

Accidentalidad en Medellín

Valentina Serrato Henao

3/11/2020

Primero se carga el archivo previamente comprimido a un .RData

```
load(file.choose())
```

```
#librerias necesarias  
require(mixlm)
```

```
## Loading required package: mixlm
```

```
## Warning: package 'mixlm' was built under R version 4.0.3
```

```
##
```

```
## Attaching package: 'mixlm'
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
##      glm, lm
```

```
library(randomForest)
```

```
## Warning: package 'randomForest' was built under R version 4.0.3
```

```
## randomForest 4.6-14
```

```
## Type rfNews() to see new features/changes/bug fixes.
```

```
library(tree)
```

```
## Warning: package 'tree' was built under R version 4.0.3
```

```
library(MASS)
```

```
## Warning: package 'MASS' was built under R version 4.0.3
```

Se divide el conjunto de datos en datos de entrenamiento(2015-2017) y datos de prueba (2018)

```
## Seleccion del conjunto de entrenamiento y el conjunto de prueba
train<-accidentalidad[-which(accidentalidad$PERIODO=="2018"),]
test<-accidentalidad[which(accidentalidad$PERIODO=="2018"),]
```

Arbol de regresión

Se plantea un modelo de arbol de regresión que permita predecir el *número de accidentes* usando como predictoras el día, el mes, el nombre del día, la comúna, el riesgo de accidentalidad del barrio, y un conjunto de variables dummy que determina si es o no un día importante.

```
library(rpart)
```

```
## Warning: package 'rpart' was built under R version 4.0.3
```

```
#Modelo de arbol de regresion para el número de accidentes
NumAccidentes_test<-test$'Numero de accidentes'
modelo1 <- rpart('Numero de accidentes'~MES+DIA+DIA_NOMBRE+COMUNA+puente_festivo+puente_reyes+semana_santa)
a1 <- predict(modelo1, test)

#MSE del conjunto de entrenamiento
msetrain<-mean((train$'Numero de accidentes' - predict(modelo1))^2)

#MSE del conjunto de prueba
msetest<- mean((NumAccidentes_test - a1)^2)

#variacion de los MSE
variacion_mse<-(msetest/msetrain)-1
msetest
```

```
## [1] 5.06317
```

```
msetrain
```

```
## [1] 5.169926
```

```
variacion_mse
```

```
## [1] -0.02064933
```

```
cat("R Cuadrado \n")
```

```
## R Cuadrado
```

```
R2 = cor(NumAccidentes_test, a1)^2
R2
```

```
## [1] 0.6667385
```



```

## Node number 3: 3186 observations,      complexity param=0.2679659
##   mean=11.11802, MSE=103.73
##   left son=6 (1725 obs) right son=7 (1461 obs)
##   Primary splits:
##       Riesgo           splits as  RLL,      improve=0.696999700, (0 missing)
##       DIA_NOMBRE       splits as  LRRRRRRR, improve=0.041514600, (0 missing)
##       puente_festivo   splits as  RL,       improve=0.014791390, (0 missing)
##       puente_semana_santa splits as RL,      improve=0.003022493, (0 missing)
##       MES              < 1.5 to the left, improve=0.002665055, (0 missing)
##   Surrogate splits:
##       DIA_NOMBRE       splits as  RLLLLLLL, agree=0.549, adj=0.016, (0 split)
##       puente_festivo   splits as  LR,       agree=0.544, adj=0.005, (0 split)
##       puente_semana_santa splits as LR,      agree=0.542, adj=0.002, (0 split)
##       festivo_entre_semana splits as LR,      agree=0.542, adj=0.001, (0 split)
##
## Node number 4: 21901 observations
##   mean=2.27323, MSE=2.1636
##
## Node number 5: 19199 observations,      complexity param=0.04482908
##   mean=4.355696, MSE=8.658261
##   left son=10 (5388 obs) right son=11 (13811 obs)
##   Primary splits:
##       Riesgo           splits as  RLR, improve=0.208111100, (0 missing)
##       COMUNA           splits as  --L---L-----LL---R--L----, improve=0.066945120, (0 missing)
##       DIA_NOMBRE       splits as  LRRRRRRR, improve=0.033984540, (0 missing)
##       puente_festivo   splits as  RL, improve=0.015483120, (0 missing)
##       puente_semana_santa splits as RL, improve=0.003449238, (0 missing)
##
## Node number 6: 1725 observations
##   mean=3.292754, MSE=4.600092
##
## Node number 7: 1461 observations,      complexity param=0.04246632
##   mean=20.35729, MSE=63.10848
##   left son=14 (209 obs) right son=15 (1252 obs)
##   Primary splits:
##       DIA_NOMBRE       splits as  LRRRRRRR, improve=0.35542780, (0 missing)
##       puente_festivo   splits as  RL,      improve=0.13151730, (0 missing)
##       puente_semana_santa splits as RL,      improve=0.02679235, (0 missing)
##       puente_reyes     splits as  RL,      improve=0.01692679, (0 missing)
##       MES              < 1.5 to the left, improve=0.01640579, (0 missing)
##
## Node number 10: 5388 observations
##   mean=2.20657, MSE=2.107106
##
## Node number 11: 13811 observations,      complexity param=0.01138711
##   mean=5.194121, MSE=8.709186
##   left son=22 (5119 obs) right son=23 (8692 obs)
##   Primary splits:
##       COMUNA           splits as  --R---R-----LL---R--R----, improve=0.073055970, (0 missing)
##       DIA_NOMBRE       splits as  LRRRRRRR, improve=0.057029500, (0 missing)
##       puente_festivo   splits as  RL, improve=0.025923330, (0 missing)
##       MES              < 1.5 to the left, improve=0.006429498, (0 missing)
##       puente_semana_santa splits as RL, improve=0.005567587, (0 missing)
##

