SNSF Report

Chiara Gilardi, Leslie O'Bray, Carla Sch?rer and Tommaso Portaluri

13 June 2018

Introduction

The Swiss National Science Foundation (SNF) is a research funding agency which disseminates yearly, on behalf of the Swiss Government, billions of CHF to the best researchers in Switzerland. This report contains a statistical analysis performed on three data sets provided by SNF, containing information on the applications for funding received in 2016, the corresponding grades and the scores given by both internal and external evaluators.

The analysis performed for SNF had a three-fold aim, corresponding to the following three research questions: 1) Is gender bias occurring at any stage of the SNSF evaluation process? Is the gender of the main applicant influencing the rating of the application? 2) To what extent the different steps of the evaluation and the different criteria within each step determine the final funding decision? 3) When an application is approved, but the budget requested is cut, how can we explain this?

The SNSF evaluation procedure is a multi-step process (involving external reviewers, internal referees, and an internal board) which takes into consideration both the track record of the applicant and the quality of the project (see Appendix for a more detailed description of the evaluation procedure).

Several studies (Witteman et al., 2017; Solans-Domenech et al., 2017) have shown that female applicants' projects get higher score when the application is blinded. Moreover, female applicants receive usually higher grades for projects and lower grades for track record. Hence, after investigating the gender dimension to identify possible biases in the evaluation procedure, the focus of the analysis will be the relative importance of of the criteria for funding (applicant's track record vs. quality of the proposal) and, also, of each step of the evaluation procedure (which opinion is more likely to determine the final decision - the external referee's or the board's?). Possible interactions between the gender dimension and the second research question will also be investigated (for instance, by taking into account also the gender of evaluator or the percentage of female referees).

Data Description

We have three data sets: Applications, External Reviewers and Internal Referees. They contain respectively information about the SNSF project funding applications, the evaluation of the applications by external peer reviewers and the evaluation of the proposals by external the internal referee and co-referee (when available). For a full description of the data & variables, please see the Appendix.

Cleaning the Data

We decide to work with only complete applications, i.e. project for which we have information from all the three data sets.

To avoid a temporal trend, we are only considering application from 2016.

In both the external and internal step, we encountered applications which had several reviews per application. For the sake of our analysis, in these scenarios we computed the mean grade for each criteria, so that each application had a “single” score for each criteria assessed on. In doing so we also introduced a new variable, PercentFemale, which calculated the percent of female reviewers out of all reviewers of a single application (ranging from 0 to 1).

All applications with a grade were converted to an ordinal factor.

Specific to each data set, this are the detailed considerations:

Applications

We decide to consider only the MainDiscipline2 because for MainDiscipline we have 118 levels, while for the other only 21.

There is one application for which we do not know the gender of the applicant, and therefore we decided to omit that observation from the analysis.

We will also not consider the variables “CallTitle”, “Professorship”, “AcademicAge”. The first one, because we consider it has nothing to add to the model. The two last, due to the fact that there are a considerable number of NA’s on those variables (around 93% of the observations).

External Reviewers

Reviewers always have the option to choose not to consider or to give the grade “0” when reviewing an application. Some might be mistakes, in others cases there might be a conflict of interest, or they might be very ambivalent about the project. Therefore, we did not considered observations with this grades.

One of the questions evaluated in the applications is “Broader impact (forms part of the assessment of scientific relevance, originality and topicality)”. For the time frame we are considering, in all the applications this grade was NA. Hence, we omit this variable from our model.

- **ProposalCombined**: We created a new variable to summarize the assessment of the scientific proposal in the external review step. This is a simple mean of the grade given for Suitability and Scientific Relevance. This helped to isolate the effect of the grade given to the scientific proposal, versus the applicant track record, as well as to ensure easier comparison with the internal review step.
- **PercentFemale**: As previously mentioned, we introduced a new variable calculating the percent of reviewers of each application that is female.

Internal Referees

There were 22 observations (1 for the time frame we are dealing with) for which only demographic information was available, no grades were given. We decide to omit those observations.

Also we decide to not consider the Referee role as a variable in our model, as the majority of the evaluations has only one referee.

- **PercentFemale**: As previously mentioned, we introduced a new variable calculating the percent of reviewers of each application that is female.

Exploratory Analysis

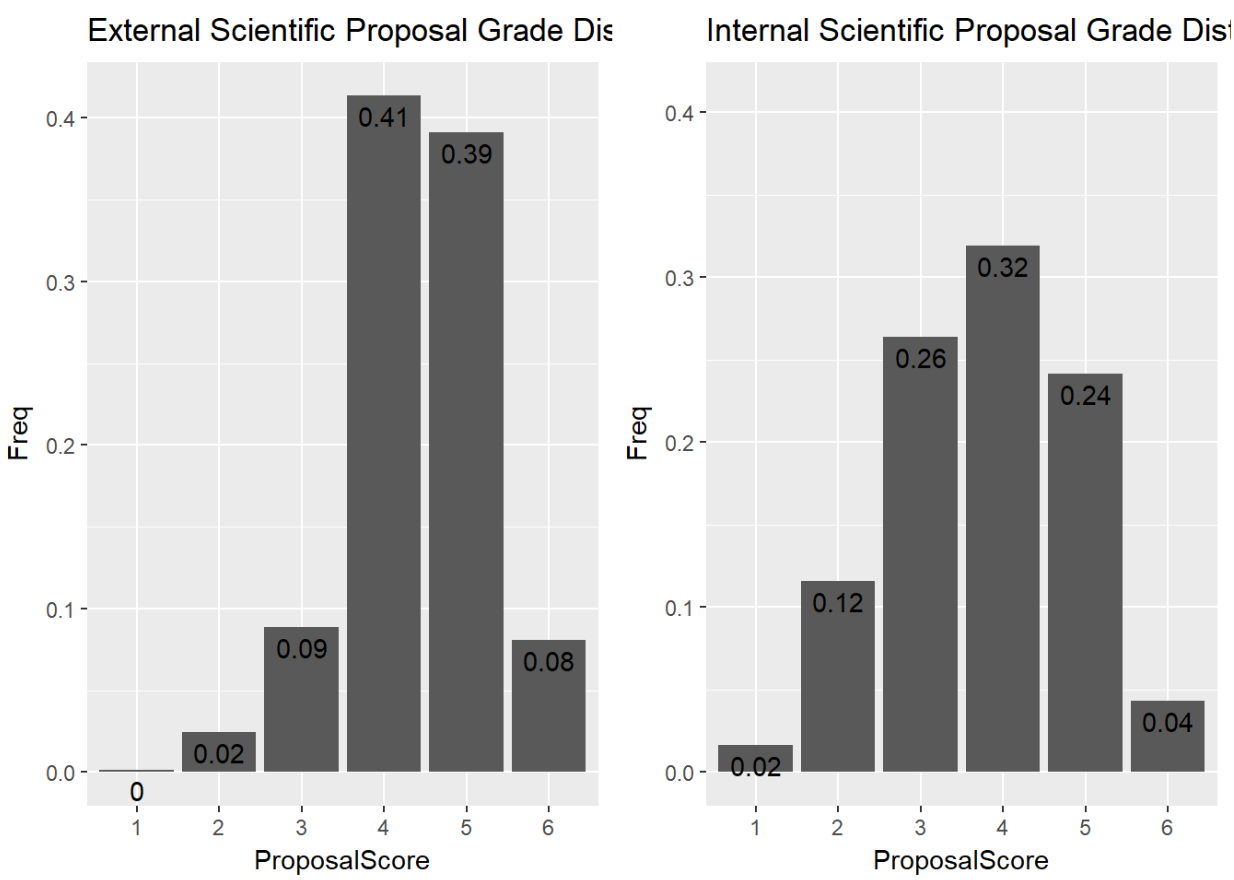
In our exploratory analysis, we discovered a few interesting insights, that relates to the findings we will discuss from our analysis.

Distribution of Grades between the External & Internal Review Step

Since the external & internal step both assess candidates on the same criteria (the strength of the scientific proposal, and the strength of the applicant), on the same ordinal scale (from poor to outstanding), we were interested to see if the distribution of grades are the same. We would expect different distributions for the Overall Grade vs the Ranking, since those have two different measurements, however we were interested to see if for the same absolute ranking, the external and internal reviewers had different perspectives on the application. After combining the Suitability & Scientific Relevance grades given to a candidate in the external review step, we can compare the average grade given for the Scientific Proposal in the two steps, as well as the grade given for the Applicant Track Record in both steps.

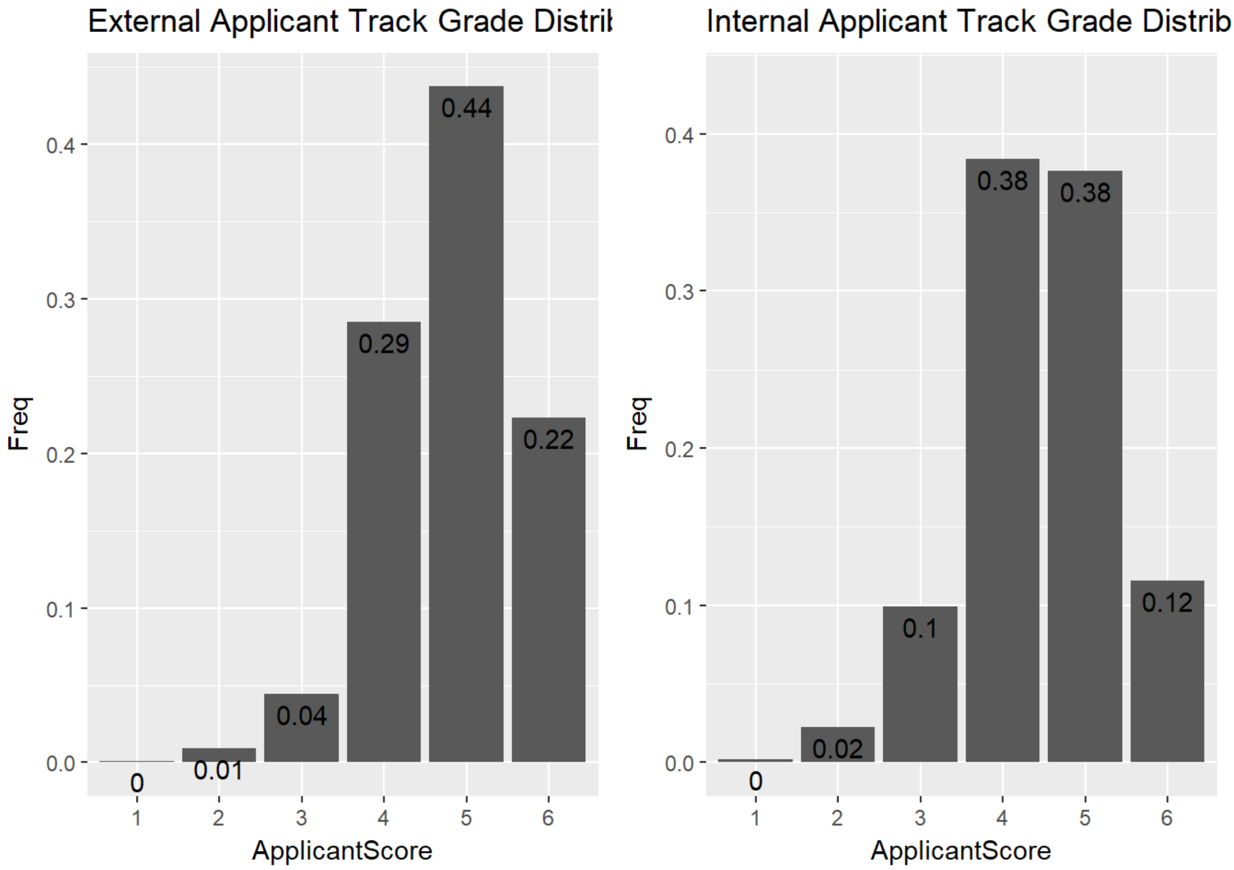
We see that the External Reviewers are more generous with their grades; for the strength of the Scientific Proposal, 48% of proposals are considered “excellent” or “outstanding”, versus only 28% in the internal review step.

```
<ScaleContinuousPosition>
 Range:
 Limits:    0 -- 0.45
```



Similarly we see the same pattern with Applicant Track Record: 66% of Applicant Track records are considered “excellent” or “outstanding” by the External Reviewers, versus merely 50% by the Internal Reviewers.

```
<ScaleContinuousPosition>
 Range:
 Limits:    0 -- 0.45
```



Since we noticed this discrepancy, we wanted to quantify how differently the grades were to one another. To assess the agreement between the two steps, for the same criteria, we used Cohen’s Kappa. Cohen’s Kappa measures the proportion of agreement between two raters assessing something on an ordinal scale, accounting for the fact that there will always be some proportion by random chance. An important specification of Cohen’s Kappa is the weight given to the measurements. If the external & internal reviewers both assessed the Applicant Track Record as “excellent”, that would be considered full agreement. However, we want to allocate partial credit if the rating is a level close to it. We used a linear weight up to distance 2, and after that gave no credit. (In this example, if one rater gave an “outstanding” or “very good”, that would be considered a distance of one and be weighted by 0.8. If the second rater assessed the Applicant Track to be “good”, which is a distance of two away from excellent, that would be weighted as 0.6). Anything with a distance of 3 or more (in this example, if the second rater gave a rating of “average”), we allocated no weight, as the difference between average and excellent is quite large.

From this, we found that there was just moderate agreement between the two steps when using the weighted kappa, for both the grades given for the Scientific Proposal & the Applicant Track Record.

```
[1] "Cohen's Kappa for Applicant Track Record"
```

```
    Call: cohen.kappa1(x = x, w = w, n.obs = n.obs, alpha = alpha, levels = levels)

    Cohen Kappa and Weighted Kappa correlation coefficients and confidence boundari
es
                     lower estimate upper
    unweighted kappa  0.21     0.24  0.28
    weighted kappa    0.35     0.42  0.49

     Number of subjects = 1623
```

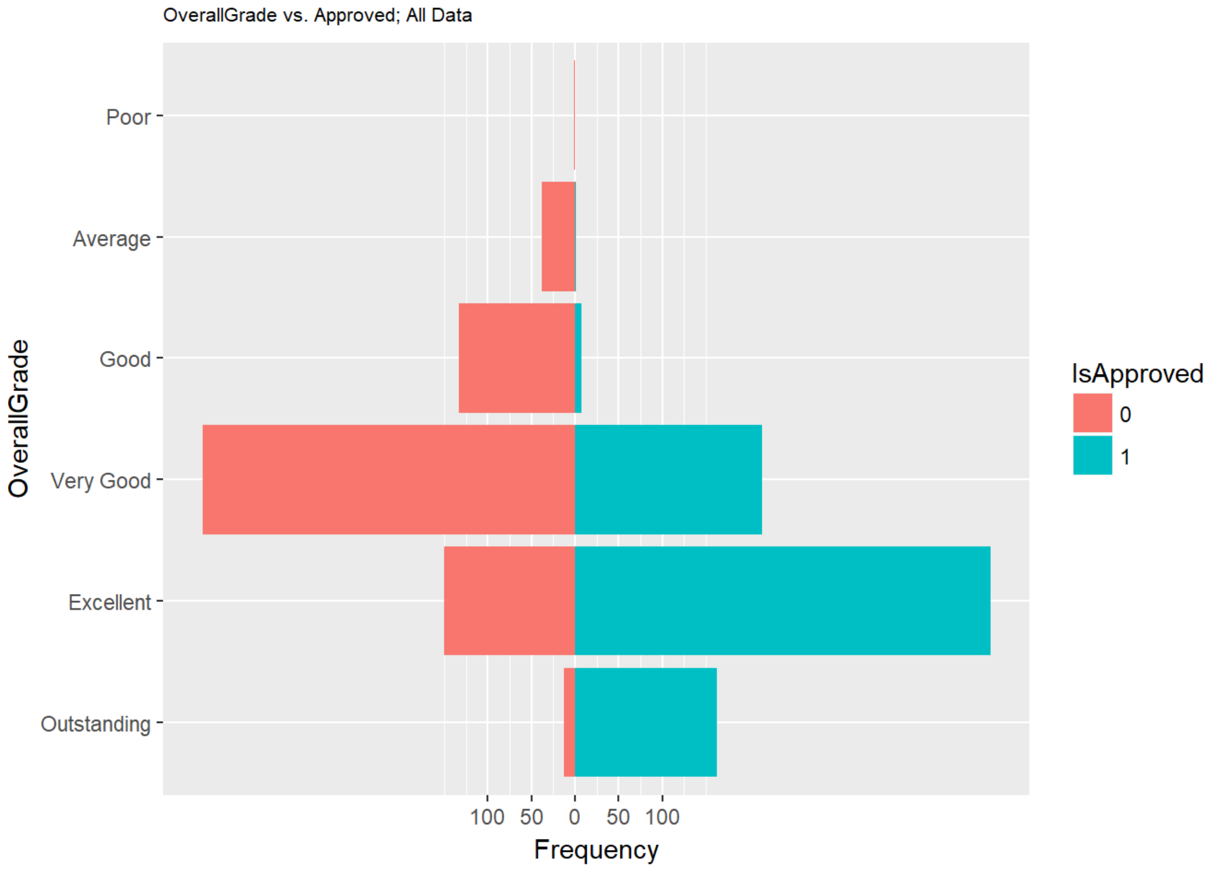
```
    [1] "Cohen's Kappa for Scientific Proposal"
```

```
    Call: cohen.kappa1(x = x, w = w, n.obs = n.obs, alpha = alpha, levels = levels)

    Cohen Kappa and Weighted Kappa correlation coefficients and confidence boundari
es
                     lower estimate upper
    unweighted kappa  0.21     0.24  0.28
    weighted kappa    0.35     0.42  0.49

     Number of subjects = 1623
```

