Master of
Statistics and
Data Science

STATISTICAL ANALYSIS OF RELIABILITY AND SURVIVAL DATA
GROUP PROJECT

Authors: Jose Nicolas Saud Miño (r0837039), Hovhannes Margaryan (r0868941), Venkatesh Viswanathan Manikantan (r0825919)
Instructor: Prof. Ingrid Van Keilegom

Spring 2022

1 Introduction

Customer churn is a major problem in the telecommunication industry. A competitive environment between different competitors, evolving technologies, a wide range of technological products and lower prices, have shifted the market towards a customer centric product rather than just offering connectivity solutions. On this regard, internet service providers (ISPs) are keen to recollect as much data as possible on their customers, in order to analyze their satisfaction and prevent unsatisfied clients of deactivating the service. Based on this premise, this study focuses on survival analysis conducted on an ISP´s customer churn dataset with means of payment as a treatment variable. Voluntary churn of customers is regarded as a variable of interest. The response variable is subject to random right censoring where a client is considered censored if his or her membership is suspended, refereed to involuntary churn (forced deactivation), or if he or she is still active in the company at the end of the study period. The Kaplan-Meier estimator is used to estimate the survival curves of the means of payment variable. Differences are observed in the Kaplan-Meier estimates of the three levels of the means of payment treatment varible: credit card, savings and deposit. Additionally, semi-parametric Cox and parametric Accelerated Failure Time models are fitted to the dataset to discover relationship between the survival time of a customer using explanatory variables.

The structure of the paper is as follows. Section 2 provides explanatory analysis of the customer churn dataset along with preliminary insights about the survival behavior of customers. Section 3 presents the results of the Kaplan-Meier estimator on the treatment variable. Section 4 discusses Cox and Accelerated Failure Time models on the churn dataset. Section 5 provides the paper´s conclusions.

2 Explanatory data analysis of customer churn dataset

In this section explanatory data analysis is conducted on the customer churn dataset. The dataset is from an Ecuadorian ISP's customer database of active and deactivated clients from years 2017 to 2022. The dataset contains $n = 47514$ observations. 19621 customers are deactivated and 27893 are active. 69.07% of the observations are censored, while the rest of the customers left the company voluntarily. The dataset has the following variables:

- status - the current status of a customer with levels deactivated and active,
- means-payment - the payment method used by a customer to pay for a the service with levels credit card, deposit, savings,
- generation - the age group a customer falls in with levels centenials, millennials, old, x generation,
- risk-score - the credit score of a customer according to a credit buro
- calls-4m - the number of times a customer called the company in the last 4 months before its last active status

- visit-4m - the number of times a customer was technically visited by the company in the last 4 months before its last active status
- time - the time a costumer is with the company in months
- voluntary-churn - the variable of interest with levels 1 and 0. 1 indicates voluntary churn while 0 indicates right censoring. Voluntary churn, is the deactivation attributed to a customer's desire of finishing the service due to reasons such as other competitors, technical problems, coverage, lower prices among others. Involuntary churn is a forced deactivation that happens when the customer stops paying the service without sharing the motive with the company and ends up with unpaid debt.

Fig. 1 shows a box-plot of status against time. It can be observed that the sample mean of deactivated customers is slightly higher than the sample mean of the active customers. Numerically, the sample mean of the active customers is 25.68, while the sample mean of the deactivated customers is 28.03. No outliers can be detected in the plot.