```

```

## Node number 14: 209 observations
##   mean=8.76555, MSE=10.76322
##
## Node number 15: 1252 observations,   complexity param=0.01120086
##   mean=22.29233, MSE=45.67173
##   left son=30 (108 obs) right son=31 (1144 obs)
##   Primary splits:
##     puente_festivo      splits as  RL,      improve=0.15116250, (0 missing)
##     puente_semana_santa splits as  RL,      improve=0.04378930, (0 missing)
##     DIA_NOMBRE          splits as  -RLRRLR, improve=0.04134922, (0 missing)
##     festivo_entre_semana splits as  RL,      improve=0.03089104, (0 missing)
##     MES                 < 1.5 to the left, improve=0.02898165, (0 missing)
##   Surrogate splits:
##     puente_semana_santa splits as  RL, agree=0.923, adj=0.111, (0 split)
##     puente_reyes        splits as  RL, agree=0.920, adj=0.074, (0 split)
##
## Node number 22: 5119 observations
##   mean=4.154718, MSE=8.04717
##
## Node number 23: 8692 observations
##   mean=5.806259, MSE=8.088097
##
## Node number 30: 108 observations
##   mean=13.74074, MSE=76.82167
##
## Node number 31: 1144 observations
##   mean=23.09965, MSE=35.17538

```

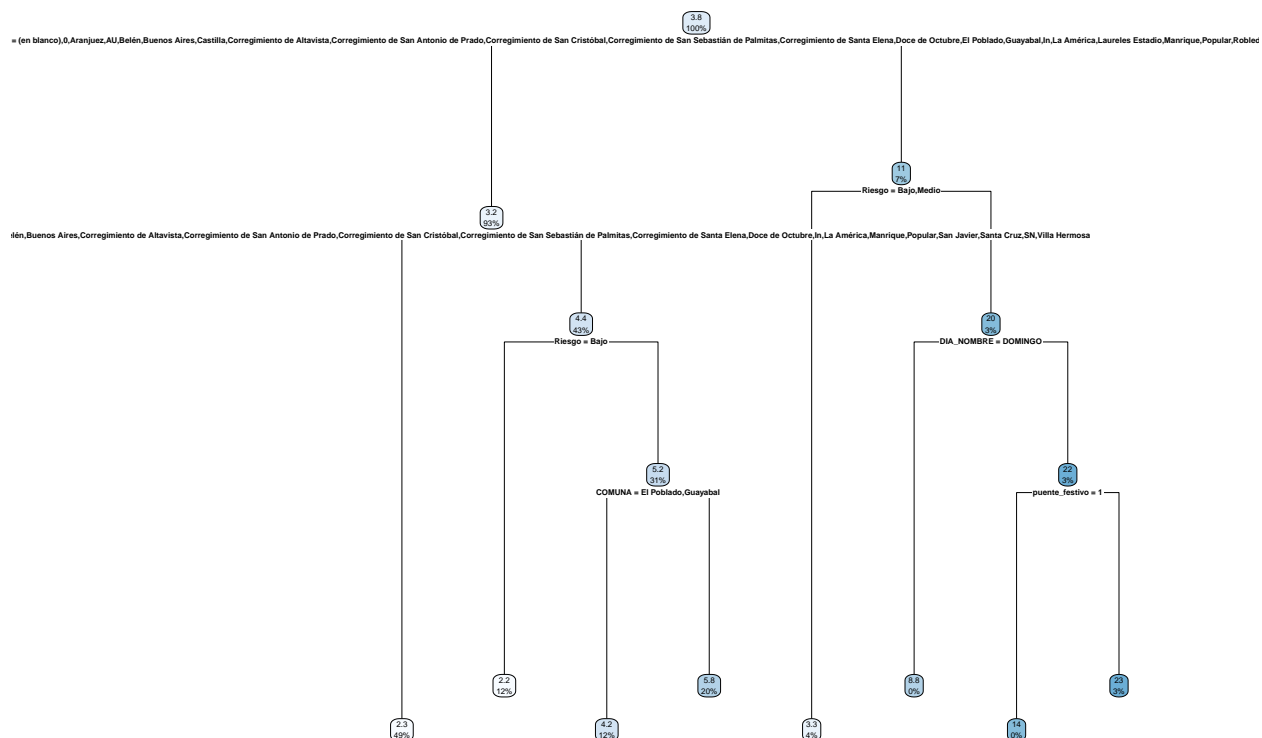
Podemos ver que la variación de los MSE de prueba y entrenamiento es del 2%, también se ve que tenemos un $R^2 = 0.6667$ y $R^2_{adj} = 0.6666$, por lo tanto el modelo sí explica en un 66.67% la variabilidad del número de accidentes.

Se utilizará el método forward para depurar un poco el modelo y encontrar uno más parsimonioso.

```
library(rpart.plot)
```

```
## Warning: package 'rpart.plot' was built under R version 4.0.3
```

```
rpart.plot(modelo1)
```



