Introduction

The aim of this assignment is to create a model that can predict the winner of the 2024 Oscars best picture category. Employing the logistic regression means the response variable must be binary, 0 or 1, indicating when a movie didn't win or won respectively. **Appendix I** shows the code which created a subset titled 'OscarsSubset' for the purposes of training the model.

Task 1

The first task required fitting a full model, interpreting one of the significant coefficients and report the confidence interval.

As seen in **Appendix II,** the response variable is 'Ch' (Winner indicator), and the explanatory variables is every other variable except for 'Year', 'Name' and 'Pic' (Picture nomination). 'Pic' provides no additional information as all movies in the data set have been nominated. Furthermore, we set the family of the GLM to be binomial (by default it has the logit link) and data to be the 'OscarSubset' data set.

As observed in the output**,** 'Dir' (Director Nomination) is a significant variable stating that its fitted estimator is 1.780. Since the model is a GLM, one can interpret this coefficient through log odds suggesting that having a director nomination increases the log odds of winning best picture by 1.780, therefore an odds ratio of exp{1.780} ≈ 5.93. So, one could say that the director nomination roughly will increase the chances of winning best picture 6 times.

One can obtain the 95% confidence interval of the significant coefficient through doing the following calculation. Let the fitted estimator be denoted by $\hat{\beta}$,

exp{$\hat{\beta} \pm 1.96 s.e(\hat{\beta})$} = exp{1.78 $\pm$ 1.96(0.6959)} ≈ (1.516,23.20) (4sf).

Task 2

This task was to use an appropriate model selection strategy to find and fit the best model. My chosen model selection strategy is the stepwise search which created 6 models, AIC and BIC backward step searches, backward search using the Likelihood Ratio Test (LRT), and their forward-step counterparts. To avoid overfitting, I didn't consider interaction terms.

In **Appendix III,** the step function iteratively does a backwards search starting with the full model and removes the variable with the lowest AIC until the AIC of the model can no longer improve. **Appendix IV** also uses the step function, but the penalty term changes for the search to use BIC, similar to **Appendix III,** it removes terms with the lowest BIC until the BIC of the overall model can no longer improve. **Appendix V uses** the drop1 function. The code conducts the LRT and removes the variable with the largest p value that is not significant. It then updates the function with the removed variable and use the drop1 function again. This process repeats until there are only left with significant variables. As seen in the output, drop1 with the LRT on fit56, observe that 'SciFi' has an insignificant p value of 0.0655, thus eliminating the variable. Then we update fit56, calling it fit57 and perform the drop1 function again. This time every variable is significant; therefore, no further variables are removed and fit57 is our model achieved through backwards selection with LRT.

Now we obtain the models through forward selection. In **Appendix VI,** I defined the Intercept model and define the scope to be every variable in the full model. Modifying the stepwise search to do forward selection starting from the intercept model, the code selects the variable with the lowest AIC and repeats until the AIC of the overall model can no longer be improved. In **Appendix VII,** the code follows the same process except with BIC. In **Appendix VIII,** we perform a stepwise search through the add1 function testing with the LRT. Each variable added is the most significant variable (smallest p-value) and update the fit with the new variable and repeat. As seen in the output, we add the variable 'Days' because it is the smallest significant variable, and when we update the model and preform add1 with LRT on it, there are no more significant variables that can be added, hence we stop there. Now we have the following models to choose from:

Model	Origin of Model	Equation
Model 1	Backwards step function with AIC	$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_5 x_{i5} + \beta_6 x_{i6} + \beta_7 x_{i7} + \beta_8 x_{i8} + \beta_9 x_{i9} + \beta_{10} x_{i10} + \beta_{11} x_{i11} + \beta_{12} x_{i12} + \beta_{13} x_{i13} + \beta_{14} x_{i14} + \beta_{15} x_{i15}$
Model 2	Backwards step function with BIC	$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_6 x_{i6}$
Model 3	drop1 function with LRT	$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_6 x_{i6} + \beta_{11} x_{i11} + \beta_{12} x_{i12} + \beta_{13} x_{i13} + \beta_{14} x_{i14}$
Model 4	Forwards step function with AIC	$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_6 x_{i6} + \beta_9 x_{i9} + \beta_{10} x_{i10} + \beta_{14} x_{i14} + \beta_{16} x_{i16} + \beta_{17} x_{i17}$
Model 5	Forwards step function with BIC	$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_6 x_{i6}$
Model 6	add1 function with LRT	$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \beta_3 x_{i3} + \beta_4 x_{i4} + \beta_6 x_{i6} + \beta_9 x_{i9} + \beta_{10} x_{i10}$
ModelFull	Full Model	
InterceptModel	Model with only the intercept term	$y_i = \beta_0$

Table 1. List of models to be selected from (See **Appendix IX** for listing of models in R).

$$x_{i1} = \begin{cases} 1, & Director\ Nomination \\ 0, & Otherwise \end{cases} \quad x_{i2} = \begin{cases} 1, & Editing\ Nominated \\ 0, & Otherwise \end{cases}$$

$$x_{i3} = \begin{cases} 1, & Dance\ Direction\ Nomination \\ 0, & Otherwise \end{cases} \quad x_{i4} = \begin{cases} 1, & Golden\ Globe\ Drama\ Winner \\ 0, & Otherwise \end{cases}$$

$$x_{i5} = \begin{cases} 1, & Golden\ Globe\ Director\ Winner \\ 0, & Otherwise \end{cases} \quad x_{i6} = \begin{cases} 1, & Producers\ Guild\ Winner \\ 0, & Otherwise \end{cases}$$

$$x_{i7} = \begin{cases} 1, & Directors\ Guild\ Winner \\ 0, & Otherwise \end{cases} \quad x_{i8} = \begin{cases} 1, & Genre\ is\ Romance \\ 0, & Otherwise \end{cases}$$

$$x_{i9} = \begin{cases} 1, & Genre\ is\ SciFi \\ 0, & Otherwise \end{cases} \quad x_{i10} = Days\ between\ release\ date\ and\ oscars\ ceremony,$$

$$x_{i11} = \begin{cases} 1, & PG\ rated \\ 0, & Otherwise \end{cases} \quad x_{i12} = \begin{cases} 1, & PG13\ Rated \\ 0, & Otherwise \end{cases}$$

$$x_{i13} = \begin{cases} 1, & R\ Rated \\ 0, & Otherwise \end{cases} \quad x_{i14} = \begin{cases} 1, & National\ Society\ of\ Film\ Critics\ Winner \\ 0, & Otherwise \end{cases}$$

$$x_{i15} = Weighted\ averaged\ IMDB\ user\ rating, x_{i16} = Roger\ Ebert\ Rating,$$
$$x_{i17} = Number\ of\ Oscar\ Nominations.$$

To select between the models, a series of ANOVA tests were performed among the nested models. If the p>0.05, then the simpler model is more adequate, and vice versa if p<0.05, (See **Appendix X** for R outputs to ANOVA tests). First, it was checked if Model 3 and Model 6 are more adequate than the full and intercept model respectively. According to the tables, Model 3 and Model 6 are more adequate than the full and intercept models. Model 2 is nested in Model 1; we conducted the ANOVA test to show that Model 1 is more adequate. Model 5 is nested in Model 4; an ANOVA test was carried out to show that Model 4 is more adequate. This leaves Model 1, 3, 4, and 6 remaining to choose from. Model 6 is nested in Model 4 and Model 3 is nested in Model 1, an ANOVA test shows that Model 4 and Model 1 are more adequate models.

This narrowed the choice to Model 1 and Model 4, however because these two models are not nested, an ANOVA test would be inappropriate. Checking the AIC and BIC of the two models (See **Appendix X**), Model 1 has an AIC and BIC of 313.9 and 384.4 respectively, while Model 4 has a respective AIC and BIC of 384.4 and 364.4. The AIC quantities are quite close together, while the BIC quantities differ by 20. Additionally Model 4 is simpler, having 5 less variables to consider than Model 1, hence justifying Model 4 as the final model.

Task 3

This section aims to find the AUC of the final model as well as the optimal threshold to from the ROC curves to later find the sensitivity of the model.

Since we are applying a logistic regression, we only consider the logit link. The code finds the AUC to be 0.9164 and the optimal threshold to be 0.1337 (4sf) (See **Appendix XI**).

	Prediction	
Observed	False	True
Didn't Win	421	87
Win	15	81

Table 2. Table of predictions by model against observed values.

From this, we can calculate the sensitivity to be $\frac{81}{81+15} = 0.84375$.

