Analysis Report of ClimMob project ProyectoCompleto

Automated Report by ClimMob

19 February, 2020

You are reading a report generated by ClimMob. This is a software package to analyze data generated by citizen science or crowdsourcing.

# Introduction

In agriculture, the local environmental conditions determine to a large degree which technological solutions are the most suitable. In dry soils, for example, drought-resistant crop varieties will outperform other varieties, but in wet soils these same varieties may do worse than most. Not only drought, but an entire range of problems including excessive heat, floods, new pests and diseases tend to intensify under climate change. This multitude of limiting factors requires multiple technological solutions, tested in diverse environments.

Citizen science is based on the cooperation of citizen scientist or observers (paid or unpaid). Researchers assign microtasks (observations, experiments) that, once completed and gathered, contribute with a great amount of information to science. One of the advantages of citizen science is that agricultural researchers can get access to many environments by crowdsourcing their experiments. As farmers contribute with their time, skills and knowledge to the investigation, researchers are able to do more tests than in a traditional setup. Also citizen scientists acquire new knowledge, abilities and information useful for future challenges of their work.

## ClimMob

The primary goal of ClimMob is to help farmers adapt to variable and changing climates. ClimMob was created as part of Bioversity International’s research in the CGIAR Research Programme on Climate Change, Agriculture, and Food Security (CCAFS). It serves to prepare and analyze citizen science experiments in which a large number of farmers observe and compare different technological options under a wide range environmental conditions (van Etten et al. 2019).

ClimMob software assigns a limited number of items (typically 3 crop varieties or agricultural practices) to each farmer, who will compare their performance. Each farmer gets a different combination of items drawn from a much larger set of items. Comparisons of this kind are thought to be a very reliable way to obtain data from human observers. Once the results of the microtasks have been collected, ClimMob builds an image of the whole set of assigned objects, combining all observations. ClimMob not only reconstructs the overall ordering of items, but also takes into account differences and similarities between observers and the conditions under which they observe. It assigns similar observers to groups that each corresponds to a different preference profile. Groups are created on the basis of whichever variables which have been collected, that are found to be significantly linked to the preferences.

ClimMob uses Plackett-Luce models to analyze ranking data (Turner et al. 2020). It automatically generates analytical reports, as well as individualized information sheets for each participant.

## How to cite

If you publish any results generated with ClimMob, you should cite a number of articles as the package builds on various contributions. Van Etten (2019) introduced the crowdsourcing philosophy behind ClimMob. It is important to mention that ClimMob is implemented in R, a free, open-source analysis software (R Core Team 2019). Methodologically, if you report on the tree results, you should mentioned that ClimMob applies the Plackett-Luce model published by Turner et al. (2020). To cite ClimMob itself, mention van Etten et al. (2020).

# Section 1: Headline Results

Overall there were 10 ’participants participating in this study. Each participant assessed 3 different items and ranked them in order of their overall preference. In addition they also provided rankings for 2 additional traits:

|  |  |
| --- | --- |
| Short name | Question |
| overall\_characteristic | Overall Characteristic |
| sobre\_usabilidad | Sobre usabilidad |
| sobre\_comunicación | Sobre comunicación |

Table 1 provides a list of the items assessed within this trial, with the frequency and percentage of participants who assessed each item.

Table 1: Frequency of items assessed

|  |  |  |
| --- | --- | --- |
| Item | Freq | Relative freq |
| Facebook | 7 | 70% |
| Odk Collect | 8 | 80% |
| Telegram | 7 | 70% |
| Whatsapp | 8 | 80% |

## Overall Differences in Rankings

Overall there were statistically significant differences found in the rankings of items in the overall ranking (p=0.00138). The best ranked items overall were Whatsapp, Telegram, Facebook, Odk Collect . Statistically significant differences were also found in the trait(s) Overall Characteristic, Sobre comunicación

A summary of the p-values testing the hypothesis that there exist differences in the rankings within each of the Plackett-Luce models fitted for each of the assessed traits, and the list of items which were significantly highest and lowest ranked overall, are summarised in Table 1.1.

Table 1.1: Summary of differences found in items by trait

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Ranking | Best Ranked | Worst Ranked | p.value | sig |
| Overall | Whatsapp, Telegram, Facebook, Odk Collect | Odk Collect, Facebook, Telegram, Whatsapp | 0.00138 | \*\* |
| Sobre usabilidad | Whatsapp, Telegram, Facebook, Odk Collect | Odk Collect, Facebook, Telegram, Whatsapp | 0.00395 | \*\* |
| Sobre comunicación | Telegram, Whatsapp, Odk Collect, Facebook | Facebook, Odk Collect, Whatsapp, Telegram | 0.18901 |  |

See Section 3 for further details.

## Effect of covariates

None of the variables tested were found to have a statistically significant relationship to the overall ranking at the 5% significance level.

A summary of the univariate signifance levels for all covariates that were able to be tested is shown in Table 1.2.1.