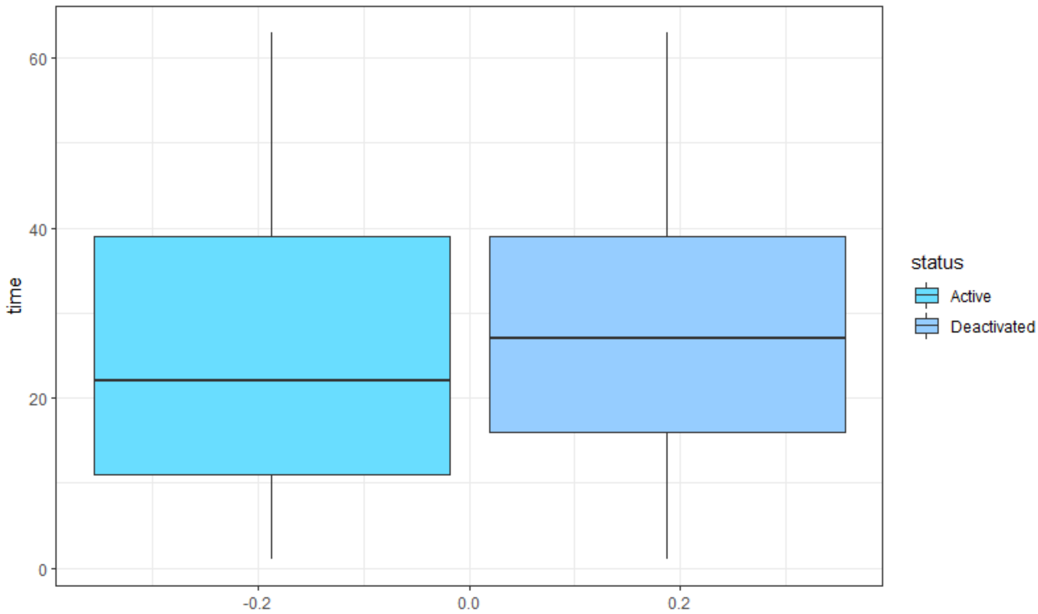
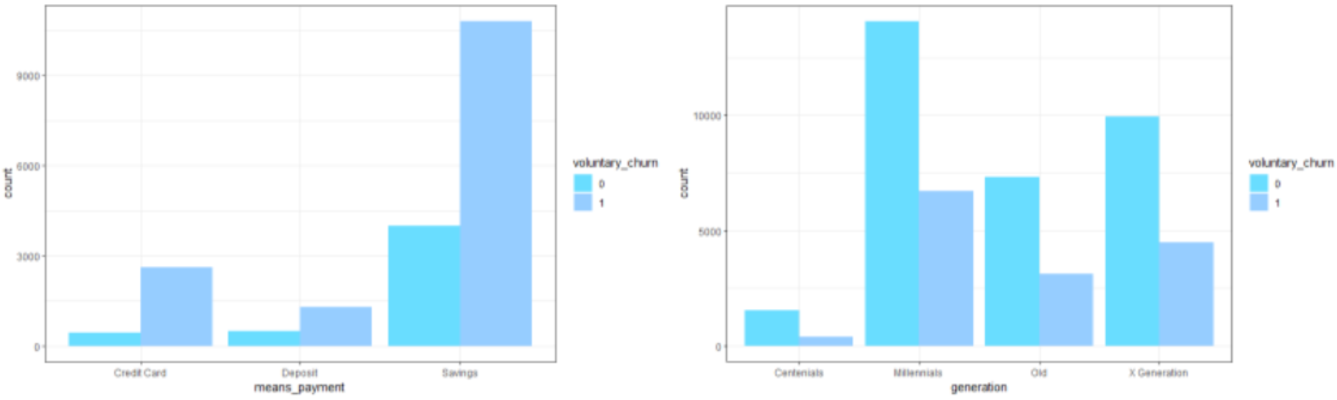


Figure 1: Box-plot of status against time

Fig. 2(a) represents a bar graph of the payment method stratified by voluntary churn. The number of observations per level of the means of payment variable is highly unbalanced and this issue is addressed in Section 4. Moreover, 73% of the customers have savings as their payment method while 19.1% of the customers use credit card as their method of payment. It can be observed that the number of customers that have voluntarily left the company is more in customers whose payment method is savings than those whose payment method is credit card or deposit. Moreover, the number of customers with deposit as their payment method and who voluntarily left the company is the lowest in comparison to the other two levels of the the means of payment variable.



(a) (b)

Figure 2: Bar graphs of means of payment (a) and generation (b) variables stratified by voluntary churn

Fig. 2(b) shows a bar graph of the generation variable stratified by voluntary churn. The number of observations per level of the generation is unbalanced. Furthermore, 43.66% percent of the customers fall in the millennials group while 30.34% of the customers are from the x generation group. The number of customers that have involuntarily left the company is more in the group of customers fall in millennials in comparison to the other levels of the variable generation. This correlates with the intrinsic behavior of each age group. Since millenials are a younger population with less stable income and more focused on price rather than quality, they tend to leave the company for better prices without paying, as opposed to older age groups such as centenials and X generation (40+ years). As a result, it is expected that this variable influences the probability of churn, depending on the age generation of each observation.

