Witcher Games sales Model

Arinze Francis

2022-06-09

Witcher Video Games sales Model

```
## function (..., list = character(), pos = -1, envir = as.environment(pos),
##     inherits = FALSE)
## {
##     dots <- match.call(expand.dots = FALSE)$...
##     if (length(dots) && !all(vapply(dots, function(x) is.symbol(x) ||
##         is.character(x), NA, USE.NAMES = FALSE)))
##         stop("... must contain names or character strings")
##     names <- vapply(dots, as.character, "")
##     if (length(names) == 0L)
##         names <- character()
##     list <- .Primitive("c")(list, names)
##     .Internal(remove(list, envir, inherits))
## }
## <bytecode: 0x0000000014d351b8>
## <environment: namespace:base>
```

Load R packages

```
## Warning: package 'ggplot2' was built under R version 4.1.2
```

```
## Warning: package 'equatiomatic' was built under R version 4.1.3
```

Import data

```
##                       Title Year      Sales.Category  Distribution.Channel
## 1  The Witcher 3: Wild Hunt 2015 Distribution method Physical distribution
## 2  The Witcher 3: Wild Hunt 2015 Distribution method  Digital distribution
## 3  The Witcher 3: Wild Hunt 2016 Distribution method Physical distribution
## 4  The Witcher 3: Wild Hunt 2016 Distribution method  Digital distribution
## 5  The Witcher 3: Wild Hunt 2017 Distribution method Physical distribution
## 6  The Witcher 3: Wild Hunt 2017 Distribution method  Digital distribution
## 7  The Witcher 3: Wild Hunt 2018 Distribution method Physical distribution
## 8  The Witcher 3: Wild Hunt 2018 Distribution method  Digital distribution
## 9  The Witcher 3: Wild Hunt 2019 Distribution method Physical distribution
## 10 The Witcher 3: Wild Hunt 2019 Distribution method  Digital distribution
## 11 The Witcher 3: Wild Hunt 2020 Distribution method Physical distribution
## 12 The Witcher 3: Wild Hunt 2020 Distribution method  Digital distribution
## 13 The Witcher 3: Wild Hunt 2015   Hardware platform                   PC
## 14 The Witcher 3: Wild Hunt 2015   Hardware platform           Playstation
## 15 The Witcher 3: Wild Hunt 2015   Hardware platform                 Xbox
## 16 The Witcher 3: Wild Hunt 2016   Hardware platform                   PC
## 17 The Witcher 3: Wild Hunt 2016   Hardware platform           Playstation
## 18 The Witcher 3: Wild Hunt 2016   Hardware platform                 Xbox
## 19 The Witcher 3: Wild Hunt 2017   Hardware platform                   PC
## 20 The Witcher 3: Wild Hunt 2017   Hardware platform           Playstation
## 21 The Witcher 3: Wild Hunt 2017   Hardware platform                 Xbox
## 22 The Witcher 3: Wild Hunt 2018   Hardware platform                   PC
## 23 The Witcher 3: Wild Hunt 2018   Hardware platform           Playstation
## 24 The Witcher 3: Wild Hunt 2018   Hardware platform                 Xbox
## 25 The Witcher 3: Wild Hunt 2019   Hardware platform                   PC
## 26 The Witcher 3: Wild Hunt 2019   Hardware platform           Playstation
## 27 The Witcher 3: Wild Hunt 2019   Hardware platform                 Xbox
## 28 The Witcher 3: Wild Hunt 2019   Hardware platform               Switch
## 29 The Witcher 3: Wild Hunt 2020   Hardware platform                   PC
## 30 The Witcher 3: Wild Hunt 2020   Hardware platform           Playstation
## 31 The Witcher 3: Wild Hunt 2020   Hardware platform                 Xbox
## 32 The Witcher 3: Wild Hunt 2020   Hardware platform               Switch
## 33 The Witcher 3: Wild Hunt 2019               Region         North America
## 34 The Witcher 3: Wild Hunt 2019               Region         South America
## 35 The Witcher 3: Wild Hunt 2019               Region                Europe
## 36 The Witcher 3: Wild Hunt 2019               Region                Africa
## 37 The Witcher 3: Wild Hunt 2019               Region                  Asia
## 38 The Witcher 3: Wild Hunt 2019               Region             Australia
## 39 The Witcher 3: Wild Hunt 2020               Region         North America
## 40 The Witcher 3: Wild Hunt 2020               Region         South America
## 41 The Witcher 3: Wild Hunt 2020               Region                Europe
## 42 The Witcher 3: Wild Hunt 2020               Region                Africa
## 43 The Witcher 3: Wild Hunt 2020               Region                  Asia
## 44 The Witcher 3: Wild Hunt 2020               Region             Australia
##    Percentage Units.sold..estimates.
## 1         68%                6120000
## 2         32%                2880000
## 3         38%                1520000
## 4         62%                2480000
## 5         24%                1080000
## 6         76%                3420000
```

```
## 7         19%                665000
## 8         81%               2835000
## 9         17%               1190000
## 10        83%               5810000
## 11        16%               1280000
## 12        84%               6720000
## 13        32%               2880000
## 14        46%               4140000
## 15        21%               1890000
## 16        42%               1680000
## 17        43%               1720000
## 18        15%                600000
## 19        50%               2250000
## 20        35%               1575000
## 21        15%                675000
## 22        54%               1890000
## 23        33%               1155000
## 24        13%                455000
## 25        53%               3710000
## 26        26%               1820000
## 27        10%                700000
## 28        11%                770000
## 29        52%               4160000
## 30        31%               2480000
## 31         7%                560000
## 32         9%                720000
## 33      25.6%               1792000
## 34       7.0%                490000
## 35      37.2%               2604000
## 36       0.3%                 21000
## 37      27.9%               1953000
## 38       2.0%                140000
## 39      26.4%               1848000
## 40       5.8%                406000
## 41      40.3%               2821000
## 42      0.30%                 21000
## 43      24.7%               1729000
## 44       2.4%                168000
```