Impact of Internal Reviewers on Funding

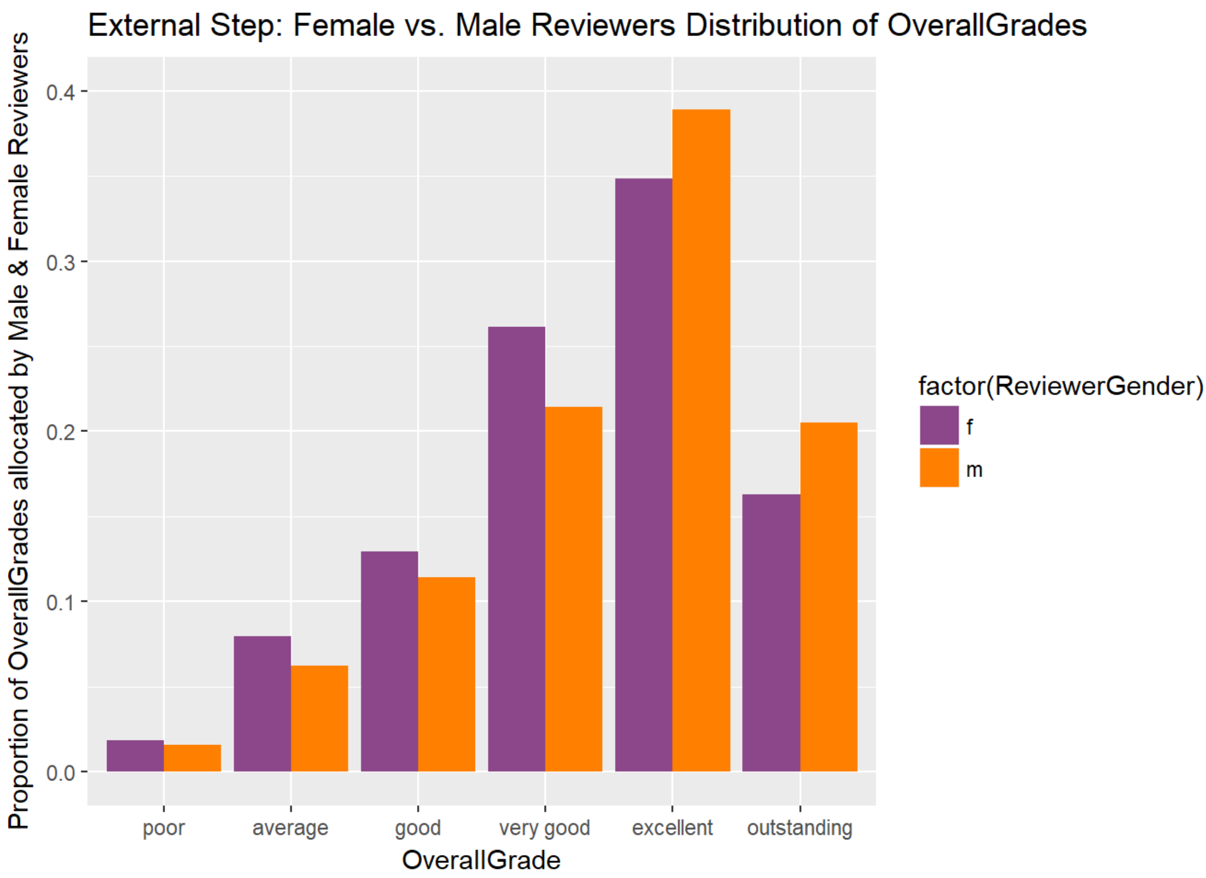
We wanted to understand if this discrepancy between grades had an impact on whether an application is funded. To do this, we visualized the summary grade given to an application, and whether that application is funded or not. As we can see here, there are several applications with an OverallGrade of “excellent” or “outstanding” that end up not approved. It highlights that not only do the Internal Reviewers give tougher grades in general than the external step, but they also consider some “excellent” and “outstanding” applications by the external reviewers to be not of the quality that deserves funding. This trend is true in all divisions and both genders, please refer to the appendix to see the specific graphic.

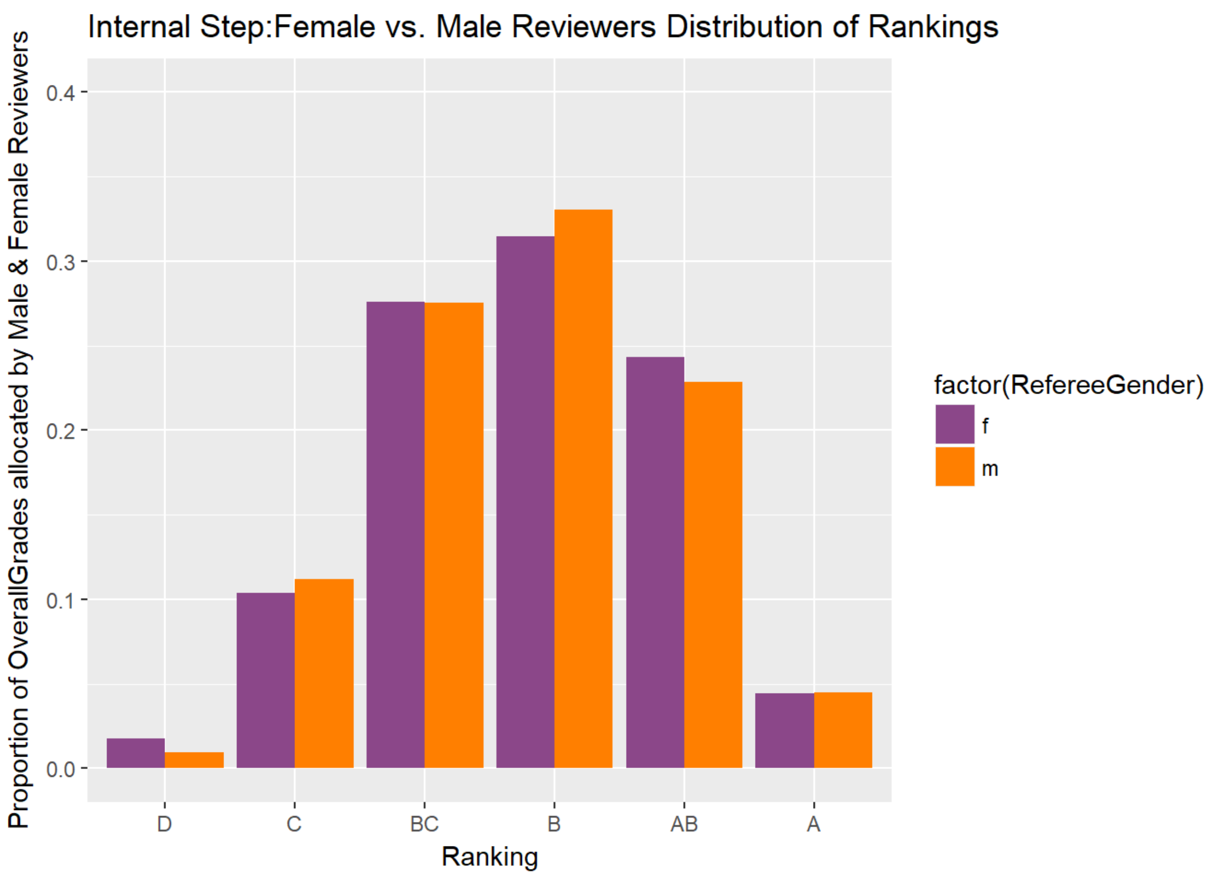
Our conclusion for this is that the internal step is very consequential, and the difference in the rating they give translates into differences in whether an application gets funded or not.



Distribution of Grades by the Gender of the Reviewer

The third interesting insight we found was when we investigated the impact of the gender of the person reviewing the data. We look at the relative frequencies of grades given by male and female reviewers, to applicants, regardless of gender. Within the external step in particular, we found that female reviewers give proportionally fewer “excellent” and “outstanding” grades, compared to their male counterparts. Within the internal step, we did not notice a particular difference, though we will consider the impact of the gender of the reviewer more rigorously in our analysis.


External Step: Female vs. Male Reviewers Distribution of OverallGrades
Proportion of OverallGrades allocated by Male & Female Reviewers
0.0
0.1
0.2
0.3
0.4
poor
average
good
very good
excellent
outstanding
OverallGrade
factor(ReviewerGender)
f
m


Internal Step:Female vs. Male Reviewers Distribution of Rankings
Proportion of OverallGrades allocated by Male & Female Reviewers
0.0
0.1
0.2
0.3
0.4
D
C
BC
B
AB
A
Ranking
factor(RefereeGender)
f
m

Gender Bias

To see if gender has an influence in any of the steps of the evaluation process, we did several things. For the external and internal steps, we first fit a logistic regression with the function glm in R, where we used IsApproved (a binary variable) as a response and demographic information of the applicant, project information and the given grades as predictors. The aim of this regression is to see if gender has an influence on the final decision from the perspective of each step.

As the final decision is determined by the different grades in the process, in order to see if gender has an influence on any of them, we fitted an Ordinal regression with the function clm of the package “ordinal” on each grade with demographic data and project information as predictors.

Analysis

External Step

Logistic Regression

Regression data: To perform the analysis, we combined in one data frame information about the applications (IsApproved, Age, Gender, Division, IsContinuation, PreviousRequest, InstType, log (AmountRequested), Semester) and about the grades given by the external reviewers (ApplicantTrack, ScientificRelevance, Suitability, OverallGrade, ProposalCombined, PercentFemale).

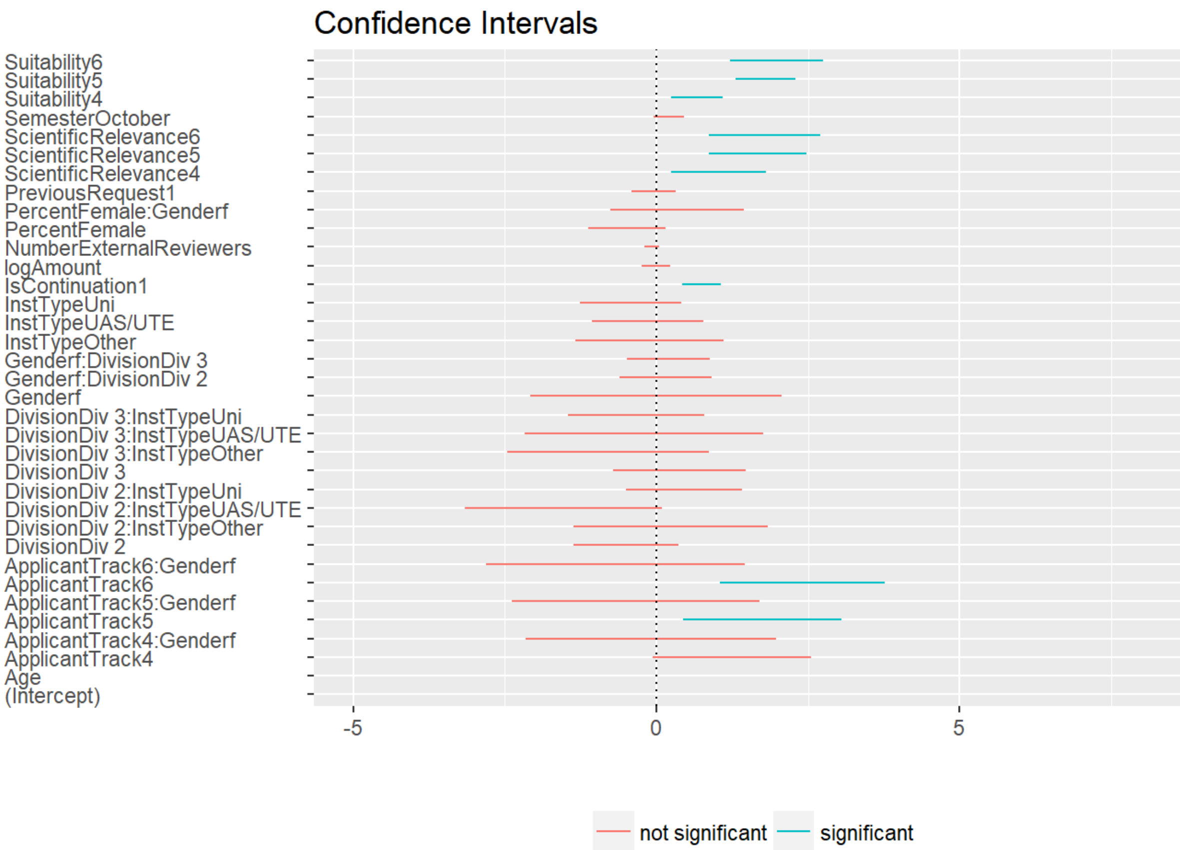
As there are almost no application approved that have grades smaller than “good”, we decide to aggregate grades “poor”, “average” and “good” to avoid perfect separation problems. All grades are considered as ordered factors.

We first fitted a full model with all the variables and the interactions between Gender and Division, PercentFemale and ApplicantTrack. Also we considered the interaction between InstType and Division. we didn’t considered OverallGrade as it is highly correlated with the grades of the applicant and the project. we achieved a pseudo-R^2 value of 0.4225, indicating that percent of the variation in the approval of the applications can be explained by the model.

```
Call:
glm(formula = IsApproved ~ . + Gender:Division + Gender:PercentFemale +
    Gender:ApplicantTrack + InstType:Division, family = "binomial",
    data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4435  -0.8248   0.3475   0.7641   2.3987

Coefficients:
                              Estimate Std. Error z value Pr(>|z|)
(Intercept)                  -2.825251   1.710909  -1.651 0.098674 .
ApplicantTrack4               1.068437   0.640934   1.667 0.095514 .
ApplicantTrack5               1.569166   0.643813   2.437 0.014797 *
ApplicantTrack6               2.240959   0.669648   3.346 0.000818 ***
ScientificRelevance4          0.973704   0.392782   2.479 0.013175 *
ScientificRelevance5          1.619786   0.407403   3.976 7.01e-05 ***
ScientificRelevance6          1.743556   0.465394   3.746 0.000179 ***
Suitability4                  0.660676   0.216996   3.045 0.002330 **
Suitability5                  1.785305   0.251556   7.097 1.27e-12 ***
Suitability6                  1.959409   0.391975   4.999 5.77e-07 ***
PercentFemale                -0.482505   0.323079  -1.493 0.135317
NumberExternalReviewers      -0.074361   0.062292  -1.194 0.232580
Age                          -0.003454   0.008290  -0.417 0.676955
Genderf                       0.093744   1.004496   0.093 0.925646
DivisionDiv 2                -0.493560   0.440134  -1.121 0.262123
DivisionDiv 3                 0.369348   0.554063   0.667 0.505017
IsContinuation1               0.740935   0.162477   4.560 5.11e-06 ***
PreviousRequest1             -0.050188   0.185346  -0.271 0.786561
InstTypeOther                -0.105453   0.621404  -0.170 0.865245
InstTypeUAS/UTE              -0.141240   0.469157  -0.301 0.763375
InstTypeUni                  -0.418471   0.425878  -0.983 0.325801
SemesterOctober               0.201023   0.129692   1.550 0.121140
logAmount                    -0.008541   0.119088  -0.072 0.942823
Genderf:DivisionDiv 2         0.150768   0.389586   0.387 0.698759
Genderf:DivisionDiv 3         0.188020   0.348181   0.540 0.589193
PercentFemale:Genderf         0.340882   0.558943   0.610 0.541948
ApplicantTrack4:Genderf      -0.194354   1.001970  -0.194 0.846198
ApplicantTrack5:Genderf      -0.443824   0.993699  -0.447 0.655136
ApplicantTrack6:Genderf      -0.766600   1.042735  -0.735 0.462229
DivisionDiv 2:InstTypeOther   0.234177   0.813566   0.288 0.773469
DivisionDiv 3:InstTypeOther  -0.796655   0.844377  -0.943 0.345434
DivisionDiv 2:InstTypeUAS/UTE -1.470105   0.820285  -1.792 0.073103 .
DivisionDiv 3:InstTypeUAS/UTE -0.195563   0.996166  -0.196 0.844363
DivisionDiv 2:InstTypeUni     0.450810   0.483178   0.933 0.350814
DivisionDiv 3:InstTypeUni    -0.333450   0.571995  -0.583 0.559920
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)
```

```
    Null deviance: 2243.9  on 1622  degrees of freedom
Residual deviance: 1626.4  on 1588  degrees of freedom
AIC: 1696.4

Number of Fisher Scoring iterations: 5
```

When selecting the variables with the AIC criteria in order to work with a small and effective model, we end up with the following predictors: ApplicantTrack, ScientificRelevance, Suitability, PercentFemale, Age, Gender, Division, IsContinuation, InstType, Semester, Division:InstType. No interaction with Gender where significant. The pseudo R^2 for this model is 0.4203 , i.e. this smaller model explains basically the same variance of the data than the former one. Non of them reveal that the information from the external reviewers explain the final decision correctly. This fact will be explored in more detail later on.



The confidence intervals which do not include zero are those for the variables: IsContinuation, Suitability (all levels from 4 to 6), ScientificRelevance (all levels from 4 to 6), and ApplicantTrack(for levels 5 and 6)

```
        ApplicantTrack5      ApplicantTrack6 ScientificRelevance4
               4.802641             9.402340             2.647732
ScientificRelevance5 ScientificRelevance6          Suitability4
               5.052007             5.717641             1.936100
           Suitability5         Suitability6       IsContinuation1
               5.961400             7.095133             2.097896
```

The interpretation of the coefficients is the following:

- If the Applicant Track grade is 5, the proposal is 4.8 times more likely to be approved than a 4 or lower
- If the Applicant Track grade is 6, the proposal is 9.4 times more likely to be approved than a 3 or lower
- If ScientificRelevance grade is 4, the proposal is 2.65 times more likely to be approved than a 3 or lower
- If ScientificRelevance grade is 5, the proposal is 5.05 times more likely to be approved than a 4 or lower
- If ScientificRelevance grade is 6, the proposal is 5.72 times more likely to be approved than a 5 or lower
- If the Suitability has grade 4, the proposal is 1.94 times more likely to be approved than a 3 or lower
- If the Suitability has grade 5, the proposal is 5.96 times more likely to be approved than a 4 or lower
- If the Suitability has grade 6, the proposal is 7.1 times more likely to be approved than a 5 or lower
- If the project is a continuation, it is 2.1 times more likely to be approved than if it's not

Notice that the coefficient for gender is not significant, since its confidence interval (-2.074,2.059) contains zero. The point estimate for gender is 0.094 and this means that a woman is 2.99 times more likely to be approved than a man.

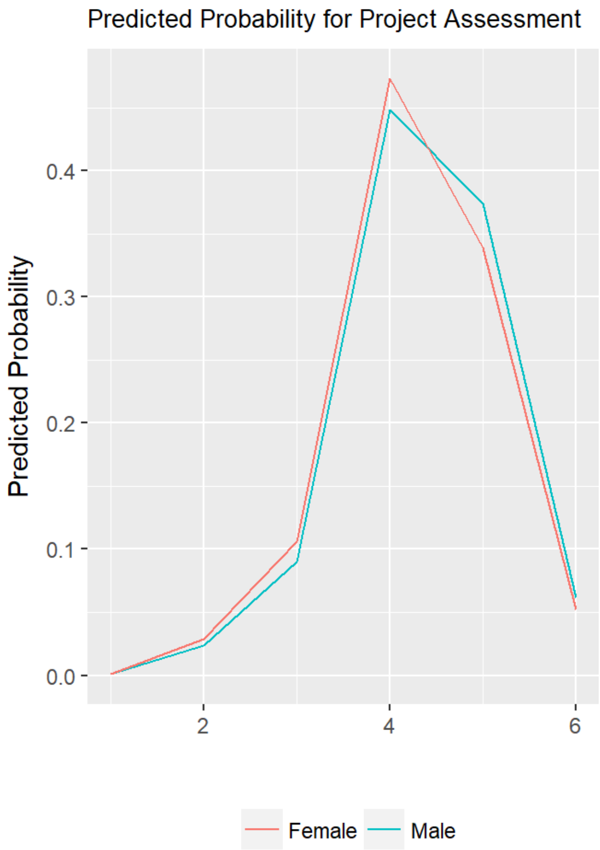
As we are interested in the effect of gender in each step of the evaluation process, we looked at the difference of the probability of being accepted between male and female given the final model. Female have a probability 0.517, while male have 0.545. Although there is a small difference, our analysis suggest that this difference is not significant.