Next, the status of customers is analyzed against the risk score. Risk score directly represents the probability of debt of a certain person, based on its financial behavior. If a person has high amount of debt, the credit score will be lower, whereas a person with no debt has a better financial history and hence have a higher credit score. The risk score range from 0 to 999. This can be observed in Fig. 3, where the distribution plot for voluntary and involuntary churn demonstrates that lower credit scores (risk score) represent a higher population of deactivated customers due to debt whereas high credit scores are more related to better customers that churn due to dissatisfaction with the service and that are those in which the company is more interest in analyzing and retaining. Thus, the variable risk-score is expected to influence in the survival probability of a customer, since lower risk scores will enhance a customer´s deactivation before the study period but for involuntary reasons, which is part of the censoring variable.

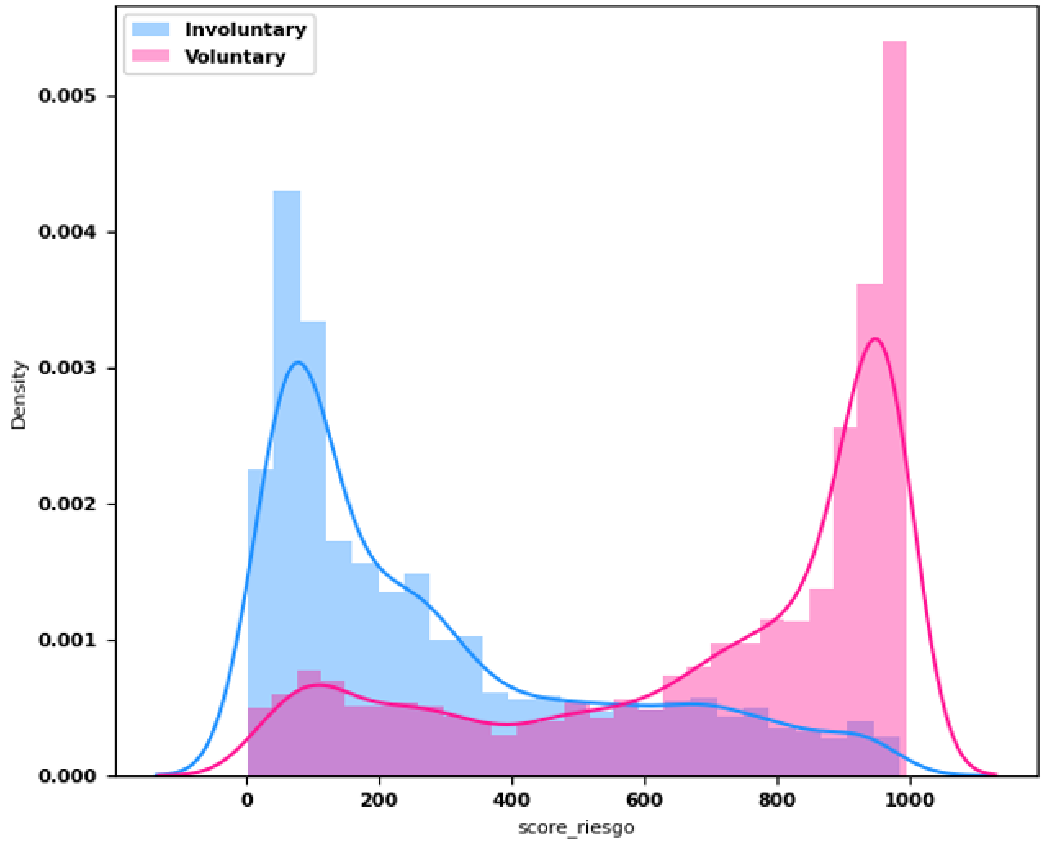


Figure 3: Distribution plot of voluntary and involuntary churn against the variable risk score

3 Estimation of survival curves of customers by Kaplan-Meier

In this section, the Kaplan-Meier estimator is computed for the means of payment variable. The Kaplan-Meier estimator for right censored data is defined as:

$$\hat{S}(t) = \prod_{j:Y_j \leq t} \frac{r_j - d_j}{r_j}, \tag{1}$$

where Y_j indicates the event time ($Y_j = min(T_j, C_j)$) with T_j being the survival time and C_j the censoring time. r_j is the size of risk set at Y_j and d_j is the number of voluntary churns at Y_j. T and C are independent since the voluntary churn is not dependant on involuntary churn of a customer. Details about these concepts are provided in Section 2. Therefore, the Kaplan-Meier estimator is appropriate in this scenario as its main assumption of independence between the survival time and censoring is not violated. Fig. 4 presents the estimated survival curves with 95% confidence intervals for each level of the means of payment variable: credit card, deposit and savings. The confidence intervals are calculated using the log-log transformation. First, it can be noticed that the estimates of the survival curves do not reach zero as the largest observation, $Y_j = 63$, is censored. Second, the estimated survival curve of deposit is higher than the survival curves of credit card and savings.

