PSTAT-175-Final-Project

William Nelson, Cole Spyksma, Alicia Qu

2025-06-13

```
library(readr)
library(tidyverse)
library(survival)
library(survminer)
library(ggplot2)
library(knitr)
```

```
# Read in data
GRACE1000 <- read_table("GRACE1000.dat", col_names = FALSE)
#GRACE1000 <- read_table("~/Downloads/GRACE1000.dat", col_names = FALSE)
GRACE1000 <- GRACE1000 %>% select(-X10)
colnames(GRACE1000) <- c("id", "days", "death", "revasc", "revascdays", "los", "age", "sysbp", "stchange
GRACE1000 <- GRACE1000 %>% mutate(
  revascdays = case_when(
    revascdays == 0 ~ 0.1,
    TRUE ~ revascdays
  )
)
```

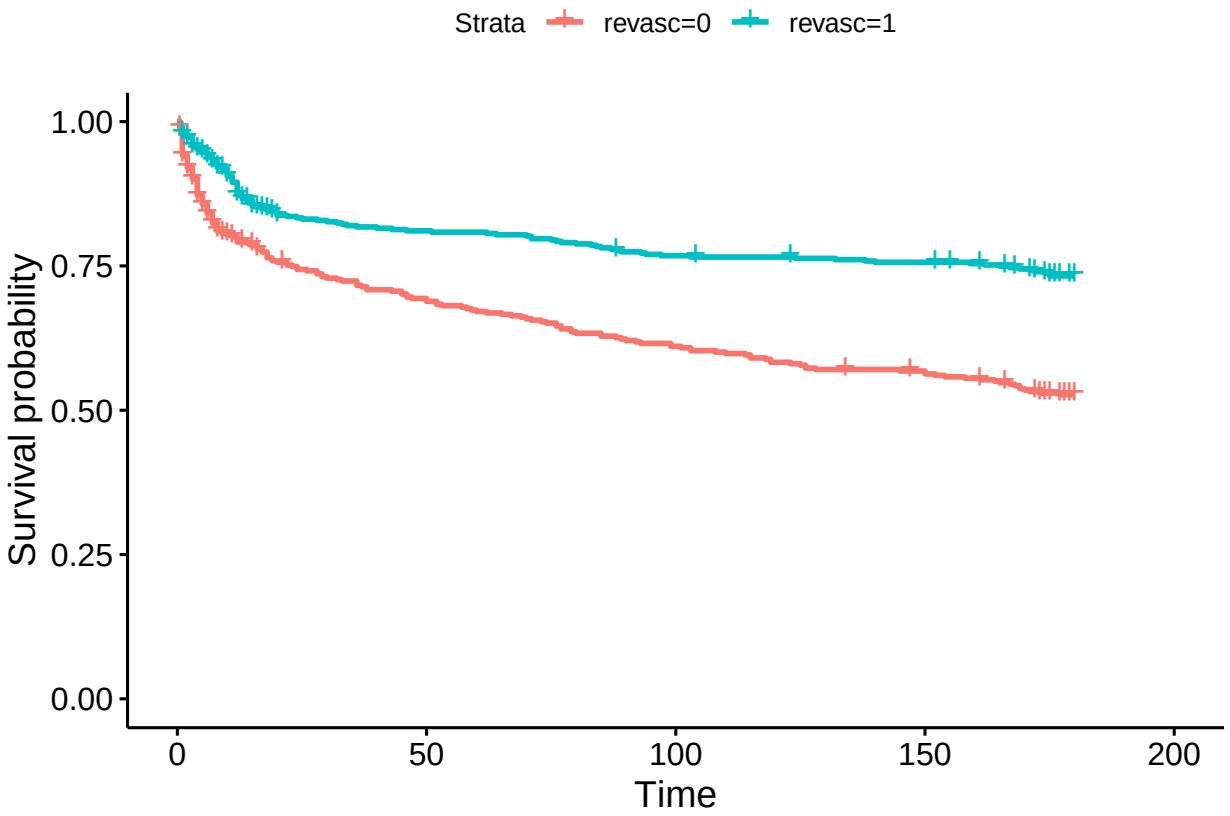
Final Project

1: Introduction

We will be analyzing the "GRACE1000" dataset from Hosmer and Lemenshow and May (2008), which contains data on 1000 patients who were part of a study on the effectiveness of revascularization by the Global Registry of Acute Coronary Events (GRACE). Our failure time of interest is the follow up time when the researchers checked-in with the patients, and our event of interest is whether the patient died during the follow-up period. Our main covariate of interest will be whether the patient had the revascularization procedure performed on them, which is coded by the "revasc" variable with value 1 if the patient had the procedure and 0 if they did not. The main research question of interest is whether revascularization is associated with higher survival rates after completing the procedure. We hypothesize that patients who receive the revascularization procedure will have higher survival rates than patients without the procedure done.

