

Arm_Guy_150Project_Assign1_due_Apr19

Arm and Guy

April 19, 2021

Assignment Instructions

- Include the names of the individuals in the group.
- Some sort of EDA (exploratory data analysis). Do not print the data, but do something that indicates you've uploaded the data and know what some of the variables are. You might have some summary statistics or a graph. This is not an extensive assignment.
- Outline the “something new” part of the assignment. You should indicate who is doing what, what resources each of you will use to learn about your new topic, and a few sentences on what the topic is or how it relates to survival analysis / the analysis at hand. Additionally, for each “new” thing, provide 1-2 sentences describing what will be challenging about learning something new

1. Names of Group Members

- a. Arm Wonghirundacha
- b. Guy Thampakkul

2. Explnatory Data Analysis and Visuals

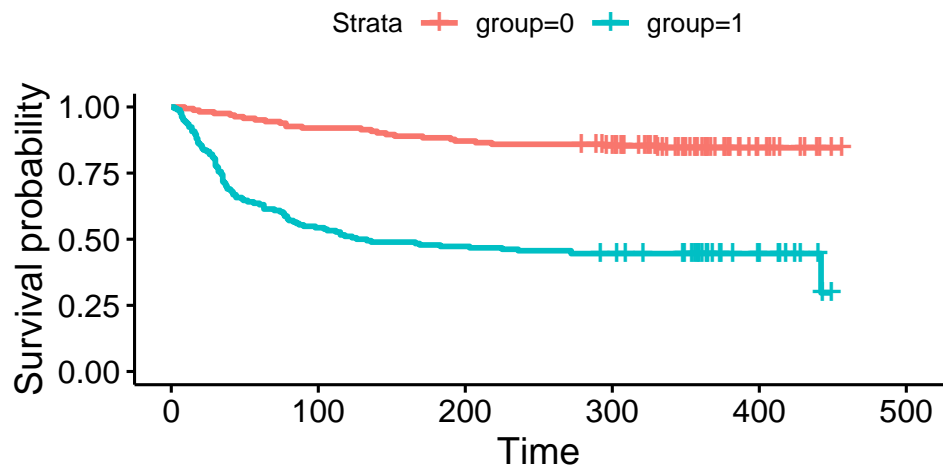
```
urlfile="https://raw.githubusercontent.com/gthampak/Arm_Guy_MATH150_Project/main/HELPdata.csv"
```

```
HELPdata <- read_csv(url(urlfile))
```

```
HELPdata_survfit <- survfit(Surv(dayslink, linkstatus) ~ group, data=HELPdata)
```

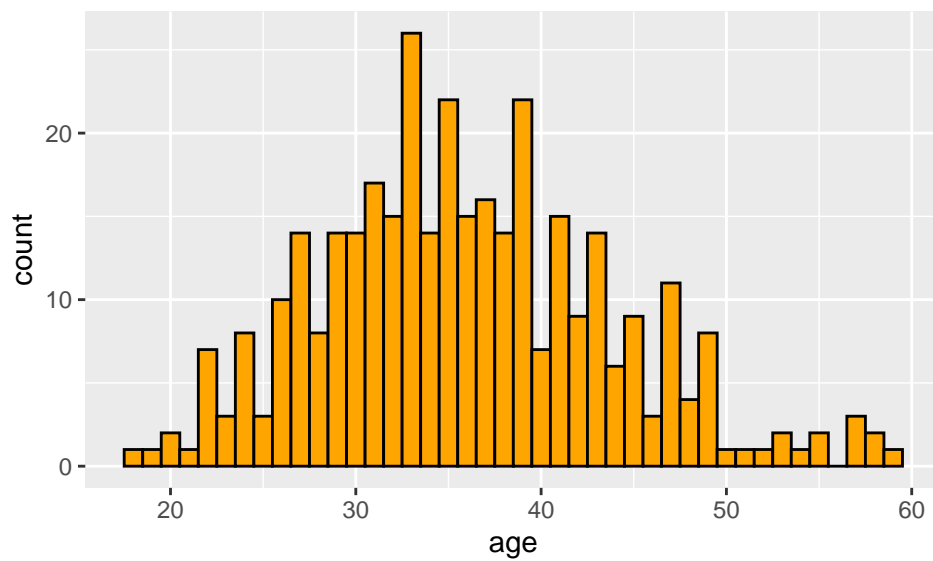
```
ggsurvplot(HELPdata_survfit, conf.type = "TRUE") +  
  ggtitle("KM-curve")
```

KM-curve



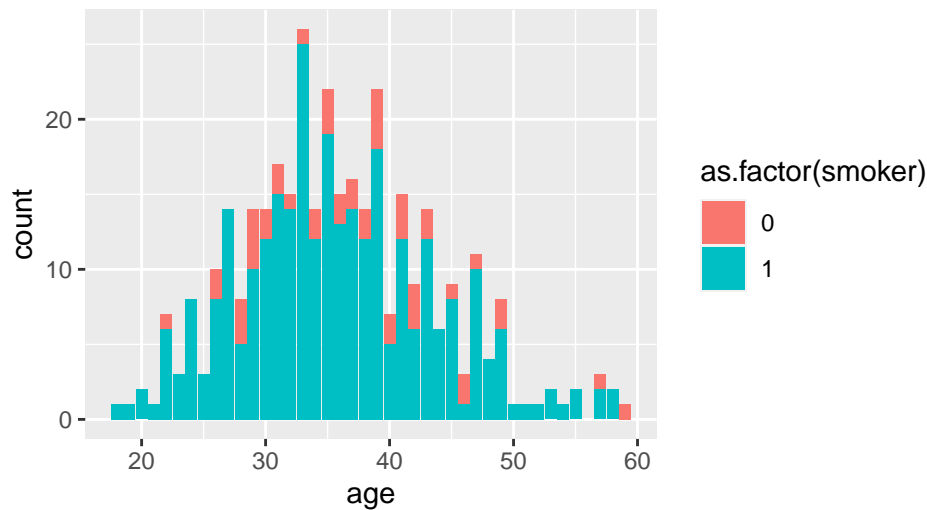
#histogram of age

```
ggplot(HELPdata, aes(x=age, fill = as.factor(smoker))) +  
  geom_histogram(binwidth = 1, color="black", fill="orange")
```



```
ggplot(data=HELPdata) +  
  geom_bar(mapping=aes(x=age, fill=as.factor(smoker))) +  
  ggtitle("Histogram of Patients' Age and Smoking")
```

Histogram of Patients' Age and Smoking



```
table(HELPdata$abuse)
```

```
##
##    0    1    2
## 101 180   64
```

3. Something New Outline

Topics:

- Investigation of the proportional hazards assumption (what does the R function `cox.zph` do?)
- Exponential or Weibull PH regression (parametric survival model)
- Deriving / detailing AIC & BIC for model selection on Cox PH
- Power analysis (a simulation?)
- Derivation of the sample size calculation for the log rank test (and application to the data)
- An analysis of the Schoenfeld residuals (how are they calculated and why is that calculation relevant?)
- Bootstrapping the survival model (what are the assumptions? what do you conclude?)
- An analysis of possible time dependent covariates (do transformations help?)

<https://cran.r-project.org/web/packages/survival/vignettes/timedep.pdf>

- An analysis / understanding / simulation of the multiple comparisons issues for assessing many different models (or other exploratory hypotheses).
- Andersen-Gill extension of the Cox PH model for time-varying covariates (available in `rms` R package).
- Another topic related to survival analysis that you find interesting.

Arm -

Guy -