ANOVA Validation on New Data

Mitch Toland

4/8/2020

Introduction

This report documents the comparison of outputs from the RCB analysis-of-variance codes implemented in two environments, Velocity and Development. There are many changes being made as the Velocity system is being rolled out. As a result there is a continuing need to ensure codes are producing the correct outputs for each of several statistical modules used by the Velocity system to check data quality, make adjustments to some measurements and to do analyses. This report addresses the outputs from the RCB module. In Phase II, the RCB module will be replaced with the ANOVA module. This is mentioned here because some people have heard of both the RCB and ANOVA modules. The RCB module is the more limited Phase I implementation of the model fitting and analysis tools.

First load needed software.

```
library(jsonlite)
library(asreml)
```

```
## Loading required package: lattice
```

```
## Licensed to: Monsanto
## Serial Number: 402084150 Expires: 31-jul-2020 (114 days)
```

```
source("/repos/RCB4Cloud/R/RCB_MainFunction.R")
#source("/repos/RCB4Cloud/R/RCB_SupportFunctions.R") # supperceded by next line of code
source("/repos/RCB4Cloud/Reference/Rscripts/RCB_SupportFunctionsWithasremlPlus.R")
source("/repos/RCB4Cloud/Reference/Rscripts/fromAPI.R")
```

Second, get the data.

```
zdir  <- "/repos/RCB4Cloud/Reference/newValidation2020_04_07/"
zfile <- "ANOVA_GLOBAL_BLUE_845-31555-EHT-outputToP360.json"
zjsn  <- fromJSON(paste0(zdir, zfile))
zdat  <- zjsn$modelOutputs$rcbBlue$input$data[[1]]
zpar  <- zjsn$modelOutputs$rcbBlue$parameters
zrslt <- zjsn$modelOutputs$rcbBlue$results
```

Third, ensure datatypes are correct as the output is convert to character strings for storage.

```
zdat$factorLevelId         <- as.character(zdat$factorLevelId)
zdat$isAnswerDeactivated   <- as.logical(zdat$isAnswerDeactivated)
zdat$isDsrDeactivated      <- as.logical(zdat$isDsrDeactivated)
zdat$isPlaceHolder         <- as.logical(zdat$isPlaceHolder)
zdat$isQaqcDeactivated     <- as.logical(zdat$isQaqcDeactivated)
zdat$isSetEntryDeactivated <- as.logical(zdat$isSetEntryDeactivated)
zdat$locationId            <- as.character(zdat$locationId)
zdat$questionCode          <- as.character(zdat$questionCode)
zdat$repId                 <- as.character(zdat$repId)
```

```
zdat$subSiteId             <- as.character(zdat$subSiteId)
zdat$value                 <- as.numeric(zdat$value)
```

Fourth, set the parameters.

```
zpar$factorLevelId         <- "factorLevelId"
zpar$isAnswerDeactivated   <- "isAnswerDeactivated"
zpar$isDsrDeactivated      <- "isDsrDeactivated"
zpar$isPlaceHolder         <- "isPlaceHolder"
zpar$isQaqcDeactivated     <- "isQaqcDeactivated"
zpar$isSetEntryDeactivated <- "isSetEntryDeactivated"
zpar$locationId            <- "locationId"
zpar$questionCode          <- "questionCode"
zpar$repId                 <- "repId"
zpar$subSiteId             <- "subSiteId"
zpar$value                 <- "value"
```

Fifth, fit the RCB model to the original input data and make summary outputs. The text echoed below indicates the model is being fit and the correspondinng output is being generated.

```
zRCB_Output <- RCB_ModelFittingFunction(zdat, zpar, "P2")
```

```
## [1] "input 104 rows"
## [1] "current question code is EHT"
## [1] "analyzing 104 rows"
## [1] " Model2: Multi LOC Single REP BLUE"
## Running ASReml using analysis type: P2
## finished in 11.97 secs
## Creating output files...
```

Finally, compare the Velocity and development outputs.

```
zzcmp <- compareAPIandRCBoutputs(zrslt, zRCB_Output, ndigits=10)
```

The next table counts the number of decrepancies between the numerical parts of the tables of least-square mean estimates of the average repsonse for each level of the treatment factor.

```
zzcmp$compareLSM
```

```
##                numValue           standardError        degreesOfFreedom
##                       0                       0                       0
##              sampleSize lowerConfidenceInterval upperConfidenceInterval
##                       0                       0                       0
```

The next table counts the number of decrepancies between the numerical parts of the tables of deltas or differences between the least-square means which were compared in the previous table.

```
zzcmp$compareDeltas
```

```
##             differences standardErrorDifference lowerConfidenceInterval
##                       0                       0                       0
## upperConfidenceInterval                  tValue        degreesOfFreedom
##                       0                       0                       0
##        pValueDifference
##                       0
```

The next table counts the number of decrepancies between the numerical estimates of the variance components (estimates of the variance associated with blocking factor and the residuals).

```
zzcmp$compareVarComp
```

```
## [1] 0
```

The next two tables are the variance component tables from Velocity and the development codes.

```
zzcmp$apiVarComp
```

```
##         row varianceEstimates constraint
## 1 subSiteId           80.2761   Positive
## 2  residual          103.9023   Positive
```

```
zzcmp$rcbVarComp
```

```
##         row varianceEstimates constraint
## 1 subSiteId  80.2760997657319   Positive
## 2  residual   103.90231625527   Positive
```

The next table counts the number of decrepancies between the numerical parts of the analysis-of-variance tables associated with the model fits.

```
zzcmp$compareAnova
```

```
##   numeratorDegreesOfFreedom denominatorDegreesOfFreedom
##                           0                           0
##       incrementalSumSquares       conditionalSumSquares
##                           0                           0
##                 probability
##                           0
```

The next two tables give the analysis-of-variance tables. Since they are small they can be recited here.

```
zzcmp$apiAnova
```

```
##      numeratorDegreesOfFreedom denominatorDegreesOfFreedom
## [1,]                         1                           2
## [2,]                        34                          67
##      incrementalSumSquares conditionalSumSquares  probability
## [1,]               614.900               614.900 1.621157e-03
## [2,]                 3.673                 3.673 2.727858e-06
```

```
zzcmp$rcbAnova
```

```
##      degreesFreedomNumerator degreesFreedomDenominator ssIncremental
## [1,]                       1                         2       614.900
## [2,]                      34                        67         3.673
##      ssConditional  probability
## [1,]       614.900 1.621157e-03
## [2,]         3.673 2.727858e-06
```

The next value is the difference between the estimated LSD-values for the model fits.

```
zzcmp$compareLSD
```

```
## [1] 0
```

Conclusion

Base on the outputs and comparisons above, the Velocity and Development outputs are the same.