More Analysis

Yuri Lavinas

May 19, 2016

Contents

Faço o ANOVA somente para os modelos "clusterizados"

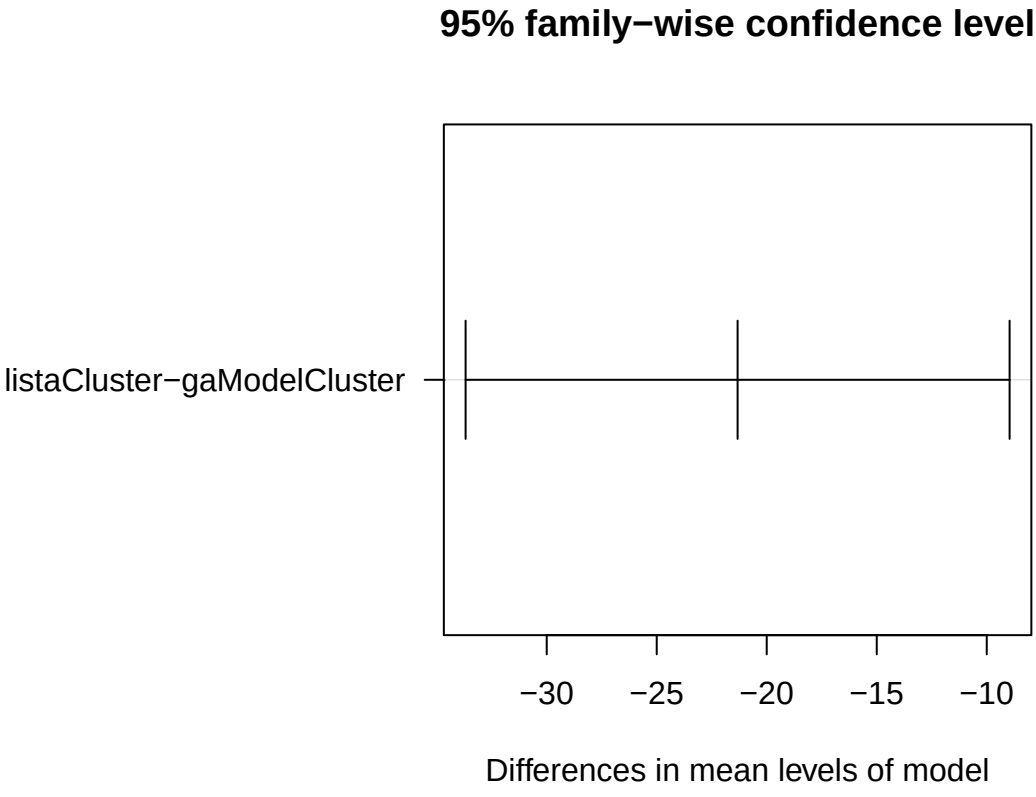
Primeiro crio o data frame somente com os modelos citados

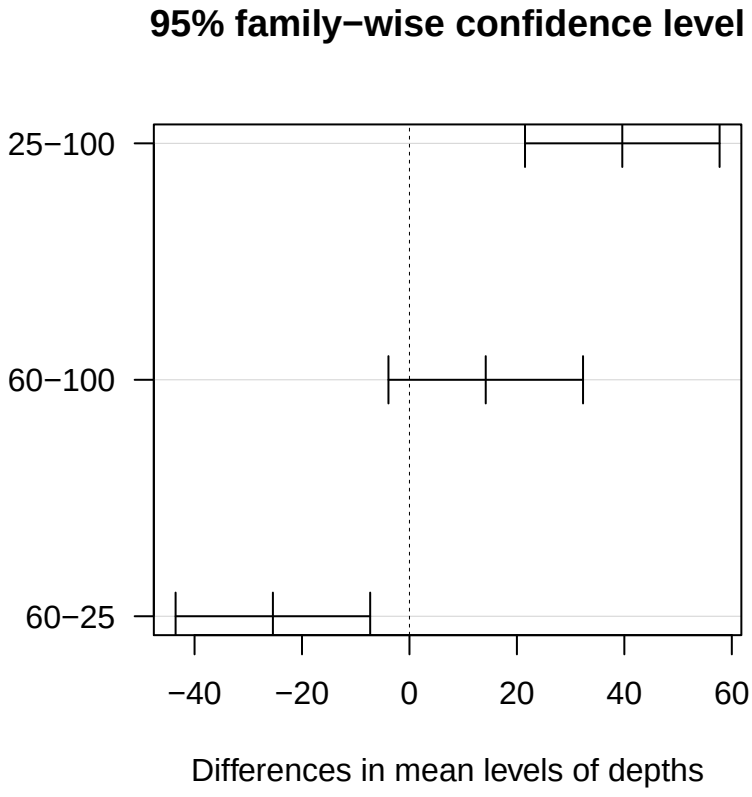
```
subTabela = finalData[finalData$model=='gaModelCluster'|finalData$model=='listaCluster',]
summary(subTabela)
```

