proj1

Jiayi Shi

2025-02-19

library(survival)  
library(survminer)  
  
data <- read.table("menopause.dat", header=FALSE, col.names=c("id", "intake\_age", "menopause\_age", "menopause", "race", "education"))  
  
data <- data %>% mutate(  
 menopause = factor(menopause, labels = c("Censored", "Menopause")),  
 race = factor(race, labels = c("White non-Hispanic", "Black non-Hispanic", "Other Ethnicity")),  
 education = factor(education, labels = c("Post-graduate", "College Graduate", "Some College", "High School or less")),  
 menopause\_time = menopause\_age - intake\_age  
)  
  
label(data$intake\_age) <- "Intake Age (years)"  
label(data$menopause\_age) <- "Menopause Age (years)"  
label(data$menopause) <- "Menopause Status"  
label(data$race) <- "Ethnicity"  
label(data$education) <- "Education Level"  
  
table1(~ intake\_age+menopause\_age+race+education | menopause, data, overall = F)

|  | Censored (N=305) | Menopause (N=75) |
| --- | --- | --- |
| **Intake Age (years)** |  |  |
| Mean (SD) | 47.4 (2.44) | 49.7 (2.62) |
| Median [Min, Max] | 46.9 [44.4, 59.6] | 49.4 [45.6, 55.8] |
| **Menopause Age (years)** |  |  |
| Mean (SD) | 51.2 (2.43) | 51.9 (2.57) |
| Median [Min, Max] | 50.8 [44.7, 64.3] | 51.9 [47.3, 56.7] |
| **Ethnicity** |  |  |
| White non-Hispanic | 248 (81.3%) | 56 (74.7%) |
| Black non-Hispanic | 24 (7.9%) | 13 (17.3%) |
| Other Ethnicity | 33 (10.8%) | 6 (8.0%) |
| **Education Level** |  |  |
| Post-graduate | 132 (43.3%) | 35 (46.7%) |
| College Graduate | 81 (26.6%) | 15 (20.0%) |
| Some College | 54 (17.7%) | 16 (21.3%) |
| High School or less | 38 (12.5%) | 9 (12.0%) |

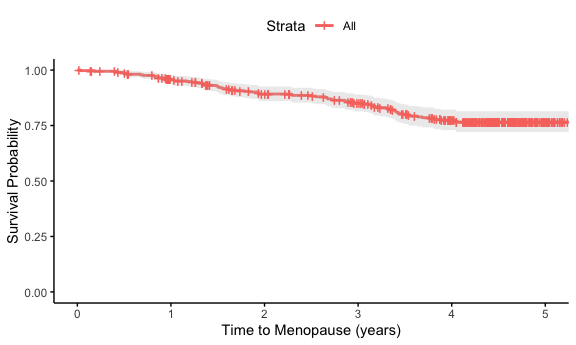
### (Ia) Estimate the median menopause time assuming an exponential distribution  
  
# Number of observations  
n <- nrow(data)  
# Estimate lambda (rate parameter)  
lambda\_hat <- 1 / mean(data$menopause\_time)  
# Estimate the median menopause time  
median\_estimate <- log(2) / lambda\_hat  
# Compute standard error using the Delta Method  
se\_median <- log(2) / (lambda\_hat \* sqrt(n))  
# Compute the 95% confidence interval  
ci\_lower <- median\_estimate - 1.96 \* se\_median  
ci\_upper <- median\_estimate + 1.96 \* se\_median  
cat("Estimated median menopause time (Exponential):", median\_estimate, "\n")

## Estimated median menopause time (Exponential): 2.408339

cat("95% CI:", c(ci\_lower, ci\_upper), "\n")

## 95% CI: 2.16619 2.650487

### (Ib) Compute Kaplan-Meier survival estimate  
df <- data %>%   
 mutate(menopause = as.numeric(menopause))  
  
km\_fit <- survfit(Surv(menopause\_time, menopause) ~ 1, data = df)  
  
# Display survival estimates in a table  
km\_table <- data.frame(time = km\_fit$time,  
 n.risk = km\_fit$n.risk,  
 n.event = km\_fit$n.event,  
 n.censor = km\_fit$n.censor,  
 surv = km\_fit$surv,  
 upper = km\_fit$upper,  
 lower = km\_fit$lower)  
  
#median\_time <- summary(km\_fit)$table["median"]  
  