Task 4

Now equipped with the final model, with the AUC is 0.92 indicating excellent discrimination, we can now predict which movie will win the best picture category for the 2024 Oscars. Since the model predicts the winner among every movie in the data set, we must scale the probabilities in the 2024 Oscars. This is done by creating a subset of the Oscars with only 2024 movies and apply the trained model to predict the outcome on this subset. We scale the probabilities by summing all the probabilities and dividing the predicted probability by the sum of probabilities for each movie and get this final table indicating that Anora is most likely to win best picture, as seen in **Appendix XII.**

Movie	Percentage chance of winning (%) (4sf)
A Complete Unknown	2.558
Anora	57.48
Concalve	1.644
Dune: Part 2	0.1090
Emilia Perez	18.96
I'm Still Here	0.3685
Nickel Boys	0.8693
The Brutalist	15.47
The Substance	0.2889
Wicked	2.249

Table 3. The probabilities of each movie winning.

Task 5

The issue with this model is that there is no constraint that in each year, movies are competing with other movies in the same year, therefore the probabilities in each year should add to 1. A suitable alternative model is called the Plackett-Luce Model.

The Placket-Luce Model works through an implicit ranking of the events, in our case the implicit ranking can be done through a variable such as the number of nominations or the weighted average IMDB rating. Suppose we had a set S where S = $\{i_1, i_2, \dots, i_n\}$. Each $i_j$is given a 'Worthiness' parameter $\theta_i > 0\ \forall i$ which is a relative scale where a higher number indicates a higher likelihood of success. In this model, $\mathbb{P}\{Event\ i_j\ takes\ place\} = {\theta_i}\Big/{\sum_{k\in S}\theta_k}$ which guarantees that the total probability in each year adds up to 1 and each year is taken independently.

APPENDIX

Appendix I

oscars <- read.csv("oscars.csv",header = TRUE)
attach(oscars)
#Create a subset to analyse the movies before 2024
OscarsSubset <- subset(oscars, ï..Year < 2024) #Unexpectedly, the attached data
#was saving the Year column as ï..Year
OscarsSubset$Ch <- ifelse(OscarsSubset$Ch == 2, 0, 1) #Redefine 2s to 0s
OscarsSubset$Ch <- factor(OscarsSubset$Ch, levels = c(0, 1), labels = c("Didn't Win", "Win")) #Redifine 0s and 1s
contrasts(OscarsSubset$Ch) <- contr.treatment(2)

#To check that the leveling is correct
str(OscarsSubset$Ch)
table(OscarsSubset$Ch)
contrasts(OscarsSubset$Ch)

```
> oscars <- read.csv("oscars.csv",header = TRUE)
> attach(oscars)
> #Create a subset to analyse the movies before 2024
> OscarsSubset <- subset(oscars, ï..Year < 2024)
> OscarsSubset$Ch <- ifelse(OscarsSubset$Ch == 2, 0, 1) #Redefine 2s to 0s
> OscarsSubset$Ch <- factor(OscarsSubset$Ch, levels = c(0, 1), labels = c("Didn't Win", "Win"))
 #Redifine 0s and 1s
> contrasts(OscarsSubset$Ch) <- contr.treatment(2)
>
> #To check that the leveling is correct
> str(OscarsSubset$Ch)
 Factor w/ 2 levels "Didn't Win","Win": 1 1 1 1 1 2 1 1 1 1 ...
 - attr(*, "contrasts")= num [1:2, 1] 0 1
  ..- attr(*, "dimnames")=List of 2
  .. ..$ : chr [1:2] "Didn't Win" "Win"
  .. ..$ : chr "2"
> table(OscarsSubset$Ch)

Didn't Win        Win
       508         96
> contrasts(OscarsSubset$Ch)
           2
Didn't Win 0
Win        1
>
```

Stats Modelling Coursework.R | oscars | Oscars2024

Filter Cols: 1 - 50

	ï..Year	Name	Ch	Nom	Pic	Dir	Aml	Afl	Ams	Afs	Scr	Ci
1	2024	A Complete Unkown	0	8	1	1	1	0	1	1	1	
2	2024	Anora	0	6	1	1	0	1	1	0	1	
3	2024	Conclave	0	8	1	0	1	0	0	1	1	
4	2024	Dune: Part Two	0	5	1	0	0	0	0	0	0	
5	2024	Emilia Perez	0	13	1	1	0	1	0	1	1	
6	2024	I'm Still Here	0	3	1	0	0	1	0	0	0	
7	2024	Nickel Boys	0	2	1	0	0	0	0	0	1	
8	2024	The Brutalist	0	10	1	1	1	0	1	1	1	
9	2024	The Substance	0	5	1	1	0	1	0	0	1	
10	2024	Wicked	0	10	1	0	0	1	0	1	0	
11	2023	American Fiction	2	5	1	1	1	0	1	0	1	

Stats Modelling Coursework.R | OscarsSubset | oscars | Oscars2024

Filter Cols: 1 - 50

	ï..Year	Name	Ch	Nom	Pic	Dir	Aml	Afl	Ams
11	2023	American Fiction	Didn't Win	5	1	1	1	0	
12	2023	Anatomy of a Fall	Didn't Win	5	1	1	0	1	
13	2023	Barbie	Didn't Win	8	1	0	0	0	
14	2023	Killers of the Flower Moon	Didn't Win	10	1	1	0	1	
15	2023	Maestro	Didn't Win	7	1	0	1	1	
16	2023	Oppenheimer	Win	7	1	1	1	0	
17	2023	Past Lives	Didn't Win	2	1	0	0	0	
18	2023	Poor Things	Didn't Win	11	1	1	0	1	
19	2023	The Holdovers	Didn't Win	5	1	0	1	1	
20	2023	The Zone of Interest	Didn't Win	5	1	1	0	0	
21	2022	All Quiet on the Western Front	Didn't Win	9	1	0	0	0	

Appendix II

```
ModelFull <- glm (Ch~Nom+Dir+Aml+Afl+Ams+Afs+Scr+Cin+Art+Cos+Sco+Son+Edi+
                  Sou+For+Anf+Eff+Mak+Dan+AD+Gdr+Gmc+Gd+Gm1+Gm2+
                  Gf1+Gf2+PGA+DGA+Action +Adventure+ Animation+
                  Biography+Comedy+Crime+Docu+Drama+ Family+ Fantasy+
                  Film.noir+History+Horror+Music+Musical+Mystery+Romance+
                  SciFi+Sport+Thriller+War+Western+Length+Days+G+PG+
                  PG13+R+ U+Ebert+NYFCC+LAFCA+NSFC+NBR+WR,
                  family = binomial, data = OscarsSubset)
```