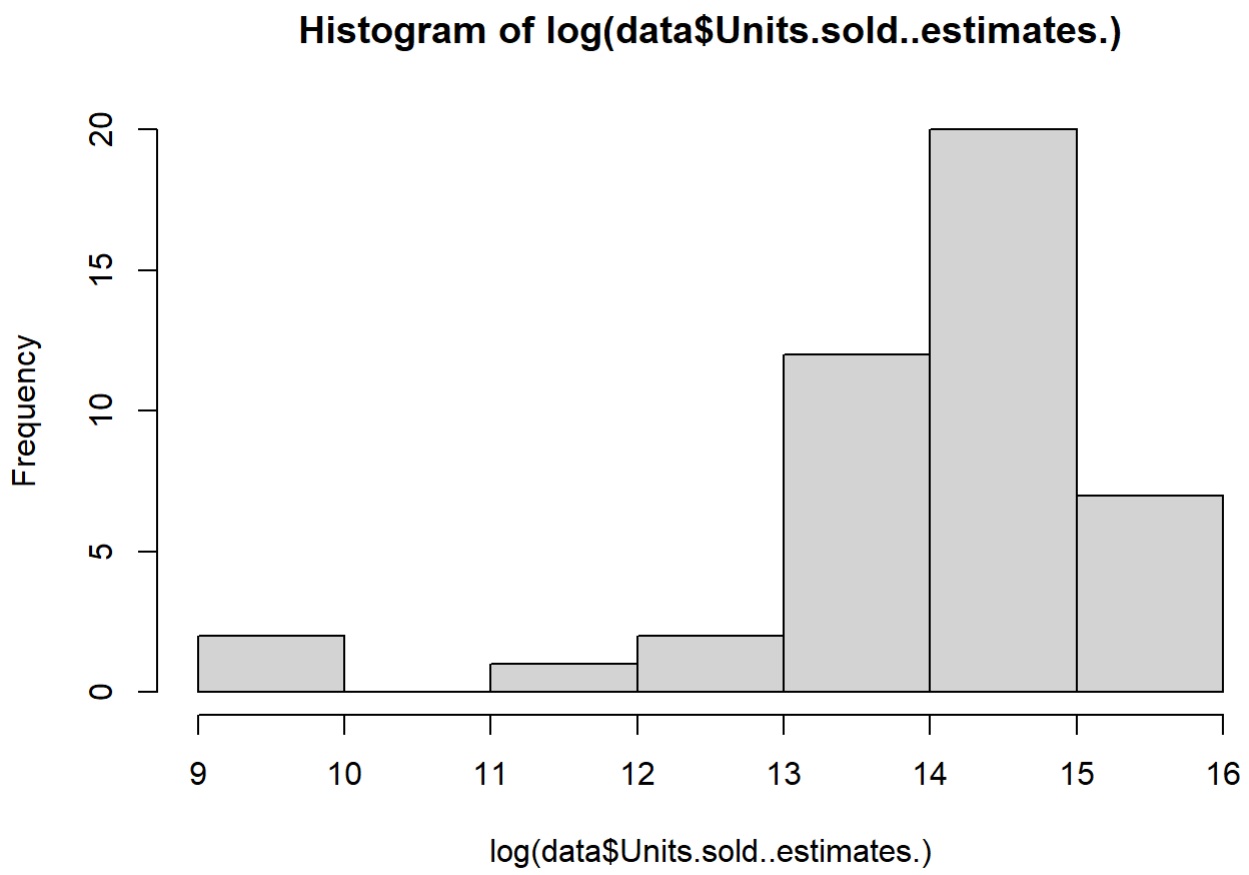
Estimate linear additive model

```
data_result <- lm(Units.sold..estimates. ~ Distribution.Channel + as.factor(Year), data = data )

summary(data_result)
```

```
##
## Call:
## lm(formula = Units.sold..estimates. ~ Distribution.Channel +
##     as.factor(Year), data = data)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2381500  -269031   -38729   215990  2906833
##
## Coefficients:
##                                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)                             760000     945218    0.80   0.4284
## Distribution.ChannelAsia                1820000    1056785    1.72   0.0965
## Distribution.ChannelAustralia            133000    1056785    0.13   0.9008
## Distribution.ChannelDigital distribution  4501500     904977    4.97 0.000033
## Distribution.ChannelEurope              2691500    1056785    2.55   0.0169
## Distribution.ChannelNorth America       1799000    1056785    1.70   0.1002
## Distribution.ChannelPC                  3239000     904977    3.58   0.0013
## Distribution.ChannelPhysical distribution  2453167     904977    2.71   0.0115
## Distribution.ChannelPlaystation         2625667     904977    2.90   0.0073
## Distribution.ChannelSouth America        427000    1056785    0.40   0.6894
## Distribution.ChannelSwitch               724000    1056785    0.69   0.4991
