additional_mlr

Gulzina Kuttubekova

12/6/2019

```r
games <- read.csv('games.csv')
sales <- read.csv('sales.csv')
```

```r
# since Sales has many zeros, log-transformation leads to -Inf. Hence we add some epsilon = 0.000001 to
sales$Sales = sales$Sales + 0.000001
```

```r
set.seed(13)
sets <- stratified(sales, c('Platform', 'Genre', 'Publisher',
                            'Developer', 'Rating', 'Decade',
                            'Platform_Company', 'Platform_Gen',
                            'Family_Platform', 'Main_Developer',
                            'Developer_Country', 'Main_Publisher',
                            'Region'), 0.7, bothSets = TRUE)

train_sales <- sets$SAMP1
test_sales <- sets$SAMP2
```

We'll try additional 3 models with fixed number of explanatory variables, and the ineraction term found significant in the hypothesis part. We fixed the number = 5, to reduce the number of coefficient parameters to be estimated and to sooth the 'curse of dimensionality'. We generated 4 sets with the combinations of explanatory variables. If we were to select the best model among all possible 5-element combinations of 10 explanatory variables we presented in the first MLR model, there would be 252 models in total. So we skipped that method and concentrated on keeping the variables that can be controlled by company developing/publishing the game. Hence, 3 main explanatory variables are fixed: {Region, Genre, Rating}. We excluded Decade since its's out of human/company control and Critic_Count. We made sure at least one model contains a Critic_Score. The remaining variables were chosen from the family of variables. We treated variables {Main_Developer, Main_Publisher} as one family, since they are usually interconnected and {Family_Platform, Platform_Company} as one family.

Hence, we have the following sets:

set1 for model3: {Region, Genre, Rating, Critic_Score, Platform_Gen} set2 for model4: {Region, Genre, Rating, Main_Publisher, Platform_Gen} set3 for model5: {Region, Genre, Rating, Main_Developer, Platform_Company} set4 for model6: {Region, Genre, Rating, } to see if the major variation in sales response variable is captured by this three only!!!!!

MLR 3 - on the dataset with zeros, interactions are kept

```r
mlr3 <- lm(log(Sales) ~ Region + Genre + Rating +
               Critic_Score + Platform_Gen +
               Region:Genre, data = train_sales)
mlr3 %>% summary()
```

```
##
## Call:
## lm(formula = log(Sales) ~ Region + Genre + Rating + Critic_Score +
##     Platform_Gen + Region:Genre, data = train_sales)
```

```
## 
## Residuals:
##      Min      1Q  Median      3Q     Max
## -13.5148  -1.5998   0.6845   2.2074  13.5205
## 
## Coefficients:
##                                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)                     -11.411873   2.013515  -5.668 1.46e-08
## RegionJP_Sales                   -6.920744   0.147951 -46.777  < 2e-16
## RegionNA_Sales                    0.996192   0.147951   6.733 1.70e-11
## RegionOther_Sales                -1.421835   0.147951  -9.610  < 2e-16
## GenreAdventure                   -1.147984   0.277779  -4.133 3.60e-05
## GenreFighting                    -0.771219   0.245196  -3.145  0.00166
## GenreMisc                        -0.976703   0.242333  -4.030 5.58e-05
## GenrePlatform                    -0.063512   0.238044  -0.267  0.78962
## GenrePuzzle                      -3.859993   0.400694  -9.633  < 2e-16
## GenreRacing                       0.320166   0.209053   1.532  0.12566
## GenreRole-Playing                -1.495986   0.190850  -7.839 4.75e-15
## GenreShooter                     -0.009377   0.178564  -0.053  0.95812
## GenreSimulation                   0.335177   0.271263   1.236  0.21661
## GenreSports                      -1.301800   0.187806  -6.932 4.27e-12
## GenreStrategy                    -0.077002   0.282069  -0.273  0.78486
## RatingE                           0.281857   2.003424   0.141  0.88812
## RatingE10+                       -0.252263   2.003768  -0.126  0.89982
## RatingK-A                         5.125818   2.833781   1.809  0.07049
## RatingM                           0.737241   2.002960   0.368  0.71282
## RatingRP                          0.928085   2.834826   0.327  0.74338
## RatingT                           0.118148   2.002747   0.059  0.95296
## Critic_Score                      0.072745   0.001901  38.274  < 2e-16
## Platform_Gen6th Gen               2.361566   0.091411  25.835  < 2e-16
## Platform_Gen7th Gen               3.238694   0.089299  36.268  < 2e-16
## Platform_Gen8th Gen               2.962798   0.111032  26.684  < 2e-16
## RegionJP_Sales:GenreAdventure     0.619435   0.392709   1.577  0.11473
## RegionNA_Sales:GenreAdventure    -0.076825   0.392709  -0.196  0.84490
## RegionOther_Sales:GenreAdventure -0.258725   0.392709  -0.659  0.51002
## RegionJP_Sales:GenreFighting      2.165859   0.344440   6.288 3.27e-10
## RegionNA_Sales:GenreFighting      0.940178   0.344440   2.730  0.00635
## RegionOther_Sales:GenreFighting   0.420253   0.344440   1.220  0.22244
## RegionJP_Sales:GenreMisc          0.040334   0.339892   0.119  0.90554
## RegionNA_Sales:GenreMisc          1.881984   0.339892   5.537 3.11e-08
## RegionOther_Sales:GenreMisc       1.397359   0.339892   4.111 3.95e-05
## RegionJP_Sales:GenrePlatform      0.052487   0.331463   0.158  0.87418
## RegionNA_Sales:GenrePlatform      0.684650   0.331463   2.066  0.03888
## RegionOther_Sales:GenrePlatform  -0.187067   0.331463  -0.564  0.57251
## RegionJP_Sales:GenrePuzzle        4.242294   0.561861   7.550 4.49e-14
## RegionNA_Sales:GenrePuzzle        3.126814   0.561861   5.565 2.65e-08
## RegionOther_Sales:GenrePuzzle     2.349035   0.561861   4.181 2.92e-05
## RegionJP_Sales:GenreRacing       -1.323832   0.290580  -4.556 5.24e-06
## RegionNA_Sales:GenreRacing       -0.353816   0.290580  -1.218  0.22338
## RegionOther_Sales:GenreRacing    -0.244347   0.290580  -0.841  0.40042
## RegionJP_Sales:GenreRole-Playing  4.568838   0.268925  16.989  < 2e-16
## RegionNA_Sales:GenreRole-Playing  0.770233   0.268925   2.864  0.00419
## RegionOther_Sales:GenreRole-Playing 0.842618  0.268925   3.133  0.00173
## RegionJP_Sales:GenreShooter      -0.570773   0.250472  -2.279  0.02269
```

```
## RegionNA_Sales:GenreShooter          -0.077856   0.250472  -0.311  0.75593
## RegionOther_Sales:GenreShooter       -0.283504   0.250472  -1.132  0.25770
## RegionJP_Sales:GenreSimulation        0.050205   0.380866   0.132  0.89513
## RegionNA_Sales:GenreSimulation       -1.260074   0.380866  -3.308  0.00094
## RegionOther_Sales:GenreSimulation    -0.415083   0.380866  -1.090  0.27579
## RegionJP_Sales:GenreSports           -0.693368   0.255160  -2.717  0.00658
## RegionNA_Sales:GenreSports            1.481018   0.255161   5.804 6.55e-09
## RegionOther_Sales:GenreSports         1.260712   0.255161   4.941 7.83e-07
## RegionJP_Sales:GenreStrategy         -0.943519   0.395530  -2.385  0.01707
## RegionNA_Sales:GenreStrategy         -2.240723   0.395531  -5.665 1.49e-08
## RegionOther_Sales:GenreStrategy      -1.300462   0.395531  -3.288  0.00101
##
## (Intercept)                          ***
## RegionJP_Sales                       ***
## RegionNA_Sales                       ***
## RegionOther_Sales                    ***
## GenreAdventure                       ***
## GenreFighting                        **
## GenreMisc                            ***
## GenrePlatform
## GenrePuzzle                          ***
## GenreRacing
## GenreRole-Playing                    ***
## GenreShooter
## GenreSimulation
## GenreSports                          ***
## GenreStrategy
## RatingE
## RatingE10+
## RatingK-A                            .
## RatingM
## RatingRP
## RatingT
## Critic_Score                         ***
## Platform_Gen6th Gen                  ***
## Platform_Gen7th Gen                  ***
## Platform_Gen8th Gen                  ***
## RegionJP_Sales:GenreAdventure
## RegionNA_Sales:GenreAdventure
## RegionOther_Sales:GenreAdventure
## RegionJP_Sales:GenreFighting         ***
## RegionNA_Sales:GenreFighting         **
## RegionOther_Sales:GenreFighting
## RegionJP_Sales:GenreMisc
## RegionNA_Sales:GenreMisc             ***
## RegionOther_Sales:GenreMisc          ***
## RegionJP_Sales:GenrePlatform
## RegionNA_Sales:GenrePlatform         *
## RegionOther_Sales:GenrePlatform
## RegionJP_Sales:GenrePuzzle           ***
## RegionNA_Sales:GenrePuzzle           ***
## RegionOther_Sales:GenrePuzzle        ***
## RegionJP_Sales:GenreRacing           ***
## RegionNA_Sales:GenreRacing
```

```
## RegionOther_Sales:GenreRacing
## RegionJP_Sales:GenreRole-Playing     ***
## RegionNA_Sales:GenreRole-Playing     **
## RegionOther_Sales:GenreRole-Playing  **
## RegionJP_Sales:GenreShooter          *
## RegionNA_Sales:GenreShooter
## RegionOther_Sales:GenreShooter
## RegionJP_Sales:GenreSimulation
## RegionNA_Sales:GenreSimulation       ***
## RegionOther_Sales:GenreSimulation
## RegionJP_Sales:GenreSports           **
## RegionNA_Sales:GenreSports           ***
## RegionOther_Sales:GenreSports        ***
## RegionJP_Sales:GenreStrategy         *
## RegionNA_Sales:GenreStrategy         ***
## RegionOther_Sales:GenreStrategy      **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.002 on 24026 degrees of freedom
## Multiple R-squared:  0.4171, Adjusted R-squared:  0.4157
## F-statistic: 301.6 on 57 and 24026 DF,  p-value: < 2.2e-16
```

```
anova(mlr3)
```

```
## Analysis of Variance Table
##
## Response: log(Sales)
##                 Df Sum Sq Mean Sq  F value    Pr(>F)
## Region           3 213709   71236 4448.923 < 2.2e-16 ***
## Genre           11   5619     511   31.902 < 2.2e-16 ***
## Rating           6   1786     298   18.588 < 2.2e-16 ***
## Critic_Score     1  17893   17893 1117.447 < 2.2e-16 ***
## Platform_Gen     3  21688    7229  451.497 < 2.2e-16 ***
## Region:Genre    33  14545     441   27.526 < 2.2e-16 ***
## Residuals    24026 384705      16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Still all 5 variables including the interactions are significant. Adj-R^2 is 0.4157, which is relatively good since we dropped half of the variables from the previous two models. We will refer to 2 those models as "full" and "reduced" to the models 3,4,5,6. Now we do some model assessment. Pics are not included here, they are approximately the same for all.

We observed the same trend on assumptions violations! We'll check all possible model assessments at the end.

MLR 4 - on the dataset with zeros, interactions are kept