Here is a plot of the survival functions of patients who had the revascularization procedure in blue and patients who did not have the procedure done in red:

```
ggsurvplot(survfit(Surv(days,death) ~ revasc, data = GRACE1000))
```



We can see that at first glance, patients who had revascularization done appear to have a higher survival probability compared to patients who did not have the procedure done. We will further explore this data with the tools we have learned:

2: Model Fitting

For this project, we plan to use AIC and BIC as model selection criteria to determine the most appropriate predictors for survival outcomes in the GRACE1000 dataset. Since this dataset contains multiple potential predictors, including age, systolic blood pressure, ST changes, and revascularization status, it's essential to balance model complexity with explanatory power. AIC and BIC provide objective measures to compare different Cox proportional hazards models, where AIC focuses more on prediction and BIC favors simpler, more interpretable models. This approach ensures that the final model explaining the time to death in relation to revascularization is both statistically robust and practically meaningful.

```
# Forward Stepwise Selection
# Possible covariates are revasc, sysbp, age, stchange

age.mod <- coxph(Surv(days,death) ~ age, data = GRACE1000)
sysbp.mod <- coxph(Surv(days,death) ~ sysbp, data = GRACE1000)
stchange.mod <- coxph(Surv(days,death) ~ stchange, data = GRACE1000)
revasc.mod <- coxph(Surv(days,death) ~ revasc, data = GRACE1000)

kable(AIC(age.mod, sysbp.mod, stchange.mod, revasc.mod))
```

	df	AIC
age.mod	1	4151.309
sysbp.mod	1	4245.999
stchange.mod	1	4252.286
revasc.mod	1	4231.999

```
kable(BIC(age.mod, sysbp.mod, stchange.mod, revasc.mod))
```

	df	BIC
age.mod	1	4155.090
sysbp.mod	1	4249.780
stchange.mod	1	4256.067
revasc.mod	1	4235.780

Upon analyzing the Cox proportional hazards models, the results indicate that age alone provides the best balance between model fit and simplicity, as evidenced by the lowest AIC value of 4151.309 and BIC value of 4155.090. In comparison, models using systolic blood pressure (AIC = 4245.999), ST changes (AIC = 4252.286), and revascularization status (AIC = 4231.999) each demonstrate higher AIC values, suggesting that these individual predictors are less informative in explaining survival outcomes. The differences between AIC values are considerable, with the age model clearly outperforming others by a margin exceeding 4-10 AIC units, reinforcing the strength of age as a predictor.

```
age.sysbp.mod <- coxph(Surv(days,death) ~ age + sysbp, data = GRACE1000)
age.stchange.mod <- coxph(Surv(days,death) ~ age + stchange, data = GRACE1000)
age.revasc.mod <- coxph(Surv(days,death) ~ age + revasc, data = GRACE1000)

kable(AIC(age.sysbp.mod, age.stchange.mod, age.revasc.mod))
```

	df	AIC
age.sysbp.mod	2	4131.179
age.stchange.mod	2	4136.352

	df	AIC
age.revasc.mod	2	4136.017

```
kable(BIC(age.sysbp.mod, age.stchange.mod, age.revasc.mod))
```

	df	BIC
age.sysbp.mod	2	4138.741
age.stchange.mod	2	4143.913
age.revasc.mod	2	4143.579

In this analysis, we compared three multivariable Cox models incorporating age with additional predictors: systolic blood pressure, ST change, and revascularization status. The model that included age and systolic blood pressure yielded the lowest AIC value (4131.179), indicating the best fit among the multivariable models tested. Models incorporating age with ST change (AIC = 4136.352) and age with revascularization (AIC = 4136.017) had slightly higher AIC values, suggesting marginally inferior fits. The consistently low AIC values across all models reinforce the importance of age as a key predictor while indicating that including systolic blood pressure may offer a slight improvement over the other two variables when combined with age.

```
sysbp.age.stchange.mod <- coxph(Surv(days,death) ~ sysbp + age + stchange, data = GRACE1000)
sysbp.age.revasc.mod <- coxph(Surv(days,death) ~ sysbp + age + revasc, data = GRACE1000)

kable(AIC(sysbp.age.stchange.mod, sysbp.age.revasc.mod))
```

	df	AIC
sysbp.age.stchange.mod	3	4117.788
sysbp.age.revasc.mod	3	4116.536

```
kable(BIC(sysbp.age.stchange.mod, sysbp.age.revasc.mod))
```

