Earth Agency Project: Progress Report

Consulting Team A, group 3

March 7th, 2021

1. Decisions made in creating the working data frame

(a) We have combined *EarthAgency_Adults_R.csv* and *EarthAgency_Children_R.csv* into one data.frame: `AC_df`. We have done this so that we can make an Adult vs. children comparison. To combined the two data sets we have done b-d.

(b) To normalize the adult and children's `invitalscores`, we have compressed the adult's scores (which ranged from 0-5), to match the children's scores (which ranged from 0-3).

original	0	1	2	3	4	5
normalized	0	0	1	2	3	3

(c) To normalize the adult and children's `inpsychscore`, we have compressed the adult's scores (which ranged from 0-5), to match the children's scores (which ranged from 0-4).

original	0	1	2	3	4	5
normalized	0	1	2	2	3	3

(d) We have added `FirstLang` to the children's records, and assigned every child a value of 1 for `FirstLang`. We're assuming that 1=using first language, and 0=not using first language.

(e) The 87th entry in the children's record's has a `MeanSeverity` of 2.67. But 2.67 is not a possible value (given that `MeanSeverity` is an average of 4 integer scores, and should thus be a multiple of 0.25). We've changed the `MeanSeverity` for that record to 2.5.

(f) The 41st children's record has no `BIoJtscore` or `AntJtscore`. We did not use this record in the data.frame.

(g) The 45th children's record has no `Agency_Language` or `SRFactsTotal`. We did not use this record in the data.frame.

(h) We have not included the independent variable `SRTotal` in the data frame. Our understanding is that this variable is measuring comprehension of the video and was designed to test if children had payed attention to the video. For the children it was a 4 question test. Of the 91 children, 4 had a score of 2, 28 a score of 3, and 59 a score of 4. And we have not filtered out any children from the data frame based on this comprehension check.

(i) There is also a `SRTotal` variable for the adults. We did not include it in the data frame. If it is also a comprehension check, it might be useful consider filtering out some of the adults based on the results. This is a table of the results:

score	6	7	8	9	10	11	12
count	1	1	15	18	40	37	20

(j) We have renamed the `Condition` levels to `Obj` (object), `Nat` (nature, animal, vitalist), `Per` (person, psychological), for ease of understanding and consistency while we were coding.

(k) For a better model fit, we have combined the original 13 levels of `MeanSeverity` into three levels, as follows:

original	0.00	0.25	0.50	0.75	1.00	1.25	1.50	1.75	2.00	2.25	2.50	2.75	3.00
combined	1	1	1	1	1	1	2	2	2	3	3	3	3

(l) For a better model fit, we have combined the original 5 levels of `BioJtscore` (the number of scenarios where the respondent used a biocentric justification), into three levels, as follows:

original	0	1	2	3	4
combined	1	1	1	2	3

(m) For a better model fit, we have combined the original 5 levels of `AntJtscore` (the number of scenarios where the respondent used an anthropocentric justification), into three levels, as follows:

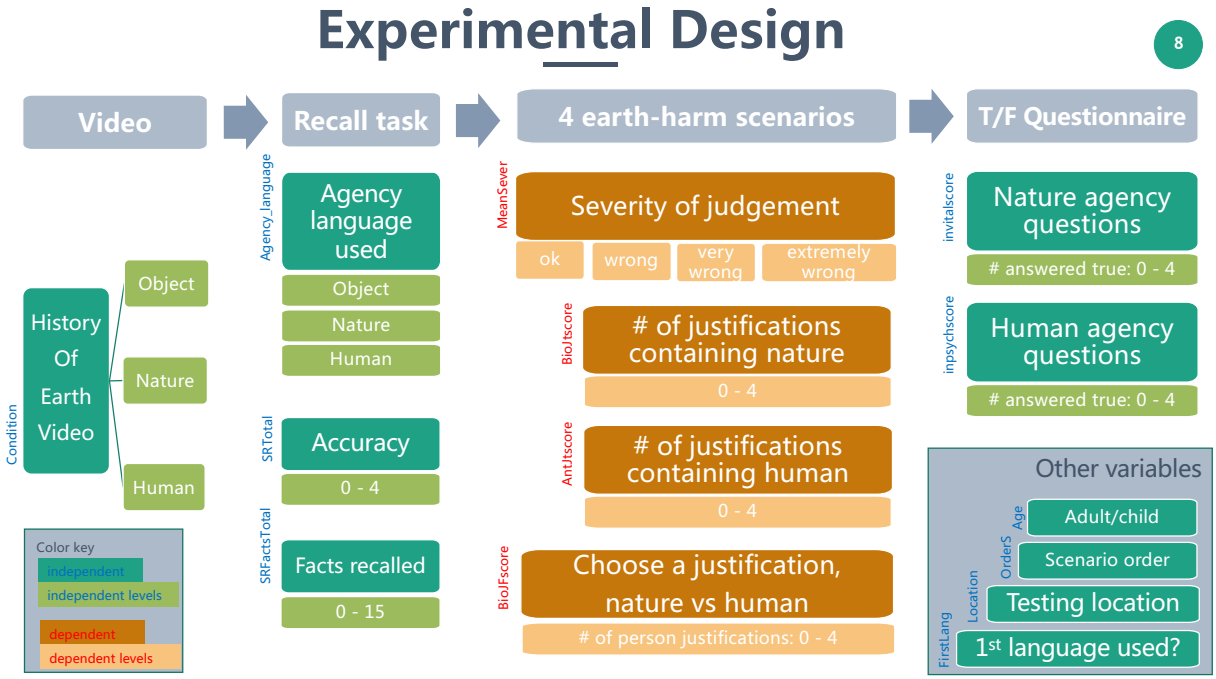
original	0	1	2	3	4
combined	1	1	1	2	3

(n) For a better model fit, we have combined the original 5 levels of `BioJFtotal` (the number of scenarios where the respondent chose biocentric when given a choice between biocentric and anthropocentric), into three levels, as follows:

original	0	1	2	3	4
combined	1	1	2	3	3

2. Experimental Design concerns

The chart below shows our understanding of the variables collected. The green are the independent variables with their levels, and the brown are the four dependent variables. As you have noted, the flow our your experiment design has intertwined the `Condition` variable with the `Agency_language`, `inpsychscore` and `invitalscore` variables. In particular, since the questionnaire was given after the participants watched the video, it's not clear that the `invitalscore` and `inpsychscore` variables are measuring the participants underlying beliefs, or the beliefs expressed in the video that they just watched. Also `Agency_language`, which is attempting to measure how the participants describe the video, overlaps with the perspective of the video watched. As such it is not clear that `Agency_language` is descriptive of the participant or of the video they just watched.



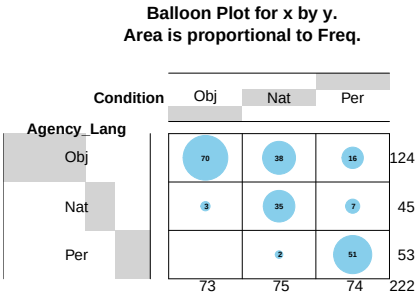
3. Correlation among independent variables

Because of our worry about the independence of the independent variables, we checked the correlations between the independent variables. The following chart (which is redundant across the diagonal), shows Pearson product-moment correlations between the numeric variables (`SRFactsTotal`, `invitalscore`, `inpsychscore`), a polychoric correlations between ordinal categorical variables (`Condition`, `Agency_Language`, `FirstLang` and `Age`), and polyserial correlations between numeric and ordinal variables. 1 or -1 is a strong correlation. 0 is no correlation.