# Plot Kaplan-Meier Curve  
ggsurvplot(km\_fit, data = df, conf.int = TRUE,   
 ggtheme = theme\_classic(), xlab="Time to Menopause (years)",   
 ylab="Survival Probability")



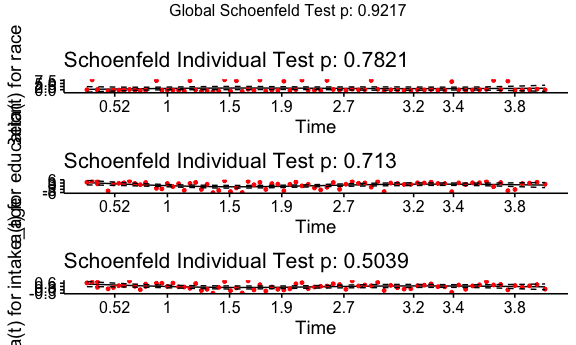
cat("Estimated median time to menopause (Kaplan-Meier):", summary(km\_fit)$table["median"], "\n")

## Estimated median time to menopause (Kaplan-Meier): NA

# Fit the Cox proportional hazards model  
cox\_model <- coxph(Surv(menopause\_time, menopause) ~ race + education + intake\_age, data = df)  
  
# Display model summary  
summary(cox\_model)

## Call:  
## coxph(formula = Surv(menopause\_time, menopause) ~ race + education +   
## intake\_age, data = df)  
##   
## n= 380, number of events= 75   
##   
## coef exp(coef) se(coef) z Pr(>|z|)   
## raceBlack non-Hispanic 0.90155 2.46343 0.33129 2.721 0.0065 \*\*   
## raceOther Ethnicity 0.01639 1.01653 0.43332 0.038 0.9698   
## educationCollege Graduate -0.88633 0.41217 0.33881 -2.616 0.0089 \*\*   
## educationSome College 0.05357 1.05503 0.30769 0.174 0.8618   
## educationHigh School or less -0.58824 0.55530 0.40087 -1.467 0.1423   
## intake\_age 0.31054 1.36417 0.03647 8.514 <2e-16 \*\*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## exp(coef) exp(-coef) lower .95 upper .95  
## raceBlack non-Hispanic 2.4634 0.4059 1.2869 4.7155  
## raceOther Ethnicity 1.0165 0.9837 0.4348 2.3766  
## educationCollege Graduate 0.4122 2.4262 0.2122 0.8007  
## educationSome College 1.0550 0.9478 0.5772 1.9283  
## educationHigh School or less 0.5553 1.8008 0.2531 1.2183  
## intake\_age 1.3642 0.7330 1.2700 1.4653  
##   
## Concordance= 0.772 (se = 0.026 )  
## Likelihood ratio test= 68.84 on 6 df, p=7e-13  
## Wald test = 77.65 on 6 df, p=1e-14  
## Score (logrank) test = 86.61 on 6 df, p=<2e-16

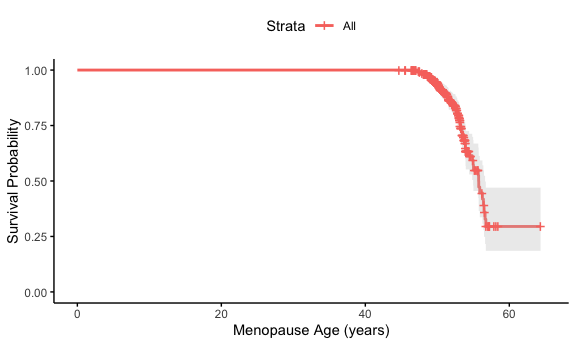
# Check the proportional hazards assumption  
cox.zph\_test = cox.zph(cox\_model)  
  
# Plot Schoenfeld residuals to visualize proportional hazards assumption  
ggcoxzph(cox.zph\_test)



### (III) Compute Kaplan-Meier for menopause\_age  
km\_fit\_age <- survfit(Surv(menopause\_age, menopause) ~ 1, data = df)  
  
