WSP modelling

Lizzie Jones

16/06/2021

Quantitative analysis/modelling

Quantitative data analysis: Descriptive and statistical - to understand variation in respondent's awareness, knowledge and attitudes towards white storks and their reintroduction.

Methods plan

- GLM approach + model selection and averaging
  - Anderson, D. and Burnham, K., 2004. Model selection and multi-model inference. Second. NY: Springer-Verlag, 63(2020), p.10.
  - Burnham, K.P., Anderson, D.R. and Huyvaert, K.P., 2011. AIC model selection and multimodel inference in behavioral ecology: some background, observations, and comparisons. Behavioral ecology and sociobiology, 65(1), pp.23-35.
- PCA/Clustering?

Exploring the response variable

Response variable = Attitudes to WS reintroduction (Composite score)

```
str(final_data$OverallAttitudeScore) # check it's a numeric column
```

```
##  num [1:3531] NA NA NA 4.93 4.64 4.86 4.57 4.86 3.29 3.93 ...
```

```
final_data %>%
     group_by(SurveyType) %>%
     summarise(sum(!is.na(OverallAttitudeScore))) ## Couting NON-NA values per
survey type
```

```
## # A tibble: 2 x 2
##   SurveyType `sum(!is.na(OverallAttitudeScore))`
##   <fct>                                    <int>
## 1 NatRep                                     743
## 2 Proactive                                 1749
```

Possible predictor variables

Factor variables

- Age (collapse further?)
- Gender (female / male)

- Urban / suburban / rural
- Highest education (collapse – e.g. degree; below degree)
- Occupation (use? If so, would need to collapse! unemployed; retired; potentially pool responses except for those who answered "environment, nature & wildlife")
- Visited Knepp (yes / no)
- Time spent in nature
- Member of conservation/environmental organisation (quite a few people listed RSPB)
- Awareness
- Heard of white stork before taking this survey?
- Heard of white stork project / reintroduction effort?

```
# Select all possible predictor vars
model_data <- final_data %>%
  dplyr::select(UniqueID_all, OverallAttitudeScore, SiteProximity, SurveyType,
         Age_short, Gender, Area_type, Education_short, Occupation_short,
         Q27_Knepp_visit, Q18_exp_nature, Q1_aware_stork, Q9_heard, KnowledgeS
core,
         Q22....Are.you.a.member.of.any.environmental..wildlife.or.conservatio
n.organisations.,
         NCI, ProCoBS, BirdInterestScore, EnvConcern.score)
nrow(model_data)
```

```
## [1] 3531
```

```
model_clean <- model_data[!is.na(model_data$OverallAttitudeScore), ]
nrow(model_clean) ## Dropped ~1100 rows due to NA in AttitudeScore
```

```
## [1] 2492
```

```
# Select numeric variables
model_clean1 <- model_clean %>%
  drop_na()
nrow(model_clean1)
```

```
## [1] 2483
```

