Milestone-3

Prananditha M,Harshika A

2022-05-10

Milestone-3

The main aim of this project is to examine the behavior of Researchers who were successful in getting a Grantsmanship and funding for their research. we have taken into account a number of Characteristics of the researchers, especially academic grant proposal tactics, Scholarly profiles, and personality traits.

Before proceeding to apply a model to the data, we have done a few Transformations, it mainly involves factorizing the Categorical data.

Number of Proposals(Np): They are divided into three levels, NP_1(1-2 proposals per year), NP_2(3-4 proposals per year), and NP_3(5 or more proposals per year). all the levels were represented as 1,2 and 3 in the dataset which I have changed to the above terms.

Funding Agency(FA): Different funding agencies are represented as numbers from 1 to 5 which I have changed to the Agencies acronym. NSF(National Science Foundation), NIH(National Institutes of Health), DOE(Department of Energy), DOD(Department of Defense), NASA(National Aeronautics and Space Administration), and OT for the rest other Agencies.

Break Frequency(BF): Break Frequency is represented as 1 for (break every 1-2 hours or less) which I have changed to BF_1 and 2(break every 3-4 hours or longer) to BF_2.

Pilot Research(PR): These levels are taken as numbers from 1 to 5 in the dataset which I have changed to PR_1(Time dedicated to Pilot Research is less than 1 month), PR_2 (1-3 months), PR_3 (3-6 months), PR_4(6-12 months), PR_5(more than 12 months).

Time of Submission(TS): changed to TS_1 (Submission on the deadline day) which was 1 in the dataset and TS_2 (Submission earlier than the deadline) which was 2.

Deadline Stress(DS): DS_1(Similar or lesser stress than a regular day) and DS_2(more stress than a regular day). they were represented as 1 and 2 in the dataset.

Research Style(RS):RS_1(hands-off) and RS_2(hands-on) were given as 1 and 2.

Before Achieving the optimized model. we have done forward, backward, and mixed selection methods to select the optimal predictors. Based on the AIC and the number of predictors an Optimal model was chosen this process is explained in detail at the end of the report in the Appendix section of the report.

Grantsmanship Analysis:
We have used Logit models 2 and 3 for this analysis. The data is divided into two categories First,**Successful Grantsmanship(S^{G30})**, in which a more permissive definition of grantsmanship is employed. Researchers who estimated their success rate to be equal to or greater than 30% belong to the most successful class ${S_1}^{G30}$ and the rest belong to the least successful class ${S_0}^{G30}$. Second,**Highly Successful Grantsmanship(S^{G50})**, where a stricter success definition is considered that is, respondents who estimate their success rate to be lower than 50% belong to the ${S_0}^{G50}$ class and rest in ${S_1}^{G50}$.

All of the Interpretation that was done is by taking a reference researcher who belongs to the ${S_1}^{G30}$ class who submits 1-2 Proposals per year(NP_1), mostly to NSF and experiences no higher stress than any regular working day. These values are taken using baseline and mean values of predictors.

Comparison Between ${S_1}^{SG30}$ and ${S_0}^{SG30}$
Looking at Figure-2a Research Tactics plot, we can see that as the number of Proposals increases the faculty's ability to belong to the most Successful Group Decreases. The probability of group Np2 and Group NP3 belonging to the ${S_1}^{SG30}$ Group drops by 25.4% and 42.9% respectively With respect to the reference researcher RR^{G30} with a p-value of 0.001. This might be happening because the more and more Proposals one come-up with every year, there might be a decrease in the quality of the proposal which will reduce your Probability of successful Grantsmanship. Furthermore, The Funding Agencies NIH, DOE, and NASA are around the same probability as the RR^{G30} but OT and DOD have a 28.9% and 26.4% higher probability than the Reference Researcher to belong to the most successful group.

if we observe Figure 2b, The Faculty who has an h-index greater than one standard deviation above the mean the probability of the researcher belonging to ${S_1}^{SG30}$ increases by 7.9%.So, we can say that the Popularity of the faculty will also play a major role in Successful Grantsmanship. An opposite trend is followed for Trait Anxiety where there will be a probability drop of 7.8% with respect to RR^{G30} of the faculty who has a Trait Anxiety greater than one Standard deviation above the mean which can be seen in Figure 2c.

Lastly, Faculty who experience higher stress than on the regular days(DS_2) on the day of the deadline have a 17.5% lower probability of belonging to ${S_1}^{SG30}$ with a higher significance of $p<0.01$.

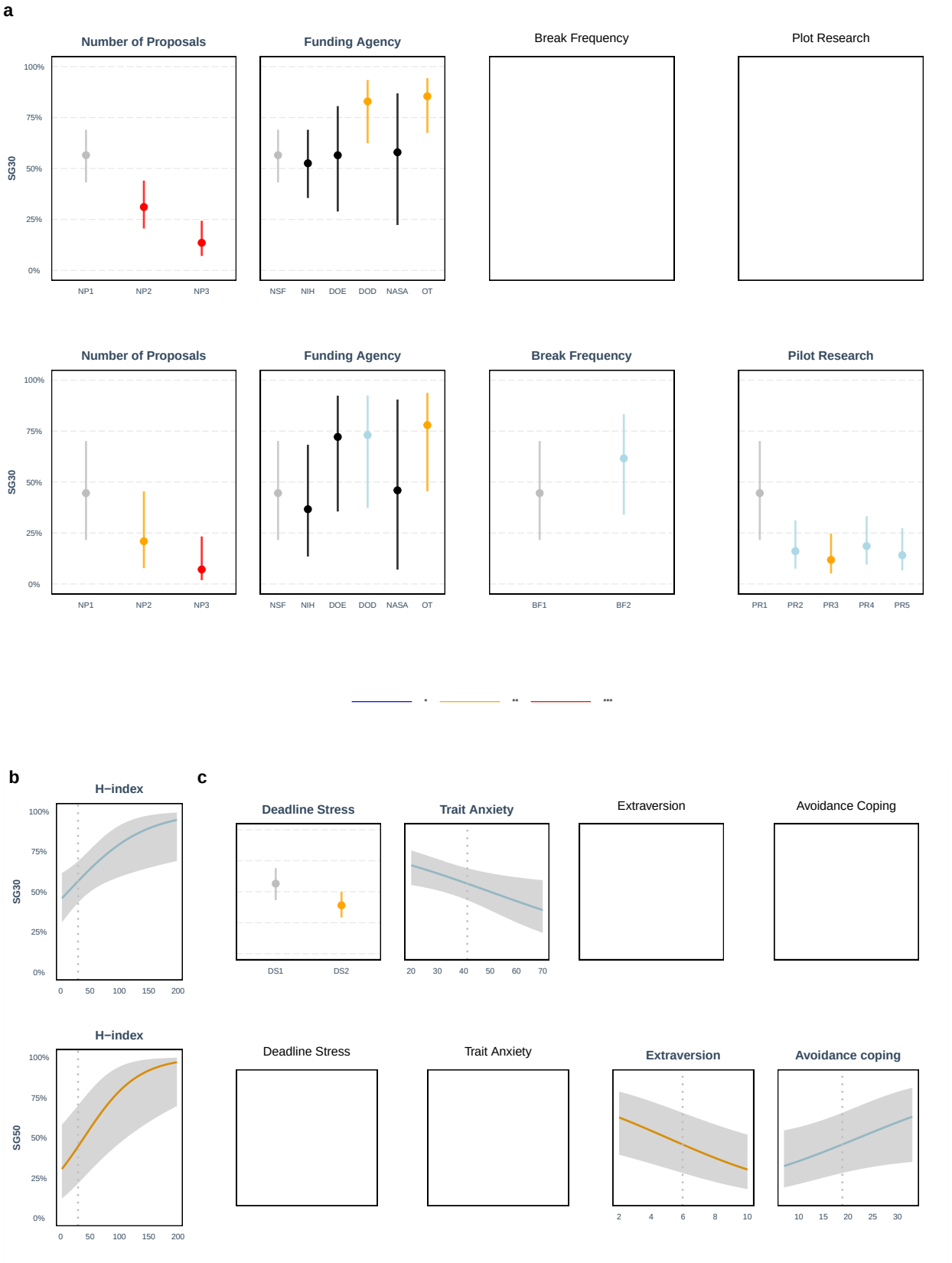
both from trait anxiety and Deadline Stress Observation, we can say that stressing will not help in Grantsmanship.