Se ajusta un modelo de regresión lineal múltiple para el número de *accidentes con solo daños*, usando como uno de los predictores el número de accidentes que resulta del modelo anterior, también el día, el mes, el nombre del día, la comúna, el riesgo de accidentalidad del barrio, y un conjunto de variables dummy que determina si es o no un día importante.

```
#Modelo de regresión lineal multiple para el número de accidentes con solo daños
NumAccidentesdaños_test<-test$'Accidente con solo daños'
modelo2<-lm('Accidente con solo daños'~MES+DIA+DIA_NOMBRE+COMUNA+puente_festivo+puente_reyes+semana_san
a2<-predict(modelo2, test)
#MSE del conjunto de entrenamiento
msetrain2<-mean((train$'Accidente con solo daños'- predict(modelo2))^2)
#MSE del conjunto de prueba
msetest2<- mean((NumAccidentesdaños_test - a2)^2)
#variacion de los MSE
variacion_mse2<-(msetest2/msetrain2)-1
msetest2
```

```
## [1] 1.014397
```

```
msetrain2
```

```
## [1] 1.036596
```

```
variacion_mse2
```

```
## [1] -0.02141495
```

```
summary(modelo2)
```

```
##
## Call:
## lm(formula = 'Accidente con solo daños' ~ MES + DIA + DIA_NOMBRE +
##     COMUNA + puente_festivo + puente_reyes + semana_santa + feria_flores +
##     puente_semana_santa + festivo_entre_semana + no_festivo_importante +
##     Riesgo + 'Numero de accidentes', data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -7.2981 -0.5539 -0.0228  0.6017  9.6401
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.4291843   0.0685458   -6.261 3.85e-10 ***
## MES              0.0068145   0.0014445    4.718 2.39e-06 ***
## DIA              0.0000341   0.0005579    0.061 0.951255
## DIA_NOMBRE1    -0.1510402   0.0124921  -12.091 < 2e-16 ***
## DIA_NOMBRE2    -0.0121809   0.0120160   -1.014 0.310722
## DIA_NOMBRE3    -0.0014694   0.0119842   -0.123 0.902414
## DIA_NOMBRE4     0.0558886   0.0119366    4.682 2.85e-06 ***
## DIA_NOMBRE5     0.0118315   0.0119346    0.991 0.321516
## DIA_NOMBRE6     0.0135517   0.0120130    1.128 0.259290
## COMUNA1         0.2995826   0.2166520    1.383 0.166738
## COMUNA2        -0.2885437   0.5662627   -0.510 0.610364
## COMUNA3        -0.2774553   0.0407198   -6.814 9.63e-12 ***
## COMUNA4        -0.0127332   0.2641004   -0.048 0.961546
## COMUNA5         0.1287637   0.0393712    3.271 0.001074 **
## COMUNA6        -0.0331051   0.0408263   -0.811 0.417441
## COMUNA7        -0.1586241   0.0396060   -4.005 6.21e-05 ***
## COMUNA8        -0.0720027   0.0642390   -1.121 0.262355
## COMUNA9         0.1449582   0.0503669    2.878 0.004003 **
## COMUNA10        0.0077416   0.0458307    0.169 0.865862
## COMUNA11        0.4103700   0.3282882    1.250 0.211295
## COMUNA12       -0.0124876   0.0543005   -0.230 0.818114
## COMUNA13       -0.4941556   0.0436854  -11.312 < 2e-16 ***
## COMUNA14        0.7212569   0.0394452   18.285 < 2e-16 ***
## COMUNA15        0.1453384   0.0403405    3.603 0.000315 ***
## COMUNA16        0.1684136   0.1025856    1.642 0.100662
## COMUNA17        0.1436803   0.0410604    3.499 0.000467 ***
## COMUNA18        0.5280426   0.0416784   12.669 < 2e-16 ***
## COMUNA19        0.3403843   0.0404641    8.412 < 2e-16 ***
## COMUNA20       -0.2845198   0.0414588   -6.863 6.85e-12 ***
## COMUNA21       -0.3001440   0.0461493   -6.504 7.92e-11 ***
## COMUNA22       -0.3886988   0.0405671   -9.582 < 2e-16 ***
## COMUNA23       -0.1611885   0.0430843   -3.741 0.000183 ***
## COMUNA24       -0.1914753   0.0464556   -4.122 3.77e-05 ***
## COMUNA25       -0.2489769   0.4907230   -0.507 0.611899
## puente_festivo1  0.0339462   0.0092288    3.678 0.000235 ***
## puente_reyes1   -0.0224406   0.0303867   -0.738 0.460215
## semana_santa1   0.0069777   0.0211452    0.330 0.741408
## feria_flores1   0.0014472   0.0147683    0.098 0.921939
```

```
## puente_semana_santa1 -0.0402875 0.0341871 -1.178 0.238627
## festivo_entre_semana1 0.1214204 0.0324081 3.747 0.000179 ***
## no_festivo_importante1 -0.0336913 0.0234840 -1.435 0.151395
## Riesgo1 0.0435822 0.0105259 4.140 3.47e-05 ***
## Riesgo2 0.0627567 0.0086473 7.257 4.01e-13 ***
## 'Numero de accidentes' 0.5208512 0.0015114 344.611 < 2e-16 ***
## ---
## Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## s: 1.019 on 44242 degrees of freedom
## Multiple R-squared: 0.8354,
## Adjusted R-squared: 0.8352
## F-statistic: 5222 on 43 and 44242 DF, p-value: < 2.2e-16
```

Podemos ver que la variación de los MSE de prueba y entrenamiento es del 2% un valor bajo, también se ve que tenemos un $R^2 = 0.8354$ y $R^2_{adj} = 0.8352$, al ser valores casi iguales entonces se puede concluir que el modelo sí explica en un 84% la variabilidad del número de accidentes con solo daños.