Moreover the estimated survival curves of credit card and savings are similar. Thus, the graph of estimated survival curves suggests that the customers that use deposit as their method of payment have higher chances of staying with the company. The implication regarding the survival of customers with deposit as their payment method can also be observed in Fig. 2(a).

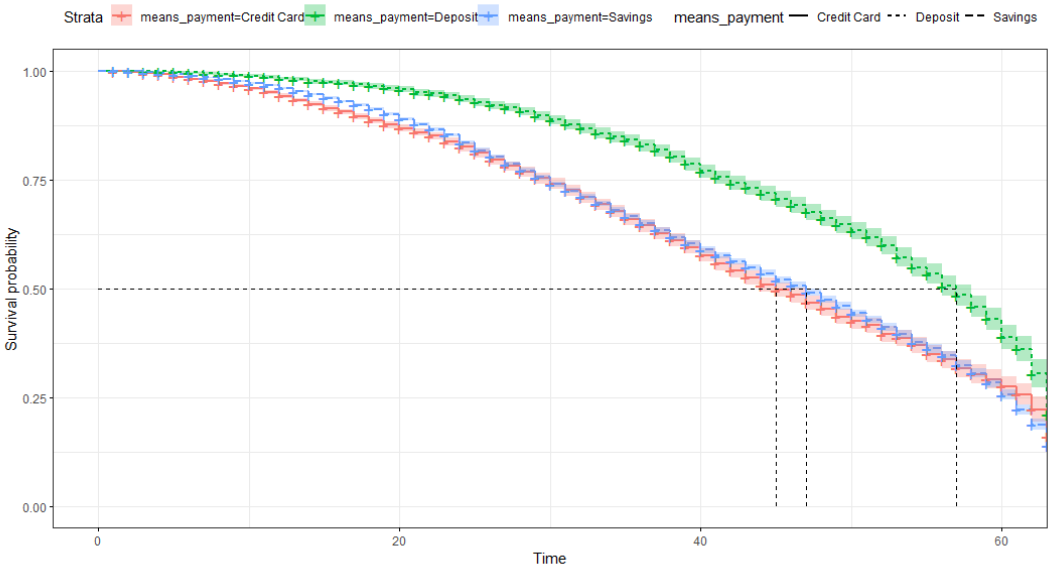


Figure 4: Estimated Kaplan-Meier survival curves for the means of payment variable

Next the mean survival time for each level of the means of payment is estimated. As mentioned in the previous paragraph, the largest observation is censored and thus the mean survival time is estimated as

$$\hat{\mu}_{t_{max}} = \int_0^{t_{max}} \hat{S}(t)dt, \qquad (2)$$

where $\hat{S}(t)$ is the Kaplan-Meier estimator and $t_{max} = 63$, the maximum survival time, for all levels of the means of payment variable. The estimated mean survival time of deposit is 50.97 is higher than the estimated mean survival time of credit card and savings which correspondingly are equal to 43.26 and 43.87. Since, $\hat{\mu}_{t_{max}}$ is not consistent 0.25, 0.5 and 0.75 quantiles are estimated by:

$$\hat{x}_p = inf\{t|\hat{S}(t) \leq 1 - p\}, \qquad (3)$$

where p indicates the quantile. 0.25, 0.5 and 0.75 quantiles and their confidence intervals for each level of the means of payment variable are presented in Table 1, Table 2 and Table 3. Table 1, Table 2 and Table 3 contain the quantile, p, the estimated pth quantile, $\hat{x}_p$ and the lower and upper bounds of 95% confidence interval of $\hat{x}_p$.