Table 1.2.1: Summary of univariate p-values for first split in Plackett-Luce tree model for the overall ranking

|  |  |
| --- | --- |
| Variable | p |
| ¿Utiliza Whatsapp? | 0.72142 |

See Section 4 for further details.

## Relationships between traits

Table 1.3 shows, for each trait in the study, the frequency with which the rankings matched with the overall ranking.

The trait which had the strongest relationship with the overall ranking was Sobre usabilidad. Overall the rankings for Sobre usabilidad matched the rankings for the overall ranking 73% of the time.

The trait which had the weakest relationship with the overall ranking was Sobre comunicación. Overall the rankings for Sobre comunicación matched the rankings for the overall ranking only 47% of the time.

Table 1.3: Relationship between individual trait assessment and overall assesment

|  |  |
| --- | --- |
| Item | Complete Ranking Agreement |
| Sobre usabilidad | 73.3% |
| Sobre comunicación | 46.7% |

See Section 5 for further details.

# Section 2: Data Summary and Exploratory Analysis of Traits

## Assessment of items

Exploratory analysis within the following section summarises results from the data directly. Given the structure of a ClimMob trail, where each participant only assesses 3 of the 4 possible items these results may be skewed if certain items were randomly assigned to face worse items than others. This is particularly a potential issue within a smaller trial, as due to the randomisation process the potential for an unbalanced assignment decreases as the sample size increases. Results from other sections, and in the overall summary use Plackett-Luce models (Turner et al. 2020), to adjust for any imbalance.

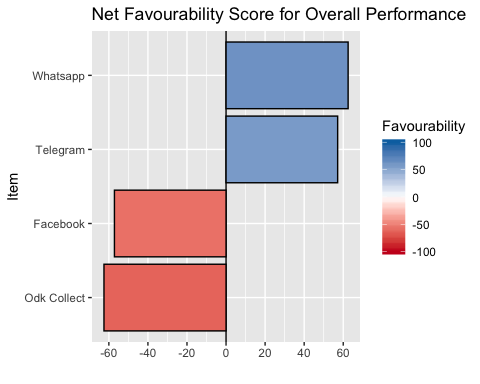
### Overall

Overall performance of each of the items is summarised in Table 2.1.

Table 2.1 - Summary of overall performance

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Item | N | Top Ranked | Bottom Ranked | Contests Won | Net Favourability Score |
| Whatsapp | 8 | 62.5% | 0% | 81.2% | 62.5 |
| Telegram | 7 | 57.1% | 0% | 78.6% | 57.1 |
| Facebook | 7 | 0% | 57.1% | 21.4% | -57.1 |
| Odk Collect | 8 | 12.5% | 75% | 18.8% | -62.5 |

This shows the percentage of participants who assessed the items as their most preferred of the 3 items they were provided, the percentage of participants who included the item as their least preferred, the percentage of ‘head to head contests’ for which the item won and the net favourability score. A score of +100 indicates the item won all ‘contests’ it was involved in, a score of 0 indicates an equal number of wins and losses, a score of -100 indicates the variety lost all contests.

*Figure 2.1* 

The item Whatsapp was the most preferred item overall being ranked highest by 62.5 of the 8 participants who assessed this item.

Other items with strong positive rankings (> 50) were Telegram.

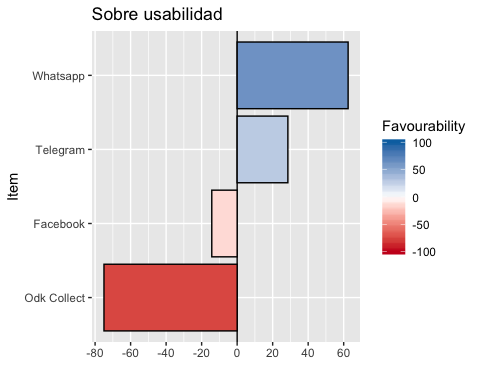
### Other Traits

Net favourability scores are shown below for the other traits assessed in this study.

**Sobre usabilidad**

Table 2.2.1: Favourability statistics for ‘sobre usabilidad’

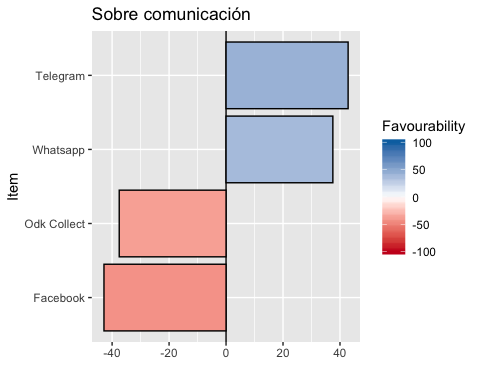
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Item | N | Top Ranked | Bottom Ranked | Contests Won | Net Favourability Score |
| Odk Collect | 8 | 12.5% | 87.5% | 12.5% | -75.0 |
| Facebook | 7 | 14.3% | 28.6% | 42.9% | -14.3 |
| Telegram | 7 | 28.6% | 0% | 64.3% | 28.6 |
| Whatsapp | 8 | 75% | 12.5% | 81.2% | 62.5 |