# Kaplan-Meier Table  
summary(km\_fit\_age)

## Call: survfit(formula = Surv(menopause\_age, menopause) ~ 1, data = df)  
##   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 47.3 367 1 0.997 0.00272 0.992 1.000  
## 47.3 366 1 0.995 0.00384 0.987 1.000  
## 47.5 363 1 0.992 0.00471 0.983 1.000  
## 47.8 361 1 0.989 0.00544 0.978 1.000  
## 47.9 360 1 0.986 0.00608 0.974 0.998  
## 48.0 358 1 0.984 0.00666 0.971 0.997  
## 48.2 354 1 0.981 0.00719 0.967 0.995  
## 48.6 346 1 0.978 0.00771 0.963 0.993  
## 48.7 345 1 0.975 0.00819 0.959 0.991  
## 48.8 340 1 0.972 0.00866 0.955 0.989  
## 49.0 329 1 0.969 0.00912 0.952 0.987  
## 49.0 326 1 0.966 0.00957 0.948 0.985  
## 49.2 315 1 0.963 0.01001 0.944 0.983  
## 49.2 313 1 0.960 0.01044 0.940 0.981  
## 49.3 311 1 0.957 0.01086 0.936 0.979  
## 49.3 310 1 0.954 0.01125 0.932 0.976  
## 49.4 298 1 0.951 0.01166 0.928 0.974  
## 49.5 292 1 0.948 0.01207 0.924 0.971  
## 49.7 278 1 0.944 0.01250 0.920 0.969  
## 50.0 257 1 0.940 0.01298 0.915 0.966  
## 50.1 253 1 0.937 0.01345 0.911 0.963  
## 50.1 252 1 0.933 0.01390 0.906 0.961  
## 50.2 245 1 0.929 0.01435 0.902 0.958  
## 50.3 244 1 0.925 0.01479 0.897 0.955  
## 50.3 234 1 0.921 0.01525 0.892 0.952  
## 50.4 229 1 0.917 0.01570 0.887 0.949  
## 50.4 227 1 0.913 0.01615 0.882 0.946  
## 50.5 218 1 0.909 0.01661 0.877 0.942  
## 50.7 208 1 0.905 0.01709 0.872 0.939  
## 50.9 196 1 0.900 0.01762 0.866 0.935  
## 51.1 183 1 0.895 0.01820 0.860 0.932  
## 51.3 172 1 0.890 0.01882 0.854 0.928  
## 51.4 170 1 0.885 0.01942 0.848 0.924  
## 51.5 163 1 0.879 0.02005 0.841 0.920  
## 51.5 160 1 0.874 0.02066 0.834 0.915  
## 51.6 159 1 0.868 0.02125 0.828 0.911  
## 51.6 156 1 0.863 0.02183 0.821 0.907  
## 51.9 142 1 0.857 0.02251 0.814 0.902  
## 52.1 134 1 0.850 0.02323 0.806 0.897  
## 52.2 128 1 0.844 0.02398 0.798 0.892  
## 52.5 117 1 0.837 0.02484 0.789 0.887  
## 52.5 114 1 0.829 0.02568 0.780 0.881  
## 52.7 110 1 0.822 0.02653 0.771 0.875  
## 52.8 107 1 0.814 0.02737 0.762 0.869  
## 52.8 106 1 0.806 0.02817 0.753 0.863  
## 52.9 101 1 0.798 0.02900 0.743 0.857  
## 53.0 96 1 0.790 0.02987 0.734 0.851  
## 53.0 94 1 0.782 0.03071 0.724 0.844  
## 53.1 90 1 0.773 0.03157 0.713 0.837  
## 53.2 86 1 0.764 0.03246 0.703 0.830  
## 53.2 84 1 0.755 0.03332 0.692 0.823  
## 53.2 83 1 0.746 0.03414 0.682 0.816  
## 53.3 80 1 0.736 0.03496 0.671 0.808  
## 53.4 75 1 0.727 0.03585 0.660 0.800  
## 53.4 74 1 0.717 0.03668 0.648 0.792  
## 53.5 73 1 0.707 0.03747 0.637 0.784  
## 53.6 67 1 0.696 0.03837 0.625 0.776  
## 53.7 64 1 0.686 0.03928 0.613 0.767  
## 53.9 54 1 0.673 0.04055 0.598 0.757  
## 53.9 51 1 0.660 0.04185 0.583 0.747  
## 53.9 50 1 0.646 0.04304 0.567 0.737  
## 54.0 48 1 0.633 0.04420 0.552 0.726  
## 54.4 34 1 0.614 0.04666 0.529 0.713  
## 54.9 30 1 0.594 0.04939 0.505 0.699  
## 55.0 28 1 0.573 0.05198 0.479 0.684  
## 55.0 27 1 0.551 0.05421 0.455 0.669  
## 55.7 21 1 0.525 0.05764 0.424 0.651  
## 55.7 20 1 0.499 0.06045 0.393 0.633  
## 55.8 19 1 0.473 0.06271 0.364 0.613  
## 55.9 18 1 0.446 0.06449 0.336 0.593  
## 56.3 16 1 0.419 0.06622 0.307 0.571  
## 56.4 15 1 0.391 0.06743 0.278 0.548  
## 56.5 13 1 0.361 0.06861 0.248 0.524  
## 56.6 11 1 0.328 0.06977 0.216 0.497  
## 56.7 10 1 0.295 0.07007 0.185 0.470

