**The impact of Blu-ray players’ price on customer satisfaction**

Coursework Assignment 1 for the Course “MN-2015 Programming for Analytics”

School of Management

Swansea, November 20th, 2017

**Table of Contents**

[1 Data 4](#_Toc498945886)

[2 Model Specification 5](#_Toc498945887)

[3 Graphical Representation 10](#_Toc498945888)

[4 From Model 2 to Model 2.1: Outliers 11](#_Toc498945889)

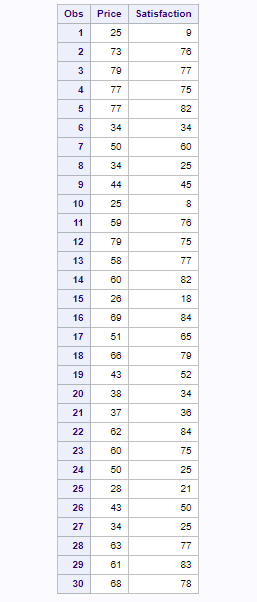
[5 Blu-ray Players Price&Hotel Guests Satisfaction: Management Report 14](#_Toc498945890)

[6 List of References 17](#_Toc498945891)

# Data

The goal of this analysis can be formulated as follows: a hotel manager wants to see if it is possible to predict customer satisfaction from the price of Blu-ray players. The data on 30 models were gathered and a permanent dataset was created in SAS with the following code:

**data** msasdata.BluRayPlayers;

input Price Satisfaction; Table 1 Blurayplayers Dataset

datalines;

25 9

73 76

79 77

77 75

77 82

34 34

50 60

34 25

44 45

25 8

59 76

79 75

58 77

60 82

26 18

69 84

51 65

66 79

43 52

38 34

37 36

62 84

60 75

50 25

28 21

43 50

34 25

63 77

61 83

68 78

;

**proc** **print** data=msasdata.BluRayPlayers; **run**;

The output is shown in Table 1:

# Model Specification

We investigate the theory that more expensive Blu-ray players tend to produce higher customer satisfaction scores, so presumably there is a positive relationship between two variables. But obviously we do not know what mathematical function describes this relationship. We can expect that at some point this function levels off because the logic behind diminishing marginal returns might apply: the most expensive Blu-ray players would have similar perceived quality with less expensive ones and produce similar customer satisfaction.

First, we need to see how the data look like. Therefore, we plot a scatter diagram of Satisfaction against Price, as well as run proc means procedure with the summary statistics on central tendency:

**proc** **gplot** data=msasdata.BluRayPlayers;

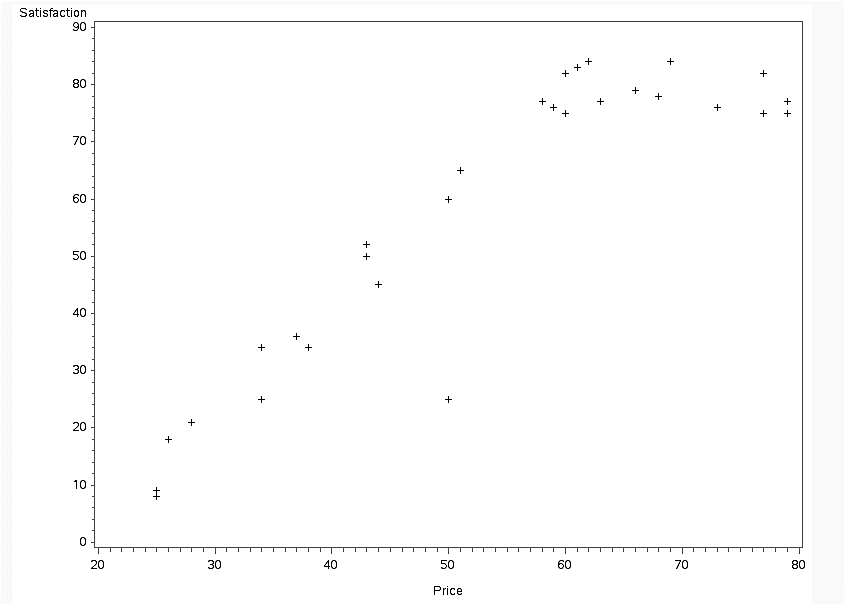
plot satisfaction\*price;

**proc** **means** data=msasdata.BluRayPlayers mean median mode maximum minimum;

**run**;

with the output presented in Figure 1 and Table 2.

