

# RCBD1 Validation

*Mitch Toland*

*2/17/2020*

## Introduction

Forty datasets are analyzed using the the phase 1 version of the ANOVA module. The whole phase 1 version is referred to as RCB since a single treatment factor randomized complete block design is the experimental design and analysis model used. The name of the function doing the actual fitting of the model to the data is called RCB\_ModelFittingFunction.

At issue is whether the version deployed in the non-production Velocity environment produces the same results as the development version available in Domino. Forty dataset run in the non-production environment are used to make the comparison. The input data, input parameters, and output are read into Domino from S3. Using the parameters, the input data are reanalyzed in the development environment in Domino. The estimates from the output means tables are compared with the results from the non-production environment to nine decimal places. Also, the analysis type is predicted for each dataset. Mismatches are counted. There are three analysis types represented in the forty datasets. These labeled “P2”, “P3” and “P4”. “P3” is the same model as “P2” but with BLUP estimation instead of BLUE estimation. “P4” uses BLUE estimation.

The results show that the two codes are producing the same results.

```
suppressMessages(library(jsonlite, quietly=TRUE))
suppressMessages(library(aws.s3, quietly=TRUE))
daeURL <- "http://cran.wustl.edu/src/contrib/Archive/dae/dae_2.7-20.tar.gz"
suppressMessages(install.packages(daeURL, repos=NULL, type="source"))
suppressMessages(library(dae, quietly=TRUE))
suppressMessages(library(dae, quietly = TRUE))
asremlPlusURL <- "http://cran.wustl.edu/src/contrib/Archive/asremlPlus/asremlPlus_2.0-12.tar.gz"
suppressMessages(install.packages(asremlPlusURL, repos=NULL, type="source"))
suppressMessages(library(asreml, quietly=TRUE))
```

```
## Licensed to: Monsanto
## Serial Number: 402084150 Expires: 31-jul-2020 (164 days)
```

```
suppressMessages(library(asremlPlus, quietly=TRUE))

source('/repos/RCB4Cloud/R/RCB_SupportFunctions.R')
source('/repos/RCB4Cloud/R/RCB_MainFunction.R')
```

```
S3BucketName <- "trait-analytics-np"
s3AnovaRunPathsDF <- read.csv("/repos/RCB4Cloud/Reference/AnovaTesting/anova-global-19-12-2019.csv")
s3AnovaRunPaths <- paste0(s3AnovaRunPathsDF[,1])
zpath0 <- s3AnovaRunPaths[1]

s3FileReader <- function(zS3BucketName, zpath0){
  zpath <- trimws(gsub("input.json", "", zpath0))
  bucket.names <- get_bucket(zS3BucketName, prefix = zpath)
  s3filePaths <- sapply(bucket.names, function(zt){zt$Key})
  names(s3filePaths) <- NULL
  s3.data <- lapply(s3filePaths, function(zt){
    curr.name <- paste("s3:/", zS3BucketName, zt, sep = "/")
```

```

    curr.data <- s3read_using(fromJSON, object = URLencode(curr.name))
    return(curr.data)
  })
  names(s3.data) <- c("anovaInputData","anovaOutputList","anovaParameters")
  return(s3.data)
}

readFit <- function(zS3BucketName, zAddress){
  zJdat      <- s3FileReader(zS3BucketName, zAddress)
  zDataIn    <- zJdat$anovaInputData$data
  zAnalysisType <- zJdat$anovaInputData$analysisType
  zJOutput    <- zJdat$anovaOutputList
  zJparams    <- zJdat$anovaParameters
  zDataIn[,zJparams$subSiteId] <- paste0(  zDataIn[, zJparams$subSiteId])
  zDataIn[,zJparams$repId]     <- paste0(  zDataIn[, zJparams$repId])
  zDataIn[,zJparams$factorLevelId] <- paste0(  zDataIn[, zJparams$factorLevelId])
  zDataIn[,zJparams$value]     <- as.numeric(zDataIn[, zJparams$value])
  zDataIn$isPlaceholder <- FALSE
  zJparams$isPlaceholder <- "isPlaceholder"
  rcbFit <- RCB_ModelFittingFunction(zDataIn, zJparams, zAnalysisType)

  if(zAnalysisType == "P1" | zAnalysisType == "P2" | zAnalysisType == "P4")
  {
    cbm <- round(as.numeric(zJOutput$blueTable$value) -
                 as.numeric(rcbFit$blueTable$value),9)
    cbs <- round(as.numeric(zJOutput$blueTable$standardError) -
                 as.numeric(rcbFit$blueTable$standardError),9)
    cbd <- round(as.numeric(zJOutput$blueTable$degreesFreedom) -
                 as.numeric(rcbFit$blueTable$degreesFreedom),9)
    cbo <- cbind(cbm,cbs,cbd)
  }
  if(zAnalysisType == "P3" | zAnalysisType == "P5")
  {
    cbm <- round(as.numeric(zJOutput$blupTable$value) -
                 as.numeric(rcbFit$blupTable$value),9)
    cbs <- round(as.numeric(zJOutput$blupTable$standardError) -
                 as.numeric(rcbFit$blupTable$standardError),9)
    cbc <- round(as.numeric(zJOutput$blupTable$count) -
                 as.numeric(rcbFit$blupTable$count),9)
    cbo <- cbind(cbm,cbs,cbc)
  }
  return(cbo)
}

readFitData <- function(zS3BucketName, zAddress){
  zJdat      <- s3FileReader(zS3BucketName, zAddress)
  zDataIn    <- zJdat$anovaInputData$data
  zAnalysisType <- zJdat$anovaInputData$analysisType
  zJOutput    <- zJdat$anovaOutputList
  zJparams    <- zJdat$anovaParameters
  zDataIn[,zJparams$subSiteId] <- paste0(  zDataIn[, zJparams$subSiteId])
  zDataIn[,zJparams$repId]     <- paste0(  zDataIn[, zJparams$repId])
  zDataIn[,zJparams$factorLevelId] <- paste0(  zDataIn[, zJparams$factorLevelId])