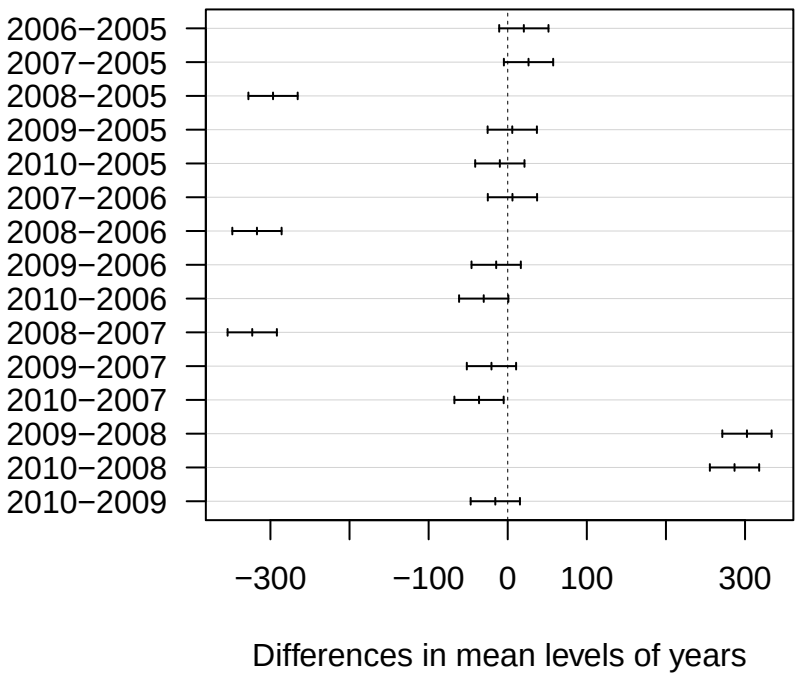
```
##  loglikeValues                  model     depths     years
##  Min.   :-2420   gaModel           :  0   100:480   2005:240
##  1st Qu.:-2032   lista             :  0   25 :480   2006:240
##  Median :-1634   hybrid_gaModel    :  0   60 :480   2007:240
##  Mean   :-1601   hybrid_listaGA_New:  0             2008:240
##  3rd Qu.:-1574   gaModelCluster    :720              2009:240
##  Max.   : -800   listaCluster      :720              2010:240
##       regions
##  Kanto    :360
##  Kansai   :360
##  Tohoku   :360
##  EastJapan:360
##
##
```

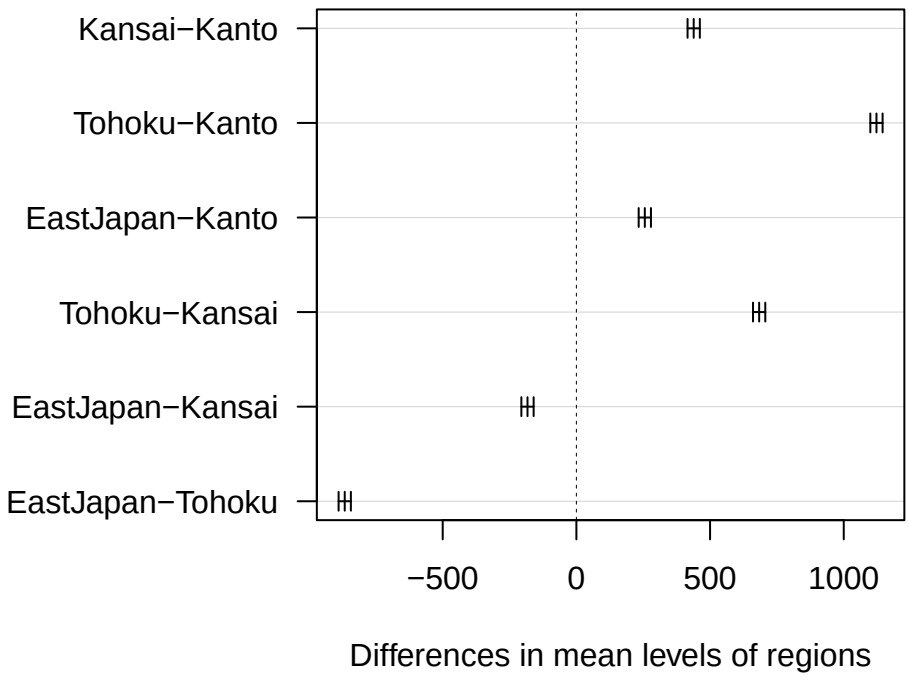
Aplico o anova, com a regressão para modelos, profundidades, anos e regiões.

```
resultANOVA = aov(loglikeValues~model+depths+years+regions , data = subTabela)
tuk = TukeyHSD(resultANOVA)
op <- par(mar = c(5,15,4,2) + 0.1)
plot(tuk,las=1)
```