summary(ModelFull)

```
> ModelFull <- glm(Ch~Nom+Dir+Aml+Afl+Ams+Afs+Scr+Cin+Art+Cos+Sco+Son+Edi+Sou+For+Anf+Eff+Mak+Dan+AD+Gdr+Gmc
+                  +Gd+Gm1+Gm2+Gf1+Gf2+PGA+DGA+Action+Adventure+Animation+Biography+Comedy+Crime+Docu+Drama+Family
+                  +Fantasy+Film.noir+History+Horror+Music+Musical+Mystery+Romance+SciFi+Sport+Thriller+War+Western
+                  +Length+Days+G+PG+PG13+R+U+Ebert+NYFCC+LAFCA+NSFC+NBR+WR, family = binomial, data = OscarsSubset)
> summary(ModelFull)

Call:
glm(formula = Ch ~ Nom + Dir + Aml + Afl + Ams + Afs + Scr + 
    Cin + Art + Cos + Sco + Son + Edi + Sou + For + Anf + Eff + 
    Mak + Dan + AD + Gdr + Gmc + Gd + Gm1 + Gm2 + Gf1 + Gf2 + 
    PGA + DGA + Action + Adventure + Animation + Biography + 
    Comedy + Crime + Docu + Drama + Family + Fantasy + Film.noir + 
    History + Horror + Music + Musical + Mystery + Romance + 
    SciFi + Sport + Thriller + War + Western + Length + Days + 
    G + PG + PG13 + R + U + Ebert + NYFCC + LAFCA + NSFC + NBR + 
    WR, family = binomial, data = OscarsSubset)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)    
(Intercept) -9.459e+00  3.063e+00  -3.088  0.00202 ** 
Nom         -1.143e-02  3.910e-01  -0.029  0.97668    
Dir          1.780e+00  6.959e-01   2.557  0.01055 *  
Aml         -1.211e-02  4.934e-01  -0.025  0.98041    
Afl          2.905e-01  5.629e-01   0.516  0.60578    
Ams          6.048e-01  4.765e-01   1.269  0.20438    
Afs          1.481e-01  4.990e-01   0.297  0.76660    
Scr          3.845e-01  6.442e-01   0.597  0.55063    
Cin         -1.487e-01  5.846e-01  -0.254  0.79921    
Art          2.350e-01  5.889e-01   0.399  0.68985    
Cos         -5.047e-01  6.502e-01  -0.776  0.43754    
Sco          2.931e-01  5.493e-01   0.534  0.59357    
Son         -1.610e-01  8.498e-01  -0.190  0.84970    
Edi          1.207e+00  5.636e-01   2.142  0.03223 *  
Sou         -5.270e-01  6.362e-01  -0.828  0.40744    
For          4.318e-01  1.465e+00   0.295  0.76824    
Anf          1.065e+00  4.817e+03   0.000  0.99982    
Eff          2.175e-01  6.460e-01   0.337  0.73630    
Mak          1.457e+00  9.438e-01   1.544  0.12257    
Dan          2.769e+00  1.628e+00   1.701  0.08895 .  
AD           1.683e+00  1.314e+00   1.281  0.20003    
Gdr          1.305e+00  4.610e-01   2.830  0.00465 ** 
Gmc         -1.290e-01  7.790e-01  -0.166  0.86850    
Gd          -2.950e+00  1.770e+00  -1.667  0.09556 .  
Gm1          1.467e+00  1.269e+00   1.157  0.24740    
Gm2         -1.193e+00  3.497e+00  -0.341  0.73293    
Gf1         -1.558e+00  1.807e+00  -0.862  0.38856    
Gf2         -1.435e+01  1.772e+03  -0.008  0.99354    
PGA          3.571e+00  4.994e-01   7.150 8.69e-13 ***
DGA          1.905e+00  1.331e+00   1.431  0.15240    
Action      -6.718e-01  7.969e-01  -0.843  0.39925    
Adventure   -2.477e-01  7.548e-01  -0.328  0.74279    
Animation   -1.172e+01  3.956e+03  -0.003  0.99764    
Biography   -5.795e-01  6.509e-01  -0.890  0.37334    
Comedy      -1.364e-01  5.373e-01  -0.254  0.79953    
Crime        1.021e+00  6.313e-01   1.617  0.10594    
Docu        -1.333e+01  3.956e+03  -0.003  0.99731    
Drama       -9.871e-01  6.238e-01  -1.582  0.11359    
Family       1.222e+00  8.718e-01   1.402  0.16102    
Fantasy     -1.039e+00  1.169e+00  -0.888  0.37443    
Film.noir   -6.244e-01  1.425e+00  -0.438  0.66130    
History      3.740e-01  7.037e-01   0.531  0.59508    
Horror      -8.303e-01  2.053e+00  -0.404  0.68593    
Music        7.516e-01  9.649e-01   0.779  0.43604    
Musical      8.186e-01  8.582e-01   0.954  0.34012    
Mystery      6.285e-01  7.954e-01   0.790  0.42944    
Romance      4.362e-01  4.183e-01   1.043  0.29704    
SciFi       -1.084e+00  1.637e+00  -0.662  0.50792    
Sport        5.060e-01  1.178e+00   0.430  0.66745    
Thriller    -8.074e-01  7.014e-01  -1.151  0.24973    
War          9.575e-01  6.376e-01   1.502  0.13317    
Western     -3.627e-02  9.434e-01  -0.038  0.96933    
Length       1.789e-03  8.491e-03   0.211  0.83311    
Days         2.381e-03  1.621e-03   1.469  0.14174    
G           -1.985e+00  2.029e+00  -0.978  0.32803    
PG          -1.135e+00  6.883e-01  -1.649  0.09922 .  
PG13        -1.424e+00  8.666e-01  -1.643  0.10038    
R           -1.438e+00  7.088e-01  -2.030  0.04241 *  
U           -1.384e+01  1.989e+03  -0.007  0.99445    
Ebert        1.325e-01  1.683e-01   0.787  0.43115    
NYFCC        1.660e-01  4.684e-01   0.354  0.72304    
LAFCA       -8.886e-01  7.154e-01  -1.242  0.21423    
NSFC         1.880e+00  7.502e-01   2.506  0.01221 *  
NBR         -1.391e-01  4.917e-01  -0.283  0.77719    
WR           5.265e-01  3.997e-01   1.317  0.18771    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 528.99  on 603  degrees of freedom
Residual deviance: 258.01  on 539  degrees of freedom
AIC: 388.01

Number of Fisher Scoring iterations: 16
```

Appendix III

AICModelBackwardsStep <- step(ModelFull, k = 2)

```
Step:  AIC=314.61
Ch ~ Dir + Ams + Edi + Mak + Dan + Gdr + Gd + PGA + DGA + Romance +
    SciFi + Days + PG + PG13 + R + NSFC + WR

          Df Deviance    AIC
- Mak      1   280.35 314.35
- Ams      1   280.56 314.56
<none>         278.61 314.61
- Romance  1   280.94 314.94
- DGA      1   280.99 314.99
- WR       1   281.05 315.05
- PG       1   281.38 315.38
- Days     1   281.85 315.85
- PG13     1   282.09 316.09
- SciFi    1   282.23 316.23
- Gd       1   282.81 316.81
- Dan      1   284.32 318.32
- R        1   285.23 319.23
- NSFC     1   285.72 319.72
- Gdr      1   286.42 320.42
- Edi      1   287.82 321.82
- Dir      1   291.50 325.50
- PGA      1   357.25 391.25

Step:  AIC=314.35

Step:  AIC=314.35
Ch ~ Dir + Ams + Edi + Dan + Gdr + Gd + PGA + DGA + Romance +
    SciFi + Days + PG + PG13 + R + NSFC + WR

          Df Deviance    AIC
- Ams      1   281.92 313.92
<none>         280.35 314.35
- Romance  1   282.64 314.64
- PG       1   282.78 314.78
- PG13     1   283.04 315.04
- WR       1   283.22 315.22
- DGA      1   283.33 315.33
- Days     1   283.51 315.51
- Gd       1   283.97 315.97
- SciFi    1   284.00 316.00
- Dan      1   286.04 318.04
- R        1   286.54 318.54
- NSFC     1   286.94 318.94
- Gdr      1   288.37 320.37
- Edi      1   290.32 322.32
- Dir      1   292.65 324.65
- PGA      1   361.67 393.67

Step:  AIC=313.92
Ch ~ Dir + Edi + Dan + Gdr + Gd + PGA + DGA + Romance + SciFi +
    Days + PG + PG13 + R + NSFC + WR

          Df Deviance    AIC
<none>         281.92 313.92
- Romance  1   284.16 314.16
- PG       1   284.62 314.62
- Days     1   284.81 314.81
- PG13     1   284.99 314.99
- DGA      1   285.19 315.19
- WR       1   285.31 315.31
- Gd       1   285.91 315.91
- SciFi    1   285.98 315.98
- Dan      1   287.63 317.63
- R        1   287.75 317.75
- NSFC     1   288.19 318.19
- Gdr      1   290.49 320.49
- Edi      1   292.39 322.39
- Dir      1   294.62 324.62
- PGA      1   366.86 396.86
```

Appendix IV

```
n <- dim(OscarsSubset)[1]
BICModelBackwardsStep <- step(ModelFull, k = log(n))
```

```
Ch ~ Dir + Edi + Dan + Gdr + Gd + PGA + SciFi + Days + NSFC

        Df Deviance    AIC
- Gd     1   299.83 357.46
- NSFC   1   302.17 359.80
- SciFi  1   302.40 360.03
- Days   1   302.62 360.26
<none>       297.93 361.97
- Dan    1   304.72 362.35
- Gdr    1   305.66 363.29
- Edi    1   308.93 366.56
- Dir    1   315.53 373.16
- PGA    1   393.46 451.10

Step:  AIC=357.46
Ch ~ Dir + Edi + Dan + Gdr + PGA + SciFi + Days + NSFC

        Df Deviance    AIC
- NSFC   1   302.95 354.18
- Days   1   304.21 355.44
- SciFi  1   304.56 355.79
- Gdr    1   306.16 357.39
<none>       299.83 357.46
- Dan    1   306.54 357.77
- Edi    1   311.33 362.56
- Dir    1   317.27 368.50
- PGA    1   395.31 446.54

Step:  AIC=354.18
Ch ~ Dir + Edi + Dan + Gdr + PGA + SciFi + Days

        Df Deviance    AIC
- Days   1   307.48 352.30
- SciFi  1   308.17 353.00
- Gdr    1   308.99 353.81
- Dan    1   309.26 354.09
<none>       302.95 354.18
- Edi    1   315.64 360.47
- Dir    1   322.77 367.60
- PGA    1   396.50 441.32

Ch ~ Dir + Edi + Dan + Gdr + PGA + SciFi

        Df Deviance    AIC
- SciFi  1   312.38 350.81
- Gdr    1   312.66 351.08
<none>       307.48 352.30
- Dan    1   313.91 352.33
- Edi    1   319.12 357.54
- Dir    1   327.15 365.57
- PGA    1   400.26 438.68

Step:  AIC=350.81
Ch ~ Dir + Edi + Dan + Gdr + PGA

       Df Deviance    AIC
- Gdr   1   318.48 350.50
<none>      312.38 350.81
- Dan   1   319.02 351.04
- Edi   1   322.55 354.56
- Dir   1   331.56 363.58
- PGA   1   402.42 434.44

Step:  AIC=350.5
Ch ~ Dir + Edi + Dan + PGA

       Df Deviance    AIC
- Dan   1   324.47 350.08
<none>      318.48 350.50
- Edi   1   332.03 357.64
- Dir   1   342.55 368.17
- PGA   1   429.82 455.44

Step:  AIC=350.08
Ch ~ Dir + Edi + PGA

       Df Deviance    AIC
<none>      324.47 350.08
- Edi   1   338.74 357.95
- Dir   1   348.27 367.48
- PGA   1   435.02 454.23
```