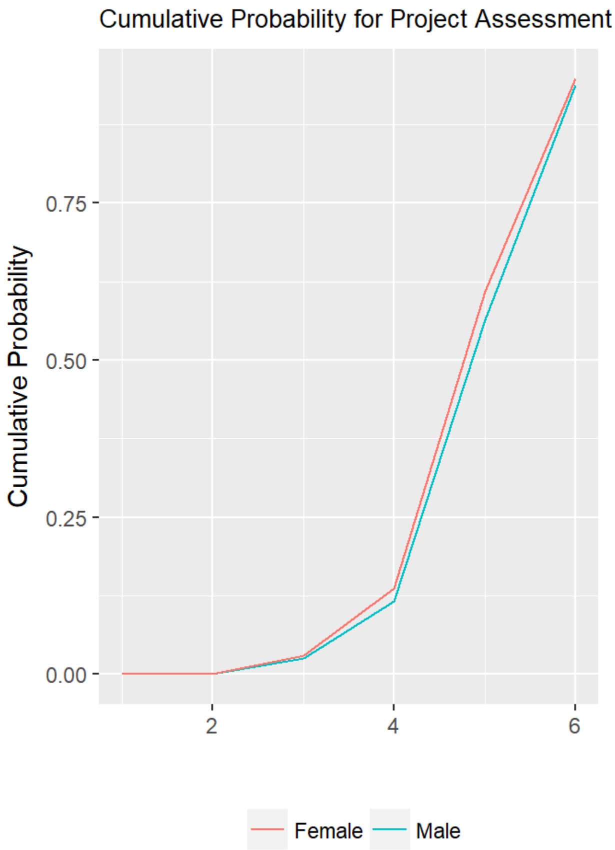
Ordinal Regression

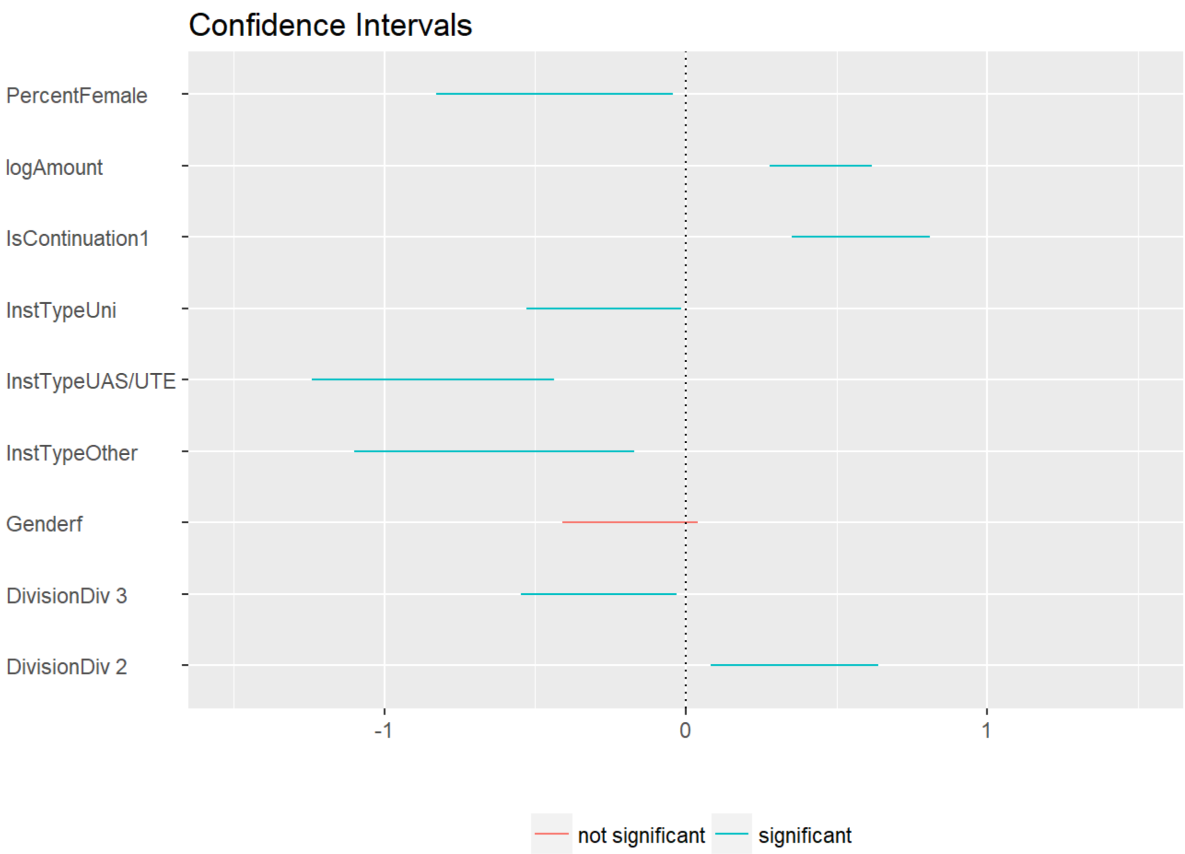
To look into the variables that influence the different grades in this part of the evaluation process, and see if the gender of the main applicant has an influence on it, we ran an ordinal regression with the function clm of the package ordinal in R. We did this for both, the applicant grade (ApplicantTrack), and the average of the grades given to the project (ProposalCombined). Also, we want to investigate which grades are more influential in the Overall Grade, which is the summary grade given by the external reviewers. In the next few paragraphs we will look into the analysis of this two regressions.

- **Project Assessment**: After fitting a full model with ProposalCombined as a response variable and different interactions, and then selecting from this model the significant variables with the AIC criteria and the help of the drop1() function in R, we end up with a model with the following predictors: Gender, Division, PercentFemale, IsContinuation, InstType and log (AmountRequested). If we fit the same model without Gender and compare it with the anova () function to the one with gender, we get a p.value of 0.1064133, meaning that for the grades given to the project, gender is not important. This is to be expected, as the project is being evaluated and not the applicant.

Overall the average difference is really small: 0.01483. This seems to suggest that there is no evidence of gender influencing the probability of achiving a certain grade. We also represented in the plots below the probabilty and cumulative probability curves of getting each grade for male and female: they follow more or less the same trend and the only difference, as we've seen from the table above, is that women are slightly more likely to get a "very good" rather than an "excellent".





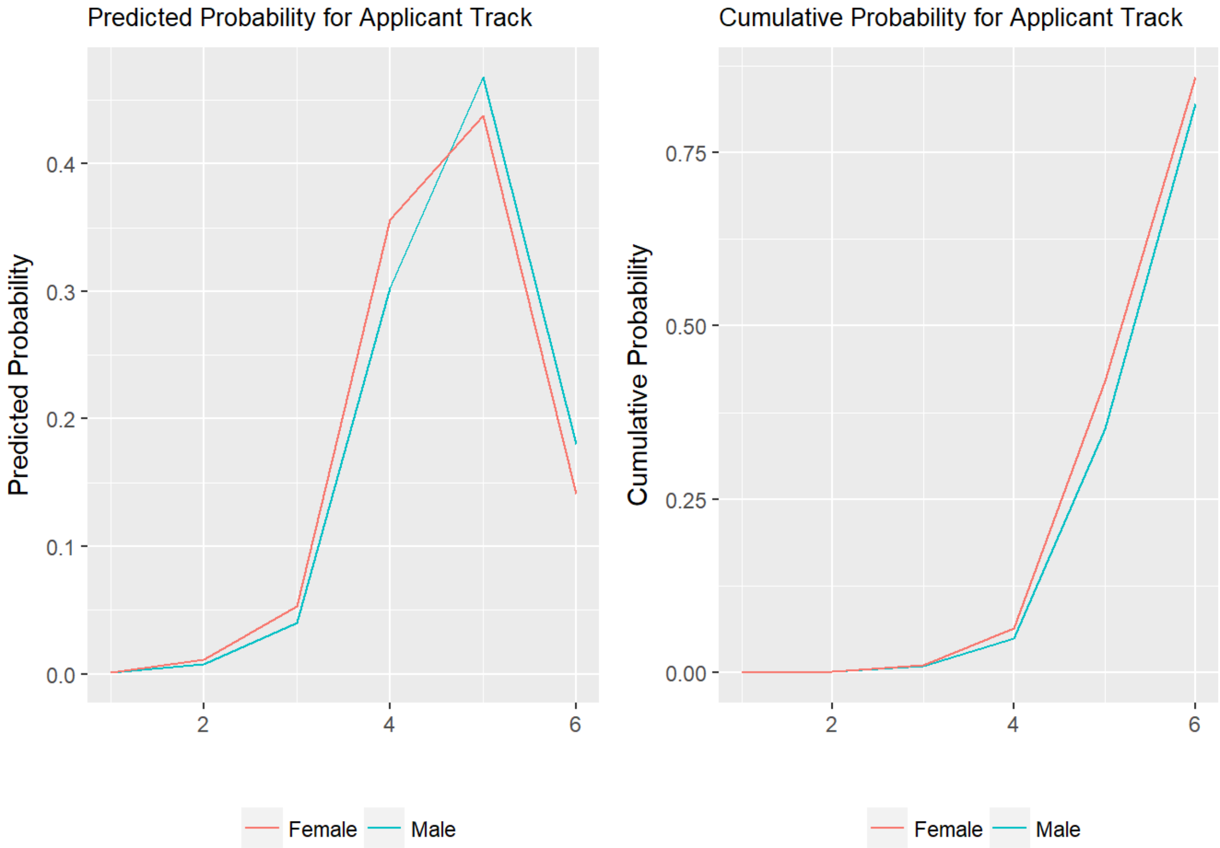


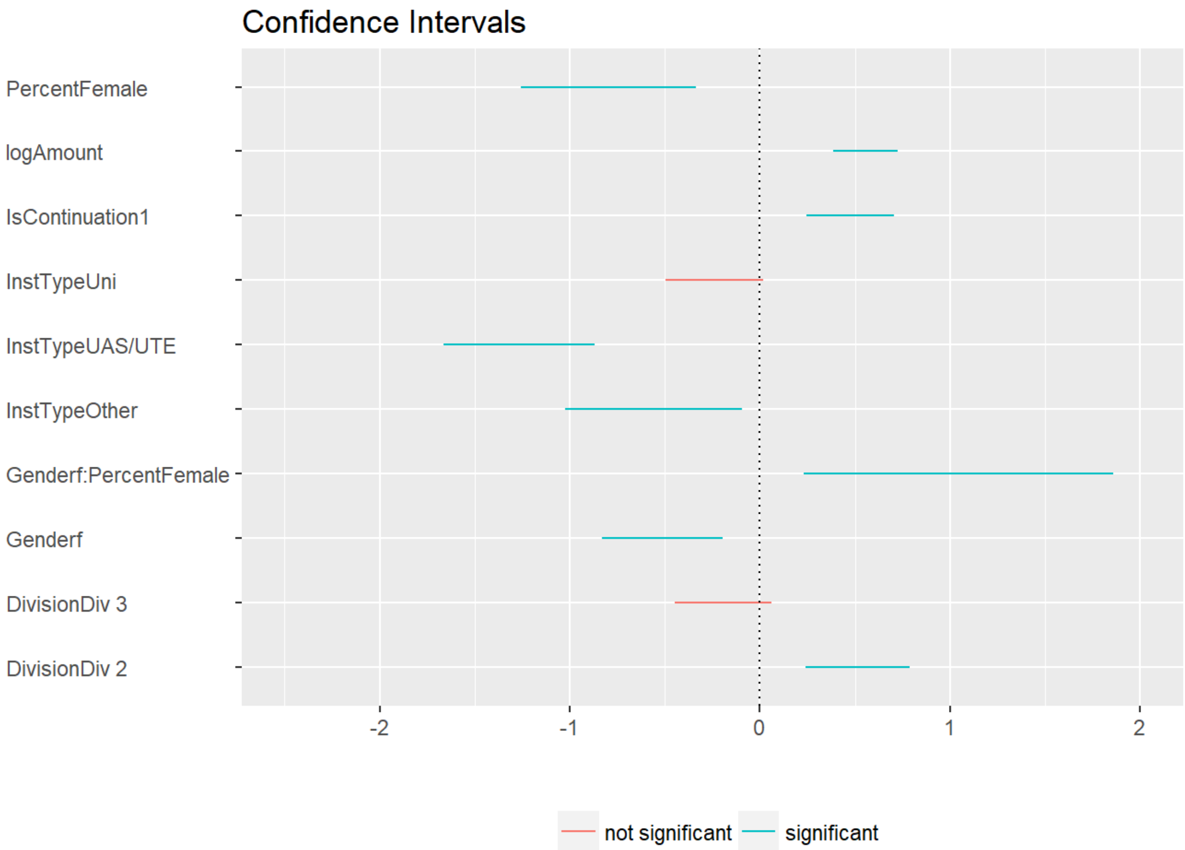
From the plot above we can see that gender seems to be significant, even if the confidence interval upper bound is really close zero. The other significant predictors are Division, the percentage of female reviewers, if the project is a continuation, the Institution type and the log(AmountRequested).

- **Applicant Track assessment**: The final model we used has ApplicantTrack as a response variable, and the following predictors: Gender, Division, PercentFemale, IsContinuation, InstType, log(AmountRequested) and the interaction between Gender and PercentFemale. Again we fitted the same model without Gender and compare it with the anova() function to the one with gender, we get a p.value of 0.0014313, meaning that for the grades given to the main applicant, gender needs to be considered in the model. In the next table we present part of the summary for this model, to see the full summary refer to the Appendix.

We computed the difference in probability of getting a specific grade for both male and female.

In the table above, we see the probabilty of getting each grade for both male and female and the difference between the two. The average difference of the cumulative probability is here 0.023, very close to zero. From the plot below we see that there is almost no difference between women and men probabilities. This seems to suggest that the small p-value from the likelihood ratio test is not really reliable to establish whether there is gender bias.





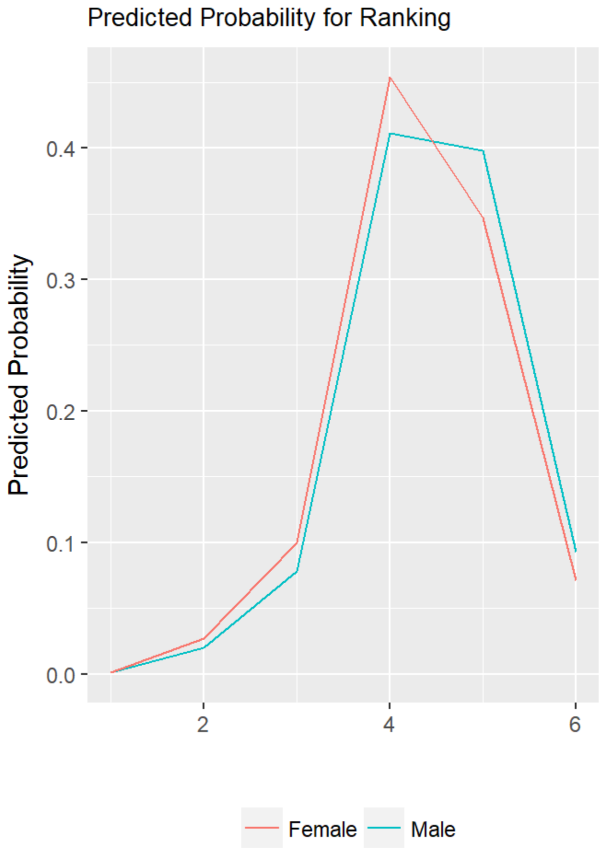
Notice that the confidence interval referring to gender do not include zero and so the corresponding coefficient is significant. The only variables that appear to be not significant in the determination of the Applicant Track grade, are the institution type and the Division. In this model, Division one is used as a base line, and from the results above, we can see that the only impact in grades, from the division point of view, is if the applicant come from Division two. Likewise, from the institution type point of view, there are different chances of reaching higher grades for applicants in institutions other than ETH and Uni.

- **Overall Grade**: This last model has Ranking as a response and Gender, Division, PercentFemale, IsContinuation, InstType, PreviousRequest and logAmount as predictors. We are not considering here the grades given to the applicant track record and to the project, as we just want to see the influence of the demographic data and the project information in each grade. A comparison of this model with the same one without gender may suggest that gender is significant to the model: p.value of 0.11397.

