ICA 5

Jonathan Thompson

2024-03-18

Customer Churn

In this assignment, you will estimate a survival analysis model on customer churn. In this dataset, you will find various characteristics about a set of phone company customers

customerID: Customer ID number gender: Provides the stated gender of the customer SeniorCitizen: States if the person is a senior citizen Partner: Does the person have a partner Dependents: Does the person have dependents tenure: States how long the person has been with the Bank PhoneService: Do they have phone service MultipleLines: Do they have multiple lines InternetService: What type of Internet Service do they have (DSL, fiber optic, none) OnlineSecurity: Do they have online security? OnlineBackup: Do they have online backup? DeviceProtection: Do they have device protection? TechSupport: Did they use tech support? StreamingTV: Do they use the internet to stream TV? StreamingMovies: Do they use the internet to stream movies? Contract: What type of contract do they have (month-to-month, one-year, or two year) PaperlessBilling: Do they use paperless biling PaymentMethod: How do they pay for services MonthlyCharges: What is their monthly charge? TotalCharges: What is their total charge for the quarter? Churn: Did they leave the company

1) Find the simple average of `tenure`. Explain why this simple average can be biased.

```
library(readr)

Churn <- read.csv ("C:/Users/catho/OneDrive/Documents/AAT/WA_Fn-UseC_-Telco-Customer-Churn.csv")
```

Answer: The average number of months a customer remains with the bank is 32.37 months. In survival analysis, it often contains censored observations that have not occurred by the end of a certain study or experiment. If a simple average is used without accounting for censoring, it may produce biased results.

2) Find the simple average of `tenure` by the following groups. Gender, SeniorCitizen, and Partner.

```
Churn$SeniorCitizen <- factor(Churn$SeniorCitizen, labels= c("Non-Senior", "Senior"))

tab1 <- aggregate(Churn$tenure, by=list(Churn$gender, Churn$SeniorCitizen, Churn$Partner), FUN=mean)

names(tab1) <-c ("Gender", "Senior Status", "Partnered", "Mean of Tenure")

knitr::kable(tab1)
```

Gender	Senior Status	Partnered	Mean of Tenure
Female	Non-Senior	No	22.87432

Gender	Senior Status	Partnered	Mean of Tenure
Male	Non-Senior	No	23.18970
Female	Senior	No	24.85938
Male	Senior	No	25.37751
Female	Non-Senior	Yes	41.72639
Male	Non-Senior	Yes	42.55436
Female	Senior	Yes	42.63710
Male	Senior	Yes	40.54154

```
library(modelsummary)
```

```
## Version 2.0.0 of 'modelsummary', to be released soon, will introduce a
##   breaking change: The default table-drawing package will be 'tinytable'
##   instead of 'kableExtra'. All currently supported table-drawing packages
##   will continue to be supported for the foreseeable future, including
##   'kableExtra', 'gt', 'huxtable', 'flextable, and 'DT'.
##
##   You can always call the 'config_modelsummary()' function to change the
##   default table-drawing package in persistent fashion. To try 'tinytable'
##   now:
##
##   config_modelsummary(factory_default = 'tinytable')
##
##   To set the default back to 'kableExtra':
##
##   config_modelsummary(factory_default = 'kableExtra')
```

```
datasummary(tenure ~ Mean*(Gender=gender)+Mean*(`Senior Citizen`= SeniorCitizen)+Mean*(Partner), data=Ch
```

3) Find the simple average of `tenure` and `MonthlyCharge` by `Contract` type.

```
datasummary((`Tenure`=tenure)+(`Monthly Charge`= MonthlyCharges)~Mean*Contract, data=Churn)
```

4) Estimate a Kaplan Meier survival model. Use gender and senior citizen as explanatory variables. Do these variables produce statistically different survival rates?

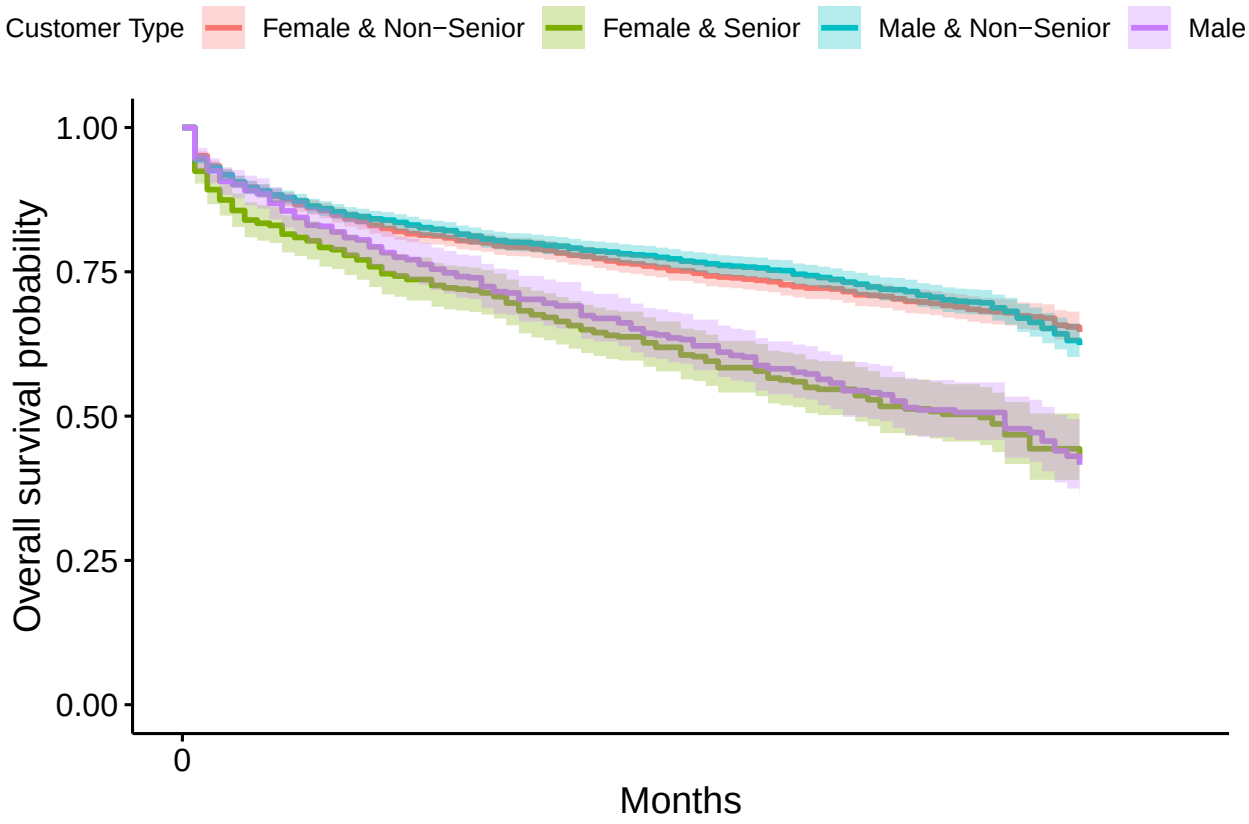
```
Churn$delta <-ifelse(Churn$Churn=="Yes", 1, 0)

survminer::ggsurvplot(
    fit = survival::survfit(survival::Surv(tenure, delta) ~ gender+SeniorCitizen, data = Churn),
    xlab = "Months",
    ylab = "Overall survival probability",
    legend.title = "Customer Type",
```

	Gender		Senior Citizen		Partner	
	Female	Male	Non-Senior	Senior	No	Yes
tenure	32.24	32.50	32.19	33.30	23.36	42.02

	Month-to-month	One year	Two year
Tenure	18.04	42.04	56.74
Monthly Charge	66.40	65.05	60.77

```
    legend.labs = c("Female & Non-Senior", "Female & Senior", "Male & Non-Senior","Male & Senior"),
    break.x.by = 100,
    censor = FALSE,
    conf.int = TRUE)
```



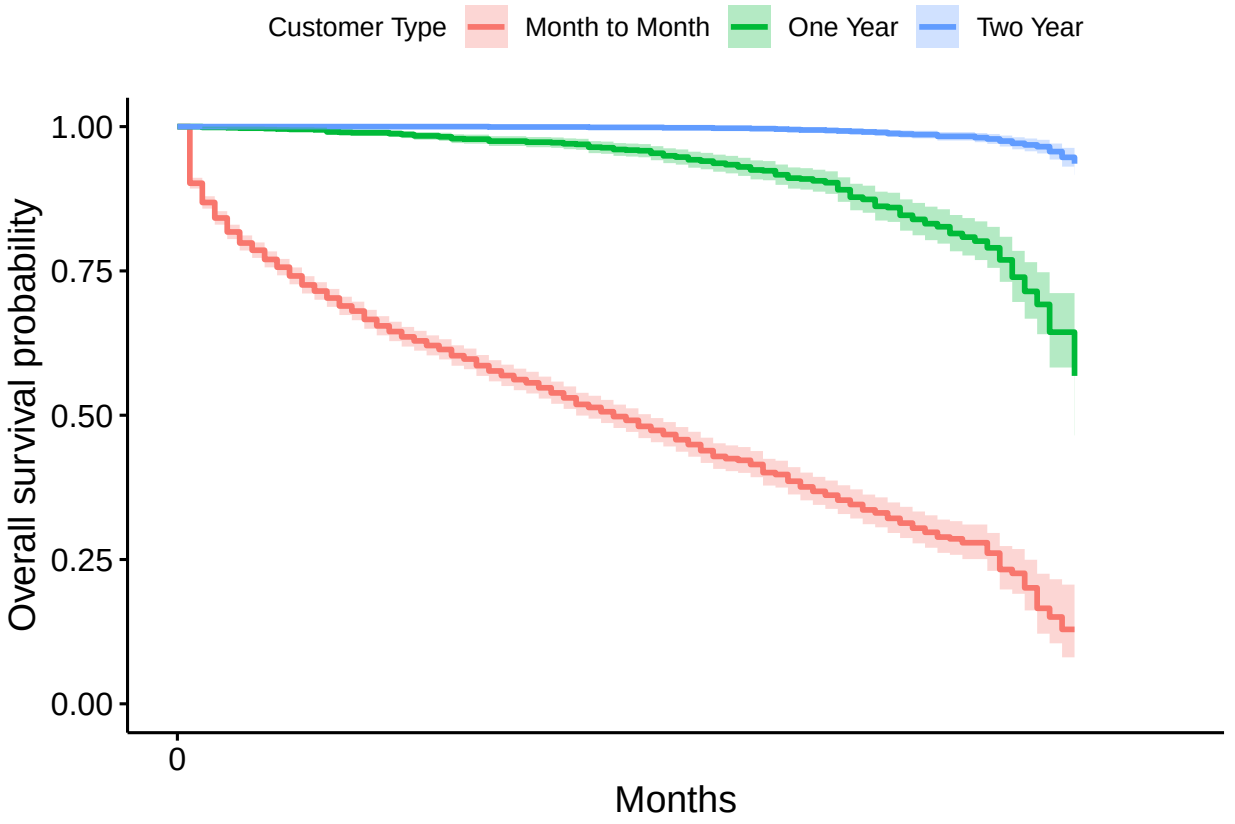
Answer: Yes, but they do not specifically differ in regards to gender. They differ in regards to if they are a Senior or Non-Senior. Both Male and Female Non-Seniors are at about the same survival probability rate at around .69 while Male & Female Seniors at at a much lower rate of about .50. As the line graph ends, both drop lower and lower. However, the Seniors of both genders drop at a much lower rate.

5) Estimate a Kaplan Meier survival model that uses contract type as the explanatory variable. Do we see a difference between contract type?

```
Churn$delta <-ifelse(Churn$Churn=="Yes", 1, 0)

survminer::ggsurvplot(
    fit = survival::survfit(survival::Surv(tenure, delta) ~ Contract, data = Churn),
    xlab = "Months",
    ylab = "Overall survival probability",
    legend.title = "Customer Type",
```

```
    legend.labs = c("Month to Month", "One Year", "Two Year"),
    break.x.by = 100,
    censor = FALSE,
    conf.int = TRUE)
```



Answer: Yes, especially with the Month-To-Month contract type. Within this contract type, it instantly drops within the first month, gradually gets lower & ends at around a probability of .15. The One-Year contract type drops mildly lower as each month accrues but drops towards the very end month(s) at around .60. The Two-Year contract type has no difference in decline until they very end month(s) in the upper .90's.

6) Estimate a Cox proportional hazard model of `tenure`. Use the following variables as explanatory variables: gender, Senior Citizen, contract type, partner, dependents, type of internet access, do they have phone service, and do they use paperless billing.

```
library(broom)
library(survival)

model_cox <- survival::coxph(
  survival::Surv(tenure, delta) ~ gender + SeniorCitizen + Contract + Partner + Dependents + InternetSe
  data = Churn)

tidy_broom <- broom::tidy(model_cox)

coef_names <- c("Male", "Senior Citizen", "One-Year Contract", "Two-Year Contract", "Yes To Partner", "
```

```r
tidy_broom$term <- coef_names

colnames(tidy_broom)[2:5] <- c("Estimate", "Standard Error", "Statistic", "P-Value")

print(tidy_broom)
```

```
## # A tibble: 10 x 5
##    term                    Estimate `Standard Error` Statistic `P-Value`
##    <chr>                      <dbl>            <dbl>     <dbl>     <dbl>
##  1 Male                     -0.0443           0.0463    -0.956 3.39e-  1
##  2 Senior Citizen           -0.0626           0.0558    -1.12  2.62e-  1
##  3 One-Year Contract        -2.03             0.0850   -23.9   8.42e-126
##  4 Two-Year Contract        -3.90             0.160    -24.4   1.36e-131
##  5 Yes To Partner           -0.611            0.0543   -11.3   2.25e- 29
##  6 Yes To Dependents        -0.0946           0.0681    -1.39  1.65e-  1
##  7 Optic Internet Service    0.360            0.0678     5.32  1.05e-  7
##  8 No To Internet Service   -0.225            0.112     -2.00  4.53e-  2
##  9 Yes To Phone Service     -0.123            0.0974    -1.26  2.08e-  1
## 10 Yes To Paperless Billing  0.123            0.0561     2.20  2.80e-  2
```