km\_age\_table <- data.frame(time = km\_fit\_age$time,   
 survival = km\_fit\_age$surv,   
 lower = km\_fit\_age$lower,   
 upper = km\_fit\_age$upper)  
  
# Kaplan-Meier Plot for menopause\_age  
ggsurvplot(km\_fit\_age, data = df, conf.int = TRUE,   
 ggtheme = theme\_classic(), xlab="Menopause Age (years)",   
 ylab="Survival Probability")



cat("Estimated median menopause age (Kaplan-Meier):", summary(km\_fit\_age)$table["median"], "years.\n")

## Estimated median menopause age (Kaplan-Meier): 55.7399 years.

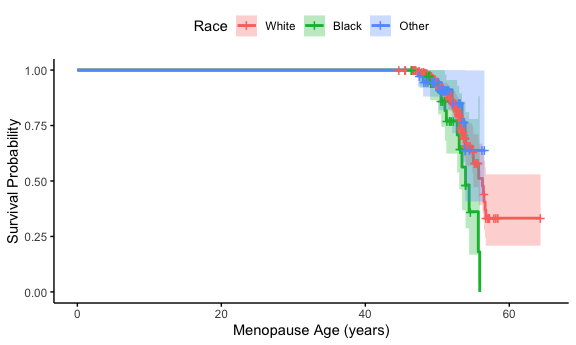
# Assume exponential distribution  
lambda\_hat <- 1 / mean(data$menopause\_age) # MLE for exponential rate  
median\_exp <- log(2) / lambda\_hat  
cat("Estimated Median (Exponential Model):", median\_exp, "years.\n")

## Estimated Median (Exponential Model): 35.57552 years.

### (IV) Test for Survival Differences by Race  
km\_fit\_race <- survfit(Surv(menopause\_age, menopause) ~ race, data = df)  
summary(km\_fit\_race)

