have Research Experience
- **COEFFICIENT INTERPRETATION**
    - The coefficient, 1.712443, refers to difference in log-odds between students who were Admitted (Admit=1) and the BASELINE, Admit=0.
        * This number is positive. This means that the log-odds for students who have research experience is 1.698 + 1.712 = 3.411 which is 1.712 more than for students with no research experience.
        * This means that students who have research experience are even more likely than students who do not have research experience to get admitted to college versus not.

Qualitative Predictors - Is it a good predictor?

4. Is `Research` a good predictor of `Admit`?

Here, `Pr(>|z|)` = $8.48 \times 10^{-5} < 0.05$, so we can conclude that we have evidence in favor of the Research Hypothesis that there is some difference between students who have research experience or not. So, Research Experience does contribute to predicting Admit.

Qualitative Predictors - Accuracy of Predictions

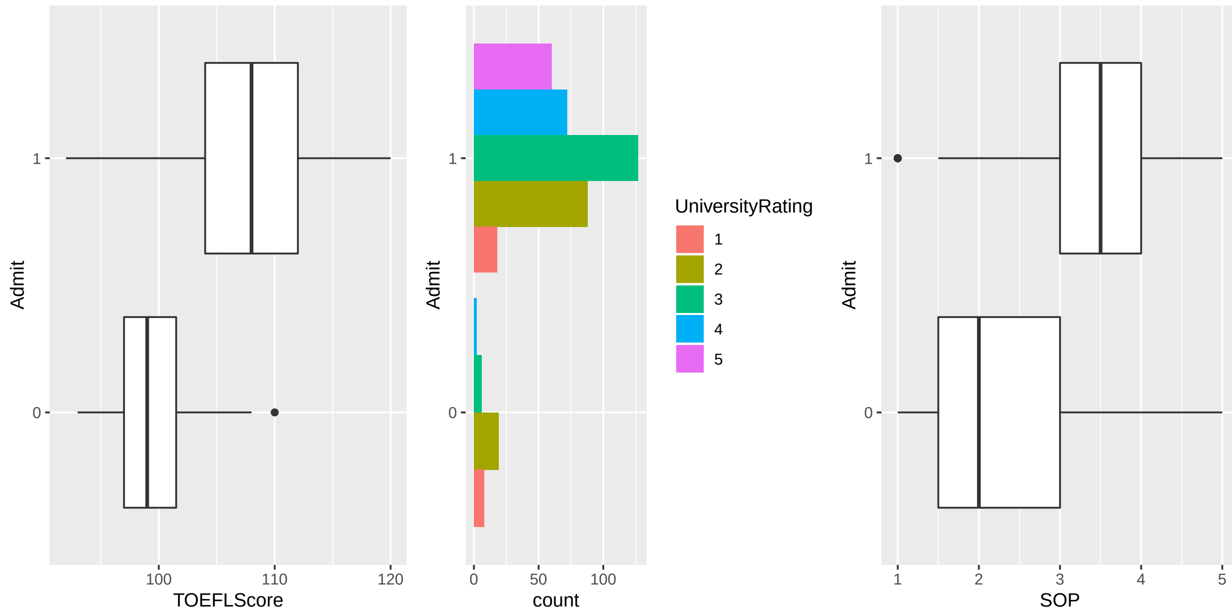
5. How good are the predictions based on your model?

- In this case, AIC = 221.92.
- This is higher than when GREScore was used (170.69) and CGPA (135.9).
- So Research Experience offers worse predictions than GREScore and CGPA.

Multiple Predictors - Relationship

Let's use TOEFLScore, University Rating and SOP to predict Admit.

```
> a=ggplot(Grad, aes(x=Admit, y=TOEFLScore))+geom_boxplot()+coord_flip()
> b=ggplot(Grad, aes(x=Admit, fill=UniversityRating))+geom_bar(position='dodge')+coord_flip()
> e=ggplot(Grad, aes(x=Admit, y=SOP))+geom_boxplot()+coord_flip()
> grid.arrange(a,b,e, nrow=1)
```



Multiple Predictors - Fit Model

```
> mod_mult <- glm(Admit ~ TOEFLScore + UniversityRating + SOP, data = Grad, family = "binomial") #Alway
> summary(mod_mult)

Call:
glm(formula = Admit ~ TOEFLScore + UniversityRating + SOP, family = "binomial",
    data = Grad)

Deviance Residuals:
     Min        1Q    Median        3Q       Max
-3.14358   0.03628   0.15799   0.35408   1.76384

Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
(Intercept)        -32.75777    6.35787  -5.152 2.57e-07 ***
TOEFLScore           0.33920    0.06543   5.184 2.17e-07 ***
UniversityRating2   -0.51711    0.60334  -0.857    0.391
UniversityRating3   -0.02778    0.79156  -0.035    0.972
UniversityRating4   -0.98167    1.11360  -0.882    0.378
UniversityRating5   14.16409 1215.86287   0.012    0.991
SOP                  0.11632    0.26887   0.433    0.665
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 237.37  on 399  degrees of freedom
Residual deviance: 159.02  on 393  degrees of freedom
AIC: 173.02
```

```
Number of Fisher Scoring iterations: 18
```

Multiple Predictors Solutions - Strength of Relationship

2. How strong is the relationship between the 3 predictors and `Admit`?

Residual Deviance is 159.02 compared to the Null Deviance of 237.37.