*Figure 2.2.1* 

**Sobre comunicación**

Table 2.2.2: Favourability statistics for ‘sobre comunicación’

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Item | N | Top Ranked | Bottom Ranked | Contests Won | Net Favourability Score |
| Facebook | 7 | 14.3% | 57.1% | 28.6% | -42.9 |
| Odk Collect | 8 | 0% | 37.5% | 31.2% | -37.5 |
| Whatsapp | 8 | 62.5% | 25% | 68.8% | 37.5 |
| Telegram | 7 | 57.1% | 14.3% | 71.4% | 42.9 |

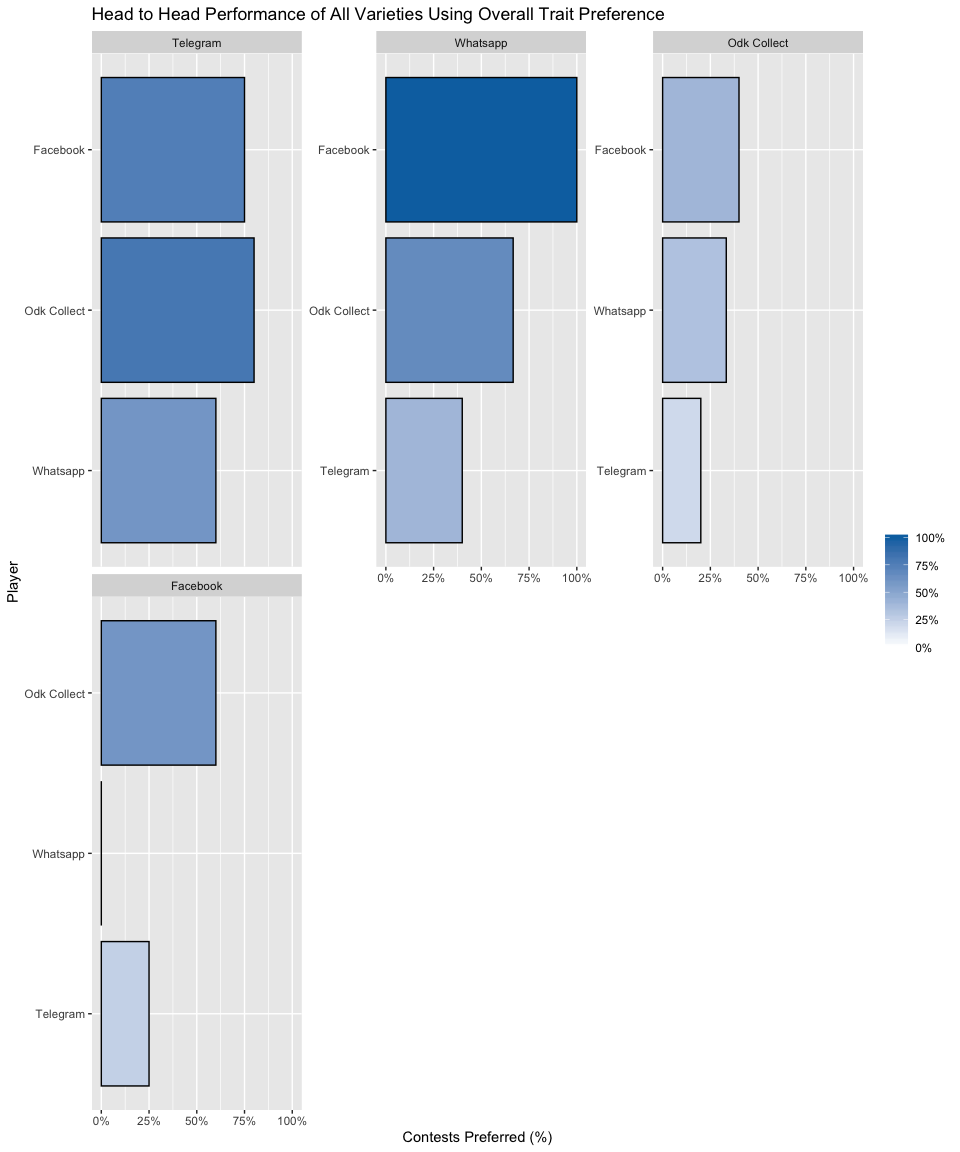
*Figure 2.2.2* 

## Pairwise Contests

Appendix B contains the full data for each pairwise comparison of the varieties, which is summarised in the plots below.