p	$\hat{x}_p$	lower	upper
0.25	30	29	31
0.5	45	44	47
0.75	62	61	63

Table 1: Estimated quantiles of survival time of credit card

p	$\hat{x}_p$	lower	upper
0.25	42	41	43
0.5	57	56	58
0.75	63	63	NA

Table 2: Estimated quantiles of survival time of deposit

p	$\hat{x}_p$	lower	upper
0.25	30	30	30
0.5	47	46	47
0.75	61	60	61

Table 3: Estimated quantiles of survival time of savings

It is observed that $\hat{x}_{0.5} = 57$ for deposit and it is higher than the median survival time of customers with credit card or savings as their payment method. This observation aligns with the previous finding of estimated mean survival time of customers that pay by deposit. Moreover, $\hat{x}_{0.25} = 30$ for credit card and savings is lower than $p = 0.25$ quantile of deposit. It is also noticed that the upper bound of the 95% confidence interval of deposit for $p = 0.75$ is NA which implies that it is not possible to calculate its value as the number of censored observations is higher than the event of interest as mentioned in Section 2. These results complement the expected behavior of a client with respect to the variable voluntary churn. Payment methods, credit card and savings follow a similar process of invoice (automatic debit from credit card or savings account) and they both represent similar types of clients and thus the survival curves, the mean and median survival time for both populations overlap. However, this phenomenon does not occur with payment type deposit, where a client's payment depends entirely on his or her willingness to pay for the service.

The difference between survival curves of the three levels of the means of payment variable is formally tested. Testing the difference of survival functions, S_j is equivalent to testing the difference of hazard functions, h_j where $h(t) = \frac{d}{dt}(-log(S(t)))$. Thus, the following hypothesis is considered:

$$\begin{aligned} H_0 &: h_1(t) = h_2(t) = h_3(t), \forall t \leq Y_r \\ H_1 &: h_i(t) \neq h_j(t), \text{for at least one} (i, j), t \leq Y_r, \end{aligned} \tag{4}$$

where $h_i, i = \overline{1, 3}$ correspondingly refers to the three levels of the means of payment variable: credit card, deposit and savings. The hypothesis is tested by the long-rank test which weights all voluntary churns equally. The results of the long-rank test are summarized in Table 4. Table 4 contains the payment method, the number of observations in each level, N, observed, O and expected, E, number of voluntary churns and values of the statistics.

level	N	Observed	Expected	$\frac{(O-E)^2}{E}$	$\frac{(O-E)^2}{V}$
credit card	9110	2625	2363	29.1	35.6
deposit	3720	1277	2079	309.4	375.4
savings	34684	10793	10253	28.4	96.6

Table 4: The result of the log-rank test applied to the hypothesis presented in Eq. 4

$\chi_2^2 = 382$ with p-value $< 2e - 16$. Thus, the null hypothesis that the hazard functions of the three levels of the means of payment variable are equal is rejected. The result of the log-rank test aligns with the observation of difference in survival curves in Fig. 4.

4 Cox and Accelerated Failure Time Models

In this section semi-parametric Cox and Accelerated Failure Time (AFT) models are applied to the customer churn dataset to explore the relationship between the survival time and explanatory variables that are present in the dataset.

The Cox model assumes that the ratio of the hazards of two observations is constant over time. Since, the dataset contains tied observations the likelihood is approximated by Efron's formula. The Cox model is initially fitted using the means of payment variable and model selection is conducted using the AIC metric in both directions by the following covariates: generation, risk-score, calls-4m and visit-4m. Table 5 demonstrates the final Cox model and contains the step, degrees of freedom, df, deviance, residual df, residual deviance and the AIC score for each step.

Step	Df	Deviance	Residual Df	Residual Deviance	AIC
means-payment	2	0	14693	283295.8	283299.8
+calls-4m	1	402.52704	14692	282893.3	282899.3
+visit-4m	1	294.60426	14691	282598.7	282606.7
+generation	3	103.18700	14688	282495.5	282509.5
+ risk-score	1	42.51505	14687	282453.0	282469.0

Table 5: Final Cox model selected using the AIC metric

Thus, the final model chosen using the AIC score contains the following covariates: means-payment, generation, risk-score, calls-4m and visit-4m. The outcome of the final model is presented in Table 6. Table 6 shows the covariate, the estimated coefficient, $\hat{\beta}$, the relative risk, $exp(\hat{\beta})$, the standard error of the coefficient, s.e.($\hat{\beta}$), 95% confidence interval of the relative risk and the p-values.