Appendix V

```
fit1 <- ModelFull
drop1(fit1,test = "LRT")       #Remove by highest p value which is Anf
fit2 <- update(fit1, .~. - Anf)
drop1(fit2,test = "LRT")       #Remove Aml
fit3 <- update(fit2, .~. - Aml)
drop1(fit3,test = "LRT")       #Remove Western
fit4 <- update(fit3, .~. - Western)
drop1(fit4,test = "LRT")       #Remove Nom
fit5 <- update(fit4, .~. - Nom)
drop1(fit5,test = "LRT")       #Remove Animation
fit6 <- update(fit5, .~. - Animation)
drop1(fit6,test = "LRT")       #Remove Gmc
fit7 <- update(fit6, .~. - Gmc)
drop1(fit7,test = "LRT")       #Remove Docu
fit8 <- update(fit7, .~. - Docu)
drop1(fit8,test = "LRT")       #Remove Length
fit9 <- update(fit8, .~. - Length)
drop1(fit9,test = "LRT")       #Remove Son
fit10 <- update(fit9, .~. - Son)
drop1(fit10,test = "LRT")      #Remove NBR
fit11 <- update(fit10, .~. - NBR)
drop1(fit11,test = "LRT")      #Remove NYFCC
fit12 <- update(fit11, .~. - NYFCC)
drop1(fit12,test = "LRT")      #Remove Adventure
fit13 <- update(fit12, .~. - Adventure)
drop1(fit13,test = "LRT")      #Remove For
fit14 <- update(fit13, .~. - For)
drop1(fit14,test = "LRT")      #Remove Gm2
fit15 <- update(fit14, .~. - Gm2)
drop1(fit15,test = "LRT")      #Remove Sport
fit16 <- update(fit15, .~. - Sport)
drop1(fit16,test = "LRT")      #Remove Afs
fit17 <- update(fit16, .~. - Afs)
drop1(fit17,test = "LRT")      #Remove Horror
fit18 <- update(fit17, .~. - Horror)
drop1(fit18,test = "LRT")      #Remove U
fit19 <- update(fit18, .~. - U)
drop1(fit19,test = "LRT")      #Remove Cin
fit20 <- update(fit19, .~. - Cin)
drop1(fit20,test = "LRT")      #Remove Art
fit21 <- update(fit20, .~. - Art)
drop1(fit21,test = "LRT")      #Remove Comedy
fit22 <- update(fit21, .~. - Comedy)
drop1(fit22,test = "LRT")      #Remove Eff
fit23 <- update(fit22, .~. - Eff)
drop1(fit23,test = "LRT")      #Remove Film.noir
fit24 <- update(fit23, .~. - Film.noir)
drop1(fit24,test = "LRT")      #Remove Ebert
fit25 <- update(fit24, .~. - Ebert)
drop1(fit25,test = "LRT")      #Remove History
fit26 <- update(fit25, .~. - History)
drop1(fit26,test = "LRT")      #Remove Music
fit27 <- update(fit26, .~. - Music)
drop1(fit27,test = "LRT")      #Remove Biography
```

```
fit28 <- update(fit27, .~. - Biography)
drop1(fit28,test = "LRT")       #Remove Gf2
fit29 <- update(fit28, .~. - Gf2)
drop1(fit29,test = "LRT")       #Remove Afl
fit30 <- update(fit29, .~. - Afl)
drop1(fit30,test = "LRT")       #Remove Scr
fit31 <- update(fit30, .~. - Scr)
drop1(fit31,test = "LRT")       #Remove Sco
fit32 <- update(fit31, .~. - Sco)
drop1(fit32,test = "LRT")       #Remove Gfl
fit33 <- update(fit32, .~. - Gfl)
drop1(fit33,test = "LRT")       #Remove Cos
fit34 <- update(fit33, .~. - Cos)
drop1(fit34,test = "LRT")       #Remove Musical
fit35 <- update(fit34, .~. - Musical)
drop1(fit35,test = "LRT")       #Remove Gf1
fit36 <- update(fit35, .~. - Gf1)
drop1(fit36,test = "LRT")       #Remove Mystery
fit37 <- update(fit36, .~. - Mystery)
drop1(fit37,test = "LRT")       #Remove Thriller
fit38 <- update(fit37, .~. - Thriller)
drop1(fit38,test = "LRT")       #Remove LAFCA
fit39 <- update(fit38, .~. - LAFCA)
drop1(fit39,test = "LRT")       #Remove Gm1
fit40 <- update(fit39, .~. - Gm1)
drop1(fit40,test = "LRT")       #Remove G
fit41 <- update(fit40, .~. - G)
drop1(fit41,test = "LRT")       #Remove Sou
fit42 <- update(fit41, .~. - Sou)
drop1(fit42,test = "LRT")       #Remove Crime
fit43 <- update(fit42, .~. - Crime)
drop1(fit43,test = "LRT")       #Remove Family
fit44 <- update(fit43, .~. - Family)
drop1(fit44,test = "LRT")       #Remove Fantasy
fit45 <- update(fit44, .~. - Fantasy)
drop1(fit45,test = "LRT")       #Remove AD
fit46 <- update(fit45, .~. - AD)
drop1(fit46,test = "LRT")       #Remove War
fit47 <- update(fit46, .~. - War)
drop1(fit47,test = "LRT")       #Remove Drama
fit48 <- update(fit47, .~. - Drama)
drop1(fit48,test = "LRT")       #Remove Action
fit49 <- update(fit48, .~. - Action)
drop1(fit49,test = "LRT")       #Remove Mak
fit50 <- update(fit49, .~. - Mak)
drop1(fit50,test = "LRT")       #Remove Ams
fit51 <- update(fit50, .~. - Ams)
drop1(fit51,test = "LRT")       #Remove Romance
fit52 <- update(fit51, .~. - Romance)
drop1(fit52,test = "LRT")       #Remove WR
fit53 <- update(fit52, .~. - WR)
drop1(fit53,test = "LRT")       #Remove Days
fit54 <- update(fit53, .~. - Days)
drop1(fit54,test = "LRT")       #Remove DGA
fit55 <- update(fit54, .~. - DGA)
drop1(fit55,test = "LRT")       #Remove Gd
```