	df	BIC
sysbp.age.stchange.mod	3	4129.130
sysbp.age.revasc.mod	3	4127.879

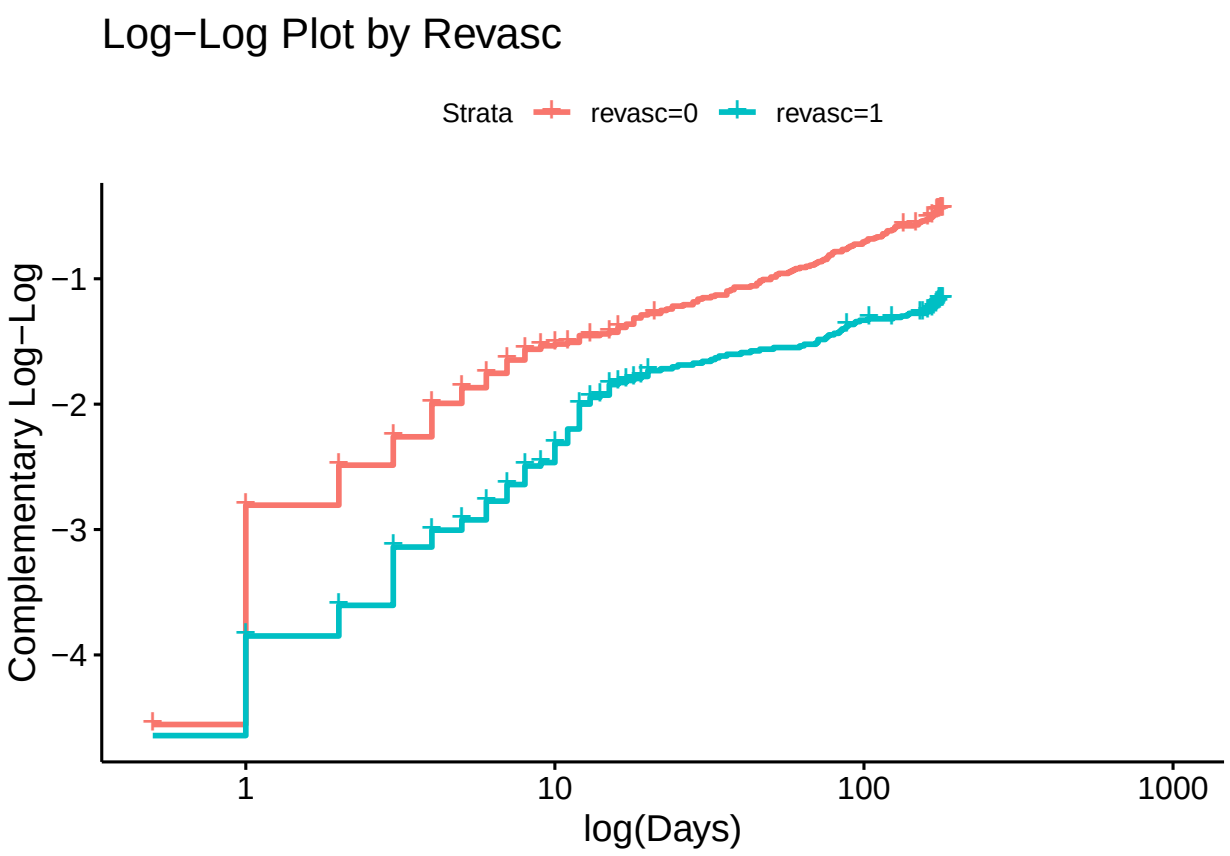
```
# See if stchange lowers AIC and BIC
all.mod <- coxph(Surv(days,death) ~ stchange + sysbp + revasc + age, data = GRACE1000)
kable(data.frame(Metric = c("AIC", "BIC"), Value = c(AIC(all.mod), BIC(all.mod))))
```

Metric	Value
AIC	4100.721
BIC	4115.844

After fitting a comprehensive Cox proportional hazards model incorporating all considered predictors: ST change, systolic blood pressure, revascularization, and age, the resulting AIC value was 4100.721 and BIC was 4115.844, both lower than those from the previous models. This indicates that including all variables yields the best-fitting model based on forward stepwise selection criteria. These results suggest that while age alone and age in combination with other predictors offer substantial predictive power, the full model provides the most accurate explanation of survival probability in this dataset, balancing model complexity with fit.

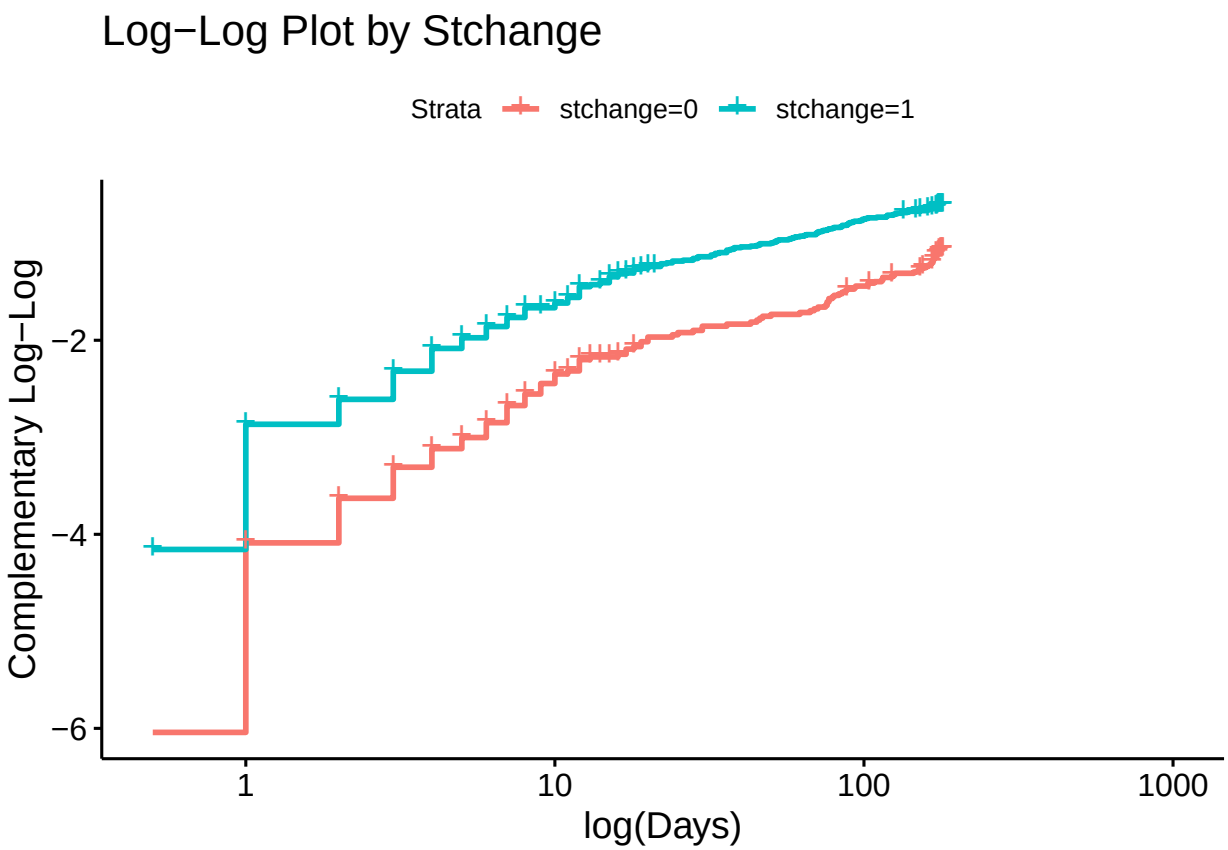
3: Check Proportional Hazards Assumptions

```
ggsurvplot(survfit(Surv(days,death) ~ revasc, data = GRACE1000),
           fun = "cloglog") +
  labs(x = "log(Days)", y = "Complementary Log-Log",
       title = "Log-Log Plot by Revasc")
```



