

Comparison of API and DEV Phase I RCBD1 Results

Mitch Toland

3/25/2020

Introduction

This report compares the results from running two styles of RCBD1 models in two environments: Velocity and development in Domino. The two styles of models are 'global' and 'local' corresponding to analyzing data as a combined dataset and as several datasets subdivided out of the whole dataset. There are 1,160 global datasets and 2,484 local datasets.

As will be demonstrated, the results run on the same data in both environments are the same. This illustrates that the Velocity implementation is the same as the development implementation.

To begin the report, needed code and software needs to be loaded.

```
library(dplyr)

##
## Attaching package: 'dplyr'
## The following objects are masked from 'package:stats':
##
##   filter, lag
## The following objects are masked from 'package:base':
##
##   intersect, setdiff, setequal, union

library(lattice)
library(asreml)

## Licensed to: Monsanto
## Serial Number: 402084150 Expires: 31-jul-2020 (128 days)
# Default git repository branch is "Mi-RCB-dev"
# Use "git checkout master" in the terminal shell to switch to the master branch

source("/repos/RCB4Cloud/R/RCB_MainFunction.R")
source("/repos/RCB4Cloud/Reference/Rscripts/RCB_SupportFunctionsWithasremlPlus.R")
source("/repos/RCB4Cloud/Reference/Rscripts/fromAPI.R")
```


The next code chunk loads lists of Velocity job IDs. These job IDs are used to get the needed data from Product 360.

```
# enter P360 access credentials before running get_ping_token

# ping_token = get_ping_token(client_id, client_secret,FALSE)

# 1Y0ePbjERgputHcELdhVlNhYtg
# 123456789012345678901234567890
#
# All needed job ids are strings 29 characters long similar to the above.
# The first and last characters confuse the
```

```

# data quiry, so extract just the 2nd through 28th characters. However, some job IDs are missing
# so skip those. The function, 'extractJobIds', does all of this and is found in fromAPI.R.

anovaGlobalJobData <- read.csv("/repos/RCB4Cloud/Reference/Data/ValidationData/anova_global_job_output_id")
anovaGlobalJobIds <- extractJobIds(anovaGlobalJobData, "job_output_id")

anovaLocalJobData <- read.csv("/repos/RCB4Cloud/Reference/Data/ValidationData/anova_local_job_output_id")
anovaLocalJobIds <- extractJobIds(anovaLocalJobData, "job_output_id")

# Compare the number of jobs run to the number of non-missing job IDs
# c(nrow(anovaGlobalJobData), length(anovaGlobalJobIds))      # reports 2005 1160
# c(nrow(anovaLocalJobData), length(anovaLocalJobIds))        # reports 6587 2484

```


Three functions are called to get the data from Product 360, extract the relevant pieces, fit a single treatment factor randomized complete block design model (RCBD1) to the input data using the same input parameters, compare the output with the output from Velocity, and then summarize the comparisons across datasets. The three functions are call_API, aovAPIvsRCB, and summaryCompare.

To save time and make it possible for this Rmarkdown to calculate, the precalculated results are read in from a file on disk.

```

##### Do global modles #####
# zmG0      <- sapply(anovaGlobalJobIds,
#                    function(zn){suppressMessages(
#                      summaryCompare(aovAPIvsRCB(call_API(zn, ping_token))))})
# zmG       <- as.data.frame(t(zmG0))
# zmG$JobId  <- rownames(zmG)
# rownames(zmG) <- NULL
# write_json(zmG, "/repos/RCB4Cloud/Reference/Data/GlobalDataValidationResults.json",pretty=TRUE, auto_
zmG <- fromJSON("/repos/RCB4Cloud/Reference/Data/GlobalDataValidationResults.json")

```


The sum of the absolute differences between 5 numerical outputs across all 1,160 datasets is used as a measure of discrepancy between the outputs from the two environments. All parts of the outputs are compared. The following table lists the 5 discrepancy measures. All 5 measures are 0 indicating that the results are the same. The default amount of rounding is 10 decimal places. Thus it can be concluded that the results are the same out to 10 decimal places, which is much more than needed in practice.

```

apply(zmG[, 1:5], 2, sum)

```

##	compareLSM	compareDeltas	compareVarComp	compareAnova	compareLSD
##	0.00e+00	0.00e+00	3.38e-08	0.00e+00	0.00e+00
#	compareLSM	compareDeltas	compareVarComp	compareAnova	compareLSD
#	0	0	0	0	0

The same process used for the global data is repeated for the local datasets.

Again, the precalculated results are read in from a file on disk.

```
##### DO localized models #####
# zml02      <- sapply(anovaLocalJobIds,
#                      function(zn){summaryCompare(aovAPIvsRCB(call_API(zn, ping_token)))})
# zml2       <- as.data.frame(t(zml02))
# zml2$JobId  <- rownames(zml2)
# rownames(zml2) <- NULL
# write_json(zml2, "/repos/RCB4Cloud/Reference/Data/LocalDataValidationResults2.json", pretty=TRUE, auto_unescape=TRUE)
zml2 <- fromJSON("/repos/RCB4Cloud/Reference/Data/LocalDataValidationResults2.json")
```


As is the done for the global datasets, the sum of the absolute differences between the 5 numerical outputs is obtained across all 2,484 datasets. The 5 total discrepancies are all 0, indicating that all parts of the numerical outputs match out to 10 decimal places. The 5 total discrepancies are given in the table below.

```
apply(zml2[, 1:5], 2, sum)
```

##	compareLSM	compareDeltas	compareVarComp	compareAnova	compareLSD
##	0	0	0	0	0
#	<i>compareLSM</i>	<i>compareDeltas</i>	<i>compareVarComp</i>	<i>compareAnova</i>	<i>compareLSD</i>
#	0	0	0	0	0

Conclusion

3,644 datasets were analyzed by the same code implemented in two computing environments. All numbers created for the output match to 10 decimal places. Thus it is concluded that the Velocity implementation matches the development code.