### Overall

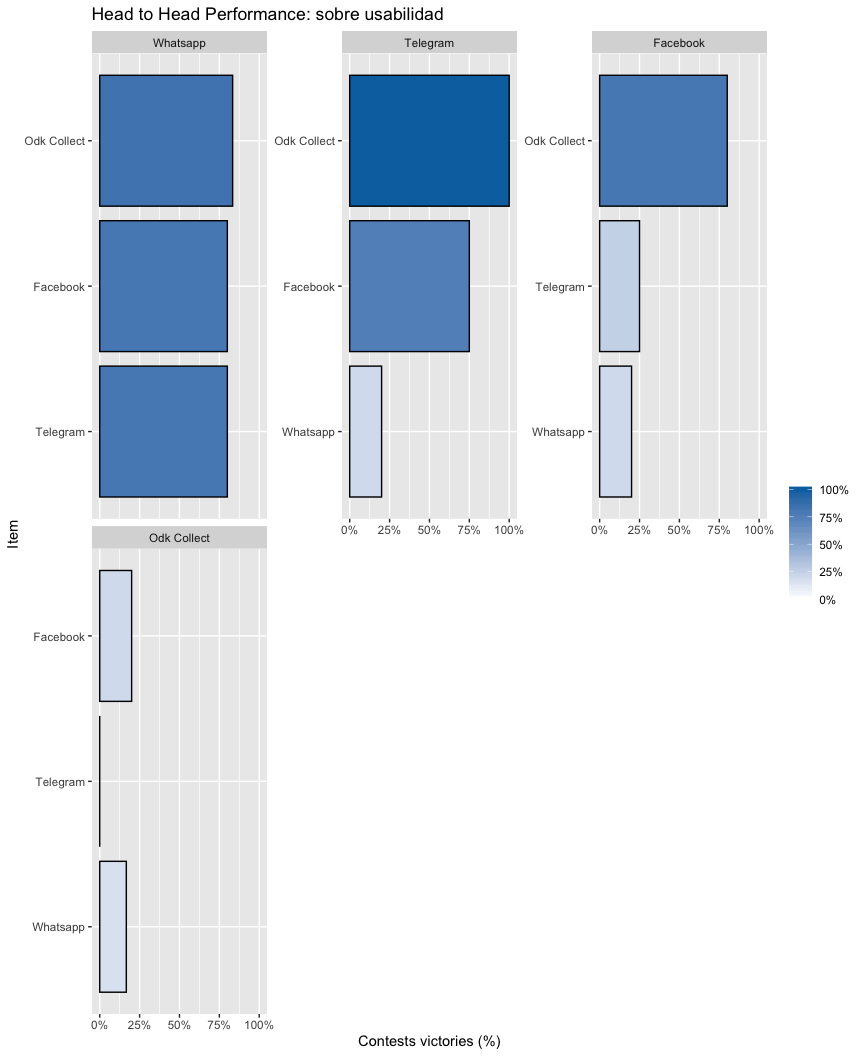
Figure 2.3 shows the outcomes of all pairwise contests between the items included in the trial. Each panel shows the performance of one item against all the other items, and shows the percentage of the times in which the panelled item was ranked above the items shown as bars. The most preferred item is shown in the top left panel and the least preferred is shown in the bottom right panel

