Project\_stat8030.R

Hp

2023-06-26

library(tidyverse)

## Warning: package 'ggplot2' was built under R version 4.3.1

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.2 ✔ readr 2.1.4  
## ✔ forcats 1.0.0 ✔ stringr 1.5.0  
## ✔ ggplot2 3.4.2 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.2 ✔ tidyr 1.3.0  
## ✔ purrr 1.0.1   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

library(MASS)

##   
## Attaching package: 'MASS'  
##   
## The following object is masked from 'package:dplyr':  
##   
## select

library(dplyr)  
library(stargazer)

##   
## Please cite as:   
##   
## Hlavac, Marek (2022). stargazer: Well-Formatted Regression and Summary Statistics Tables.  
## R package version 5.2.3. https://CRAN.R-project.org/package=stargazer

library(caret)

## Loading required package: lattice  
##   
## Attaching package: 'caret'  
##   
## The following object is masked from 'package:purrr':  
##   
## lift

library(leaps)  
library(ggplot2)  
library(readr)  
lego\_sets <- read\_csv("C:/Users/Hp/Downloads/archive (2)/lego\_sets.csv")

## Rows: 12261 Columns: 14  
## ── Column specification ────────────────────────────────────────────────────────  
## Delimiter: ","  
## chr (7): ages, prod\_desc, prod\_long\_desc, review\_difficulty, set\_name, theme...  
## dbl (7): list\_price, num\_reviews, piece\_count, play\_star\_rating, prod\_id, st...  
##   
## ℹ Use `spec()` to retrieve the full column specification for this data.  
## ℹ Specify the column types or set `show\_col\_types = FALSE` to quiet this message.

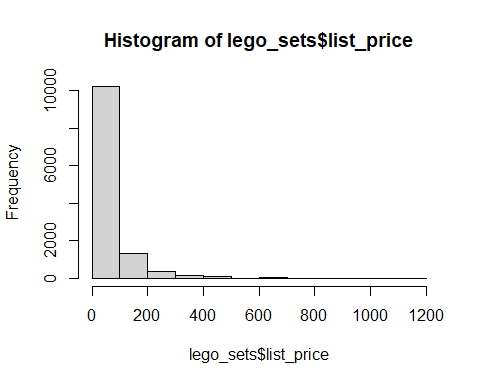
View(lego\_sets)  
  
#descriptive analytics   
str(lego\_sets)

## spc\_tbl\_ [12,261 × 14] (S3: spec\_tbl\_df/tbl\_df/tbl/data.frame)  
## $ ages : chr [1:12261] "6-12" "6-12" "6-12" "12+" ...  
## $ list\_price : num [1:12261] 30 20 13 100 80 ...  
## $ num\_reviews : num [1:12261] 2 2 11 23 14 7 37 24 23 11 ...  
## $ piece\_count : num [1:12261] 277 168 74 1032 744 ...  
## $ play\_star\_rating : num [1:12261] 4 4 4.3 3.6 3.2 3.7 3.7 4.4 3.6 3.6 ...  
## $ prod\_desc : chr [1:12261] "Catapult into action and take back the eggs from the Piggy Trike!" "Launch a flying attack and rescue the eggs from the Piggy Plane!" "Chase the piggy with lightning-fast Chuck and rescue the eggs!" "Explore the architecture of the United States Capitol Building!" ...  
## $ prod\_id : num [1:12261] 75823 75822 75821 21030 21035 ...  
## $ prod\_long\_desc : chr [1:12261] "Use the staircase catapult to launch Red into the air and race after the piggy, who is speeding off in the Pigg"| \_\_truncated\_\_ "Pilot Pig has taken off from Bird Island with 4 eggs in his plane and is making his escape. Shoot Red into the "| \_\_truncated\_\_ "Pitch speedy bird Chuck against the Piggy Car. Chase the car to try and take back the eggs, but watch out for i"| \_\_truncated\_\_ "Discover the architectural secrets of the iconic United States Capitol Building with this amazing LEGO® Archite"| \_\_truncated\_\_ ...  
## $ review\_difficulty: chr [1:12261] "Average" "Easy" "Easy" "Average" ...  
## $ set\_name : chr [1:12261] "Bird Island Egg Heist" "Piggy Plane Attack" "Piggy Car Escape" "United States Capitol Building" ...  
## $ star\_rating : num [1:12261] 4.5 5 4.3 4.6 4.6 4.9 4.2 4.7 4.7 4.8 ...  
## $ theme\_name : chr [1:12261] "Angry Birds™" "Angry Birds™" "Angry Birds™" "Architecture" ...  
## $ val\_star\_rating : num [1:12261] 4 4 4.1 4.3 4.1 4.4 4.1 4.3 4.1 4.5 ...  
## $ country : chr [1:12261] "US" "US" "US" "US" ...  
## - attr(\*, "spec")=  
## .. cols(  
## .. ages = col\_character(),  
## .. list\_price = col\_double(),  
## .. num\_reviews = col\_double(),  
## .. piece\_count = col\_double(),  
## .. play\_star\_rating = col\_double(),  
## .. prod\_desc = col\_character(),  
## .. prod\_id = col\_double(),  
## .. prod\_long\_desc = col\_character(),  
## .. review\_difficulty = col\_character(),  
## .. set\_name = col\_character(),  
## .. star\_rating = col\_double(),  
## .. theme\_name = col\_character(),  
## .. val\_star\_rating = col\_double(),  
## .. country = col\_character()  
## .. )  
## - attr(\*, "problems")=<externalptr>