```
mlr4 <- lm(log(Sales) ~ Region + Genre + Rating +
              Main_Publisher + Platform_Gen +
              Region:Genre + Region:Main_Publisher, data = train_sales)
mlr4 %>% summary()
```

```
##
```

```
## Call:
## lm(formula = log(Sales) ~ Region + Genre + Rating + Main_Publisher +
##     Platform_Gen + Region:Genre + Region:Main_Publisher, data = train_sales)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -13.8720 -1.4894  0.6537  2.1869 14.3245
##
## Coefficients:
##                                    Estimate Std. Error t value
## (Intercept)                        -3.978290   1.991041  -1.998
## RegionJP_Sales                     -9.066782   0.293172 -30.926
## RegionNA_Sales                      1.392576   0.293172   4.750
## RegionOther_Sales                  -1.150469   0.293172  -3.924
## GenreAdventure                     -1.125365   0.275819  -4.080
## GenreFighting                      -0.378711   0.246521  -1.536
## GenreMisc                          -0.949404   0.240692  -3.944
## GenrePlatform                       0.239772   0.236642   1.013
## GenrePuzzle                        -3.382076   0.398663  -8.484
## GenreRacing                         0.631171   0.207625   3.040
## GenreRole-Playing                  -1.223873   0.193627  -6.321
## GenreShooter                       -0.088007   0.176757  -0.498
## GenreSimulation                     0.374421   0.269865   1.387
## GenreSports                        -1.018004   0.192986  -5.275
## GenreStrategy                       0.238106   0.280714   0.848
## RatingE                            -1.584122   1.980538  -0.800
## RatingE10+                         -1.875049   1.980889  -0.947
## RatingK-A                           3.728663   2.803138   1.330
## RatingM                            -0.442697   1.979995  -0.224
## RatingRP                            0.365740   2.867536   0.128
## RatingT                            -1.324303   1.979876  -0.669
## Main_PublisherEA                    0.565703   0.244009   2.318
## Main_PublisherMicrosoft             1.307546   0.404975   3.229
## Main_PublisherNamco                -0.794004   0.283885  -2.797
## Main_PublisherNintendo              1.339245   0.318745   4.202
## Main_PublisherOther                -1.369821   0.208681  -6.564
## Main_PublisherSega                 -0.242578   0.261750  -0.927
## Main_PublisherSONY                  0.008897   0.299818   0.030
## Main_PublisherSquare Enix           0.724243   0.405886   1.784
## Main_PublisherTencent              -1.256549   1.233287  -1.019
## Main_PublisherUbisoft              -0.262300   0.254553  -1.030
## Platform_Gen6th Gen                 1.927270   0.090316  21.339
## Platform_Gen7th Gen                 2.629058   0.087766  29.955
## Platform_Gen8th Gen                 2.401379   0.110799  21.673
## RegionJP_Sales:GenreAdventure      -0.034243   0.389972  -0.088
## RegionNA_Sales:GenreAdventure      -0.064966   0.389972  -0.167
## RegionOther_Sales:GenreAdventure   -0.259382   0.389972  -0.665
## RegionJP_Sales:GenreFighting        1.632943   0.346553   4.712
## RegionNA_Sales:GenreFighting        0.923843   0.346553   2.666
## RegionOther_Sales:GenreFighting     0.556275   0.346553   1.605
## RegionJP_Sales:GenreMisc           -0.377621   0.337744  -1.118
## RegionNA_Sales:GenreMisc            1.930569   0.337744   5.716
## RegionOther_Sales:GenreMisc         1.403657   0.337744   4.156
## RegionJP_Sales:GenrePlatform       -0.422986   0.329798  -1.283
```