In this Complementary Log-Log plot of the treatment and control groups, the curves for patients who underwent revascularization (revasc=1) and those who did not (revasc=0) are presented. The relatively parallel nature of the curves suggests that the proportional hazards assumption holds reasonably well for the revascularization variable, as the log-log curves for the two groups do not cross and maintain a fairly consistent vertical separation throughout the range of time. Overall, the log-log plot supports the use of the Cox proportional hazards model with revascularization as a predictor

```
ggsurvplot(survfit(Surv(days,death) ~ stchange, data = GRACE1000),
           fun = "cloglog") +
  labs(x = "log(Days)", y = "Complementary Log-Log",
       title = "Log-Log Plot by Stchange")
```




```
# also very parallel lines, no assumptions violated
```

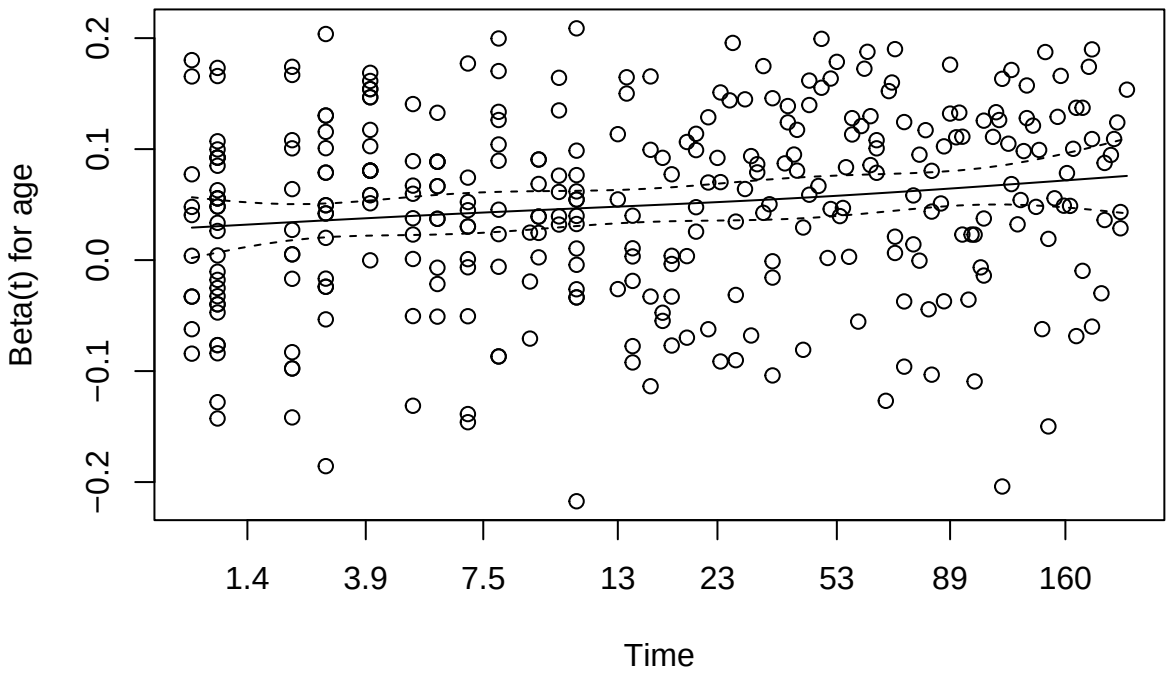
The Complementary Log-Log plot for Stchange also displays curves that do not cross into higher time values, meaning the proportional hazards assumption also holds here with the different Stchange values.