Comparison Between ${S_1}^{G50}$ and ${S_0}^{G50}$
The Reference Researcher for this group RR^{G50} has the same traits as RR^{G30} but RR^{G50} has additional traits such as Extraversion (6.0) and Trait Anxiety(18.8).

The same trend as ${S_1}^{G30}$ follows here, with the increase in the number of Proposals, the probability of the faculty belonging to the most successful class ${S_1}^{G50}$ Decreases. With respect to the RR^{G50}, NP_2 group has 25.4% and NP_3 has a 42.7% Probability drop for belonging to the most successful group. The Funding Agencies DOD and OT Groups have a 26.4% and 31.0% higher probability than the RR^{G50} to belong to ${S_1}^{G50}$. The rest Groups have the similar Probability as RR^{G50}.

Next, The faculty who doesn't take frequent breaks while working(BF_2) has an 18.1% higher probability to belong to ${S_1}^{G50}$.this might be because frequent breaks might bring in more distraction. Contrary to what we believe, Researcher who had performed a high amount of pilot research has less probability to belong to the most Successful Group with $p<0.05$. The worst case is PR_3 with a 36.4% probability drop, which might mean that 3-4 months of work on new Research is not a good enough amount of time for a successful Grantsmanship.

Figure-2


a
Number of Proposals
Funding Agency
Break Frequency
Plot Research
SG30
100%
75%
50%
25%
0%
NP1
NP2
NP3
NSF
NIH
DOE
DOD
NASA
OT
BF1
BF2
PR1
PR2
PR3
PR4
PR5
Pilot Research
*
**
***
b
H−index
SG50
0
50
100
150
200
c
Deadline Stress
Trait Anxiety
Extraversion
Avoidance Coping
DS1
DS2
20
30
40
50
60
70
2
4
6
8
10
15
20
25

Furthermore, The faulty with an h-index greater than one standard deviation above the mean has a 10.2% higher probability of belonging to the most Successful Group. and Lastly, as Expected, Extraverion and Avoidance coping have an opposite trend. where Faulty with Extraverion score with one Standard deviation above the mean has 13.1% less probability and Avoidance coping Score with one Standard deviation above mean has 9.1% higher probability of belonging to $S_1{}^{G50}$.

Grant Funding Analysis: We have used Logit models 4 and 5 for this analysis. By using Different Definition of Success we have divided the data into two groups, First, **Well-Funded Research Operations($S^{\$75}$)** where a faculty belong to the most successful class $\mathbf{S_1{}^{\$75}}$ if he/she estimates their grant funding to be 75% or greater of their needs and the rest belong to $\mathbf{S_0{}^{\$75}}$. Second, **Fully-funded Research Operations($S^{\$\$}$)** where a faculty belong to the most successful group $\mathbf{S_1{}^{\$\$}}$ if his/her research cost is Funded fully else they belong to $\mathbf{S_0{}^{\$\$}}$

The Reference Researcher $RR^{\$75}$, in this case, is similar to the above but with 42% of the time in a week is dedicated to the research with an openness score of 7.5

Comparison of $\mathbf{S_1{}^{\$75}}$ and $\mathbf{S_0{}^{\$75}}$
if we look at figure 3a, The Funding agencies NIH, NASA, and OT have a similar probability of Reference Researcher $RR^{\$75}$ of belonging to the $S_1{}^{F75}$ Group. and The DOD and DOE groups have a 28.0% and 28.6% higher probability of belonging to $S_1{}^{\$75}$. The faculty who dedicate one standard deviation above the meantime in a week for Research have a higher probability of 8.3% belonging to the $S_1{}^{\$75}$.So, the more amount of work dedicated to research, the more quality work is produced in turn increasing the successful funding.