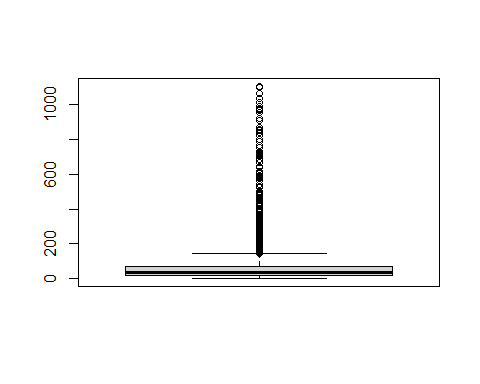
summary(lego\_sets)

## ages list\_price num\_reviews piece\_count   
## Length:12261 Min. : 2.272 Min. : 1.00 Min. : 1.0   
## Class :character 1st Qu.: 19.990 1st Qu.: 2.00 1st Qu.: 97.0   
## Mode :character Median : 36.588 Median : 6.00 Median : 216.0   
## Mean : 65.142 Mean : 16.83 Mean : 493.4   
## 3rd Qu.: 70.192 3rd Qu.: 13.00 3rd Qu.: 544.0   
## Max. :1104.870 Max. :367.00 Max. :7541.0   
## NA's :1620   
## play\_star\_rating prod\_desc prod\_id prod\_long\_desc   
## Min. :1.000 Length:12261 Min. : 630 Length:12261   
## 1st Qu.:4.000 Class :character 1st Qu.: 21034 Class :character   
## Median :4.500 Mode :character Median : 42069 Mode :character   
## Mean :4.338 Mean : 59837   
## 3rd Qu.:4.800 3rd Qu.: 70922   
## Max. :5.000 Max. :2000431   
## NA's :1775   
## review\_difficulty set\_name star\_rating theme\_name   
## Length:12261 Length:12261 Min. :1.800 Length:12261   
## Class :character Class :character 1st Qu.:4.300 Class :character   
## Mode :character Mode :character Median :4.700 Mode :character   
## Mean :4.514   
## 3rd Qu.:5.000   
## Max. :5.000   
## NA's :1620   
## val\_star\_rating country   
## Min. :1.000 Length:12261   
## 1st Qu.:4.000 Class :character   
## Median :4.300 Mode :character   
## Mean :4.229   
## 3rd Qu.:4.700   
## Max. :5.000   
## NA's :1795