ZPH Plot for Age:

```
kable(as.data.frame(cox.zph(age.mod)$table))
```

	chisq	df	p
age	7.758003	1	0.0053475
GLOBAL	7.758003	1	0.0053475

```
plot(cox.zph(age.mod))
```



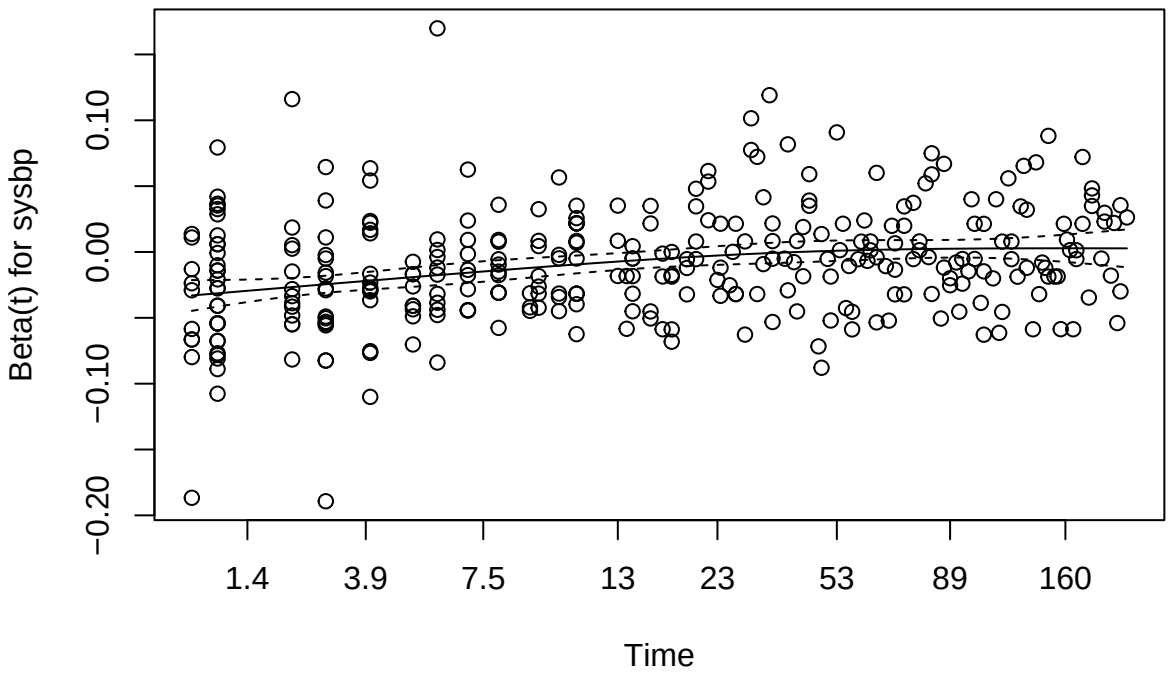
The statistical ouput shows a p-value of 0.0053 for both age and the global test, which is highly significant ($p < 0.05$). This indicates a violation of the proportional hazards assumption for age, meaning that the effect of age on the hazard is not constant over time. This zph plot visualizes Schoenfeld residuals for age over time in the Cox model. The residuals appear randomly scattered around the horizontal axis, and the smoothed curve does not display a clear trend. This pattern suggests that the proportional hazards assumption for age holds reasonably well. Although we can't see it in the graph, with the p-value being that low we will still assume the ph assumptions are being violated.

ZPH Plot for Systolic Blood Pressure:

```
kable(as.data.frame(cox.zph(sysbp.mod)$table))
```

	chisq	df	p
sysbp	33.18016	1	0
GLOBAL	33.18016	1	0

```
plot(cox.zph(sysbp.mod))
```



Here we see Systolic Blood Pressure also violating the proportional hazards assumption with a p_value of 8.4e-09. Once again it doesn't appear to have any clear correlation in the graph but with a p_value that low we can not ignore that there is a clear time-varying effect in the model.

4: Conclusions

Our main scientific question of interest is whether the revascularization procedure significantly increases patients' survival probabilities, and while we have a general idea, we can take a look at the results of our models:

```
summary(revasc.mod)
```

```
## Call:
## coxph(formula = Surv(days, death) ~ revasc, data = GRACE1000)
##
##   n= 1000, number of events= 324
##
##           coef exp(coef) se(coef)      z Pr(>|z|)
## revasc -0.7149    0.4892   0.1144 -6.249 4.13e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##        exp(coef) exp(-coef) lower .95 upper .95
## revasc    0.4892      2.044     0.391    0.6122
##
## Concordance= 0.589  (se = 0.014 )
## Likelihood ratio test= 40.46  on 1 df,   p=2e-10
```

```
## Wald test            = 39.05  on 1 df,   p=4e-10
## Score (logrank) test = 40.73  on 1 df,   p=2e-10
```

```
exp(coef(revasc.mod) + c(-1.96,1.96) * sqrt(revasc.mod$var[1,1]))
```

```
## [1] 0.3909648 0.6122096
```

```
exp(confint(revasc.mod))
```

```
##            2.5 %    97.5 %
## revasc 0.3909664 0.6122071
```