```
## RegionNA_Sales:GenrePlatform                       0.727877   0.329798    2.207
## RegionOther_Sales:GenrePlatform                    -0.249954   0.329798   -0.758
## RegionJP_Sales:GenrePuzzle                          2.718858   0.559876    4.856
## RegionNA_Sales:GenrePuzzle                          3.148916   0.559876    5.624
## RegionOther_Sales:GenrePuzzle                       2.393262   0.559876    4.275
## RegionJP_Sales:GenreRacing                         -1.414129   0.288881   -4.895
## RegionNA_Sales:GenreRacing                         -0.352749   0.288881   -1.221
## RegionOther_Sales:GenreRacing                      -0.295705   0.288881   -1.024
## RegionJP_Sales:GenreRole-Playing                    3.658082   0.273166   13.391
## RegionNA_Sales:GenreRole-Playing                    0.879337   0.273166    3.219
## RegionOther_Sales:GenreRole-Playing                 0.918298   0.273166    3.362
## RegionJP_Sales:GenreShooter                        -0.345416   0.248007   -1.393
## RegionNA_Sales:GenreShooter                        -0.040480   0.248007   -0.163
## RegionOther_Sales:GenreShooter                     -0.292776   0.248007   -1.181
## RegionJP_Sales:GenreSimulation                      0.069205   0.379093    0.183
## RegionNA_Sales:GenreSimulation                     -1.120508   0.379093   -2.956
## RegionOther_Sales:GenreSimulation                  -0.338159   0.379093   -0.892
## RegionJP_Sales:GenreSports                         -0.248450   0.264012   -0.941
## RegionNA_Sales:GenreSports                          1.568093   0.264012    5.939
## RegionOther_Sales:GenreSports                       1.062534   0.264012    4.025
## RegionJP_Sales:GenreStrategy                       -1.206164   0.393972   -3.062
## RegionNA_Sales:GenreStrategy                       -2.112518   0.393972   -5.362
## RegionOther_Sales:GenreStrategy                    -1.214470   0.393972   -3.083
## RegionJP_Sales:Main_PublisherEA                     0.593101   0.344565    1.721
## RegionNA_Sales:Main_PublisherEA                    -0.667219   0.344565   -1.936
## RegionOther_Sales:Main_PublisherEA                 -0.069210   0.344565   -0.201
## RegionJP_Sales:Main_PublisherMicrosoft             1.014900   0.572027    1.774
## RegionNA_Sales:Main_PublisherMicrosoft            -1.792808   0.572027   -3.134
## RegionOther_Sales:Main_PublisherMicrosoft         -1.846906   0.572027   -3.229
## RegionJP_Sales:Main_PublisherNamco                 3.574551   0.401134    8.911
## RegionNA_Sales:Main_PublisherNamco                -0.668865   0.401134   -1.667
## RegionOther_Sales:Main_PublisherNamco             -0.863438   0.401134   -2.152
## RegionJP_Sales:Main_PublisherNintendo              8.251061   0.447936   18.420
## RegionNA_Sales:Main_PublisherNintendo             -0.726427   0.447936   -1.622
## RegionOther_Sales:Main_PublisherNintendo          -0.381802   0.447936   -0.852
## RegionJP_Sales:Main_PublisherOther                 2.598460   0.294576    8.821
## RegionNA_Sales:Main_PublisherOther                -0.111433   0.294576   -0.378
## RegionOther_Sales:Main_PublisherOther             -0.303021   0.294576   -1.029
## RegionJP_Sales:Main_PublisherSega                  2.163689   0.369433    5.857
## RegionNA_Sales:Main_PublisherSega                 -0.603227   0.369433   -1.633
## RegionOther_Sales:Main_PublisherSega              -0.048022   0.369433   -0.130
## RegionJP_Sales:Main_PublisherSONY                  3.650524   0.423400    8.622
## RegionNA_Sales:Main_PublisherSONY                 -0.420851   0.423400   -0.994
## RegionOther_Sales:Main_PublisherSONY               1.091206   0.423400    2.577
## RegionJP_Sales:Main_PublisherSquare Enix           4.785374   0.573512    8.344
## RegionNA_Sales:Main_PublisherSquare Enix          -1.343233   0.573512   -2.342
## RegionOther_Sales:Main_PublisherSquare Enix       -0.382335   0.573512   -0.667
## RegionJP_Sales:Main_PublisherTencent               1.638395   1.722067    0.951
## RegionNA_Sales:Main_PublisherTencent              -4.021626   1.722067   -2.335
## RegionOther_Sales:Main_PublisherTencent           -1.477582   1.722067   -0.858
## RegionJP_Sales:Main_PublisherUbisoft               0.771721   0.359481    2.147
## RegionNA_Sales:Main_PublisherUbisoft              -0.942881   0.359481   -2.623
## RegionOther_Sales:Main_PublisherUbisoft           -0.564856   0.359481   -1.571
##                                                    Pr(>|t|)
```