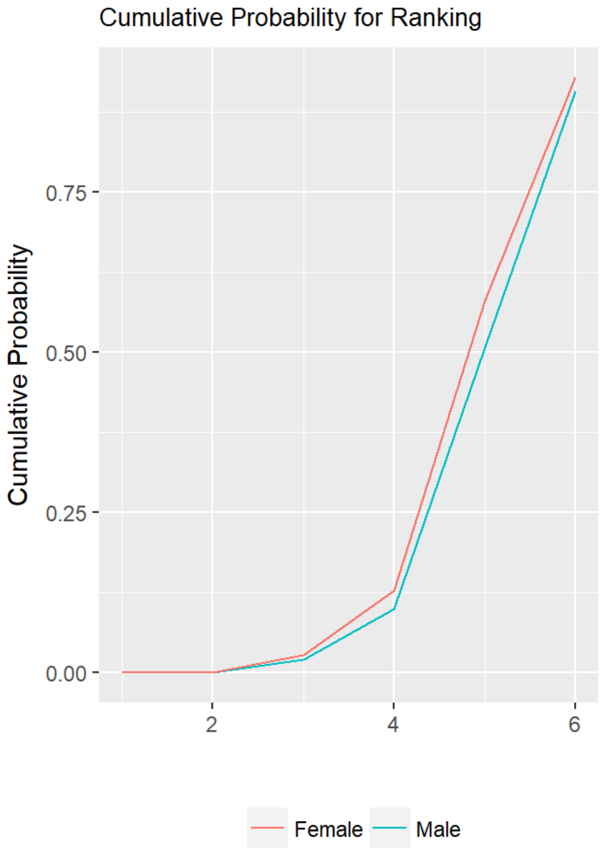
	Male <dbl>	**Female** <dbl>	**Difference** <dbl>
poor	0.001	0.001	0.000
average	0.020	0.027	-0.007
good	0.078	0.100	-0.022
very good	0.411	0.454	-0.043
excellent	0.398	0.347	0.051
outstanding	0.093	0.071	0.022

6 rows

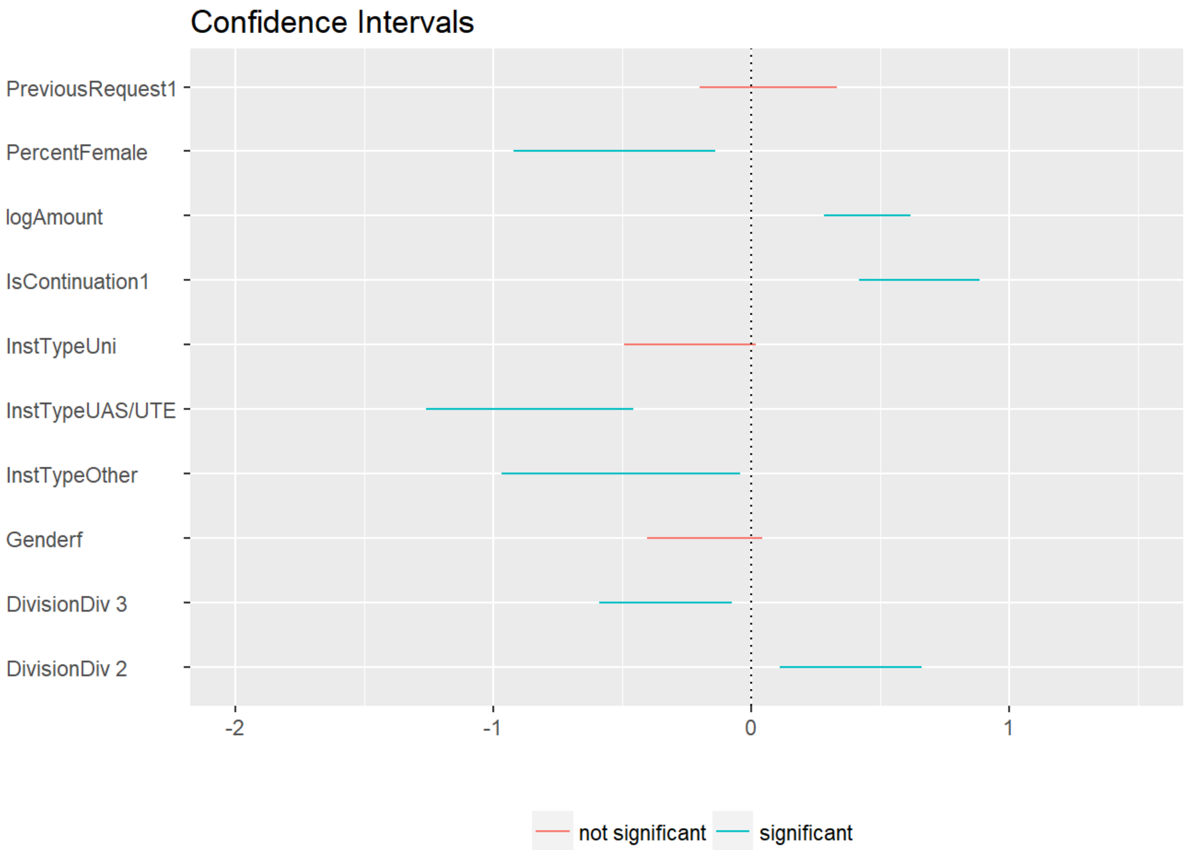
The predicted probabilities of achieving certain grade for male and female is shown in the next table. The average difference of the cumulative probability is here as well close to zero (0.02417). Notice that the only difference is that female applicants are mole likely to get a “good” grade rather than a “very good”, compared to male applicants. From the cumulative probability plot below, we see that the trend is the same for both genders and that the difference is not relevant.



Predicted Probability for Ranking
Predicted Probability
0.0
0.1
0.2
0.3
0.4
2
4
6
Female
Male



Cumulative Probability for Ranking
Cumulative Probability
0.00
0.25
0.50
0.75
2
4
6
Female
Male



Confidence Intervals
PreviousRequest1
PercentFemale
logAmount
IsContinuation1
InstTypeUni
InstTypeUAS/UTE
InstTypeOther
Genderf
DivisionDiv 3
DivisionDiv 2
-2
-1
0
1
not significant
significant

Here gender seems to be not significant, since its confidence interval include zero. However the difference between the upper bound and zero is really small. The other significant variables in this model are the percentage of female referees, log(AmountRequested), Is Continuation, the institution type and Division.

Internal Step

Logistic Regression for approval

Regression data: as explained before, in order to perform the analysis, we combined in a single data frame all the information about applications (IsApproved, Age, Gender, Division, IsContinuation, PreviousRequest, InstType, log(AmountRequested), Semester) and about grades given by the internal referees (ApplicantTrack, ProjectAssessment, Ranking, PercentFemale).

We had again a perfect separation problem, due to the fact that there are very few approved applications with grades worse than “good”. We aggregated the grades which were in category “poor”, “average” and “good” in a unique category for both the applicant track record and the project assessment. All grades are considered as factors. Since the proportion of female referees who evaluate the application takes only values 0, 0.5 and 1, we consider it as factor.

We first fitted a full model with all the available variables and the interactions between Gender and Division, Gender and PercentFemale, Gender and ApplicantTrack. Moreover, we considered the interaction between InstType and Division. We didn’t include the Ranking grades into the model, since they are highly correlated with the single grades for the applicant and the project. We achieved a pseudo-R^2 value of 0.6914, meaning that the variation in the binary variable Y (approved or not) can be explained for more than half by this model including only the internal step information.

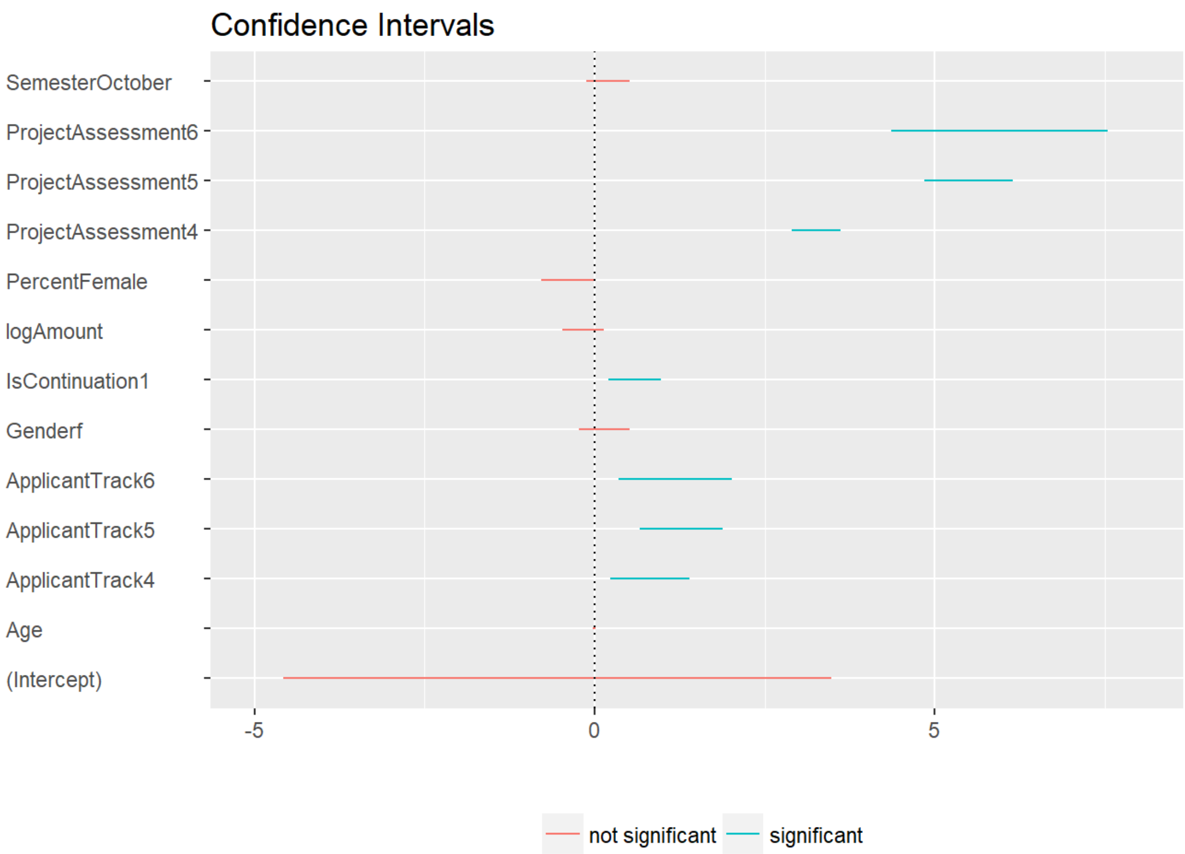
Then, we did variable selection using the AIC criteria in order to obtain a small and effective model as we did for the external step. The remaining predictors are Gender, Age, Semester, IsContinuation, PercentFemale, ApplicantTrack, ProjectAssessment and log(AmountRequested). Interactions were removed from the model because not significant.

```
Call:
glm(formula = IsApproved ~ Gender + Age + Semester + IsContinuation +
    PercentFemale + ApplicantTrack + ProjectAssessment + logAmount,
    family = "binomial", data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6763  -0.4069   0.2040   0.5752   2.6648

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)        -0.55687    2.05505  -0.271  0.78641
Genderf             0.14332    0.19051   0.752  0.45187
Age                -0.01377    0.01017  -1.354  0.17565
SemesterOctober     0.19430    0.16517   1.176  0.23943
IsContinuation1     0.58642    0.19803   2.961  0.00306 **
PercentFemale      -0.38897    0.20067  -1.938  0.05258 .
ApplicantTrack4     0.80144    0.29538   2.713  0.00666 **
ApplicantTrack5     1.25919    0.30685   4.104 4.07e-05 ***
ApplicantTrack6     1.16908    0.42310   2.763  0.00572 **
ProjectAssessment4  3.24147    0.18380  17.636  < 2e-16 ***
ProjectAssessment5  5.46123    0.32697  16.703  < 2e-16 ***
ProjectAssessment6  5.66171    0.77317   7.323 2.43e-13 ***
logAmount          -0.17391    0.15548  -1.118  0.26335
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2243.9  on 1622  degrees of freedom
Residual deviance: 1073.2  on 1610  degrees of freedom
AIC: 1099.2

Number of Fisher Scoring iterations: 6
```

The pseudo R^2 for this model is 0.686 , i.e. this reduced model explains basically the same proportion of variance of the data as the previous model. Even if we removed some variables, the model still explains almost 70% of the variation of the variable IsApproved. It therefore seems that the internal grades are significant predictors for the final funding decision.

In order to assess if gender is a significant predictor for this regression, we computed the 95% confidence intervals for the coefficients of our model. We can see them in the following plot: those that are red include 0 and therefore they are not significant. Those in blue don’t include zero and so they are significant at a 5% level.

```
'data.frame':   13 obs. of  3 variables:
 $ l  : num  -4.5783 -0.2284 -0.0338 -0.1292 0.2008 ...
 $ u  : num  3.48201 0.51915 0.00612 0.51893 0.97778 ...
 $ col: Factor w/ 2 levels "not significant",..: 1 1 1 1 2 1 2 2 2 2 ...
```



The confidence intervals which includes zero are those for the variables: IsContinuation, ApplicantTrack (all levels from 4 to 6) and ProjectAssessment (all levels from 4 to 6). Notice that the coefficient for gender is not significant, since its confidence interval (-0.2284,0.5191) contains zero. The point estimate for gender is 0.1433246 and this means that a woman is 1.1541044 times more likely to be approved than a man.

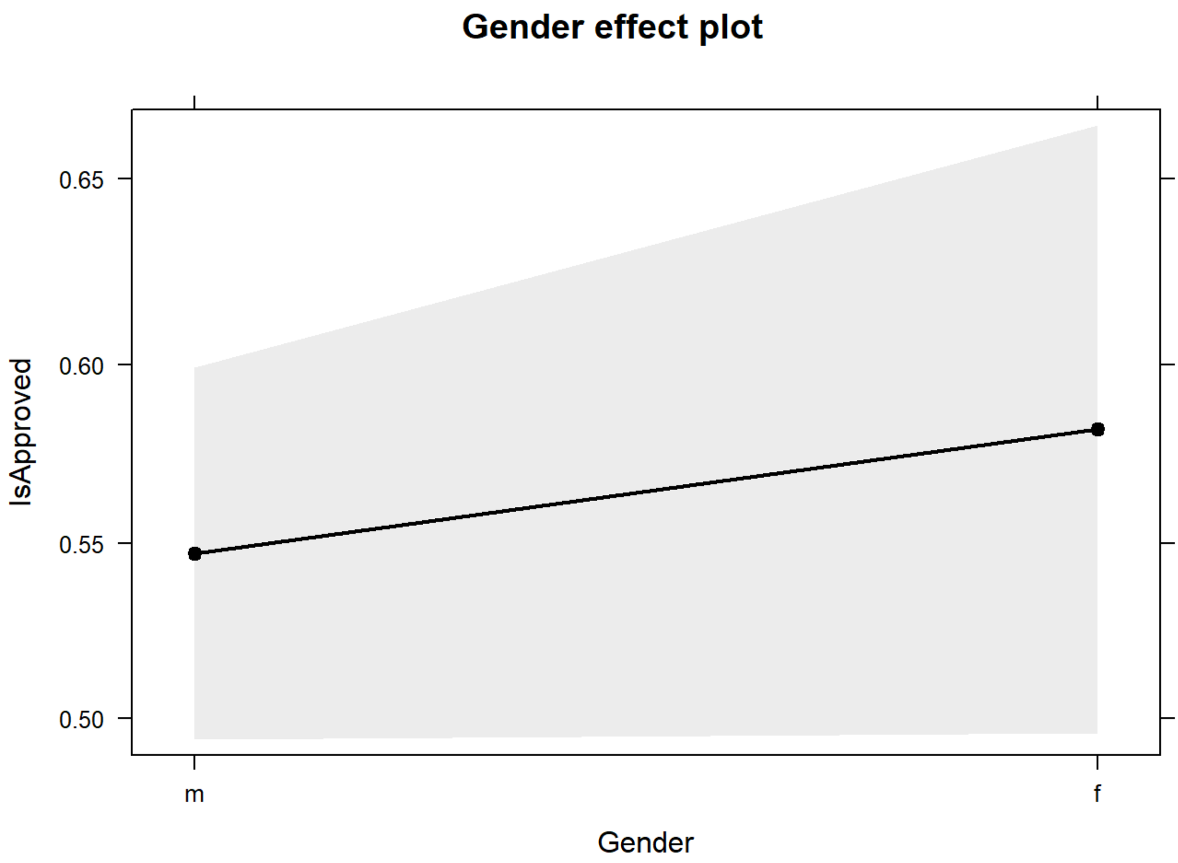
```
     IsContinuation1    ApplicantTrack4    ApplicantTrack5
            1.797546           2.228747           3.522561
     ApplicantTrack6 ProjectAssessment4 ProjectAssessment5
            3.219040          25.571365         235.386692
  ProjectAssessment6
          287.640939
```

We also gave an interpretation to the coefficients of the significant variables:

- If the project is a continuation, it is 1.8 times more likely to be approved than if it’s not
- If the Applicant Track grade is 4, the proposal is 2.23 times more likely to be approved than a 3 or lower
- If the Applicant Track grade is 5, the proposal is 3.52 times more likely to be approved than a 3 or lower
- If the Applicant Track grade is 6, the proposal is 3.22 times more likely to be approved than a 3 or lower
- If the Project Assessment grade is 4, the proposal is 25.57 times more likely to be approved than a 3 or lower
- If the Project Assessment grade is 5, the proposal is 235.39 times more likely to be approved than a 3 or lower

- If the Project Assessment grade is 6, the proposal is 287.64 times more likely to be approved than a 3 or lower

As said before, our focus is the effect of gender in each step of the evaluation process. We checked the difference in the probability of being accepted between male and female: women have probability 0.582 of the project being approved, while men have 0.547. Surprisingly, the probability of being funded seems to be higher for female applicants, even if the difference is too small to be significant. We can clearly see that from the Gender effect plot, where the confidence intervals are overlapping and the line is almost horizontal. From this initial analysis we can say that it doesn't seem that referees are biased against women.



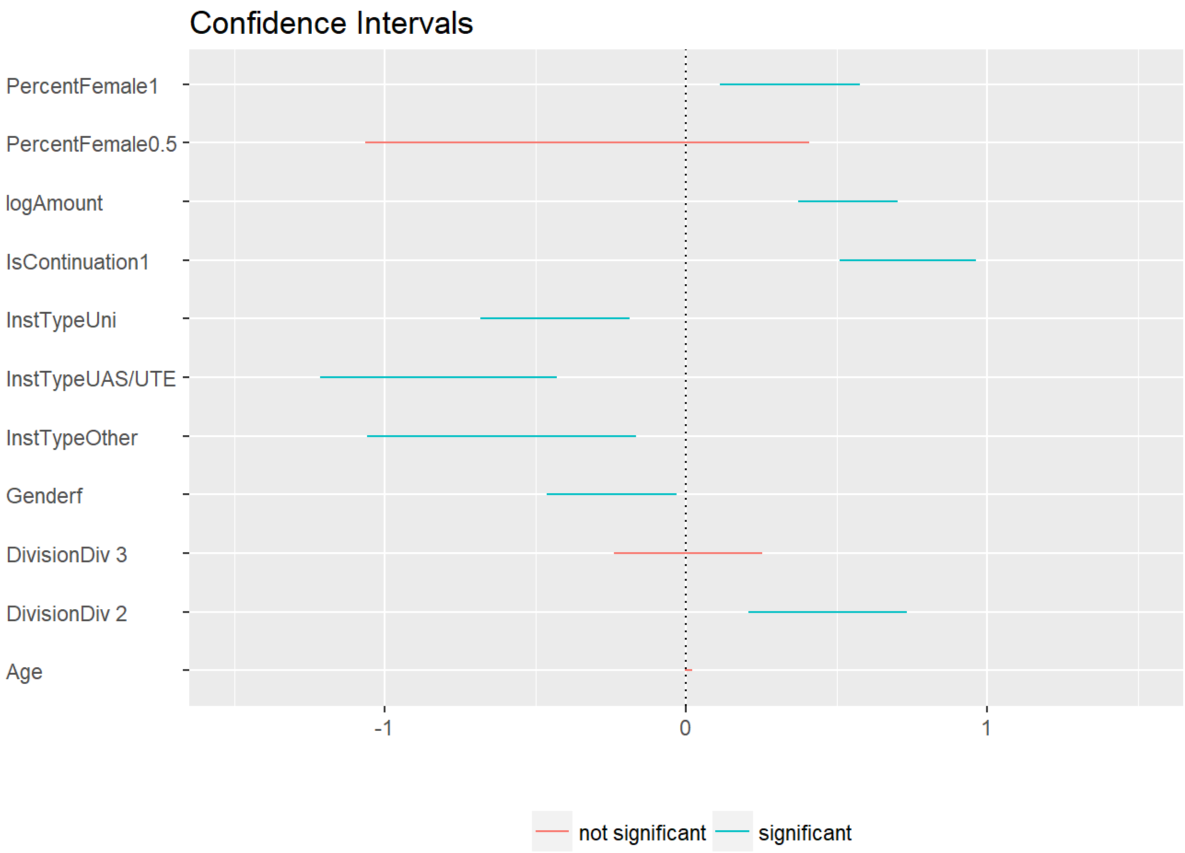
Ordinal Regression

To see which variables influence the different grades in the second step of the evaluation process and check if the gender of the main applicant has an influence on it, we ran an ordinal regression with the function clm of the package ordinal in R. We did this for both, the applicant grade (ApplicantTrack) and grade given to the project (ProjectAssessment). In order to avoid a multiple testing problem, we considered applying a Bonferroni correction, accounting for the number of test performed.