Figure 1 Price vs Satisfaction



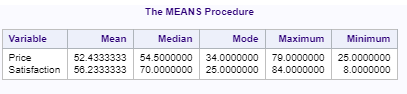
It seems that our expectation is likely to be confirmed and the data might demonstrate the decline in the customer satisfaction responsiveness to the change in Blu-ray players’ price, while moving further up from the mid-price ones. Moreover, since the medians are higher than the means, more of the data are to the right of the means and closer to them. Whereas the first part of the data could possibly follow a linear function, the second half levels out, with the price varying from 54.5 to 79 and the respective customer satisfaction laying in between 70 and 84. So, the following question arises: do these data follow a linear relationship?

Table 2 Summary Statistics on Price and Satisfaction

To answer this question, we need to check the major assumptions of a linear regression method which are:

1. Residuals should be completely random; in other words, they do not follow any pattern at all (Derby, 2010).
2. Residuals should fit a normal distribution (Derby, 2010).

It would be necessary enough to check the first assumption only because randomness in the residuals is crucial for the model to be right. To do so, we run a proc reg procedure and plot residuals against one of the variables - either price, satisfaction or the predicted values:

**proc** **reg** data=msasdata.BluRayPlayers noprint;

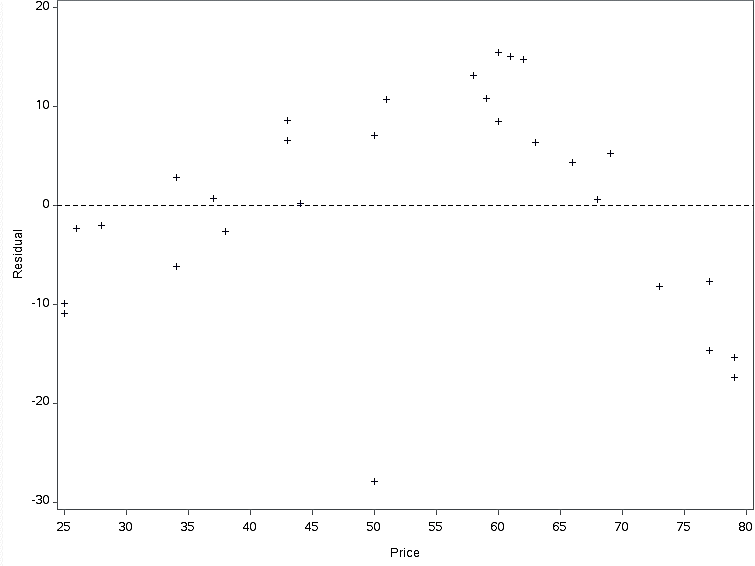
model satisfaction = price;

plot residual.\*price;

**run**;

which produces the output in Figure 2:

Figure 2 Price vs Simple Linear Regression Residuals



On this plot, we see an obvious pattern that data points reproduce. That is a bit blurred downward sloping concave curve. Thus, we conclude that a line would not be an accurate representation of the data, but a cubic polynomial function would become a better alternative. This function has a form (Support.sas.com, 2017):

To proceed further, we need to either modify the linear regression model or our data appropriately. With the DATA step, we add the quadratic and cubic terms pricesq and pricecube to account for polynomial effects in the MODEL statement in PROC REG (Support.sas.com, 2017). We output regression and ANOVA results. Since we have more than one regressor, a fit plot is not created; instead, we output a plot of residuals against predictions (Support.sas.com, 2017):

**data** msasdata.players;

set msasdata.blurayplayers;

pricesq=price\*price;

pricecube=price\*price\*price;

ods graphics on;

**proc** **reg** data=msasdata.players plots=predictions(X=Price);

model Satisfaction= price pricesq pricecube;

**run;**

The results can be found in Table 3 and Figures 3 to 4.

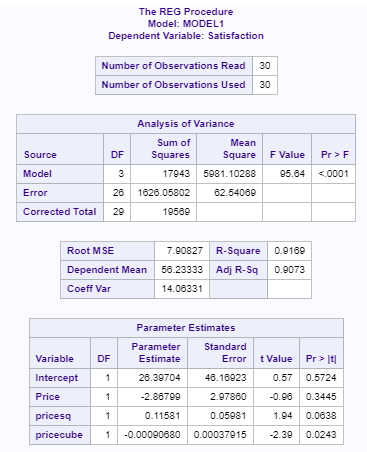
The F-Test in Table 3 shows that the model has explanatory power, with p<0.0001. According to the coefficient of determination, the model explains 92% of the variation in Satisfaction. However, parameter estimates are not significant at 95% CL (their p-values are higher than 0.05). We assume that it might be either due to the wrong model specification or not enough data points.

Table 3 ANOVA and Regression Results Model 1

Figure 3 shows a plot of price against residuals. Now we see that the residuals are distributed randomly, between positive and negative values. So, there is no pattern as it could be observed before in Figure 2.

Figure 4 Predicted Vs Dependent Model 1

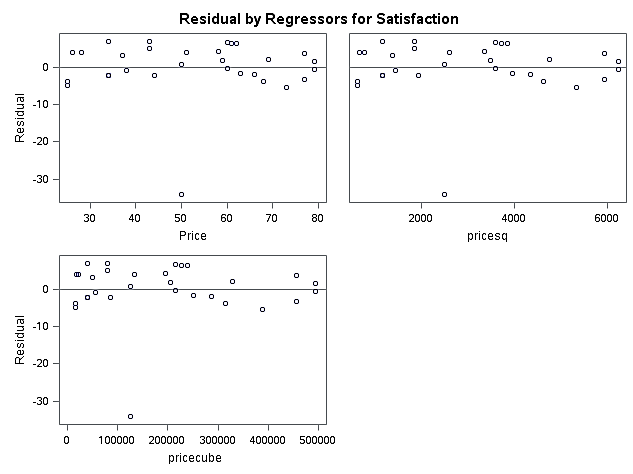
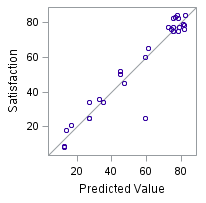


Figure 3 Price vs Cubic Regression Residuals Model 1

Moreover, the plot of the dependent versus the predicted variable in Figure 5 exhibits that they lie close to each other. Therefore, we can make an overall conclusion that this model better fits the data.

However, we are going to remove Price as a regressor to see if other regression coefficients become significant. The last line in the previous code is changed to model Satisfaction= pricesq pricecube;

The ANOVA and regression output are shown below in Table 4:

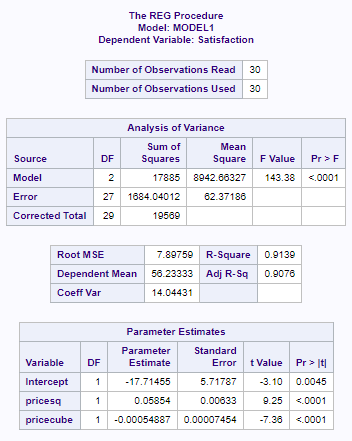
As can be seen, the regression coefficients are now statistically significant, with p-values <.0001. Importantly, the model still demonstrates a very good fit to data, with R-square=0.9139 and no patterns in the residuals.

Table 4 ANOVA and Regression Results Model 2

We write down a regression equation from the table of Parameter Estimates:

Satisfaction=-17.71455+0.05854Price2-0.00055Price3.

The both regression coefficients are significantly different from zero, with P-Values<0.0001, so there is some evidence that there is a relationship between price of Blue-ray players and customer satisfaction. This relationship is positive until the function reaches its maximum, at (0.05854\*2)\*Price-(0.00055\*3)\*Price2=0, P=70.96, and then it becomes negative. The intercept parameter shows by how much the customer satisfaction would fall if there was no blue-ray player, estimated at -17.71455. In addition, the customer satisfaction score is expected to change at the rate = (0.05854\*2)\*Price-(0.00055\*3)\*Price2.

# Graphical Representation

The Fit Plot for Satisfaction shown in Figure 5 was created by requesting the PLOTS=PREDICTIONS option in the previous proc reg code.

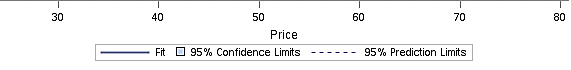
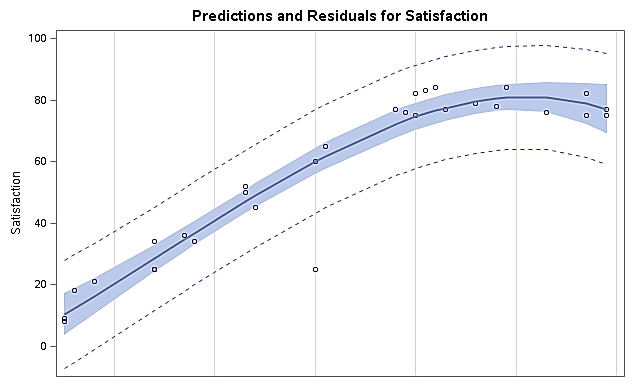
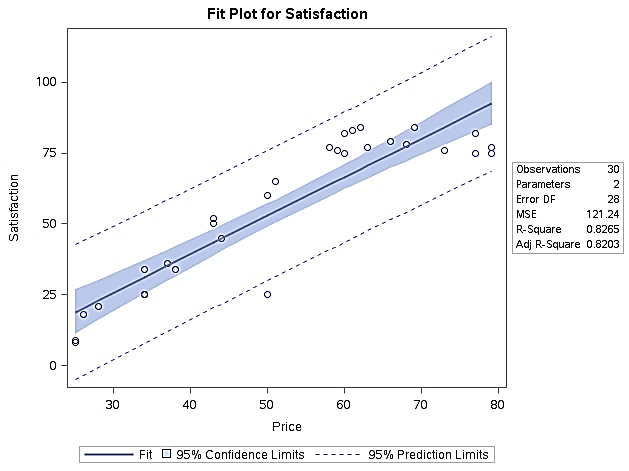


Figure 5 Fit Plot For Satisfaction Model 2

Figure 6 Fit Plot For Satisfaction Model 2



.

This plot contains of a scatter plot of Satisfaction against Price, the fitted cubic line as well as 95% confidence and prediction limits. It clearly shows that from the price of around 60 the customer satisfaction score levels out or even starts decreasing slightly. In meantime, for the Blu-ray players’ prices up to 60 the increase in customer satisfaction is rapid.

We can also observe that there is not much scatter around the fitted line, which would either mean it is a reasonable model to put in or there is just not much variation in the data. The latter can be explained by the sample size which is only 30 observations. In comparison to the straight-line fit in Figure 6, the cubic curve better fits the data, with almost all the observations contained in the inner confidence limits. Overall, we conclude that it is a reasonable model to predict customer satisfaction, however, more observations might be needed for its further validation.

# From Model 2 to Model 2.1: Outliers

There is a data point on the graph in Figure 5 that lies out of the outer confidence limits much below the line. It appears to be an “outlier” which seems to drag the fitted line down. We do not believe that this point is contributing to the model. But we conduct a formal test with the following code:

ods graphics on;

**proc** **reg** data=msasdata.players plots=predictions(X=Price) noprint;

model Satisfaction= pricesq pricecube;

output out=msasdata.outlier p=predict residual=resid rstudent=delresid lcl=lowercl ucl=uppercl;

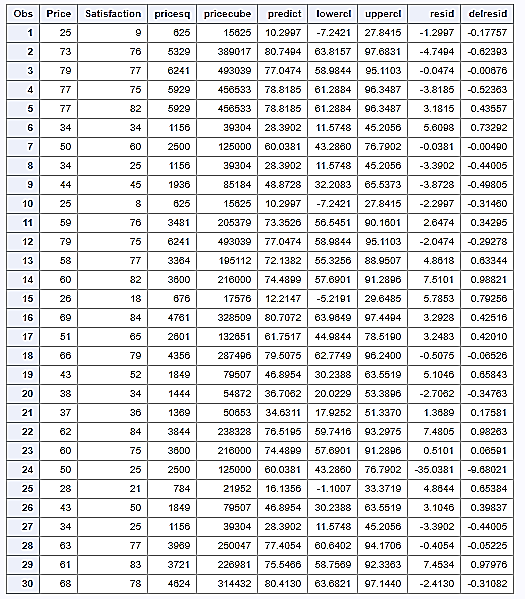
**proc** **print** data=msasdata.outlier;

**run**;

**quit**;

ods graphics off;

Table 5 Regression data Model 2

This code outputs a dataset called ‘outlier’. It includes a “studentized deletion residual” whose large values (magnitude 3 or more) indicate an outlier (Stat.purdue.edu, 2011). Clearly, the data point mentioned above is an observation 24, with Delresid = -9.68021. Therefore, we confirm that it is an outlier.

We remove it from the dataset and redo the calculation of the regression line, along with the Fit Plot for Satisfaction:

**data** msasdata.players2;

set msasdata.outlier;

if -**2.5**<delresid<**2.5**;

keep satisfaction price pricesq pricecube;

ods graphics on;

**proc** **reg** data=msasdata.players2 plots=predictions(X=Price);

model Satisfaction= pricesq pricecube;

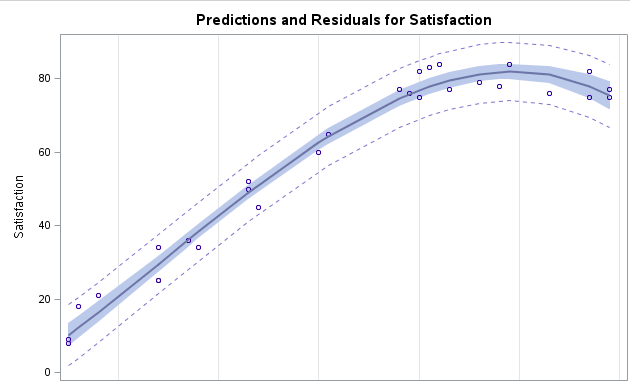


Figure 7 Fit Plot For Satisfaction Model 2.1 without Outlier

As displayed in Figure 7, the outlier disappeared. The new regression line is Satisfaction=-20.06264+0.06371\*Price2-0.000613\*Price3. New ANOVA results show that the model improved, with R-square=0.9803 in comparison to 0.9139 before.

We check the data for other outliers, by creating a dataset called outlier2:

ods graphics on;

**proc** **reg** data=msasdata.players2 plots=predictions(X=Price) noprint;

model Satisfaction= pricesq pricecube;

output out=msasdata.outlier2 p=predict residual=resid rstudent=delresid lcl=lowercl ucl=uppercl;

**proc** **print** data=msasdata.outlier2;

**run**;

**quit**;

ods graphics off;

According to Table 6, we do not have any candidates for outliers. Some might think about removing the observations 9 or 15. These are the extremes out of existing values, which does not mean that they are impossible. In addition, there is not much variation in the model. Therefore, we would keep the model how it is, without removing the variation further.

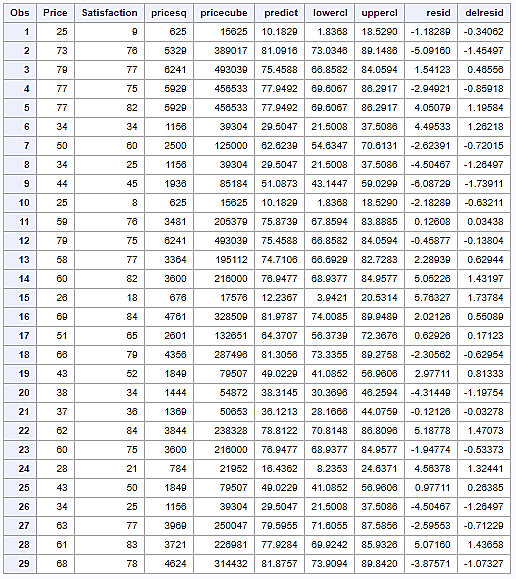


Table 6 Regression Data Model 2.1 without Outlier

After deleting the outlier, the model has become more accurate, with a better R-squared. The model fit and, consequently, its results were biased by the outlier, and therefore, the curve was not representative of the dataset. But we still need to understand the consequences for the business of such random observations. In this case, an outlier “below” the line is bad for the hotel – the satisfaction score of at least one customer is much less than the model predicts. In contrast, an outlier “above” the line is good – at least one customer would score his/her satisfaction way higher than the average for the same player.

# Blu-ray Players Price&Hotel Guests Satisfaction: Management Report

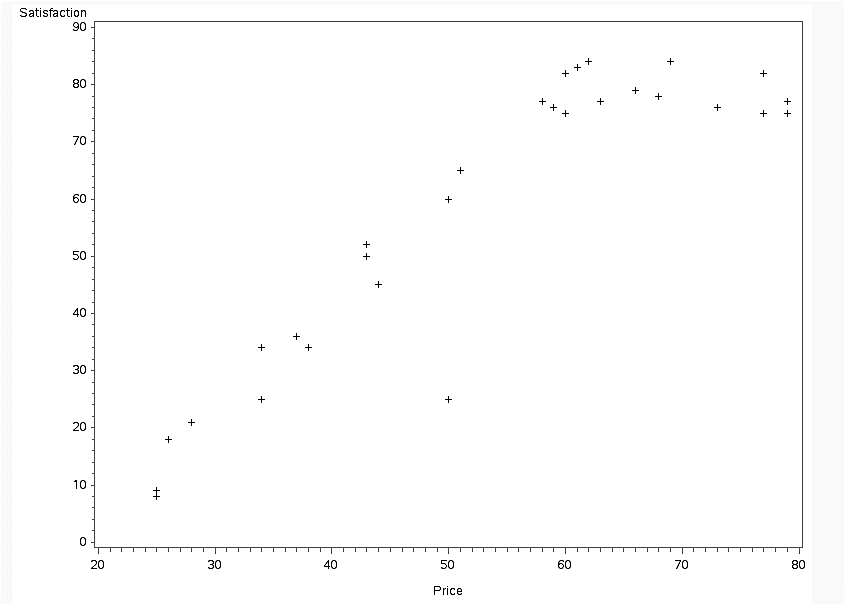
We investigated the hypothesis that there is a connection between the price of Blu-ray players and the customer satisfaction of hotel guests. To test this hypothesis, we asked regular guests about 30 models of Blu-ray players.

Figure 1 (repeated) Price vs Satisfaction

We see the collected data on a scatter plot in Figure 1. The data seem to confirm our hypothesis. To prove it scientifically, we applied a regression method. After conducting some formal procedures, we found out that a cubic line Satisfaction=-20.06264+0.06371Price2-0.00061273 Price3 would capture this data the best.

The equation of this line shows that there is a relationship between the price of Blu-ray players and the customer satisfaction because the coefficients of Price are not equal to zero. Moreover, all of them are statistically significant (i.e. their p-values are small enough to say there is one chance in 10,000 that the discussed relationship is random). So, there is a strong evidence to support the hypothesis (Kholopova and Berghoff, 2016).

This model explains 98 % of variation in the recorded customer satisfaction, so it can be used to decide on appropriate Blu-ray players. But it is important to note that we found an ‘outlier’ in the data – some customers like this “outlier” might respond unusually to the Blu-ray player’s characteristics and still show low satisfaction, with disregard to the relatively high price.

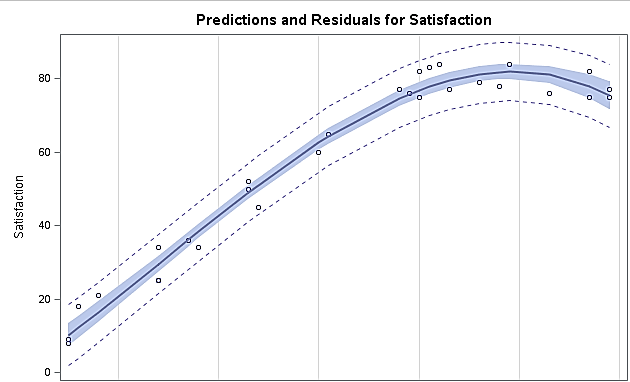
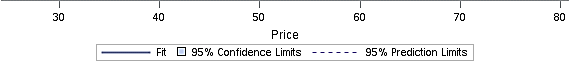


Figure 7 (repeated) Fit Plot For Satisfaction Model 2.1 without Outlier

Using the plot in Figure 7 as a guidance, it is clear that up until the price of around 69-70 the customer satisfaction score increases, and surprisingly enough from 69-70 it starts declining. It is because at Price=69.32 the curve reaches its maximum. Importantly, the derivative of this curve (0.05854\*2)\*Price-(0.00055\*3)\*Price2 shows that the customer satisfaction is very responsive to the price changes at lower prices and then closer to the middle this responsiveness declines, so that in the price range from 60 to 70 its marginal change is becoming quite small. Therefore, it would be reasonable to buy Blu-ray players with the price in a range from 60 to 69.32 which would provide guests with similar satisfaction levels. For example, the model predicts that S(60)=76.94368

S(63)=79.591

S(70)=81.94997

S(75)=79.81

S(80)=73.9636

In conclusion, to keep overhead costs in a reasonable range, but still improve customer satisfaction significantly, it is highly recommended to buy the middle-priced Blu-ray players, and not the most expensive ones. Interestingly enough, the most expensive ones could even provide a lower satisfaction score than the mid-price ones. The obtained model helps predicting customer satisfaction depending on the quality of Blu-ray players which is very useful in such a customer-oriented branch as hotel business. However, it is recommended to gather more data to verify the results on larger samples.

# List of References

Derby, N. (2010). *Getting Correct Results from PROC REG.* [online] Support.sas.com. Available at: http://support.sas.com/resources/papers/proceedings10/270-2010.pdf [Accessed 20 Nov. 2017].

Kholopova, M. and Berghoff, V. (2016). *The impact of level of education on age at first marriage.* Reutlingen: Reutlingen University.

Stat.purdue.edu. (2011). *Outliers & Influential Observations.* [online] Available at: http://www.stat.purdue.edu/~ghobbs/STAT\_512/Lecture\_Notes/Regression/Topic\_17.pdf [Accessed 20 Nov. 2017].

Support.sas.com. (2017). *Polynomial Regression.* [online] Available at: https://support.sas.com/documentation/cdl/en/statug/63033/HTML/default/viewer.htm#statug\_reg\_sect004.htm [Accessed 20 Nov. 2017].