fit56 <- update(fit55, .~. - Gd)
drop1(fit56,test = "LRT")   #Remove SciFi
fit57 <- update(fit56, .~. - SciFi)
drop1(fit57,test = "LRT")   #We stop here because now we have only significant #variables

```
> fit54 <- update(fit53, .~. - Days)
> drop1(fit54,test = "LRT")        #Remove DGA
Single term deletions

Model:
Ch ~ Dir + Edi + Dan + Gdr + Gd + PGA + DGA + SciFi + PG + PG13 +
    R + NSFC
       Df Deviance    AIC    LRT  Pr(>Chi)    
<none>      289.63 315.63                     
Dir     1   304.82 328.82 15.189 9.725e-05 ***
Edi     1   300.99 324.99 11.359  0.000751 ***
Dan     1   294.77 318.77  5.143  0.023340 *  
Gdr     1   296.79 320.79  7.158  0.007464 ** 
Gd      1   293.27 317.27  3.634  0.056594 .  
PGA     1   386.03 410.03 96.401 < 2.2e-16 ***
DGA     1   293.06 317.06  3.428  0.064081 .  
SciFi   1   293.54 317.54  3.907  0.048072 *  
PG      1   293.69 317.69  4.056  0.044022 *  
PG13    1   293.74 317.74  4.104  0.042774 *  
R       1   297.26 321.26  7.625  0.005755 ** 
NSFC    1   296.94 320.94  7.307  0.006869 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> fit55 <- update(fit54, .~. - DGA)
> drop1(fit55,test = "LRT")        #Remove Gd
Single term deletions

Model:
Ch ~ Dir + Edi + Dan + Gdr + Gd + PGA + SciFi + PG + PG13 + R +
    NSFC
       Df Deviance    AIC     LRT  Pr(>Chi)    
<none>      293.06 317.06                      
Dir     1   308.97 330.97  15.912 6.636e-05 ***
Edi     1   304.60 326.60  11.540 0.0006813 ***
Dan     1   298.14 320.14   5.079 0.0242191 *  
Gdr     1   300.17 322.17   7.107 0.0076804 ** 
Gd      1   294.16 316.16   1.096 0.2951492    
PGA     1   394.46 416.46 101.400 < 2.2e-16 ***
SciFi   1   296.04 318.04   2.982 0.0841739 .  
PG      1   297.55 319.55   4.490 0.0340866 *  
PG13    1   297.63 319.63   4.574 0.0324598 *  
R       1   299.98 321.98   6.922 0.0085135 ** 
NSFC    1   300.68 322.68   7.624 0.0057585 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> fit56 <- update(fit55, .~. - Gd)
> drop1(fit56,test = "LRT")        #Remove SciFi
Single term deletions

Model:
Ch ~ Dir + Edi + Dan + Gdr + PGA + SciFi + PG + PG13 + R + NSFC
       Df Deviance    AIC     LRT  Pr(>Chi)    
<none>      294.16 316.16                      
Dir     1   310.04 330.04  15.885 6.731e-05 ***
Edi     1   306.35 326.35  12.191 0.0004801 ***
Dan     1   299.16 319.16   5.003 0.0252997 *  
Gdr     1   300.37 320.37   6.210 0.0127030 *  
PGA     1   395.90 415.90 101.747 < 2.2e-16 ***
SciFi   1   297.55 317.55   3.393 0.0654886 .  
PG      1   298.58 318.58   4.423 0.0354502 *  
PG13    1   299.05 319.05   4.892 0.0269835 *  
R       1   301.64 321.64   7.480 0.0062379 ** 
NSFC    1   300.93 320.93   6.772 0.0092599 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> fit57 <- update(fit56, .~. - SciFi)
> drop1(fit57,test = "LRT")        #We stop here because now we have only significant variables
Single term deletions

Model:
Ch ~ Dir + Edi + Dan + Gdr + PGA + PG + PG13 + R + NSFC
       Df Deviance    AIC     LRT  Pr(>Chi)    
<none>      297.55 317.55                      
Dir     1   312.61 330.61  15.058 0.0001043 ***
Edi     1   308.50 326.50  10.949 0.0009367 ***
Dan     1   302.69 320.69   5.139 0.0233919 *  
Gdr     1   304.81 322.81   7.258 0.0070594 ** 
PGA     1   397.82 415.82 100.268 < 2.2e-16 ***
PG      1   302.10 320.10   4.549 0.0329368 *  
PG13    1   303.88 321.88   6.335 0.0118370 *  
R       1   305.22 323.22   7.675 0.0056006 ** 
NSFC    1   305.04 323.04   7.490 0.0062045 ** 
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> 
```

Appendix VI

InterceptModel <- glm(Ch~1 , family = binomial, data = OscarsSubset)
scope = formula(ModelFull)
AICModelForwardsStep <- step(InterceptModel, scope = scope, direction = 'forward', k = 2)

```
Step:  AIC=317.83
Ch ~ PGA + Dir + Edi + Dan + Gdr + SciFi + Days + NSFC

            Df Deviance    AIC
+ Ebert      1   296.39 316.39
+ Romance    1   296.49 316.49
+ Nom        1   297.00 317.00
+ R          1   297.36 317.36
+ Ams        1   297.51 317.51
+ Drama      1   297.52 317.52
<none>           299.83 317.83
+ Gd         1   297.93 317.93
+ LAFCA      1   298.48 318.48
+ WR         1   298.51 318.51
+ Musical    1   298.62 318.62
+ Gm1        1   298.63 318.63
+ Scr        1   298.71 318.71
+ Thriller   1   298.85 318.85
+ Horror     1   298.87 318.87
+ Family     1   298.94 318.94
+ Comedy     1   298.95 318.95
+ Action     1   299.02 319.02
+ PG         1   299.07 319.07
+ Sco        1   299.21 319.21
+ Aml        1   299.21 319.21
+ PG13       1   299.24 319.24
+ Gf2        1   299.27 319.28
+ Afs        1   299.28 319.28
+ War        1   299.29 319.29
+ DGA        1   299.31 319.31
+ Mak        1   299.31 319.31
+ AD         1   299.39 319.39
+ Western    1   299.41 319.41
+ Art        1   299.49 319.49
+ Gf1        1   299.50 319.50
+ NYFCC      1   299.51 319.51
+ Gmc        1   299.53 319.53
+ Sport      1   299.55 319.55
+ Eff        1   299.56 319.56
+ G          1   299.56 319.56
+ Fantasy    1   299.56 319.56
+ Cos        1   299.59 319.59
+ For        1   299.60 319.60
+ Length     1   299.61 319.61
+ U          1   299.62 319.62
+ History    1   299.66 319.66
+ Biography  1   299.70 319.70
+ Afl        1   299.70 319.70
+ Film.noir  1   299.73 319.73
+ Music      1   299.74 319.74
+ Mystery    1   299.75 319.75
+ Docu       1   299.76 319.76
+ Crime      1   299.77 319.77
+ Animation  1   299.77 319.77
+ Son        1   299.77 319.77
+ Adventure  1   299.78 319.78
+ Sou        1   299.78 319.78
+ Anf        1   299.79 319.79
+ Cin        1   299.81 319.81
+ Gm2        1   299.82 319.82
+ NBR        1   299.82 319.82

Step:  AIC=316.39
Ch ~ PGA + Dir + Edi + Dan + Gdr + SciFi + Days + NSFC + Ebert

            Df Deviance    AIC
+ Nom        1   293.94 315.94
+ Romance    1   294.03 316.03
+ Ams        1   294.17 316.17
+ WR         1   294.19 316.19
<none>           296.39 316.39
+ Gm1        1   294.82 316.82
+ Drama      1   294.82 316.82
+ Gd         1   295.03 317.03
+ Mak        1   295.10 317.10
+ Scr        1   295.24 317.24
+ DGA        1   295.26 317.26
+ Musical    1   295.42 317.42
+ Comedy     1   295.49 317.49
+ PG         1   295.68 317.68
+ Thriller   1   295.68 317.68
+ Action     1   295.71 317.71
+ LAFCA      1   295.72 317.72
+ Horror     1   295.73 317.73
+ R          1   295.75 317.75
+ Gf2        1   295.85 317.85
+ Eff        1   295.89 317.89
+ Gmc        1   295.91 317.91
```

```
 + Family     1   295.93 317.93
 + Western    1   295.97 317.97
 + War        1   295.97 317.97
 + Afs        1   295.98 317.98
 + Sco        1   295.98 317.98
 + Cos        1   296.03 318.03
 + Length     1   296.06 318.06
 + Fantasy    1   296.08 318.08
 + G          1   296.10 318.10
 + AD         1   296.14 318.14
 + For        1   296.14 318.14
 + Aml        1   296.15 318.15
 + Crime      1   296.15 318.16
 + NYFCC      1   296.21 318.21
 + Gf1        1   296.22 318.22
 + Sport      1   296.23 318.23
 + Music      1   296.24 318.24
 + U          1   296.25 318.25
 + Son        1   296.26 318.26
 + Art        1   296.26 318.26
 + Biography  1   296.27 318.27
 + PG13       1   296.30 318.30
 + History    1   296.30 318.30
 + Sou        1   296.30 318.30
 + Docu       1   296.33 318.33
+ Mystery     1   296.35 318.35
+ Animation   1   296.36 318.36
+ Afl         1   296.36 318.36
+ NBR         1   296.36 318.36
+ Anf         1   296.37 318.36
+ Adventure   1   296.37 318.37
+ Cin         1   296.38 318.38
+ Gm2         1   296.39 318.39
+ Film.noir   1   296.39 318.39

Step:  AIC=315.94
Ch ~ PGA + Dir + Edi + Dan + Gdr + SciFi + Days + NSFC + Ebert +
    Nom

            Df Deviance    AIC
<none>          293.94 315.94
+ Romance    1   291.96 315.96
+ WR         1   291.97 315.97
+ Gd         1   292.21 316.21
+ Drama      1   292.36 316.36
+ Gm1        1   292.62 316.62
+ Ams        1   292.67 316.67
+ Sou        1   292.79 316.79
+ Comedy     1   292.90 316.90
+ PG         1   292.95 316.95
+ Cin        1   293.19 317.19
+ Scr        1   293.22 317.22
+ LAFCA      1   293.26 317.26
+ Action     1   293.31 317.31
+ DGA        1   293.38 317.38
+ Mak        1   293.39 317.39
+ Musical    1   293.40 317.40
+ Gf2        1   293.43 317.43
+ Fantasy    1   293.44 317.43
+ Western    1   293.44 317.44
+ Horror     1   293.50 317.51
+ G          1   293.52 317.52
+ Family     1   293.52 317.52
+ Thriller   1   293.55 317.54
+ R          1   293.56 317.56
+ War        1   293.58 317.58
+ Art        1   293.63 317.63
+ Crime      1   293.66 317.66
+ Gmc        1   293.68 317.68
+ Music      1   293.69 317.69
+ AD         1   293.71 317.71
+ PG13       1   293.75 317.75
+ NYFCC      1   293.76 317.76
+ U          1   293.79 317.79
+ Sport      1   293.79 317.79
+ Gf1        1   293.81 317.81
 + Biography  1   293.81 317.81
 + For        1   293.83 317.83
 + Mystery    1   293.85 317.85
 + Eff        1   293.86 317.86
 + Docu       1   293.89 317.89
 + History    1   293.89 317.89
 + Afl        1   293.90 317.90
 + Animation  1   293.90 317.90
 + Anf        1   293.91 317.91
 + Length     1   293.92 317.92
 + Aml        1   293.92 317.92
 + Film.noir  1   293.93 317.93
 + Cos        1   293.93 317.93
 + Gm2        1   293.94 317.94
 + Afs        1   293.94 317.94
 + Sco        1   293.94 317.94
 + Son        1   293.94 317.94
 + Adventure  1   293.94 317.94
 + NBR        1   293.94 317.94
```

