ICA 5

Kwaku Atuahene

2024-24-10

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.

The average tenure of a customer at this bank is 32.37 months. The simple average of `tenure` might be biased since it might not take into account that some observations have been censored.

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

```
datasummary(tenure~Mean*gender+Mean*SeniorCitizen+Mean*Partner, data = bank)
```

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

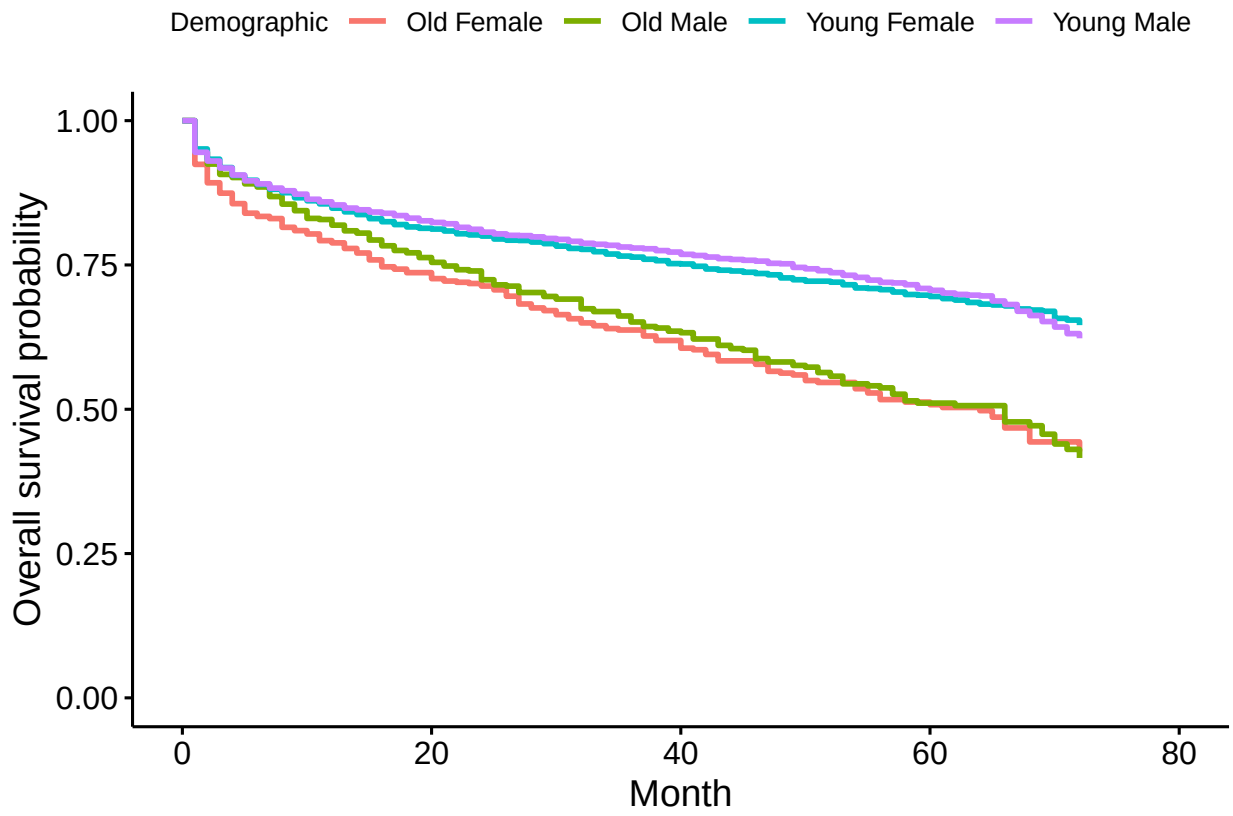
```
datasummary(tenure + MonthlyCharges ~Mean*Contract, data = bank)
```

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

	gender		SeniorCitizen		Partner	
	Female	Male	No	Yes	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
MonthlyCharges	66.40	65.05	60.77

```
bank$Demographic <- "Young Female"
bank$Demographic[bank$gender=="Male" & bank$SeniorCitizen=="No"]<- "Young Male"
bank$Demographic[bank$gender=="Male" & bank$SeniorCitizen=="Yes"]<- "Old Male"
bank$Demographic[bank$gender=="Female" & bank$SeniorCitizen=="Yes"]<- "Old Female"
survminer::ggsurvplot(
    fit = survival::survfit(survival::Surv(tenure, delta) ~ Demographic, data = bank),
    xlab = "Month",
    ylab = "Overall survival probability",
    legend.title = "Demographic",
    legend.labs = c("Old Female"="Old Female","Old Male"="Old Male","Young Female"="Young Female", "Your
    censor = FALSE)
```



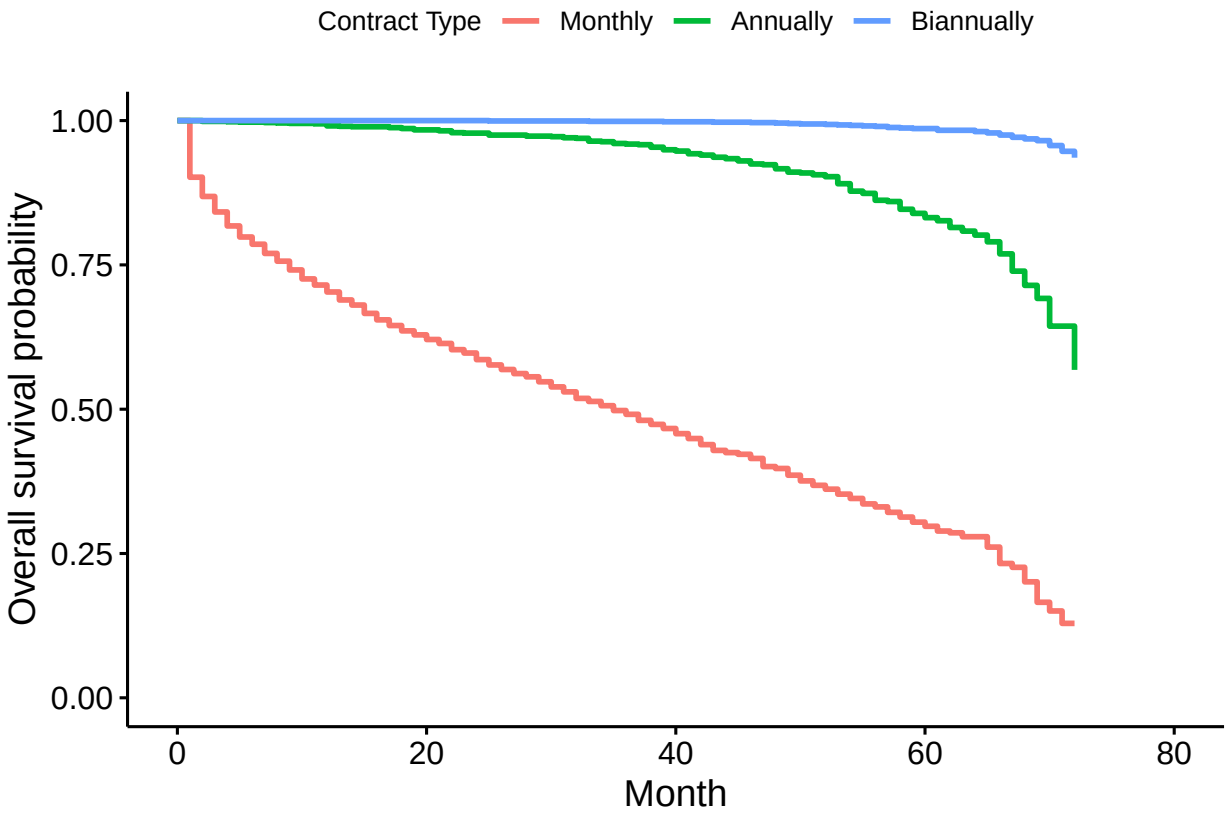

```
bank$Demographic <- NULL
```

Yes, there appears to be a statistically different survival rates across these variables particularly Senior Citizenship Status. In the beginning there isn't much of a difference but as the months go by the survival probabilities of Senior Citizens drop deeper than the Non-Senior Citizens. This is probably because the Senior Citizens are lost more rapidly due to death. In terms of gender, the difference is more pronounced in-group (i.e. within Senior Citizenship Status) but it not as huge as that of the Senior Citizenship Status.

We can simply say that Males have slightly higher survival probabilities than females more pronounced amongst Senior citizens than amongst the younger generation.

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

```
survminer::ggsurvplot(
    fit = survival::survfit(survival::Surv(tenure, delta) ~ Contract, data = bank),
    xlab = "Month",
    ylab = "Overall survival probability",
    legend.title = "Contract Type",
    legend.labs = c("Month-to-month"="Monthly", "One year"="Annually", "Two year"="Biannually"),
    censor = FALSE)
```



Yes, there's a difference between Contract types. We can see that longer Contract types have higher survival probabilities. This is probably due to the fact that there is more to lose in the a longer contractual agreement than a shorter contractual agreement, so customers tend to stay through the period unless they are willing to forgo their financial commitment.

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

```
f1<-survival::coxph(survival::Surv(tenure, delta) ~ gender+SeniorCitizen+Dependents+PhoneService +Inter
modelsummary(list("Cox Proportional Hazard Model"=f1), coef_rename = c("genderMale"="Gender (Male)","Se
```

	Cox Proportional Hazard Model
Gender (Male)	−0.044
	(0.046)
Senior Citizen (Yes)	−0.063
	(0.056)
Dependents (Yes)	−0.095
	(0.068)
Phone Service (Yes)	−0.123
	(0.097)
Internet Service (Fiber Optic)	0.360***
	(0.068)
Internet Service (No)	−0.225*
	(0.112)
Paperless Billing (Yes)	0.123*
	(0.056)
Contract type (One Year)	−2.029***
	(0.085)
Contract type (Two Year)	−3.902***
	(0.160)
Partner (Yes)	−0.611***
	(0.054)
Num.Obs.	7043
AIC	28 462.5
BIC	28 531.1
RMSE	0.54

+ p <0.1, * p <0.05, ** p <0.01, *** p <0.001

Yes they have Phone Service & Paperless Billing; those with Phone Service are less likely to leave as compared to those without and the opposite can be said about those with Paperless Billing. Having Phone Service is not statistically different (significant) from not having Phone Service in this model whilst on the other hand having Paperless Billing is statistically different from not having. Whether they have No Internet or Fiber Optic Internet is very important (statistically significant) and the length of their contractual agreement as well; with longer periods being more statistically significant and less likely to leave. We also see that Partners are statistically significant to this model and are less likely to leave as compared to singles. Gender, Senior Citizenship Status & Dependents Status don't seem to be statistically significant.