Se utilizará el método forward para depurar un poco el modelo y encontrar uno más parsimonioso.

```
forward(modelo2)
```

```
## Forward selection, alpha-to-enter: 0.2
##
## Full model: 'Accidente con solo daños' ~ MES + DIA + DIA_NOMBRE + COMUNA +
##     puente_festivo + puente_reyes + semana_santa + feria_flores +
##     puente_semana_santa + festivo_entre_semana + no_festivo_importante +
##     Riesgo + 'Numero de accidentes'
## <environment: 0x000000004ff3eea8>

## Warning: Using formula(x) is deprecated when x is a character vector of length > 1.
##   Consider formula(paste(x, collapse = " ")) instead.

## Warning: Using formula(x) is deprecated when x is a character vector of length > 1.
##   Consider formula(paste(x, collapse = " ")) instead.

##           Step    RSS      AIC  R2pred      Cp    F value
## COMUNA          1 190272 64611.4 0.31736 139137.962 8.2479e+02
## Riesgo          1 226602 72304.0 0.31736 139137.962 5.1115e+03
## 'Numero de accidentes' 1 51061 6307.9 0.31736 139137.962 1.9761e+05
## Riesgo1         2 173755 60593.9 0.37655 123224.004 2.1036e+03
## 'Numero de accidentes'1 2 46442 2158.7 0.37655 123224.004 1.3707e+05
## 'Numero de accidentes'2 3 46181 1913.2 0.83425 278.020 1.2226e+05
## DIA_NOMBRE      4 45959 1712.5 0.83500 76.717 3.5517e+01
## MES             5 45937 1693.4 0.83507 57.696 2.1011e+01
## festivo_entre_semana 6 45923 1681.3 0.83512 45.592 1.4102e+01
## puente_festivo   7 45911 1671.8 0.83515 36.088 1.1505e+01
## no_festivo_importante 8 45909 1671.8 0.83515 36.050 2.0376e+00
##           Pr(>F)
## COMUNA          < 2.2e-16 ***
## Riesgo          < 2.2e-16 ***
## 'Numero de accidentes' < 2.2e-16 ***
## Riesgo1         < 2.2e-16 ***
```



```
## 'Numero de accidentes'1 < 2.2e-16 ***
## 'Numero de accidentes'2 < 2.2e-16 ***
## DIA_NOMBRE < 2.2e-16 ***
## MES 4.58e-06 ***
## festivo_entre_semana 0.0001734 ***
## puente_festivo 0.0006948 ***
## no_festivo_importante 0.1534551
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Call:
## lm(formula = 'Accidente con solo daños' ~ COMUNA + Riesgo + 'Numero de accidentes' +
##     DIA_NOMBRE + MES + festivo_entre_semana + puente_festivo +
##     no_festivo_importante, data = train)
##
## Coefficients:
##             (Intercept)             COMUNA1             COMUNA2
##             -0.477048             0.299758             -0.287300
##             COMUNA3             COMUNA4             COMUNA5
##             -0.277248             -0.013731             0.128882
##             COMUNA6             COMUNA7             COMUNA8
##             -0.032941             -0.158548             -0.071564
##             COMUNA9             COMUNA10            COMUNA11
##             0.144686             0.007632             0.406816
##             COMUNA12            COMUNA13            COMUNA14
##             -0.012209            -0.493864             0.721331
##             COMUNA15            COMUNA16            COMUNA17
##             0.145171             0.168236             0.143718
##             COMUNA18            COMUNA19            COMUNA20
##             0.528629             0.340588             -0.284452
##             COMUNA21            COMUNA22            COMUNA23
##             -0.299883            -0.388436             -0.161117
##             COMUNA24            COMUNA25             Riesgo1
##             -0.191502            -0.248269             0.043755
##             Riesgo2 'Numero de accidentes'             DIA_NOMBRE1
##             0.062618             0.520786             -0.151468
##             DIA_NOMBRE2            DIA_NOMBRE3            DIA_NOMBRE4
##             -0.011016            -0.002536             0.055911
##             DIA_NOMBRE5            DIA_NOMBRE6             MES
##             0.011813             0.013043             0.006547
##     festivo_entre_semana1     puente_festivo1     no_festivo_importante1
##             0.121798             0.029053             -0.033512
```

Se ajusta un modelo de regresión lineal múltiple para el número de *accidentes con heridos*, usando como uno de los predictores el número de accidentes que resulta del primer modelo, también el día, el mes, el nombre del día, la comúna, el riesgo de accidentalidad del barrio, y un conjunto de variables dummy que determina si es o no un día importante.

```
#Modelo de regresión lineal multiple para el número de accidentes con heridos
NumAccidentesheridos_test<-test$'Accidentes con Heridos'
modelo3<-lm('Accidentes con Heridos'~MES+DIA+DIA_NOMBRE+COMUNA+puente_festivo+puente_reyes+semana_santa
a3<-predict(modelo3, test)
#MSE del conjunto de entrenamiento
```

```

msetrain3<-mean((train$'Accidentes con Heridos' - predict(modelo3))^2)
#MSE del conjunto de prueba
msetest3<- mean((NumAccidentesheridos_test - a3)^2)
#variacion de los MSE
variacion_mse3<-(msetrain3/msetest3)-1
msetest3

```

```
## [1] 1.004133
```

```
msetrain3
```

```
## [1] 1.029467
```

```
variacion_mse3
```

```
## [1] 0.02522995
```

```
summary(modelo3)
```

```

##
## Call:
## lm(formula = 'Accidentes con Heridos' ~ MES + DIA + DIA_NOMBRE +
##      COMUNA + puente_festivo + puente_reyes + semana_santa + feria_flores +
##      puente_semana_santa + festivo_entre_semana + no_festivo_importante +
##      Riesgo + 'Numero de accidentes', data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.4791 -0.6043  0.0232  0.5508  7.4496
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)      0.4118924   0.0683097    6.030 1.65e-09 ***
## MES              -0.0064576   0.0014395   -4.486 7.28e-06 ***
## DIA              -0.0001534   0.0005560   -0.276 0.782548
## DIA_NOMBRE1       0.1413750   0.0124490   11.356 < 2e-16 ***
## DIA_NOMBRE2       0.0175322   0.0119746    1.464 0.143169
## DIA_NOMBRE3      -0.0019363   0.0119429   -0.162 0.871202
## DIA_NOMBRE4      -0.0547757   0.0118955   -4.605 4.14e-06 ***
## DIA_NOMBRE5      -0.0073285   0.0118935   -0.616 0.537781
## DIA_NOMBRE6      -0.0165969   0.0119716   -1.386 0.165645
## COMUNA1          -0.3814604   0.2159058   -1.767 0.077270 .
## COMUNA2           0.3046012   0.5643123    0.540 0.589356
## COMUNA3           0.2711304   0.0405795    6.681 2.39e-11 ***
## COMUNA4          -0.0451907   0.2631907   -0.172 0.863672
## COMUNA5          -0.1158523   0.0392356   -2.953 0.003151 **
## COMUNA6           0.0404964   0.0406857    0.995 0.319572
## COMUNA7           0.1657098   0.0394696    4.198 2.69e-05 ***
## COMUNA8           0.0702066   0.0640178    1.097 0.272790
## COMUNA9          -0.1553568   0.0501934   -3.095 0.001968 **
## COMUNA10         -0.0062812   0.0456728   -0.138 0.890616