The summary of the revasc.mod Cox model indicates that revascularization is a highly significant predictor of survival ($p < 0.0001$), with a hazard ratio of 0.4892 and a 95% confidence interval of [0.391, 0.612]. This means that patients who underwent revascularization were approximately 51% less likely to die compared to those who did not undergo the procedure. The hazard ratio being less than 1 signifies a protective effect of revascularization.

```
summary(stchange.mod)
```

```
## Call:
## coxph(formula = Surv(days, death) ~ stchange, data = GRACE1000)
##
##   n= 1000, number of events= 324
##
##            coef exp(coef) se(coef)    z Pr(>|z|)
## stchange 0.5189    1.6802   0.1185 4.38 1.19e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## stchange      1.68     0.5952     1.332     2.119
##
## Concordance= 0.57  (se = 0.013 )
## Likelihood ratio test= 20.17  on 1 df,   p=7e-06
## Wald test            = 19.19  on 1 df,   p=1e-05
## Score (logrank) test = 19.62  on 1 df,   p=9e-06
```

```
exp(coef(stchange.mod) + c(-1.96, 1.96) * sqrt(stchange.mod$var[1, 1]))
```

```
## [1] 1.332036 2.119339
```

```
exp(confint(stchange.mod))
```

```
##             2.5 %  97.5 %
## stchange 1.332041 2.11933
```

Patients with an ST change on their initial ECG have a 68% higher hazard of mortality compared to those without such deviations, after controlling for time. This is a clinically meaningful result, as it highlights that ST changes detected early in an ECG are strongly associated with worse survival outcomes. The statistical significance ($p < 0.00001$) and a confidence interval for the hazard ratio of [1.332, 2.119] (an interval greater than 1) suggest that this association is reliable.

5: Advanced Methods

Treating Revascularization as a Time-Dependent Covariate

When trying to answer the question of whether the revascularization procedure is associated with higher survival rates, we run into a problem with our data in that not all of our observations had the revascularization

procedure on day 0 or the first day they were a part of the study. This means that if a patient had the revascularization done on the 5th day (revascdays = 5), then our models are counting the first 5 days as this patient already having the procedure done, when in reality it has not happened until later. To fix this issue with the data, we will treat "revasc" as a binary time-dependent covariate that changes at some point in the study. To implement the time-varying covariate, we will split every observation that had the revascularization procedure done into 2 observations where the first observation has a start time of 0 and a stop time of the day they had the procedure done, and the second observation will have a start time of the day the procedure was done and a stop time of when the observation was censored. This way our models will properly understand that patients have not had the procedure done until the day given by "revascdays", and therefore are not a part of the treatment group until then.

We will manipulate the data to achieve this:

```
library(dplyr)
library(tidyr)

grace2 <- GRACE1000 %>%
  mutate(
    start_1 = 0,
    stop_1 = pmin(revascdays, days, na.rm = TRUE),
    status_1 = ifelse(days <= revascdays, death, 0),
    revasc_status_1 = 0,  # Renamed for clarity

    start_2 = revascdays,
    stop_2 = ifelse(revasc == 1, days, NA),
    status_2 = ifelse(revasc == 1 & days > revascdays, death, NA),
    revasc_status_2 = 1  # Renamed for clarity
  ) %>%
  pivot_longer(
    cols = c(start_1, stop_1, status_1, revasc_status_1, start_2, stop_2, status_2, revasc_status_2),
    names_to = c("variable", "interval"),
    names_pattern = "([a-z_]+)_([12])"
  ) %>%
  pivot_wider(
    names_from = variable,
    values_from = value
  ) %>%
  filter(!is.na(stop)) %>%
  arrange(id, interval) %>%
  filter(!(start==stop))

# Check output
kable(head(grace2))
```

id	days	death	revasc	revascdays	los	age	sysbp	stchange	interval	start	stop	status	revasc_status
1	180	0	0	180.0	9	28	107	1	1	0.0	180.0	0	0
2	5	0	1	0.1	5	32	121	1	1	0.0	0.1	0	0
2	5	0	1	0.1	5	32	121	1	2	0.1	5.0	0	1
3	2	0	0	2.0	2	33	150	0	1	0.0	2.0	0	0
4	5	0	1	2.0	5	35	172	0	1	0.0	2.0	0	0
4	5	0	1	2.0	5	35	172	0	2	2.0	5.0	0	1

With our data organized to represent revascularization status as a time-varying covariate, we can apply our best fitting model from above to our adjusted data and look at the results:

```
allmod2 <- coxph(Surv(start,stop,status) ~ stchange + sysbp + revasc + age, data = grace2)
summary(allmod2)
```

```
## Call:
## coxph(formula = Surv(start, stop, status) ~ stchange + sysbp +
##     revasc + age, data = grace2)
##
##   n= 1510, number of events= 319
##
##               coef exp(coef)  se(coef)      z Pr(>|z|)
## stchange  0.482253  1.619720  0.119681  4.029 5.59e-05 ***
## sysbp    -0.007979  0.992053  0.001928 -4.139 3.49e-05 ***
## revasc   -0.482493  0.617243  0.118606 -4.068 4.74e-05 ***
## age       0.042746  1.043673  0.004727  9.044  < 2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## stchange    1.6197     0.6174    1.2811    2.0479
## sysbp       0.9921     1.0080    0.9883    0.9958
## revasc      0.6172     1.6201    0.4892    0.7788
## age         1.0437     0.9582    1.0340    1.0534
##
## Concordance= 0.708  (se = 0.014 )
## Likelihood ratio test= 172.3  on 4 df,   p=<2e-16
## Wald test            = 168  on 4 df,   p=<2e-16
## Score (logrank) test = 175.4  on 4 df,   p=<2e-16
```