```
## (Intercept)                              0.045718 *
## RegionJP_Sales                            < 2e-16 ***
## RegionNA_Sales                            2.05e-06 ***
## RegionOther_Sales                         8.73e-05 ***
## GenreAdventure                            4.52e-05 ***
## GenreFighting                             0.124498
## GenreMisc                                 8.02e-05 ***
## GenrePlatform                             0.310962
## GenrePuzzle                               < 2e-16 ***
## GenreRacing                               0.002369 **
## GenreRole-Playing                         2.65e-10 ***
## GenreShooter                              0.618562
## GenreSimulation                           0.165321
## GenreSports                               1.34e-07 ***
## GenreStrategy                             0.396326
## RatingE                                   0.423809
## RatingE10+                                0.343868
## RatingK-A                                 0.183473
## RatingM                                   0.823082
## RatingRP                                  0.898510
## RatingT                                   0.503577
## Main_PublisherEA                          0.020438 *
## Main_PublisherMicrosoft                   0.001245 **
## Main_PublisherNamco                       0.005163 **
## Main_PublisherNintendo                    2.66e-05 ***
## Main_PublisherOther                       5.34e-11 ***
## Main_PublisherSega                        0.354064
## Main_PublisherSONY                        0.976328
## Main_PublisherSquare Enix                 0.074379 .
## Main_PublisherTencent                     0.308279
## Main_PublisherUbisoft                     0.302817
## Platform_Gen6th Gen                       < 2e-16 ***
## Platform_Gen7th Gen                       < 2e-16 ***
## Platform_Gen8th Gen                       < 2e-16 ***
## RegionJP_Sales:GenreAdventure             0.930030
## RegionNA_Sales:GenreAdventure             0.867693
## RegionOther_Sales:GenreAdventure          0.505975
## RegionJP_Sales:GenreFighting              2.47e-06 ***
## RegionNA_Sales:GenreFighting              0.007686 **
## RegionOther_Sales:GenreFighting           0.108471
## RegionJP_Sales:GenreMisc                  0.263548
## RegionNA_Sales:GenreMisc                  1.10e-08 ***
## RegionOther_Sales:GenreMisc               3.25e-05 ***
## RegionJP_Sales:GenrePlatform              0.199659
## RegionNA_Sales:GenrePlatform              0.027321 *
## RegionOther_Sales:GenrePlatform           0.448519
## RegionJP_Sales:GenrePuzzle                1.20e-06 ***
## RegionNA_Sales:GenrePuzzle                1.88e-08 ***
## RegionOther_Sales:GenrePuzzle             1.92e-05 ***
## RegionJP_Sales:GenreRacing                9.88e-07 ***
## RegionNA_Sales:GenreRacing                0.222065
## RegionOther_Sales:GenreRacing             0.306023
## RegionJP_Sales:GenreRole-Playing          < 2e-16 ***
## RegionNA_Sales:GenreRole-Playing          0.001288 **
```

```
## RegionOther_Sales:GenreRole-Playing          0.000776 ***
## RegionJP_Sales:GenreShooter                  0.163704
## RegionNA_Sales:GenreShooter                  0.870347
## RegionOther_Sales:GenreShooter               0.237807
## RegionJP_Sales:GenreSimulation               0.855149
## RegionNA_Sales:GenreSimulation               0.003122 **
## RegionOther_Sales:GenreSimulation            0.372390
## RegionJP_Sales:GenreSports                   0.346686
## RegionNA_Sales:GenreSports                   2.90e-09 ***
## RegionOther_Sales:GenreSports                5.73e-05 ***
## RegionJP_Sales:GenreStrategy                 0.002204 **
## RegionNA_Sales:GenreStrategy                 8.30e-08 ***
## RegionOther_Sales:GenreStrategy              0.002054 **
## RegionJP_Sales:Main_PublisherEA              0.085209 .
## RegionNA_Sales:Main_PublisherEA              0.052830 .
## RegionOther_Sales:Main_PublisherEA           0.840807
## RegionJP_Sales:Main_PublisherMicrosoft       0.076040 .
## RegionNA_Sales:Main_PublisherMicrosoft       0.001726 **
## RegionOther_Sales:Main_PublisherMicrosoft    0.001245 **
## RegionJP_Sales:Main_PublisherNamco            < 2e-16 ***
## RegionNA_Sales:Main_PublisherNamco           0.095441 .
## RegionOther_Sales:Main_PublisherNamco        0.031369 *
## RegionJP_Sales:Main_PublisherNintendo         < 2e-16 ***
## RegionNA_Sales:Main_PublisherNintendo        0.104876
## RegionOther_Sales:Main_PublisherNintendo     0.394024
## RegionJP_Sales:Main_PublisherOther            < 2e-16 ***
## RegionNA_Sales:Main_PublisherOther           0.705225
## RegionOther_Sales:Main_PublisherOther        0.303647
## RegionJP_Sales:Main_PublisherSega            4.78e-09 ***
## RegionNA_Sales:Main_PublisherSega            0.102515
## RegionOther_Sales:Main_PublisherSega         0.896578
## RegionJP_Sales:Main_PublisherSONY             < 2e-16 ***
## RegionNA_Sales:Main_PublisherSONY            0.320242
## RegionOther_Sales:Main_PublisherSONY         0.009965 **
## RegionJP_Sales:Main_PublisherSquare Enix      < 2e-16 ***
## RegionNA_Sales:Main_PublisherSquare Enix     0.019183 *
## RegionOther_Sales:Main_PublisherSquare Enix 0.504998
## RegionJP_Sales:Main_PublisherTencent         0.341405
## RegionNA_Sales:Main_PublisherTencent         0.019533 *
## RegionOther_Sales:Main_PublisherTencent      0.390885
## RegionJP_Sales:Main_PublisherUbisoft         0.031822 *
## RegionNA_Sales:Main_PublisherUbisoft         0.008724 **
## RegionOther_Sales:Main_PublisherUbisoft      0.116124
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 3.953 on 23987 degrees of freedom
## Multiple R-squared:  0.432,  Adjusted R-squared:  0.4297
## F-statistic:   190 on 96 and 23987 DF,  p-value: < 2.2e-16
```

```r
anova(mlr4)
```

```
## Analysis of Variance Table
##
## Response: log(Sales)
```

```
##                         Df Sum Sq Mean Sq  F value    Pr(>F)
## Region                   3 213709   71236 4558.572 < 2.2e-16 ***
## Genre                   11   5619     511   32.688 < 2.2e-16 ***
## Rating                   6   1786     298   19.046 < 2.2e-16 ***
## Main_Publisher          10  21643    2164  138.495 < 2.2e-16 ***
## Platform_Gen             3  14361    4787  306.334 < 2.2e-16 ***
## Region:Genre            33  14521     440   28.158 < 2.2e-16 ***
## Region:Main_Publisher   30  13464     449   28.720 < 2.2e-16 ***
## Residuals            23987 374842      16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Still all 5 variables including the interactions are significant. Adj-R^2 is 0.4297, which is relatively good since we dropped half of the variables from the previous two models. Asj-R^2 is larger than for MLR3. Now we do some model assessment. The same plots.