Ordinal regression for Project Assessment

Project Assessment: we fitted the full model with ProjectAssessment as response variable and then selected from the significant variables with the AIC criteria and the help of the drop1() function in R. We end up with a model with the following predictors: Gender, Division, PercentFemale, Age, IsContinuation, InstType and log(AmountRequested). If we fit the same model without Gender and compare it to the one with gender with the anova() function, we get a p.value of 0.0249831, meaning that Gender doesn't seem to be a significant predictor for the grades given to the project by the internal referees.However, if we account for the number of test performed and we apply a Bonferroni correction to this p-value it won't be significant anymore.

Below you see the predicted confidence interval for all the coefficients in our model: all the variables are significant (except for one level of Division and one for PercetFemale). Notice that the upper bound of the gender confidence interval is really close to zero. From this results, we cannot say that there isclear evidence of gender bias.

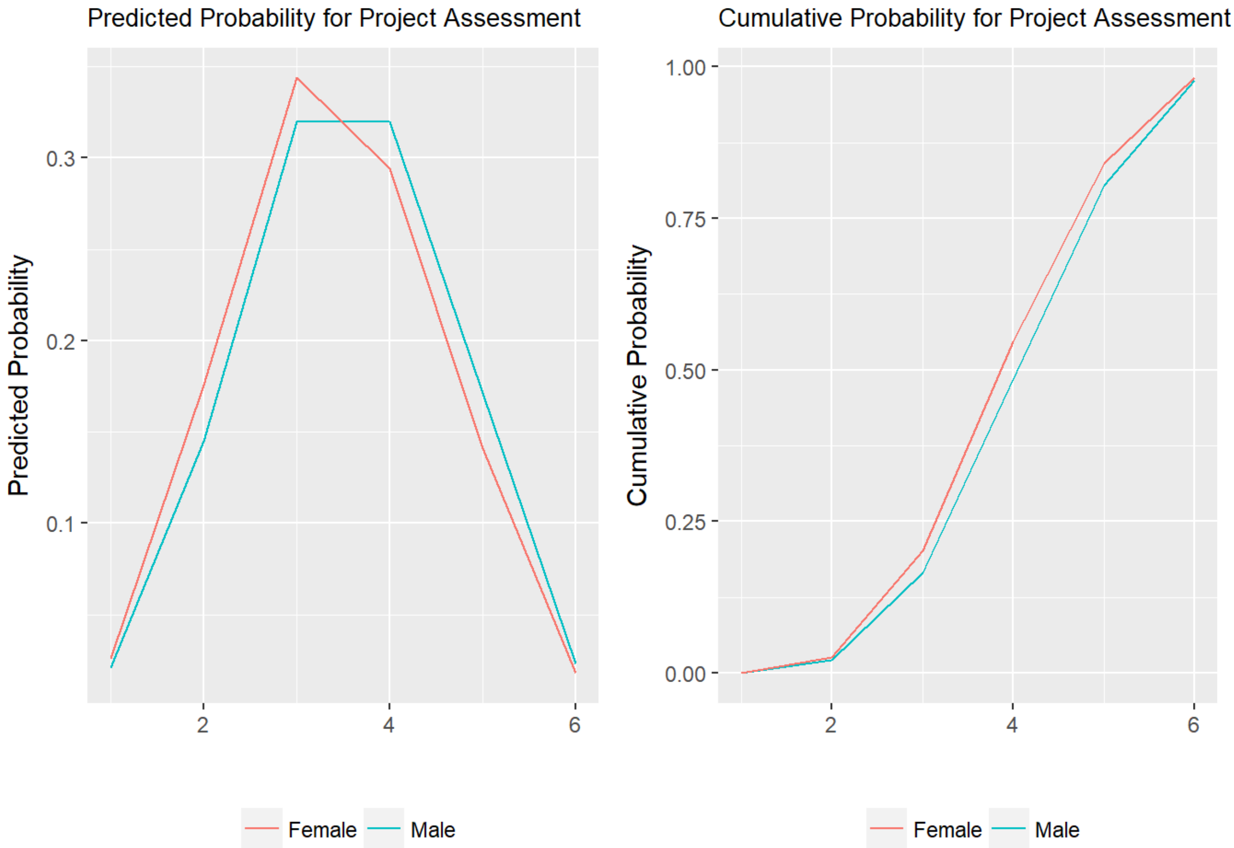


Since we have gender as a predictor in our model, wecan show the effect it has on the grades. For this purpose, we estimated the probability of falling in the different categories for each gender and the difference between men and women. The result is presented in the following table.

	Male <dbl>	Female <dbl>	Difference <dbl>
poor	0.021	0.026	-0.006
average	0.145	0.176	-0.031
good	0.320	0.344	-0.025
very good	0.320	0.294	0.026
excellent	0.171	0.141	0.031
outstanding	0.023	0.018	0.005

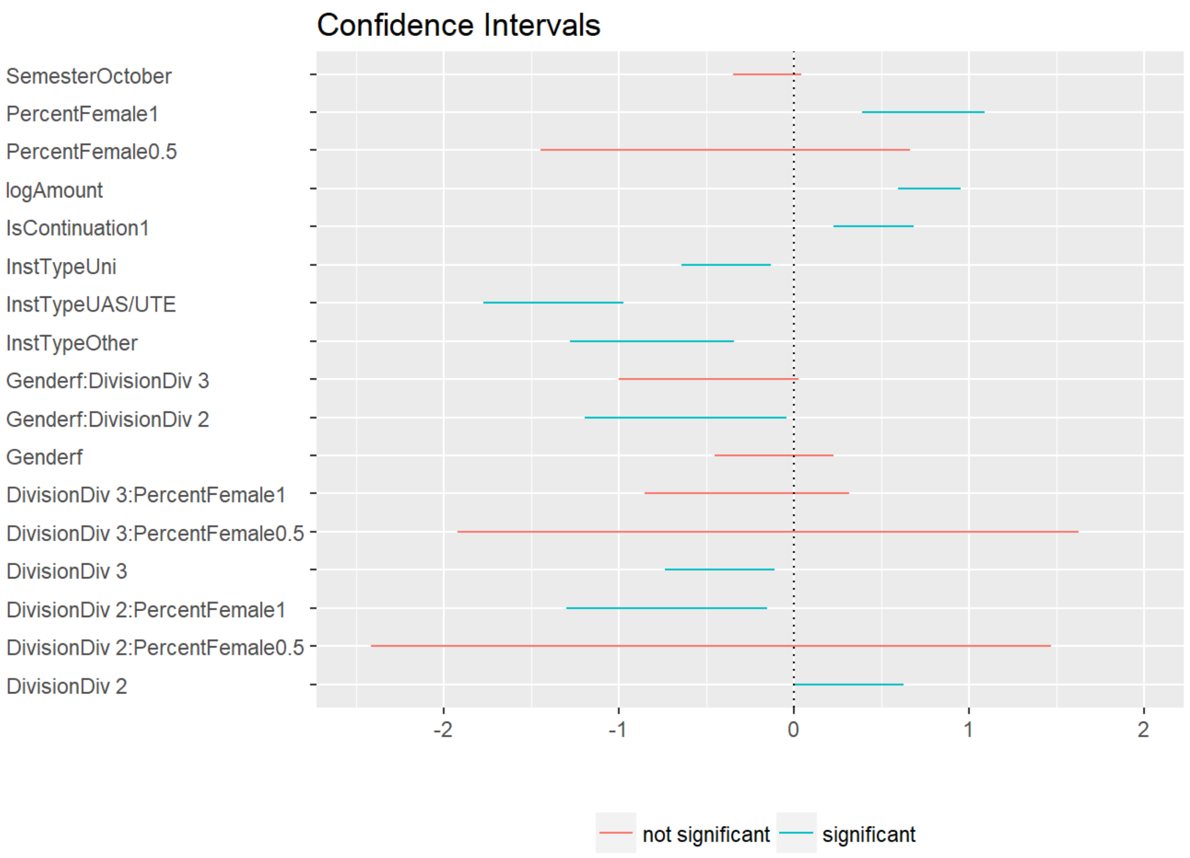
6 rows

Overall the average difference is really small: 0.02067. This seems to suggest that there is no evidence of gender influencing the probability of achiving a certain grade. We also represented in the plots below the probabilty and cumulative probability curves of getting each grade for male and female: they follow more or less the same trend and the only difference, as we've seen from the table above, is that women are slightly more likely to get a “very good” rather than an “excellent”.



Ordinal regression for Applicant Track

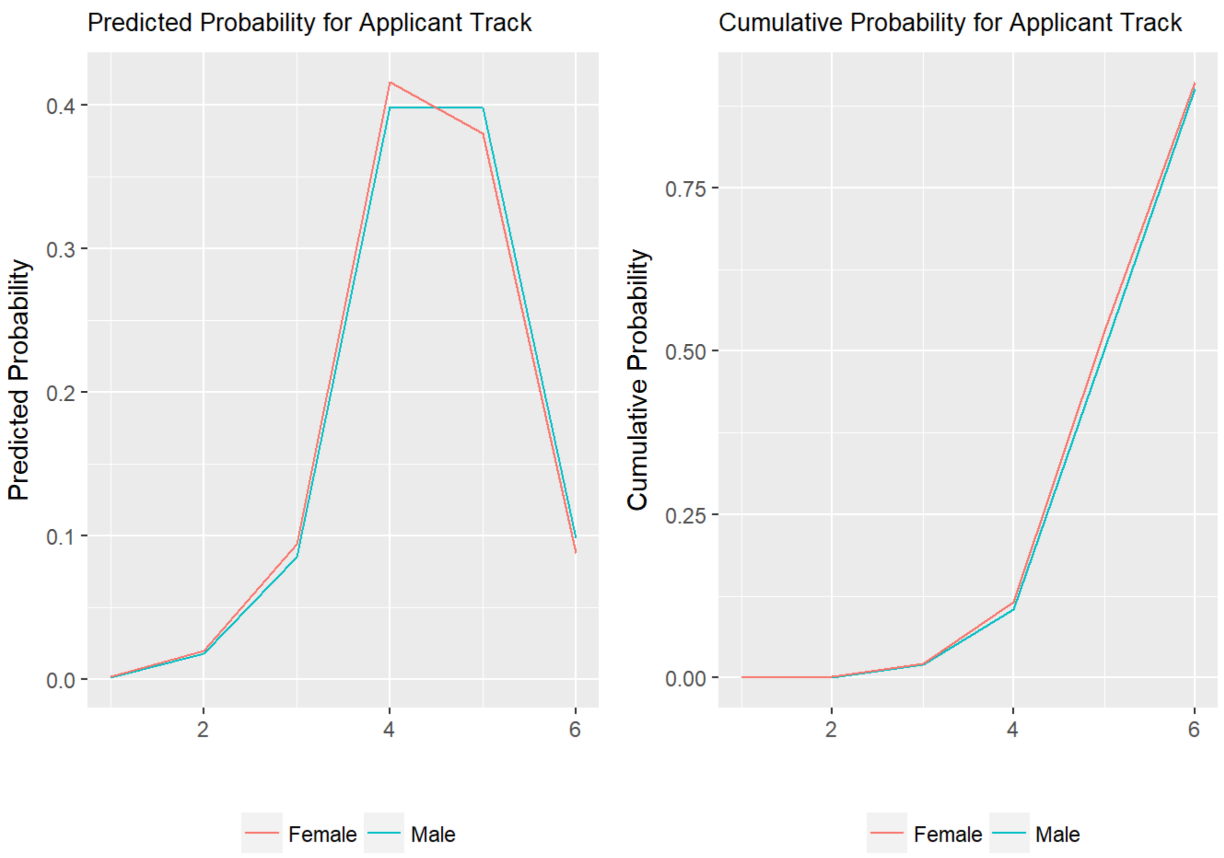
- **Applicant Track assessment**: the model we used has ApplicantTrack as a response variable and the following predictors: Gender, Division, PercentFemale, IsContinuation, InstType, log(AmountRequested), Semester, the interaction between Gender and Divion and the interaction between Division and PercentFemale. Again we fitted the same model without Gender and compare it with the anova() function to the one with gender, we get a p.value of 310^{-4}, meaning that for the grades given to the main applicant track record, gender shouldn't be included in the model. We can also see from the plot below that the confidence interval of gender includes 0 and so the coefficient is not significant. Notice that the significant variables seem to be: Division, Insitution type, IsConitnuation and the percentage of female reviewers. Finally, the interaction between Division and the percentage of female seems to have some importance too.



	Male <dbl>	Female <dbl>	Difference <dbl>
poor	0.001	0.002	0.000
average	0.018	0.020	-0.002
good	0.085	0.094	-0.009
very good	0.399	0.416	-0.017
excellent	0.398	0.380	0.019
outstanding	0.098	0.088	0.010

6 rows

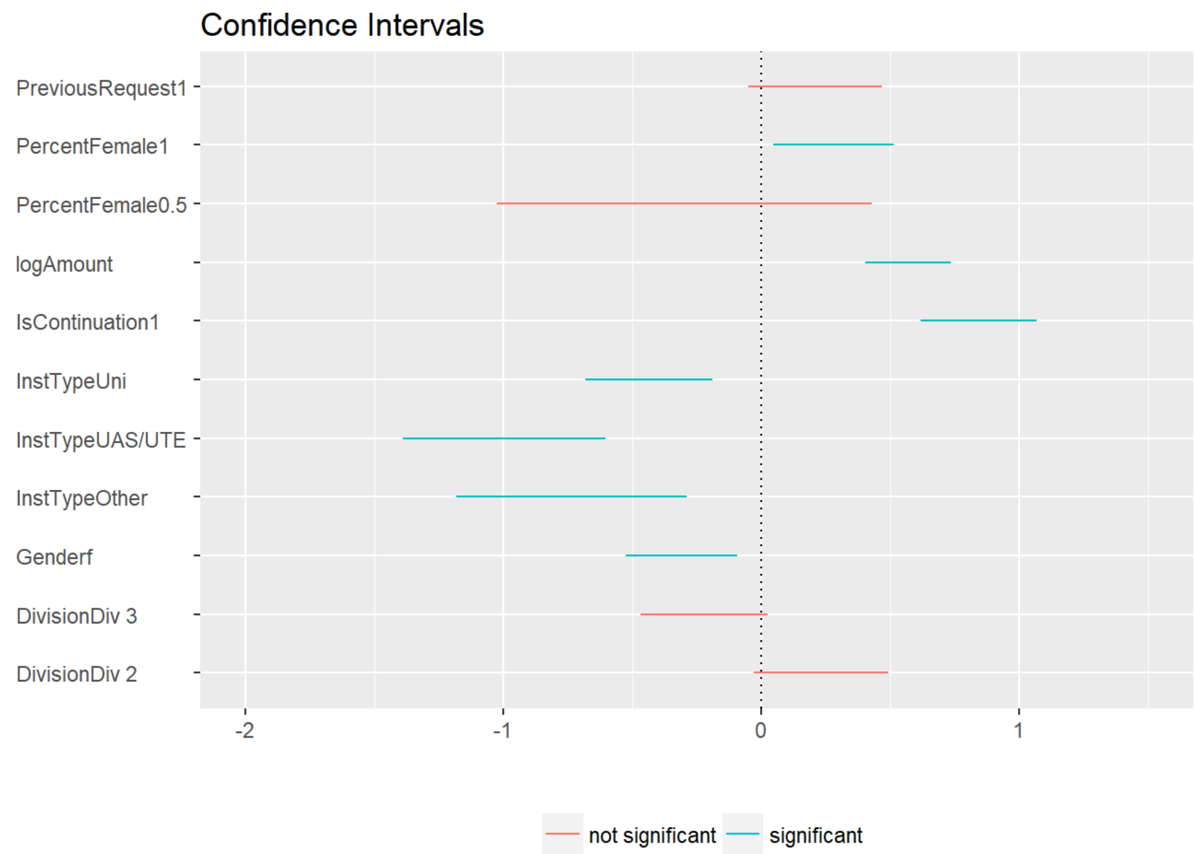
We computed the probability of getting each grade for both male and female and the difference between the two. The average difference of the cumulative probability is here 0.0095, very close to zero. From the plot below we see that there is almost no difference between women and men probabilities. This seems to suggest that the small p-value from the likelihood ratio test is not reliable enough to establish whether there is gender bias.



Ordinal regression for Ranking

Ranking: This last model has Ranking as a response and Gender, Division, PercentFemale, IsContinuation, InstType, PreviousRequest and logAmount as predictors. We are not considering here the grades given to the applicant track record and to the project, as we just want to see the influence of the demographic data and the project information in each grade. A comparison of this model with the same one without gender may suggest that gender is significant to the model: p.value of 0.0048605. However, if we account for the number of tests done the significance disappears immediately.