	Condition	Agency_Language	SRFactsTotal	invitalscore	inpsychscore	FirstLang	Age
Condition	1.00	0.85	0.01	0.09	0.19	0.01	0.03
Agency_Language	0.85	1.00	0.10	0.04	0.24	0.14	0.18
SRFactsTotal	0.01	0.10	1.00	0.10	-0.24	-0.19	-0.52
invitalscore	0.09	0.04	0.10	1.00	0.43	-0.11	-0.22
inpsychscore	0.19	0.24	-0.24	0.43	1.00	0.15	0.38
FirstLang	0.01	0.14	-0.19	-0.11	0.15	1.00	0.88
Age	0.03	0.18	-0.52	-0.22	0.38	0.88	1.00

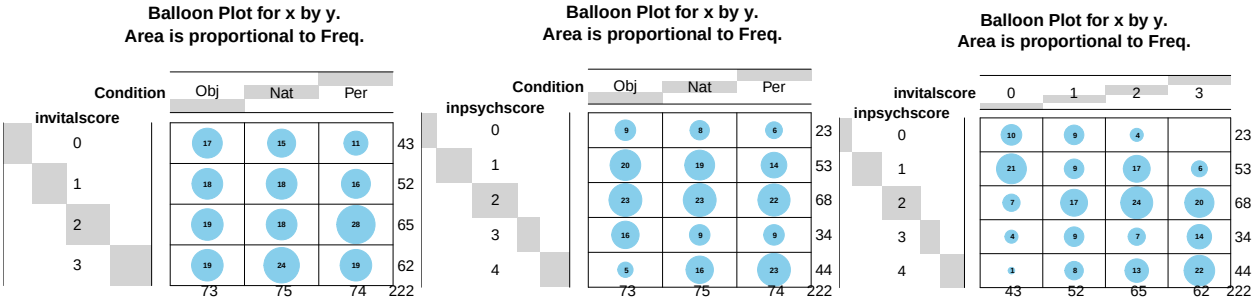
3a. Correlation between `Condition` and `Agency_language`

`Condition` and `Agency_language` had the strongest correlation and the p-value of the Pearson's chi-square test was effectively zero ($p = 7.3{*}10^{-36}$). Because of this, and because of our concerns about the experimental design, we have not included `Agency_language` in the model fits below. But we are going to explore using PCA (principal component analysis) to combine these two independent variables.



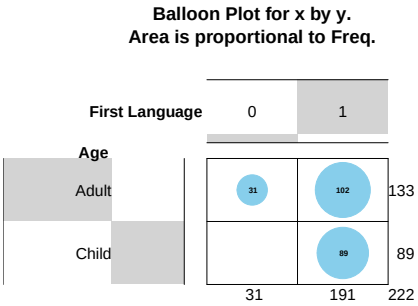
3b. Correlation between `Condition`, `invitalscore`, and `inpsychscore`

A Chi-square test shows that there is not a significant correlation between `Condition` and `invitalscore` (p=0.54), and there is a slightly significant correlation between `Condition` and `inpsychscore` (p=0.04). But the correlation between `invitalscore` and `inpsychscore` is highly significant ($p = 5.0 * 10^{-4}$). And we are going to explore using PCA (principal component analysis) to combine these two independent variables (`invitalscore` and `inpsychsocre`). But in the models below we have included both.



3c. Correlation between `FirstLang` and `Age`

Because we are assuming that the 89 children in the study are all using their first language, there is a strong correlation between `FirstLang` and `Age` ($p = 2.4 * 10^{-6}$). And for contectual reasons, we have left both of these variables in the models below.



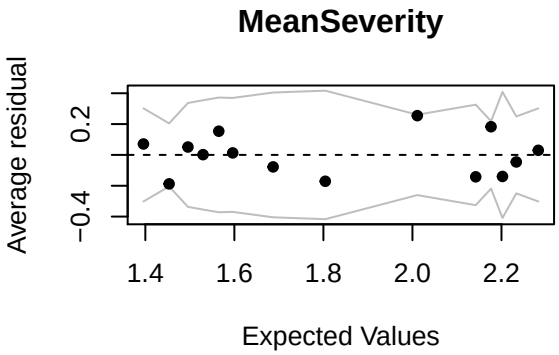
4a. Linear models

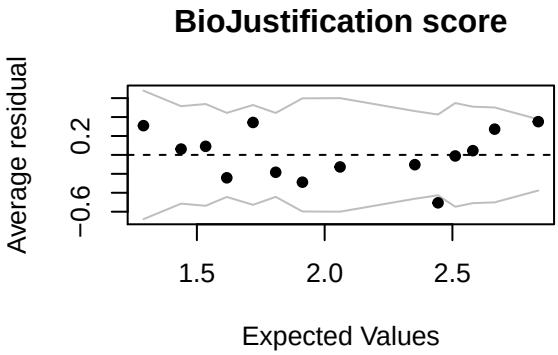
We need to interpret the linear models and the binnedplots. And give a rationale for why we aren't using these models.