We observed the same trend on assumptions violations! We'll check all possible model assessments at the end.

MLR 5

```
mlr5 <- lm(log(Sales) ~ Region + Genre + Rating +
             Main_Developer + Platform_Company +
             Region:Genre + Region:Main_Developer, data = train_sales)
mlr5 %>% summary()
```

```
##
## Call:
## lm(formula = log(Sales) ~ Region + Genre + Rating + Main_Developer +
##     Platform_Company + Region:Genre + Region:Main_Developer,
##     data = train_sales)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -13.2371  -1.6283   0.7269   2.2487  14.5427
##
## Coefficients:
##                                  Estimate Std. Error t value
## (Intercept)                     -0.698485   2.041483  -0.342
## RegionJP_Sales                  -9.559806   0.494572 -19.329
## RegionNA_Sales                   0.946733   0.494572   1.914
## RegionOther_Sales               -1.560112   0.494572  -3.154
## GenreAdventure                  -1.187446   0.279776  -4.244
## GenreFighting                   -0.437775   0.248800  -1.760
## GenreMisc                       -0.650100   0.244517  -2.659
## GenrePlatform                    0.145087   0.240695   0.603
## GenrePuzzle                     -3.115719   0.404688  -7.699
## GenreRacing                      0.544072   0.211506   2.572
## GenreRole-Playing               -1.474056   0.196765  -7.491
## GenreShooter                    -0.003031   0.180031  -0.017
## GenreSimulation                  0.258891   0.273637   0.946
## GenreSports                     -0.866438   0.202490  -4.279
## GenreStrategy                   -0.241303   0.282857  -0.853
## RatingE                         -2.989187   2.013927  -1.484
```

```
## RatingE10+                               -2.997091  2.013820  -1.488
## RatingK-A                                 0.729259  2.848074   0.256
## RatingM                                  -1.805132  2.012993  -0.897
## RatingRP                                 -3.389905  2.848164  -1.190
## RatingT                                  -2.791573  2.012956  -1.387
## Main_DeveloperEA                         -0.112133  0.386055  -0.290
## Main_DeveloperMicrosoft                   1.143546  0.519704   2.200
## Main_DeveloperNamco                      -1.131715  0.510168  -2.218
## Main_DeveloperNintendo                    0.701583  0.446887   1.570
## Main_DeveloperOther                      -1.622627  0.346564  -4.682
## Main_DeveloperSega                       -1.097455  0.459521  -2.388
## Main_DeveloperSONY                       -1.416919  0.431976  -3.280
## Main_DeveloperSquare Enix                -0.177250  0.583433  -0.304
## Main_DeveloperTencent                     0.025513  1.012479   0.025
## Main_DeveloperUbisoft                    -0.477983  0.407696  -1.172
## Platform_CompanyNintendo                  0.702300  0.076266   9.209
## Platform_CompanySega                     -3.550869  0.544674  -6.519
## Platform_CompanySONY                      1.692284  0.064044  26.424
## RegionJP_Sales:GenreAdventure             0.316777  0.395477   0.801
## RegionNA_Sales:GenreAdventure            -0.140121  0.395477  -0.354
## RegionOther_Sales:GenreAdventure         -0.320572  0.395477  -0.811
## RegionJP_Sales:GenreFighting              1.882704  0.349635   5.385
## RegionNA_Sales:GenreFighting              0.892520  0.349635   2.553
## RegionOther_Sales:GenreFighting           0.431717  0.349635   1.235
## RegionJP_Sales:GenreMisc                 -0.113705  0.343144  -0.331
## RegionNA_Sales:GenreMisc                  1.900433  0.343144   5.538
## RegionOther_Sales:GenreMisc               1.306393  0.343144   3.807
## RegionJP_Sales:GenrePlatform             -0.424369  0.335830  -1.264
## RegionNA_Sales:GenrePlatform              0.641277  0.335830   1.910
## RegionOther_Sales:GenrePlatform          -0.364550  0.335830  -1.086
## RegionJP_Sales:GenrePuzzle                3.268878  0.567878   5.756
## RegionNA_Sales:GenrePuzzle                3.043271  0.567878   5.359
## RegionOther_Sales:GenrePuzzle             2.205608  0.567878   3.884
## RegionJP_Sales:GenreRacing               -1.382806  0.293758  -4.707
## RegionNA_Sales:GenreRacing               -0.385789  0.293758  -1.313
## RegionOther_Sales:GenreRacing            -0.360305  0.293758  -1.227
## RegionJP_Sales:GenreRole-Playing          3.827080  0.277764  13.778
## RegionNA_Sales:GenreRole-Playing          0.699028  0.277764   2.517
## RegionOther_Sales:GenreRole-Playing       0.817431  0.277764   2.943
## RegionJP_Sales:GenreShooter              -0.562793  0.252672  -2.227
## RegionNA_Sales:GenreShooter              -0.053512  0.252672  -0.212
## RegionOther_Sales:GenreShooter           -0.306622  0.252672  -1.214
## RegionJP_Sales:GenreSimulation            0.023376  0.384865   0.061
## RegionNA_Sales:GenreSimulation           -1.236968  0.384865  -3.214
## RegionOther_Sales:GenreSimulation        -0.446277  0.384865  -1.160
## RegionJP_Sales:GenreSports               -0.493968  0.277704  -1.779
## RegionNA_Sales:GenreSports                1.465637  0.277704   5.278
## RegionOther_Sales:GenreSports             0.846822  0.277704   3.049
## RegionJP_Sales:GenreStrategy             -1.143051  0.398334  -2.870
## RegionNA_Sales:GenreStrategy             -2.285722  0.398334  -5.738
## RegionOther_Sales:GenreStrategy          -1.302385  0.398334  -3.270
## RegionJP_Sales:Main_DeveloperEA           1.153818  0.545640   2.115
## RegionNA_Sales:Main_DeveloperEA          -0.139972  0.545640  -0.257
## RegionOther_Sales:Main_DeveloperEA        0.657726  0.545640   1.205
```

```
## RegionJP_Sales:Main_DeveloperMicrosoft          1.869750    0.733757    2.548
## RegionNA_Sales:Main_DeveloperMicrosoft         -0.953175    0.733757   -1.299
## RegionOther_Sales:Main_DeveloperMicrosoft      -0.519933    0.733757   -0.709
## RegionJP_Sales:Main_DeveloperNamco              5.678721    0.720901    7.877
## RegionNA_Sales:Main_DeveloperNamco             -0.102157    0.720901   -0.142
## RegionOther_Sales:Main_DeveloperNamco           0.002679    0.720901    0.004
## RegionJP_Sales:Main_DeveloperNintendo           7.724546    0.629933   12.262
## RegionNA_Sales:Main_DeveloperNintendo           0.354391    0.629933    0.563
## RegionOther_Sales:Main_DeveloperNintendo        0.705866    0.629933    1.121
## RegionJP_Sales:Main_DeveloperOther              2.744419    0.489753    5.604
## RegionNA_Sales:Main_DeveloperOther              0.158245    0.489753    0.323
## RegionOther_Sales:Main_DeveloperOther           0.068341    0.489753    0.140
## RegionJP_Sales:Main_DeveloperSega               4.095186    0.648790    6.312
## RegionNA_Sales:Main_DeveloperSega               0.508639    0.648790    0.784
## RegionOther_Sales:Main_DeveloperSega            1.253502    0.648790    1.932
## RegionJP_Sales:Main_DeveloperSONY               4.572915    0.608455    7.516
## RegionNA_Sales:Main_DeveloperSONY               0.181613    0.608455    0.298
## RegionOther_Sales:Main_DeveloperSONY            1.848963    0.608455    3.039
## RegionJP_Sales:Main_DeveloperSquare Enix        6.323578    0.824711    7.668
## RegionNA_Sales:Main_DeveloperSquare Enix        0.346780    0.824711    0.420
## RegionOther_Sales:Main_DeveloperSquare Enix    -0.088689    0.824711   -0.108
## RegionJP_Sales:Main_DeveloperTencent            0.805294    1.430940    0.563
## RegionNA_Sales:Main_DeveloperTencent           -2.674838    1.430940   -1.869
## RegionOther_Sales:Main_DeveloperTencent        -0.836591    1.430940   -0.585
## RegionJP_Sales:Main_DeveloperUbisoft            1.083740    0.576252    1.881
## RegionNA_Sales:Main_DeveloperUbisoft           -0.667981    0.576252   -1.159
## RegionOther_Sales:Main_DeveloperUbisoft        -0.038915    0.576252   -0.068
##                                                Pr(>|t|)
## (Intercept)                                    0.732244
## RegionJP_Sales                                  < 2e-16 ***
## RegionNA_Sales                                 0.055600 .