## Call: survfit(formula = Surv(menopause\_age, menopause) ~ race, data = df)  
##   
## race=White non-Hispanic   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 47.3 292 1 0.997 0.00342 0.990 1.000  
## 47.8 289 1 0.993 0.00484 0.984 1.000  
## 47.9 288 1 0.990 0.00593 0.978 1.000  
## 48.2 284 1 0.986 0.00686 0.973 1.000  
## 48.6 278 1 0.983 0.00769 0.968 0.998  
## 48.7 277 1 0.979 0.00844 0.963 0.996  
## 48.8 273 1 0.976 0.00914 0.958 0.994  
## 49.0 263 1 0.972 0.00983 0.953 0.991  
## 49.2 255 1 0.968 0.01051 0.948 0.989  
## 49.2 253 1 0.964 0.01114 0.943 0.986  
## 49.3 251 1 0.960 0.01174 0.938 0.984  
## 49.3 250 1 0.956 0.01230 0.933 0.981  
## 49.4 241 1 0.953 0.01288 0.928 0.978  
## 49.5 237 1 0.948 0.01344 0.923 0.975  
## 49.7 226 1 0.944 0.01402 0.917 0.972  
## 50.0 205 1 0.940 0.01469 0.911 0.969  
## 50.1 201 1 0.935 0.01534 0.905 0.966  
## 50.2 196 1 0.930 0.01599 0.899 0.962  
## 50.3 187 1 0.925 0.01666 0.893 0.958  
## 50.4 182 1 0.920 0.01732 0.887 0.955  
## 50.4 180 1 0.915 0.01796 0.881 0.951  
## 50.7 165 1 0.910 0.01869 0.874 0.947  
## 50.9 155 1 0.904 0.01947 0.866 0.943  
## 51.4 135 1 0.897 0.02045 0.858 0.938  
## 51.5 130 1 0.890 0.02142 0.849 0.933  
## 51.5 128 1 0.883 0.02235 0.840 0.928  
## 51.6 127 1 0.876 0.02323 0.832 0.923  
## 51.6 124 1 0.869 0.02410 0.823 0.918  
## 51.9 113 1 0.861 0.02508 0.814 0.912  
## 52.2 101 1 0.853 0.02624 0.803 0.906  
## 52.5 92 1 0.844 0.02755 0.791 0.899  
## 52.5 89 1 0.834 0.02882 0.780 0.893  
## 52.7 85 1 0.824 0.03011 0.767 0.885  
## 52.8 82 1 0.814 0.03137 0.755 0.878  
## 52.9 77 1 0.804 0.03270 0.742 0.870  
## 53.0 73 1 0.793 0.03405 0.729 0.862  
## 53.1 68 1 0.781 0.03549 0.714 0.854  
## 53.2 66 1 0.769 0.03687 0.700 0.845  
## 53.2 64 1 0.757 0.03821 0.686 0.836  
## 53.2 63 1 0.745 0.03945 0.672 0.827  
## 53.3 61 1 0.733 0.04065 0.657 0.817  
## 53.5 57 1 0.720 0.04192 0.642 0.807  
## 53.6 52 1 0.706 0.04334 0.626 0.796  
## 53.7 50 1 0.692 0.04471 0.610 0.786  
## 53.9 40 1 0.675 0.04682 0.589 0.773  
## 54.0 39 1 0.657 0.04872 0.569 0.760  
## 54.9 26 1 0.632 0.05300 0.536 0.745  
## 55.0 24 1 0.606 0.05696 0.504 0.728  
## 55.0 23 1 0.580 0.06027 0.473 0.711  
## 55.7 17 1 0.545 0.06566 0.431 0.691  
## 55.8 16 1 0.511 0.06985 0.391 0.668  
## 56.3 15 1 0.477 0.07304 0.354 0.644  
## 56.4 14 1 0.443 0.07536 0.318 0.618  
## 56.5 12 1 0.406 0.07760 0.279 0.591  
## 56.6 11 1 0.369 0.07885 0.243 0.561  
## 56.7 10 1 0.332 0.07914 0.208 0.530  
##   
## race=Black non-Hispanic   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 47.5 36 1 0.972 0.0274 0.9200 1.000  
## 49.0 31 1 0.941 0.0407 0.8644 1.000  
## 50.1 24 1 0.902 0.0547 0.8006 1.000  
## 50.5 23 1 0.862 0.0649 0.7442 0.999  
## 51.1 19 1 0.817 0.0757 0.6814 0.980  
## 51.3 18 1 0.772 0.0840 0.6234 0.955  
## 52.8 12 1 0.707 0.0986 0.5383 0.930  
## 53.0 11 1 0.643 0.1086 0.4619 0.895  
## 53.4 8 1 0.563 0.1212 0.3689 0.858  
## 53.9 7 1 0.482 0.1278 0.2870 0.811  
## 54.4 4 1 0.362 0.1417 0.1678 0.780  
## 55.7 2 1 0.181 0.1462 0.0371 0.882  
## 55.9 1 1 0.000 NaN NA NA  
##   
## race=Other Ethnicity   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 47.3 39 1 0.974 0.0253 0.926 1.000  
## 48.0 37 1 0.948 0.0358 0.880 1.000  
## 50.3 26 1 0.912 0.0496 0.819 1.000  
## 52.1 15 1 0.851 0.0748 0.716 1.000  
## 53.4 10 1 0.766 0.1051 0.585 1.000  
## 53.9 6 1 0.638 0.1457 0.408 0.998

ggsurvplot(km\_fit\_race, data = df, conf.int = TRUE,   
 ggtheme = theme\_classic(), xlab="Menopause Age (years)",   
 ylab="Survival Probability", legend.title="Race", legend.labs=c("White", "Black", "Other"))



# Generalized Log-rank test  
## Gehan-Breslow-Wilcoxon Test (Early Failures Weighted Higher)  
survdiff\_test <- survdiff(Surv(menopause\_age, menopause) ~ race, data = df, rho = 1)  
print(survdiff\_test)

## Call:  
## survdiff(formula = Surv(menopause\_age, menopause) ~ race, data = df,   
## rho = 1)  
##   
## N Observed Expected (O-E)^2/E (O-E)^2/V  
## race=White non-Hispanic 304 45.01 48.33 0.229 1.345  
## race=Black non-Hispanic 37 10.28 5.78 3.500 4.527  
## race=Other Ethnicity 39 5.18 6.35 0.215 0.283  
##   
## Chisq= 4.6 on 2 degrees of freedom, p= 0.1