## Distribution.ChannelXbox                1290667     904977    1.43   0.1653
## as.factor(Year)2016                    -1982000     668370   -2.97   0.0063
## as.factor(Year)2017                    -1782000     668370   -2.67   0.0128
## as.factor(Year)2018                    -2182000     668370   -3.26   0.0030
## as.factor(Year)2019                     -818708     617715   -1.33   0.1962
## as.factor(Year)2020                     -659292     617715   -1.07   0.2953
##
## (Intercept)
## Distribution.ChannelAsia                .
## Distribution.ChannelAustralia
## Distribution.ChannelDigital distribution  ***
## Distribution.ChannelEurope              *
## Distribution.ChannelNorth America
## Distribution.ChannelPC                  **
## Distribution.ChannelPhysical distribution *
## Distribution.ChannelPlaystation         **
## Distribution.ChannelSouth America
## Distribution.ChannelSwitch
## Distribution.ChannelXbox
## as.factor(Year)2016                     **
## as.factor(Year)2017                     *
## as.factor(Year)2018                     **
## as.factor(Year)2019
## as.factor(Year)2020
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 1060000 on 27 degrees of freedom
## Multiple R-squared:  0.724,  Adjusted R-squared:  0.561
## F-statistic: 4.44 on 16 and 27 DF,  p-value: 0.000335
```

```
hist(log(data$Units.sold..estimates.))
```