From the summary output of our model we can still see all our covariates have significant effects of the on the time till an event, with p-values much smaller than our significance level of 0.05. Our overall likelihood ratio test also reports an extremely small p-value which tells us that the covariates in the model are providing the necessary information to estimate time until an event. We can also look at the ANOVA output to check the significance of each covariate individually as they are added:

```
anova(allmod2)
```

```
## Analysis of Deviance Table
##  Cox model: response is Surv(start, stop, status)
## Terms added sequentially (first to last)
##
##           loglik  Chisq Df Pr(>|Chi|)
## NULL     -2100.7
## stchange -2091.0 19.339  1  1.094e-05 ***
## sysbp    -2080.6 20.884  1  4.878e-06 ***
## revasc   -2060.8 39.570  1  3.165e-10 ***
## age      -2014.5 92.493  1  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Similar to the summary output, we see that all four covariates have statistically significant p-values for their Chi-Squared log-likelihood test, and each of the four covariates provide significant information to the model.

Next, we want to make sure that our non-revasc covariates do not have a significant interaction with the interval variable - if they do, this would tell us that the effects of the covariate can be attributed to time. We will fit models with an interaction term between the interval signifier and each covariate, and analyze the results of these models:

```r
ageInteractMod <- coxph(Surv(start, stop, status) ~ interval * age, data = grace2)
summary(ageInteractMod)
```

```
## Call:
## coxph(formula = Surv(start, stop, status) ~ interval * age, data = grace2)
##
##   n= 1510, number of events= 319
##
##                    coef exp(coef)  se(coef)      z Pr(>|z|)
## interval2     -1.742213  0.175132  0.777551 -2.241    0.025 *
## age            0.041637  1.042516  0.005839  7.131    1e-12 ***
## interval2:age  0.021487  1.021720  0.010461  2.054    0.040 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##               exp(coef) exp(-coef) lower .95 upper .95
## interval2        0.1751     5.7100   0.03815    0.8039
## age              1.0425     0.9592   1.03065    1.0545
## interval2:age    1.0217     0.9787   1.00098    1.0429
##
## Concordance= 0.664  (se = 0.015 )
## Likelihood ratio test= 127.5  on 3 df,   p=<2e-16
## Wald test            = 108.8  on 3 df,   p=<2e-16
## Score (logrank) test = 116.9  on 3 df,   p=<2e-16
```

We see in this summary output that the age interaction model shows a significant interaction between interval and age, with a p-value of 0.04 just below our significance threshold. Because of the significance of the interaction, we conclude that the effect of the age covariate on survival time varies depending on the value of the interval, suggesting that the age covariate has some underlying time correlation and its inclusion in the model may be problematic.

```r
sysbpInteractMod <- coxph(Surv(start, stop, status) ~ interval * sysbp, data = grace2)
summary(sysbpInteractMod)
```

```
## Call:
## coxph(formula = Surv(start, stop, status) ~ interval * sysbp,
##     data = grace2)
##
##   n= 1510, number of events= 319
##
##                      coef exp(coef)  se(coef)      z Pr(>|z|)
## interval2       -1.367907  0.254639  0.599488 -2.282   0.0225 *
## sysbp           -0.012009  0.988063  0.002470 -4.861 1.17e-06 ***
## interval2:sysbp  0.007032  1.007056  0.004260  1.650   0.0988 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                 exp(coef) exp(-coef) lower .95 upper .95
## interval2          0.2546      3.927   0.07864    0.8245
## sysbp              0.9881      1.012   0.98329    0.9929
## interval2:sysbp    1.0071      0.993   0.99868    1.0155
##
## Concordance= 0.593  (se = 0.017 )
## Likelihood ratio test= 37.09  on 3 df,   p=4e-08
## Wald test            = 39.13  on 3 df,   p=2e-08
```

```
## Score (logrank) test = 39.05  on 3 df,   p=2e-08
```

The sysbp interaction model does not show a significant interaction between interval and sysbp, with a p-value of 0.0988. This p-value is above our threshold and leads us to conclude that there is no interaction effect between the sysbp covariate and the interval signifier.