Appendix VII

BICModelForwardsStep <- step(InterceptModel, scope = scope, direction = 'forward', k = log(n))

```
Step:  AIC=357.95
Ch ~ PGA + Dir

             Df Deviance    AIC
+ Edi         1   324.47 350.08
+ Gdr         1   329.99 355.60
+ Nom         1   330.58 356.20
+ Dan         1   332.03 357.64
<none>            338.74 357.95
+ SciFi       1   334.37 359.99
+ Musical     1   334.67 360.28
+ NSFC        1   334.86 360.47
+ Ams         1   334.89 360.50
+ Length      1   335.38 361.00
+ Gm1         1   335.76 361.38
+ NYFCC       1   335.90 361.52
+ WR          1   336.13 361.74
+ Scr         1   336.49 362.11
+ Romance     1   336.50 362.11
+ Days        1   336.56 362.18
+ Ebert       1   336.59 362.21
+ Sco         1   336.62 362.23
+ Family      1   337.07 362.69
+ War         1   337.22 362.83
+ Art         1   337.24 362.86
+ Aml         1   337.29 362.90
+ Mak         1   337.35 362.96
+ PG13        1   337.38 363.00
+ Fantasy     1   337.49 363.11
+ DGA         1   337.54 363.15
+ PG          1   337.74 363.35
+ R           1   337.90 363.51
+ Gf2         1   337.93 363.55
+ Cos         1   337.98 363.60
+ History     1   338.04 363.66
+ Crime       1   338.12 363.74
+ Cin         1   338.16 363.78
+ Sou         1   338.18 363.80
+ Action      1   338.19 363.80
+ U           1   338.19 363.81
+ NBR         1   338.22 363.83
+ Thriller    1   338.32 363.93
+ Drama       1   338.41 364.03
+ Mystery     1   338.44 364.05
+ Music       1   338.50 364.12
+ AD          1   338.50 364.12
+ Eff         1   338.55 364.16
+ Sport       1   338.57 364.18
+ Afs         1   338.57 364.18
+ Gf1         1   338.57 364.18
+ Horror      1   338.60 364.21
+ Film.noir   1   338.60 364.21
+ Gd          1   338.60 364.21
+ Gmc         1   338.60 364.22
+ Biography   1   338.61 364.22
+ Animation   1   338.65 364.26
+ G           1   338.66 364.27
+ Anf         1   338.68 364.29
+ Comedy      1   338.71 364.32
+ Docu        1   338.71 364.33
+ Western     1   338.73 364.34
+ For         1   338.73 364.35
+ Adventure   1   338.74 364.35
+ LAFCA       1   338.74 364.36
+ Afl         1   338.74 364.36
+ Gm2         1   338.74 364.36
+ Son         1   338.74 364.36
```

```
Step:  AIC=350.08
Ch ~ PGA + Dir + Edi

            Df Deviance    AIC
<none>           324.47 350.08
+ Dan        1   318.48 350.50
+ SciFi      1   318.50 350.52
+ Gdr        1   319.02 351.04
+ Ebert      1   320.14 352.16
+ Days       1   321.06 353.07
+ Romance    1   321.07 353.09
+ Musical    1   321.12 353.14
+ NSFC       1   321.54 353.56
+ Scr        1   321.61 353.63
+ Ams        1   321.84 353.85
+ NYFCC      1   321.86 353.88
+ PG13       1   322.31 354.33
+ Thriller   1   322.60 354.61
+ Fantasy    1   322.61 354.63
+ Gm1        1   322.74 354.76
+ Nom        1   322.84 354.86
+ R          1   322.92 354.94
+ Action     1   322.99 355.01
+ Drama      1   323.34 355.36
+ Family     1   323.37 355.39
+ Aml        1   323.41 355.42
+ DGA        1   323.46 355.48
+ PG         1   323.56 355.58
+ Gf2        1   323.56 355.58
+ Length     1   323.63 355.65
+ War        1   323.72 355.73
+ WR         1   323.78 355.80
+ U          1   323.91 355.92
+ Gf1        1   323.95 355.97
+ Comedy     1   323.98 356.00
+ History    1   324.00 356.02
+ Mak        1   324.00 356.02
+ AD         1   324.08 356.09
+ NBR        1   324.11 356.13
+ Horror     1   324.14 356.16
+ Afs        1   324.19 356.20
+ Art        1   324.20 356.21
+ Crime      1   324.20 356.22
+ Sport      1   324.22 356.23
+ Sco        1   324.22 356.24
+ Film.noir  1   324.31 356.33
+ Gd         1   324.31 356.33
+ Sou        1   324.32 356.33
+ G          1   324.34 356.36
+ LAFCA      1   324.37 356.39
+ Cos        1   324.37 356.39
+ Adventure  1   324.38 356.40
+ Gm2        1   324.39 356.41
+ Animation  1   324.41 356.43
+ Mystery    1   324.42 356.44
+ Son        1   324.42 356.44
+ Gmc        1   324.43 356.44
+ Music      1   324.43 356.45
+ Anf        1   324.43 356.45
+ Western    1   324.44 356.46
+ Eff        1   324.45 356.46
+ Biography  1   324.45 356.47
+ Docu       1   324.45 356.47
+ For        1   324.45 356.47
+ Cin        1   324.46 356.48
+ Afl        1   324.46 356.48
```