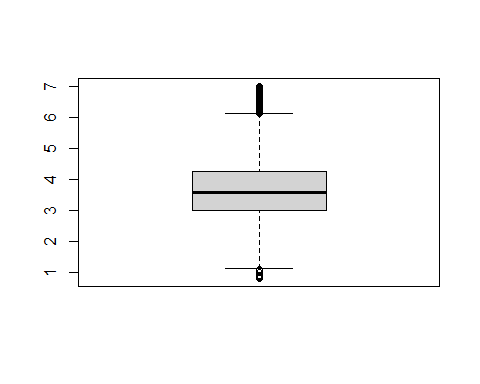
#histogram   
hist(lego\_sets$list\_price)



boxplot(lego\_sets$list\_price,width = 0.7)



boxplot(log(lego\_sets$list\_price))



#correlations   
numeric\_data <- lego\_sets[, sapply(lego\_sets, is.numeric)]  
cor\_matrix <- cor(numeric\_data)  
print(cor\_matrix)

## list\_price num\_reviews piece\_count play\_star\_rating prod\_id  
## list\_price 1.0000000 NA 0.8696299 NA 0.3886331  
## num\_reviews NA 1 NA NA NA  
## piece\_count 0.8696299 NA 1.0000000 NA 0.2177165  
## play\_star\_rating NA NA NA 1 NA  
## prod\_id 0.3886331 NA 0.2177165 NA 1.0000000  
## star\_rating NA NA NA NA NA  
## val\_star\_rating NA NA NA NA NA  
## star\_rating val\_star\_rating  
## list\_price NA NA  
## num\_reviews NA NA  
## piece\_count NA NA  
## play\_star\_rating NA NA  
## prod\_id NA NA  
## star\_rating 1 NA  
## val\_star\_rating NA 1

cor(lego\_sets$list\_price,lego\_sets$piece\_count)

## [1] 0.8696299

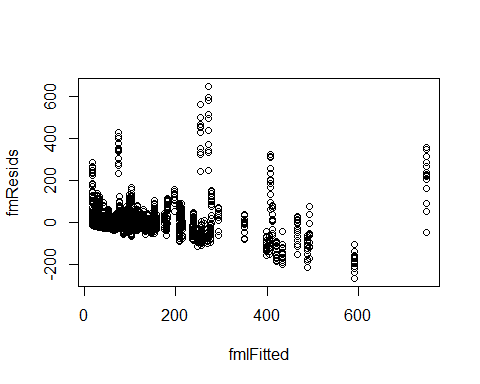
#first\_model  
attach(lego\_sets)  
fm<-lm(list\_price~piece\_count)  
summary(fm)

##   
## Call:  
## lm(formula = list\_price ~ piece\_count)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -267.46 -14.38 -6.45 6.97 650.69   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.732e+01 4.778e-01 36.26 <2e-16 \*\*\*  
## piece\_count 9.691e-02 4.969e-04 195.03 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 45.41 on 12259 degrees of freedom  
## Multiple R-squared: 0.7563, Adjusted R-squared: 0.7562   
## F-statistic: 3.804e+04 on 1 and 12259 DF, p-value: < 2.2e-16

sm<-lm(list\_price~prod\_id+piece\_count+play\_star\_rating+num\_reviews+review\_difficulty+star\_rating+country)  
#sm <- lm(list\_price ~ ., data = lego\_sets)  
summary(sm)