## RegionOther_Sales                              0.001610 **
## GenreAdventure                                 2.20e-05 ***
## GenreFighting                                  0.078497 .
## GenreMisc                                      0.007849 **
## GenrePlatform                                  0.546657
## GenrePuzzle                                    1.42e-14 ***
## GenreRacing                                    0.010107 *
## GenreRole-Playing                              7.05e-14 ***
## GenreShooter                                   0.986568
## GenreSimulation                                0.344102
## GenreSports                                    1.89e-05 ***
## GenreStrategy                                  0.393617
## RatingE                                        0.137754
## RatingE10+                                     0.136695
## RatingK-A                                      0.797912
## RatingM                                        0.369867
## RatingRP                                       0.233977
## RatingT                                        0.165515
## Main_DeveloperEA                               0.771467
## Main_DeveloperMicrosoft                        0.027789 *
## Main_DeveloperNamco                            0.026542 *
## Main_DeveloperNintendo                         0.116444
## Main_DeveloperOther                            2.86e-06 ***
```

```
## Main_DeveloperSega                            0.016936 *
## Main_DeveloperSONY                            0.001039 **
## Main_DeveloperSquare Enix                     0.761278
## Main_DeveloperTencent                         0.979897
## Main_DeveloperUbisoft                         0.241048
## Platform_CompanyNintendo                       < 2e-16 ***
## Platform_CompanySega                          7.21e-11 ***
## Platform_CompanySONY                           < 2e-16 ***
## RegionJP_Sales:GenreAdventure                 0.423140
## RegionNA_Sales:GenreAdventure                 0.723111
## RegionOther_Sales:GenreAdventure              0.417606
## RegionJP_Sales:GenreFighting                  7.32e-08 ***
## RegionNA_Sales:GenreFighting                  0.010695 *
## RegionOther_Sales:GenreFighting               0.216930
## RegionJP_Sales:GenreMisc                      0.740374
## RegionNA_Sales:GenreMisc                      3.09e-08 ***
## RegionOther_Sales:GenreMisc                   0.000141 ***
## RegionJP_Sales:GenrePlatform                  0.206371
## RegionNA_Sales:GenrePlatform                  0.056206 .
## RegionOther_Sales:GenrePlatform               0.277702
## RegionJP_Sales:GenrePuzzle                    8.70e-09 ***
## RegionNA_Sales:GenrePuzzle                    8.44e-08 ***
## RegionOther_Sales:GenrePuzzle                 0.000103 ***
## RegionJP_Sales:GenreRacing                    2.52e-06 ***
## RegionNA_Sales:GenreRacing                    0.189098
## RegionOther_Sales:GenreRacing                 0.220009
## RegionJP_Sales:GenreRole-Playing               < 2e-16 ***
## RegionNA_Sales:GenreRole-Playing              0.011855 *
## RegionOther_Sales:GenreRole-Playing           0.003255 **
## RegionJP_Sales:GenreShooter                   0.025932 *
## RegionNA_Sales:GenreShooter                   0.832276
## RegionOther_Sales:GenreShooter                0.224944
## RegionJP_Sales:GenreSimulation                0.951568
## RegionNA_Sales:GenreSimulation                0.001311 **
## RegionOther_Sales:GenreSimulation             0.246236
## RegionJP_Sales:GenreSports                    0.075292 .
## RegionNA_Sales:GenreSports                    1.32e-07 ***
## RegionOther_Sales:GenreSports                 0.002296 **
## RegionJP_Sales:GenreStrategy                  0.004114 **
## RegionNA_Sales:GenreStrategy                  9.68e-09 ***
## RegionOther_Sales:GenreStrategy               0.001079 **
## RegionJP_Sales:Main_DeveloperEA               0.034473 *
## RegionNA_Sales:Main_DeveloperEA               0.797545
## RegionOther_Sales:Main_DeveloperEA            0.228052
## RegionJP_Sales:Main_DeveloperMicrosoft        0.010835 *
## RegionNA_Sales:Main_DeveloperMicrosoft        0.193945
## RegionOther_Sales:Main_DeveloperMicrosoft     0.478586
## RegionJP_Sales:Main_DeveloperNamco            3.49e-15 ***
## RegionNA_Sales:Main_DeveloperNamco            0.887312
## RegionOther_Sales:Main_DeveloperNamco         0.997035
## RegionJP_Sales:Main_DeveloperNintendo          < 2e-16 ***
## RegionNA_Sales:Main_DeveloperNintendo         0.573722
## RegionOther_Sales:Main_DeveloperNintendo      0.262495
## RegionJP_Sales:Main_DeveloperOther            2.12e-08 ***
```

```
## RegionNA_Sales:Main_DeveloperOther          0.746613
## RegionOther_Sales:Main_DeveloperOther       0.889024
## RegionJP_Sales:Main_DeveloperSega           2.80e-10 ***
## RegionNA_Sales:Main_DeveloperSega           0.433059
## RegionOther_Sales:Main_DeveloperSega        0.053364 .
## RegionJP_Sales:Main_DeveloperSONY           5.86e-14 ***
## RegionNA_Sales:Main_DeveloperSONY           0.765338
## RegionOther_Sales:Main_DeveloperSONY        0.002378 **
## RegionJP_Sales:Main_DeveloperSquare Enix    1.82e-14 ***
## RegionNA_Sales:Main_DeveloperSquare Enix    0.674133
## RegionOther_Sales:Main_DeveloperSquare Enix 0.914362
## RegionJP_Sales:Main_DeveloperTencent        0.573595
## RegionNA_Sales:Main_DeveloperTencent        0.061595 .
## RegionOther_Sales:Main_DeveloperTencent     0.558793
## RegionJP_Sales:Main_DeveloperUbisoft        0.060029 .
## RegionNA_Sales:Main_DeveloperUbisoft        0.246394
## RegionOther_Sales:Main_DeveloperUbisoft     0.946159
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.023 on 23987 degrees of freedom
## Multiple R-squared:  0.4118, Adjusted R-squared:  0.4094
## F-statistic: 174.9 on 96 and 23987 DF,  p-value: < 2.2e-16
```

```
anova(mlr5)
```

```
## Analysis of Variance Table
##
## Response: log(Sales)
##                        Df Sum Sq Mean Sq  F value    Pr(>F)
## Region                  3 213709   71236 4401.901 < 2.2e-16 ***
## Genre                  11   5619     511   31.565 < 2.2e-16 ***
## Rating                  6   1786     298   18.392 < 2.2e-16 ***
## Main_Developer         10  15285    1528   94.450 < 2.2e-16 ***
## Platform_Company        3  12698    4233  261.551 < 2.2e-16 ***
## Region:Genre           33  14521     440   27.190 < 2.2e-16 ***
## Region:Main_Developer  30   8143     271   16.774 < 2.2e-16 ***
## Residuals           23987 388183      16
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Still all 5 variables including the interactions are significant. Adj-R^2 is 0.4094, dropped from mlr3 and mlr4. Model violations are the same.

MLR 6