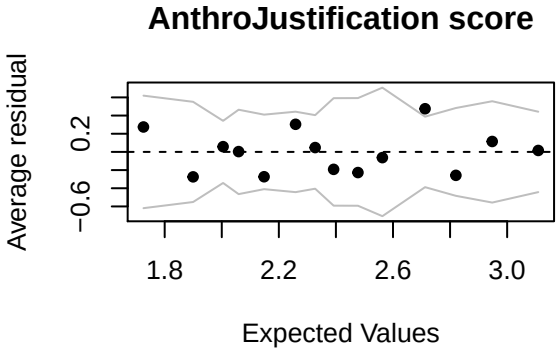
```
Call:
lm(formula = MeanSever ~ Condition + SRFactsTotal + invitalscore +
    inpsychscore + FirstLang + Age, data = AC_df)

Residuals:
     Min       1Q   Median       3Q      Max
-1.69195 -0.44206  0.01201  0.41579  1.42598

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.59835    0.17997   8.881 2.72e-16 ***
ConditionNat  0.05820    0.10935   0.532   0.5951
ConditionPer  0.07237    0.11116   0.651   0.5157
SRFactsTotal  0.01943    0.01592   1.221   0.2234
invitalscore  0.01161    0.05014   0.231   0.8172
inpsychscore  0.01280    0.04580   0.279   0.7802
FirstLang1   -0.28690    0.13608  -2.108   0.0362 *
AgeChild      0.72846    0.11856   6.144 3.87e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.6609 on 214 degrees of freedom
Multiple R-squared:  0.1976,    Adjusted R-squared:  0.1714
F-statistic: 7.531 on 7 and 214 DF,  p-value: 4.144e-08

Call:
lm(formula = BioJtscore ~ Condition + SRFactsTotal + invitalscore +
    inpsychscore + FirstLang + Age, data = AC_df)

Residuals:
    Min      1Q  Median      3Q     Max
-2.4924 -0.6939  0.1608  0.6146  2.7648

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.53184    0.28421   5.390 1.86e-07 ***
ConditionNat -0.01138    0.17268  -0.066  0.94753
ConditionPer  0.02327    0.17554   0.133  0.89465
SRFactsTotal  0.06933    0.02513   2.758  0.00631 **
invitalscore -0.01261    0.07919  -0.159  0.87365
inpsychscore  0.04838    0.07233   0.669  0.50427
FirstLang1   -0.40170    0.21489  -1.869  0.06295 .
AgeChild      1.11615    0.18723   5.961 1.02e-08 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.044 on 214 degrees of freedom
Multiple R-squared:  0.1856,    Adjusted R-squared:  0.159
F-statistic: 6.969 on 7 and 214 DF,  p-value: 1.748e-07
```