```

```
## COMUNA11      -0.3976333  0.3271574  -1.215  0.224213
## COMUNA12      0.0169490  0.0541134   0.313  0.754120
## COMUNA13      0.4971154  0.0435350  11.419 < 2e-16 ***
## COMUNA14     -0.7036044  0.0393094 -17.899 < 2e-16 ***
## COMUNA15     -0.1393602  0.0402016  -3.467  0.000528 ***
## COMUNA16     -0.1749994  0.1022323  -1.712  0.086944 .
## COMUNA17     -0.1339403  0.0409190  -3.273  0.001064 **
## COMUNA18     -0.5246591  0.0415348 -12.632 < 2e-16 ***
## COMUNA19     -0.3237815  0.0403247  -8.029  1.00e-15 ***
## COMUNA20      0.2861871  0.0413160   6.927  4.36e-12 ***
## COMUNA21      0.3071802  0.0459904   6.679  2.43e-11 ***
## COMUNA22      0.4036221  0.0404273   9.984 < 2e-16 ***
## COMUNA23      0.1635935  0.0429359   3.810  0.000139 ***
## COMUNA24      0.1921081  0.0462956   4.150  3.34e-05 ***
## COMUNA25      0.2624756  0.4890327   0.537  0.591461
## puente_festivo1 -0.0307303  0.0091970  -3.341  0.000834 ***
## puente_reyes1   0.0259633  0.0302821   0.857  0.391239
## semana_santa1  -0.0042150  0.0210724  -0.200  0.841460
## feria_flores1   0.0091067  0.0147174   0.619  0.536071
## puente_semana_santa1 0.0338915  0.0340694   0.995  0.319849
## festivo_entre_semana1 -0.1233307  0.0322965  -3.819  0.000134 ***
## no_festivo_importante1 0.0326076  0.0234032   1.393  0.163536
## Riesgo1        -0.0477085  0.0104897  -4.548  5.43e-06 ***
## Riesgo2        -0.0634423  0.0086175  -7.362  1.84e-13 ***
## 'Numero de accidentes' 0.4731846  0.0015062 314.156 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## s: 1.015 on 44242 degrees of freedom
## Multiple R-squared:  0.7803,
## Adjusted R-squared:  0.78
## F-statistic: 3653 on 43 and 44242 DF,  p-value: < 2.2e-16
```

Podemos ver que la variación de los MSE de prueba y entrenamiento es del 2% un valor bajo, también se ve que tenemos un $R^2 = 0.7803$ y $R^2_{adj} = 0.78$, al ser valores casi iguales entonces se puede concluir que el modelo sí explica en un 78% la variabilidad del número de accidentes con heridos.

Se utilizará el método forward para depurar un poco el modelo y encontrar uno más parsimonioso.

```
forward(modelo3)
```

```
## Forward selection, alpha-to-enter: 0.2
##
## Full model: 'Accidentes con Heridos' ~ MES + DIA + DIA_NOMBRE + COMUNA +
##   puente_festivo + puente_reyes + semana_santa + feria_flores +
##   puente_semana_santa + festivo_entre_semana + no_festivo_importante +
##   Riesgo + 'Numero de accidentes'
## <environment: 0x000000004feb14a0>

## Warning: Using formula(x) is deprecated when x is a character vector of length > 1.
##   Consider formula(paste(x, collapse = " ")) instead.

## Warning: Using formula(x) is deprecated when x is a character vector of length > 1.
##   Consider formula(paste(x, collapse = " ")) instead.
```

```

##          Step    RSS      AIC  R2pred          Cp    F value
## COMUNA          1 162802 57706.5 0.21473 113751.211 4.8572e+02
## Riesgo          1 178953 61849.4 0.21473 113751.211 3.5281e+03
## 'Numero de accidentes' 1 50664 5962.5 0.21473 113751.211 1.3706e+05
## Riesgo1         2 148527 53646.3 0.28350 99902.062 2.1269e+03
## 'Numero de accidentes'1 2 46116 1847.3 0.28350 99902.062 1.1199e+05
## 'Numero de accidentes'2 3 45835 1580.3 0.77884 250.581 9.9157e+04
## DIA_NOMBRE       4 45640 1403.2 0.77972 73.092 3.1554e+01
## MES              5 45620 1386.1 0.77981 55.918 1.9165e+01
## festivo_entre_semana 6 45605 1373.5 0.77987 43.390 1.4526e+01
## puente_festivo    7 45595 1366.2 0.77991 36.093 9.2970e+00
## no_festivo_importante 8 45593 1366.3 0.77991 36.208 1.8856e+00
##          Pr(>F)
## COMUNA          < 2.2e-16 ***
## Riesgo          < 2.2e-16 ***
## 'Numero de accidentes' < 2.2e-16 ***
## Riesgo1         < 2.2e-16 ***
## 'Numero de accidentes'1 < 2.2e-16 ***
## 'Numero de accidentes'2 < 2.2e-16 ***
## DIA_NOMBRE       < 2.2e-16 ***
## MES              1.202e-05 ***
## festivo_entre_semana 0.0001384 ***
## puente_festivo    0.0022966 **
## no_festivo_importante 0.1697115
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

##
## Call:
## lm(formula = 'Accidentes con Heridos' ~ COMUNA + Riesgo + 'Numero de accidentes' +
##     DIA_NOMBRE + MES + festivo_entre_semana + puente_festivo +
##     no_festivo_importante, data = train)
##
## Coefficients:
##          (Intercept)          COMUNA1          COMUNA2
##          0.467936          -0.380959          0.302680
##          COMUNA3          COMUNA4          COMUNA5
##          0.270976          -0.043790          -0.115997
##          COMUNA6          COMUNA7          COMUNA8
##          0.040295          0.165578          0.069749
##          COMUNA9          COMUNA10         COMUNA11
##          -0.155123          -0.006185          -0.393760
##          COMUNA12         COMUNA13         COMUNA14
##          0.016449          0.496827          -0.703766
##          COMUNA15         COMUNA16         COMUNA17
##          -0.139300          -0.174543          -0.134042
##          COMUNA18         COMUNA19         COMUNA20
##          -0.525142          -0.323943          0.286092
##          COMUNA21         COMUNA22         COMUNA23
##          0.306979          0.403345          0.163428
##          COMUNA24         COMUNA25          Riesgo1
##          0.192018          0.261539          -0.047832
##          Riesgo2 'Numero de accidentes' DIA_NOMBRE1
##          -0.063341          0.473226          0.141478