*Figure 2.3* 

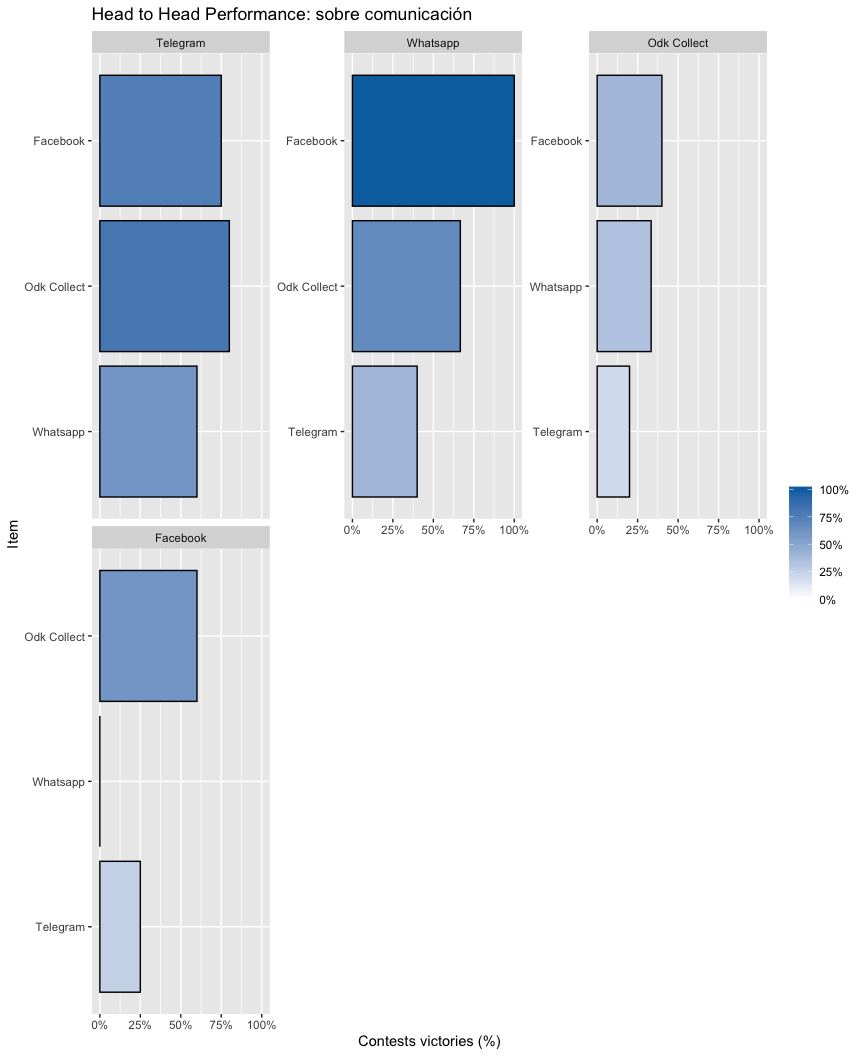
### Other Traits

Results from the pairwise contests of the other individual traits assessed are shown below.

**Sobre usabilidad**

*Figure 2.4.1* 

**Sobre comunicación**

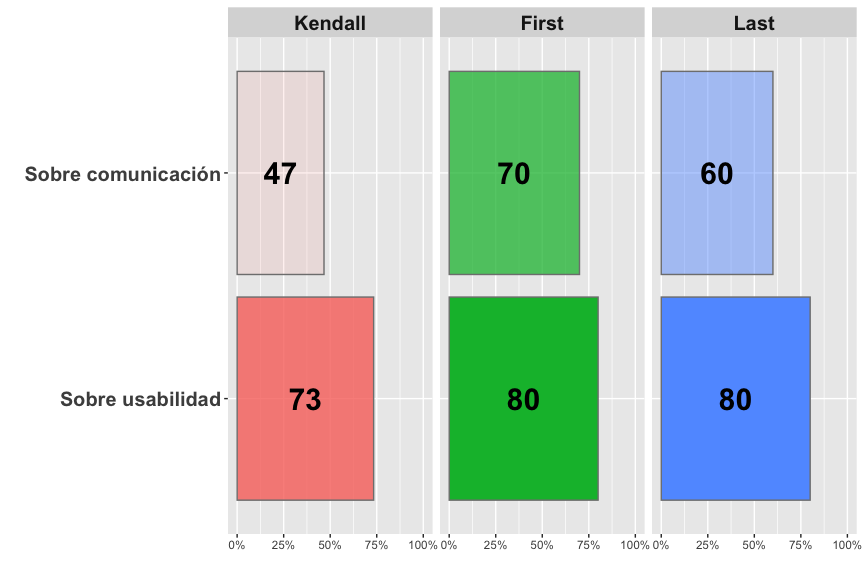
*Figure 2.4.2* 

## Relationship between traits

Table 2.5 shows the relationship between the individual trait rankings and the overall rankings. Complete agreement represents the percentage of respondents for which the ranking of the 3 items in respect to the trait is an exact match to the overall ranking. Best and worst agreement represents the percentage for which the best and worst item for the trait matched the overall best and worst. Complete ranking agreement shows the proportion of correlation on the full ranking with the overall performance as baseline using the Kendall correlation coefficient (Kendall 1938).

Table 2.5: Relationship between individual trait assessment and overall assessment

|  |  |  |  |
| --- | --- | --- | --- |
| Item | Complete Ranking Agreement | Agreement with Overall Best | Agreement with Overall Worst |
| Sobre usabilidad | 73.3% | 80% | 80% |
| Sobre comunicación | 46.7% | 70% | 60% |

*Figure 2.5* 

Overall the trait which had the strongest relationship with the overall ranking was Sobre usabilidad, with identical rankings being given as the overall ranking 73.33% of the time.

# Section 3: Data Summary & Exploratory Analysis of Explanatory Variables

|  |  |
| --- | --- |
| Short name | Question |
| registration\_REG\_p3p | ¿Utiliza Whatsapp? |

## Overall Ranking

Table 4.1 shows the results from the likelihood ratio test from the Plackett-Luce model for overall rankings of the different items. The hypothesis being tested is that there is no difference in the assessments of any of the different items.

Table 4.1: Likelihood ratio test results from overall model

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| model | logLikelihood | DF | Statistic | Pr(>Chisq) |  |
| NULL | -17.91759 | 30 | NA | NA |  |
| Overall | -10.12635 | 27 | 15.58248 | 0.0013808 | \*\* |

p=0.0013808. This suggests that there is evidence of a difference between items

Figure 4.1 shows the estimates of the model coefficients with 84% confidence intervals. The purpose of this graph is to be able to best distinguish between the relative strength of each of the items assessed. As such the coefficient estimates themselves are not directly interpretable, but it can be concluded that a higher value for the coefficient indicates that a variety has been more preferred. The 84% confidence width is chosen so that non-overlapping confidence intervals could be interpreted as indicating significant differences at the 5% significance level. This may not match exactly with the mean separation groupings, as these groupings also take into account multiple testing through the Benjamini and Hochberg adjustment.