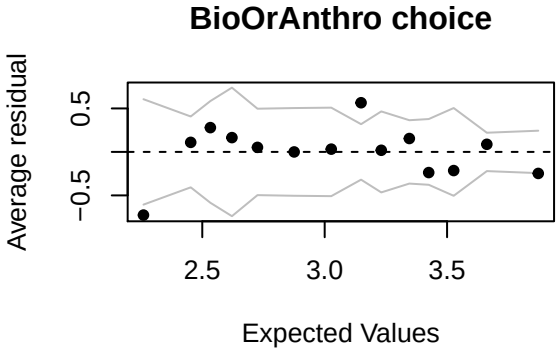
```
Call:
lm(formula = AntJtscore ~ Condition + SRFactsTotal + invitalscore +
    inpsychscore + FirstLang + Age, data = AC_df)

Residuals:
     Min       1Q   Median       3Q      Max
-2.58773 -0.71166 -0.03065  0.85416  2.23616

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   1.845842   0.278763   6.622 2.82e-10 ***
ConditionNat -0.209865   0.169373  -1.239 0.216676
ConditionPer -0.275448   0.172178  -1.600 0.111122
SRFactsTotal  0.009204   0.024653   0.373 0.709273
invitalscore  0.111711   0.077671   1.438 0.151820
inpsychscore  0.118660   0.070940   1.673 0.095852 .
FirstLang1   -0.021672   0.210778  -0.103 0.918201
AgeChild      0.617294   0.183645   3.361 0.000919 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 1.024 on 214 degrees of freedom
Multiple R-squared:  0.1338,    Adjusted R-squared:  0.1055
F-statistic: 4.724 on 7 and 214 DF,  p-value: 5.959e-05


Call:
lm(formula = BioJFtotal ~ Condition + SRFactsTotal + invitalscore +
    inpsychscore + FirstLang + Age, data = AC_df)

Residuals:
    Min      1Q  Median      3Q     Max
-2.6461 -0.6242  0.2526  0.6808  1.6597

Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)   2.69136    0.26312  10.229  < 2e-16 ***
ConditionNat  0.22307    0.15987   1.395  0.16435
ConditionPer -0.03234    0.16251  -0.199  0.84247
SRFactsTotal  0.06373    0.02327   2.739  0.00668 **
invitalscore  0.08002    0.07331   1.092  0.27626
inpsychscore -0.09731    0.06696  -1.453  0.14763
FirstLang1    0.32873    0.19895   1.652  0.09993 .
AgeChild     -0.56200    0.17334  -3.242  0.00138 **
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.9662 on 214 degrees of freedom
Multiple R-squared:  0.1988,    Adjusted R-squared:  0.1726
F-statistic: 7.584 on 7 and 214 DF,  p-value: 3.619e-08
```

4b. Binned residual plots









##

4c. ANOVA for linear model of Mean Severity

```
Analysis of Variance Table

Model 1: MeanSever ~ Condition + SRFactsTotal + invitalscore + inpsychscore +
    FirstLang + Age
Model 2: MeanSever ~ Condition + SRFactsTotal + invitalscore + inpsychscore +
    FirstLang
  Res.Df     RSS Df Sum of Sq     F    Pr(>F)
1    214  93.471
2    215 109.959 -1   -16.488 37.75 3.873e-09 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

5a. Mean Severity Ordinal model with combined levels

Do we want to do an F1 test aside from misclassification error to evaluate the model??

```
original levels

   0 0.25  0.5 0.75    1 1.25  1.5 1.75    2 2.25  2.5 2.75    3
   1    2    6   18   17   21   22   24   29   23   24   16   19

combined levels
```

```
 1  2  3
65 75 82

              Value Std. Error t value p value
ConditionNat  0.286      0.321   0.891   0.373
ConditionPer  0.248      0.318   0.778   0.437
SRFactsTotal  0.058      0.047   1.239   0.215
invitalscore -0.016      0.144  -0.111   0.912
inpsychscore  0.019      0.131   0.148   0.882
FirstLang1   -0.787      0.395  -1.994   0.046
AgeChild      2.049      0.362   5.659   0.000
1|2          -0.430      0.519  -0.828   0.408
2|3           1.266      0.526   2.408   0.016

              1  2  3
1            42 14  9
2            31 13 31
3            17 10 55
prop.correct 65 17 67

Misclassification error is: 0.5045045
```

5b. Biocentric Justification score, Ordinal model with combined levels

```
original levels

 0  1  2  3  4
19 57 61 62 23

combined levels

  1   2   3
137  62  23

              Value Std. Error t value p value
ConditionNat -0.059      0.356  -0.166   0.868
ConditionPer  0.196      0.358   0.548   0.584
SRFactsTotal  0.163      0.055   2.978   0.003
invitalscore -0.010      0.159  -0.066   0.947
inpsychscore -0.011      0.147  -0.073   0.942
FirstLang1   -0.474      0.473  -1.002   0.316
AgeChild      2.144      0.418   5.123   0.000
1|2           1.816      0.627   2.899   0.004
2|3           3.736      0.673   5.548   0.000