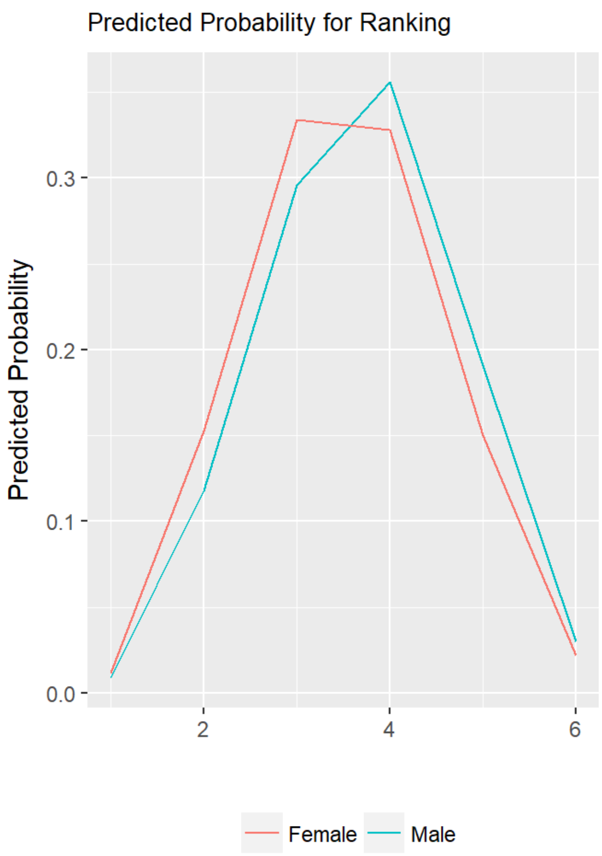
We can also see from the plot below that the upper bound of the confidence interval of gender is not so far from 0. This also suggests that the previous p-value is not completely trustable.

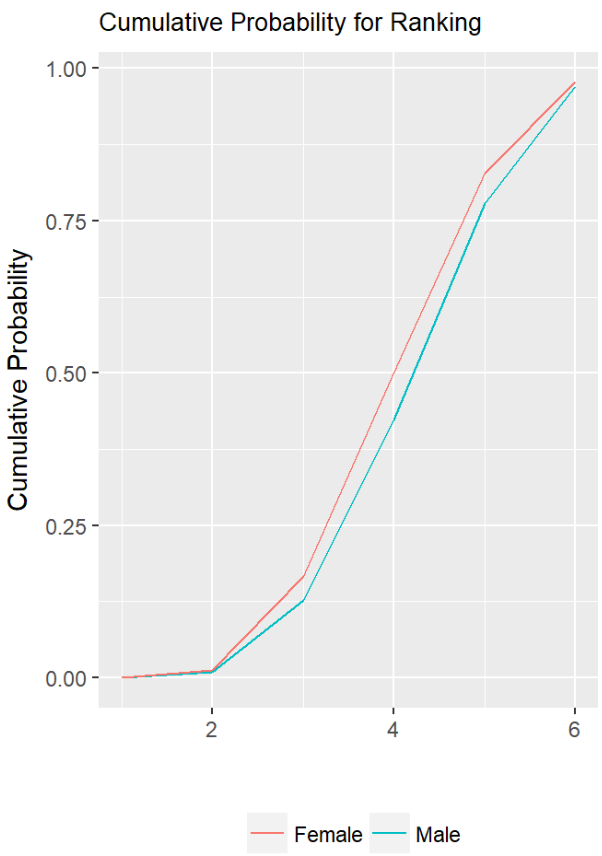


	Male	Female	Difference
	<dbl>	<dbl>	<dbl>
poor	0.009	0.012	-0.003
average	0.118	0.153	-0.035
good	0.296	0.334	-0.038
very good	0.356	0.328	0.028
excellent	0.191	0.150	0.041
outstanding	0.030	0.022	0.008

6 rows

The predicted probabilities of achieving certain grade for male and female is shown in the next table. The average difference of the cumulative probability is here as well close to zero (0.0255). Notice that the only difference is that female applicants are mole likely to get a “good” grade rather than a “very good”, compared to male applicants. From the cumulative probability plot below, we see that the trend is the same for both genders and that the difference is not relevant.





Results

External Step

The external step model is not a good explanation of the variation in the approval of applications (around 42%). We nevertheless look into the influence of gender at this stage, but couldn't find evidence of its effect on the final decision.

When looking at the different grades in this step, we found that gender has a small influence in the Application track assessment, but as this variable is far less important than the project assessment in the determination of the Overall Grade, the influence vanishes out.

Internal Step

The internal model is a quite good explanation of the variation in the approval of applications (around 70% of the variability is explained). When looking at each criteria of evaluation, we found that gender seems to have a small influence in the grade given, however when we checked the variable importance it always turned out to be one of the least important variable in the model.

Considering all the analysis that we have done so far, we can say that there is no evidence of gender bias in the funding decision at the Swiss National Science Foundation. Gender is never an important variable in all the regressions we performed and the difference in predicted probability between male and female applicants is not relevant.

Relative Importance of the Different Steps

Our second research question was to assess the relative importance of each step in the process, and the relative importance of each criteria within each step.

Analysis

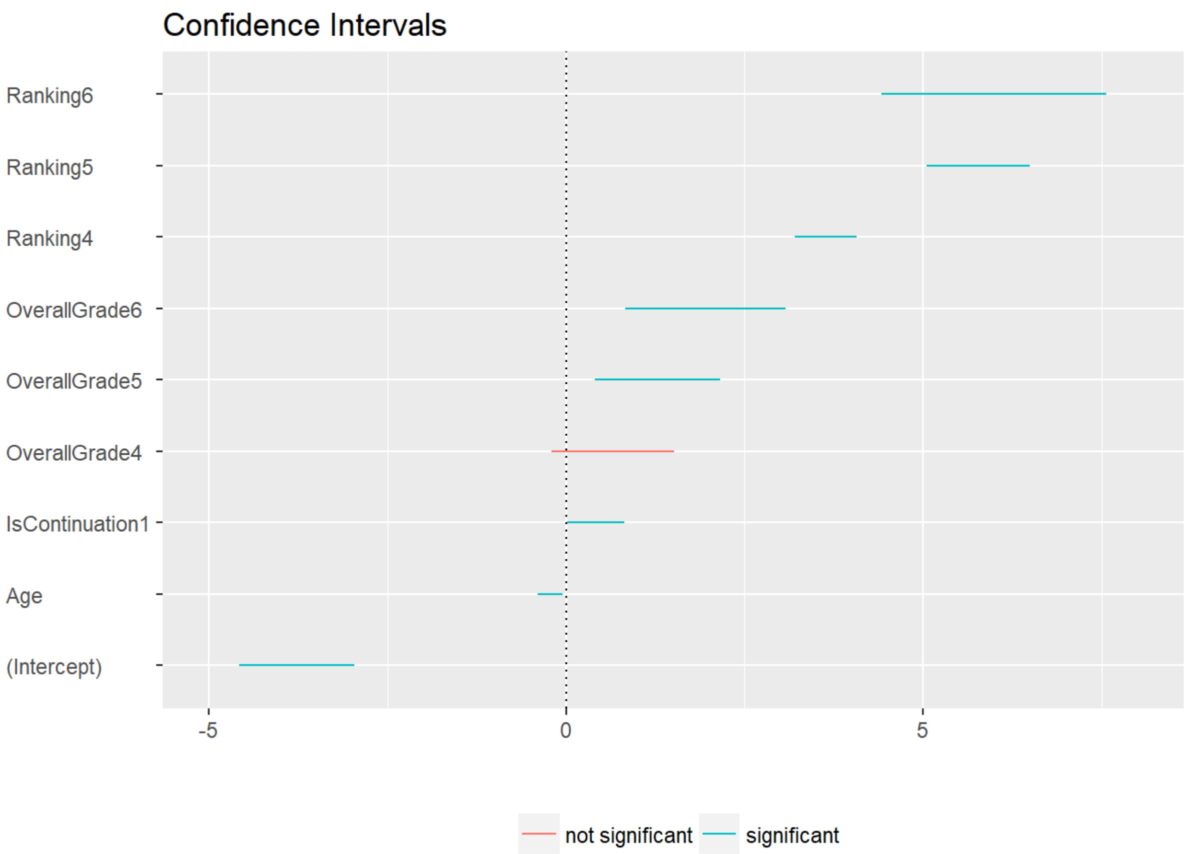
Most Important Step

To approach the question of which step in the process is most important, we first fit a logistic regression with IsApproved as our binary response variable using the glm function in R. We fit a full model with all potential demographic predictors and interactions across the different steps of the process, and the summary grade given to an application in the external (OverallGrade) and internal (Ranking) step. To address the first part of the question (relative importance of each step in the process), we used only the summary grade in each step due to the correlation between the individual grades given within each step and the summary grade given. With the full model (predictors: Gender, Division, Age, IsContinuation, InstType, log(AmountRequested), PercentFemale, Ranking, OverallGrade, Gender:Division, PercentFemale:Gender), we achieved a pseudo-R^2 value of 0.7251, indicating that percent of the variation in Y can be explained by the model.

As our goal was to explain the most important factors, we then did backwards variable selection using the AIC. This left us with a model with only 4 predictors: Ranking, OverallGrade, Age, and IsContinuation. The pseudo R^2 measure of this model is 0.7234, which indicates that this simplified model nearly explains exactly as much variance in the data as the full model, and so we can be content to use just the small model.

Now that we've reduced our model to 4 predictors, we wanted to understand exactly how important each of those predictors are to the final funding decision. To do this, we looked looked at the confidence intervals of the coefficients to see which had the largest impact. To do this, we needed to first standardize our continuous variables (Age). When we plot the confidence intervals of our coefficients, we can see that the Ranking has by far the largest coefficinet, and thus the biggest impact on the final funding decision.

```
'data.frame':   9 obs. of  3 variables:
 $ l  : num  -4.5798 -0.3937 0.0238 3.2037 5.0482 ...
 $ u  : num  -2.9578 -0.0533 0.8199 4.0746 6.4937 ...
 $ col: Factor w/ 2 levels "not significant",..: 2 2 2 2 2 2 1 2 2
```



Most Important Criteria Within Each Step

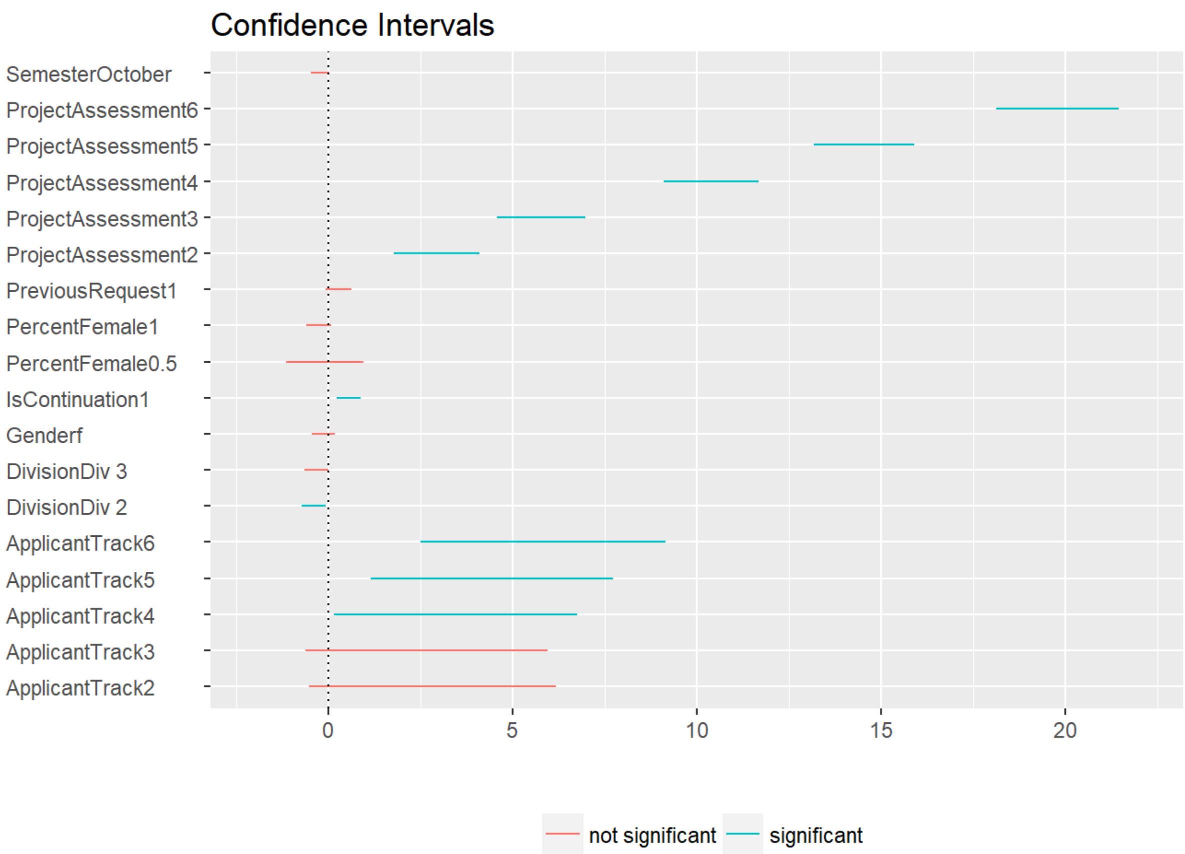
The second aspect of this question was to identify what was the most important criteria within each step. To understand this, we again did a permutation test of the different predictors determining the summary grade given in each the external and the internal review step. We used the Ordinal Regressions from earlier: one for the external OverallGrade using the demographic data, Scientific Proposal grade, and Applicant Track grade as predictors, and a second one predicting the Ranking using the demographic data, Scientific Proposal grade, and Applicant Track grade as predictors.

Internal step

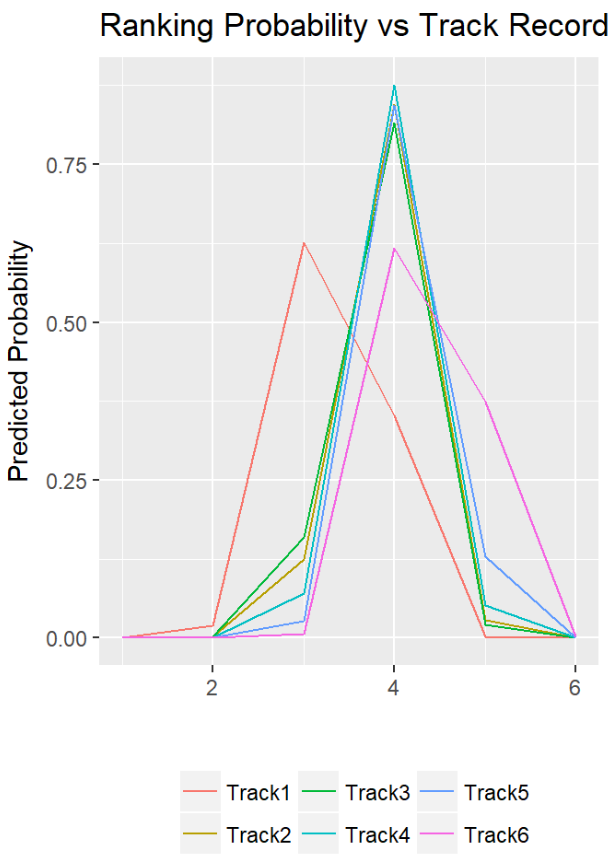
In order to check which variable has the biggest influence on the final grade given by the internal referees, we fitted an ordinal regression using the “Ranking” grade as multinomial response and the Applicant Track grade, the Project Assessment grade and some demographic data as predictors. We started as always fitting the full model and we did variable selection using the AIC. The explanatory variables included in the final model are: ApplicantTrack, ProjectAssessment, Gender, Division, PercentFemale, IsContinuation, PreviousRequest and Semester.

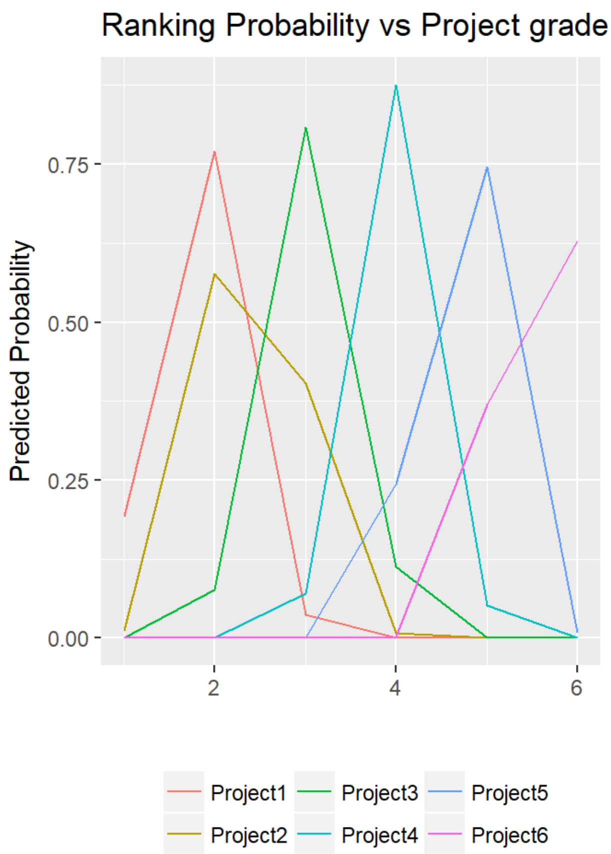
We also computed the predicted probabilities for each of the Ranking grade, varying the Applicant Track grade (plot on the left) and the Project Assessment grade (plot on the right). It is clear that the same variation in grade implies different changes in probabilities: when the project grade changes, the probability variation is much bigger. This confirms what we found before: the quality of the project has a greater influence on the final ranking, compared to the track record.

```
    formula:
    Ranking ~ Gender + Division + PercentFemale + ProjectAssessment + ApplicantTrac
k + IsContinuation + PreviousRequest + Semester
    data:    data

     link  threshold nobs logLik  AIC     niter max.grad cond.H
     logit flexible  1623 -903.45 1852.89 8(0)  9.63e-10 1.5e+04

    Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
    Genderf            -0.13388    0.15488  -0.864 0.387379
    DivisionDiv 2      -0.38768    0.16597  -2.336 0.019497 *
    DivisionDiv 3      -0.32121    0.17131  -1.875 0.060784 .
    PercentFemale0.5   -0.09689    0.54083  -0.179 0.857818
    PercentFemale1     -0.25787    0.16968  -1.520 0.128572
    ProjectAssessment2  2.89349    0.58765   4.924 8.49e-07 ***
    ProjectAssessment3  5.72942    0.61107   9.376  < 2e-16 ***
    ProjectAssessment4 10.34641    0.65075  15.899  < 2e-16 ***
    ProjectAssessment5 14.50192    0.68990  21.020  < 2e-16 ***
    ProjectAssessment6 19.73523    0.84756  23.285  < 2e-16 ***
    ApplicantTrack2     2.53814    1.67610   1.514 0.129946
    ApplicantTrack3     2.37557    1.64346   1.445 0.148326
    ApplicantTrack4     3.15947    1.64503   1.921 0.054781 .
    ApplicantTrack5     4.15085    1.64861   2.518 0.011810 *
    ApplicantTrack6     5.55011    1.66405   3.335 0.000852 ***
    IsContinuation1     0.55162    0.17049   3.235 0.001214 **
    PreviousRequest1    0.28234    0.17894   1.578 0.114596
    SemesterOctober    -0.22630    0.13094  -1.728 0.083948 .
    ---
    Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Threshold coefficients:
        Estimate Std. Error z value
    1|2    2.008      1.640   1.225
    2|3    6.697      1.739   3.852
    3|4   11.229      1.754   6.403
    4|5   16.688      1.775   9.404
    5|6   22.650      1.814  12.483
```



From the plot above we can see that the variable whose coefficients are at the biggest distance from 0 is the Project Assessment. The grades given to the track record of the applicant are significant too, since the confidence intervals don't include 0 but we can see that the effect is smaller.