Covariate	$\hat{\beta}$	$exp(\hat{\beta})$	s.e.($\hat{\beta}$)	95% CI	p-value
credit card	0.059	1.062	0.022	[1.016, 1.108]	0.006
deposit	-0.522	0.593	0.029	[0.559, 0.628]	<2e-16
millennials	0.0271	1.027	0.051	[0.928, 1.137]	0.599
old	-0.183	0.832	0.0535	[0.749, 0.924]	0.0006
x generation	-0.122	0.884	0.0524	[0.798, 0.980]	0.019
risk-score	0.00017	1.00017	0.00003	[1.0001, 1.0002]	1e-10
calls-4m	0.087	1.091	0.0032	[1.084, 1.098]	<2e-16
visit-4m	-0.338	0.713	0.02142	[0.683, 0.743]	<2e-16

Table 6: The outcome of the final Cox model

p-values presented in Table 6 suggest that the covariate deposit, calls-4m and visit-4m are significant. The relative risk of a customer with payment method deposit is 0.593 relative to a customer that pays by savings, in the centenials age group with zero risk score, calls and visits. This complements the finding that deposits is a significant factor and it increases the survival probability as indicated in Section 3. Since the relative risk for the population with this type of payment is lower, the survival curve doesn´t decrease as fast and longer permanence can be expected within the company for this population segment.

On the other hand, since the number of observations in each level of the means of payment variable is unbalanced as indicated in Section 2, Propensity Weighting is applied to the dataset given generation, risk-score, calls and visits. Afterwards, the weights are given to the Cox model to check if the unbalancedness of the datasets affects the results of the Cox model. The results of the model are summarized in Table 7.

Covariate	$\hat{\beta}$	$exp(\hat{\beta})$	s.e.($\hat{\beta}$)	95% CI	p-value
credit card	0.082	1.086	0.01404871	[1.034, 1.141]	0.001
deposit	-0.539	0.583	0.013	[0.550, 0.617]	<2e-16
millennials	0.050	1.051	0.031	[0.87, 1.271]	0.601
old	-0.139	0.869	0.033	[0.717, 1.054]	0.157
x generation	-0.104	0.9	0.0325	[0.744, 1.089]	0.282
risk-score	0.0002	1.0002	0.00001	[1.0001, 1.0002]	1e-10
calls-4m	0.08	1.084	0.0021	[1.068 1.1003]	<2e-16
visit-4m	-0.372	0.689	0.015	[0.643, 0.737]	<2e-16

Table 7: The results of the Cox model with Propensity Weighting

The comparison of Table 6 and Table 7 implies that the significance of the covariates is not modified by Propensity Weighting. Thus, it can be concluded that the unbalancedness of the dataset does not affect the results produced by the Cox model.

Next, the assumption of the Cox model regarding constant relative risks is formally assessed. The results of the test are summarised in Table 8. Table 8 shows the covariate, the χ^2 value, df and the p-value for each covariate.

Covariate	χ^2	df	p-value
means-payment	128.42	2	<2e-16
generation	11.77	3	0.008
risk-score	0.49	1	0.48
calls-4m	1.17	1	0.279
visit-4m	5.86	1	0.01
global	156.05	8	<2e-16

Table 8: Results of assessing the assumption of the Cox model: the p-value of the means of payment is significant

Covariate	χ^2	df	p-value
generation	10.71	3	0.013
risk-score	7.17	1	0.007
calls-4m	5.07	1	0.024
visit-4m	0.02	1	0.87
means-payment:time	17.760	3	0.00049
global	43.17	9	0.000002

Table 9: Results of assessing the assumption of the Cox model with interaction with time added on the means of payment variable

It is observed that the p-value of the means of payment variable is significant and hence the assumption of the constant relative risk is violated. To remedy the situation, interaction with time is added on the means of payment in the Cox model and the test is re-conducted on the new model. Table 9 represents the results of assessing if the relative risks are constant over time on the new model. It is noted that the p-value of the means of payment variable with an interaction with time is non-significant in comparison to the means of payment in Table 8.

Next, AFT model is fitted to the data, using the same covariates as in the Cox model. In order to evaluate whether the selected covariates influence or not the survival function by accelerating or slowing a survival's probability of voluntary churn in the company, the following model is considered:

$$S_i(t) = S_0 exp(\theta^t(X_i)t) \tag{5}$$

where S_0 is the baseline survival function and θ is the vector of coeffcients. $exp\theta^t(X_i)$ is the acceleration factor and it enhances the increase or decrease of the survival function in time. AFT model is fitted assuming different distributions such as log-normal, weibull, exponential and log-logistic for S_0. AIC score is used to compare the AFT models as summarized in Table 10.