Mean separation analysis was also conducted to indicate, using letters, which items are significantly more preferred than others: when items have at least one letter in common, there is not enough evidence from the experiment to be confident about their relative order of preference at the 5% significance level.

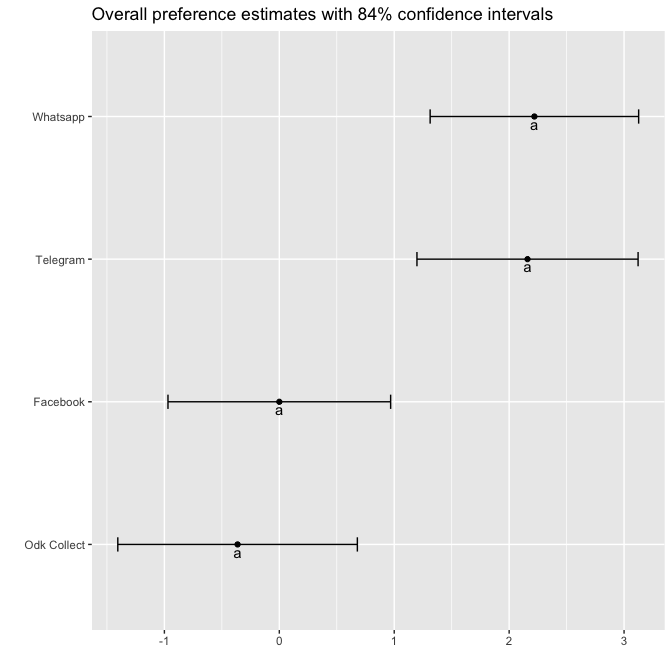


Figure 4.1 - Overall ranking: model coefficients and mean separation.

The same information as Figure 4.1 is shown in Table 4.2 below

Table 4.2 - Model coefficients and mean separation of items at 5% level with BH adjustment

|  |  |  |  |
| --- | --- | --- | --- |
|  | Estimate | quasiSE | Group |
| Whatsapp | 2.22 | 0.65 | a |
| Telegram | 2.16 | 0.68 | a |
| Facebook | 0.00 | 0.69 | a |
| Odk Collect | -0.36 | 0.74 | a |

Table 4.3 and Figure 4.2 use the coefficients from the model to estimate the probability of each item being considered to be the top ranked item in a direct comparison between all of the possible items

Table 4.3: Percentage probability of being the highest ranked overall.

|  |  |  |
| --- | --- | --- |
|  | Item | Win probability |
| Whatsapp | Whatsapp | 47% |
| Telegram | Telegram | 44.3% |
| Facebook | Facebook | 5.1% |
| Odk Collect | Odk Collect | 3.6% |

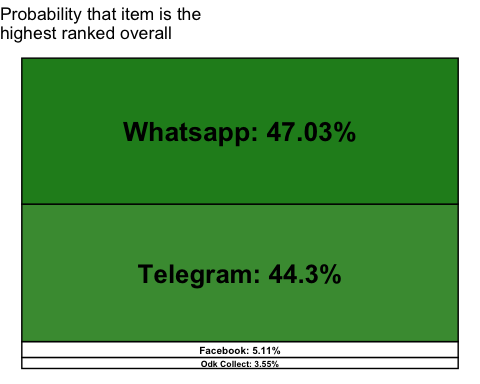


Figure 4.2 - Overall Ranking: Probability of Being The Highest Ranked Overall.

# Section 5: Plackett-Luce Models of Other Traits

**Sobre usabilidad**

Table 5.1.1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| model | logLikelihood | DF | Statistic | Pr(>Chisq) |  |
| NULL | -17.91759 | 30 | NA | NA |  |
| Sobre usabilidad | -11.24701 | 27 | 13.34117 | 0.003954 | \*\* |

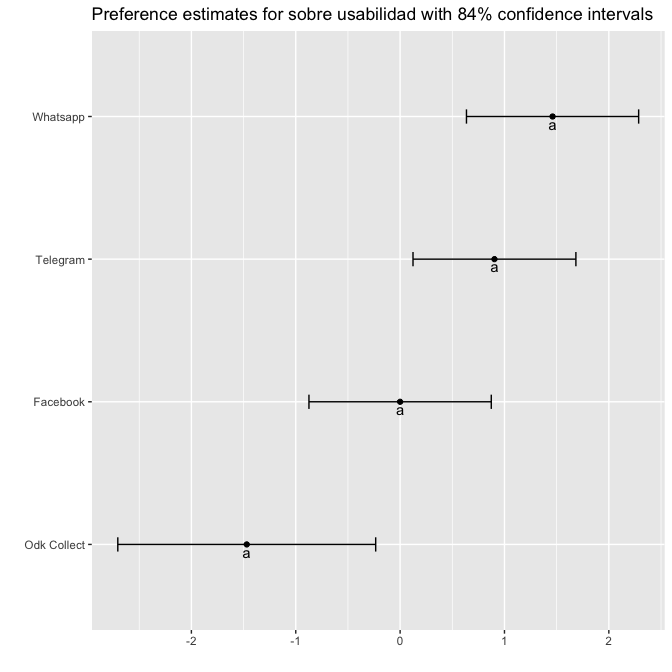
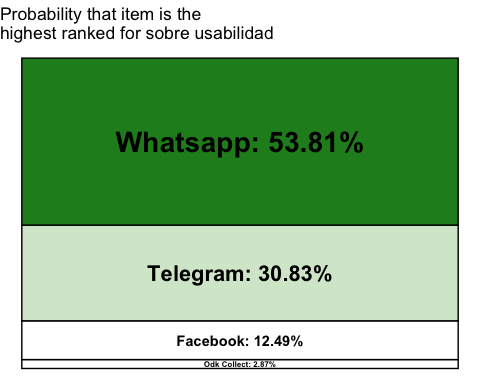