```

```

zDataIn[,zJparams$value]      <- as.numeric(zDataIn[, zJparams$value])
zDataIn$isPlaceholder         <- FALSE
zJparams$isPlaceholder        <- "isPlaceholder"
rcbFit                         <- RCB_ModelFittingFunction(zDataIn, zJparams, zAnalysisType)
olst <- list(dataIn            = zDataIn,
             output           = zJOutput,
             parms            = zJparams,
             analysisType     = zAnalysisType,
             rcbFit           = rcbFit)

return(olst)
}

modelCheck <- function(zS3BucketName, zAddress){
  zJdat      <- s3FileReader(zS3BucketName, zAddress)
  zDataIn    <- zJdat$anovaInputData$data
  zAnalysisType <- zJdat$anovaInputData$analysisType
  zJOutput    <- zJdat$anovaOutputList
  zJparams    <- zJdat$anovaParameters
  zDataIn[,zJparams$subSiteId] <- paste0(  zDataIn[, zJparams$subSiteId])
  zDataIn[,zJparams$repId]     <- paste0(  zDataIn[, zJparams$repId])
  zDataIn[,zJparams$factorLevelId] <- paste0(  zDataIn[, zJparams$factorLevelId])
  zDataIn[,zJparams$value]      <- as.numeric(zDataIn[, zJparams$value])
  agFormula <- paste0(zJparams$repId, "~", zJparams$subSiteId)
  repsPerSubSite <- aggregate(as.formula(agFormula), data=zDataIn, FUN=function(zc){length(unique(zc))})
  # print(agFormula)
  # print(repsPerSubSite)
  numberOfSubSites <- nrow(repsPerSubSite)
  maxNumberOfReps <- max(repsPerSubSite[, zJparams$repId])
  if(numberOfSubSites == 1 & maxNumberOfReps > 1) {PredictedAnalysisType <- "P1"}
  if(numberOfSubSites > 1 & maxNumberOfReps == 1) {PredictedAnalysisType <- "P2_or_P3"}
  if(numberOfSubSites > 1 & maxNumberOfReps > 1) {PredictedAnalysisType <- "P4_or_P5"}
  output <- list(AnalysisPredicted = PredictedAnalysisType,
                 AnalysisUsed      = zAnalysisType)
  return(output)
}

```

```

#chksa <- unlist(lapply(s3AnovaRunPaths, function(zt){sum(abs(readFit(S3BucketName, zt)))}))
chkdc <- do.call(rbind,lapply(s3AnovaRunPaths, function(zt){apply(readFit(S3BucketName, zt),2,function(

```

```

## [1] "input 426 rows"
## [1] "current question code MST"
## [1] "analyzing 426 rows"
## [1] " Model2: Multi LOC Single REP BLUE"

## Running ASReml using analysis type: P2

## finished in 11.89 secs

## Creating output files...

## [1] "input 419 rows"
## [1] "current question code YLD"
## [1] "analyzing 419 rows"
## [1] " Model2: Multi LOC Single REP BLUE"