Distribution	log-normal	weibull	exponential	log-logistic
AIC	151145.3	148791.0	159412.9	149421.5

Table 10: AFT models using log-normal, weibull, exponential and log-logistic as assumptions for the baseline survival function

It is observed that the lowest AIC score has the AFT model with Weibull distribution. The results of the AFT model with Weibull distribution is presented in Table 11. Table 11 contains the covariate, the coefficient, $\hat{\theta}$, the standard error of the coefficient, s.e($\hat{\theta}$), p-value and 95% CI of the coefficient.

Covariate	$\hat{\theta}$	s.e($\hat{\theta}$)	p-value	95% CI
intercept	4.01	0.024	< 2e-16	[3.962, 4.058]
credit card	-0.027	0.01	0.00582	[-0.047, -0.008]
deposit	0.234	0.013	< 2e-16	[0.207, 0.261]
millenials	-0.011	0.023	0.63781	[-0.057, 0.035]
old	0.084	0.024	0.00056	[0.0365, 0.132]
x generation	0.057	0.023	0.01604	[0.0107, 0.104]
risk-score	-0.00007	0.00001	5.8e-10	[-0.0001, -0.00005]
calls-4m	-0.039	0.001	< 2e-16	[-0.042, -0.036]
visit-4m	0.155	0.009	< 2e-16	[0.136, 0.174]

Table 11: Output of Weibull AFT model

Similar results to the Cox model are observed for the AFT model with Weibull distribution. Deposit, calls-4m and visit-4m are statistically significant. The similarity of the AFT model with Weibull distribution is not surprising as in general, the AFT model with Weibull distribution is equivalent to the Cox model. Moreover, the AFT model with Weibull distribution is equivalent to linear model with standard extreme value error distribution.

5 Conclusion

Analysis of the ISP's customer churn dataset was conducted including exploratory data analysis and modelling approaches of the survival probabilities of voluntary churn in order to model and explain the customer's time journey throughout the company in terms of its probability of voluntary deactivation in time. A descriptive analysis of the variables present in the dataset demonstrated that the means of payment variable had considerable differences between the voluntary and involuntary churns. These hypotheses were formally tested using survival estimator models such as Kaplan-Meier, Cox and AFT. All of the models reached the same conclusion that the survival functions were not equivalent between the levels of the means of payment variable. The Kaplan-Meier estimator showed that the median and mean survival time of customers that pay by deposit was higher than the median and mean survival times of customers that pay by credit card or savings. The difference between survival curves of the three levels of the means of payment variable were proven by the long rank test. Additionally, the Cox and Weibull AFT models demonstrated that deposit, the number of calls and visits were statistically significant.

6 References

[1] Cox, D.R. and Oakes, D. (1984). Analysis of survival data, Chapman and Hall, New York.

[2] Grambsch, P. M., Therneau, T. M. (1994). Proportional Hazards Tests and Diagnostics Based on Weighted Residuals. Biometrika, 81(3), 515–526. https://doi.org/10.2307/2337123