Appendix VIII

```
#Now we do the model with add1 function against LRT
Fit1 <- InterceptModel
add1(Fit1, scope = scope, test = 'LRT') #We add variables with the most significant
                                        #(smallest) p values, in this case it's PGA
Fit2 <- update(Fit1, .~. + PGA)
add1(Fit2, scope = scope, test = 'LRT') #Add Dir
Fit3 <- update(Fit2, .~. + Dir)
add1(Fit3, scope = scope, test = 'LRT') #Add Edi
Fit4 <- update(Fit3, .~. + Edi)
add1(Fit4, scope = scope, test = 'LRT') #Add Dan
Fit5 <- update(Fit4, .~. + Dan)
add1(Fit5, scope = scope, test = 'LRT') #Add Gdr
Fit6 <- update(Fit5, .~. + Gdr)
add1(Fit6, scope = scope, test = 'LRT') #Add SciFi
Fit7 <- update(Fit6, .~. + SciFi)
add1(Fit7, scope = scope, test = 'LRT') #Add Days
Fit8 <- update(Fit7, .~. + Days)
add1(Fit8, scope = scope, test = 'LRT') #No more significant variables to be added
```

```
> Fit7 <- update(Fit6, .~. + SciFi)
> add1(Fit7, scope = scope, test = 'LRT') #Add Days
Single term additions

Model:
Ch ~ PGA + Dir + Edi + Dan + Gdr + SciFi
          Df Deviance    AIC    LRT Pr(>Chi)
<none>         307.48 321.48
Nom        1   306.53 322.53 0.9503  0.32965
Aml        1   307.18 323.18 0.2937  0.58783
Afl        1   307.40 323.40 0.0797  0.77768
Ams        1   305.87 321.87 1.6063  0.20501
Afs        1   307.16 323.16 0.3187  0.57241
Scr        1   306.14 322.14 1.3346  0.24799
Cin        1   307.48 323.48 0.0009  0.97606
Art        1   307.24 323.23 0.2434  0.62178
Cos        1   307.43 323.43 0.0499  0.82330
Sco        1   307.37 323.37 0.1081  0.74230
Son        1   307.47 323.47 0.0051  0.94318
Sou        1   307.27 323.27 0.2037  0.65175
For        1   307.47 323.47 0.0075  0.93097
Anf        1   307.44 323.44 0.0345  0.85271
Eff        1   307.23 323.23 0.2523  0.61545
Mak        1   307.06 323.05 0.4236  0.51512
AD         1   306.87 322.87 0.6118  0.43410
Gmc        1   307.04 323.04 0.4339  0.51010
Gd         1   306.83 322.83 0.6516  0.41955
Gm1        1   305.00 321.00 2.4762  0.11558
Gm2        1   307.46 323.46 0.0195  0.88892
Gf1        1   307.47 323.47 0.0124  0.91150
Gf2        1   307.00 323.00 0.4751  0.49064
DGA        1   306.43 322.43 1.0495  0.30563
Action     1   306.54 322.54 0.9427  0.33159
Adventure  1   307.44 323.44 0.0350  0.85152
Animation  1   307.43 323.43 0.0518  0.82002
Biography  1   307.41 323.41 0.0657  0.79777
Comedy     1   306.47 322.47 1.0038  0.31639
Crime      1   307.42 323.42 0.0552  0.81422
Docu       1   307.46 323.46 0.0172  0.89561
Drama      1   304.91 320.91 2.5703  0.10889
Family     1   306.66 322.66 0.8179  0.36580
Fantasy    1   307.17 323.17 0.3090  0.57827
Film.noir  1   307.45 323.45 0.0307  0.86097
History    1   307.13 323.13 0.3458  0.55652
Horror     1   306.98 322.98 0.5002  0.47940
Music      1   307.38 323.38 0.1024  0.74901
Musical    1   306.31 322.31 1.1643  0.28058
Mystery    1   307.23 323.23 0.2529  0.61507
Romance    1   304.75 320.75 2.7282  0.09859 .
Sport      1   307.15 323.15 0.3291  0.56619
Thriller   1   306.55 322.55 0.9250  0.33617
War        1   306.83 322.83 0.6483  0.42071
Western    1   307.39 323.40 0.0827  0.77361
Length     1   307.42 323.42 0.0616  0.80405
Days       1   302.95 318.95 4.5271  0.03336 *
```

```
G            1    307.31 323.31 0.1687   0.68123
PG           1    306.68 322.68 0.8001   0.37108
PG13         1    306.37 322.37 1.1123   0.29158
R            1    305.95 321.95 1.5290   0.21626
U            1    307.28 323.28 0.1937   0.65983
Ebert        1    304.64 320.64 2.8358   0.09219 .
NYFCC        1    306.09 322.09 1.3902   0.23837
LAFCA        1    307.48 323.47 0.0035   0.95316
NSFC         1    304.21 320.21 3.2635   0.07084 .
NBR          1    307.31 323.31 0.1713   0.67892
WR           1    305.71 321.71 1.7636   0.18417
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
> Fit8 <- update(Fit7, .~. + Days)
> add1(Fit8, scope = scope, test = 'LRT') #No more significant variables to be added
Single term additions

Model:
Ch ~ PGA + Dir + Edi + Dan + Gdr + SciFi + Days
          Df Deviance    AIC      LRT Pr(>Chi)
<none>        302.95 318.95
Nom        1   300.92 318.92 2.03298   0.15392
Aml        1   302.58 320.58 0.36635   0.54500
Afl        1   302.86 320.86 0.09282   0.76062
Ams        1   300.95 318.95 2.00122   0.15717
Afs        1   302.25 320.25 0.69984   0.40284
Scr        1   301.75 319.75 1.20560   0.27221
Cin        1   302.94 320.94 0.00630   0.93673
Art        1   302.72 320.72 0.22919   0.63213
Cos        1   302.82 320.83 0.12579   0.72284
Sco        1   302.70 320.70 0.25047   0.61674
Son        1   302.92 320.92 0.02994   0.86263
Sou        1   302.87 320.87 0.07664   0.78190
For        1   302.95 320.95 0.00380   0.95087
Anf        1   302.91 320.91 0.04225   0.83714
Eff        1   302.66 320.66 0.28823   0.59136
Mak        1   302.36 320.36 0.59149   0.44184
AD         1   302.61 320.61 0.34028   0.55967
Gmc        1   302.60 320.60 0.35229   0.55282
Gd         1   302.17 320.17 0.78400   0.37592
Gm1        1   301.07 319.07 1.87672   0.17071
Gm2        1   302.94 320.94 0.01298   0.90928
Gf1        1   302.87 320.86 0.08648   0.76870
Gf2        1   302.60 320.60 0.34924   0.55454
DGA        1   302.02 320.02 0.92972   0.33494
Action     1   301.97 319.97 0.97678   0.32300
Adventure  1   302.94 320.94 0.01158   0.91432
Animation  1   302.90 320.90 0.05545   0.81383
Biography  1   302.88 320.88 0.06864   0.79333
Comedy     1   301.93 319.93 1.01958   0.31262
Crime      1   302.88 320.88 0.07595   0.78287
Docu       1   302.88 320.88 0.06647   0.79655
Drama      1   300.84 318.84 2.11297   0.14606
Family     1   302.19 320.19 0.76240   0.38258
Fantasy    1   302.77 320.77 0.17660   0.67431
Film.noir  1   302.90 320.90 0.04775   0.82703
History    1   302.46 320.46 0.49043   0.48374
Horror     1   302.23 320.23 0.71671   0.39723
Music      1   302.79 320.79 0.16014   0.68902
Musical    1   302.07 320.07 0.87849   0.34862
Mystery    1   302.76 320.76 0.19341   0.66010
Romance    1   300.56 318.56 2.38773   0.12229
Sport      1   302.51 320.51 0.44319   0.50559
Thriller   1   302.11 320.11 0.83900   0.35968
War        1   302.47 320.47 0.47998   0.48843
Western    1   302.68 320.68 0.27068   0.60287
Length     1   302.75 320.75 0.20050   0.65432
G          1   302.84 320.84 0.11442   0.73517
PG         1   302.30 320.30 0.64976   0.42020
PG13       1   302.47 320.47 0.47807   0.48930
R          1   301.62 319.62 1.32639   0.24945
U          1   302.73 320.73 0.22017   0.63891
Ebert      1   301.33 319.33 1.62225   0.20278
NYFCC      1   302.01 320.01 0.93700   0.33305
LAFCA      1   302.94 320.94 0.01228   0.91177
NSFC       1   299.83 317.83 3.12445   0.07713 .
NBR        1   302.91 320.91 0.04046   0.84059
WR         1   300.77 318.77 2.17980   0.13983
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

Appendix IX

Model1 <- AICModelBackwardsStep
Model2 <- BICModelBackwardsStep
Model3 <- fit57
Model4 <- AICModelForwardsStep
Model5 <- BICModelForwardsStep
Model6 <- Fit8

```
> Model1

Call:  glm(formula = Ch ~ Dir + Edi + Dan + Gdr + Gd + PGA + DGA + Romance + 
    SciFi + Days + PG + PG13 + R + NSFC + WR, family = binomial, 
    data = OscarsSubset)

Coefficients:
(Intercept)          Dir          Edi          Dan          Gdr           Gd          PGA  
  -9.428192     1.574330     1.169340     3.131095     1.170446    -2.476237     3.320186  
        DGA      Romance        SciFi         Days           PG         PG13            R  
   1.733160     0.545207    -2.375553     0.002247    -0.865544    -0.980035    -1.041853  
       NSFC           WR  
   1.348854     0.590864  

Degrees of Freedom: 603 Total (i.e. Null);  588 Residual
Null Deviance:      529 
Residual Deviance: 281.9        AIC: 313.9
> Model2

Call:  glm(formula = Ch ~ Dir + Edi + PGA, family = binomial, data = OscarsSubset)

Coefficients:
(Intercept)          Dir          Edi          PGA  
     -4.696        1.936        1.192        3.172  

Degrees of Freedom: 603 Total (i.e. Null);  600 Residual
Null Deviance:      529 
Residual Deviance: 324.5        AIC: 332.5
> Model3

Call:  glm(formula = Ch ~ Dir + Edi + Dan + Gdr + PGA + PG + PG13 + 
    R + NSFC, family = binomial, data = OscarsSubset)

Coefficients:
(Intercept)          Dir          Edi          Dan          Gdr          PGA           PG  
    -4.2732       1.6477       1.1344       2.8947       0.9961       3.4102      -1.0559  
       PG13            R         NSFC  
    -1.3522      -1.0656       1.4158  

Degrees of Freedom: 603 Total (i.e. Null);  594 Residual
Null Deviance:      529 
Residual Deviance: 297.5        AIC: 317.5
> Model4

Call:  glm(formula = Ch ~ PGA + Dir + Edi + Dan + Gdr + SciFi + Days + 
    NSFC + Ebert + Nom, family = binomial, data = OscarsSubset)