Predictor correlation matrix

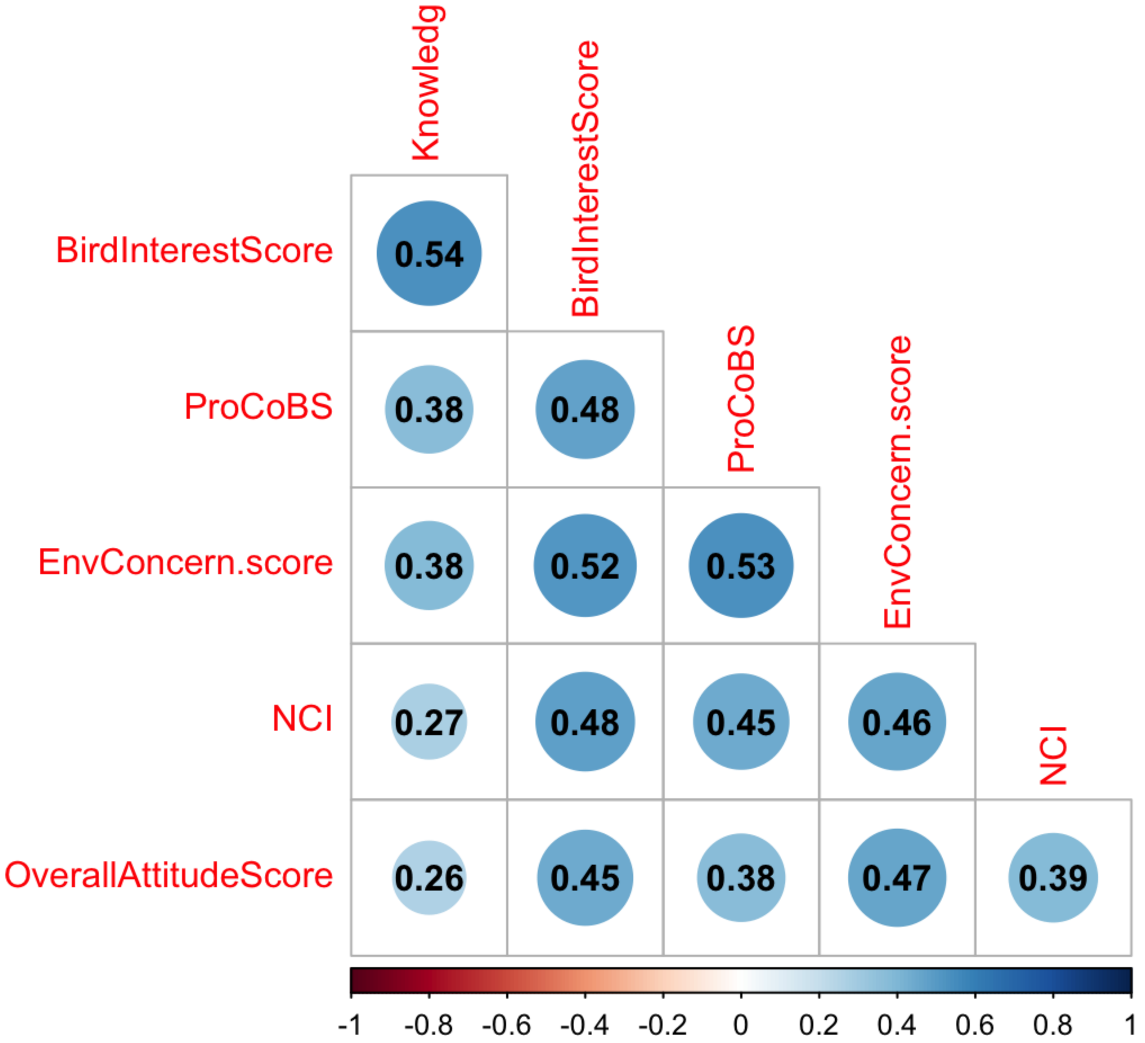
Numeric variables

- Contact and connection with nature; general environmental attitudes and behaviour
- Nature Connection Index (composite score)
- Environmental concern (composite score)
- General attitude towards birds (composite score)

```
# Select numeric variables
model_numeric <- model_clean1 %>%
  dplyr::select_if(., is.numeric) %>%
  dplyr::select(., -UniqueID_all) %>%
  drop_na()
nrow(model_numeric)
```

```
## [1] 2483
```

```
# Create corrlation matrix
model.cor = cor(model_numeric, method = c("spearman"))
res1 <- cor.mtest(model_numeric, conf.level = .95)
corrplot::corrplot(model.cor, p.mat = res1$p, method = "circle", type = "lowe
r", insig='blank',
         addCoef.col ='black', order = "AOE", diag=FALSE)
```




```
### All variables moderately correlated but non significant
```