# Cox score test  
cox\_model\_age <- coxph(Surv(menopause\_age, menopause) ~ race + intake\_age, data = df)  
summary(cox\_model\_age)

## Call:  
## coxph(formula = Surv(menopause\_age, menopause) ~ race + intake\_age,   
## data = df)  
##   
## n= 380, number of events= 75   
##   
## coef exp(coef) se(coef) z Pr(>|z|)   
## raceBlack non-Hispanic 0.7631 2.1450 0.3123 2.443 0.01455 \*   
## raceOther Ethnicity -0.1091 0.8966 0.4323 -0.252 0.80068   
## intake\_age -0.1797 0.8355 0.0600 -2.995 0.00275 \*\*  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## exp(coef) exp(-coef) lower .95 upper .95  
## raceBlack non-Hispanic 2.1450 0.4662 1.1630 3.9561  
## raceOther Ethnicity 0.8966 1.1153 0.3842 2.0922  
## intake\_age 0.8355 1.1968 0.7428 0.9398  
##   
## Concordance= 0.611 (se = 0.029 )  
## Likelihood ratio test= 15.3 on 3 df, p=0.002  
## Wald test = 14.73 on 3 df, p=0.002  
## Score (logrank) test = 15.1 on 3 df, p=0.002

### (V) Cox Regression for Menopause Age with Race, adjusting for education and intake\_age  
cox\_model\_age <- coxph(Surv(menopause\_age, menopause) ~ race + education, data = df)  
summary(cox\_model\_age)

## Call:  
## coxph(formula = Surv(menopause\_age, menopause) ~ race + education,   
## data = df)  
##   
## n= 380, number of events= 75   
##   
## coef exp(coef) se(coef) z Pr(>|z|)   
## raceBlack non-Hispanic 0.9130 2.4917 0.3348 2.727 0.0064 \*\*  
## raceOther Ethnicity -0.1182 0.8885 0.4337 -0.273 0.7852   
## educationCollege Graduate -0.8236 0.4389 0.3223 -2.556 0.0106 \*   
## educationSome College -0.0591 0.9426 0.3105 -0.190 0.8490   
## educationHigh School or less -0.7686 0.4637 0.4058 -1.894 0.0582 .   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## exp(coef) exp(-coef) lower .95 upper .95  
## raceBlack non-Hispanic 2.4917 0.4013 1.2927 4.8027  
## raceOther Ethnicity 0.8885 1.1255 0.3797 2.0790  
## educationCollege Graduate 0.4389 2.2786 0.2334 0.8253  
## educationSome College 0.9426 1.0609 0.5129 1.7323  
## educationHigh School or less 0.4637 2.1568 0.2093 1.0271  
##   
## Concordance= 0.595 (se = 0.038 )  
## Likelihood ratio test= 14.52 on 5 df, p=0.01  
## Wald test = 14.67 on 5 df, p=0.01  
## Score (logrank) test = 15.17 on 5 df, p=0.01

# (Vb) Relative risk estimates for Black vs Other controlling for education  
  
dd <- df  
dd$race <- relevel(factor(df$race), ref = "Other Ethnicity")  
  
# Fit the Cox model with new reference group  
cox\_model\_ref\_other <- coxph(Surv(menopause\_age, menopause) ~ race + education, data = dd)  
  
# Display model summary  
summary(cox\_model\_ref\_other)

## Call:  
## coxph(formula = Surv(menopause\_age, menopause) ~ race + education,   
## data = dd)  
##   
## n= 380, number of events= 75   
##   
## coef exp(coef) se(coef) z Pr(>|z|)   
## raceWhite non-Hispanic 0.1182 1.1255 0.4337 0.273 0.7852   
## raceBlack non-Hispanic 1.0312 2.8044 0.5045 2.044 0.0410 \*  
## educationCollege Graduate -0.8236 0.4389 0.3223 -2.556 0.0106 \*  
## educationSome College -0.0591 0.9426 0.3105 -0.190 0.8490   
## educationHigh School or less -0.7686 0.4637 0.4058 -1.894 0.0582 .  
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## exp(coef) exp(-coef) lower .95 upper .95  
## raceWhite non-Hispanic 1.1255 0.8885 0.4810 2.6336  
## raceBlack non-Hispanic 2.8044 0.3566 1.0433 7.5383  
## educationCollege Graduate 0.4389 2.2786 0.2334 0.8253  
## educationSome College 0.9426 1.0609 0.5129 1.7323  
## educationHigh School or less 0.4637 2.1568 0.2093 1.0271  
##   
## Concordance= 0.595 (se = 0.038 )  
## Likelihood ratio test= 14.52 on 5 df, p=0.01  
## Wald test = 14.67 on 5 df, p=0.01  
## Score (logrank) test = 15.17 on 5 df, p=0.01