##   
## Call:  
## lm(formula = list\_price ~ prod\_id + piece\_count + play\_star\_rating +   
## num\_reviews + review\_difficulty + star\_rating + country)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -345.81 -13.89 -4.09 8.15 432.86   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.097e+01 4.355e+00 11.703 < 2e-16 \*\*\*  
## prod\_id 1.246e-04 2.567e-06 48.535 < 2e-16 \*\*\*  
## piece\_count 8.644e-02 7.312e-04 118.218 < 2e-16 \*\*\*  
## play\_star\_rating 8.664e+00 8.424e-01 10.285 < 2e-16 \*\*\*  
## num\_reviews 1.794e-02 1.437e-02 1.249 0.211643   
## review\_difficultyChallenging 1.188e+01 1.874e+00 6.342 2.37e-10 \*\*\*  
## review\_difficultyEasy -1.240e+01 1.004e+00 -12.342 < 2e-16 \*\*\*  
## review\_difficultyVery Challenging 3.912e+01 1.532e+01 2.554 0.010657 \*   
## review\_difficultyVery Easy -9.926e+00 1.523e+00 -6.518 7.48e-11 \*\*\*  
## star\_rating -1.553e+01 1.045e+00 -14.863 < 2e-16 \*\*\*  
## countryAU -2.480e+00 2.797e+00 -0.886 0.375404   
## countryBE 4.703e+00 2.802e+00 1.678 0.093313 .   
## countryCA -1.311e+01 2.568e+00 -5.105 3.38e-07 \*\*\*  
## countryCH -2.075e-01 2.795e+00 -0.074 0.940823   
## countryCZ 6.699e+00 2.827e+00 2.370 0.017819 \*   
## countryDE 3.246e+00 2.827e+00 1.148 0.250856   
## countryDN -1.014e+01 2.793e+00 -3.630 0.000284 \*\*\*  
## countryES 3.246e+00 2.827e+00 1.148 0.250856   
## countryFI 1.251e+01 2.827e+00 4.426 9.69e-06 \*\*\*  
## countryFR 2.186e+00 2.804e+00 0.780 0.435612   
## countryGB 1.080e+00 2.792e+00 0.387 0.698915   
## countryIE 4.023e-01 2.796e+00 0.144 0.885590   
## countryIT 6.809e-01 2.825e+00 0.241 0.809542   
## countryLU 4.793e+00 2.827e+00 1.695 0.090025 .   
## countryNL 4.923e+00 2.789e+00 1.765 0.077527 .   
## countryNO 1.151e+01 2.827e+00 4.073 4.68e-05 \*\*\*  
## countryNZ 5.482e+00 2.811e+00 1.950 0.051199 .   
## countryPL 5.053e+00 2.855e+00 1.770 0.076797 .   
## countryPT 2.779e+00 2.830e+00 0.982 0.326049   
## countryUS -1.053e+01 2.566e+00 -4.103 4.10e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 43.13 on 10155 degrees of freedom  
## (2076 observations deleted due to missingness)  
## Multiple R-squared: 0.8106, Adjusted R-squared: 0.8101   
## F-statistic: 1499 on 29 and 10155 DF, p-value: < 2.2e-16

#diagnostic   
  
fmResids <- fm$residuals  
fmlFitted <- fm$fitted.value  
  
plot(fmlFitted,fmResids)



dev.new(width = 10, height = 8)  
par(mar = c(5, 5, 2, 2))  
  
  
smResids <- sm$residuals  
smlFitted <- sm$fitted.value  
plot(smlFitted,smResids)  
qqnorm(fmResids)  
qqnorm(smResids)  
  
#extension   
summary(sm)