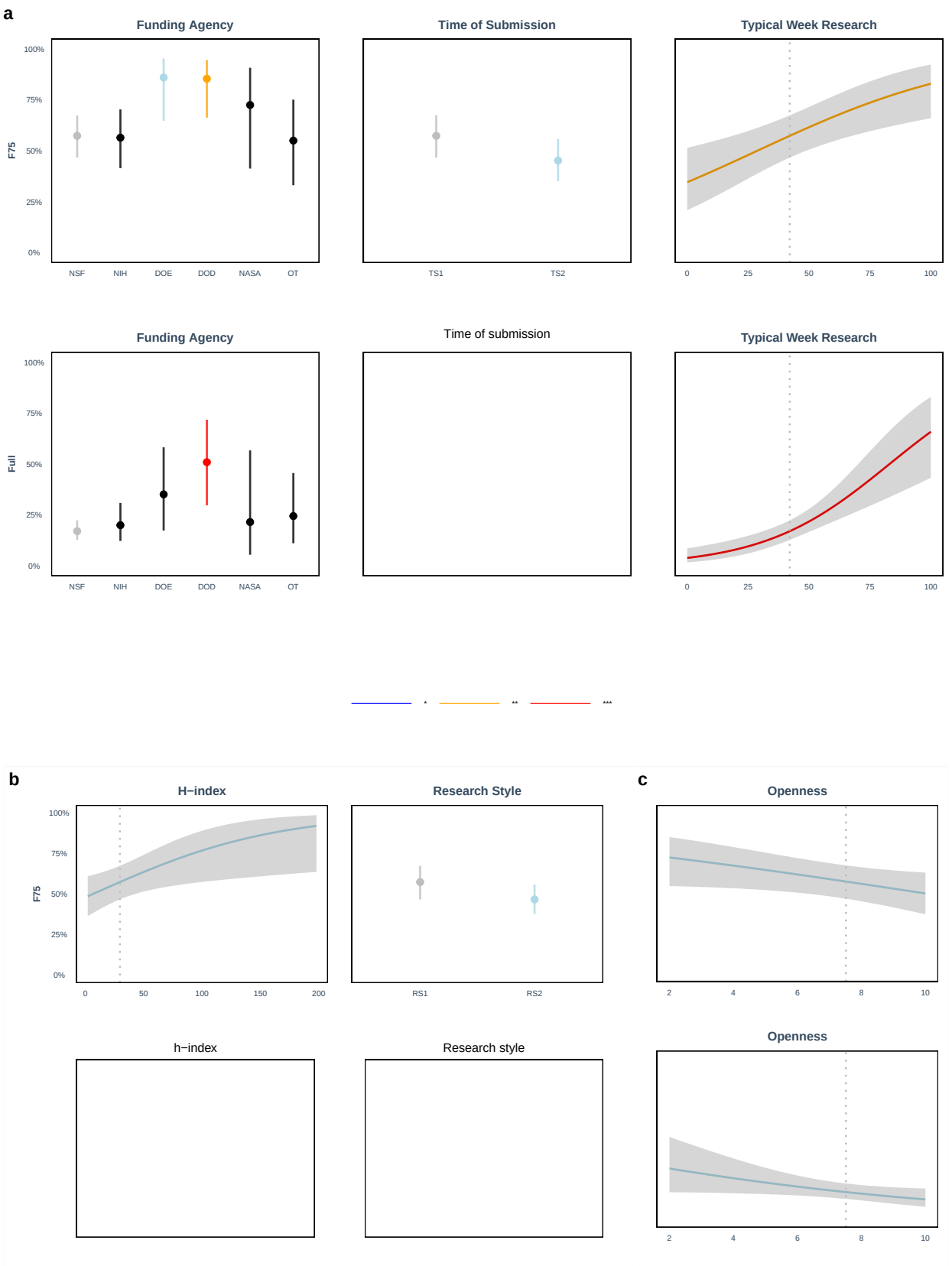
Furthermore, the faculty who submit the proposal before the deadline has a 12.2% less probability to belong to $S_1{}^{\$75}$. This might be the case because they might have produced less quality work by finishing the proposal early and not unitizing the extra amount of time for refining and going over the proposal again to make it better.

h-index in every case has shown the same positive impact. here, faculty with an h-index higher than one standard deviation above the mean has a 6.5% higher probability with respect to $RR^{\$75}$ of belonging to the $S_1{}^{\$75}$ group. Contrary to popular belief, adopting a hands-on Research style(RS_2) has shown a negative impact on the Funding. RS_2 group has a probability 10.8% lower than $RR^{\$75}$ to belong to the $S_1{}^{\$75}$ group. this can be seen in figure 3b.

Lastly, the Openness score has a negative impact on the funding. the faculty who has an openness score one standard deviation above the mean has a 5.3% less probability of belonging to the $S_1{}^{\$75}$ group.

Comparison between $\mathbf{S_1{}^{\$\$}}$ and $\mathbf{S_0{}^{\$\$}}$
The reference Researcher($RR^{\$\$}$) here is similar to $RR^{\$75}$. The Funding agencies Group DOD has a higher probability than $RR^{\$\$}$ to belong to the most successful Group. Coming to Typical Week of Research, the faculty who dedicated time that falls in one SD above the mean has a higher probability of 10% to belong to the $S_1{}^{\$\$}$ Group. Just as in the case of $S^{\$75}$, the Openness score has a negative impact on the Funding. The group whose score falls in one SD above the mean Openness score has a less probability than the $RR^{\$\$}$ to belong to the most successful group.

Figure-3



Combined Analysis(Grantsmanship and Grant Funding) Here, we combine the above two analyses.

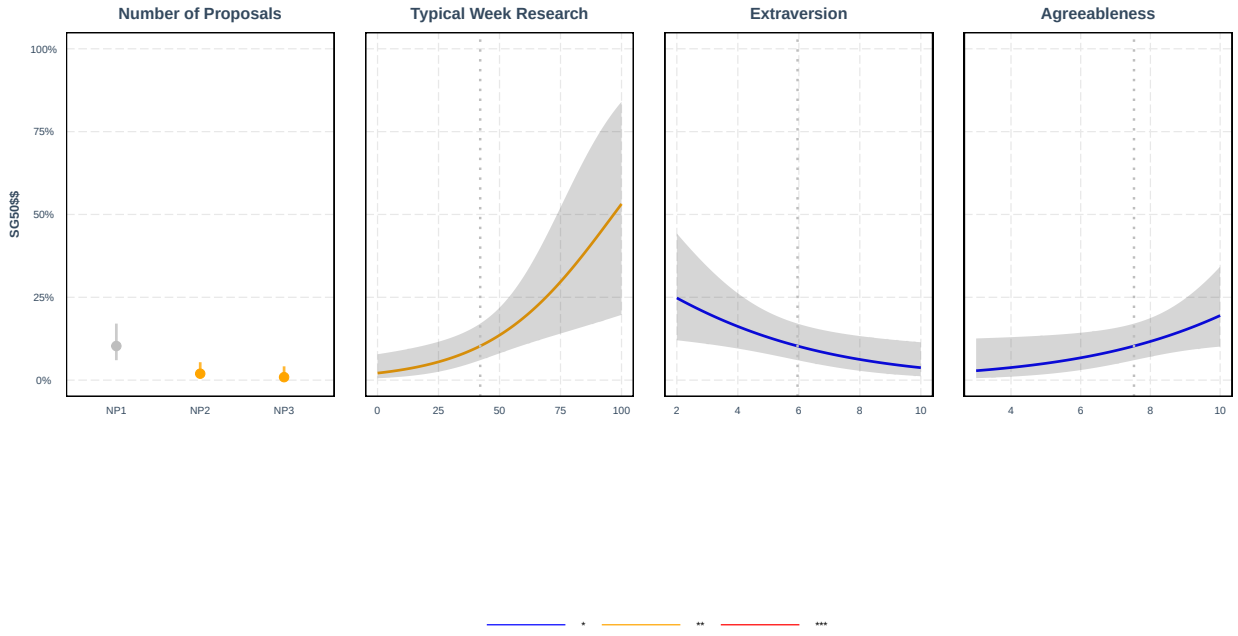
Logit model 6 is used for this analysis. The Response variable is formed by considering faculty with Grantsmanship above 50% and fully funded to Group ($S_1{}^{G50\$\$}$) and the rest to Group ($S_0{}^{G50\$\$}$).