```

```

## Running ASReml using analysis type: P2
## finished in 1.04 secs
## Creating output files...
## [1] "input 426 rows"
## [1] "current question code MST"
## [1] "analyzing 426 rows"
## [1] " Model2: Multi LOC Single REP BLUE"
## Running ASReml using analysis type: P2
## finished in 1.03 secs
## Creating output files...
## [1] "input 425 rows"
## [1] "current question code GWT"
## [1] "analyzing 425 rows"
## [1] " Model2: Multi LOC Single REP BLUE"
## Running ASReml using analysis type: P2
## finished in 1.04 secs
## Creating output files...
## [1] "input 425 rows"
## [1] "current question code GWT"
## [1] "analyzing 425 rows"
## [1] " Model2: Multi LOC Single REP BLUE"
## Running ASReml using analysis type: P2
## finished in 1.04 secs
## Creating output files...
## [1] "input 426 rows"
## [1] "current question code MST"
## [1] "analyzing 426 rows"
## [1] " Model2: Multi LOC Single REP BLUE"
## Running ASReml using analysis type: P2
## finished in 1.03 secs
## Creating output files...
## [1] "input 419 rows"
## [1] "current question code YLD"
## [1] "analyzing 419 rows"
## [1] " Model2: Multi LOC Single REP BLUE"
## Running ASReml using analysis type: P2
## finished in 1.04 secs
## Creating output files...
## [1] "input 136 rows"
## [1] "current question code HVPOP"
## [1] "analyzing 136 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"

```

```

## Running ASReml using analysis type: P4
## finished in 1.03 secs
## Creating output files...
## [1] "input 138 rows"
## [1] "current question code EHT"
## [1] "analyzing 138 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.03 secs
## Creating output files...
## [1] "input 171 rows"
## [1] "current question code YLD"
## [1] "analyzing 171 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.04 secs
## Creating output files...
## [1] "input 171 rows"
## [1] "current question code MST"
## [1] "analyzing 171 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.04 secs
## Creating output files...
## [1] "input 150 rows"
## [1] "current question code PHT"
## [1] "analyzing 150 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.05 secs
## Creating output files...
## [1] "input 153 rows"
## [1] "current question code YLD"
## [1] "analyzing 153 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.04 secs
## Creating output files...
## [1] "input 162 rows"
## [1] "current question code HVPOP"
## [1] "analyzing 162 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"

```

```

## Running ASReml using analysis type: P4
## finished in 1.04 secs
## Creating output files...
## [1] "input 137 rows"
## [1] "current question code MST"
## [1] "analyzing 137 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.03 secs
## Creating output files...
## [1] "input 137 rows"
## [1] "current question code PHT"
## [1] "analyzing 137 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.03 secs
## Creating output files...
## [1] "input 138 rows"
## [1] "current question code EHT"
## [1] "analyzing 138 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.03 secs
## Creating output files...
## [1] "input 137 rows"
## [1] "current question code MST"
## [1] "analyzing 137 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.03 secs
## Creating output files...
## [1] "input 137 rows"
## [1] "current question code PHT"
## [1] "analyzing 137 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.04 secs
## Creating output files...
## [1] "input 138 rows"
## [1] "current question code EHT"
## [1] "analyzing 138 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"

```

```

## Running ASReml using analysis type: P4
## finished in 1.05 secs
## Creating output files...
## [1] "input 153 rows"
## [1] "current question code YLD"
## [1] "analyzing 153 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.03 secs
## Creating output files...
## [1] "input 137 rows"
## [1] "current question code HVPOP"
## [1] "analyzing 137 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.04 secs
## Creating output files...
## [1] "input 139 rows"
## [1] "current question code EHT"
## [1] "analyzing 139 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.04 secs
## Creating output files...
## [1] "input 162 rows"
## [1] "current question code HVPOP"
## [1] "analyzing 162 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.05 secs
## Creating output files...
## [1] "input 153 rows"
## [1] "current question code PHT"
## [1] "analyzing 153 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.05 secs
## Creating output files...
## [1] "input 171 rows"
## [1] "current question code YLD"
## [1] "analyzing 171 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"

```

```

## Running ASReml using analysis type: P4
## finished in 1.04 secs
## Creating output files...
## [1] "input 171 rows"
## [1] "current question code MST"
## [1] "analyzing 171 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.04 secs
## Creating output files...
## [1] "input 137 rows"
## [1] "current question code PHT"
## [1] "analyzing 137 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.04 secs
## Creating output files...
## [1] "input 138 rows"
## [1] "current question code EHT"
## [1] "analyzing 138 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.04 secs
## Creating output files...
## [1] "input 137 rows"
## [1] "current question code MST"
## [1] "analyzing 137 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.03 secs
## Creating output files...
## [1] "input 137 rows"
## [1] "current question code HVPOP"
## [1] "analyzing 137 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.04 secs
## Creating output files...
## [1] "input 153 rows"
## [1] "current question code YLD"
## [1] "analyzing 153 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"

