Paired Desing

Yuri Lavinas

May 27, 2016

Contents

Statistical Analysis

Agora faço o Paired Design t.test aplicando para todas as combinações possíveis de modelos, em todas as regiões e profundidades, para todos os anos.

Baseado nos arquivos que explicam o Paired Desing, escrevi o código a seguir. Porém não entendi porque ao fazer desta forma pode ser considerado um teste pareado. Os slides comparam duas formas de realizar este tipo de teste. Uma delas tem *seta* um parametro da função com **True**, explicitando que é um teste pareado. Já para o outra forma, esse parametro fica com **False**.

```
summary(finalData)
```

```
##  loglikeValues                 model     depths      years
##  Min.   :-3158   gaModel           :720   100:1440   2005:720
##  1st Qu.:-2079   lista             :720   25 :1440   2006:720
##  Median :-1679   hybrid_gaModel    :720   60 :1440   2007:720
##  Mean   :-1702   hybrid_listaGA_New:720              2008:720
##  3rd Qu.:-1602   gaModelCluster    :720              2009:720
##  Max.   : -800   listaCluster      :720              2010:720
##       regions
##  Kanto    :1080
##  Kansai   :1080
##  Tohoku   :1080
##  EastJapan:1080
##
##
```

```
# Summarize the n=30 repeated measures on each Problem:Algorithm combination by their mean value
ttestPaired= function(region){
    subTabela = finalData[finalData$depths==25&finalData$regions==region,]
    aggfinaldata<-aggregate(loglikeValues~years:model, data=subTabela,FUN=mean)
    # Perform paired t-test
    cat('in', region, 'the t.test between the models gaModel and lista is: ')
    difTimes<-with(aggfinaldata,loglikeValues[1:6]-loglikeValues[7:12])
    print(t.test(difTimes))
    cat('in', region, 'the t.test between the models gaModel and hybrid_gaModel is: ')
    difTimes<-with(aggfinaldata,loglikeValues[1:6]-loglikeValues[13:18])
```

```
    print(t.test(difTimes))
    cat('in', region, 'the t.test between the models gaModel and hybrid_listaGA_New is: ')
    difTimes<-with(aggfinaldata,loglikeValues[1:6]-loglikeValues[19:24])
    print(t.test(difTimes))
    cat('in', region, 'the t.test between the models gaModel and gaModelCluster is: ')
    difTimes<-with(aggfinaldata,loglikeValues[1:6]-loglikeValues[25:30])
    print(t.test(difTimes))
    cat('in', region, 'the t.test between the models gaModel and listaCluster is: ')
    difTimes<-with(aggfinaldata,loglikeValues[1:6]-loglikeValues[31:36])
    print(t.test(difTimes))
}

    ttestPaired('Kansai')
```

```
## in Kansai the t.test between the models gaModel and lista is:
##  One Sample t-test
##
## data:  difTimes
## t = 10.637, df = 5, p-value = 0.000127
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  24.06675 39.40531
## sample estimates:
## mean of x
##  31.73603
##
## in Kansai the t.test between the models gaModel and hybrid_gaModel is:
##  One Sample t-test
##
## data:  difTimes
## t = 1.1955, df = 5, p-value = 0.2855
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -4.693085 12.852947
## sample estimates:
## mean of x
##  4.079931
##
## in Kansai the t.test between the models gaModel and hybrid_listaGA_New is:
##  One Sample t-test
##
## data:  difTimes
## t = 17.138, df = 5, p-value = 1.238e-05
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   79.70227 107.83100
## sample estimates:
## mean of x
##  93.76664
##
## in Kansai the t.test between the models gaModel and gaModelCluster is:
##  One Sample t-test
##
```

```
## data:  difTimes
## t = -3.2157, df = 5, p-value = 0.02358
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -12.753743  -1.422024
## sample estimates:
## mean of x
## -7.087883
##
## in Kansai the t.test between the models gaModel and listaCluster is:
##  One Sample t-test
##
## data:  difTimes
## t = 4.7105, df = 5, p-value = 0.005287
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   6.07558 20.67227
## sample estimates:
## mean of x
##  13.37392
```

```
    ttestPaired('Tohoku')
```

```
## in Tohoku the t.test between the models gaModel and lista is:
##  One Sample t-test
##
## data:  difTimes
## t = -1.622, df = 5, p-value = 0.1657
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -13.769220   3.115127
## sample estimates:
## mean of x
## -5.327047
##
## in Tohoku the t.test between the models gaModel and hybrid_gaModel is:
##  One Sample t-test
##
## data:  difTimes
## t = 6.624, df = 5, p-value = 0.001181
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  25.53039 57.91218