Reference Researcher here ($RR^{G50\$\$}$), has an Agreeable score of 7.5 and an extraversion score of 6.0 and submits 1-2 proposals a year and devotes 42% of the time in a week to Research.

If we see in the below figure-4, Follows a similar trend which we have seen in Grantsmanship for the Number of proposals. The Probability of a faculty decreases with an increase in the submission of a number of proposals per year. i.e, of NP_2 there is a fall of 8.4% and for NP_3 it is 9.4% with respect to $RR^{G50\$\$}$.Next, For a faculty with a typical week of Research on SD above the mean has a 7% higher probability of belonging to The most Successful Group.

Finally, Extraversion and Agreeableness scores follow an opposite trend. The faculty with more Extroversion score has less probability of belonging to $S_1{}^{G50\$\$}$ and the opposite is true of Agreeableness.

Figure-4



Appendix Before arriving at the Optimized model, we have done backward, forward, and step-wise selection methods to remove the predictors which are not useful for the prediction of the appropriate response variable.

```
Logit(RV)~WH+BF+NP+FA+AP+AR+DWH+T+TWR+DWR+Rank+RS+H+DS+NASA+TA+E+A+C+N+O+AC+EC+TC
```

The above equation is the full model(logit 1) which contains both categorical and continuous variables.

Logit-2:(S^{G30}) Starting with the above equation, we have applied backward selection on it which gave me the following model as the final result after removing features based on AIC.it has an AIC score of 424.93.

```
##
## Call:
## glm(formula = SR ~ NP + FA + DWH + T + RS + H + DS + TA, family = "binomial",
##     data = SG30)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9666  -0.7499  -0.4682   0.8009   2.4752
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  3.274779   1.042061   3.143  0.00167 **
## NPNP2       -1.029792   0.276937  -3.719  0.00020 ***
## NPNP3       -2.043547   0.369292  -5.534 3.14e-08 ***
## FADOE       -1.470377   0.757056  -1.942  0.05211 .
## FANASA      -1.303464   0.911606  -1.430  0.15276
## FANIH       -1.628428   0.570952  -2.852  0.00434 **
## FANSF       -1.385549   0.507181  -2.732  0.00630 **
## FAOT        -0.043459   0.680704  -0.064  0.94909
## DWH         -0.578448   0.380622  -1.520  0.12857
## TTS2         0.407418   0.265909   1.532  0.12548
## RSRS2       -0.402941   0.255856  -1.575  0.11528
## H            0.015209   0.006142   2.476  0.01329 *
## DSDS2       -0.448631   0.297825  -1.506  0.13198
## TA          -0.027886   0.013203  -2.112  0.03467 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 495.86  on 402  degrees of freedom
## Residual deviance: 396.93  on 389  degrees of freedom
## AIC: 424.93
##
## Number of Fisher Scoring iterations: 4
```

we have observed that all the models gave the same result and with similar AIC, so, the result was only given in the report once. we can see that there are few predictors in the results which are not that significant for the response variable so we went on removing them one by one while checking the AIC.

The below is the result for the model after removing "DWH" (**SR ~ NP + DS + H + FA + TA + T + RS**) it has an AIC of 425.24, although it is higher than the previous one, it is just by one unit. a one-unit change in the AIC will is not a very big difference compared to the benefit of having less number of Predictors in the model.

```
##
## Call:
## glm(formula = SR ~ NP + DS + H + FA + TA + T + RS, family = "binomial",
##     data = SG30)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0572  -0.7691  -0.4682   0.8388   2.4895
##
```

```
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.93684    0.67729   1.383 0.166598
## NPNP2       -1.04986    0.27585  -3.806 0.000141 ***
## NPNP3       -2.12058    0.36751  -5.770 7.92e-09 ***
## DSDS2       -0.61511    0.27508  -2.236 0.025343 *
## H            0.01515    0.00608   2.491 0.012734 *
## FADOD        1.38871    0.50896   2.729 0.006361 **
## FADOE       -0.07253    0.58991  -0.123 0.902146
## FANASA       0.04998    0.79877   0.063 0.950108
## FANIH       -0.22773    0.33213  -0.686 0.492930
## FAOT         1.41546    0.50449   2.806 0.005020 **
## TA          -0.02804    0.01314  -2.134 0.032813 *
## TTS2         0.42086    0.26429   1.592 0.111288
## RSRS2       -0.40529    0.25482  -1.591 0.111716
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 495.86  on 402  degrees of freedom
## Residual deviance: 399.24  on 390  degrees of freedom
## AIC: 425.24
##
## Number of Fisher Scoring iterations: 5
```

We can still see that T and RS are not that significant for the model so, we have removed T(**SR** ~ **NP** + **DS** + **H** + **FA** + **TA** + **RS**) first. below is the result, it has an AIC of AIC: 425.79, still not a significant difference.

```
##
## Call:
## glm(formula = SR ~ NP + DS + H + FA + TA + RS, family = "binomial",
##     data = SG30)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9829  -0.7612  -0.4812   0.8476   2.5936
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.292478   0.635906   2.032 0.042103 *
## NPNP2       -1.063985   0.274534  -3.876 0.000106 ***
## NPNP3       -2.164759   0.366576  -5.905 3.52e-09 ***
## DSDS2       -0.729162   0.266000  -2.741 0.006121 **
## H            0.015239   0.006078   2.507 0.012163 *
## FADOD        1.315187   0.501690   2.622 0.008754 **
## FADOE       -0.065143   0.595852  -0.109 0.912943
## FANASA       0.066427   0.810197   0.082 0.934656
## FANIH       -0.186489   0.331077  -0.563 0.573245
## FAOT         1.471712   0.495070   2.973 0.002952 **
## TA          -0.028945   0.013065  -2.215 0.026734 *
## RSRS2       -0.409013   0.253807  -1.612 0.107068
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 495.86  on 402  degrees of freedom
## Residual deviance: 401.79  on 391  degrees of freedom
## AIC: 425.79
##
## Number of Fisher Scoring iterations: 5
```