# Extract coefficients and standard errors  
coef\_black <- coef(cox\_model\_age)["raceBlack non-Hispanic"]  
coef\_other <- coef(cox\_model\_age)["raceOther Ethnicity"]  
SE\_black <- summary(cox\_model\_age)$coef["raceBlack non-Hispanic", "se(coef)"]  
SE\_other <- summary(cox\_model\_age)$coef["raceOther Ethnicity", "se(coef)"]  
  
# Compute log HR ratio and standard error  
log\_HR\_ratio <- coef\_black - coef\_other  
SE\_log\_HR\_ratio <- sqrt(SE\_black^2 + SE\_other^2)  
  
# Compute 95% CI  
lower\_CI <- exp(log\_HR\_ratio - 1.96 \* SE\_log\_HR\_ratio)  
upper\_CI <- exp(log\_HR\_ratio + 1.96 \* SE\_log\_HR\_ratio)  
  
# Print results  
cat("HR Ratio (Black vs. Other Ethnicity):", exp(log\_HR\_ratio), "\n")

## HR Ratio (Black vs. Other Ethnicity): 2.80437

cat("95% Confidence Interval: (", lower\_CI, ",", upper\_CI, ")\n")

## 95% Confidence Interval: ( 0.9581297 , 8.20817 )

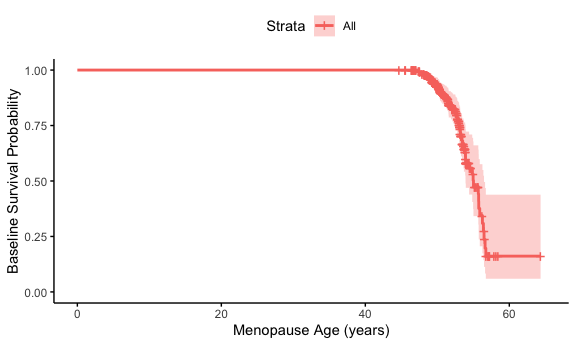
vcov\_matrix <- vcov(cox\_model\_age)  
vcov\_matrix["raceBlack non-Hispanic", "raceOther Ethnicity"]

## [1] 0.02285186

# (Vc) Baseline survival function for White non-Hispanic with Post-graduate education  
base\_surv <- survfit(cox\_model\_age, newdata = data.frame(race = "White non-Hispanic", education = "Post-graduate"))  
summary(base\_surv)