```

```
##          DIA_NOMBRE2          DIA_NOMBRE3          DIA_NOMBRE4
##          0.016781          -0.001003          -0.054600
##          DIA_NOMBRE5          DIA_NOMBRE6          MES
##          -0.007091          -0.016414          -0.006230
## festivo_entre_semana1      puente_festivo1 no_festivo_importante1
##          -0.123126          -0.026037          0.032126
```

Se ajusta un modelo de regresión lineal múltiple para el número de *accidentes con muertos*, usando como uno de los predictores el número de accidentes que resulta del primer modelo, también el día, el mes, el nombre del día, la comuna, el riesgo de accidentalidad del barrio, y un conjunto de variables dummy que determina si es o no un día importante.

```
#Modelo de regresión lineal multiple para el número de accidentes con muertos
NumAccidentesmuerτος_test<-test$'Accidente de muertos'
modelo4<-lm('Accidente de muertos'~MES+DIA+DIA_NOMBRE+COMUNA+puente_festivo+puente_reyes+semana_santa+f
a4<-predict(modelo4, test)
#MSE del conjunto de entrenamiento
msetrain4<-mean((train$'Accidente de muertos'- predict(modelo4))^2)
#MSE del conjunto de prueba
msetest4<- mean((NumAccidentesmuerτος_test - a4)^2)
#variacion de los MSE
variacion_mse4<-(msetrain4/msetest4)-1
msetest4
```

```
## [1] 0.03690023
```

```
msetrain4
```

```
## [1] 0.02219283
```

```
variacion_mse4
```

```
## [1] -0.398572
```

```
summary(modelo4)
```

```
##
## Call:
## lm(formula = 'Accidente de muertos' ~ MES + DIA + DIA_NOMBRE +
##      COMUNA + puente_festivo + puente_reyes + semana_santa + feria_flores +
##      puente_semana_santa + festivo_entre_semana + no_festivo_importante +
##      Riesgo + 'Numero de accidentes', data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -0.29619 -0.02730 -0.01528 -0.00614  2.96689
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    1.729e-02  1.003e-02   1.724  0.08470 .
## MES           -3.569e-04  2.114e-04  -1.689  0.09131 .
```

```

## DIA                1.193e-04  8.163e-05  1.462  0.14375
## DIA_NOMBRE1        9.665e-03  1.828e-03  5.288  1.24e-07 ***
## DIA_NOMBRE2       -5.351e-03  1.758e-03 -3.044  0.00234 **
## DIA_NOMBRE3        3.406e-03  1.754e-03  1.942  0.05211 .
## DIA_NOMBRE4       -1.113e-03  1.747e-03 -0.637  0.52399
## DIA_NOMBRE5       -4.503e-03  1.746e-03 -2.579  0.00992 **
## DIA_NOMBRE6        3.045e-03  1.758e-03  1.732  0.08319 .
## COMUNA1            8.188e-02  3.170e-02  2.583  0.00980 **
## COMUNA2          -1.606e-02  8.286e-02 -0.194  0.84633
## COMUNA3            6.325e-03  5.958e-03  1.062  0.28843
## COMUNA4            5.792e-02  3.864e-02  1.499  0.13389
## COMUNA5          -1.291e-02  5.761e-03 -2.241  0.02501 *
## COMUNA6          -7.391e-03  5.974e-03 -1.237  0.21597
## COMUNA7          -7.086e-03  5.795e-03 -1.223  0.22145
## COMUNA8            1.796e-03  9.399e-03  0.191  0.84846
## COMUNA9            1.040e-02  7.370e-03  1.411  0.15825
## COMUNA10         -1.460e-03  6.706e-03 -0.218  0.82760
## COMUNA11         -1.274e-02  4.803e-02 -0.265  0.79089
## COMUNA12         -4.461e-03  7.945e-03 -0.562  0.57444
## COMUNA13         -2.960e-03  6.392e-03 -0.463  0.64333
## COMUNA14         -1.765e-02  5.772e-03 -3.058  0.00223 **
## COMUNA15         -5.978e-03  5.903e-03 -1.013  0.31116
## COMUNA16          6.586e-03  1.501e-02  0.439  0.66084
## COMUNA17         -9.740e-03  6.008e-03 -1.621  0.10499
## COMUNA18         -3.383e-03  6.098e-03 -0.555  0.57902
## COMUNA19         -1.660e-02  5.921e-03 -2.804  0.00505 **
## COMUNA20         -1.667e-03  6.066e-03 -0.275  0.78343
## COMUNA21         -7.036e-03  6.753e-03 -1.042  0.29742
## COMUNA22         -1.492e-02  5.936e-03 -2.514  0.01194 *
## COMUNA23         -2.405e-03  6.304e-03 -0.382  0.70283
## COMUNA24         -6.328e-04  6.797e-03 -0.093  0.92583
## COMUNA25         -1.350e-02  7.180e-02 -0.188  0.85088
## puente_festivo1    -3.216e-03  1.350e-03 -2.382  0.01724 *
## puente_reyes1     -3.523e-03  4.446e-03 -0.792  0.42819
## semana_santa1     -2.763e-03  3.094e-03 -0.893  0.37190
## feria_flores1     -1.055e-02  2.161e-03 -4.884  1.04e-06 ***
## puente_semana_santa1  6.396e-03  5.002e-03  1.279  0.20103
## festivo_entre_semana1  1.910e-03  4.742e-03  0.403  0.68705
## no_festivo_importante1  1.084e-03  3.436e-03  0.315  0.75247
## Riesgo1            4.126e-03  1.540e-03  2.679  0.00738 **
## Riesgo2            6.856e-04  1.265e-03  0.542  0.58792
## 'Numero de accidentes' 5.964e-03  2.212e-04 26.969 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## s: 0.149 on 44242 degrees of freedom
## Multiple R-squared: 0.03088,
## Adjusted R-squared: 0.02994
## F-statistic: 32.78 on 43 and 44242 DF,  p-value: < 2.2e-16

```

Podemos ver que la variación de los MSE de prueba y entrenamiento es del 40%, también se ve que tenemos un $R^2 = 0.03$, esto indica que el modelo no es adecuado por lo cual se buscara mejorar la predicción con el uso de otros métodos