95% family−wise confidence level
listaCluster−gaModelCluster
−30
−25
−20
−15
−10
Differences in mean levels of model


95% family−wise confidence level
25−100
60−100
60−25
−40
−20
0
20
40
60
Differences in mean levels of depths


95% family−wise confidence level
2006−2005
2007−2005
2008−2005
2009−2005
2010−2005
2007−2006
2008−2006
2009−2006
2010−2006
2008−2007
2009−2007
2010−2007
2009−2008
2010−2008
2010−2009
−300
−100
0
100
300
Differences in mean levels of years




```r
print(tuk)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = loglikeValues ~ model + depths + years + regions, data = subTabela)
##
## $model
##                                  diff       lwr       upr     p adj
## listaCluster-gaModelCluster -21.32415 -33.68752 -8.960766 0.0007354
##
## $depths
##             diff        lwr       upr     p adj
## 25-100  39.61761  21.507440 57.727781 0.0000010
## 60-100  14.19939  -3.910778 32.309563 0.1571677
## 60-25  -25.41822 -43.528389 -7.308048 0.0029173
##
## $years
##                  diff         lwr          upr     p adj
## 2006-2005   20.354382  -10.796654   51.5054187 0.4246300
## 2007-2005   26.305360   -4.845677   57.4563962 0.1534544
## 2008-2005 -296.689791 -327.840828 -265.5387549 0.0000000
## 2009-2005    5.739671  -25.411366   36.8907070 0.9951573
## 2010-2005   -9.985246  -41.136283   21.1657901 0.9428367
## 2007-2006    5.950977  -25.200059   37.1020139 0.9942631
```

```
## 2008-2006 -317.044174 -348.195210 -285.8931372 0.0000000
## 2009-2006  -14.614712  -45.765748   16.5363247 0.7633058
## 2010-2006  -30.339629  -61.490665    0.8114078 0.0613859
## 2008-2007 -322.995151 -354.146188 -291.8441147 0.0000000
## 2009-2007  -20.565689  -51.716726   10.5853472 0.4124771
## 2010-2007  -36.290606  -67.441643   -5.1395697 0.0116890
## 2009-2008  302.429462  271.278425  333.5804984 0.0000000
## 2010-2008  286.704545  255.553509  317.8555814 0.0000000
## 2010-2009  -15.724917  -46.875953   15.4261195 0.7021719
##
## $regions
##                        diff       lwr       upr p adj
## Kansai-Kanto       438.7355  415.8101  461.6609     0
## Tohoku-Kanto      1122.3814 1099.4560 1145.3068     0
## EastJapan-Kanto    255.9241  232.9987  278.8495     0
## Tohoku-Kansai      683.6459  660.7205  706.5713     0
## EastJapan-Kansai -182.8114 -205.7368 -159.8860     0
## EastJapan-Tohoku -866.4573 -889.3827 -843.5319     0
```

Como sugerido, fiz o t.test com a opção Paired = T para uma única região. Acredito que analisar só uma região faça com que a análise não seja independente para essa variável, mas não entendo porque assim eu consigo ver o efeito de um tipo de modelo.

Primeiro crio o novo data.frame para termos dados da região de Kansai, escolhi arbitrariamente

```
subTabela = finalData[finalData$regions=='Kansai',]
summary(subTabela)
```

```
##  loglikeValues                  model     depths     years
##  Min.   :-1703   gaModel           :180   100:360   2005:180
##  1st Qu.:-1639   lista             :180   25 :360   2006:180
##  Median :-1617   hybrid_gaModel    :180   60 :360   2007:180
##  Mean   :-1626   hybrid_listaGA_New:180             2008:180
##  3rd Qu.:-1609   gaModelCluster    :180             2009:180
##  Max.   :-1601   listaCluster      :180             2010:180
##       regions
##  Kanto    :   0
##  Kansai   :1080
##  Tohoku   :   0
##  EastJapan:   0
##
##
```