Recall that: This Residual Deviance for when GGPA was used (135.59), when GREScore was used (166.69) and when Research Experience was used (217.92). This model has abetter fit than GREScore and Research Experience, but not as good as CGPA.

Multiple Predictors - Effect of Predictors

3. What is the effect of the 3 predictors on `Admit`?

- **Recall:** we are predicting the log-odds for a student being admitted to grad school versus not based on Research Experience

```
      (Intercept)        TOEFLScore UniversityRating2 
     -32.75776724        0.33920249       -0.51710824 
UniversityRating3 UniversityRating4 UniversityRating5 
      -0.02778198       -0.98166585       14.16409287 
              SOP 
       0.11632492 
```

- Quantitative Predictors
  - **TOEFLScore:** log-odds is 0.3392 which is > 0 which means the ODDS $= e^{0.339} > 1$. This means that as the TOEFLScore increases, the odds for being admitted versus not increases by $e^{0.339} = 1.403$ (or the log-odds increases by 0.339). This means it is more likely for students to be admitted as their TOEFLScore increases.
  - **SOP:** log-odds is 0.1163 which is > 0 which means the ODDS $= e^{0.1163} > 1$. This means that as the SOP increases, the odds for being admitted versus not increases by $e^{0.1163} = 1.1232$ (or the log-odds increases by 0.1163). This means it is more likely for students to be admitted as their SOP increases.

Overall: TOEFLScore has a bigger effect on Admit than SOP since the log-odds (and hence the odds) for Admit versus not is bigger.

Multiple Predictors - Effect of Predictors

- Qualitative Predictors
  - Since twe have coefficients for University Rating = 2,3,4,5, the baseline is University Rating = 1.
- **INTERCEPT INTERPRETATION**
  - The intercept cannot be interpreted like before because there are other things in the model besised UniversityRating
- **COEFFICIENT INTERPRETATION**

    - The coefficient for UniversityRating 2 is -0.5171. this refers to difference in log-odds for Admitted (Admit=1) compared to the BASELINE, Admit=0, for students who go to university with rating 2 compared to rating 1
        * This number is negative. This means that the log-odds for students who have research experience is 0.5171 LESS than for students who come from a univeristy with rating 1. (OR the odds is $e^{0.5171} = 1.677$ LESS)
        * This means that students who come from a University with rating 1 are less likely to be admitted versus not than students who come from a university with rating 1.

- *LIKEWISE:*
    - The coefficient for UniversityRating 2 is -0.5171, for 3 is -0.02778, for 4 is -0.9817. These students who come from universities with these ratings are less likely to be admitted versus not than students who come from a university with rating 1.
    - The coefficient for rating 5 is 14.164, so students who come from a university with rating 5 are more likely to be admitted to to grad school versus not than students who come from a university with rating 1.

Multiple Predictors - Is it a good predictor?

4. Are the 3 variables good predictors of `Admit`?

`Pr(>|z|)` for the following variables are all > 0.05 except for TOEFLScore.

- TOEFLScore is the only good predictor of Admit. This is the only variable we detect a meaningful difference

- This would explain why the interpretations of the effects for UniversityRatings are not as expected. The odds/log/odds for Ratings 2,3,4 and 5 are all essenatially not different than Rating 1.

Multiple Predictors - Accuracy of Predictions

5. How good are the predictions based on your model?

- In this case, AIC = 173.02
- This is higher than when GREScore was used (170.69) and CGPA (135.9).
- But less than wehen research Experience was used (221.92).

So, CGPA is still the better model for predictions!

Multiple Predictors - Model/Variable selection

Again, Let's use the `stepAIC`function.

```
> library(MASS)
> sel=stepAIC(mod_mult, trace = FALSE)
> sel$anova
Stepwise Model Path
Analysis of Deviance Table

Initial Model:
```

```
Admit ~ TOEFLScore + UniversityRating + SOP

Final Model:
Admit ~ TOEFLScore

                Step Df  Deviance Resid. Df Resid. Dev
1                                       393   159.0156
2 - UniversityRating  4 2.9767224       397   161.9923
3              - SOP  1 0.2941248       398   162.2864
       AIC
1 173.0156
2 167.9923
3 166.2864
```

We see that the best model using stepwise selection and using AIC as the criteria uses only TOEFLScore to predict Admit (instead of using all of TOEFLScore, SOP and UniversityRating)

You can add `direction="forward"` or `direction="backward"` in the `stepAIC()` function to change algorithm.