The result after removing RS($\mathbf{SR} \sim \mathbf{NP + DS + H + FA + TA}$) is shown below, it has an AIC of 426.38. but there are no predictors to remove now and comparing all the models AIC and the number of predictors, we found this model to be the final model.it has 5 predictors which is less compared to all the models we have checked and has an AIC only one-unit greater than the model with least AIC which is acceptable.

```
##
## Call:
## glm(formula = SR ~ NP + DS + H + FA + TA, family = "binomial",
##     data = SG30)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -2.0417  -0.7889  -0.4860   0.8675   2.5292
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  1.061692   0.616950   1.721 0.085274 .
## NPNP2       -1.058489   0.273260  -3.874 0.000107 ***
## NPNP3       -2.116194   0.364541  -5.805 6.43e-09 ***
## DSDS2       -0.708862   0.264606  -2.679 0.007386 **
## H            0.015509   0.006084   2.549 0.010806 *
## FADOD        1.316337   0.504125   2.611 0.009024 **
## FADOE       -0.001807   0.592027  -0.003 0.997564
## FANASA       0.057120   0.788450   0.072 0.942247
## FANIH       -0.161031   0.329344  -0.489 0.624880
## FAOT         1.506769   0.490554   3.072 0.002129 **
## TA          -0.030411   0.013047  -2.331 0.019761 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 495.86  on 402  degrees of freedom
## Residual deviance: 404.38  on 392  degrees of freedom
## AIC: 426.38
##
## Number of Fisher Scoring iterations: 5
```

SO Logit-2 is $\mathbf{SR} \sim \mathbf{NP + DS + H + FA + TA}$ and below is the Parameters estimate.

	Predictor	Prob_wise	Odds_ratio	std_error	Z_val	p_val
1	(Intercept)	0.743	1.061692346	0.616949502	1.720873983	0.08527369 .
2	NPNP2	0.258	−1.058489084	0.273260471	−3.873553610	0.00010726 ***
3	NPNP3	0.108	−2.116193709	0.364541198	−5.805087930	1e−08 ***
4	DSDS2	0.330	−0.708861859	0.264606491	−2.678928459	0.00738582 **
5	H	0.504	0.015508654	0.006084454	2.548898053	0.01080639 *
6	FADOD	0.789	1.316337151	0.504124573	2.611134672	0.00902423 **
7	FADOE	0.500	−0.001807304	0.592027280	−0.003052737	0.99756427
8	FANASA	0.514	0.057120427	0.788450500	0.072446433	0.94224663
9	FANIH	0.460	−0.161031306	0.329344225	−0.488945285	0.62488043
10	FAOT	0.819	1.506769089	0.490553627	3.071568540	0.00212937 **
11	TA	0.492	−0.030411115	0.013047142	−2.330864020	0.01976053 *

Logit-3(S^{G50}) Same above Procedure is followed for this analysis too, we have started with Logit-1 equation and preceded with forward, backward and step-wise selection methods. Just as above all the selection methods have give me the exactly the same results. below is the summary of it.

```
##
## Call:
## glm(formula = SR ~ BF + NP + FA + AR + H + DS + E + AC + TC,
##     family = "binomial", data = SG50)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7456  -0.5416  -0.3385  -0.1819   3.0842
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.137361   1.624346  -0.700  0.48380
## BFBF2        0.715502   0.334340   2.140  0.03235 *
## NPNP2       -1.030998   0.353445  -2.917  0.00353 **
## NPNP3       -2.346719   0.530585  -4.423 9.74e-06 ***
## FADOE        0.031801   0.806799   0.039  0.96856
## FANASA      -1.147385   1.259285  -0.911  0.36222
## FANIH       -1.582539   0.684656  -2.311  0.02081 *
## FANSF       -1.212161   0.580089  -2.090  0.03665 *
## FAOT         0.329312   0.720980   0.457  0.64785
## ARPR2       -1.431195   0.640805  -2.233  0.02552 *
## ARPR3       -1.822639   0.662316  -2.752  0.00592 **
## ARPR4       -1.270992   0.606813  -2.095  0.03621 *
## ARPR5       -1.663969   0.623994  -2.667  0.00766 **
## H            0.018070   0.007472   2.418  0.01560 *
```

```
## DSDS2       -0.564842   0.333251  -1.695  0.09009 .
## E           -0.257018   0.081684  -3.146  0.00165 **
## AC           0.070176   0.034171   2.054  0.04001 *
## TC           0.074570   0.046684   1.597  0.11019
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 352.78  on 402  degrees of freedom
## Residual deviance: 268.31  on 385  degrees of freedom
## AIC: 304.31
##
## Number of Fisher Scoring iterations: 6
```

it has an AIC of 304.31, and there are few insignificant predictors, so i have removed then one after the other checking the AIC. Below is the summary of the model after removing TC($\mathbf{SR} \sim \mathbf{NP} + \mathbf{FA} + \mathbf{H} + \mathbf{E} + \mathbf{BF} + \mathbf{DS} + \mathbf{AC} + \mathbf{AR}$) it has got the AIC of 305.07, As mentioned above it not that much of a difference compared to the above method.

```
##
## Call:
## glm(formula = SR ~ NP + FA + H + E + BF + DS + AC + AR, family = "binomial",
##     data = SG50)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.7804  -0.5652  -0.3373  -0.1853   3.0361
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.401126   0.866185  -0.463  0.64330
## NPNP2       -1.090248   0.350141  -3.114  0.00185 **
## NPNP3       -2.320346   0.530839  -4.371 1.24e-05 ***
## FADOD        1.217326   0.570722   2.133  0.03293 *
## FADOE        1.150949   0.637401   1.806  0.07097 .
## FANASA      -0.079129   1.143931  -0.069  0.94485
## FANIH       -0.331883   0.465450  -0.713  0.47582
## FAOT         1.523960   0.522836   2.915  0.00356 **
## H            0.019637   0.007385   2.659  0.00784 **
## E           -0.239514   0.080540  -2.974  0.00294 **
## BFBF2        0.777168   0.330880   2.349  0.01883 *
## DSDS2       -0.624192   0.331079  -1.885  0.05939 .
## AC           0.071610   0.034170   2.096  0.03611 *
## ARPR2       -1.421208   0.633325  -2.244  0.02483 *
## ARPR3       -1.751765   0.656173  -2.670  0.00759 **
## ARPR4       -1.240427   0.600665  -2.065  0.03891 *
## ARPR5       -1.587982   0.618220  -2.569  0.01021 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 352.78  on 402  degrees of freedom
```

```
## Residual deviance: 271.07  on 386  degrees of freedom
## AIC: 305.07
##
## Number of Fisher Scoring iterations: 6
```