##   
## Call:  
## lm(formula = list\_price ~ prod\_id + piece\_count + play\_star\_rating +   
## num\_reviews + review\_difficulty + star\_rating + country)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -345.81 -13.89 -4.09 8.15 432.86   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 5.097e+01 4.355e+00 11.703 < 2e-16 \*\*\*  
## prod\_id 1.246e-04 2.567e-06 48.535 < 2e-16 \*\*\*  
## piece\_count 8.644e-02 7.312e-04 118.218 < 2e-16 \*\*\*  
## play\_star\_rating 8.664e+00 8.424e-01 10.285 < 2e-16 \*\*\*  
## num\_reviews 1.794e-02 1.437e-02 1.249 0.211643   
## review\_difficultyChallenging 1.188e+01 1.874e+00 6.342 2.37e-10 \*\*\*  
## review\_difficultyEasy -1.240e+01 1.004e+00 -12.342 < 2e-16 \*\*\*  
## review\_difficultyVery Challenging 3.912e+01 1.532e+01 2.554 0.010657 \*   
## review\_difficultyVery Easy -9.926e+00 1.523e+00 -6.518 7.48e-11 \*\*\*  
## star\_rating -1.553e+01 1.045e+00 -14.863 < 2e-16 \*\*\*  
## countryAU -2.480e+00 2.797e+00 -0.886 0.375404   
## countryBE 4.703e+00 2.802e+00 1.678 0.093313 .   
## countryCA -1.311e+01 2.568e+00 -5.105 3.38e-07 \*\*\*  
## countryCH -2.075e-01 2.795e+00 -0.074 0.940823   
## countryCZ 6.699e+00 2.827e+00 2.370 0.017819 \*   
## countryDE 3.246e+00 2.827e+00 1.148 0.250856   
## countryDN -1.014e+01 2.793e+00 -3.630 0.000284 \*\*\*  
## countryES 3.246e+00 2.827e+00 1.148 0.250856   
## countryFI 1.251e+01 2.827e+00 4.426 9.69e-06 \*\*\*  
## countryFR 2.186e+00 2.804e+00 0.780 0.435612   
## countryGB 1.080e+00 2.792e+00 0.387 0.698915   
## countryIE 4.023e-01 2.796e+00 0.144 0.885590   
## countryIT 6.809e-01 2.825e+00 0.241 0.809542   
## countryLU 4.793e+00 2.827e+00 1.695 0.090025 .   
## countryNL 4.923e+00 2.789e+00 1.765 0.077527 .   
## countryNO 1.151e+01 2.827e+00 4.073 4.68e-05 \*\*\*  
## countryNZ 5.482e+00 2.811e+00 1.950 0.051199 .   
## countryPL 5.053e+00 2.855e+00 1.770 0.076797 .   
## countryPT 2.779e+00 2.830e+00 0.982 0.326049   
## countryUS -1.053e+01 2.566e+00 -4.103 4.10e-05 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 43.13 on 10155 degrees of freedom  
## (2076 observations deleted due to missingness)  
## Multiple R-squared: 0.8106, Adjusted R-squared: 0.8101   
## F-statistic: 1499 on 29 and 10155 DF, p-value: < 2.2e-16

smp<-lm(list\_price~prod\_id+I(prod\_id^2)+piece\_count+prod\_id:play\_star\_rating+play\_star\_rating+num\_reviews+review\_difficulty+star\_rating+country)  
  
summary(smp)

##   
## Call:  
## lm(formula = list\_price ~ prod\_id + I(prod\_id^2) + piece\_count +   
## prod\_id:play\_star\_rating + play\_star\_rating + num\_reviews +   
## review\_difficulty + star\_rating + country)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -316.45 -13.71 -4.04 8.21 450.25   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.416e+01 4.675e+00 9.447 < 2e-16 \*\*\*  
## prod\_id 2.207e-04 2.543e-05 8.677 < 2e-16 \*\*\*  
## I(prod\_id^2) 3.398e-12 8.948e-12 0.380 0.704145   
## piece\_count 8.729e-02 7.482e-04 116.669 < 2e-16 \*\*\*  
## play\_star\_rating 9.838e+00 8.685e-01 11.328 < 2e-16 \*\*\*  
## num\_reviews 1.126e-02 1.451e-02 0.776 0.437798   
## review\_difficultyChallenging 1.096e+01 1.880e+00 5.831 5.66e-09 \*\*\*  
## review\_difficultyEasy -1.185e+01 1.008e+00 -11.752 < 2e-16 \*\*\*  
## review\_difficultyVery Challenging 3.843e+01 1.530e+01 2.512 0.012033 \*   
## review\_difficultyVery Easy -9.482e+00 1.531e+00 -6.193 6.13e-10 \*\*\*  
## star\_rating -1.517e+01 1.049e+00 -14.468 < 2e-16 \*\*\*  
## countryAU -2.484e+00 2.794e+00 -0.889 0.373908   
## countryBE 4.697e+00 2.798e+00 1.678 0.093288 .   
## countryCA -1.313e+01 2.564e+00 -5.121 3.09e-07 \*\*\*  
## countryCH -2.089e-01 2.791e+00 -0.075 0.940334   
## countryCZ 6.684e+00 2.823e+00 2.368 0.017918 \*   
## countryDE 3.232e+00 2.823e+00 1.145 0.252328   
## countryDN -1.014e+01 2.789e+00 -3.636 0.000278 \*\*\*  
## countryES 3.232e+00 2.823e+00 1.145 0.252328   
## countryFI 1.250e+01 2.823e+00 4.427 9.67e-06 \*\*\*  
## countryFR 2.185e+00 2.800e+00 0.780 0.435147   
## countryGB 1.080e+00 2.788e+00 0.387 0.698581   
## countryIE 4.018e-01 2.792e+00 0.144 0.885582   
## countryIT 6.673e-01 2.821e+00 0.237 0.813032   
## countryLU 4.778e+00 2.823e+00 1.692 0.090595 .   
## countryNL 4.923e+00 2.785e+00 1.768 0.077150 .   
## countryNO 1.150e+01 2.823e+00 4.073 4.68e-05 \*\*\*  
## countryNZ 5.474e+00 2.807e+00 1.950 0.051223 .   
## countryPL 5.045e+00 2.851e+00 1.769 0.076853 .   
## countryPT 2.762e+00 2.826e+00 0.977 0.328373   
## countryUS -1.056e+01 2.563e+00 -4.120 3.83e-05 \*\*\*  
## prod\_id:play\_star\_rating -2.406e-05 4.395e-06 -5.475 4.48e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 43.07 on 10153 degrees of freedom  
## (2076 observations deleted due to missingness)  
## Multiple R-squared: 0.8112, Adjusted R-squared: 0.8106   
## F-statistic: 1407 on 31 and 10153 DF, p-value: < 2.2e-16