Agora faço o t.test como disse acima, aplicando para todas as combinações possíveis de modelos.

```
t.test(finalData$model=='gaModel', finalData$model=='lista', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "gaModel" and finalData$model == "lista"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
```

```
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='gaModel', finalData$model=='hybrid_gaModel', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "gaModel" and finalData$model == "hybrid_gaModel"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='gaModel', finalData$model=='hybrid_listaGA_New', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "gaModel" and finalData$model == "hybrid_listaGA_New"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='gaModel', finalData$model=='gaModelCluster', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "gaModel" and finalData$model == "gaModelCluster"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='gaModel', finalData$model=='listaCluster', paired = T)
```

```
##
```

```
##  Paired t-test
##
## data:  finalData$model == "gaModel" and finalData$model == "listaCluster"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='lista', finalData$model=='gaModel', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "lista" and finalData$model == "gaModel"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='lista', finalData$model=='hybrid_gaModel', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "lista" and finalData$model == "hybrid_gaModel"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='lista', finalData$model=='hybrid_listaGA_New', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "lista" and finalData$model == "hybrid_listaGA_New"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='lista', finalData$model=='gaModelCluster', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "lista" and finalData$model == "gaModelCluster"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='lista', finalData$model=='listaCluster', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "lista" and finalData$model == "listaCluster"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='hybrid_gaModel', finalData$model=='lista', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "hybrid_gaModel" and finalData$model == "lista"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='hybrid_gaModel', finalData$model=='gaModel', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "hybrid_gaModel" and finalData$model == "gaModel"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='hybrid_gaModel', finalData$model=='hybrid_listaGA_New', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "hybrid_gaModel" and finalData$model == "hybrid_listaGA_New"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='hybrid_gaModel', finalData$model=='gaModelCluster', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "hybrid_gaModel" and finalData$model == "gaModelCluster"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='hybrid_gaModel', finalData$model=='listaCluster', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "hybrid_gaModel" and finalData$model == "listaCluster"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='hybrid_listaGA_New', finalData$model=='lista', paired = T)
```

```
##
##  Paired t-test
```

```
##
## data:  finalData$model == "hybrid_listaGA_New" and finalData$model == "lista"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='hybrid_listaGA_New', finalData$model=='gaModel', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "hybrid_listaGA_New" and finalData$model == "gaModel"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='hybrid_listaGA_New', finalData$model=='hybrid_gaModel', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "hybrid_listaGA_New" and finalData$model == "hybrid_gaModel"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='hybrid_listaGA_New', finalData$model=='gaModelCluster', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "hybrid_listaGA_New" and finalData$model == "gaModelCluster"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```r
t.test(finalData$model=='hybrid_listaGA_New', finalData$model=='listaCluster', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "hybrid_listaGA_New" and finalData$model == "listaCluster"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```r
t.test(finalData$model=='gaModelCluster', finalData$model=='lista', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "gaModelCluster" and finalData$model == "lista"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```r
t.test(finalData$model=='gaModelCluster', finalData$model=='gaModel', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "gaModelCluster" and finalData$model == "gaModel"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```r
t.test(finalData$model=='gaModelCluster', finalData$model=='hybrid_gaModel', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "gaModelCluster" and finalData$model == "hybrid_gaModel"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
```

```
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='gaModelCluster', finalData$model=='hybrid_listaGA_New', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "gaModelCluster" and finalData$model == "hybrid_listaGA_New"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='gaModelCluster', finalData$model=='listaCluster', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "gaModelCluster" and finalData$model == "listaCluster"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='listaCluster', finalData$model=='lista', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "listaCluster" and finalData$model == "lista"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='listaCluster', finalData$model=='gaModel', paired = T)
```

```
##
##  Paired t-test
```

```
##
## data:  finalData$model == "listaCluster" and finalData$model == "gaModel"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='listaCluster', finalData$model=='hybrid_gaModel', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "listaCluster" and finalData$model == "hybrid_gaModel"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='listaCluster', finalData$model=='hybrid_listaGA_New', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "listaCluster" and finalData$model == "hybrid_listaGA_New"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```

```
t.test(finalData$model=='listaCluster', finalData$model=='gaModelCluster', paired = T)
```

```
##
##  Paired t-test
##
## data:  finalData$model == "listaCluster" and finalData$model == "gaModelCluster"
## t = 0, df = 4319, p-value = 1
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.01722335  0.01722335
## sample estimates:
## mean of the differences
##                       0
```