Se plantea un Modelo de regresion lineal multiple para el número de accidentes con heridos o muertos

```
## Creando la variable Accidentes_heridos_o_muertos  
require(dplyr)
```

```
## Loading required package: dplyr
```

```
##
```

```
## Attaching package: 'dplyr'
```

```
## The following object is masked from 'package:MASS':
```

```
##
```

```
## select
```

```
## The following object is masked from 'package:randomForest':
```

```
##
```

```
## combine
```

```
## The following object is masked from 'package:mixlm':
```

```
##
```

```
## tally
```

```
## The following objects are masked from 'package:stats':
```

```
##
```

```
## filter, lag
```

```
## The following objects are masked from 'package:base':
```

```
##
```

```
## intersect, setdiff, setequal, union
```

```
accidentalidad<-accidentalidad %>%
```

```
  select(puente_festivo,puente_reyes,semana_santa,feria_flores,puente_semana_santa,festivo_entre_semana
```

```
  mutate(accidentalidad,accidentes_heridos_o_muertos='Accidente de muertos'+`Accidentes con Heridos`)
```

```
## Seleccion del conjunto de entrenamiento y el conjunto de prueba
```

```
train<-accidentalidad[-which(accidentalidad$PERIODO=="2018"),]
```

```
test<-accidentalidad[which(accidentalidad$PERIODO=="2018"),]
```

```
accidentes_heridos_o_muertos_test<- test$accidentes_heridos_o_muertos
```

```
modelo5<-lm(accidentes_heridos_o_muertos~MES+DIA+DIA_NOMBRE+COMUNA+puente_festivo+puente_reyes+semana_s
```

```
#MSE del conjunto de entrenamiento
```

```
msetrain5<-mean((train$accidentes_heridos_o_muertos- predict(modelo5))^2)
```

```
#MSE del conjunto de prueba
```

```
a5<-predict(modelo5, test)
```

```
msetest5<- mean((accidentes_heridos_o_muertos_test - a5)^2)
```

```
#variacion de los MSE
```

```
variacion_mse5<-(msetest5/msetrain5)-1
```

```
msetest5
```

```
## [1] 1.014397
```

```
msetrain5
```

```
## [1] 1.036596
```

```
variacion_mse5
```

```
## [1] -0.02141495
```

```
summary(modelo5)
```

```
##
```

```
## Call:
```

```
## lm(formula = accidentes_heridos_o_muertos ~ MES + DIA + DIA_NOMBRE +  
##      COMUNA + puente_festivo + puente_reyes + semana_santa + feria_flores +  
##      puente_semana_santa + festivo_entre_semana + no_festivo_importante +  
##      Riesgo + 'Numero de accidentes', data = train)
```

```
##
```

```
## Residuals:
```

```
##      Min       1Q   Median       3Q      Max  
## -9.6401 -0.6017  0.0228  0.5539  7.2981
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept)    0.4291843  0.0685458   6.261 3.85e-10 ***  
## MES            -0.0068145  0.0014445  -4.718 2.39e-06 ***  
## DIA            -0.0000341  0.0005579  -0.061 0.951255  
## DIA_NOMBRE1     0.1510402  0.0124921  12.091 < 2e-16 ***  
## DIA_NOMBRE2     0.0121809  0.0120160   1.014 0.310722  
## DIA_NOMBRE3     0.0014694  0.0119842   0.123 0.902414  
## DIA_NOMBRE4    -0.0558886  0.0119366  -4.682 2.85e-06 ***  
## DIA_NOMBRE5    -0.0118315  0.0119346  -0.991 0.321516  
## DIA_NOMBRE6    -0.0135517  0.0120130  -1.128 0.259290  
## COMUNA1        -0.2995826  0.2166520  -1.383 0.166738  
## COMUNA2         0.2885437  0.5662627   0.510 0.610364  
## COMUNA3         0.2774553  0.0407198   6.814 9.63e-12 ***  
## COMUNA4         0.0127332  0.2641004   0.048 0.961546  
## COMUNA5        -0.1287637  0.0393712  -3.271 0.001074 **  
## COMUNA6         0.0331051  0.0408263   0.811 0.417441  
## COMUNA7         0.1586241  0.0396060   4.005 6.21e-05 ***  
## COMUNA8         0.0720027  0.0642390   1.121 0.262355  
## COMUNA9        -0.1449582  0.0503669  -2.878 0.004003 **  
## COMUNA10        -0.0077416  0.0458307  -0.169 0.865862  
## COMUNA11        -0.4103700  0.3282882  -1.250 0.211295  
## COMUNA12         0.0124876  0.0543005   0.230 0.818114  
## COMUNA13         0.4941556  0.0436854  11.312 < 2e-16 ***
```



```
## COMUNA14          -0.7212569  0.0394452 -18.285 < 2e-16 ***
## COMUNA15          -0.1453384  0.0403405  -3.603 0.000315 ***
## COMUNA16          -0.1684136  0.1025856  -1.642 0.100662
## COMUNA17          -0.1436803  0.0410604  -3.499 0.000467 ***
## COMUNA18          -0.5280426  0.0416784 -12.669 < 2e-16 ***
## COMUNA19          -0.3403843  0.0404641  -8.412 < 2e-16 ***
## COMUNA20           0.2845198  0.0414588   6.863 6.85e-12 ***
## COMUNA21           0.3001440  0.0461493   6.504 7.92e-11 ***
## COMUNA22           0.3886988  0.0405671   9.582 < 2e-16 ***
## COMUNA23           0.1611885  0.0430843   3.741 0.000183 ***
## COMUNA24           0.1914753  0.0464556   4.122 3.77e-05 ***
## COMUNA25           0.2489769  0.4907230   0.507 0.611899
## puente_festivo1    -0.0339462  0.0092288  -3.678 0.000235 ***
## puente_reyes1      0.0224406  0.0303867   0.738 0.460215
## semana_santa1     -0.0069777  0.0211452  -0.330 0.741408
## feria_flores1     -0.0014472  0.0147683  -0.098 0.921939
## puente_semana_santa1 0.0402875  0.0341871   1.178 0.238627
## festivo_entre_semana1 -0.1214204  0.0324081  -3.747 0.000179 ***
## no_festivo_importante1 0.0336913  0.0234840   1.435 0.151395
## Riesgo1           -0.0435822  0.0105259  -4.140 3.47e-05 ***
## Riesgo2           -0.0627567  0.0086473  -7.257 4.01e-13 ***
## 'Numero de accidentes' 0.4791488  0.0015114 317.020 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## s: 1.019 on 44242 degrees of freedom
## Multiple R-squared:  0.7834,
## Adjusted R-squared:  0.7832
## F-statistic: 3722 on 43 and 44242 DF,  p-value: < 2.2e-16
```

```
## Seleccion del conjunto de entrenamiento y el conjunto de prueba
```

```
train<-accidentalidad[-which(accidentalidad$PERIODO=="2018"),]
test<-accidentalidad[which(accidentalidad$PERIODO=="2018"),]
```

```
accidentes_heridos_o_muertos_test<- test$accidentes_heridos_o_muertos
modelo5<-lm(accidentes_heridos_o_muertos~MES+DIA+DIA_NOMBRE+COMUNA+puente_festivo+puente_reyes+semana_santa)
#MSE del conjunto de entrenamiento
msetrain5<-mean((train$accidentes_heridos_o_muertos- predict(modelo5))^2)
#MSE del conjunto de prueba
a5<-predict(modelo5, test)
msetest5<- mean((accidentes_heridos_o_muertos_test - a5)^2)
#variacion de los MSE
variacion_mse5<-(msetest5/msetrain5)-1
msetest5
```

```
## [1] 1.014397
```

```
msetrain5
```

```
## [1] 1.036596
```

```
variacion_mse5
```

```
## [1] -0.02141495
```

```
summary(modelo5)
```

```
##
## Call:
## lm(formula = accidentes_heridos_o_muertos ~ MES + DIA + DIA_NOMBRE +
##      COMUNA + puente_festivo + puente_reyes + semana_santa + feria_flores +
##      puente_semana_santa + festivo_entre_semana + no_festivo_importante +
##      Riesgo + 'Numero de accidentes', data = train)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -9.6401 -0.6017  0.0228  0.5539  7.2981
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)    0.4291843   0.0685458   6.261 3.85e-10 ***
## MES            -0.0068145   0.0014445  -4.718 2.39e-06 ***
## DIA            -0.0000341   0.0005579  -0.061 0.951255
## DIA_NOMBRE1     0.1510402   0.0124921  12.091 < 2e-16 ***
## DIA_NOMBRE2     0.0121809   0.0120160   1.014 0.310722
## DIA_NOMBRE3     0.0014694   0.0119842   0.123 0.902414
## DIA_NOMBRE4    -0.0558886   0.0119366  -4.682 2.85e-06 ***
## DIA_NOMBRE5    -0.0118315   0.0119346  -0.991 0.321516
## DIA_NOMBRE6    -0.0135517   0.0120130  -1.128 0.259290
## COMUNA1        -0.2995826   0.2166520  -1.383 0.166738
## COMUNA2         0.2885437   0.5662627   0.510 0.610364
## COMUNA3         0.2774553   0.0407198   6.814 9.63e-12 ***
## COMUNA4         0.0127332   0.2641004   0.048 0.961546
## COMUNA5        -0.1287637   0.0393712  -3.271 0.001074 **
## COMUNA6         0.0331051   0.0408263   0.811 0.417441
## COMUNA7         0.1586241   0.0396060   4.005 6.21e-05 ***
## COMUNA8         0.0720027   0.0642390   1.121 0.262355
## COMUNA9        -0.1449582   0.0503669  -2.878 0.004003 **
## COMUNA10        -0.0077416   0.0458307  -0.169 0.865862
## COMUNA11        -0.4103700   0.3282882  -1.250 0.211295
## COMUNA12         0.0124876   0.0543005   0.230 0.818114
## COMUNA13         0.4941556   0.0436854  11.312 < 2e-16 ***
## COMUNA14        -0.7212569   0.0394452 -18.285 < 2e-16 ***
## COMUNA15        -0.1453384   0.0403405  -3.603 0.000315 ***
## COMUNA16        -0.1684136   0.1025856  -1.642 0.100662
## COMUNA17        -0.1436803   0.0410604  -3.499 0.000467 ***
## COMUNA18        -0.5280426   0.0416784 -12.669 < 2e-16 ***
## COMUNA19        -0.3403843   0.0404641  -8.412 < 2e-16 ***
## COMUNA20         0.2845198   0.0414588   6.863 6.85e-12 ***
## COMUNA21         0.3001440   0.0461493   6.504 7.92e-11 ***
## COMUNA22         0.3886988   0.0405671   9.582 < 2e-16 ***
## COMUNA23         0.1611885   0.0430843   3.741 0.000183 ***
## COMUNA24         0.1914753   0.0464556   4.122 3.77e-05 ***
## COMUNA25         0.2489769   0.4907230   0.507 0.611899
```

```

## puente_festivo1      -0.0339462  0.0092288  -3.678  0.000235 ***
## puente_reyes1        0.0224406  0.0303867   0.738  0.460215
## semana_santa1       -0.0069777  0.0211452  -0.330  0.741408
## feria_flores1       -0.0014472  0.0147683  -0.098  0.921939
## puente_semana_santa1  0.0402875  0.0341871   1.178  0.238627
## festivo_entre_semana1 -0.1214204  0.0324081  -3.747  0.000179 ***
## no_festivo_importante1 0.0336913  0.0234840   1.435  0.151395
## Riesgo1              -0.0435822  0.0105259  -4.140  3.47e-05 ***
## Riesgo2              -0.0627567  0.0086473  -7.257  4.01e-13 ***
## 'Numero de accidentes' 0.4791488  0.0015114 317.020 < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## s: 1.019 on 44242 degrees of freedom
## Multiple R-squared:  0.7834,
## Adjusted R-squared:  0.7832
## F-statistic:  3722 on 43 and 44242 DF,  p-value: < 2.2e-16

```