               1  2 3
1            128  9 0
2             44 17 1
3              9 14 0
prop.correct  93 27 0

Misclassification error is: 0.3468468
```

5c. Anthropocentric Justification score, Ordinal model with combined levels

```
original levels

 0  1  2  3  4
 9 36 76 61 40

combined levels

  1   2   3
121  61  40

              Value Std. Error t value p value
ConditionNat -0.278      0.328  -0.848   0.396
ConditionPer -0.527      0.334  -1.576   0.115
SRFactsTotal  0.055      0.050   1.105   0.269
invitalscore  0.089      0.151   0.589   0.556
inpsychscore  0.195      0.139   1.403   0.161
FirstLang1    0.353      0.461   0.767   0.443
AgeChild      1.301      0.354   3.680   0.000
1|2           1.566      0.598   2.621   0.009
2|3           3.065      0.624   4.908   0.000

               1  2  3
1            104 15  2
2             45  9  7
3             21 11  8
prop.correct  86 15 20

Misclassification error is: 0.454955
```

5d. Biocentric choice score, Ordinal model with combined levels

```
original levels

  0   1   2   3   4
  5  16  41  60 100

combined levels

  1   2   3
 21  41 160

              Value Std. Error t value p value
ConditionNat  0.041      0.400   0.102   0.919
ConditionPer -0.279      0.404  -0.691   0.490
SRFactsTotal  0.195      0.065   3.012   0.003
invitalscore  0.377      0.185   2.041   0.041
```

```
inpsychscore -0.394      0.174  -2.258   0.024
FirstLang1    0.967      0.518   1.868   0.062
AgeChild     -1.030      0.433  -2.378   0.017
1|2          -1.696      0.654  -2.592   0.010
2|3          -0.153      0.631  -0.243   0.808

              1 2   3
1             3 5  13
2             0 3  38
3             1 2 157
prop.correct 14 7  98

Misclassification error is: 0.2657658
```

6. ANOVA for ordinal models

This is code that Maggie added to check the ANOVA for the polr() models. Not sure if this is correct (Zihuan said that we needed to check how you do ANOVA for ordinal models). And this is just checking the Age variable. Not the Condition variable.

```
Likelihood ratio tests of ordinal regression models

Response: comMeanSever
                                                                      Model
1       Condition + SRFactsTotal + invitalscore + inpsychscore + FirstLang
2 Condition + SRFactsTotal + invitalscore + inpsychscore + FirstLang + Age
  Resid. df Resid. Dev   Test    Df LR stat.      Pr(Chi)
1       214   472.3211
2       213   438.0412 1 vs 2     1  34.2799 4.772864e-09

Likelihood ratio tests of ordinal regression models

Response: comBioJtscore
                                                                      Model
1       Condition + SRFactsTotal + invitalscore + inpsychscore + FirstLang
2 Condition + SRFactsTotal + invitalscore + inpsychscore + FirstLang + Age
  Resid. df Resid. Dev   Test    Df LR stat.      Pr(Chi)
1       214   386.1977
2       213   356.5992 1 vs 2     1 29.59845 5.314746e-08

Likelihood ratio tests of ordinal regression models

Response: comAntJtscore
                                                                      Model
1       Condition + SRFactsTotal + invitalscore + inpsychscore + FirstLang
2 Condition + SRFactsTotal + invitalscore + inpsychscore + FirstLang + Age
  Resid. df Resid. Dev   Test    Df LR stat.      Pr(Chi)
1       214   423.6080
2       213   409.5921 1 vs 2     1 14.01596 0.0001812655

Likelihood ratio tests of ordinal regression models
```

```
Response: comBioJFtotal
                                                                      Model
1       Condition + SRFactsTotal + invitalscore + inpsychscore + FirstLang
2 Condition + SRFactsTotal + invitalscore + inpsychscore + FirstLang + Age
  Resid. df Resid. Dev   Test    Df LR stat.    Pr(Chi)
1       214   299.7488
2       213   293.8512 1 vs 2     1 5.897596 0.01516157
```