Coefficients:
(Intercept)          PGA          Dir          Edi          Dan          Gdr        SciFi  
  -5.618512     3.205198     1.501600     0.976435     3.096655     0.904102    -2.105131  
       Days         NSFC        Ebert          Nom  
   0.002574     1.243580    -0.162326     0.114767  

Degrees of Freedom: 603 Total (i.e. Null);  593 Residual
Null Deviance:      529 
Residual Deviance: 293.9        AIC: 315.9
> Model5

Call:  glm(formula = Ch ~ PGA + Dir + Edi, family = binomial, data = OscarsSubset)

Coefficients:
(Intercept)          PGA          Dir          Edi  
     -4.696        3.172        1.936        1.192  

Degrees of Freedom: 603 Total (i.e. Null);  600 Residual
Null Deviance:      529 
Residual Deviance: 324.5        AIC: 332.5
> Model6

Call:  glm(formula = Ch ~ PGA + Dir + Edi + Dan + Gdr + SciFi + Days, 
    family = binomial, data = OscarsSubset)

Coefficients:
(Intercept)          PGA          Dir          Edi          Dan          Gdr        SciFi  
  -5.362246     3.184004     1.853256     1.196892     3.307927     0.910085    -2.340620  
       Days  
   0.002627  

Degrees of Freedom: 603 Total (i.e. Null);  596 Residual
Null Deviance:      529 
Residual Deviance: 303  AIC: 319
```

Appendix X

#First we test the fit models against the full and null models respectively
anova(Model3,ModelFull,test="LRT") #So Model3 is more adequate than the full #model

```
> anova(Model3,ModelFull,test="LRT")
Analysis of Deviance Table

Model 1: Ch ~ Dir + Edi + Dan + Gdr + PGA + PG + PG13 + R + NSFC
Model 2: Ch ~ Nom + Dir + Aml + Afl + Ams + Afs + Scr + Cin + Art + Cos +
    Sco + Son + Edi + Sou + For + Anf + Eff + Mak + Dan + AD +
    Gdr + Gmc + Gd + Gm1 + Gm2 + Gf1 + Gf2 + PGA + DGA + Action +
    Adventure + Animation + Biography + Comedy + Crime + Docu +
    Drama + Family + Fantasy + Film.noir + History + Horror +
    Music + Musical + Mystery + Romance + SciFi + Sport + Thriller +
    War + Western + Length + Days + G + PG + PG13 + R + U + Ebert +
    NYFCC + LAFCA + NSFC + NBR + WR
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       594     297.55
2       539     258.01 55   39.541   0.9424
```

anova(Model6,InterceptModel,test="LRT") #So Model6 is more adequate

```
> anova(Model6,InterceptModel,test="LRT")
Analysis of Deviance Table

Model 1: Ch ~ PGA + Dir + Edi + Dan + Gdr + SciFi + Days
Model 2: Ch ~ 1
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       596     302.95
2       603     528.99 -7  -226.04 < 2.2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

#Note that Model 2 is nested in Model 1

anova(Model2,Model1,test="LRT") #Model1 is more adequate

```
> anova(Model2,Model1,test="LRT")
Analysis of Deviance Table

Model 1: Ch ~ Dir + Edi + PGA
Model 2: Ch ~ Dir + Edi + Dan + Gdr + Gd + PGA + DGA + Romance + SciFi +
    Days + PG + PG13 + R + NSFC + WR
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       600     324.47
2       588     281.92 12    42.55 2.691e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

#Model5 and model 4 are nested

anova(Model5,Model4,test="LRT") #Model4 is more adequate

```
> anova(Model5,Model4,test="LRT")
Analysis of Deviance Table

Model 1: Ch ~ PGA + Dir + Edi
Model 2: Ch ~ PGA + Dir + Edi + Dan + Gdr + SciFi + Days + NSFC + Ebert +
    Nom
  Resid. Df Resid. Dev Df Deviance  Pr(>Chi)
1       600     324.47
2       593     293.94  7   30.526 7.603e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1
```

#So now we only need to test between Model3, Model6, Model1 and Model4

#Model6 is nested in Model4 and Model3 is nested in Model11 anova(Model6,Model4,test="LRT") #Model4 is more adequate

```
> anova(Model6,Model4,test="LRT")
Analysis of Deviance Table

Model 1: Ch ~ PGA + Dir + Edi + Dan + Gdr + SciFi + Days
Model 2: Ch ~ PGA + Dir + Edi + Dan + Gdr + SciFi + Days + NSFC + Ebert +
    Nom
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       596     302.95
2       593     293.94  3   9.0088  0.02917 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

anova(Model3,Model1,test="LRT") #Model1 is more adequate

```
> anova(Model3,Model1,test="LRT")
Analysis of Deviance Table

Model 1: Ch ~ Dir + Edi + Dan + Gdr + PGA + PG + PG13 + R + NSFC
Model 2: Ch ~ Dir + Edi + Dan + Gdr + Gd + PGA + DGA + Romance + SciFi +
    Days + PG + PG13 + R + NSFC + WR
  Resid. Df Resid. Dev Df Deviance Pr(>Chi)
1       594     297.55
2       588     281.92  6   15.631  0.01588 *
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

c(AIC(Model1),AIC(Model4))
c(BIC(Model1),BIC(Model4))

```
> c(AIC(Model1),AIC(Model4))
[1] 313.9184 315.9423
> c(BIC(Model1),BIC(Model4))
[1] 384.3756 364.3816
```

Appendix XI

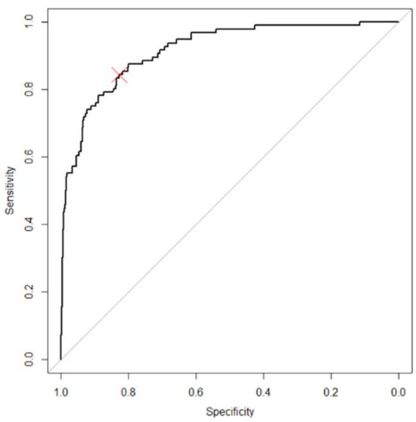
ModelFinal <- Model4
install.packages("pROC")
library(pROC)
fitted_prob <- predict(ModelFinal,type="response")
ROC <- roc(OscarsSubset$Ch,fitted_prob,plot=TRUE) #Check if ROC curve is plotting
ROC$auc

```
> ROC$auc
Area under the curve: 0.9164
```

ind <- which.min((ROC$sensitivities-1)^2+(ROC$specificities-1)^2)
ROC$thresholds[ind]

```
> ind <- which.min((ROC$sensitivities-1)^2+(ROC$specificities-1)^2)
> ROC$thresholds[ind]
[1] 0.1337347
```

ROC <- roc(OscarsSubset$Ch,fitted_prob,plot=TRUE)
points(ROC$specificities[ind],ROC$sensitivities[ind],pch=4,
cex=3,col="red") #ROC curve with optimal threshold marked



fitted_c <- fitted_prob>0.1337347
table(OscarsSubset$Ch,fitted_c)

```
           fitted_c
            FALSE TRUE
Didn't Win    421   87
Win            15   81
```

Appendix XII

Oscars2024 <- subset(oscars, ï..Year == 2024)

	ï..Year	Name	Ch	Nom	Pic	Dir	Aml	Afl	Ams	Afs	Scr	Cin
1	2024	A Complete Unkown	0	8	1	1	1	0	1	1	1	0
2	2024	Anora	0	6	1	1	0	1	1	0	1	0
3	2024	Conclave	0	8	1	0	1	0	0	1	1	0
4	2024	Dune: Part Two	0	5	1	0	0	0	0	0	0	1
5	2024	Emilia Perez	0	13	1	1	0	1	0	1	1	1
6	2024	I'm Still Here	0	3	1	0	0	1	0	0	0	0
7	2024	Nickel Boys	0	2	1	0	0	0	0	0	1	0
8	2024	The Brutalist	0	10	1	1	1	0	1	1	1	1
9	2024	The Substance	0	5	1	1	0	1	0	0	1	0
10	2024	Wicked	0	10	1	0	0	1	0	1	0	0

prediction <- predict(ModelFinal,newdata = Oscars2024, type = "response")

ScaledPrediction <- prediction/sum(prediction)

PercentagePrediction <- ScaledPrediction * 100

ProbabilityTable <- data.frame(Movie=Oscars2024$Name,
Probability=PercentagePrediction)
print(ProbabilityTable)

```
> print(ProbabilityTable)
                Movie Probability
1  A Complete Unkown   2.5580514
2              Anora  57.4818498
3           Conclave   1.6440418
4     Dune: Part Two   0.1090333
5       Emilia Perez  18.9646235
6     I'm Still Here   0.3684625
7        Nickel Boys   0.8693420
8      The Brutalist  15.4666780
9      The Substance   0.2889486
10            Wicked   2.2489690
```