Table 5. 1 .2 Sobre usabilidad model parameter estimates

|  |  |  |  |
| --- | --- | --- | --- |
|  | Estimate | quasiSE | Group |
| Whatsapp | 1.46 | 0.59 | a |
| Telegram | 0.90 | 0.56 | a |
| Facebook | 0.00 | 0.62 | a |
| Odk Collect | -1.47 | 0.88 | a |



**Sobre comunicación**

Table 5.2.1

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| model | logLikelihood | DF | Statistic | Pr(>Chisq) |  |
| NULL | -17.91759 | 30 | NA | NA |  |
| Sobre comunicación | -15.52995 | 27 | 4.775292 | 0.18901 |  |

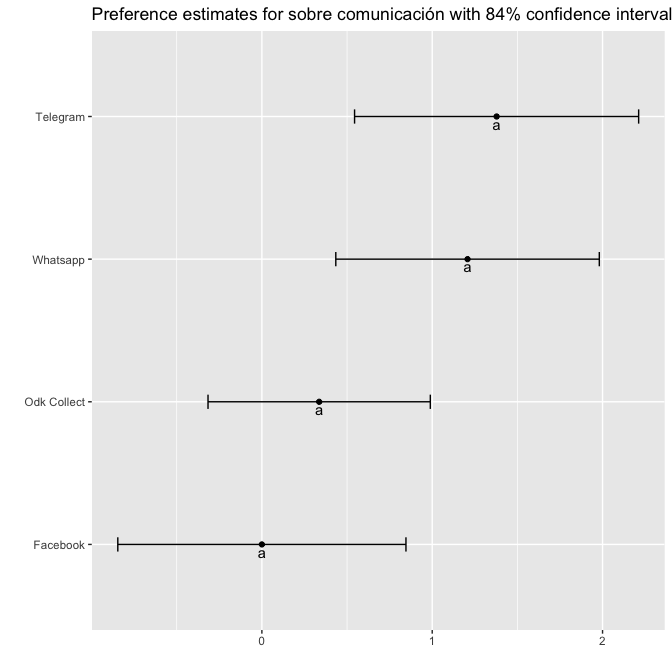
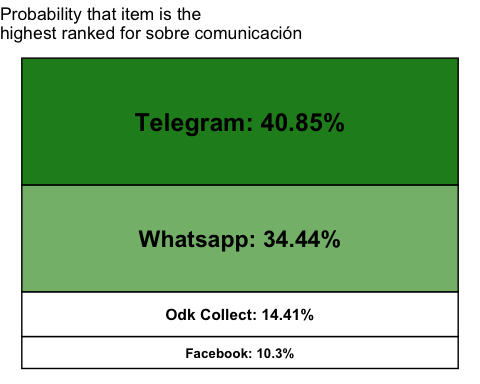


Table 5. 2 .2 Sobre comunicación model parameter estimates

|  |  |  |  |
| --- | --- | --- | --- |
|  | Estimate | quasiSE | Group |
| Telegram | 1.38 | 0.59 | a |
| Whatsapp | 1.21 | 0.55 | a |
| Odk Collect | 0.34 | 0.46 | a |
| Facebook | 0.00 | 0.60 | a |