Specify a multiplicative model

```
data_result_11 <- lm(log(Units.sold..estimates.) ~ Distribution.Channel + as.factor(Year), data
= data )

summary(data_result_11)
```

```
##
## Call:
## lm(formula = log(Units.sold..estimates.) ~ Distribution.Channel +
##     as.factor(Year), data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -0.8541 -0.1169  0.0155  0.1058  0.8300
##
## Coefficients:
##                                        Estimate Std. Error t value
## (Intercept)                              10.410      0.304   34.26
## Distribution.ChannelAsia                  4.472      0.340   13.16
## Distribution.ChannelAustralia             1.988      0.340    5.85
## Distribution.ChannelDigital distribution  5.318      0.291   18.28
## Distribution.ChannelEurope                4.860      0.340   14.31
## Distribution.ChannelNorth America         4.462      0.340   13.13
## Distribution.ChannelPC                    4.961      0.291   17.05
## Distribution.ChannelPhysical distribution 4.387      0.291   15.08
## Distribution.ChannelPlaystation           4.679      0.291   16.08
## Distribution.ChannelSouth America         3.056      0.340    8.99
## Distribution.ChannelSwitch                3.568      0.340   10.50
## Distribution.ChannelXbox                  3.669      0.291   12.61
## as.factor(Year)2016                      -0.821      0.215   -3.82
## as.factor(Year)2017                      -0.761      0.215   -3.54
## as.factor(Year)2018                      -1.071      0.215   -4.99
## as.factor(Year)2019                      -0.472      0.199   -2.37
## as.factor(Year)2020                      -0.444      0.199   -2.23
##                                                    Pr(>|t|)
## (Intercept)                             < 0.0000000000000002 ***
## Distribution.ChannelAsia                 0.00000000000029079 ***
## Distribution.ChannelAustralia            0.00000312332316640 ***
## Distribution.ChannelDigital distribution < 0.0000000000000002 ***
## Distribution.ChannelEurope               0.00000000000004026 ***
## Distribution.ChannelNorth America        0.00000000000030606 ***
## Distribution.ChannelPC                   0.00000000000000056 ***
## Distribution.ChannelPhysical distribution 0.00000000000001132 ***
## Distribution.ChannelPlaystation          0.00000000000000236 ***
## Distribution.ChannelSouth America        0.00000000130933692 ***
## Distribution.ChannelSwitch               0.00000000004900283 ***
## Distribution.ChannelXbox                 0.00000000000079138 ***
## as.factor(Year)2016                                  0.00071 ***
## as.factor(Year)2017                                  0.00146 **
## as.factor(Year)2018                      0.00003158073652826 ***
## as.factor(Year)2019                                  0.02492 *
## as.factor(Year)2020                                  0.03395 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.34 on 27 degrees of freedom
## Multiple R-squared:  0.953,  Adjusted R-squared:  0.926
## F-statistic: 34.5 on 16 and 27 DF,  p-value: 0.0000000000000731
```