Global model

We now generate the global model. Remember, this is a saturated model with all of the fixed effects and their interesting interactions. There are no random effects in this model so we use a linear model.

```
head(model_clean, 20)
```

```
##    UniqueID_all OverallAttitudeScore SiteProximity SurveyType Age_short Gen
der
## 4             4                 4.93     Not local  Proactive     18-44   M
ale
## 5             5                 4.64     Not local  Proactive     18-44 Fem
ale
## 6             6                 4.86         Local  Proactive     18-44 Fem
ale
## 7             7                 4.57     Not local  Proactive     45-64 Fem
ale
## 8             8                 4.86     Not local  Proactive     18-44   M
ale
## 9             9                 3.29     Not local  Proactive     45-64   M
ale
## 10           10                 3.93     Not local  Proactive     45-64   M
ale
## 11           11                 4.71     Not local  Proactive     18-44 Fem
ale
## 12           12                 4.36     Not local  Proactive     18-44   M
ale
## 13           13                 3.79     Not local  Proactive     18-44   M
ale
## 14           14                 4.71     Not local  Proactive     45-64   M
ale
## 17           17                 3.86     Not local  Proactive     18-44 Fem
ale
## 18           18                 4.29     Not local  Proactive     45-64 Fem
ale
## 19           19                 3.93     Not local  Proactive       65+   M
ale
## 20           20                 4.29     Not local  Proactive     18-44 Fem
ale
## 21           21                 4.64     Not local  Proactive     18-44   M
ale
## 22           22                 4.86     Not local  Proactive     18-44   M
ale
## 23           23                 4.57     Not local  Proactive     18-44 Fem
ale
## 24           24                 3.86     Not local  Proactive     18-44 Fem
ale
## 26           26                 4.00     Not local  Proactive       65+   M
ale
##    Area_type    Education_short   Occupation_short Q27_Knepp_visit
## 4      Urban University graduate Environment/Nature             Yes
## 5      Rural University graduate Environment/Nature             Yes
## 6  Sub-urban University graduate  Non-environmental              No
## 7      Rural High school/College            Student              No
## 8      Urban University graduate  Non-environmental              No
## 9      Rural University graduate  Prefer not to say              No
## 10 Sub-urban High school/College  Non-environmental              No
## 11     Urban University graduate  Non-environmental             Yes
```

```
## 12 Sub-urban University graduate Environment/Nature             No
## 13     Rural University graduate  Non-environmental             No
## 14 Sub-urban High school/College Unemployed/Retired             No
## 17 Sub-urban University graduate          Education             No
## 18     Rural University graduate  Non-environmental            Yes
## 19 Sub-urban High school/College Unemployed/Retired             No
## 20     Urban University graduate Environment/Nature             No
## 21     Urban High school/College  Non-environmental             No
## 22 Sub-urban University graduate  Non-environmental             No
## 23     Urban University graduate            Student             No
## 24 Sub-urban University graduate              Other             No
## 26     Urban University graduate Unemployed/Retired            Yes
##       Q18_exp_nature Q1_aware_stork Q9_heard KnowledgeScore
## 4  Every day, 7 days            Yes      Yes            6.1
## 5  Every day, 7 days            Yes      Yes            5.6
## 6  Every day, 7 days            Yes      Yes            3.0
## 7           3-4 days            Yes      Yes            2.7
## 8           1-2 days            Yes      Yes            2.0
## 9  Every day, 7 days            Yes      Yes            6.8
## 10          5-6 days            Yes      Yes            5.7
## 11          1-2 days            Yes      Yes            4.9
## 12 Every day, 7 days            Yes      Yes            5.6
## 13 Every day, 7 days            Yes      Yes            3.5
## 14 Every day, 7 days            Yes       No            4.0
## 17          3-4 days            Yes      Yes            3.5
## 18          3-4 days            Yes      Yes            3.1
## 19          1-2 days            Yes       No            2.4
## 20          5-6 days            Yes      Yes            4.5
## 21          3-4 days            Yes      Yes            3.6
## 22 Every day, 7 days            Yes      Yes            4.5
## 23          3-4 days            Yes      Yes            5.2
## 24 Every day, 7 days            Yes      Yes            2.4
## 26          5-6 days            Yes      Yes            7.0
##    Q22....Are.you.a.member.of.any.environmental..wildlife.or.conservation.o
rganisations.
