

# DS 223: Marketing Analytics

## Homework 3 - Survival Analysis

Alexander Shahramanyan

April 28, 2024

```
# Read the CSV file
telco <- read.csv("telco.csv")
telco$churn = ifelse(telco$churn=='Yes', 1, 0)

head(telco)
```

```
##   ID region tenure age  marital address income          ed
## 1  1 Zone 2     13  44   Married      9     64      College degree
## 2  2 Zone 3     11  33   Married      7    136  Post-undergraduate degree
## 3  3 Zone 3     68  52   Married     24    116 Did not complete high school
## 4  4 Zone 2     33  33 Unmarried     12     33      High school degree
## 5  5 Zone 2     23  30   Married      9     30 Did not complete high school
## 6  6 Zone 2     41  39 Unmarried     17     78      High school degree
##  retire gender voice internet forward  custcat churn
## 1      No   Male    No        No      Yes Basic service      1
## 2      No   Male   Yes        No      Yes Total service      1
## 3      No  Female   No        No      No  Plus service      0
## 4      No  Female   No        No      No  Basic service      1
## 5      No   Male   No        No      Yes  Plus service      0
## 6      No  Female   No        No      No  Plus service      0
```

```
surv_obj = Surv(time=telco$tenure, event=telco$churn)
reg_models <- list()

for(distribution in names(survreg.distributions)){
  # get the regression model
  reg_m = survreg(formula=surv_obj~1, dist=distribution)

  # print the summary
  # summary(reg_m)

  # add reg_m to reg_models
  reg_models[[distribution]] <- reg_m
}
```

```
# Initialize an empty list for storing plots
plot_list <- list()

for (distribution in names(survreg.distributions)) {
  reg_m <- reg_models[[distribution]]

  probs <- seq(.1, .9, length=9)
  pred <- predict(reg_m, type="quantile", p=1-probs, newdata=data.frame(1))

  df <- data.frame(Time=pred, Probabilities=probs)
```

```

# Generate the plot for current distribution
p <- ggplot(df, aes(x = Time, y = Probabilities)) +
  geom_line() +
  labs(title = paste("Survival Curve |", distribution),
        x = "Survival Time",
        y = "Survival Probability")

# Store the plot in the list
plot_list[[distribution]] <- p
}

# Combine the plots into a grid (4x3) and leave the last two positions blank
plot_grid <- wrap_plots(plot_list, nrow = 4, ncol = 3) +
  plot_spacer() + plot_spacer()

# Print the combined plot grid
print(plot_grid)

```