#Feature\_selection  
step<-stepAIC(smp,direction= "forward",trace=FALSE)  
step$anova

## Stepwise Model Path   
## Analysis of Deviance Table  
##   
## Initial Model:  
## list\_price ~ prod\_id + I(prod\_id^2) + piece\_count + prod\_id:play\_star\_rating +   
## play\_star\_rating + num\_reviews + review\_difficulty + star\_rating +   
## country  
##   
## Final Model:  
## list\_price ~ prod\_id + I(prod\_id^2) + piece\_count + prod\_id:play\_star\_rating +   
## play\_star\_rating + num\_reviews + review\_difficulty + star\_rating +   
## country  
##   
##   
## Step Df Deviance Resid. Df Resid. Dev AIC  
## 1 10153 18838097 76682.91

step1<-stepAIC(smp,direction= "backward",trace=FALSE)  
step1$anova

## Stepwise Model Path   
## Analysis of Deviance Table  
##   
## Initial Model:  
## list\_price ~ prod\_id + I(prod\_id^2) + piece\_count + prod\_id:play\_star\_rating +   
## play\_star\_rating + num\_reviews + review\_difficulty + star\_rating +   
## country  
##   
## Final Model:  
## list\_price ~ prod\_id + piece\_count + play\_star\_rating + review\_difficulty +   
## star\_rating + country + prod\_id:play\_star\_rating  
##   
##   
## Step Df Deviance Resid. Df Resid. Dev AIC  
## 1 10153 18838097 76682.91  
## 2 - I(prod\_id^2) 1 267.5583 10154 18838364 76681.05  
## 3 - num\_reviews 1 1284.1252 10155 18839649 76679.75

smnp<-lm(list\_price~prod\_id+piece\_count+prod\_id:play\_star\_rating+play\_star\_rating+review\_difficulty+star\_rating+country)  
summary(smnp)