it has still got some insignificant predictors to it like DS(**SR** ~ **NP + FA + H + E + BF + AC + AR**). so i have removed it.

```
##
## Call:
## glm(formula = SR ~ NP + FA + H + E + BF + AC + AR, family = "binomial",
##     data = SG50)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6541  -0.5517  -0.3618  -0.1856   2.9466
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.739805   0.847869  -0.873  0.38291
## NPNP2       -1.109366   0.347451  -3.193  0.00141 **
## NPNP3       -2.352486   0.532853  -4.415 1.01e-05 ***
## FADOD        1.215892   0.575289   2.114  0.03456 *
## FADOE        1.169823   0.627242   1.865  0.06218 .
## FANASA       0.054759   1.124187   0.049  0.96115
## FANIH       -0.328073   0.463724  -0.707  0.47927
## FAOT         1.480385   0.517699   2.860  0.00424 **
## H            0.021883   0.007337   2.982  0.00286 **
## E           -0.226719   0.078564  -2.886  0.00390 **
## BFBF2        0.691553   0.324237   2.133  0.03294 *
## AC           0.065148   0.033571   1.941  0.05231 .
## ARPR2       -1.433783   0.630057  -2.276  0.02287 *
## ARPR3       -1.791062   0.651747  -2.748  0.00599 **
## ARPR4       -1.257765   0.598854  -2.100  0.03570 *
## ARPR5       -1.589493   0.613702  -2.590  0.00960 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 352.78  on 402  degrees of freedom
## Residual deviance: 274.58  on 387  degrees of freedom
## AIC: 306.58
##
## Number of Fisher Scoring iterations: 6
```

The AIC of this model is 306.58, if we compare the two above models, AC which was significant in the above model is not anymore after removing DS. so i have removed it to from the model below is the summary(**SR** ~ **NP + FA + H + E + BF + AR**)

```
##
## Call:
## glm(formula = SR ~ NP + FA + H + E + BF + AR, family = "binomial",
```

```
##     data = SG50)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6327  -0.5583  -0.3716  -0.1998   3.0352
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.112530   0.726945   0.155  0.87698
## NPNP2       -1.079896   0.345701  -3.124  0.00179 **
## NPNP3       -2.334106   0.528247  -4.419 9.93e-06 ***
## FADOD        1.275880   0.576484   2.213  0.02688 *
## FADOE        1.302022   0.609770   2.135  0.03274 *
## FANASA      -0.210313   1.133933  -0.185  0.85286
## FANIH       -0.301756   0.461597  -0.654  0.51329
## FAOT         1.482684   0.516513   2.871  0.00410 **
## H            0.019810   0.007163   2.765  0.00568 **
## E           -0.183699   0.074633  -2.461  0.01384 *
## BFBF2        0.613464   0.319434   1.920  0.05480 .
## ARPR2       -1.176672   0.613326  -1.919  0.05505 .
## ARPR3       -1.519735   0.631145  -2.408  0.01604 *
## ARPR4       -1.074928   0.588699  -1.826  0.06786 .
## ARPR5       -1.355213   0.596967  -2.270  0.02320 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 352.78  on 402  degrees of freedom
## Residual deviance: 278.41  on 388  degrees of freedom
## AIC: 308.41
##
## Number of Fisher Scoring iterations: 6
```

After removing AC, i AIC has changed to 308.41, which is more than two units from our previous model which is considered high. so after comparing all the three models and their AIC's and number of predictors, we have come to a conclusion that the second model is the final model.The last model might has had the least number of predictors but the comparing the AIC The last model as mentioned above more than two unit difference is AIC is considered high.

so, logit-3 is **SR** ~ **NP + FA + H + E + BF + AC + AR** and below is the Parameters estimate

	Predictor	Prob_wise	Odds_ratio	std_error	Z_val	p_val
1	(Intercept)	0.323	−0.73980549	0.847868889	−0.87254704	0.38291
2	NPNP2	0.248	−1.10936563	0.347451347	−3.19286610	0.00141 **
3	NPNP3	0.087	−2.35248623	0.532853104	−4.41488697	1e−05 ***
4	FADOD	0.771	1.21589202	0.575288930	2.11353280	0.03456 *
5	FADOE	0.763	1.16982323	0.627241890	1.86502726	0.06218 .
6	FANASA	0.514	0.05475864	1.124187384	0.04870953	0.96115
7	FANIH	0.419	−0.32807311	0.463724370	−0.70747438	0.47927
8	FAOT	0.815	1.48038455	0.517699252	2.85954548	0.00424 **
9	H	0.505	0.02188252	0.007337126	2.98243731	0.00286 **
10	E	0.444	−0.22671882	0.078563602	−2.88579968	0.0039 **
11	BFBF2	0.666	0.69155344	0.324237475	2.13286093	0.03294 *
12	AC	0.516	0.06514843	0.033571359	1.94059555	0.05231 .
13	ARPR2	0.193	−1.43378305	0.630056700	−2.27564130	0.02287 *
14	ARPR3	0.143	−1.79106185	0.651747155	−2.74809308	0.00599 **
15	ARPR4	0.221	−1.25776512	0.598854440	−2.10028521	0.0357 *
16	ARPR5	0.169	−1.58949329	0.613701857	−2.59000893	0.0096 **

Logit-4($S^{\$75}$) we have started with Logit-1 equation and preceded with forward, backward and step-wise selection methods. Just as above all the selection methods have give me the exactly the same results. below is the summary of it.

```
##
## Call:
## glm(formula = FC ~ FA + T + TWR + RS + H + TA + E + O, family = "binomial",
##     data = F75)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9735  -1.0145  -0.6707   1.1196   1.9882
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.633027   0.965847   2.726  0.00641 **
## FADOE        0.200931   0.795718   0.253  0.80064
## FANASA      -0.642206   0.829469  -0.774  0.43879
## FANIH       -1.509412   0.587101  -2.571  0.01014 *
## FANSF       -1.442779   0.542853  -2.658  0.00787 **
## FAOT        -1.474862   0.672143  -2.194  0.02822 *
## TTS2        -0.543443   0.221893  -2.449  0.01432 *
## TWR          0.021334   0.006960   3.065  0.00217 **
## RSRS2       -0.436695   0.220332  -1.982  0.04748 *
## H            0.010820   0.005573   1.942  0.05219 .
```

```
## TA          -0.016566   0.011046  -1.500  0.13368
## E           -0.098280   0.049702  -1.977  0.04800 *
## O           -0.108308   0.061463  -1.762  0.07804 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 557.96  on 402  degrees of freedom
## Residual deviance: 504.29  on 390  degrees of freedom
## AIC: 530.29
##
## Number of Fisher Scoring iterations: 4
```