Conclusions and Managerial Implications

The overall model is highly significant. 95.3% of the fluctuations of the units sold can be explained by the model. All Distribution Channel Methods had a significant effect on Unit sold (p < .05.) including the various years too. Witcher games sales benefited greatly from The Digital Distribution channel (5.31), followed by the PC channel (4.96) and the Europe Channel (4.86). Australia distribution Channel had the lowest contribution of 1.99. If Distribution Channel-Digital Distribution is increased by one percent, the units sold is to increase by 5.10 percent (holding all other variables constant) and etc.

Additional Insights

```
tidy_result<-tidy(data_result_11)
tidy_result
```

```
## # A tibble: 17 x 5
##    term                                  estimate std.error statistic  p.value
##    <chr>                                    <dbl>     <dbl>     <dbl>    <dbl>
##  1 (Intercept)                             10.4       0.304     34.3  8.67e-24
##  2 Distribution.ChannelAsia                 4.47      0.340     13.2  2.91e-13
##  3 Distribution.ChannelAustralia            1.99      0.340      5.85 3.12e- 6
##  4 Distribution.ChannelDigital distribution  5.32      0.291     18.3  9.86e-17
##  5 Distribution.ChannelEurope               4.86      0.340     14.3  4.03e-14
##  6 Distribution.ChannelNorth America        4.46      0.340     13.1  3.06e-13
##  7 Distribution.ChannelPC                   4.96      0.291     17.1  5.57e-16
##  8 Distribution.ChannelPhysical distribution  4.39      0.291     15.1  1.13e-14
##  9 Distribution.ChannelPlaystation          4.68      0.291     16.1  2.36e-15
## 10 Distribution.ChannelSouth America        3.06      0.340      8.99 1.31e- 9
## 11 Distribution.ChannelSwitch               3.57      0.340     10.5  4.90e-11
## 12 Distribution.ChannelXbox                 3.67      0.291     12.6  7.91e-13
## 13 as.factor(Year)2016                     -0.821     0.215     -3.82 7.06e- 4
## 14 as.factor(Year)2017                     -0.761     0.215     -3.54 1.46e- 3
## 15 as.factor(Year)2018                     -1.07      0.215     -4.99 3.16e- 5
## 16 as.factor(Year)2019                     -0.472     0.199     -2.37 2.49e- 2
## 17 as.factor(Year)2020                     -0.444     0.199     -2.23 3.40e- 2
```

```
# MULTICOLLINEARITY

vif(data_result_11)
```

```
##                       GVIF Df GVIF^(1/(2*Df))
## Distribution.Channel 1.636 11           1.023
## as.factor(Year)      1.636  5           1.050
```

```
# Interpretation: From the GVIF,there is no multivariate problem as the explanatory variables we
re below 5.
```