## 4
Yes
## 5
Yes
## 6
Yes
## 7
No
## 8
No
## 9
Yes
## 10
Yes
## 11
No
```

```
## 12
Yes
## 13
Yes
## 14
Yes
## 17
Yes
## 18
Yes
## 19
No
## 20
Yes
## 21
Yes
## 22
Yes
## 23
No
## 24
Yes
## 26
Yes
##    NCI ProCoBS BirdInterestScore EnvConcern.score
## 4   43      19                17               10
## 5  100      23                19               10
## 6   59      17                17               10
## 7   59      16                17               10
## 8   59      18                19               10
## 9  100      28                20               10
## 10  45      22                15               10
## 11  40      20                16               10
## 12 100      24                20               10
## 13  53      16                18                8
## 14  49      18                20               10
## 17  75      18                18               10
## 18  59      16                15               10
## 19  55      18                16                8
## 20  59      24                19               10
## 21  62      21                20               10
## 22 100      20                17               10
## 23  34      15                16               10
## 24  34      21                16               10
## 26  83      24                20               10
```

```
# Create saturated model (potentially SurveyType or locality as a random effec
t?)
global_model <- glm(OverallAttitudeScore ~ SiteProximity + SurveyType +
         Age_short + Gender + Area_type + Education_short + Occupation_short +
         Q27_Knepp_visit + Q18_exp_nature + Q1_aware_stork + Q9_heard + Knowle
dgeScore +
         Q22....Are.you.a.member.of.any.environmental..wildlife.or.conservatio
n.organisations. +
         NCI + ProCoBS + BirdInterestScore + EnvConcern.score, data = model_cl
ean1)
summary(global_model)
```

```
##
## Call:
## glm(formula = OverallAttitudeScore ~ SiteProximity + SurveyType +
##     Age_short + Gender + Area_type + Education_short + Occupation_short +
##     Q27_Knepp_visit + Q18_exp_nature + Q1_aware_stork + Q9_heard +
##     KnowledgeScore + Q22....Are.you.a.member.of.any.environmental..wildlif
e.or.conservation.organisations. +
##     NCI + ProCoBS + BirdInterestScore + EnvConcern.score, data = model_clea
n1)
##
## Deviance Residuals:
##      Min        1Q    Median        3Q       Max
## -2.98646  -0.24504   0.08231   0.34999   1.37574
##
## Coefficients:
##
Estimate
## (Intercept)
1.9250392
## SiteProximityNot local
0.0309636
## SurveyTypeProactive
0.1413150
## Age_short45-64
-0.0195080
## Age_short65+
0.0050121
## Age_shortN/A
-0.2888474
## GenderMale
-0.0736924
## GenderN/A
-0.1402716
## Area_typeSub-urban
-0.0123917
## Area_typeUrban
-0.0462044
## Education_shortNo formal quals.
-0.0305604
## Education_shortOther
-0.2211145
## Education_shortUniversity graduate
-0.0140019
## Occupation_shortEnvironment/Nature
-0.2498903
## Occupation_shortNon-environmental
-0.0172591
## Occupation_shortOther
-0.0212587
## Occupation_shortPrefer not to say
-0.1263891
```

```
## Occupation_shortStudent
-0.0808504
## Occupation_shortUnemployed/Retired
-0.0419222
## Q27_Knepp_visitYes
0.0110452
## Q18_exp_nature3-4 days
-0.0273542
## Q18_exp_nature5-6 days
-0.0166776
## Q18_exp_natureEvery day, 7 days
-0.0497188
## Q18_exp_natureNone
0.0038165
## Q1_aware_storkYes
0.1490855
## Q9_heardNot sure
-0.0849757
## Q9_heardYes
0.0140659
## KnowledgeScore
-0.0231093
## Q22....Are.you.a.member.of.any.environmental..wildlife.or.conservation.orga
nisations.Yes -0.0666779
## NCI
0.0020069
## ProCoBS
0.0085734
## BirdInterestScore
0.0459183
## EnvConcern.score
0.1316478
##
Std. Error
## (Intercept)
0.1048284
## SiteProximityNot local
0.0290671
## SurveyTypeProactive
0.0368516
## Age_short45-64
0.0260297
## Age_short65+
0.0370880
## Age_shortN/A
0.1980219
## GenderMale
0.0226040