## Call: survfit(formula = cox\_model\_age, newdata = data.frame(race = "White non-Hispanic",   
## education = "Post-graduate"))  
##   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 47.3 367 1 0.997 0.00314 0.9907 1.000  
## 47.3 366 1 0.994 0.00447 0.9850 1.000  
## 47.5 363 1 0.991 0.00552 0.9798 1.000  
## 47.8 361 1 0.987 0.00642 0.9749 1.000  
## 47.9 360 1 0.984 0.00722 0.9702 0.999  
## 48.0 358 1 0.981 0.00796 0.9656 0.997  
## 48.2 354 1 0.978 0.00866 0.9611 0.995  
## 48.6 346 1 0.975 0.00933 0.9565 0.993  
## 48.7 345 1 0.971 0.00997 0.9521 0.991  
## 48.8 340 1 0.968 0.01059 0.9476 0.989  
## 49.0 329 1 0.965 0.01123 0.9430 0.987  
## 49.0 326 1 0.961 0.01184 0.9384 0.985  
## 49.2 315 1 0.958 0.01247 0.9336 0.983  
## 49.2 313 1 0.954 0.01308 0.9289 0.980  
## 49.3 311 1 0.951 0.01368 0.9242 0.978  
## 49.3 310 1 0.947 0.01425 0.9196 0.975  
## 49.4 298 1 0.943 0.01484 0.9148 0.973  
## 49.5 292 1 0.940 0.01544 0.9099 0.970  
## 49.7 278 1 0.936 0.01606 0.9048 0.968  
## 50.0 257 1 0.932 0.01675 0.8993 0.965  
## 50.1 253 1 0.927 0.01742 0.8938 0.962  
## 50.1 252 1 0.923 0.01808 0.8883 0.959  
## 50.2 245 1 0.919 0.01875 0.8828 0.956  
## 50.3 244 1 0.914 0.01939 0.8772 0.953  
## 50.3 234 1 0.910 0.02006 0.8715 0.950  
## 50.4 229 1 0.905 0.02073 0.8657 0.947  
## 50.4 227 1 0.901 0.02140 0.8599 0.944  
## 50.5 218 1 0.896 0.02210 0.8539 0.941  
## 50.7 208 1 0.891 0.02284 0.8476 0.937  
## 50.9 196 1 0.886 0.02364 0.8408 0.934  
## 51.1 183 1 0.880 0.02452 0.8336 0.930  
## 51.3 172 1 0.874 0.02547 0.8259 0.926  
## 51.4 170 1 0.868 0.02640 0.8182 0.922  
## 51.5 163 1 0.862 0.02736 0.8102 0.918  
## 51.5 160 1 0.856 0.02831 0.8023 0.913  
## 51.6 159 1 0.850 0.02922 0.7944 0.909  
## 51.6 156 1 0.843 0.03015 0.7864 0.905  
## 51.9 142 1 0.837 0.03116 0.7777 0.900  
## 52.1 134 1 0.829 0.03222 0.7687 0.895  
## 52.2 128 1 0.822 0.03330 0.7594 0.890  
## 52.5 117 1 0.814 0.03450 0.7493 0.885  
## 52.5 114 1 0.806 0.03570 0.7390 0.879  
## 52.7 110 1 0.798 0.03692 0.7284 0.873  
## 52.8 107 1 0.789 0.03813 0.7178 0.868  
## 52.8 106 1 0.781 0.03931 0.7072 0.861  
## 52.9 101 1 0.772 0.04052 0.6961 0.855  
## 53.0 96 1 0.762 0.04181 0.6845 0.849  
## 53.0 94 1 0.753 0.04309 0.6729 0.842  
## 53.1 90 1 0.743 0.04437 0.6609 0.835  
## 53.2 86 1 0.733 0.04571 0.6484 0.828  
## 53.2 84 1 0.722 0.04702 0.6359 0.821  
## 53.2 83 1 0.712 0.04829 0.6236 0.813  
## 53.3 80 1 0.702 0.04959 0.6108 0.806  
## 53.4 75 1 0.690 0.05102 0.5972 0.798  
## 53.4 74 1 0.679 0.05237 0.5837 0.790  
## 53.5 73 1 0.668 0.05365 0.5703 0.782  
## 53.6 67 1 0.656 0.05506 0.5561 0.773  
## 53.7 64 1 0.643 0.05643 0.5418 0.764  
## 53.9 54 1 0.629 0.05843 0.5241 0.754  
## 53.9 51 1 0.614 0.06039 0.5061 0.744  
## 53.9 50 1 0.599 0.06226 0.4884 0.734  
## 54.0 48 1 0.582 0.06418 0.4691 0.723  
## 54.4 34 1 0.558 0.06839 0.4389 0.710  
## 54.9 30 1 0.531 0.07289 0.4054 0.695  
## 55.0 28 1 0.503 0.07679 0.3727 0.678  
## 55.0 27 1 0.475 0.07997 0.3414 0.661  
## 55.7 21 1 0.442 0.08427 0.3039 0.642  
## 55.7 20 1 0.408 0.08738 0.2683 0.621  
## 55.8 19 1 0.376 0.08916 0.2364 0.599  
## 55.9 18 1 0.344 0.09037 0.2057 0.576  
## 56.3 16 1 0.309 0.09088 0.1734 0.550  
## 56.4 15 1 0.273 0.09065 0.1428 0.524  
## 56.5 13 1 0.236 0.08930 0.1127 0.496  
## 56.6 11 1 0.197 0.08708 0.0827 0.468  
## 56.7 10 1 0.161 0.08226 0.0594 0.438

ggsurvplot(base\_surv, conf.inf = T, data=df,ggtheme = theme\_classic(), xlab="Menopause Age (years)", ylab="Baseline Survival Probability")



# (Vd) Check proportional hazards assumption  
cox.zph(cox\_model\_age)

## chisq df p  
## race 3.15 2 0.21  
## education 0.85 3 0.84  
## GLOBAL 4.49 5 0.48