Model validation

Assumption 1: Outliers

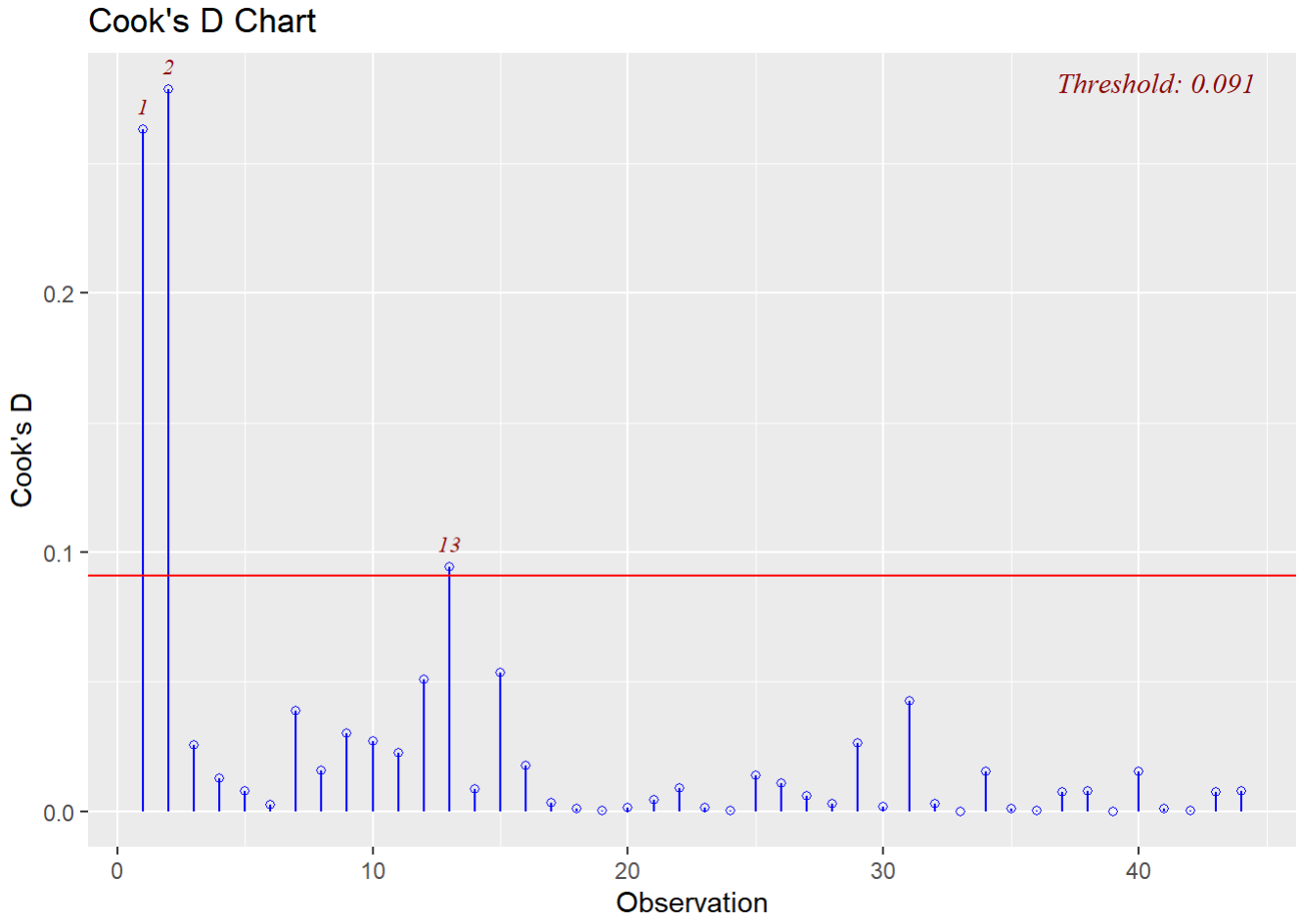
```
library(olsrr)
```

```
##
## Attaching package: 'olsrr'
```

```
## The following object is masked from 'package:equatiomatic':
##
##     hsb
```

```
## The following object is masked from 'package:datasets':
##
##     rivers
```

```
data_result_11 %>% ols_plot_cooksd_chart()
```




```
# Interpretations: 1, 2 and 13 are outliers as they are above the 0.091 threshold
```

Assumption 2: Normality

```
data_result_11 %>% ols_test_normality()
```

```
## -----------------------------------------------
##        Test             Statistic       pvalue
## -----------------------------------------------
## Shapiro-Wilk              0.948          0.0466
## Kolmogorov-Smirnov        0.1169         0.5458
## Cramer-von Mises          9.1024         0.0000
## Anderson-Darling          0.7773         0.0402
## -----------------------------------------------
```

```
# Interpretation: Shapiro-Wilk test is below 0.05 meaning the data significantly deviate from a
 normal distribution.
```

Assumption 3: Heteroskedasticity

```
data_result_11 %>% ols_test_breusch_pagan()
```

```
##
##  Breusch Pagan Test for Heteroskedasticity
##  -----------------------------------------
##  Ho: the variance is constant
##  Ha: the variance is not constant
##
##                          Data
##  -----------------------------------------------------
##  Response : log(Units.sold..estimates.)
##  Variables: fitted values of log(Units.sold..estimates.)
##
##         Test Summary
##  ---------------------------
##  DF            =    1
##  Chi2          =    11.3089
##  Prob > Chi2   =    0.0007714
```

```
# Interpretation: the p-value is below an appropriate threshold (e.g. p < 0.05) therefore the nu
ll hypothesis of homoskedasticity is rejected and heteroskedasticity assumed.
```