## GenderN/A
0.1260233
## Area_typeSub-urban
```

```
0.0258798
## Area_typeUrban
0.0310454
## Education_shortNo formal quals.
0.0801647
## Education_shortOther
0.0793529
## Education_shortUniversity graduate
0.0245173
## Occupation_shortEnvironment/Nature
0.0469196
## Occupation_shortNon-environmental
0.0386381
## Occupation_shortOther
0.0508600
## Occupation_shortPrefer not to say
0.0803485
## Occupation_shortStudent
0.0708754
## Occupation_shortUnemployed/Retired
0.0439091
## Q27_Knepp_visitYes
0.0286365
## Q18_exp_nature3-4 days
0.0305464
## Q18_exp_nature5-6 days
0.0334786
## Q18_exp_natureEvery day, 7 days
0.0333505
## Q18_exp_natureNone
0.0540408
## Q1_aware_storkYes
0.0355500
## Q9_heardNot sure
0.0581903
## Q9_heardYes
0.0312539
## KnowledgeScore
0.0085784
## Q22....Are.you.a.member.of.any.environmental..wildlife.or.conservation.orga
nisations.Yes  0.0275251
## NCI
0.0005082
## ProCoBS
0.0025724
## BirdInterestScore
0.0052108
## EnvConcern.score
0.0110773
##
t value
```

```
## (Intercept)
18.364
## SiteProximityNot local
1.065
## SurveyTypeProactive
3.835
## Age_short45-64
-0.749
## Age_short65+
0.135
## Age_shortN/A
-1.459
## GenderMale
-3.260
## GenderN/A
-1.113
## Area_typeSub-urban
-0.479
## Area_typeUrban
-1.488
## Education_shortNo formal quals.
-0.381
## Education_shortOther
-2.786
## Education_shortUniversity graduate
-0.571
## Occupation_shortEnvironment/Nature
-5.326
## Occupation_shortNon-environmental
-0.447
## Occupation_shortOther
-0.418
## Occupation_shortPrefer not to say
-1.573
## Occupation_shortStudent
-1.141
## Occupation_shortUnemployed/Retired
-0.955
## Q27_Knepp_visitYes
0.386
## Q18_exp_nature3-4 days
-0.895
## Q18_exp_nature5-6 days
-0.498
## Q18_exp_natureEvery day, 7 days
-1.491
## Q18_exp_natureNone
0.071
## Q1_aware_storkYes
4.194
## Q9_heardNot sure
```

```
-1.460
## Q9_heardYes
0.450
## KnowledgeScore
-2.694
## Q22....Are.you.a.member.of.any.environmental..wildlife.or.conservation.orga
nisations.Yes  -2.422
## NCI
3.949
## ProCoBS
3.333
## BirdInterestScore
8.812
## EnvConcern.score
11.884
##
Pr(>|t|)
## (Intercept)
< 2e-16
## SiteProximityNot local
0.286871
## SurveyTypeProactive
0.000129
## Age_short45-64
0.453658
## Age_short65+
0.892513
## Age_shortN/A
0.144786
## GenderMale
0.001129
## GenderN/A
0.265791
## Area_typeSub-urban
0.632110
## Area_typeUrban
0.136805
## Education_shortNo formal quals.
0.703073
## Education_shortOther
0.005370
## Education_shortUniversity graduate
0.567983
## Occupation_shortEnvironment/Nature
1.10e-07
## Occupation_shortNon-environmental
0.655142
## Occupation_shortOther
0.675995
## Occupation_shortPrefer not to say
0.115845
```

```
## Occupation_shortStudent
0.254090
## Occupation_shortUnemployed/Retired
0.339798
## Q27_Knepp_visitYes
0.699750
## Q18_exp_nature3-4 days
0.370610
## Q18_exp_nature5-6 days
0.618417
## Q18_exp_natureEvery day, 7 days
0.136143
## Q18_exp_natureNone
0.943703
## Q1_aware_storkYes
2.84e-05
## Q9_heardNot sure
0.144334
## Q9_heardYes
0.652712
## KnowledgeScore
0.007111
## Q22....Are.you.a.member.of.any.environmental..wildlife.or.conservation.orga
nisations.Yes 0.015489
## NCI
8.07e-05
## ProCoBS
0.000872
## BirdInterestScore
< 2e-16
## EnvConcern.score
< 2e-16
##
## (Intercept)
***
## SiteProximityNot local
## SurveyTypeProactive
***
## Age_short45-64
## Age_short65+
## Age_shortN/A
## GenderMale
**
## GenderN/A
## Area_typeSub-urban
## Area_typeUrban
## Education_shortNo formal quals.
## Education_shortOther
**
## Education_shortUniversity graduate
## Occupation_shortEnvironment/Nature
```

```
***
## Occupation_shortNon-environmental
## Occupation_shortOther
## Occupation_shortPrefer not to say
## Occupation_shortStudent
## Occupation_shortUnemployed/Retired
## Q27_Knepp_visitYes
## Q18_exp_nature3-4 days
## Q18_exp_nature5-6 days
## Q18_exp_natureEvery day, 7 days
## Q18_exp_natureNone
## Q1_aware_storkYes
***
## Q9_heardNot sure
## Q9_heardYes
## KnowledgeScore
**
## Q22....Are.you.a.member.of.any.environmental..wildlife.or.conservation.orga
nisations.Yes *
## NCI
***
## ProCoBS
***
## BirdInterestScore
***
## EnvConcern.score
***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for gaussian family taken to be 0.2767955)
##
##     Null deviance: 1006.63  on 2482  degrees of freedom
## Residual deviance:  678.15  on 2450  degrees of freedom
## AIC: 3891.9
##
## Number of Fisher Scoring iterations: 2
```