```
stchangeInteractMod <- coxph(Surv(start, stop, status) ~ interval * stchange, data = grace2)
summary(stchangeInteractMod)
```

```
## Call:
## coxph(formula = Surv(start, stop, status) ~ interval * stchange,
##     data = grace2)
##
##   n= 1510, number of events= 319
##
##                         coef exp(coef) se(coef)      z Pr(>|z|)
## interval2          -0.46981   0.62512  0.21395 -2.196 0.028102 *
## stchange            0.51958   1.68132  0.14616  3.555 0.000378 ***
## interval2:stchange  0.05277   1.05418  0.25341  0.208 0.835053
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##                    exp(coef) exp(-coef) lower .95 upper .95
## interval2             0.6251     1.5997    0.4110    0.9508
## stchange              1.6813     0.5948    1.2625    2.2391
## interval2:stchange    1.0542     0.9486    0.6415    1.7323
##
## Concordance= 0.588  (se = 0.015 )
## Likelihood ratio test= 32.29  on 3 df,   p=5e-07
## Wald test            = 30.82  on 3 df,   p=9e-07
## Score (logrank) test = 32.09  on 3 df,   p=5e-07
```

The stchange interaction model does not show a significant interaction between interval and stchange, with a p-value of 0.835. This p-value is above our threshold and leads us to conclude that there is no interaction effect between the stchange covariate and the interval signifier.

Because of the significant interaction between age and interval in ageInteractMod, we fit a new model on the time-varying data without the age covariate.

```
allmod3 <- coxph(Surv(start,stop,status) ~ sysbp + stchange + revasc, data = grace2)
summary(allmod3)
```

```
## Call:
## coxph(formula = Surv(start, stop, status) ~ sysbp + stchange +
##     revasc, data = grace2)
##
##   n= 1510, number of events= 319
##
##                coef exp(coef)  se(coef)      z Pr(>|z|)
## sysbp    -0.009139  0.990903  0.002022 -4.520 6.18e-06 ***
## stchange  0.515455  1.674401  0.119413  4.317 1.58e-05 ***
## revasc   -0.712608  0.490364  0.115122 -6.190 6.02e-10 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
##          exp(coef) exp(-coef) lower .95 upper .95
## sysbp       0.9909     1.0092    0.9870    0.9948
```

```
## stchange    1.6744     0.5972    1.3250    2.1159
## revasc      0.4904     2.0393    0.3913    0.6145
##
## Concordance= 0.65  (se = 0.016 )
## Likelihood ratio test= 79.79  on 3 df,   p=<2e-16
## Wald test            = 76.85  on 3 df,   p=<2e-16
## Score (logrank) test = 78.66  on 3 df,   p=<2e-16
```

We can now calculate a 95% confidence interval for the hazard ratio between the treatment and control groups:

```
# hazard ratio for revasc
exp(coef(allmod3)[3] + c(-1.96, 1.96) * sqrt(allmod3$var[3, 3]))
```

```
## [1] 0.3913134 0.6144861
```

The time-varying model still indicates that revascularization is a highly significant predictor of survival ($p < 0.0001$), with a 95% confidence interval of the hazard ratio being [0.3913134, 0.6144861]. The hazard ratio being less than 1 signifies a protective effect of revascularization.

Conclusion

This analysis of the GRACE1000 dataset examined the relationship between revascularization and survival in patients with acute coronary events. Initial exploratory analysis, including Kaplan-Meier curves, suggested that patients undergoing revascularization had better survival probabilities compared to those who did not. Cox proportional hazards models, selected using AIC and BIC criteria, consistently highlighted age as a strong predictor of mortality, with lower AIC and BIC values compared to other covariates such as systolic blood pressure, ST changes, and revascularization status. However, a comprehensive model including all predictors provided the best overall fit, balancing complexity and explanatory power. The analysis of the proportional hazards assumption revealed that while age and systolic blood pressure violated the PH assumption, revascularization and ST changes appeared to satisfy it, supporting their use in Cox models.

Advanced modeling approaches addressed the time-dependent nature of revascularization by splitting patient records at the time of treatment. After accounting for the potential interaction between the effects of the other covariates and the time interval, the updated model still showed that revascularization significantly reduced the hazard of death ($HR < 1$, $p < 0.0001$), affirming its protective effect even after accounting for the timing of the procedure. This suggests that the survival benefit of revascularization was consistent regardless of when it was administered during hospitalization. In summary, this project demonstrates that revascularization is associated with a substantial and statistically significant improvement in survival, supporting its role as an effective treatment in acute coronary events.

References

Hosmer, D.W. and Lemeshow, S. and May, S. (2008) Applied Survival Analysis: Regression Modeling of Time to Event Data: Second Edition, John Wiley and Sons Inc., New York, NY

Wickham H (2016). *ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. ISBN 978-3-319-24277-4, https://ggplot2.tidyverse.org.

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemund G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). "Welcome to the tidyverse." *Journal of Open Source Software*, 4(43), 1686. doi:10.21105/joss.01686.

Kassambara A, Kosinski M, Biecek P (2024). *survminer: Drawing Survival Curves using 'ggplot2'*. R package version 0.5.0, https://rpkgs.datanovia.com/survminer/index.html.

Wickham H, Hester J, Bryan J (2024). *readr: Read Rectangular Text Data*. R package version 2.1.5, https://github.com/tidyverse/readr, https://readr.tidyverse.org.

Terry M. Therneau, Patricia M. Grambsch (2000). *Modeling Survival Data: Extending the Cox Model*. Springer, New York. ISBN 0-387-98784-3.

Xie Y (2025). *knitr: A General-Purpose Package for Dynamic Report Generation in R*. R package version 1.50, https://yihui.org/knitr/.