##   
## Call:  
## lm(formula = list\_price ~ prod\_id + piece\_count + prod\_id:play\_star\_rating +   
## play\_star\_rating + review\_difficulty + star\_rating + country)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -316.46 -13.70 -4.05 8.20 450.69   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 4.385e+01 4.543e+00 9.652 < 2e-16 \*\*\*  
## prod\_id 2.280e-04 1.891e-05 12.054 < 2e-16 \*\*\*  
## piece\_count 8.751e-02 6.863e-04 127.516 < 2e-16 \*\*\*  
## play\_star\_rating 9.822e+00 8.676e-01 11.322 < 2e-16 \*\*\*  
## review\_difficultyChallenging 1.128e+01 1.846e+00 6.111 1.03e-09 \*\*\*  
## review\_difficultyEasy -1.182e+01 1.008e+00 -11.728 < 2e-16 \*\*\*  
## review\_difficultyVery Challenging 3.784e+01 1.529e+01 2.476 0.013311 \*   
## review\_difficultyVery Easy -9.323e+00 1.519e+00 -6.139 8.61e-10 \*\*\*  
## star\_rating -1.516e+01 1.046e+00 -14.496 < 2e-16 \*\*\*  
## countryAU -2.482e+00 2.793e+00 -0.889 0.374198   
## countryBE 4.699e+00 2.798e+00 1.679 0.093110 .   
## countryCA -1.315e+01 2.564e+00 -5.127 3.00e-07 \*\*\*  
## countryCH -2.084e-01 2.791e+00 -0.075 0.940487   
## countryCZ 6.692e+00 2.823e+00 2.371 0.017764 \*   
## countryDE 3.239e+00 2.823e+00 1.148 0.251196   
## countryDN -1.014e+01 2.789e+00 -3.637 0.000277 \*\*\*  
## countryES 3.239e+00 2.823e+00 1.148 0.251196   
## countryFI 1.250e+01 2.823e+00 4.430 9.52e-06 \*\*\*  
## countryFR 2.188e+00 2.800e+00 0.782 0.434443   
## countryGB 1.079e+00 2.788e+00 0.387 0.698713   
## countryIE 4.004e-01 2.792e+00 0.143 0.885956   
## countryIT 6.762e-01 2.821e+00 0.240 0.810583   
## countryLU 4.786e+00 2.823e+00 1.696 0.090003 .   
## countryNL 4.922e+00 2.785e+00 1.768 0.077151 .   
## countryNO 1.151e+01 2.823e+00 4.076 4.61e-05 \*\*\*  
## countryNZ 5.478e+00 2.807e+00 1.951 0.051055 .   
## countryPL 5.056e+00 2.851e+00 1.773 0.076206 .   
## countryPT 2.769e+00 2.826e+00 0.980 0.327099   
## countryUS -1.057e+01 2.562e+00 -4.125 3.73e-05 \*\*\*  
## prod\_id:play\_star\_rating -2.424e-05 4.374e-06 -5.542 3.07e-08 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 43.07 on 10155 degrees of freedom  
## (2076 observations deleted due to missingness)  
## Multiple R-squared: 0.8112, Adjusted R-squared: 0.8106   
## F-statistic: 1504 on 29 and 10155 DF, p-value: < 2.2e-16

summary(fm)

##   
## Call:  
## lm(formula = list\_price ~ piece\_count)  
##   
## Residuals:  
## Min 1Q Median 3Q Max   
## -267.46 -14.38 -6.45 6.97 650.69   
##   
## Coefficients:  
## Estimate Std. Error t value Pr(>|t|)   
## (Intercept) 1.732e+01 4.778e-01 36.26 <2e-16 \*\*\*  
## piece\_count 9.691e-02 4.969e-04 195.03 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## Residual standard error: 45.41 on 12259 degrees of freedom  
## Multiple R-squared: 0.7563, Adjusted R-squared: 0.7562   
## F-statistic: 3.804e+04 on 1 and 12259 DF, p-value: < 2.2e-16

#prediction\_model  
  
piece\_count\_predictions <- data.frame(piece\_count = c(10, 20, 30))  
predict(fm,piece\_count\_predictions)

## 1 2 3   
## 18.29346 19.26259 20.23173

fm<-lm(list\_price~piece\_count)  
piece\_count\_predictions <- data.frame(piece\_count = c(10, 20, 30))  
  
prediction\_interval <- predict(fm, newdata = piece\_count\_predictions, interval = "confidence", level = 0.95)  
prediction\_interval

## fit lwr upr  
## 1 18.29346 17.36180 19.22512  
## 2 19.26259 18.33582 20.18937  
## 3 20.23173 19.30977 21.15369