```
mlr6 <- lm(log(Sales) ~ Region + Genre + Rating + Region:Genre,
           data = train_sales)
mlr6 %>% summary()
```

```
##
## Call:
## lm(formula = log(Sales) ~ Region + Genre + Rating + Region:Genre,
```

13

```
##      data = train_sales)
##
## Residuals:
##      Min      1Q   Median      3Q      Max
## -12.2757  -1.9603   0.7555   2.3260  13.3127
##
## Coefficients:
##                                        Estimate Std. Error t value Pr(>|t|)
## (Intercept)                            -2.29530    2.10319  -1.091 0.275134
## RegionJP_Sales                         -6.90841    0.15537 -44.465  < 2e-16
## RegionNA_Sales                          1.01096    0.15537   6.507 7.82e-11
## RegionOther_Sales                      -1.40796    0.15537  -9.062  < 2e-16
## GenreAdventure                         -1.32692    0.29167  -4.549 5.41e-06
## GenreFighting                          -0.47501    0.25736  -1.846 0.064949
## GenreMisc                              -0.68693    0.25435  -2.701 0.006923
## GenrePlatform                           0.09340    0.24967   0.374 0.708344
## GenrePuzzle                            -3.31051    0.42047  -7.873 3.60e-15
## GenreRacing                             0.36711    0.21923   1.675 0.094039
## GenreRole-Playing                      -1.25014    0.20005  -6.249 4.20e-10
## GenreShooter                           -0.01567    0.18733  -0.084 0.933341
## GenreSimulation                         0.03804    0.28422   0.134 0.893542
## GenreSports                            -0.68744    0.19646  -3.499 0.000468
## GenreStrategy                          -0.62605    0.29446  -2.126 0.033509
## RatingE                                -1.43935    2.10281  -0.684 0.493673
## RatingE10+                             -1.67434    2.10283  -0.796 0.425906
## RatingK-A                               2.40745    2.97417   0.809 0.418262
## RatingM                                -0.84230    2.10236  -0.401 0.688687
## RatingRP                               -3.04408    2.97502  -1.023 0.306218
## RatingT                                -1.55824    2.10215  -0.741 0.458540
## RegionJP_Sales:GenreAdventure           0.60410    0.41239   1.465 0.142973
## RegionNA_Sales:GenreAdventure          -0.09249    0.41239  -0.224 0.822536
## RegionOther_Sales:GenreAdventure       -0.27981    0.41239  -0.679 0.497457
## RegionJP_Sales:GenreFighting            2.14715    0.36171   5.936 2.96e-09
## RegionNA_Sales:GenreFighting            0.91640    0.36171   2.534 0.011298
## RegionOther_Sales:GenreFighting         0.40330    0.36171   1.115 0.264859
## RegionJP_Sales:GenreMisc                0.03098    0.35693   0.087 0.930833
## RegionNA_Sales:GenreMisc                1.87083    0.35693   5.241 1.61e-07
## RegionOther_Sales:GenreMisc             1.39497    0.35693   3.908 9.32e-05
## RegionJP_Sales:GenrePlatform            0.04515    0.34808   0.130 0.896790
## RegionNA_Sales:GenrePlatform            0.67668    0.34808   1.944 0.051902
## RegionOther_Sales:GenrePlatform        -0.19954    0.34808  -0.573 0.566471
## RegionJP_Sales:GenrePuzzle              4.23664    0.59002   7.180 7.15e-13
## RegionNA_Sales:GenrePuzzle              3.12473    0.59002   5.296 1.19e-07
## RegionOther_Sales:GenrePuzzle           2.32849    0.59002   3.946 7.96e-05
## RegionJP_Sales:GenreRacing             -1.34426    0.30514  -4.405 1.06e-05
## RegionNA_Sales:GenreRacing             -0.38180    0.30514  -1.251 0.210874
## RegionOther_Sales:GenreRacing          -0.27598    0.30514  -0.904 0.365780
## RegionJP_Sales:GenreRole-Playing        4.56304    0.28241  16.158  < 2e-16
## RegionNA_Sales:GenreRole-Playing        0.76039    0.28241   2.693 0.007096
## RegionOther_Sales:GenreRole-Playing     0.83688    0.28241   2.963 0.003046
## RegionJP_Sales:GenreShooter            -0.57809    0.26303  -2.198 0.027969
## RegionNA_Sales:GenreShooter            -0.08984    0.26303  -0.342 0.732684
## RegionOther_Sales:GenreShooter         -0.29190    0.26303  -1.110 0.267104
## RegionJP_Sales:GenreSimulation          0.04291    0.39996   0.107 0.914562
```