We also computed the predicted probabilities for each of the Ranking grade, varying the Applicant Track grade (plot on the left) and the Project Assessment grade (plot on the right). It is clear that the same variation in grade implies different changes in probabilities: when the project grade changes, the probability variation is much bigger. This confirms what we found before: the quality of the project has a greater influence on the final ranking, compared to the track record.

Results

Budget Cuts

Analysis

Results

Conclusion

Appendix

Detailed Data Description

Applications

- **AmountRequested**: Rounded to the next 10k CHF
- **AmountGranted**: Rounded to the next 10k CHF
- **IsApproved**: 1 if the application was approved, 0 if it was rejected
- **GradeFinal**: Comparative ranking of the application as determined by the evaluation body (the division of the National Research Council). A: “belongs to the 10% best percent”; AB: “10% are worse, 75% are better”; B: “50% are worse, 25% are better”; BC: " 25% are worse, 50% are better“; C”10% are worse, 75% are better“; D:”90% of the applications are better"

- **Division**: Evaluation Body in which the application was evaluated. Division 1 evaluates Social Sciences and Humanities; Division 2 Mathematics, Natural Sciences and Engineering; Division 3 Biology and Medicine

- **MainDiscipline**: as chosen by the applicant from the SNSF discipline list
- **MainDisciplineLevel2**: category in the SNF discipline list grouping disciplines into fields of research
- **CallTitle**: Call for proposals under which the application was submitted. Applications from the same Call are evaluated together, i.e. in competition to each other
- **CallEndDate**: Submission deadline of the Call
- **ResponsibleApplicantAcademicAgeAtSubmission**: Years since the applicant’s PhD at time of submission; data only available since mid 2016
- **ResponsibleApplicantAgeAtSubmission**: Biological age of the applicant at time of submission; data only available since mid 2016
- **ResponsibleApplicantProfessorshipType**: employment situation of the applicant at time of submission; data only available since mid 2016
- **Gender**: of the main applicant
- **NationalityIsoCode**: Nationality of the main applicant
- **IsHasPreviousProjectRequested**: 0 if it is the applicant’s first application at the SNSF, 1 if not
- **InstType**: Type of institution where the applicant is employed
- **IsContinuation**: 1 if the project is a thematic continuation of a previously approved project, 0 if not
- **ProjectID**: Anonymized identifier of the application

Referee Grades

- **Question**: Evaluation criterion
- **QuestionRating**: The (co-)referee’s assessment of the evaluation criterion
- **OverallRanking**: The (co-)referee’s overall comparative ranking of the application. A: “belongs to the 10% best percent”; same scale as the GradeFinal
- **RefereeRole**: Some applications have one referee evaluation, some have two. The role indicates who was the primary and who was the secondary referee (also called co-referee)
- **RefereeGender**

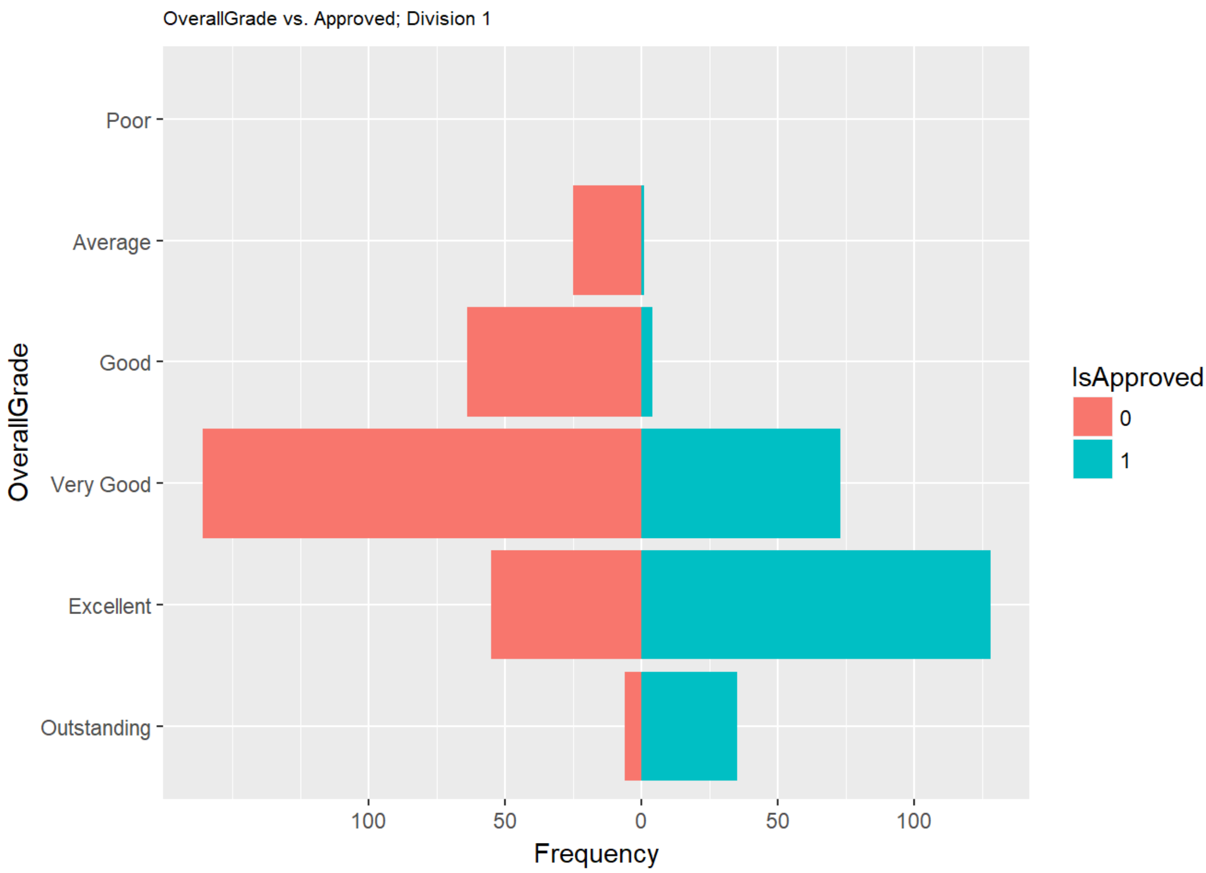
- **IDs**: Anonymized identifiers of the application, the referee and the evaluation by the referee

Reviews

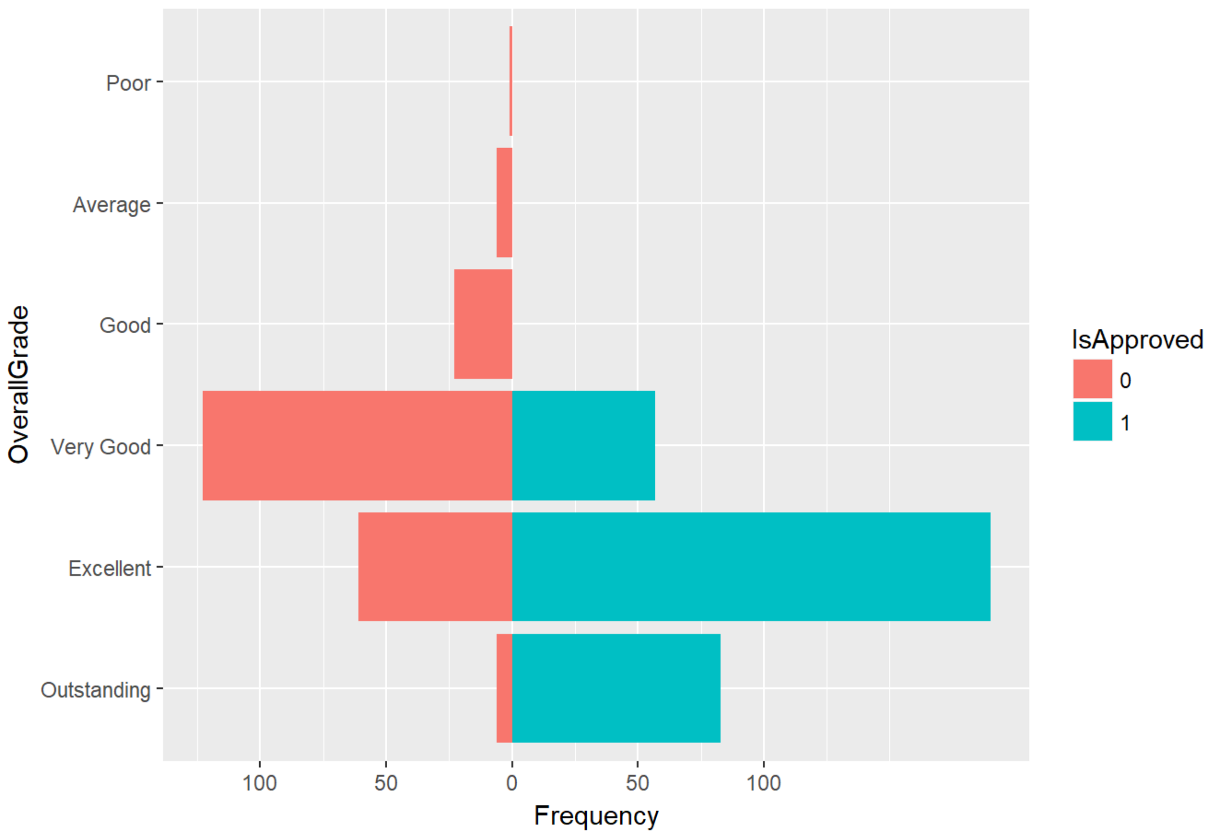
- **Question**: Evaluation criterion
- **QuestionRating**: The external reviewer's assessment of the evaluation criterion
- **OverallGrade**: The external reviewer's overall assessment of the application
- **SourcePerson**: Who suggested the reviewer?
- **Gender**
- **Country**: where the reviewer is located. Not always known
- **EmailEnding**: ending of the reviewer's email address. Might be used as an approximation of the country where the reviewer is located in cases where this data is missing
- **IDs**: Anonymized identifiers of the application, the reviewer and the review

Exploratory Analysis

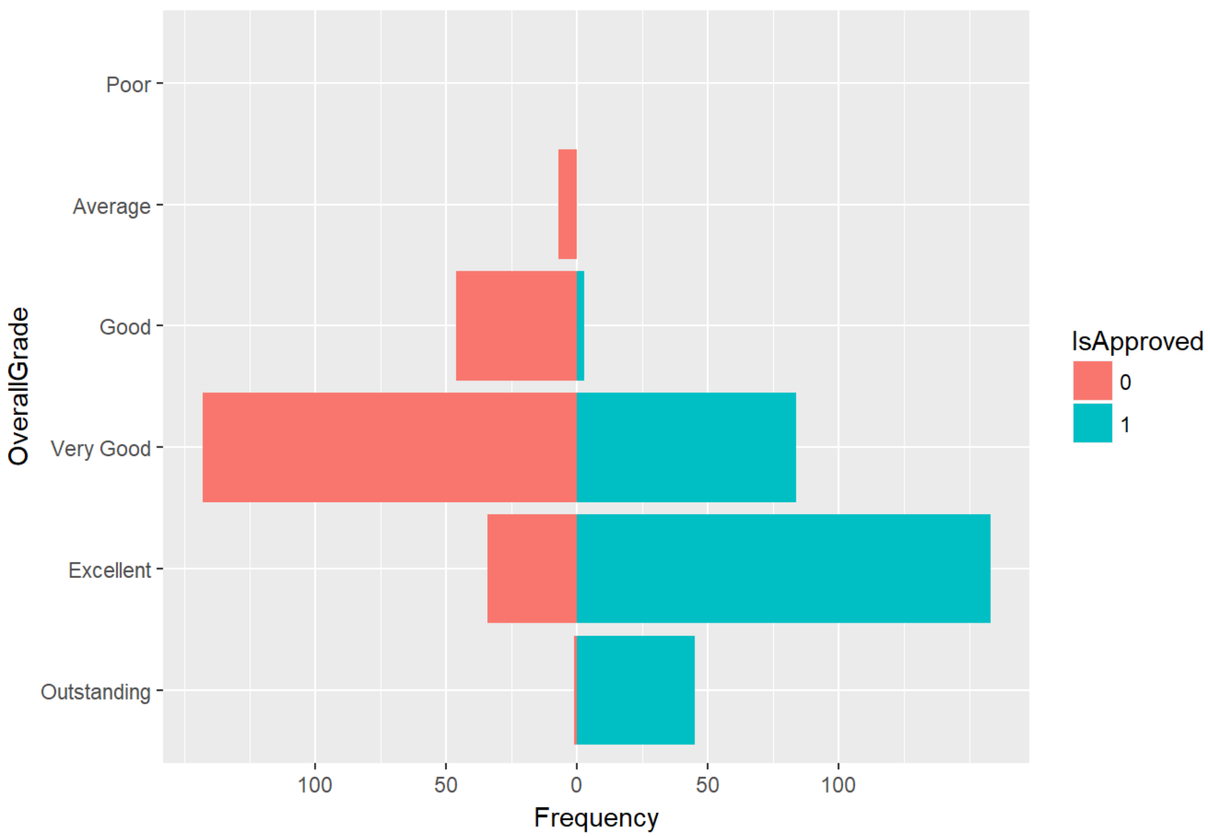
OverallGrade vs. IsApproved, by Division