The AIC of the above model is 530.29, there are few insignificant predictors present so after removing one of them TA($\mathbf{FC} \sim \mathbf{FA} + \mathbf{T} + \mathbf{TWR} + \mathbf{RS} + \mathbf{H} + \mathbf{E} + \mathbf{O}$),below is the result.

```
##
## Call:
## glm(formula = FC ~ FA + T + TWR + RS + H + E + O, family = "binomial",
##     data = F75)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9762  -1.0093  -0.7256   1.1352   1.9535
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.360078   0.636782   0.565  0.57176
## FADOD        1.447319   0.541528   2.673  0.00753 **
## FADOE        1.616276   0.605352   2.670  0.00759 **
## FANASA       0.748336   0.654768   1.143  0.25308
## FANIH       -0.023322   0.284823  -0.082  0.93474
## FAOT        -0.081497   0.440745  -0.185  0.85330
## TTS2        -0.492391   0.218031  -2.258  0.02392 *
## TWR          0.021775   0.006925   3.144  0.00167 **
## RSRS2       -0.441436   0.219544  -2.011  0.04436 *
## H            0.012379   0.005498   2.252  0.02434 *
## E           -0.084595   0.048620  -1.740  0.08188 .
## O           -0.112096   0.061392  -1.826  0.06786 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 557.96  on 402  degrees of freedom
## Residual deviance: 506.56  on 391  degrees of freedom
## AIC: 530.56
##
## Number of Fisher Scoring iterations: 4
```

there is only a 0.27 increase in the AIC which is insignificant, the next predictor i have removed is E($\mathbf{FC} \sim \mathbf{FA} + \mathbf{T} + \mathbf{TWR} + \mathbf{RS} + \mathbf{H} + \mathbf{O}$), below is the result

```
##
## Call:
## glm(formula = FC ~ FA + T + TWR + RS + H + O, family = "binomial",
##     data = F75)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.9118  -1.0140  -0.6824   1.1528   1.8474
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -0.110662   0.575011  -0.192  0.84739
## FADOD        1.471609   0.540101   2.725  0.00644 **
## FADOE        1.520466   0.596143   2.551  0.01076 *
## FANASA       0.673052   0.649964   1.036  0.30043
## FANIH       -0.037829   0.283622  -0.133  0.89389
## FAOT        -0.095856   0.439791  -0.218  0.82746
## TTS2        -0.488214   0.217315  -2.247  0.02467 *
## TWR          0.022207   0.006915   3.212  0.00132 **
## RSRS2       -0.432951   0.218442  -1.982  0.04748 *
## H            0.012837   0.005485   2.340  0.01927 *
## O           -0.120590   0.061126  -1.973  0.04852 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 557.96  on 402  degrees of freedom
## Residual deviance: 509.61  on 392  degrees of freedom
## AIC: 531.61
##
## Number of Fisher Scoring iterations: 4
```

the AIC has increased to 531.61. which is acceptable. and if we observe the two models above, the openness score was insignificant in the first model but became significant after removing Extraversion(E). so, After comparing the AIC values and number of predictors we have decided that the above model is best suited for this analysis because it has the least number of predictors and AIC with only one-unit difference compared to the model with least AIC.

so, Logit-4 is **FC ~ FA + T + TWR + RS + H + O** and below is the parameter estimate

	Predictor	Prob_wise	Odds_ratio	std_error	Z_val	p_val
1	(Intercept)	0.472	−0.11066238	0.575010546	−0.1924528	0.84739
2	FADOD	0.813	1.47160860	0.540101289	2.7246900	0.00644 **
3	FADOE	0.821	1.52046573	0.596143345	2.5505036	0.01076 *
4	FANASA	0.662	0.67305207	0.649964180	1.0355218	0.30043
5	FANIH	0.491	−0.03782897	0.283622369	−0.1333779	0.89389
6	FAOT	0.476	−0.09585561	0.439790797	−0.2179573	0.82746
7	TTS2	0.380	−0.48821435	0.217315451	−2.2465699	0.02467 *
8	TWR	0.506	0.02220679	0.006914599	3.2115806	0.00132 **
9	RSRS2	0.393	−0.43295119	0.218441910	−1.9819969	0.04748 *
10	H	0.503	0.01283690	0.005485412	2.3401892	0.01927 *
11	O	0.470	−0.12058971	0.061125823	−1.9728112	0.04852 *

Logit-5($S^{\$\$}$)

Even for this Analysis we have got the same results for all the three selection methods.below is the summary of it.

```
##
## Call:
## glm(formula = FC ~ FA + T + TWR + RS + O + TC, family = "binomial",
##     data = full)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5430  -0.7173  -0.5335  -0.3289   2.4630
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -1.85207    1.18524  -1.563  0.11814
## FADOE       -0.63374    0.67752  -0.935  0.34959
## FANASA      -1.19887    0.93091  -1.288  0.19780
## FANIH       -1.44538    0.55926  -2.584  0.00975 **
## FANSF       -1.62151    0.49447  -3.279  0.00104 **
## FAOT        -1.09857    0.67339  -1.631  0.10280
## TTS2        -0.38176    0.26261  -1.454  0.14603
## TWR          0.03949    0.00851   4.640 3.48e-06 ***
## RSRS2       -0.41016    0.25833  -1.588  0.11234
## O           -0.15429    0.07389  -2.088  0.03679 *
```

```
## TC           0.06333    0.03487   1.816  0.06933 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 430.53  on 402  degrees of freedom
## Residual deviance: 382.88  on 392  degrees of freedom
## AIC: 404.88
##
## Number of Fisher Scoring iterations: 4
```

AIC of the above model is 404.88, have we have few insignificant features.so, first i have removed RS($\mathbf{FC} \sim \mathbf{TWR} + \mathbf{FA} + \mathbf{O} + \mathbf{TC} + \mathbf{T}$) and checked the AIC, it was 405.39

```
##
## Call:
## glm(formula = FC ~ TWR + FA + O + TC + T, family = "binomial",
##     data = full)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6364  -0.7192  -0.5399  -0.3269   2.5425
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.719172   1.136730  -3.272  0.00107 **
## TWR          0.038998   0.008488   4.594 4.34e-06 ***
## FADOD        1.610567   0.492890   3.268  0.00108 **
## FADOE        1.052867   0.510600   2.062  0.03921 *
## FANASA       0.409281   0.819650   0.499  0.61754
## FANIH        0.214909   0.341723   0.629  0.52941
## FAOT         0.522334   0.522790   0.999  0.31773
## O           -0.153825   0.073602  -2.090  0.03662 *
## TC           0.063941   0.034674   1.844  0.06517 .
## TTS2        -0.387816   0.261329  -1.484  0.13781
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 430.53  on 402  degrees of freedom
## Residual deviance: 385.39  on 393  degrees of freedom
## AIC: 405.39
##
## Number of Fisher Scoring iterations: 4
```