```
## RegionNA_Sales:GenreSimulation     -1.27372   0.39996  -3.185 0.001451
## RegionOther_Sales:GenreSimulation  -0.42336   0.39996  -1.059 0.289832
## RegionJP_Sales:GenreSports         -0.70128   0.26795  -2.617 0.008871
## RegionNA_Sales:GenreSports          1.45427   0.26795   5.427 5.77e-08
## RegionOther_Sales:GenreSports       1.22789   0.26795   4.583 4.62e-06
## RegionJP_Sales:GenreStrategy       -0.96319   0.41536  -2.319 0.020407
## RegionNA_Sales:GenreStrategy       -2.27322   0.41536  -5.473 4.47e-08
## RegionOther_Sales:GenreStrategy    -1.33909   0.41536  -3.224 0.001266
##
## (Intercept)
## RegionJP_Sales                       ***
## RegionNA_Sales                       ***
## RegionOther_Sales                    ***
## GenreAdventure                       ***
## GenreFighting                        .
## GenreMisc                            **
## GenrePlatform
## GenrePuzzle                          ***
## GenreRacing                          .
## GenreRole-Playing                    ***
## GenreShooter
## GenreSimulation
## GenreSports                          ***
## GenreStrategy                        *
## RatingE
## RatingE10+
## RatingK-A
## RatingM
## RatingRP
## RatingT
## RegionJP_Sales:GenreAdventure
## RegionNA_Sales:GenreAdventure
## RegionOther_Sales:GenreAdventure
## RegionJP_Sales:GenreFighting         ***
## RegionNA_Sales:GenreFighting         *
## RegionOther_Sales:GenreFighting
## RegionJP_Sales:GenreMisc
## RegionNA_Sales:GenreMisc             ***
## RegionOther_Sales:GenreMisc          ***
## RegionJP_Sales:GenrePlatform
## RegionNA_Sales:GenrePlatform         .
## RegionOther_Sales:GenrePlatform
## RegionJP_Sales:GenrePuzzle           ***
## RegionNA_Sales:GenrePuzzle           ***
## RegionOther_Sales:GenrePuzzle        ***
## RegionJP_Sales:GenreRacing           ***
## RegionNA_Sales:GenreRacing
## RegionOther_Sales:GenreRacing
## RegionJP_Sales:GenreRole-Playing     ***
## RegionNA_Sales:GenreRole-Playing     **
## RegionOther_Sales:GenreRole-Playing  **
## RegionJP_Sales:GenreShooter          *
## RegionNA_Sales:GenreShooter
## RegionOther_Sales:GenreShooter
```

```
## RegionJP_Sales:GenreSimulation
## RegionNA_Sales:GenreSimulation       **
## RegionOther_Sales:GenreSimulation
## RegionJP_Sales:GenreSports           **
## RegionNA_Sales:GenreSports           ***
## RegionOther_Sales:GenreSports        ***
## RegionJP_Sales:GenreStrategy         *
## RegionNA_Sales:GenreStrategy         ***
## RegionOther_Sales:GenreStrategy      **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.202 on 24030 degrees of freedom
## Multiple R-squared:  0.3571, Adjusted R-squared:  0.3556
## F-statistic: 251.8 on 53 and 24030 DF,  p-value: < 2.2e-16
```

```
anova(mlr6)
```

```
## Analysis of Variance Table
##
## Response: log(Sales)
##                 Df Sum Sq Mean Sq  F value    Pr(>F)
## Region           3 213709   71236 4034.335 < 2.2e-16 ***
## Genre           11   5619     511   28.929 < 2.2e-16 ***
## Rating           6   1786     298   16.856 < 2.2e-16 ***
## Region:Genre    33  14521     440   24.919 < 2.2e-16 ***
## Residuals    24030 424310      18
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Still all 3 variables including the interactions are significant. Adj-R^2 is 0.3556, decreased from all previous cases. Model violations are the same.

We observed the same trend on assumptions violations! We'll check all possible model assessments at the end.

Model Comparison

Predictions and RMSE:

```
# rmse on train data
sqrt(sum((mlr3$fitted.values - train_sales$Sales)^2) / length(mlr3$fitted.values))
```

```
## [1] 6.700798
```

```
predicted_sales <- predict(mlr3, test_sales[,c(3,5,10,14,19)])
```

```
# rmse on test data
sqrt(sum((predicted_sales - test_sales$Sales)^2) / length(predicted_sales))
```

```
## [1] 6.53495
```

```
# rmse on train data
sqrt(sum((mlr4$fitted.values - train_sales$Sales)^2) / length(mlr4$fitted.values))
```

```
## [1] 6.731926
```

```r
predicted_sales <- predict(mlr4, test_sales[,c(3,10,14,18,19)])

# rmse on test data
sqrt(sum((predicted_sales - test_sales$Sales)^2) / length(predicted_sales))
```

[1] 6.807955

```r
# rmse on train data
sqrt(sum((mlr5$fitted.values - train_sales$Sales)^2) / length(mlr5$fitted.values))
```

[1] 6.694827

```r
predicted_sales <- predict(mlr5, test_sales[,c(3, 10, 16, 13, 19)])

# rmse on test data
sqrt(sum((predicted_sales - test_sales$Sales)^2) / length(predicted_sales))
```

[1] 6.713769

```r
# rmse on train data
sqrt(sum((mlr6$fitted.values - train_sales$Sales)^2) / length(mlr6$fitted.values))
```

[1] 6.598686

```r
predicted_sales <- predict(mlr6, test_sales[,c(3, 10, 19)])

# rmse on test data
sqrt(sum((predicted_sales - test_sales$Sales)^2) / length(predicted_sales))
```

[1] 6.718925

Create dataframe for model assessment:

```r
df <- data.frame(AIC = c(AIC(mlr3), AIC(mlr4), AIC(mlr5), AIC(mlr6)),
                 BIC = c(BIC(mlr3), BIC(mlr4), BIC(mlr5), BIC(mlr6)),
                 AdjRsq = c(0.4157, 0.4297, 0.4094, 0.3556),
                 trainRMSE = c(6.700798, 6.731926, 6.694827, 6.598686),
                 testRMSE = c(6.53495, 6.807955, 6.713769, 6.718925))
row.names(df) <- c('MLR3', 'MLR4', 'MLR5', 'MLR6')
df
```

```
##              AIC      BIC AdjRsq trainRMSE testRMSE
## MLR3 135200.5 135677.7 0.4157  6.700798 6.534950
## MLR4 134653.0 135445.7 0.4297  6.731926 6.807955
## MLR5 135495.3 136288.0 0.4094  6.694827 6.713769
## MLR6 137552.4 137997.3 0.3556  6.598686 6.718925
```