7 Appendix

7.1 R Code

```
########################## Libraries ##########################
rm(list=ls())
library(readxl)
library(openxlsx)
library(dplyr)
library(ggplot2)
library(stringr)
library(MASS)
library(zoo)
library(stringi)
library(ramify)
library("KMsurv")
library("survival")
library("survminer")
library(WeightIt)
dataset <- read_excel("survival_churn_dataset.xlsx")

########################## Section 2 ##########################
head(dataset)
n <- dim(dataset)[1]

sum(dataset$status"Deactivated")
sum(dataset$status=="Active")
```

```
100*sum(dataset$voluntary_churn==1)/n
summary(dataset[dataset$voluntary_churn==1, ])
a <- dataset %>% filter(dataset$status == "Active")
mean(a$time)
b <- dataset %>% filter(dataset$status == "Deactivated")
mean(b$time)

sum(dataset$means_payment=="Savings") / n
sum(dataset$means_payment=="Credit Card") / n

sum(dataset$generation =="Millennials") / n
sum(dataset$generation =="X Generation") / n


g <- ggplot(aes(y=time), data=dataset)+
  geom_boxplot(aes(fill=status))+
  theme_bw() +
  scale_fill_manual(values=c("#69DDFF", "#96CDFF"))

g <- ggplot(aes(means_payment, fill = voluntary_churn),
           data=dataset %>% filter(dataset$status == "Deactivated"))+
  geom_bar(position = "dodge")+theme_bw()+
  scale_fill_manual(values=c("#69DDFF", "#96CDFF"))
g <- ggplot(aes(generation, fill = voluntary_churn),
           data=dataset)+
  geom_bar(position = "dodge")+theme_bw()+
  scale_fill_manual(values=c("#69DDFF", "#96CDFF"))

########################## Section 3 ##########################
str(dataset)
dataset$voluntary_churn<-as.numeric(dataset$voluntary_churn)
surv_m <- survfit(Surv(time, voluntary_churn) ~ means_payment, data = dataset,
                  type="kaplan-meier",  conf.type="log-log")
g <- ggsurvplot(surv_m,
              pval = FALSE, conf.int = TRUE,
              risk.table = FALSE,
              risk.table.col = "means_payment",
              linetype = "means_payment",
              surv.median.line = "hv",
              ggtheme = theme_bw())
survival:::survmean(surv_m, rmean="common")
quantile(surv_m, 0.25)
quantile(surv_m, 0.5)
quantile(surv_m, 0.75)
survdiff(Surv(time, voluntary_churn) ~ means_payment, data=dataset, rho=0)

########################## Section 4 ##########################
########################## Cox ################################
dataset$means_payment <- as.factor(dataset$means_payment)
dataset$means_payment <- relevel(dataset$means_payment,
                                 ref="Savings")
Scope = list(upper = .~ ((means_payment) + (generation)
                        + risk_score + calls_4m
```

```
                          +visit_4m),
             lower = .~((means_payment)))
fit_0 <- coxph(Surv(time, voluntary_churn) ~
                      (means_payment), data=dataset)
step <- stepAIC(fit_0, Scope,  direction = "both")
step$anova
model_final <- coxph(Surv(time, voluntary_churn) ~
                       (means_payment) + (generation)
                     + risk_score + calls_4m
                     +visit_4m, data=dataset)
exp(confint(model_final, level=0.95))

########################## Propensity Weighting ################
w.out <- weightit(means_payment ~ generation +
                    risk_score + calls_4m+visit_4m, data = dataset,
                  focal = "Savings", estimand = "ATT")
w<-w.out$weights

model_final_weighted <- coxph(Surv(time, voluntary_churn) ~
                       (means_payment) + (generation)
                     + risk_score + calls_4m
                     +visit_4m, data=dataset, weights = w)
exp(confint(model_final_weighted, level=0.95))

cox_model_assumption_test <- cox.zph(model_final, terms=TRUE)
print(cox_model_assumption_test)

model_final_with_time <- coxph(Surv(time, voluntary_churn) ~
                                 means_payment:time + generation
                               + risk_score + calls_4m
                               +visit_4m, data=dataset)
model_final_with_time
cox_model_assumption_test <- cox.zph(model_final_with_time, terms=TRUE)
print(cox_model_assumption_test)

########################## AFT ############################
logn = survreg(Surv(time, voluntary_churn) ~
                 (means_payment) + (generation)
               + risk_score + calls_4m
               +visit_4m, data=dataset, dist="lognormal")
weib = survreg(Surv(time, voluntary_churn) ~
                 (means_payment) + (generation)
               + risk_score + calls_4m
               +visit_4m, data=dataset, dist="weibull")
expon = survreg(Surv(time, voluntary_churn) ~
                  (means_payment) + (generation)
                + risk_score + calls_4m
                +visit_4m, data=dataset, dist="exponential")
loglogist = survreg(Surv(time, voluntary_churn) ~
                      (means_payment) + (generation)
                    + risk_score + calls_4m
                    +visit_4m, data=dataset, dist="loglogistic")
logn_aic <- extractAIC(logn)[2]
```

```r
weib_aic <- extractAIC(weib)[2]
expon_aic <- extractAIC(expon)[2]
loglogist_aic <- extractAIC(loglogist)[2]
AIC_parametric = c(logn_aic, weib_aic, expon_aic, loglogist_aic)
names(AIC_parametric) = c( "log(normal)", "weibull", "exp", "log(logistic)")
summary(weib)
confint(weib, level=0.95)
```