```



```

## Running ASReml using analysis type: P4
## finished in 1.04 secs
## Creating output files...
## [1] "input 139 rows"
## [1] "current question code EHT"
## [1] "analyzing 139 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.04 secs
## Creating output files...
## [1] "input 171 rows"
## [1] "current question code YLD"
## [1] "analyzing 171 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.04 secs
## Creating output files...
## [1] "input 162 rows"
## [1] "current question code HVPOP"
## [1] "analyzing 162 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.04 secs
## Creating output files...
## [1] "input 153 rows"
## [1] "current question code PHT"
## [1] "analyzing 153 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.04 secs
## Creating output files...
## [1] "input 171 rows"
## [1] "current question code MST"
## [1] "analyzing 171 rows"
## [1] " Model4: Multi LOC MULTI REP BLUE"
## Running ASReml using analysis type: P4
## finished in 1.04 secs
## Creating output files...
## [1] "input 425 rows"
## [1] "current question code MST"
## [1] "analyzing 425 rows"
## [1] " Model3: Multi LOC SINGLE REP BLUP"

```

```
## Running ASReml using analysis type: P3
## finished in 1.03 secs
## Creating output files...
## [1] "input 425 rows"
## [1] "current question code MST"
## [1] "analyzing 425 rows"
## [1] " Model3: Multi LOC SINGLE REP BLUP"
## Running ASReml using analysis type: P3
## finished in 1.04 secs
## Creating output files...
## [1] "input 425 rows"
## [1] "current question code MST"
## [1] "analyzing 425 rows"
## [1] " Model3: Multi LOC SINGLE REP BLUP"
## Running ASReml using analysis type: P3
## finished in 1.04 secs
## Creating output files...
lapModelCheck <- lapply(s3AnovaRunPaths, function(zt){modelCheck(S3BucketName, zt)})
lapModelCheck <- as.data.frame(do.call(rbind, lapModelCheck))
colnames(chkdc) <- c("NumNEMeans", "NumNESE", "NumNEDF")
zdfo <- as.data.frame(cbind(lapModelCheck, chkdc))
zdfo$AnalysisPredicted <- as.character(zdfo$AnalysisPredicted)
zdfo$AnalysisUsed <- as.character(zdfo$AnalysisUsed)
```

As the following table shows the results from the two computing environments match.

The table is printed in four sections to get it to display.

```
print(zdfo[1:10, ])
```

##	AnalysisPredicted	AnalysisUsed	NumNEMeans	NumNESE	NumNEDF
## 1	P2_or_P3	P2	0	0	0
## 2	P2_or_P3	P2	0	0	0
## 3	P2_or_P3	P2	0	0	0
## 4	P2_or_P3	P2	0	0	0
## 5	P2_or_P3	P2	0	0	0
## 6	P2_or_P3	P2	0	0	0
## 7	P2_or_P3	P2	0	0	0
## 8	P4_or_P5	P4	0	0	0
## 9	P4_or_P5	P4	0	0	0
## 10	P4_or_P5	P4	0	0	0

```
print(zdfo[11:20, ])
```

##	AnalysisPredicted	AnalysisUsed	NumNEMeans	NumNESE	NumNEDF
## 11	P4_or_P5	P4	0	0	0
## 12	P4_or_P5	P4	0	0	0
## 13	P4_or_P5	P4	0	0	0

## 14	P4_or_P5	P4	0	0	0
## 15	P4_or_P5	P4	0	0	0
## 16	P4_or_P5	P4	0	0	0
## 17	P4_or_P5	P4	0	0	0
## 18	P4_or_P5	P4	0	0	0
## 19	P4_or_P5	P4	0	0	0
## 20	P4_or_P5	P4	0	0	0

```
print(zdfo[21:30, ])
```

##	AnalysisPredicted	AnalysisUsed	NumNEMeans	NumNESE	NumNEDF
## 21	P4_or_P5	P4	0	0	0
## 22	P4_or_P5	P4	0	0	0
## 23	P4_or_P5	P4	0	0	0
## 24	P4_or_P5	P4	0	0	0
## 25	P4_or_P5	P4	0	0	0
## 26	P4_or_P5	P4	0	0	0
## 27	P4_or_P5	P4	0	0	0
## 28	P4_or_P5	P4	0	0	0
## 29	P4_or_P5	P4	0	0	0
## 30	P4_or_P5	P4	0	0	0

```
print(zdfo[31:40, ])
```

##	AnalysisPredicted	AnalysisUsed	NumNEMeans	NumNESE	NumNEDF
## 31	P4_or_P5	P4	0	0	0
## 32	P4_or_P5	P4	0	0	0
## 33	P4_or_P5	P4	0	0	0
## 34	P4_or_P5	P4	0	0	0
## 35	P4_or_P5	P4	0	0	0
## 36	P4_or_P5	P4	0	0	0
## 37	P4_or_P5	P4	0	0	0
## 38	P2_or_P3	P3	0	0	0
## 39	P2_or_P3	P3	0	0	0
## 40	P2_or_P3	P3	0	0	0