Model selection

Caveats to model selection

Depends on the models included in the candidate set. You can't identify a model as being the "best" fit to the data if you didn't include the model to begin with! The parameter estimates and predictions arising from the "best" model or set of best models should be biologically meaningful. Need to decide whether to use model selection or common inferential statistics (e.g. based on P-values). Techniques that rely on both approaches are possible (e.g. backward variable selection followed by averaging of top models), such as the example provided above.

```
# #### MODEL SELECTION USING MUMIN PACKAGE
# options(na.action = na.fail)
# # Check assumptions of lmm model (using cur_lmm2)
# vif(global_model)
# # plot_model(full_trend_dev)
# attitude_dredge <- dredge(global_model)
# summary(get.models(attitude_dredge, 1)[[1]])
# plot(attitude_dredge, labAsExpr = TRUE)
#
# #### Use a all-subsets model subsetting approach find a confidence set of mo
dels, recalculating weights each time:
# attitude_del <- subset(attitude_dredge, delta <= 2, recalc.weights = TRUE) #
delta(AIC) cutoff
# attitude_del
```

Model averaging

But how much evidence do we actually have that this is the best model? We have over XXX models so it's unlikely that only one model explains the data. From the dredge output we can see there is little difference in the AIC and weights of the first few models. Is there really much of a difference between two models who's AIC differ by only 0.14 points? How do we decide which model(s) to interpret? Statisticians have thought about this problem and it turns out that models with delta AIC (or other criterion) less than 2 are considered to be just as good as the top model and thus we shouldn't just discount them. Alternatively, we could use the weights: if a model has weight greater or equal to 95% then it is likely to be the top model. Otherwise we can generate a "credibility" set consisting of all models whose cumulative sum of AIC weights is 0.95. In any case, the point is that we have no good reason to exclude models other than the top one when the next models after it are likely to be just as good. To get around this, we can perform what's called model averaging (AKA multi-model inference), which allows us to average the parameter estimates across multiple models and avoids the issue of model uncertainty. Let's do this below by averaging all models with a delta AIC <= 2.

Key references

- Harrison XA, Donaldson L, Correa-Cano ME, Evans J, Fisher DN, Goodwin CED, Robinson BS, Hodgson DJ, Inger R. 2018. A brief introduction to mixed effects modelling and multi-model inference in ecology. PeerJ 6:e4794 https://doi.org/10.7717/peerj.4794 (https://doi.org/10.7717/peerj.4794)

```
### Model averaging the top XXX models according to the delta AIC value
# attitude_aves <- model.avg(get.models(attitude_del, subset = delta < 2))
# summary(attitude_aves)
# sjPlot::plot_model(attitude_aves, type = "est",
#                    show.values = TRUE, value.offset = .3, title = "Model ave
raged results") #  + ylim(-.05, 0.05)
# ggplot2::ggsave(filename = "Attitude_averaging_table.png", width = 9, height
= 15, dpi = 300)
#
# # Print coefficients in table
# attitude_confint <-as.table(round(confint(attitude_aves), 3))
# attitude_ave <- round(summary(attitude_aves)$coefmat.subset, 3)
# stargazer(attitude_ave, digits=3, title="Model averaged results", type = "ht
ml", out="Attitude_averages_table.doc")
#
# # Create tab_df table of model averaged estimates
# attitude_ave1 <- tibble::rownames_to_column(as.data.frame(attitude_ave), "Pr
edictor")
# sjPlot::tab_df(attitude_ave1, title = "Model averaged results", alternate.ro
ws = TRUE, digits=3, use.viewer = TRUE)
```