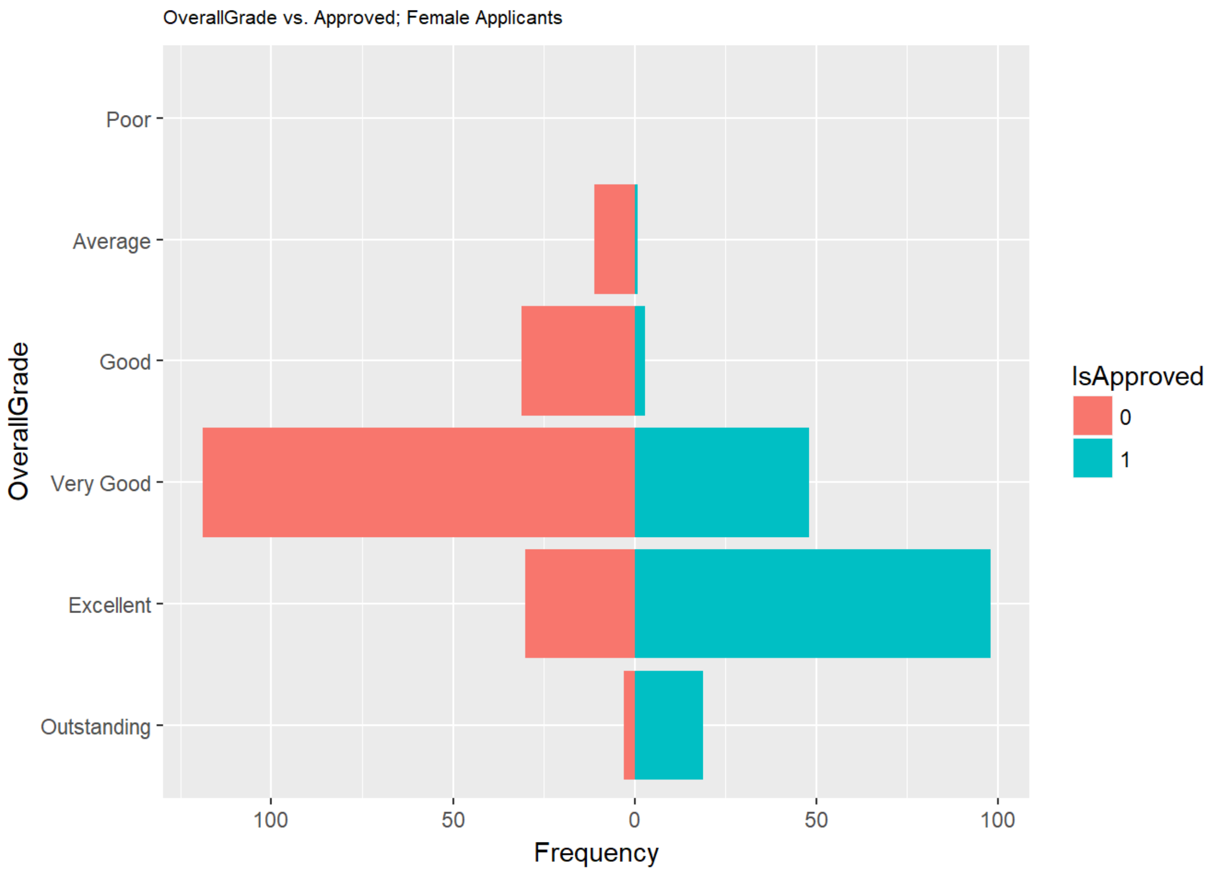
OverallGrade vs. Approved; Division 2

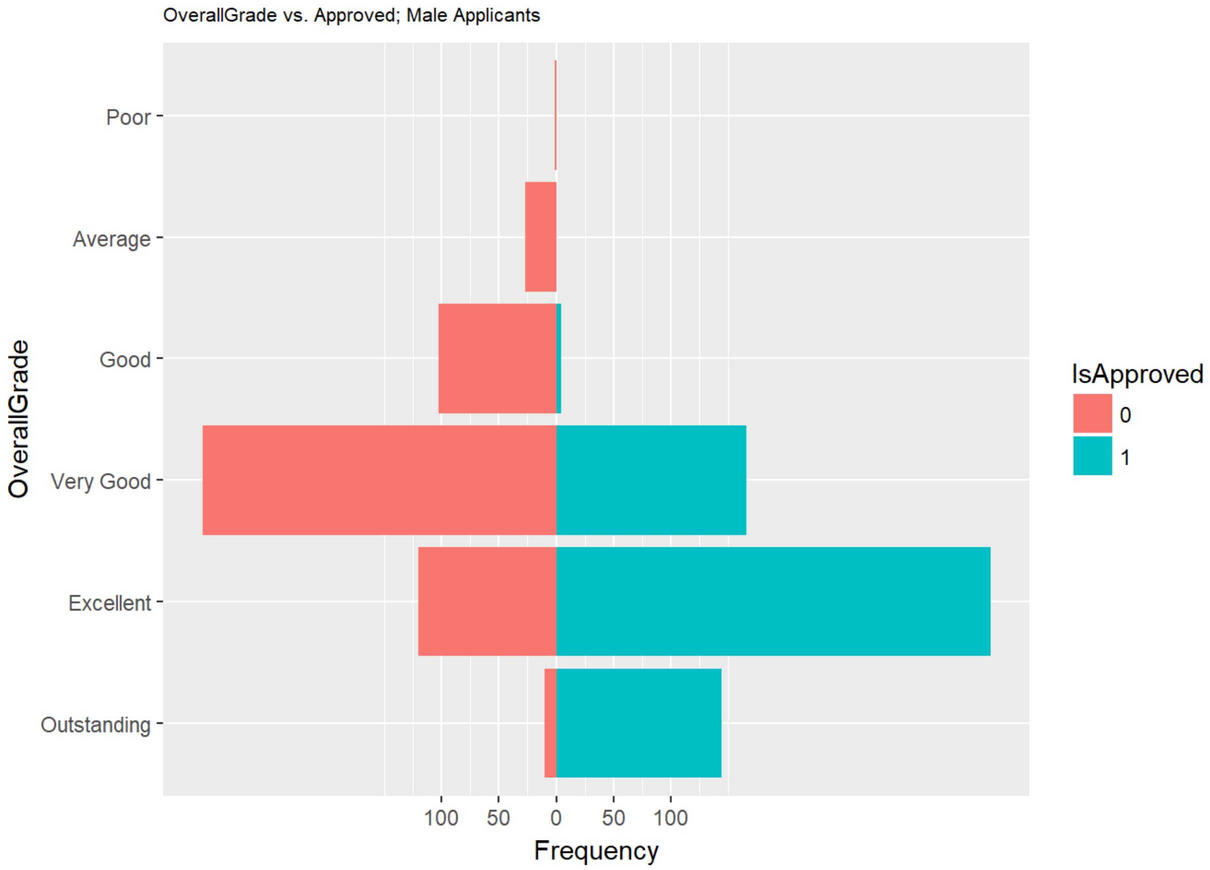


OverallGrade vs. Approved; Division 3



OverallGrade vs. IsApproved, by Gender





External Logistic Regression

- Summary of the final model

```
Call:
glm(formula = IsApproved ~ ApplicantTrack + ScientificRelevance +
    Suitability + PercentFemale + Age + Gender + Division + IsContinuation +
    InstType + Semester + Division:InstType, family = "binomial",
    data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.3819  -0.8248   0.3575   0.7624   2.4032

Coefficients:
                               Estimate Std. Error z value Pr(>|z|)
(Intercept)                   -3.029987   0.761163  -3.981 6.87e-05 ***
ApplicantTrack4                0.969484   0.503628   1.925 0.054229 .
ApplicantTrack5                1.392174   0.510720   2.726 0.006412 **
ApplicantTrack6                2.016852   0.537235   3.754 0.000174 ***
ScientificRelevance4           0.986281   0.391741   2.518 0.011813 *
ScientificRelevance5           1.631665   0.406202   4.017 5.90e-05 ***
ScientificRelevance6           1.776094   0.464162   3.826 0.000130 ***
Suitability4                   0.657609   0.215775   3.048 0.002306 **
Suitability5                   1.781169   0.250140   7.121 1.07e-12 ***
Suitability6                   1.986616   0.390191   5.091 3.55e-07 ***
PercentFemale                 -0.418836   0.261696  -1.600 0.109494
Age                           -0.004531   0.007821  -0.579 0.562333
Genderf                       -0.112332   0.150209  -0.748 0.454557
DivisionDiv 2                 -0.516975   0.431020  -1.199 0.230364
DivisionDiv 3                  0.387478   0.543470   0.713 0.475864
IsContinuation1                0.720271   0.158877   4.534 5.80e-06 ***
InstTypeOther                 -0.174505   0.615954  -0.283 0.776940
InstTypeUAS/UTE               -0.159980   0.464778  -0.344 0.730690
InstTypeUni                   -0.472194   0.423065  -1.116 0.264368
SemesterOctober                0.183584   0.123977   1.481 0.138664
DivisionDiv 2:InstTypeOther    0.321877   0.807478   0.399 0.690174
DivisionDiv 3:InstTypeOther   -0.743797   0.841546  -0.884 0.376779
DivisionDiv 2:InstTypeUAS/UTE -1.416883   0.812385  -1.744 0.081141 .
DivisionDiv 3:InstTypeUAS/UTE -0.191572   0.987872  -0.194 0.846235
DivisionDiv 2:InstTypeUni      0.517898   0.479246   1.081 0.279853
DivisionDiv 3:InstTypeUni     -0.279593   0.569949  -0.491 0.623740
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2243.9  on 1622  degrees of freedom
Residual deviance: 1630.2  on 1597  degrees of freedom
AIC: 1682.2

Number of Fisher Scoring iterations: 5
```

Ordinal external Regression

Project grades (ProposalCombined)

Summary of the final model:

```
    formula:
    ProposalCombined ~ Gender + Division + PercentFemale + IsContinuation + InstTyp
e + logAmount
    data:    external_regression_data

     link  threshold nobs logLik   AIC     niter max.grad cond.H
     logit flexible  1623 -1949.71 3927.42 7(0)  8.06e-10 4.9e+05

    Coefficients:
                    Estimate Std. Error z value Pr(>|z|)
    Genderf         -0.18423    0.11413  -1.614  0.10649
    DivisionDiv 2    0.36126    0.14190   2.546  0.01090 *
    DivisionDiv 3   -0.28708    0.13153  -2.183  0.02907 *
    PercentFemale   -0.43375    0.20023  -2.166  0.03029 *
    IsContinuation1  0.57973    0.11661   4.971 6.65e-07 ***
    InstTypeOther   -0.63461    0.23714  -2.676  0.00745 **
    InstTypeUAS/UTE -0.83721    0.20535  -4.077 4.56e-05 ***
    InstTypeUni     -0.27226    0.13099  -2.078  0.03767 *
    logAmount        0.44651    0.08627   5.176 2.27e-07 ***
    ---
    Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Threshold coefficients:
        Estimate Std. Error z value
    1|2   -1.280      1.310  -0.977
    2|3    1.798      1.114   1.613
    3|4    3.416      1.108   3.084
    4|5    5.709      1.112   5.132
    5|6    8.174      1.123   7.276
```

Odd Ratios and Confidence intervals:

Applicant Track

Summary of the final model:

```
    formula:
    ApplicantTrack ~ Gender + Division + PercentFemale + IsContinuation + InstType
+ logAmount + Gender:PercentFemale
    data:    external_regression_data

     link  threshold nobs logLik   AIC     niter max.grad cond.H
     logit flexible  1623 -1895.84 3821.69 7(0)  7.99e-12 5.1e+05

    Coefficients:
                          Estimate Std. Error z value Pr(>|z|)
    Genderf              -0.51334    0.16102  -3.188 0.001433 **
    DivisionDiv 2         0.51317    0.14066   3.648 0.000264 ***
    DivisionDiv 3        -0.19141    0.12973  -1.475 0.140095
    PercentFemale        -0.79751    0.23535  -3.389 0.000703 ***
    IsContinuation1       0.47308    0.11730   4.033 5.51e-05 ***
    InstTypeOther        -0.55931    0.23642  -2.366 0.017996 *
    InstTypeUAS/UTE      -1.26531    0.20288  -6.237 4.46e-10 ***
    InstTypeUni          -0.23820    0.13069  -1.823 0.068343 .
    logAmount             0.55195    0.08643   6.386 1.70e-10 ***
    Genderf:PercentFemale  1.04403    0.41418   2.521 0.011711 *
    ---
    Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Threshold coefficients:
        Estimate Std. Error z value
    1|2  -0.7523     1.4897  -0.505
    2|3   2.0350     1.1325   1.797
    3|4   3.8149     1.1106   3.435
    4|5   6.1707     1.1115   5.551
    5|6   8.2987     1.1208   7.404
```

Odd Ratios and Confidence intervals:

Overall Grade

Summary of the final model:

```
formula: 
OverallGrade ~ Gender + PercentFemale + Division + IsContinuation + PreviousReq
uest + InstType + logAmount
data:    external_regression_data

 link  threshold nobs logLik   AIC     niter max.grad cond.H 
 logit flexible  1623 -1989.35 4008.69 8(0)  1.46e-12 5.2e+05

Coefficients:
                 Estimate Std. Error z value Pr(>|z|)    
Genderf          -0.18044    0.11418  -1.580  0.11405    
PercentFemale    -0.52959    0.20014  -2.646  0.00814 ** 
DivisionDiv 2     0.38571    0.14093   2.737  0.00620 ** 
DivisionDiv 3    -0.33079    0.13064  -2.532  0.01134 *  
IsContinuation1   0.65198    0.11936   5.462 4.70e-08 ***
PreviousRequest1  0.06592    0.13524   0.487  0.62593    
InstTypeOther    -0.50460    0.23624  -2.136  0.03268 *  
InstTypeUAS/UTE  -0.85837    0.20480  -4.191 2.77e-05 ***
InstTypeUni      -0.23708    0.13043  -1.818  0.06912 .  
logAmount         0.44998    0.08581   5.244 1.57e-07 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold coefficients:
    Estimate Std. Error z value
1|2   -1.877      1.492  -1.257
2|3    1.845      1.120   1.648
3|4    3.478      1.113   3.126
4|5    5.727      1.117   5.129
5|6    7.971      1.126   7.077
```

Odd Ratios and Confidence intervals:

Internal Logistic Regression

- Summary of the final model

```
Call:
glm(formula = IsApproved ~ Gender + Age + Semester + IsContinuation +
    PercentFemale + ApplicantTrack + ProjectAssessment + logAmount,
    family = "binomial", data = data)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.6763  -0.4069   0.2040   0.5752   2.6648

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)        -0.55687    2.05505  -0.271  0.78641
Genderf             0.14332    0.19051   0.752  0.45187
Age                -0.01377    0.01017  -1.354  0.17565
SemesterOctober     0.19430    0.16517   1.176  0.23943
IsContinuation1     0.58642    0.19803   2.961  0.00306 **
PercentFemale      -0.38897    0.20067  -1.938  0.05258 .
ApplicantTrack4     0.80144    0.29538   2.713  0.00666 **
ApplicantTrack5     1.25919    0.30685   4.104 4.07e-05 ***
ApplicantTrack6     1.16908    0.42310   2.763  0.00572 **
ProjectAssessment4  3.24147    0.18380  17.636  < 2e-16 ***
ProjectAssessment5  5.46123    0.32697  16.703  < 2e-16 ***
ProjectAssessment6  5.66171    0.77317   7.323 2.43e-13 ***
logAmount          -0.17391    0.15548  -1.118  0.26335
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2243.9  on 1622  degrees of freedom
Residual deviance: 1073.2  on 1610  degrees of freedom
AIC: 1099.2

Number of Fisher Scoring iterations: 6
```

- Effect plots

Ordinal internal Regressions

Project Assessment

Summary of the final model:

````
```
    formula: 
    ProjectAssessment ~ Gender + Division + PercentFemale + Age + IsContinuation + 
InstType + logAmount
    data:    internal_regression_data

     link  threshold nobs logLik   AIC     niter max.grad cond.H 
     logit flexible  1623 -2351.71 4735.43 6(0)  3.09e-09 7.9e+06

    Coefficients:
                      Estimate Std. Error z value Pr(>|z|)    
    Genderf          -0.247430   0.110411  -2.241 0.025027 *  
    DivisionDiv 2     0.469429   0.133517   3.516 0.000438 ***
    DivisionDiv 3     0.005880   0.125754   0.047 0.962709    
    PercentFemale0.5 -0.323785   0.374886  -0.864 0.387759    
    PercentFemale1    0.342865   0.118669   2.889 0.003862 ** 
    Age               0.008575   0.005778   1.484 0.137757    
    IsContinuation1   0.735298   0.116026   6.337 2.34e-10 ***
    InstTypeOther    -0.612199   0.227930  -2.686 0.007233 ** 
    InstTypeUAS/UTE  -0.822447   0.200334  -4.105 4.04e-05 ***
    InstTypeUni      -0.435290   0.126598  -3.438 0.000585 ***
    logAmount         0.536188   0.084074   6.378 1.80e-10 ***
    ---
    Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Threshold coefficients:
        Estimate Std. Error z value
    1|2    3.094      1.124   2.752
    2|3    5.342      1.112   4.805
    3|4    6.901      1.114   6.193
    4|5    8.379      1.120   7.478
    5|6   10.693      1.135   9.422
```
````

Odd Ratios and Confidence intervals:

Applicant Track

Summary of the final model:

```
formula:
ApplicantTrack ~ Gender + Division + PercentFemale + IsContinuation + InstType
+ Semester + logAmount + Gender:Division + Division:PercentFemale
data:    internal_regression_data

 link  threshold nobs logLik   AIC     niter max.grad cond.H
 logit flexible  1623 -2015.18 4074.36 6(0)  8.98e-07 5.3e+05

Coefficients:
                               Estimate Std. Error z value Pr(>|z|)
Genderf                        -0.11293    0.17342  -0.651  0.51491
DivisionDiv 2                   0.31714    0.15845   2.002  0.04533 *
DivisionDiv 3                  -0.42406    0.15984  -2.653  0.00798 **
PercentFemale0.5               -0.38864    0.53455  -0.727  0.46721
PercentFemale1                  0.74102    0.17834   4.155 3.25e-05 ***
IsContinuation1                 0.45664    0.11667   3.914 9.08e-05 ***
InstTypeOther                  -0.81089    0.23809  -3.406  0.00066 ***
InstTypeUAS/UTE                -1.36995    0.20413  -6.711 1.93e-11 ***
InstTypeUni                    -0.38640    0.12925  -2.990  0.00279 **
SemesterOctober                -0.15198    0.09799  -1.551  0.12092
logAmount                       0.77310    0.09005   8.585  < 2e-16 ***
Genderf:DivisionDiv 2          -0.61647    0.29392  -2.097  0.03596 *
Genderf:DivisionDiv 3          -0.48649    0.26215  -1.856  0.06349 .
DivisionDiv 2:PercentFemale0.5 -0.47640    0.98345  -0.484  0.62809
DivisionDiv 3:PercentFemale0.5 -0.14851    0.90013  -0.165  0.86896
DivisionDiv 2:PercentFemale1   -0.72419    0.29309  -2.471  0.01348 *
DivisionDiv 3:PercentFemale1   -0.26792    0.29780  -0.900  0.36830
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Threshold coefficients:
    Estimate Std. Error z value
1|2    3.146      1.274   2.470
2|3    5.775      1.147   5.035
3|4    7.539      1.140   6.615
4|5    9.699      1.147   8.458
5|6   11.903      1.161  10.250
```

Odd Ratios and Confidence intervals:

Ranking

Summary of the final model:

```
    formula:
    Ranking ~ Gender + PercentFemale + Division + IsContinuation + PreviousRequest
+ InstType + logAmount
    data:    internal_regression_data

     link  threshold nobs logLik   AIC     niter max.grad cond.H
     logit flexible  1623 -2305.97 4643.94 7(0)  2.34e-12 5.1e+05

    Coefficients:
                     Estimate Std. Error z value Pr(>|z|)
    Genderf          -0.30985    0.11009  -2.814 0.004887 **
    PercentFemale0.5 -0.29477    0.36906  -0.799 0.424460
    PercentFemale1    0.27939    0.11925   2.343 0.019132 *
    DivisionDiv 2     0.23153    0.13271   1.745 0.081057 .
    DivisionDiv 3    -0.22028    0.12586  -1.750 0.080084 .
    IsContinuation1   0.84141    0.11480   7.329 2.32e-13 ***
    PreviousRequest1  0.20979    0.13204   1.589 0.112090
    InstTypeOther    -0.73369    0.22738  -3.227 0.001252 **
    InstTypeUAS/UTE  -0.99497    0.20053  -4.962 6.99e-07 ***
    InstTypeUni      -0.43536    0.12620  -3.450 0.000561 ***
    logAmount         0.56949    0.08423   6.761 1.37e-11 ***
    ---
    Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Threshold coefficients:
        Estimate Std. Error z value
    1|2    2.511      1.112   2.258
    2|3    5.266      1.086   4.850
    3|4    6.882      1.088   6.324
    4|5    8.451      1.095   7.717
    5|6   10.677      1.110   9.622
```

Odd Ratios and Confidence intervals:

Relative importance within Internal step

Summary of the final model:

```
    formula:
    Ranking ~ Gender + Division + PercentFemale + ProjectAssessment + ApplicantTrac
k + IsContinuation + PreviousRequest + Semester
    data:    data

     link  threshold nobs logLik  AIC     niter max.grad cond.H
     logit flexible  1623 -903.45 1852.89 8(0)  9.63e-10 1.5e+04

    Coefficients:
                       Estimate Std. Error z value Pr(>|z|)
    Genderf            -0.13388    0.15488  -0.864 0.387379
    DivisionDiv 2      -0.38768    0.16597  -2.336 0.019497 *
    DivisionDiv 3      -0.32121    0.17131  -1.875 0.060784 .
    PercentFemale0.5   -0.09689    0.54083  -0.179 0.857818
    PercentFemale1     -0.25787    0.16968  -1.520 0.128572
    ProjectAssessment2  2.89349    0.58765   4.924 8.49e-07 ***
    ProjectAssessment3  5.72942    0.61107   9.376  < 2e-16 ***
    ProjectAssessment4 10.34641    0.65075  15.899  < 2e-16 ***
    ProjectAssessment5 14.50192    0.68990  21.020  < 2e-16 ***
    ProjectAssessment6 19.73523    0.84756  23.285  < 2e-16 ***
    ApplicantTrack2     2.53814    1.67610   1.514 0.129946
    ApplicantTrack3     2.37557    1.64346   1.445 0.148326
    ApplicantTrack4     3.15947    1.64503   1.921 0.054781 .
    ApplicantTrack5     4.15085    1.64861   2.518 0.011810 *
    ApplicantTrack6     5.55011    1.66405   3.335 0.000852 ***
    IsContinuation1     0.55162    0.17049   3.235 0.001214 **
    PreviousRequest1    0.28234    0.17894   1.578 0.114596
    SemesterOctober    -0.22630    0.13094  -1.728 0.083948 .
    ---
    Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

    Threshold coefficients:
        Estimate Std. Error z value
    1|2    2.008      1.640   1.225
    2|3    6.697      1.739   3.852
    3|4   11.229      1.754   6.403
    4|5   16.688      1.775   9.404
    5|6   22.650      1.814  12.483
```

Odd Ratios and Confidence intervals: