University of Tartu
Faculty of Science and Technology
Institute of Mathematics and Statistics

Kristjan sirge

Measuring the effectiveness of casino bonuses using linear regression models

Financial and Actuarial Mathematics

Master Thesis (30EAP)

Supervisor Meelis Käärik

Tartu 2016

Contents

1 Introduction 4

2 Casinos and Playtech 5

3 Casino customer marketing 6

3.1 Attraction 6

3.2 Conversion 7

3.3 Retention 7

3.4 Reactivation 8

4 Multivariate Linear Regression 9

4.1 Correlation Matrix 9

4.2 Linear Model 9

4.3 Estimating parameters 10

4.4 Significance of the model 10

4.5 Significance of the models parameters 11

4.6 Model Fit 12

4.7 Multicollinearity 12

4.8 Residual analysis 13

4.8.1 Precision of the model 13

4.8.2 Model assumption check 13

4.8.3 Model diagnostics 14

4.8.4 Finding influentials 14

5 Data 16

5.1 Anonymity 16

5.2 Data obtainment 16

5.3 Player events 17

5.4 Bonuses 18

5.4.1 Cash bonuses 18

5.4.2 Pre-wager redeemable bonuses 18

5.4.3 Pre-wager Non-Redeemable Bonuses 19
5.4.4 After-wager Bonuses 19
5.4.5 Bonus triggering types 19

6 Analysis 21
6.1 Formulation of the problem 21
6.2 Analysis of dependent variable 21
6.2.1 Players without any significant gaming 24
6.2.2 Players with significant gaming 27
6.3 Analysis of bonuses 30
6.3.1 Trigger types 30
6.3.2 Wagering methods 33
6.3.3 Wagering coefficients 35
6.3.4 Other bonus attributes 36
6.3.5 Interplay of variables 38
6.3.6 Combination of variables 39

7 Conclusions 46

Appendices 49

A Hints 49

B Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks 51

Kasiino boonuste efektiivsuse mõõtmine kasutades lineaarset regressiooni

Lühikokkuvõte. Käesolev magistritöö annab ülevaate lineaarsest regressioonist ja regressiooni mudelite rakendamisest reaalsete kasiino mängijate rahalistele andmetele. Töös uuritakse mängijatele boonuste andmist ja sellega seostuvaid muutusi mängijate kahjumis. Eesmärgiks on leida kõige kasulikum boonusetüüp kasiinole. Meeles tuleb pidada, et järgnev magistritöö ei saa anda täiuslikku vastust sellele küsimusele, vaid see on üks tööriist, mida saavad kasutada kasiinode turundusosakonna töötajad.

Märksõnad: Regressioonanalüüs, finantsmatemaatika, lineaarsed mudelid, hasartmängud

CERCS teaduseriala: Statistika, operatsioonanalüüs, programmeerimine, finants- ja kindlustusmatemaatika (P160)

Measuring the effectiveness of casino bonuses using linear regression models

Abstract. The following master thesis gives an overview of the linear multivariate regression and the implementation of the models for casino players real monetary data. With these models the handing out of bonuses to players and the changes in players net losses are observed. The final result would be the most profitable bonus type for the casino. It is important to note, that this master thesis can't give a perfect answer to this question, but is another tool used by the casino's marketing department.

Keywords: Regression analysis, financial mathematics, linear models, gambling

CERCS research specialisation: Statistics, operation research, programming, actuarial mathematics (P160)

1 Introduction

In today's society, it is virtually impossible to be unaware of the amount of advertising in our lives. Advertising is just one of the incentives companies use to influence the tastes and spending habits of the consumer. With all such motivators, there are two big questions. Do these motivators have an effect and if they do, what kind of effect do they have?

One such usage of motivators is in the gambling industry, where one of the main tools casinos use are bonuses. Bonuses are small incentives made to both draw in new customers and keep the existing user base playing. Therefore, the types of bonuses given to players have diversified, with the aim of them being tailor-made for the player. These bonuses are a good example, since they contain numerical data on which to analyse the effectiveness and the differences of the motivators.

In the first part of the master thesis, we will give an introduction and overview into both Playtech and how the online casinos market their product to their customers. For this, the 4 phases of casino customer's life cycle are defined and the specific promotions are listed.

In the second part of the master thesis, we will give the overview of linear multivariate regression. This will include estimating the parameters, significance of the model and its parameters, the fit of model and residual analysis. The process in which these models are selected will be used on the data set.

In the third part of the thesis, we will look at the data that will be analysed in this thesis. We have two main data models. First of them will consist of the lifespan of players that registered with the chosen casino during the chosen period. The lifespan will consist of the signing up and monetary actions during the period. The second data set will consist of all of the bonuses that are in use during the time period. Alongside the data, we will give an overview of the process required for obtaining the data and some transformations needed to make it usable in the analysis.

In the fourth part of the thesis the analysis of the data with the models will be done. For this, the analysis gives an overview of the problem, the parameters used, the analysis part itself and the conclusions. The analysis is punctuated by tables and matrices used.

The appendix includes hints for quicker analysis/formatting in both R and TexStudio.

2 Casinos and Playtech

Since the beginning of recorded history, there have always been places where people have gotten their entertainment with games of chance. Today, we call them casinos and they are bigger than ever, mainly thanks to the ever-increasing online gambling market. Since the market is thriving, the competition has risen between casinos to increase and maintain their playerbases. One such example of such a company is Playtech, with whom this master thesis is written in collaboration with.

Playtech was founded in 1999 in Tartu, Estonia by entrepreneurs from the casino, software engineering and multimedia industries. Market analysis and product definition followed and in 2001 Playtech welcomed its first Casino licensee. Since then Playtech has grown by leaps and bounds adding more and more products to its portfolio including Live Casino, Bingo network, iPoker network, land-based offering, Videobet and Mobile Casino. Today Playtech is the world's leading supplier of online gaming software, having more than 3600 employees in 12 countries. Playtech as a software development company does not operate their own casinos, but rather rents their software to different operators.

Playtech offers an extensive range of games on their own platform, Playtech IMS. Besides games, Playtech IMS delivers ancillary services, which encompass the training and development needs of a gaming operator, including: online marketing, customer support, full CRM capabilities, fully-managed poker and bingo networks, sports betting, trading room services, hosting and disaster recovery services, and payment processing and advisory services management tools needed to interact and manage their players throughout their life cycle. Because of all of these services, the importance of analysing the effectiveness of different marketing strategies has become vitally important.

3 Casino customer marketing

To maintain a successful business, the casino has to use its marketing tools to both attract new customers and maintain old ones. For this, a casino player's life cycle is segmented into 4 different phases: **attraction, conversion, retention** and **reactivation**. For every phase, there are certain steps that the casino is following for its marketing campaigns: setting up expectations, coming up with a plan, monitoring the effectiveness of the actions and then refining the actions.

3.1 Attraction

In the attraction phase, the player is made aware of the casino and is persuaded to visit the website. This phase ends with the player registering with the casino.
To make the player aware of the casino, several methods are used:

- different kinds of advertisements (media, billboards, direct mail);
- using affiliate programs (an affiliate earns money by creating customers to the company with its own marketing efforts);
- using the casinos own players to do the promoting e.g. refer-a-friend program ("Get a friend to join and you both receive a bonus!"[1]);
- leads (possible player lists that are sold by different companies and then contacting them directly).

When the player decides to visit the web page, then they are directed to a **landing page**, which is a simple and attractive page different from a regular homepage. A landing page has a lot more information specifically tailored to the funnel, from which the player came from. It displays different bonuses that the player can receive when signing up and depositing money. This page is specifically made with the intention of getting the player to sign up and deposit. With affiliates, the page is specifically designed with the affiliate in mind, allowing for specialized and effective landing pages.

In the end, a player does not stick to one casino for a long period of time, which is why a licensee might have several brands. After a few weeks, the company starts pushing the player to switch brands, which will retain the player for a longer period of time.

[1] A more detailed view of the different bonuses in Playtech is given in chapter "Data"

3.2 Conversion

When the player has signed up, he has not yet made any profits for the casino; rather he has gained a sign-up bonus, which can turn into a loss in some cases. Therefore, the main goal of the conversion is to get the player to deposit some money and start playing. For many players this is natural, but some methods are made to both increase the percentage of players converted and to drive the deposit amounts up.

- First deposit bonus - the main tool in this scenario. Since most players don't deposit more than once, the idea is to make the first deposit amount as large as possible. Therefore, deposit multipliers are used. These bonuses will be the largest that the player can receive.
- Some brands give the option of adding your credit card number and the amount of the first deposit on registration.
- Pop-up messages in the clients reminding the player to deposit with links to the cashier.
- When the deposit has not happened in a certain amount of time after registration, the player will be contacted by every possible channel available and influenced to deposit.
- In more extreme cases, a more desirable bonus will be offered to the player.

3.3 Retention

In the beginning of the retention phase, the players have made a deposit and even played a few games. The purpose of phase is to keep the players playing more and longer, with the possibility of extra deposits. In this phase, the promotions themselves are less important than how they are presented. The message communicated to the player is that they can win and will win.

Almost every operator has an operation calendar schedule, which is used to having some kind of promotion every day. This calendar has two kinds of events: regular and seasonal.

- **Segmentation:** by this time, the player has visited the web page and the client for a couple of times, which means that a lot of information about them is known. Therefore, a lot more specialized views and promotions can be given. Valuable players will start receiving more expensive contact methods e.g. phone calls.

- Usage of **acquisition products**, which are used to lure players in, before starting to promote more profitable products.
- Pushing similar types of games to the player. For example, since a blackjack player might not be interested in roulette, different types of blackjack games are marketed.
- Small banners which include little games to play when waiting for larger games to finish.
- Players using multiple platforms generate more profit, therefore other platforms are promoted. For example, people playing in online casinos receive mobile promotions.
- Bonuses with opt-in conditions, which are used to segment the players even more. Not every opt-in condition is shown to all the players.
- Leaderboards to keep track of different statistics, with incentives to the leader.

3.4 Reactivation

The reactivation phase is for bringing back former casino players, who have stopped playing for some reason. This phase uses several ideas from the retention phase, but the key component is personalized contact with the player, offering bonuses for coming back.

4 Multivariate Linear Regression

This chapter is written using reference [6] as the basis. For a more detailed overview of this subject, references [2] and [10] are given.

4.1 Correlation Matrix

Before creating linear models, it is important to investigate the correlation between the variables. When there is no correlation between the variables, there is no point creating a linear model. For this a **correlation matrix** is used, which investigates the dependence between multiple variables at the same time. The correlation matrix of k random variables $X_1, ..., X_k$ is the $k \times k$ matrix where the correlation coefficient between X_i is and X_j, denoted by $corr(X_i, X_j)$.
When looking at a correlation matrix, the dictum "correlation does not imply causation", which states that correlation cannot be used to infer a causal relationship between the variables, needs to be remembered.

4.2 Linear Model

A **linear model** is a linear function of the parameters of a model. A typical linear model is written as

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon},$$

where $\mathbf{y}$ is a n-length vector of the dependent variable's observations $\mathbf{y} = (y_1, ..., y_n)^T$, k is the number of different explanatory variables (without including the free term), $\mathbf{X}$ is a $n \times p$ matrix of values of explanatory variables $(p = k + 1)$, $\boldsymbol{\beta}$ is a p-length parameter vector (including the free term) and $\boldsymbol{\varepsilon}$ is a n-length vector containing noise variables.
For an i-th dependent variable, the single model has the following form

$$y_i = \beta_0 + \beta_1 x_{1i} + +\beta_2 x_{2i} + ... + +\beta_k x_{ki} + \varepsilon_i.$$

The assumptions on the model are the following:

1. The errors $\boldsymbol{\varepsilon}$ don't have an offset $\mathbf{E}\boldsymbol{\varepsilon} = \mathbf{0}$.
2. *Homoscedasticity.* The errors $\boldsymbol{\varepsilon}$ have a constant variance $\mathbf{D}\boldsymbol{\varepsilon} = \boldsymbol{\sigma^2}$.
3. *Independence of observations.* For errors corresponding to different observations, $cov(\varepsilon_i, \varepsilon_j) = 0$ or $cov(\mathbf{y_i}, \mathbf{y_j}) = 0$, $(i \neq j)$.

4. *Normally distributed errors.* The errors ε follow a multivariate normal distribution, $\boldsymbol{\varepsilon} \sim N(\mathbf{0}, \sigma^2\mathbf{I})$. Equivalently,
$$\mathbf{y} \sim \mathcal{N}(\mathbf{X}\boldsymbol{\beta}, \sigma^{\mathbf{2}}\mathbf{I}).$$

5. *Linear dependence on the explanatory variables.* Follows directly from 1. and 3.
$$\mathbf{E}(\mathbf{y}) = \mathbf{X}\boldsymbol{\beta}.$$

4.3 Estimating parameters

For estimating the model's parameters, one of the most common approaches is to use the **method of least squares**. In this case, $\hat{\boldsymbol{\beta}}$ (the estimator for $\boldsymbol{\beta}$) is found by minimizing the sum of squares of errors $\boldsymbol{\varepsilon}$. In other words, the following equation is minimized with respect to $\boldsymbol{\beta}$:

$$\mathbf{SSE}(\boldsymbol{\beta}) := \sum_{\mathbf{i=1}}^{\mathbf{n}} \varepsilon_{\mathbf{i}}^{\mathbf{2}} = \boldsymbol{\varepsilon}^{\mathbf{T}}\boldsymbol{\varepsilon} = (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})^{\mathbf{T}}(\mathbf{y} - \mathbf{X}\boldsymbol{\beta}).$$

By solving this equation, we get the normal equation

$$\mathbf{X}^{\mathbf{T}}\mathbf{X}\boldsymbol{\beta} = \mathbf{X}^{\mathbf{T}}\mathbf{y}.$$

When the matrix $\mathbf{X}^{\mathbf{T}}\mathbf{X}$ in invertible, the previous equation has a solution

$$\hat{\boldsymbol{\beta}} := (\mathbf{X}^{\mathbf{T}}\mathbf{X})^{-\mathbf{1}}\mathbf{X}^{\mathbf{T}}\mathbf{y}.$$

From this we can find the prediction of the model $\hat{\mathbf{y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$.

4.4 Significance of the model

When creating a new linear model, we have to check if the model is statistically significant. The following pair of hypotheses has to be checked:

$$\begin{cases} H_0 : \boldsymbol{\beta} = 0, & \text{i.e. every parameter is equal to zero,} \\ & \text{ the model is not statistically significant.} \\ H_1 : \boldsymbol{\beta} \neq 0, & \text{i.e. at least one parameter is not zero,} \\ & \text{ the model is statistically significant.} \end{cases}$$

To check the significance of the model, the following sum of squares are defined:

- The sum of squares due to errors **SSE**, given in the previous section.
- The sum of squares due to regression $\mathbf{SSR} = (\hat{\mathbf{y}} - \bar{\mathbf{y}})^{\mathbf{T}}(\hat{\mathbf{y}} - \bar{\mathbf{y}})$, where $\bar{\mathbf{y}} = \mathbf{\frac{1}{n}\sum_{i=1}^{n} y_i}$.
- Total sum of squares of errors $\mathbf{SST} = (\mathbf{y} - \bar{\mathbf{y}})^{\mathbf{T}}(\mathbf{y} - \bar{\mathbf{y}})$. It is obvious that $\mathbf{SST} = \mathbf{SSR} + \mathbf{SSE}$.

These are used to define an F-statistic

$$F = \frac{\mathbf{SSE}/p}{\mathbf{SST}/(n-k)},$$

where F has an F distribution $F \sim F_{p,n-k}$ when the errors have a normal distribution. An F-test is any statistical test in which the test statistic has an F-distribution under the null hypothesis. The null hypothesis is rejected if the F calculated from the data is greater than the critical value of the F-distribution for some desired false-rejection probability (e.g. 0.05).

4.5 Significance of the models parameters

Besides the whole model, we have to look if a single parameter is statistically significant. The pair of hypotheses needed for this is:

$$\begin{cases} H_0 : \beta_i = 0, & \text{i.e. the parameter is not statistically significant.} \\ H_1 : \beta_i \neq 0, & \text{i.e. the parameter is statistically significant.} \end{cases}$$

This hypothesis is checked using a $\mathbf{t}$-statistic. A $\mathbf{t}$-statistic for an estimator $\hat{\boldsymbol{\beta}}$ is defined as

$$\mathbf{t}_{\hat{\boldsymbol{\beta}}} = \frac{\hat{\boldsymbol{\beta}}}{\hat{\sigma}},$$

where $\hat{\sigma}$ is the estimator of the standard deviation of $\hat{\boldsymbol{\beta}}$.
In case of H_0, the $\mathbf{t}$-statistic has a Student's $\mathbf{t}$-distribution. The null hypothesis is rejected if the $\mathbf{t}_{\hat{\boldsymbol{\beta}}}$ calculated from the data is greater than the critical value of the Student's $\mathbf{t}$-distribution for some desired false-rejection probability (e.g. 0.05). When an argument is not statistically significant, it needs to be removed from the model.

4.6 Model Fit

When a suitable model has been found, the next step is to check how well the data fits to the model. The main barometer for this is **coefficient of determination**, denoted by R^2. R^2 is defined as $R^2 = \frac{\mathbf{SSR}}{\mathbf{SST}} = 1 - \frac{\mathbf{SSE}}{\mathbf{SST}}$. R^2 is often interpreted as the proportion of dependent variation "explained" by the regressors in the model. Thus, $R^2 = 1$ indicates that the fitted model explains all variability in $\mathbf{y}$, while $R^2 = 0$ indicates no "linear" relationship. An interior value such as $R^2 = 0.7$ may be interpreted as follows: "70% of the variance in the dependent variable can be explained by the explanatory variables in the model. The remaining 30% can be attributed to unknown, lurking variables or inherent variability."

4.7 Multicollinearity

The following subsection, besides the aforementioned sources, uses [9].

When two or more explanatory variables have a strong dependence, meaning one can be linearly predicted from the other with a non-trivial degree of accuracy, then this is called **multicollinearity**. In such a case, it is not possible to find $\hat{\boldsymbol{\beta}}$ (since it's not possible to find an inverse matrix $(\mathbf{X^TX})^{\mathbf{-1}}$), therefore the estimations are imprecise. For describing multicollinearity, correlation coefficient r, which measures the linear correlation between two variables, is used. Multicollinearity is implied by:

- large correlation coefficients ($r > 0.95$) in the correlation matrix;
- correlation between the different explanatory variables is bigger than the correlation between the explanatory variables and the dependent variable.

The two most common ways to measure multicollinearity are:

- **Tolerance** TOL_i shows how large part of a variable's variability is not described by the other variables. $TOL_i = 1 - R_i^2$, where R_i^2 is the model's coefficient of determination, where the i-th variable is defined by the others. The smaller the tolerance, the more the variable depends on the others.
- **Variance inflation factor** VIF provides an index that measures how much the variance of an estimated regression coefficient is increased because of collinearity, $VIF = 1/TOL$.

Multicollinearity is considered large, if $TOL < 0.10$ or $VIF > 10$.
The following methods are used to alleviate multicollinearity:

- Dropping one of the variables. An explanatory variable may be dropped to produce a model with significant coefficients. However, this results in lost information, so this is not a preferred solution.

- Increasing the size of the data set. This is the preferred solution. More data can produce more precise parameter estimates (with lower standard errors). In practice, this is quite complicated to implement.

- *Ridge regression.* To find $(\mathbf{X^T X})^{-\mathbf{1}}$, a small constant $\alpha\mathbf{I}$ is added to the formula $(\mathbf{X^T X} + \alpha\mathbf{I})\hat{\boldsymbol{\beta}} = \mathbf{X^T y}$. This enables a direct numerical solution $\hat{\boldsymbol{\beta}} = (\mathbf{X^T X} + \alpha\mathbf{I})^{-\mathbf{1}}\mathbf{X^T y}$.

4.8 Residual analysis

The **residuals** of the model is defined as the difference between the measurements and the prediction of the model

$$\hat{\boldsymbol{\varepsilon}} = \mathbf{y} - \hat{\mathbf{y}} = \mathbf{y} - \mathbf{X}\hat{\boldsymbol{\beta}}.$$

Residuals are analyzed for several reasons.

4.8.1 Precision of the model

To predict the precision of the model we have to look at the variance of the errors:

$$\mathbf{MSE} = \frac{\hat{\boldsymbol{\varepsilon}}^T\hat{\boldsymbol{\varepsilon}}}{n-k} = \frac{\mathbf{SSE}(\hat{\boldsymbol{\beta}})}{n-k} = \frac{\hat{\sigma_\varepsilon^2}}{n-k}$$

where $n-k$ is the **degrees of freedom** for errors.
The precision of the model is the square root of the mean square error, called **root-mean-squared error (RMSE)**.

4.8.2 Model assumption check

- *Normally distributed errors.* The most common ways to check this are Shapiro-Wilk and Kolmogorov-Smirnov test. A sample of the dependent variables is taken and for both tests the following hypotheses pair is used:

$$\begin{cases} H_0 : \text{The sample comes from a normally distributed population} \\ H_1 : \text{The sample does not come from a normally distributed population.} \end{cases}$$

The null hypothesis is rejected if the p-value is less than the chosen α level (e.g. 0.05).

- *Homoscedasticity of errors.* White test and Breusch/Pagan test are used
- *Independence of observations.* Durbin-Watson test is used, which calculates the Durbin-Watson statistic $W \in [0 : 4]$. If $W = 2$, then there is no autocorrelation. When $W > 2$, there is a positive autocorrelation, when $W < 2$, then a negative autocorrelation.

4.8.3 Model diagnostics

The goal of the model diagnostic is to find the *outliers* of the model. There are several types of outliers: regular outliers, leverages and influentials. Each one has their own methods for finding.

- For measuring *outliers* standardized residuals and/or studentized residuals are used. When a standardized residual has been found, its value is divided with its standard deviation. If the value of the standardized residual is greater than 2, then the corresponding observation is an outlier.
- *Leverage* score for the i-th data unit is defined as h_{ii}, where $\mathbf{H} = (h_{ij}) = \mathbf{X}(\mathbf{X^T X})^{\mathbf{-1}}\mathbf{X^T}$ is the hat matrix. When $h_{ii} > \frac{2(p+1)}{n}$, then the observation is a leverage.

4.8.4 Finding influentials

There are several ways in which outliers can have influence on the model:

- **Influence on model's parameters**
  The most common statistic to measuring this is Cook's D statistic. Cook's D statistic is defined as

$$D_i = \frac{\boldsymbol{\varepsilon_i}}{m\mathbf{MSE}}\frac{h_{ii}}{(1-h_{ii})^2}.$$

  When $D_i > F_{0.5;p;n-p}$ then the i-th outlier influences the models parameters. ($F_{0.5;p;n-p}$ is the median of the F-statistic)

- **Influence on a specific parameter**. For this, $DFBETAS$ (standardized difference of the beta) is used. When $DFBETAS_{ji} > \frac{2}{\sqrt{n}}$, then the i observation has an effect on the j-th explanatory variable.

- **Influence on the estimation**. Predicted residuals are used. Predicted residuals are values that are the differences between the actual value of the model and the value of the model, when the i-th observation is removed. Predicted residuals are usually larger than normal residuals. When the predicted residual is negative, the model overvalues the dependent variable (without taking in to account the i-th variable). When positive, then the model undervalues. Models that are over-fitted tend to give small residuals for observations included in the model, but large residuals for observations that are excluded.

NOTE! For all of the models shown in the analysis chapter, it is important to note that even though finding outliers and removing them from the data set would improve the model fit, then this can have a significant effect on the financial calculations. When the outlier happens because of incorrectly inserted data,it needs to be removed. If the outlier happens because of the players unnatural behaviour, it should not be removed. In the analysis, models with and without the outliers are given.

5 Data

5.1 Anonymity

Data used in this thesis is based on the actual user activities in the real-world gaming operations. However, the data cannot be connected in any way with the actual users or with Playtech's clients. For this there are several tasks in both the obtainment and preparation phases.

- Anything that identifies users or clients by name or code is separated from the data. Instead, salted hashes are used as identifiers.
- Manipulation of the monetary figures, converting them to virtual currency using random currency rate fixed during the data export.
- Keycoding the data dimension names and properties using alphabetic keys. The dependent variable is noted as *netloss* in case of net loss after receiving first bonus. The explanatory variables are given either one or more characters. The mapping of the keys to variable names is only known to the relevant people and is not part of the dataset.

5.2 Data obtainment

Since this thesis should have a big practical value, it is important for future readers to understand how to receive the necessary confidential data from the company. Therefore, the following steps are given to both understand and hasten the future requests.

1. The person writes a request for the security team detailing the data needed, the structure of the data exported and also the planned usage of such data.
2. For receiving the correct data, the requester needs to write a correct piece of code of a selected database querying language. If the requester is not an expert in said code, some research from the internet or help from the experts in the company is needed. Since the company experts are busy with their own tasks, the requester cannot ask them to write the code, but them reviewing and improving the code is always a possibility.

3. The person uses the code to get data from some kind of database, where the numbers aren't taken from real life. A testing database is perfect for such task. Since the code most likely has to be improved or rewritten several times, using this for the real database is a time consuming task, which can have a performance impact. When using a test database, the person can both change some of the parameters, which limit the amount of data received and not have an effect on the real systems.

4. User requests access or gives the code to a person with access to real life databases and they download the necessary data. If the code is given to another person, it is important for them to understand, how the data should be compiled (for example the character, which is used as a separator).

5.3 Player events

The data used will describe a lifespan of numerous players that have registered in a Playtech's licensee's casino during one year. The data includes events that describe the player's behaviour (signing up, different payments and gaming) and promotional offers to Real winthe players (getting and redeeming bonuses).

- **Signup** - The first event for any player is signing up for a casino.
- **Deposit** - Event is created when a player deposits money into their account. For us the important data is the amount deposited.
- **Withdraw** - Event is created when a player withdraws money from their account. For us the important data is the amount withdrawn.
- **Get bonus** - Event is created when the player receives a bonus. For us the important data is the bonus amount received and the bonus details
- **Real bet** - Event is created when the player bets with real money at least once in the hour. For us the important data is the sum of (real bets - cancelled real bets) per hour.
- **Bonus bet** - Event is created when the player bets with bonus money at least once in the hour. For us the important data is the sum of (bonus bets - cancelled bonus bets) per hour.
- - Event is created when the player wins with real money bets at least once per hour.For us the important data is the real wins per hour.

- **Bonus win** - Event is created when the player wins with bonus money bets at least once per hour. For us the important data is the bonus wins per hour.
- **Redeem bonus** - Event is created when a player completes the bonus requirements and bonus money is converted to real money. For us the important data is the bonus amount redeemed and the bonus details.

The **timestamp** of the event is registered for every event. This measures the amount of hours passed in the player life cycle e.g. from the signup of the player. This is done to convert every player's life cycle into a single timeframe. In case of this thesis, the **timestamp** is converted so that the time starts from the hour the player receives their first bonus.
Regarding these events, the most important amount in use is the **Net loss**, which shows the real monetary loss of the player from the casino's standpoint. The formula for this is:

$$Netloss = Realbet - Realwin - Redeembonus$$

5.4 Bonuses

Playtech offers their licensees the possibility to create their own custom bonuses with variables that best suit their needs. The amount of said variables that can be changed are too many to list in this thesis. All of the bonuses can be divided into large subcategories.

5.4.1 Cash bonuses

Cash bonus is a bonus, that when accepted by a player, is added to the player's real balance. The player can then either withdraw the funds or use it for gaming. This bonus is also called a pre-wager immediately redeemable bonus.

5.4.2 Pre-wager redeemable bonuses

A redeemable bonus, when accepted by the player, is added to the player's bonus balance, which can be used for wagering on casino games only. Once the wagering requirements (automatically calculated by the system) are met, the remaining bonus amount (including linked pending winnings) is moved to the player's real balance. Wagering requirements can be set as:

- multiples of a deposit;
- multiples of the bonus issued;
- sum of the multiples of both the deposit and bonus amount;

Wagering coefficients can be set per game (e.g. different coefficients for blackjack and roulette) to determine the contribution of the wagering on the specific game to the wagering requirements of the active bonus.

5.4.3 Pre-wager Non-Redeemable Bonuses

A non-redeemable bonus is added to the player's bonus balance, which can be used for wagering on casino games only. Only the winnings are redeemed. The bonus amount can never be cashed out. Wagering requirements can be set only if pending winnings are used.

5.4.4 After-wager Bonuses

After-wager bonuses are bonuses with wagering requirements that are issued to the player's pending bonus balance, which cannot be used by the player. T When the wagering requirements are met, the bonus amount is moved to the player's real balance, and can then be used for wagering or be cashed out.

5.4.5 Bonus triggering types

Bonuses can be triggered by many ways. Two main types are:

- bonuses that are given manually to the player;
- bonuses that are given automatically by a player action.

Automatic bonus can be triggered by several player action types:

- **Signup** - Bonus is given when the player registers.
- **Deposit** - Bonus is given when the player deposits a certain amount of money to his balance.
- **Promotional code** - Bonus is given when the player inserts a promotional code.

- **Custom event** - Bonus is triggered by a custom event that is created by a licensee.
- **Buy bonus** - Player "buys a bonus", e.g. by trading real money for a larger amount of bonus money.
- **Bonus completion** - After completing the requirements of a previous bonus, the player is given another bonus.
- **Campaign** - Campaign Manager segments the players and decides on which players receive which bonuses.
- **Gameplay** - Player is given a bonus after certain pre-specified events in gameplay.

6 Analysis

6.1 Formulation of the problem

It is very important to see if receiving a bonus would increase the net loss of the player. Therefore, we need to compare the player net loss on the hour where he received his first bonus with the net loss after a certain period of time. For our check we will have two periods, firstly 5 days and secondly 30 days.

Dependent variable *netloss*: Net loss player received during the period after receiving the first bonus.
Therefore the problem can be written as: *Dependence of net loss from playtime on set number of days after receiving the first bonus.*

NOTE! Since for statistical tables, some data aggregation is natural (otherwise the data sets being used will be increased tens of times), then this creates a minor visibility problem in the hour where the first bonus is received (there is no visibility for what amounts were played before the first bonus and which after).

6.2 Analysis of dependent variable

Before creating models, an analysis of the dependent variable is needed. Checks for both Shapiro-Wilk and Kolmogorov-Smirnov test are used. For both timeframes, the tests show that the sample does not come from a normally distributed population. Here we have to note that for larger data sets, both Shapiro-Wilk and Kolmogorov-Smirnov tests have a tendency to reject the null hypothesis. Therefore, a look at the histogram of the *netloss* is needed.

Figure 1: Histogram of *netloss* for 5 days

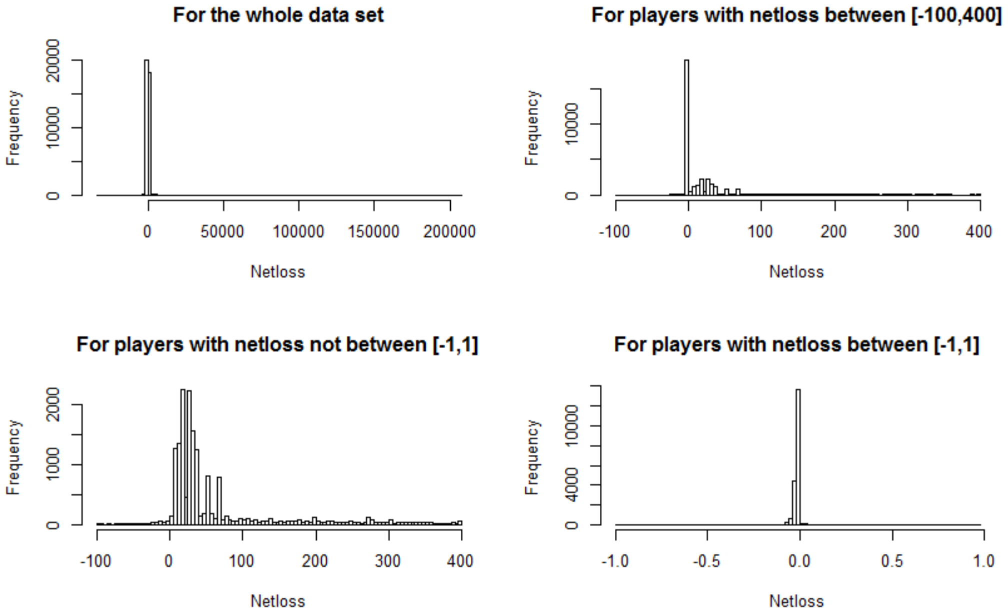


Figure 2: Histogram of *netloss* for 30 days

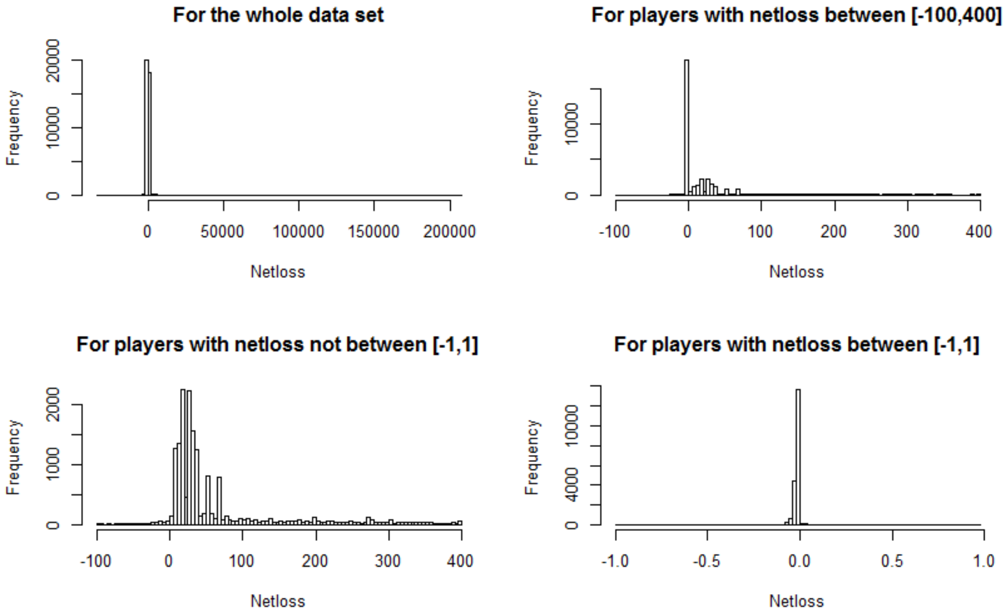


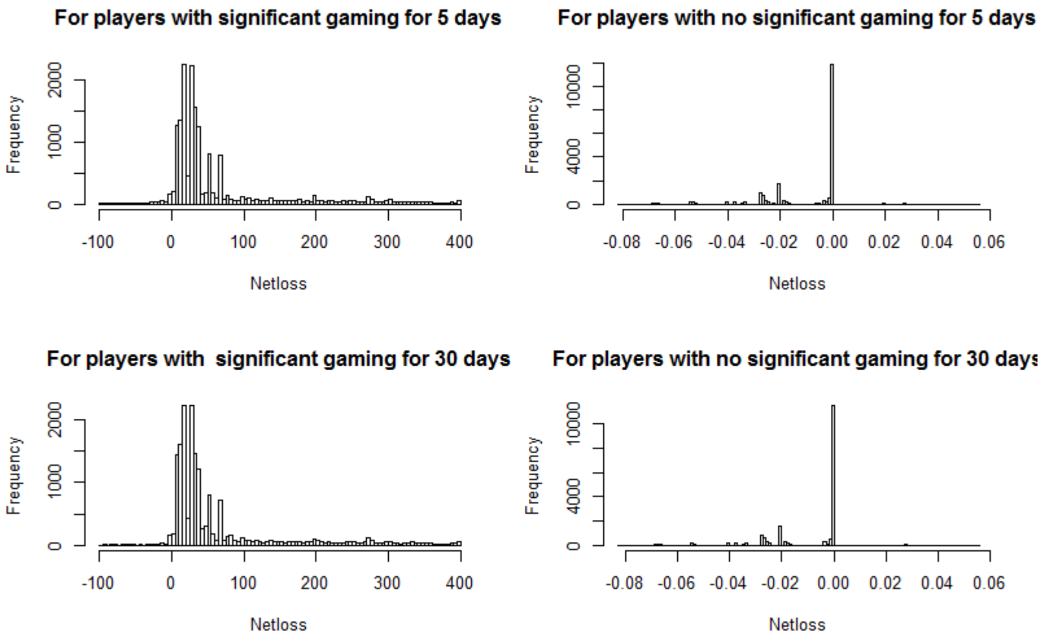
Figure 1 (for 5 days) and Figure 2 (for 30 days) give us 4 different histograms on the same data set with different subsets. From the upper two histograms, a large subset is seen that has a *netloss* near 0.

When looking at this group, a large part of the subset are players who have a *netloss* of 0. These are players, who have received a sign-up bonus for registering an account, but have not played any games. These players will be removed from the significant players and analysed separately.

Another significant group here are players who have received a very minor bonus, which has then been redeemed. The presumption here is that this is some kind of licensee trick. These players will also be removed from the significant players and analysed separately.

This leaves a very minor group that has done significant gaming, but their luck has let their *netloss* stay near 0. There players will not be removed from the significant players.

Figure 3: Histogram of *netloss* for final subsets



Therefore, the significant players dataset consists of players, whose *netloss* during the period is not between $[-1; 1]$ or whose *netloss* is between $[-1; 1]$, but they have betted more than 0.1. The rest of the players go to the set without any significant gaming. The histograms for those data sets are given in Figure 3.

6.2.1 Players without any significant gaming

Before going to the bonus type analysis, we must take a look at the variables that have the biggest effect on *netloss*. With this we are validating the data set. When the formula for *netloss* includes the variables that define it, the R^2 should be 1.

Table 1: Correlation matrix after 5 days of receiving the first bonus for players without any significant gaming

	netloss	*b*	*c*	*d*	*e*	*f*	*g*	*h*
netloss	1.00	0.05	0.04	-0.19	-0.94	0.15	-0.14	0.02
b	0.05	1.00	0.05	-0.33	0.07	0.00	-0.17	0.09
c	0.04	0.05	1.00	-0.29	0.01	-0.01	-0.10	0.30
d	-0.19	-0.33	-0.29	1.00	-0.16	0.04	0.50	-0.13
e	-0.94	0.07	0.01	-0.16	1.00	-0.17	-0.03	0.00
f	0.15	0.00	-0.01	0.04	-0.17	1.00	0.02	0.00
g	-0.14	-0.17	-0.10	0.50	-0.03	0.02	1.00	-0.33
h	0.02	0.09	0.30	-0.13	0.00	0.00	-0.33	1.00

In case of the timeframe of 5 days, the correlation matrix in Table 1 gives that the e variable has a correlation which is significantly stronger than others, therefore it should be the main variable. When taking into account the correlations between the different explanatory variables, d and g are added to the model. Since g is not statistically significant in this model, therefore only e and d are selected for the final model.

Table 2: Model for *netloss* after 5 days of receiving the first bonus for players without any significant gaming

	Model 1	Model 2
(Intercept)	-0.0000	0.0003^*
	(0.0000)	(0.0000)
e	-0.9976^*	-0.9416^*
	(0.0004)	(0.0026)
d	-0.9503^*	
	(0.0012)	
N	19127	19127
R^2	0.9964	0.8769
adj. R^2	0.9964	0.8769
Resid. sd	0.0009	0.0054

Standard errors in parentheses

* indicates significance at $p < 0.05$

Table 2 shows that Model 1 almost perfectly describes the variability in the data. For comparison, Model 2 shows that a large part (close to 88 %) of the model is defined by the e variable. For a better fit in Model 1, a check for outliers is done. All of the outliers found seem to look like normal players, therefore the decision is made to not remove those. The intercept in this case is very close to zero, which shows that besides e and d, barely any actions was done with this data set. The final model is:

$$netloss = -0.9976e - 0.9503d + \varepsilon_1.$$

Table 3: Correlation matrix after 30 days of receiving the first bonus for players without any significant gaming

	netloss	*b*	*c*	*d*	*e*	*f*	*g*	*h*
netloss	1.00	-0.28	0.04	-0.19	-0.94	0.13	-0.14	0.03
b	-0.28	1.00	0.04	-0.17	0.34	-0.06	-0.07	0.06
c	0.04	0.04	1.00	-0.33	0.01	-0.01	-0.19	0.44
d	-0.19	-0.17	-0.33	1.00	-0.16	0.04	0.50	-0.18
e	-0.94	0.34	0.01	-0.16	1.00	-0.15	-0.04	0.01
f	0.13	-0.06	-0.01	0.04	-0.15	1.00	0.02	0.00
g	-0.14	-0.07	-0.19	0.50	-0.04	0.02	1.00	-0.43
h	0.03	0.06	0.44	-0.18	0.01	0.00	-0.43	1.00

In case of the timeframe of 30 days, the correlation matrix in Table 3 again shows that the e has has an almost perfect decreasing linear relationship with *netloss*. b, d and f are also selected for the model, but b and f are not statistically significant.

Table 4: Model for *netloss* after 30 days of receiving the first bonus for players without any significant gaming

	Model 3	Model 3a
(Intercept)	-0.0000	0.0002^*
	(0.0000)	(0.0000)
e	-0.9966^*	-0.9417^*
	(0.0005)	(0.0026)
d	-0.9323^*	
	(0.0014)	
N	17918	17918
R^2	0.9952	0.8796
adj. R^2	0.9952	0.8796
Resid. sd	0.0010	0.0052

Standard errors in parentheses

* indicates significance at $p < 0.05$

From Table 4, Model 3 describes almost perfectly the variability in the data. Model 3a shows that again, a large part (close to 88 %) of the model is defined by the e variable. A check for outliers reveals none which should be removed (since the data set includes players who have not done any significant playing, it is normal to not find any outliers in the data). The final model for this case is:

$$netloss = -0.9966e - 0.9323d + \varepsilon_1.$$

For these players, the time after receiving the bonus does not not matter, since their playtime is very short. They have a minor effect on the main profit/loss for the company and sometimes are created because of licensee tricks. Therefore, no further research for this data set will be done.

6.2.2 Players with significant gaming

For this data set, the presence of gaming should give a lot more extra weight to more explanatory variables, therefore the correlation between *netloss* and other variables should be larger than in the data set for non-significant players.

Table 5: Correlation matrix after 5 days of receiving the first bonus for main data set

	netloss	*b*	*c*	*d*	*e*	*f*	*g*	*h*
netloss	1.00	0.11	0.66	-0.70	-0.01	0.76	-0.33	0.30
b	0.11	1.00	0.25	-0.24	0.06	0.20	-0.19	0.19
c	0.66	0.25	1.00	-1.00	0.01	0.66	-0.67	0.66
d	-0.70	-0.24	-1.00	1.00	-0.01	-0.69	0.66	-0.65
e	-0.01	0.06	0.01	-0.01	1.00	0.02	-0.02	0.02
f	0.76	0.20	0.66	-0.69	0.02	1.00	-0.36	0.34
g	-0.33	-0.19	-0.67	0.66	-0.02	-0.36	1.00	-1.00
h	0.30	0.19	0.66	-0.65	0.02	0.34	-1.00	1.00

In case of the timeframe of 5 days, the correlation matrix in Table 5 shows that for the model the useful variables are c, d and f. Multicollinearity of c and d requires that the model can have only one or the other.

Table 6: Model for *netloss* after 5 days of receiving the first bonus for main data set

	Model 4	Model 5	Model 6
(Intercept)	-142.27^{*}	-137.50^{*}	-71.04^{*}
	(12.04)	(11.76)	(9.7)
f	1.90^{*}	1.74^{*}	1.58^{*}
	(0.02)	(0.02)	(0.01)
c	0.02^{*}		
	(0.0005)		
d		-0.03^{*}	-0.01^{*}
		(0.0004)	(0.0004)
N	19211	19211	19207
R^2	0.62	0.64	0.42
adj. R^2	0.62	0.64	0.42
Resid. sd	1640.87	1602.01	1306.59

Standard errors in parentheses

* indicates significance at $p < 0.05$

Table 6 shows that Model 4 and Model 5 are quite similar in the structure and the fit of the model, with Model 5 having a fit that is slightly better. Therefore, we select Model 5 for an outlier check. 4 outliers were found and removed, with the final Model 6 being:

$$netloss = 1.58f - 0.01d - 71.04 + \varepsilon_1.$$

The 4 outliers were all players whose *netloss* was a lot larger than the other players, being larger than 100 000 for all of these players. These are players who played and lost a lot. They received numerous bonuses, but their luck was so bad, that they barely redeemed any of the winnings or bonuses and they lost all of the bonus amounts when playing.

Table 7: Correlation matrix after 30 days of receiving the first bonus for main data set

	netloss	*b*	*c*	*d*	*e*	*f*	*g*	*h*
netloss	1.00	0.06	0.83	-0.85	0.03	0.59	-0.58	0.61
b	0.06	1.00	0.13	-0.13	0.06	0.19	-0.15	0.15
c	0.83	0.13	1.00	-1.00	0.10	0.78	-0.89	0.90
d	-0.85	-0.13	-1.00	1.00	-0.10	-0.77	0.87	-0.89
e	0.03	0.06	0.10	-0.10	1.00	0.12	-0.34	0.33
f	0.59	0.19	0.78	-0.77	0.12	1.00	-0.79	0.79
g	-0.58	-0.15	-0.89	0.87	-0.34	-0.79	1.00	-1.00
h	0.61	0.15	0.90	-0.89	0.33	0.79	-1.00	0.61

In case of the timeframe of 30 days, the correlation matrix in Table 7 shows many variables with significant correlation. However, since there are several pairs of variables with perfect decreasing linear correlation, then the only certain variable chosen is f. g and h have a strong multicollinearity with each other, so do c and d, therefore only h is selected for the models and even eliminated from the final model. c and d are reviewed with separate models. This will give insight on which one of c or d should be selected for the final model.

Table 8: Model for *netloss* after 30 days of receiving the first bonus for main data set

	Model 7	Model 8	Model 9	Model 10
(Intercept)	100.21^*	88.22	117.50	-58.67^*
	(48.77)	(54.29)	(60.35)	(16.78)
f	0.22^*	0.35^*	-1.15^*	1.20^*
	(0.04)	(0.04)	(0.04)	(0.01)
d	-0.10^*		-0.07^*	-0.02^*
	(0.0005)		(0.0004)	(0.0003)
h	-0.26^*	-0.27^*		
	(0.0025)	(0.0029)		
c		0.11^*		
		(0.0006)		
N	20420	20420	20420	20416
R^2	0.83	0.78	0.73	0.65
adj. R^2	0.83	0.78	0.73	0.65
Resid. sd	6917.26	7699.57	8559.41	2373.26

Standard errors in parentheses

* indicates significance at $p < 0.05$

From Table 8, Model 7 and Model 8 are quite similar to each other, but Model 7 has a slightly better fit, therefore c is chosen for the later models. Because variable h has a danger of high multicollinearity, it has been also been removed. While checking for outliers in Model 9, 4 of them were found and removed, therefore Model 10 would be the final model:

$$netloss = -1.20f - 0.02d - 58.67 + \varepsilon_1.$$

As for the smaller timeframe, the 4 outliers were all players whose *netloss* was in the extremes, being larger than 300 000 for all of these players and smaller than -150 000 for one player. They were either very unlucky or unnaturally lucky. The lucky player would be reviewed by the casino.

For the bonuses themselves, the unlucky players received numerous bonuses, however their luck was so bad, that they barely redeemed any of the winnings or bonuses and they lost all of the bonus amounts when playing. The lucky player received several bonuses as well, but compared to the unlucky players, he mainly played with real money, negating the bonus money.

The models have a better fit for the longer timeframe, which again is to be expected. There is more gameplay data for all players in 30 days, which makes it both easier to guess their future values and the larger exceptions in player *netloss* (like jackpot winnings) have a smaller effect.

To conclude, for the significant players subset, the important variable for different bonus types is f and we can disregard e (given that for both models, e was not important, even though it was one of the variables which calculated $netloss$), therefore for bonus type analysis we are looking at the given bonuses to a player.

6.3 Analysis of bonuses

6.3.1 Trigger types

For the bonus analysis, we first look at bonus trigger types. In this data set certain bonus trigger types are not used, therefore those can be removed from the correlation matrix immediately.

Table 9: Correlation matrix after 5 days of receiving the first bonus for triggers

	$netloss$	i	j	k	l	m	n
$netloss$	1.00	-0.05	0.20	0.00	0.15	0.09	0.19
i	-0.05	1.00	-0.29	-0.01	-0.08	-0.42	-0.02
j	0.20	-0.29	1.00	-0.01	0.27	0.43	0.03
k	0.00	-0.01	-0.01	1.00	-0.01	-0.01	0.00
l	0.15	-0.08	0.27	-0.01	1.00	0.09	-0.01
m	0.09	-0.42	0.43	-0.01	0.09	1.00	0.01
n	0.19	-0.02	0.03	0.00	-0.01	0.01	1.00

For 5 days, the correlation matrix in Table 9 gives that j, n and l are the variables with correlations worth looking at.

Table 10: Model for *netloss* after 5 days of receiving the first bonus by triggers

	Model 11	Model 12
(Intercept)	-124.93^{*}	-109.37^{*}
	(21.44)	(13.52)
j	291.18^{*}	239.92^{*}
	(12.89)	(8.13)
n	6477.96^{*}	4881.28^{*}
	(239.31)	(153.59)
l	220.40^{*}	209.58^{*}
	(14.39)	(9.07)
N	19211	19207
R^2	0.08	0.13
adj. R^2	0.08	0.13
Resid. sd	2542.44	1602.86

Standard errors in parentheses

* indicates significance at $p < 0.05$

From Table 10, Model 11 gives the model for the whole dataset and Model 11 for the data set with 4 outliers removed (the same outliers as in the dependent variable analysis). For both models, the assumptions are proven. j and n both have a large change in their parameters when removing the outliers. The final model in this case would be Model 12:

$$netloss = 239.92j + 4881.28n + 209.58l - 109.37 + \varepsilon_1.$$

The bonuses with variable n are hugely profitable compared to other bonuses, which is to be expected, because these are bonuses that reward longer and loyal customers and are triggered by actions that cannot be done by first time players. So what might seem like rewards for the players are actually incentives for them to continue playing more often. The intercept shows that players, who did not receive such bonuses, have a negative *netloss*. Therefore, players, who have not accepted or not received such bonus types, generate a loss to the casino.

The bonuses with trigger type l have a very minor effect compared to other types, when looking at the change when removing the outlier. This confirms the expectations about these bonuses, because players who play a lot, should not be receiving these. Because these bonuses are triggered by players with lower playtime, the fact that they are so profitable shows that the people creating such bonuses know what they are doing.

Table 11: Correlation matrix after 30 days of receiving the first bonus for triggers

	netloss	*i*	*j*	*k*	*l*	*m*	*n*
netloss	1.00	-0.02	0.16	0.00	0.07	0.02	0.05
i	-0.02	1.00	-0.28	-0.01	-0.10	-0.17	-0.02
j	0.16	-0.28	1.00	0.00	0.41	0.28	0.04
k	0.00	-0.01	0.00	1.00	0.01	0.01	0.00
l	0.07	-0.10	0.41	0.01	1.00	0.10	0.00
m	0.02	-0.17	0.28	0.01	0.10	1.00	0.01
n	0.05	-0.02	0.04	0.00	0.00	0.01	1.00

For 30 days, the correlation matrix in Table 11 shows that the only variable with any sort of significant correlation is j. Therefore, the model in this case will be a simple one.

Table 12: Model for *netloss* after 30 days of receiving the first bonus by triggers

	Model 13	Model 14
(Intercept)	-676.88^*	-64.29^*
	(125.13)	(29.86)
j	1618.22^*	620.45^*
	(68.22)	(16.38)
N	20420	20416
R^2	0.03	0.07
adj. R^2	0.03	0.07
Resid. sd	16317.75	3890.51

Standard errors in parentheses

* indicates significance at $p < 0.05$

From Table 12, Model 13 gives the model for the whole dataset and Model 14 for the data set with 4 outliers removed (the same ones as in the dependent variable analysis). For both models, the assumptions are proven. The final model in this case would be Model 14:

$$netloss = 620.45j - 64.29 + \varepsilon_1.$$

For a longer period of time, the j and the intercept variables parameters change significantly when removing the outliers, suggesting the outliers received these kind of bonuses numerous times.

6.3.2 Wagering methods

The wagering methods are looked upon next. Some wagering methods are used quite often, some barely at all (two of the wagering methods are used less than 10 times during the whole data set).

Table 13: Correlation matrix after 5 days of receiving the first bonus for wagering methods

	netloss	*o*	*p*	*r*	*s*	*t*	*u*
netloss	1.00	0.06	0.16	0.00	0.13	0.02	0.00
o	0.06	1.00	0.37	-0.01	-0.37	0.19	0.02
p	0.16	0.37	1.00	0.00	-0.12	0.05	0.02
r	0.00	-0.01	0.00	1.00	0.00	-0.01	0.00
s	0.13	-0.37	-0.12	0.00	1.00	-0.12	-0.01
t	0.02	0.19	0.05	-0.01	-0.12	1.00	-0.01
u	0.00	0.02	0.02	0.00	-0.01	-0.01	1.00

For 5 days, the correlation matrix in Table 13 gives that p and s seem to be the only variables to have any sort of meaningful correlation.

Table 14: Model for *netloss* after 5 days of receiving the first bonus for wagering methods

	Model 15	Model 16
(Intercept)	-129.30^*	-113.85^*
	(22.32)	(14.15)
p	307.97^*	271.50^*
	(12.11)	(7.68)
s	248.93^*	218.19^*
	(11.52)	(7.31)
N	19211	19207
R^2	0.05	0.09
adj. R^2	0.05	0.09
Resid. sd	2590.77	1642.26

Standard errors in parentheses

* indicates significance at $p < 0.05$

From Table 14, Model 15 gives the model for the whole dataset and Model 16 for the data set with the same 4 outliers removed. For both models, the assumptions are proven. The final model in this case would be Model 16:

$$netloss = 271.5p + 218.19s - 113.85 + \varepsilon_1.$$

This model is logical since the bonus amounts given for the bonus type p are twice as large as the bonus types for amount s, but considering that the parameter is not twice as large, then the bonuses with wagering method s are less risky for the casino.

Table 15: Correlation matrix after 30 days of receiving the first bonus for wagering methods

	netloss	*o*	*p*	*r*	*s*	*t*	*u*	
netloss	1.00	0.04	0.11	0.00	0.00	0.04	0.00	0.01
o	0.04	1.00	0.48	0.00	-0.01	-0.18	0.19	0.17
p	0.11	0.48	1.00	0.00	0.00	0.01	0.08	0.14
q	0.00	0.00	0.00	1.00	0.00	0.01	-0.01	0.00
r	0.00	-0.01	0.00	0.00	1.00	0.00	-0.01	0.00
s	0.04	-0.18	0.01	0.01	0.00	1.00	-0.06	-0.01
t	0.00	0.19	0.08	-0.01	-0.01	-0.06	1.00	-0.01
u	0.01	0.17	0.14	0.00	0.00	-0.01	-0.01	1.00

For 30 days, the correlation matrix in Table 15 shows that the only variable with any sort of significant correlation is p.

Table 16: Model for *netloss* after 30 days of receiving the first bonus for wagering methods

	Model 17	Model 18
(Intercept)	191.28	199.53^*
	(117.21)	(27.72)
p	450.36^*	261.80^*
	(28.99)	(6.87)
N	20420	20416
R^2	0.01	0.07
adj. R^2	0.01	0.07
Resid. sd	16444.13	3888.96

Standard errors in parentheses

* indicates significance at $p < 0.05$

From Table 16, Model 17 gives the model for the whole dataset and Model 18 for the data set with the familiar 4 outliers removed. For both models, the assumptions are proven. The final model is

$$netloss = 261.80p + 199.53 + \varepsilon_1.$$

When comparing Model 16 and Model 18, the intercept has turned from negative to positive, so even though the p parameter looks similar to the smaller timeframe, the intercept change shows significant increase in *netloss*.

6.3.3 Wagering coefficients

Two types of wagering coefficients are used in some bonuses (but not all).

Table 17: Correlation matrix after 5 days of receiving the first bonus for wagering coefficients

	netloss	w	v
netloss	1.00	0.01	0.07
w	0.01	1.00	0.50
v	0.07	0.50	1.00

Table 18: Correlation matrix after 30 days of receiving the first bonus for wagering coefficients

	netloss	w	v
netloss	1.00	0.01	0.05
w	0.01	1.00	0.44
v	0.05	0.44	1.00

From Table 17 and Table 18 we can see that individually there is no correlation between the coefficients and *netloss*. This might be due to the fact that for more than half of the given bonuses, the wagering coefficients are not used. For this, some filtering of the bonuses is required.

In the following correlation matrices, only the bonuses that had these wagering coefficient are used (the average of such coefficients).

Table 19: Correlation matrix after 5 days of receiving the first bonus for bonuses with wagering coefficients

	netloss	w_α	v_α
totalnetloss	1.00	0.00	0.00
w_α	0.00	1.00	0.88
v_α	0.00	0.88	1.00

Table 20: Correlation matrix after 30 days of receiving the first bonus for bonuses with wagering coefficients

	netloss	w_α	v_α
totalnetloss	1.00	0.00	0.00
w_α	0.00	1.00	0.94
v_α	0.00	0.94	1.00

From Table 19 and Table 20, the matrices show that there is absolutely no correlation between explanatory variables and *netloss*, but there is significant correlation between the variables themselves. Therefore we can presume that assigned bonuses quite often have used both coefficients. Furthermore, the fact that there is no correlation suggest research with non-linear models, because of the expectation that the coefficients have significant correlation with *netloss*.

6.3.4 Other bonus attributes

As stated, each bonus has many more attributes that define it, some of them are binary, some have numerical value. In this these we will look at three of the most widely used ones.

Table 21: Correlation matrix after 5 days of receiving the first bonus for other bonus attributes

	netloss	x	y	z
netloss	1.00	0.00	0.19	0.15
x	0.00	1.00	-0.07	-0.08
y	0.19	-0.07	1.00	0.89
z	0.15	-0.08	0.89	1.00

For 5 days, the correlation matrix in Table 21 gives that y and z have significant correlations between each other, therefore only one of the attributes can be chosen for the outlier check and final model.

Table 22: Model for *netloss* after 5 days of receiving the first bonus by other attributes

	Model 19	Model 19a	Model 20
(Intercept)	-57.40^{*}	25.12	-65.12^{*}
	(21.61)	(21.14)	(15.02)
y	239.02^{*}		226.58^{*}
	(9.01)		(6.27)
z		206.05^{*}	
		(9.70)	
N	19211	19211	19208
R^2	0.04	0.02	0.06
adj. R^2	0.04	0.02	0.06
Resid. sd	2609.57	2626.28	1814.07

Standard errors in parentheses

* indicates significance at $p < 0.05$

From Table 22, Model 19 and Model 19a give the models for the whole dataset. The coefficient of determination for both these models is very small, but Model 19 has the larger one, therefore a check for outliers will be done for this model. 3 outliers are removed (one outlier that came out in other models does not come out in this model) and for all models, the assumptions are proven. The final model is Model 20:

$$netloss = 226.58y - 65.12 + \varepsilon_1.$$

In this model it is shown that players who have received bonuses with variable y in them are actually the main generators of *netloss*. Without it, the average *netloss* is negative.

Table 23: Correlation matrix after 30 days of receiving the first bonus for other bonus attributes

	netloss	x	y	z
netloss	1.00	0.01	0.11	0.10
x	0.01	1.00	-0.02	-0.02
y	0.11	-0.02	1.00	0.98
z	0.10	-0.02	0.98	1.00

For 30 days, the correlation matrix in Table 23 gives that y and z have significant correlation between each other, therefore only y is chosen because of the larger correlation between it and *netloss*.

Table 24: Model for *netloss* after 30 days of receiving the first bonus by other attributes

	Model 21	Model 22
(Intercept)	-166.72	-36.70
	(123.12)	(28.90)
y	381.21^*	236.96^*
	(23.77)	(5.59)
N	20420	20416
R^2	0.01	0.08
adj. R^2	0.01	0.08
Resid. sd	16437.86	3858.82

Standard errors in parentheses

* indicates significance at $p < 0.05$

From Table 24, Model 21 gives the model for the whole dataset and Model 22 for the data set with the familiar 4 outliers removed. For both models, the assumptions are proven. The final model is

$$netloss = 236.96y - 36.70 + \varepsilon_1.$$

The model shows that again the bonuses with variable y are profitable, but the profitability of these bonuses is decreased, which suggests variable y being more profitable short-term.

For both time periods variable x has barely any correlation with *netloss*, even though it is used quite often. This confirms the licensees expectations because this parameter is enabled for the convenience, not for any reasonable expectation of *netloss* change.

6.3.5 Interplay of variables

When combining the correlation matrices for triggers, wagering requirements and other attributes, only the variables that had any significant correlation with *netloss* in the previous models are chosen.

Table 25: Correlation matrix after 5 days of receiving the first bonus interplay between bonus triggers, wagering methods and other attributes

	$netloss$	l	j	n	p	s	y	z
$netloss$	1.00	0.15	0.20	0.19	0.16	0.13	0.19	0.15
l	0.15	1.00	0.27	-0.01	0.53	0.28	0.57	0.55
j	0.20	0.27	1.00	0.03	0.60	0.51	0.71	0.54
n	0.19	-0.01	0.03	1.00	0.01	0.06	0.02	0.00
p	0.16	0.53	0.60	0.01	1.00	-0.12	0.85	0.94
s	0.13	0.28	0.51	0.06	-0.12	1.00	0.06	-0.24
y	0.19	0.57	0.71	0.02	0.85	0.06	1.00	0.89
z	0.15	0.55	0.54	0.00	0.94	-0.24	0.89	1.00

Table 26: Correlation matrix after 30 days of receiving the first bonus for interplay between bonus triggers, wagering methods and other attributes

	$netloss$	j	p	y	z
$netloss$	1.00	0.16	0.11	0.11	0.10
j	0.16	1.00	0.49	0.56	0.47
p	0.11	0.49	1.00	0.95	0.97
y	0.11	0.56	0.95	1.00	0.98
z	0.10	0.47	0.97	0.98	1.00

For both 5 days and 30 days the correlation matrices in Table 25 and Table 26 show that the triggers, wagering requirements and other attributes have a significant multicollinearity. In this case the variables can't be looked at separately to create models. This is in line with the bonus structure, where each bonus should have both a trigger type and wagering method alongside other attributes. Therefore we have to look at the combined effects of previous variables to create more complex linear models.

6.3.6 Combination of variables

Since bonuses will have both a bonus trigger and a wagering method, the first combination to look over are combinations of such.

Going back to the trigger types and wagering requirements, we have to go through all of the types and requirements, to check what in these cases were actually used for significant amount of times. From the trigger types, we have to look for i, j, l, m and from the wagering requirements o, p, s, t. The reason for this in unknown, but is not desired by Playtech.

For better understanding, the models with the combinations of the variables are noted as two or more letter character strings, which has the characters of the solitary properties.

NOTE! Here it is important to note that the correlation between the combination of explanatory variables and the dependent variable does not depend on the correlation between explanatory variable and dependent variable.

Table 27: Correlation matrix after 5 days of receiving the first bonus when combining wagering methods and triggers

	netloss	*lp*	*js*	*jp*	*l*	*p*	*j*	*s*
netloss	1.00	0.12	0.13	0.14	0.15	0.16	0.20	0.13
lp	0.12	1.00	-0.01	0.30	0.69	0.74	0.22	-0.03
js	0.13	-0.01	1.00	-0.08	0.12	-0.07	0.66	0.86
jp	0.14	0.30	-0.08	1.00	0.24	0.86	0.69	-0.15
l	0.15	0.69	0.12	0.24	1.00	0.53	0.27	0.28
p	0.16	0.74	-0.07	0.86	0.53	1.00	0.60	-0.12
j	0.20	0.22	0.66	0.69	0.27	0.60	1.00	0.51
s	0.13	-0.03	0.86	-0.15	0.28	-0.12	0.51	1.00

NOTE! When trying to add other attributes, wagering methods and triggers together for the models, the combinations of the variables are lot more numerous, therefore the largest of the correlation matrices are not shown, rather the combinations which have significant correlations with the dependent variable are given.

From Matrix 27, we can see, that the combinations of different variables have a high correlation between them and the single variables, therefore single variables cannot be added to the model.

Table 28: Model for *netloss* after 5 days of receiving the first bonus when combining wagering methods and triggers

	Model 25	Model 26
(Intercept)	0.31	11.21
	(20.43)	(12.96)
lp	279.19*	312.72*
	(23.86)	(15.14)
js	350.47*	304.49*
	(17.52)	(11.12)
jp	300.32*	221.41*
	(18.03)	(11.46)
N	19211	19207
R^2	0.05	0.09
adj. R^2	0.05	0.09
Resid. sd	2595.22	1646.77

Standard errors in parentheses

* indicates significance at $p < 0.05$

From Table 28, Model 25 gives the model for the whole dataset and Model 26 for the data set with the familiar 4 outliers removed. For both models, the assumptions are proven. The final model is

$$netloss = 312.72lp + 304.49js + 221.41jp + 11.21 + \varepsilon_1.$$

When comparing this model to models from Table 10, then the profitability of the triggers is reversed from the older model (j being better than l). For bonuses with j, the main wagering methods are s and p. For bonuses with trigger type l, the profitability of this combination is higher than in the older model, therefore this combination increases the profitability.

For trigger types, comparing them to the model in Table 14, we can see that wagering method s becomes a lot more profitable when combining with trigger type j.

Table 29: Correlation matrix after 30 days of receiving the first bonus when combining wagering methods and triggers

	netloss	*jp*	*p*	*j*
netloss	1.00	0.17	0.11	0.16
jp	0.17	1.00	0.63	0.70
p	0.11	0.63	1.00	0.49
j	0.16	0.70	0.49	1.00

For 30 days, only j and p have a combination that is significant and we cannot use the single variables for this model because of the multicollinearity problems.

Table 30: Model for *netloss* after 30 days of receiving the first bonus when combining wagering methods and triggers

	Model 27	Model 28
(Intercept)	-364.38^*	175.56^*
	(119.96)	(29.20)
jp	2257.53^*	563.32^*
	(92.72)	(22.76)
N	20420	20416
R^2	0.03	0.03
adj. R^2	0.03	0.03
Resid. sd	16306.04	3965.88

Standard errors in parentheses

* indicates significance at $p < 0.05$

From Table 30, Model 27 gives the model for the whole dataset and Model 28 for the data set with the familiar 4 outliers removed. For both models, the assumptions are proven. Here we can see that the model is actually quite weak, mainly because of the limiting number of variables going into the model. The final model is

$$netloss = 563.32jp + 175.56\varepsilon_1.$$

Table 30 shows that the outliers have significantly changed the models parameters, creating such large losses for themselves with this combination that they have managed to turn receiving other types of bonuses profitable for the players.

When adding the combinations for wagering coefficients and other bonus attributes, the complexity and length of the model has the possibility of becoming even larger. Again, we can disregard x variable because of the weak correlation and z because of the perfect linear relationship between it and y.

Table 31: Correlation matrix after 5 days of receiving the first bonus when combining wagering methods and other attributes

	netloss	*sy*	*py*	*oy*	*p*	*y*
netloss	1.00	0.09	0.12	-0.01	0.16	0.19
sy	0.09	1.00	-0.02	-0.08	-0.04	0.24
py	0.12	-0.02	1.00	-0.02	0.78	0.64
oy	-0.01	-0.08	-0.02	1.00	0.03	0.13
p	0.16	-0.04	0.78	0.03	1.00	0.85
y	0.19	0.24	0.64	0.13	0.85	1.00

From Matrix 31, we can see that for 5 days, the combinations of different variables have a high correlation between them and the single variables, therefore those cannot be added to the model. The only variable with any sort of meaningful correlation is py.

Table 32: Correlation matrix after 5 days of receiving the first bonus when combining bonus triggers, wagering methods and other attributes

	netloss	*lpy*	*jpy*	*jsy*	*lp*	*js*	*jp*	*y*
netloss	1.00	0.12	0.14	0.12	0.12	0.13	0.14	0.19
lpy	0.12	1.00	0.30	-0.01	1.00	-0.01	0.30	0.62
jpy	0.14	0.30	1.00	-0.08	0.30	-0.08	1.00	0.75
jsy	0.12	-0.01	-0.08	1.00	-0.01	0.81	-0.08	0.28
lp	0.12	1.00	0.30	-0.01	1.00	-0.01	0.30	0.62
js	0.13	-0.01	-0.08	0.81	-0.01	1.00	-0.08	0.20
jp	0.14	0.30	1.00	-0.08	0.30	-0.08	1.00	0.75
y	0.19	0.62	0.75	0.28	0.62	0.20	0.75	1.00

For 5 days, the correlation matrix in Table 32 shows that the only variables to include in the models are the ones that are the most specific. When looking at combinations though, we can see that for some pairs, the addition of y does nothing for the correlations, which shows that y is present for all such bonuses.

Table 33: Model for *netloss* after 5 days of receiving the first bonus when combining bonus triggers, wagering methods and other attributes

	Model 29	Model 29a	Model 30
(Intercept)	189.93^*	22.39	21.74
	(19.15)	(20.32)	(12.87)
py	9.08^*		
	(0.55)		
lpy		280.94^*	313.74^*
		(23.91)	(15.15)
jpy		294.31^*	218.53^*
		(18.06)	(11.46)
jsy		415.40^*	399.88^*
		(23.31)	(14.77)
N	19211	19211	19207
R^2	0.01	0.04	0.08
adj. R^2	0.01	0.04	0.08
Resid. sd	2638.26	2600.71	1647.45

Standard errors in parentheses

* indicates significance at $p < 0.05$

From Table 33, Model 29 and Model 29a are two models for the whole dataset. Model 29 shows the combinations of wagering methods and other attributes, while Model 29a also adds wagering methods. For both models, the assumptions are proven. Here Model 29 shows that removing wagering methods from the model has a negative effect on the fit of the model compared to Model 29a. Therefore, when removing outliers, we only look at Model 29a, and Model 30 has 4 of the familiar outliers removed. The final model is:

$$netloss = 313.74lpy + 218.53jpy + 399.88jsy + 21.74 + \varepsilon_1.$$

When comparing this model to Model 25, it has a slightly worse R^2. Looking at the parameters, we can see only minor changes, except for jsy, for which the profitability is significantly larger.

For 30 days, y again is the only other bonus attribute we can add to the model.

Table 34: Correlation matrix after 30 days of receiving the first bonus when combining bonus triggers, wagering methods and other attributes

	netloss	*jpy*	*jp*	*p*	*j*
netloss	1.00	0.17	0.17	0.11	0.16
jpy	0.17	1.00	1.00	0.63	0.70
jp	0.17	1.00	1.00	0.63	0.70
p	0.11	0.63	0.63	1.00	0.49
j	0.16	0.70	0.70	0.49	1.00

For 30 days, the correlation matrix in Table 34 shows that adding y has absolutely no difference on the correlations, therefore we can reason, that if the bonus has the trigger j and wagering method p, then it will also have variable y.

When wanting to add wagering coefficients to the model, then for both 5 and 30 days there is not a significant variable in that list, therefore these models cannot give us any information.

When wanting to add all of the different variables shown previously into a model, then again for both 5 and 30 days there is not a significant variable in that list, therefore these models cannot give us any information. Therefore the largest combination of different variables that is significant in our models is 3.

7 Conclusions

Bonuses are a viable tool for any casino for player retention because of low cost, effectiveness in most cases and stimulating nature. For a lot of people, they stimulate a part of the human mind with the joy of prizes, thrill of gambling and the enjoyment of the luck being on their side.

As seen from the different correlation matrices, the correlations between *netloss* and the different bonus variables are quite weak, with the correlations falling between 0.1 and 0.2. Meanwhile the correlations between the explanatory variables are quite large, which in most cases can be attributed to the fact that a bonus does have several of these properties.

Increasing the timeframe weakens the fit of the linear models. In most models, the correlations weaken for most of the variables and the models usually have fewer significant variables. In some cases, when it was possible to create a model for 5 days, the same could not be said for 30 days. This is because during 30 days, players receive more bonuses, with attributes that overlap. The more bonuses they receive, the less one single variable has an effect on *netloss*.

If the question should arise why we are looking at models with such minor R^2 (half of the models have this under 0.1), then we have to remember that the monetary amounts for these timeframes are in the tens, if not hundreds of millions. When we can predict somewhere close to 10 percent of these numbers, this is still a very significant sum.

It is clear that some bonus parameters are used more often than others, which naturally skews the results. This can be because of the preference of the licensee or the regulations of the given casino's country. Playtech found this interesting because their goal is to provide a system where all bonus types are used as equally as possible.

When looking at the bigger subsets of bonus parameters (e.g. trigger types, wagering methods), then it is clear that the biggest increase for R^2 in the different models are given by the wagering methods. Wagering coefficients meanwhile are the weakest, being not usable in any models.

Wagering coefficients not having an effect on *netloss* is something that is opposite to the expectations. The logic here being: the bigger the wagering coefficient, the longer it takes the player to play through the bonus and the more *netloss* he creates. In this data set, there is no correlation for this, therefore a more thorough investigation should be done on Playtech side.

In several cases, removing the outliers has a significant change in the models parameters. Therefore some bonus attributes are very top-heavy, e.g. very profitable for players who either win or lose a lot of money. Financial people of the licensees should take notice on this, because such players make it more difficult to understand the profitability. This happens because the players are able to trigger them again and again.

When looking at parameters for all models, the ones which have to do with bonus attributes all show the profitability of these bonus types. There was not a single bonus which was stacked for the player or was abused by the playerbase.

The bonus attributes which are the most linear are the trigger type j, wagering requirement p and the other attribute y. Therefore, when giving a bonus with these three parameters, the models *netloss* can be predicted with the least variance.

Linear regression is not the most suitable to analyse the change of *netloss*, but looking at the amount of data, there are several other questions which could be analysed by this method. Coming back to *netloss*, using non-linear regression could give us a better overview for some models, perhaps owning to the fact that the profitability of certain numerical coefficients should start decreasing after a certain period.

References

[1] Bertolt, *Beautiful Correlation Tables in R*, Myowelt blog, [http://myowelt.blogspot.com.ee/2008/04/beautiful-correlation-tables-in-r.html], 2008

[2] R. Christensen *Plane Answers to Complex Questions - The Theory of Linear Models: Fourth edition*, University of New Mexico, 2011

[3] D. B. Dahl, *xtable: Export Tables to LaTeX or HTML. R package version 1.8-0.*, [https://CRAN.R-project.org/package=xtable], 2015

[4] M. Dowle, A. Srinivasan, T. Short, S. Lianoglou with contributions from R. Saporta and E. Antonyan, *data.table: Extension of Data.frame. R package version 1.9.6.*, [https://CRAN.R-project.org/package=data.table], 2015

[5] J. Fox and S. Weisberg *An R Companion to Applied Regression, Second Edition. Thousand Oaks CA: Sage.*, [http://socserv.socsci.mcmaster.ca/jfox/Books/Companion], 2011

[6] E. Käärik, *E-kursuse "Andmeanalüüs II" materjalid*, Tartu Ülikool, 2013

[7] M. Malecki, *apsrtable: apsrtable model-output formatter for social science. R package version 0.8-8.*, [https://CRAN.R-project.org/package=apsrtable], 2012

[8] S. Pakin, *The Comprehensive LATEX Symbol List*, [http://tug.ctan.org/info/symbols/comprehensive/symbols-a4.pdf], 2015

[9] Wikipedia contributors, *Multicollinearity*, Wikipedia, The Free Encyclopedia, [https://en.wikipedia.org/wiki/Multicollinearity], 10 May 2016

[10] X. Yan, X. G. Su, *Linear Regression Analysis: Theory and Computing*, University of Missouri, University of Central Florida, 2009

Appendices

A Hints

- RStudio is a free and open source integrated development environment for R. The advantages of RStudio are an improved text editor and a better overview on all of the data sets, formulas and figures.

- A list of symbols and the corresponding LATEX commands is given in [8].

- For faster formatting in R, *data.table* [4] is an alternative to *data.frame*.

- For formatted LaTex tables, usage of *apsrtable* [7] package is recommended. This will output important information from the *summary* function output in *R*. There is the possibility of putting outputs for two different models side-by-side.

- For a formatted LaTeX correlation matrix, the following *R* code is used (taken from [1]).

```
corstarsl <- function(x){
  require(Hmisc)
  x <- as.matrix(x)
  R <- rcorr(x)$r
  p <- rcorr(x)$P

  ## trunctuate the matrix that holds the correlations to two
      decimal
  R <- format(round(cbind(rep(-1.11, ncol(x)), R), 2))[,-1]

  ## build a new matrix that includes the correlations with
      their appropriate stars
  Rnew <- matrix(paste(R, sep=""), ncol=ncol(x))
  diag(Rnew) <- paste(diag(R), " ", sep="")
  rownames(Rnew) <- colnames(x)
  colnames(Rnew) <- paste(colnames(x), "", sep="")

  return(Rnew)
}
```

This creates a function *corstarsl*, which gives the correlation matrix of the selected columns. Using the *xtable* [3] package, this outputs the LaTeX code for the matrix.

- When selecting a large amount of variables for the model, it is normal that some of them are not statistically significant. But removing all of the insignificant variables at once is not the preferred action, regarding the dependence that these variables can have for each other and the significant variables. Therefore, a good solution is to use the R command *step*. With the *step* command, R goes through the model step by step, eliminating variables one by one and finally settling on the best possible model based on AIC (Akaike information criterion).

- Another way to change the model quickly is to use *update* command. *Update* command allows a new model to be created by describing the changes in the parameters compared to the previous model. This cleans up the code by not needing copy/paste for long formulas. For example *update(*$y \sim x_1, \sim . + x_2$*)* will give you the formula $y \sim x_1 + x_2$.

- For Durbin-Watson test it is possible to use both *dwtest* function in the *lmtest* package or *dwt* function in the *car* [5] package. *Dwt* function has some sort of a large optimization problem, where R seems to receive a big memory leak, therefore *dwtest* is recommended.

- Using the R code

```
opar <- par(mfrow = c(2,2), oma = c(0, 0, 1.1, 0))
plot(model, las = 1)
```

gives 4 plots for different aspects for the linear model: Residuals vs Fitted, Normal Q-Q, Scale Location and Residuals vs Leverage. This makes it easier to search for outliers, since the clearest outliers are shown with the observation number.

B Lihtlitsents lõputöö reprodutseerimiseks ja lõputöö üldsusele kättesaadavaks tegemiseks

Mina, Kristjan Sirge,

1. annan Tartu Ülikoolile tasuta loa (lihtlitsentsi) enda loodud teose *Measuring the effectiveness of casino bonuses using linear regression models*, mille juhendaja on Meelis Käärik,

   (a) reprodutseerimiseks säilitamise ja üldsusele kättesaadavaks tegemise eesmärgil, sealhulgas digitaalarhiivi DSpace-is lisamise eesmärgil kuni autoriõiguse kehtivuse tähtaja lõppemiseni;

   (b) üldsusele kättesaadavaks tegemiseks Tartu Ülikooli veebikeskkonna kaudu, sealhulgas digitaalarhiivi DSpace´i kaudu kuni autoriõiguse kehtivuse tähtaja lõppemiseni.

2. olen teadlik, et punktis 1 nimetatud õigused jäävad alles ka autorile.

3. kinnitan, et lihtlitsentsi andmisega ei rikuta teiste isikute intellektuaalomandi ega isikuandmete kaitse seadusest tulenevaid õigusi.

Tartu, 10.05.2016