next, i have removed T($\mathbf{FC} \sim \mathbf{TWR} + \mathbf{FA} + \mathbf{O} + \mathbf{TC}$),then it gave us the AIC 405.61, below is the result

```
##
## Call:
## glm(formula = FC ~ TWR + FA + O + TC, family = "binomial", data = full)
##
```

```
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6848  -0.7154  -0.5422  -0.3424   2.4680
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.680192   1.133284  -3.247 0.001165 **
## TWR          0.038358   0.008468   4.530 5.91e-06 ***
## FADOD        1.633091   0.491825   3.320 0.000899 ***
## FADOE        0.986897   0.504601   1.956 0.050489 .
## FANASA       0.336094   0.817598   0.411 0.681018
## FANIH        0.148044   0.338664   0.437 0.662009
## FAOT         0.453996   0.519286   0.874 0.381971
## O           -0.163267   0.073011  -2.236 0.025339 *
## TC           0.059684   0.034413   1.734 0.082858 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 430.53  on 402  degrees of freedom
## Residual deviance: 387.61  on 394  degrees of freedom
## AIC: 405.61
##
## Number of Fisher Scoring iterations: 4
```

we, still have one insignificant feature that is TC(**FC** ~ **TWR + FA + O**), after removing it we got the below model.

```
##
## Call:
## glm(formula = FC ~ TWR + FA + O, family = "binomial", data = full)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5998  -0.7247  -0.5456  -0.3676   2.5441
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.102705   0.653579  -3.217 0.001294 **
## TWR          0.038722   0.008388   4.616  3.9e-06 ***
## FADOD        1.621710   0.488825   3.318 0.000908 ***
## FADOE        0.970498   0.507132   1.914 0.055658 .
## FANASA       0.289574   0.816047   0.355 0.722702
## FANIH        0.195893   0.334923   0.585 0.558622
## FAOT         0.456031   0.513920   0.887 0.374886
## O           -0.148075   0.072433  -2.044 0.040923 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 430.53  on 402  degrees of freedom
## Residual deviance: 390.73  on 395  degrees of freedom
```

```
## AIC: 406.73
##
## Number of Fisher Scoring iterations: 4
```

we, got an AIC of 406.73. we don't have any more features to remove so, after comparing all the models AIC and number of predictors, we have choose the above model which has an AIC of 406.73 with 3 predictors as our final model.

so,Logit-5 is $\mathbf{FC \sim TWR + FA + O}$ and below is the parameter estimate

	Predictor	Prob_wise	Odds_ratio	std_error	Z_val	p_val
1	(Intercept)	0.109	−2.10270451	0.653578934	−3.2172159	0.00129441 **
2	TWR	0.510	0.03872225	0.008387917	4.6164318	3.9e−06 ***
3	FADOD	0.835	1.62170953	0.488825270	3.3175649	0.00090806 ***
4	FADOE	0.725	0.97049800	0.507131708	1.9137001	0.05565849 .
5	FANASA	0.572	0.28957427	0.816046818	0.3548501	0.72270192
6	FANIH	0.549	0.19589311	0.334923042	0.5848899	0.55862175
7	FAOT	0.612	0.45603115	0.513920319	0.8873577	0.37488635
8	O	0.463	−0.14807489	0.072432710	−2.0443097	0.04092296 *

Logit-6 ($S^{G50\$\$}$) The result of all the feature selection models is below

```
##
## Call:
## glm(formula = jj ~ NP + TWR + RS + E + A + O, family = "binomial",
##     data = SG50f)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.2788  -0.3312  -0.1975  -0.1091   2.9913
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.99823    1.59404  -1.881 0.059985 .
## NPNP2       -1.90548    0.59197  -3.219 0.001287 **
## NPNP3       -2.82918    0.82248  -3.440 0.000582 ***
## TWR          0.04454    0.01386   3.214 0.001308 **
```

```
## RSRS2       -0.75774    0.46237  -1.639 0.101250
## E           -0.26825    0.11286  -2.377 0.017463 *
## A            0.34244    0.15073   2.272 0.023098 *
## O           -0.21291    0.13171  -1.616 0.105986
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 187.42  on 402  degrees of freedom
## Residual deviance: 143.83  on 395  degrees of freedom
## AIC: 159.83
##
## Number of Fisher Scoring iterations: 7
```

The above model has a AIC of 159.83 with 6 parameters. below is the result after removing RS insignificant Predictor(**jj** ~ **NP + TWR + E + A + O**)

```
##
## Call:
## glm(formula = jj ~ NP + TWR + E + A + O, family = "binomial",
##     data = SG50f)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.3672  -0.3448  -0.2152  -0.1185   2.8780
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -3.20092    1.60989  -1.988 0.046780 *
## NPNP2       -1.86899    0.58617  -3.188 0.001430 **
## NPNP3       -2.63463    0.80038  -3.292 0.000996 ***
## TWR          0.04210    0.01367   3.081 0.002065 **
## E           -0.26569    0.11326  -2.346 0.018980 *
## A            0.31537    0.14894   2.118 0.034215 *
## O           -0.20333    0.13139  -1.547 0.121750
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 187.42  on 402  degrees of freedom
## Residual deviance: 146.52  on 396  degrees of freedom
## AIC: 160.52
##
## Number of Fisher Scoring iterations: 7
```

we have got an AIC of 160.52, we can see that Openness is insignificant so we have removed it. and below is the result.(**jj** ~ **NP + TWR + E + A**)

```
##
## Call:
## glm(formula = jj ~ NP + TWR + E + A, family = "binomial", data = SG50f)
```

```
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.5266  -0.3487  -0.2163  -0.1286   3.0476
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -4.51626    1.40527  -3.214  0.00131 **
## NPNP2       -1.75461    0.57739  -3.039  0.00237 **
## NPNP3       -2.54042    0.80291  -3.164  0.00156 **
## TWR          0.03958    0.01353   2.925  0.00344 **
## E           -0.26673    0.11071  -2.409  0.01598 *
## A            0.30213    0.14720   2.053  0.04012 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 187.42  on 402  degrees of freedom
## Residual deviance: 148.95  on 397  degrees of freedom
## AIC: 160.95
##
## Number of Fisher Scoring iterations: 7
```

This model has an AIC of 160.95 with no further insignificant predictors so, we have stopped the search here. after looking at the three models AIC and Number of predictors we can considered the last model with AIC 160.95 with 4 predictors as our final model for this analysis because it has the less number of predictors and only one unit high AIC which is acceptable.

so, Logit-6 is **jj** ~ **NP + TWR + E + A** and the parameter estimate table is below

	Predictor	Prob_wise	Odds_ratio	std_error	Z_val	p_val
1	(Intercept)	0.011	−4.51625802	1.40526702	−3.213808	0.00131 **
2	NPNP2	0.147	−1.75461336	0.57739401	−3.038849	0.00237 **
3	NPNP3	0.073	−2.54041618	0.80290767	−3.164020	0.00156 **
4	TWR	0.510	0.03957682	0.01352892	2.925350	0.00344 **
5	E	0.434	−0.26673346	0.11070707	−2.409362	0.01598 *
6	A	0.575	0.30212657	0.14719833	2.052514	0.04012 *