## sample estimates:
## mean of x
##  41.72128
##
## in Tohoku the t.test between the models gaModel and hybrid_listaGA_New is:
##  One Sample t-test
##
## data:  difTimes
## t = 3.3329, df = 5, p-value = 0.02071
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
```

```
##   8.308453 64.338805
## sample estimates:
## mean of x
##  36.32363
##
## in Tohoku the t.test between the models gaModel and gaModelCluster is:
##  One Sample t-test
##
## data:  difTimes
## t = -9.4035, df = 5, p-value = 0.0002294
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -36.89030 -21.05111
## sample estimates:
## mean of x
## -28.97071
##
## in Tohoku the t.test between the models gaModel and listaCluster is:
##  One Sample t-test
##
## data:  difTimes
## t = -6.257, df = 5, p-value = 0.001529
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -33.37542 -13.93769
## sample estimates:
## mean of x
## -23.65656
```

```
    ttestPaired('EastJapan')
```

```
## in EastJapan the t.test between the models gaModel and lista is:
##  One Sample t-test
##
## data:  difTimes
## t = 1.9129, df = 5, p-value = 0.114
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -10.07453  68.68307
## sample estimates:
## mean of x
##  29.30427
##
## in EastJapan the t.test between the models gaModel and hybrid_gaModel is:
##  One Sample t-test
##
## data:  difTimes
## t = 6.5282, df = 5, p-value = 0.001262
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   61.94934 142.42401
## sample estimates:
## mean of x
##  102.1867
```

```
##
## in EastJapan the t.test between the models gaModel and hybrid_listaGA_New is:
##  One Sample t-test
##
## data:  difTimes
## t = 11.564, df = 5, p-value = 8.482e-05
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  156.1337 245.3855
## sample estimates:
## mean of x
##  200.7596
##
## in EastJapan the t.test between the models gaModel and gaModelCluster is:
##  One Sample t-test
##
## data:  difTimes
## t = -8.8802, df = 5, p-value = 0.0003012
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -109.10009  -60.11634
## sample estimates:
## mean of x
## -84.60822
##
## in EastJapan the t.test between the models gaModel and listaCluster is:
##  One Sample t-test
##
## data:  difTimes
## t = -5.4451, df = 5, p-value = 0.002837
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -93.31209 -33.46295
## sample estimates:
## mean of x
## -63.38752
```

```
    ttestPaired('Kanto')
```

```
## in Kanto the t.test between the models gaModel and lista is:
##  One Sample t-test
##
## data:  difTimes
## t = 4.1215, df = 5, p-value = 0.00916
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##   4.985032 21.509870
## sample estimates:
## mean of x
##  13.24745
##
## in Kanto the t.test between the models gaModel and hybrid_gaModel is:
##  One Sample t-test
##
```

```
## data:  difTimes
## t = 1.3808, df = 5, p-value = 0.2259
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -1.542245  5.122177
## sample estimates:
## mean of x
##  1.789966
##
## in Kanto the t.test between the models gaModel and hybrid_listaGA_New is:
##  One Sample t-test
##
## data:  difTimes
## t = 5.8073, df = 5, p-value = 0.002136
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  23.43230 60.65227
## sample estimates:
## mean of x
##  42.04228
##
## in Kanto the t.test between the models gaModel and gaModelCluster is:
##  One Sample t-test
##
## data:  difTimes
## t = -3.3043, df = 5, p-value = 0.02137
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -9.408156 -1.175006
## sample estimates:
## mean of x
## -5.291581
##
## in Kanto the t.test between the models gaModel and listaCluster is:
##  One Sample t-test
##
## data:  difTimes
## t = 1.1659, df = 5, p-value = 0.2963
## alternative hypothesis: true mean is not equal to 0
## 95 percent confidence interval:
##  -3.474332  9.241409
## sample estimates:
## mean of x
##  2.883539
```

Statistical Analysis - Checking Results

Como fiquei bastante em dúvida em relação a forma como fiz os testes, tentei um outra abordagem. Como obtive resultados iguais para um caso de teste, entendi que são iguais.

Ambas abordagens são baseadas no slides do Felipe Campelo, que o Claus enviou.

Conclusions

To explore the results and conclusions from the tests from the ANOVA, we applied a paired Student's t test.