# Section 6: Plackett-Luce models with explatory variables

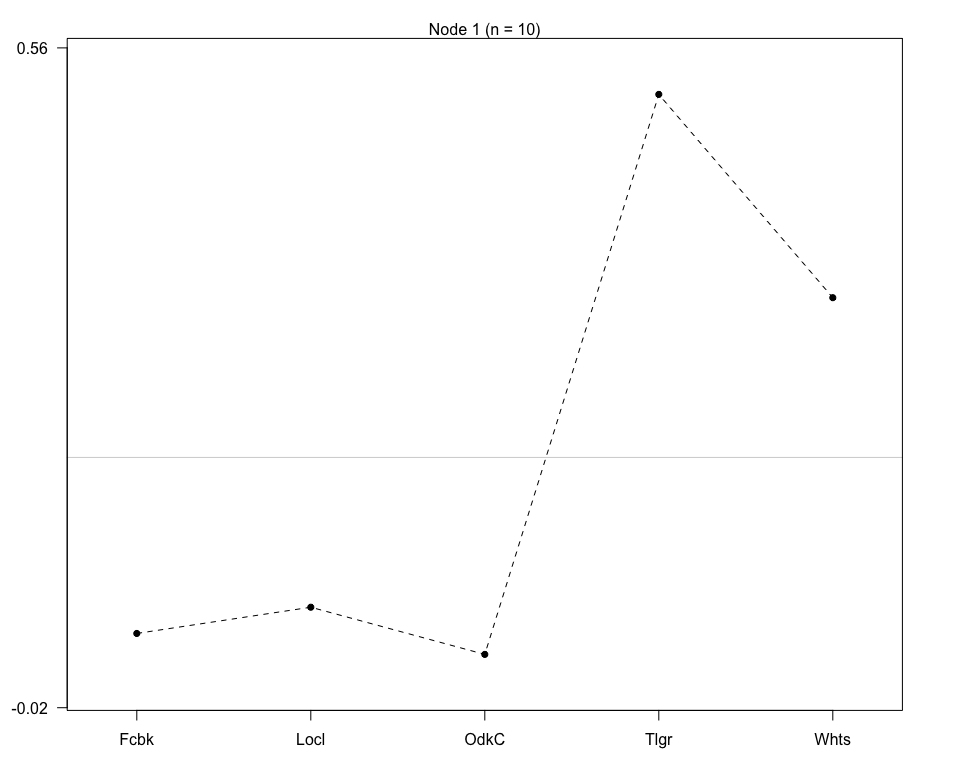
## Overall ranking

A recursive partitioning method (Strobl, Malley, and Tutz 2009) was used to determine which of the explanatory variables, if any, had significant relationships with the rankings. This approach identifies sub-groups in the data for which the rankings of the different varieties are significantly different to each other. Table 4.1 shows the p-values for each of the covariates tested, one-by-one, showing whether or not the covariate could be used to define sub-groups with significantly different rankings.

Table 6.1: Univariate p-values for first split in Plackett-Luce tree model for the overall ranking

|  |  |
| --- | --- |
| Variable | p |
| ¿Utiliza Whatsapp? | 0.72142 |

Figure 6.1 shows the partitioning of the rankings based on the most significantly different sub-groups which could be identified from the data using a sig\_level\*100% significance level. At the top of the tree is the full dataset, then working down through the different levels of the tree shows the combinations of variables which define each subgroup. The model parameters are shown for the final subgroups (“terminal nodes”) in the plots at the bottoms of the tree.

 No significant explanatory variables identified.

The highest and lowest performing item within each sub-group is identified within Table 6.2.

The model coefficient estimates, along with 84% confidence intervals are provided in Figure 6.2. This will help identification of which items were better suited to particular sub-groups identified by the analysis.

No significant explanatory variables identified.

Table 6.3 outlines the p-values for each covariate at each of the nodes in the tree, outlining whether an additional significant split could be determined from within the existing sub-group at that node.

Table 6.3: p-values for effect of each covariate at each node

|  |  |  |  |
| --- | --- | --- | --- |
| Variable | Node | statistic | p |
| ¿Utiliza Whatsapp? | 1 | 7.652219 | 0.72142 |

# References

Kendall, M. G. 1938. “A new measure of ranking correlation.” *Biometrika* 30 (1-2): 81–93. <https://doi.org/10.1093/biomet/30.1-2.81>.

R Core Team. 2019. “R: A language and environment for statistical computing. version 3.6.2.” Vienna, Austria: CRAN R Project. <https://r-project.org/>.

Strobl, Carolin, James Malley, and Gerhard Tutz. 2009. “An introduction to recursive partitioning: rationale, application, and characteristics of classification and regression trees, bagging, and random forests.” *Psychological Methods* 14 (4): 323–48. <https://doi.org/10.1037/a0016973>.

Turner, Heather L, Jacob van Etten, David Firth, and Ioannis Kosmidis. 2020. “Modelling rankings in R: the PlackettLuce package.” *Computational Statistics*. <https://doi.org/10.1007/s00180-020-00959-3>.

van Etten, Jacob, Eskender Beza, Lluís Calderer, Kees Van Duijvendijk, Carlo Fadda, Basazen Fantahun, Yosef Gebrehawaryat Kidane, et al. 2019. “First experiences with a novel farmer citizen science approach: crowdsourcing participatory variety selection through on-farm triadic comparisons of technologies (tricot).” *Experimental Agriculture* 55 (S1): 275–96. <https://doi.org/10.1017/S0014479716000739>.

van Etten, Jacob, Carlos Quirós, Kauê de Sousa, Jonathan Steinke, Sam Dumble, Brandon Madriz, Allan Coto, et al. 2020. “ClimMob: Software to support experimental citizen science in agriculture.” <https://climmob.net/climmob3>.