proj2

Jiayi Shi

2025-03-08

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%) |

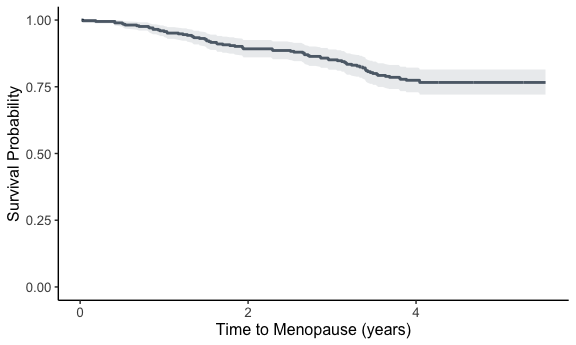
### (Ia) Estimate the median menopause time assuming an exponential distribution  
  
# Number of observations  
n <- nrow(data)  
# Estimate lambda (rate parameter)  
lambda\_hat <- 1 / mean(data$menopause\_time)  
# Estimate the median menopause time  
median\_estimate <- log(2) / lambda\_hat  
# Compute 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)  
  
# Extract survival data from the Kaplan-Meier fit  
km\_data <- data.frame(time = km\_fit$time,  
 survival = km\_fit$surv,  
 lower = km\_fit$lower,  
 upper = km\_fit$upper)  
  
# KM survival curve  
ggplot(km\_data, aes(x = time, y = survival)) +  
 geom\_step(size = 1, color = "#2C3E50") +   
 geom\_ribbon(aes(ymin = lower, ymax = upper), fill = "#BDC3C7", alpha = 0.3) +  
 scale\_y\_continuous(limits = c(0:1))+  
 labs(x = "Time to Menopause (years)",  
 y = "Survival Probability") +  
 theme\_classic() +   
 theme(axis.text = element\_text(size = 10),  
 axis.title = element\_text(size = 12))



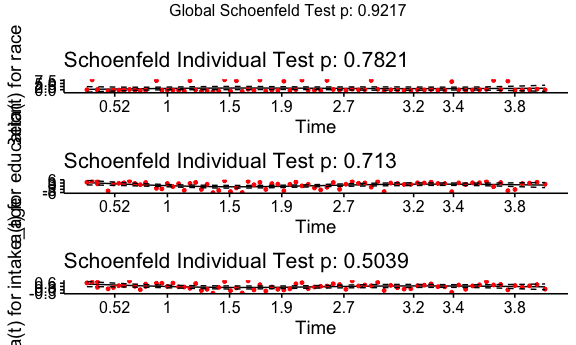
cat("Estimated median time to menopause (Kaplan-Meier):", summary(km\_fit)$table["median"], "\n")

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

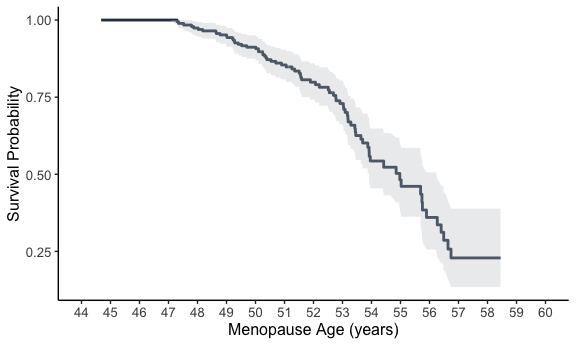
# Fit the Cox proportional hazards model  
cox\_model <- coxph(Surv(menopause\_time, menopause) ~ race + education + intake\_age, data = df)  
  
# Display model summary  
summary(cox\_model)

## 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

# Plot Schoenfeld residuals to visualize proportional hazards assumption  
cox.zph\_test <- cox.zph(cox\_model)  
ggcoxzph(cox.zph\_test)



### (III) Compute Kaplan-Meier for menopause\_age  
km\_fit\_age <- survfit(Surv(intake\_age, menopause\_age, menopause) ~ 1, data = df)  
  
# Extract survival data from the Kaplan-Meier fit  
km\_data <- data.frame(time = km\_fit\_age$time,  
 survival = km\_fit\_age$surv,  
 lower = km\_fit\_age$lower,  
 upper = km\_fit\_age$upper)  
  
# Plot using ggplot  
ggplot(km\_data, aes(x = time, y = survival)) +  
 geom\_step(size = 1, color = "#2C3E50") + # Survival curve  
 geom\_ribbon(aes(ymin = lower, ymax = upper), fill = "#BDC3C7", alpha = 0.3) + # Confidence interval shading  
 scale\_x\_continuous(breaks = seq(44, 60, by = 1), limits = c(44, 60)) +   
 labs(x = "Menopause Age (years)",  
 y = "Survival Probability") +  
 theme\_classic() +  
 theme(axis.text = element\_text(size = 10),  
 axis.title = element\_text(size = 12))



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

## Estimated median menopause age (Kaplan-Meier): 54.97604 years.

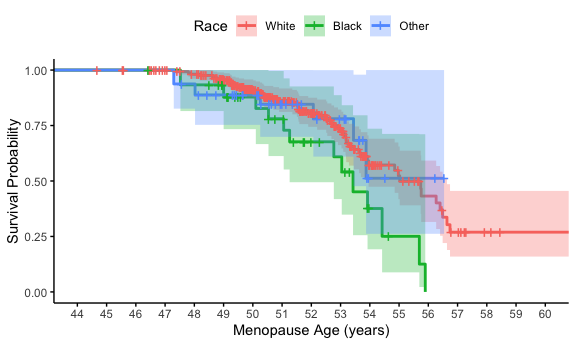
# Assume exponential distribution  
lambda\_hat <- 1 / mean(data$menopause\_age) # MLE for exponential rate  
median\_exp <- log(2) / lambda\_hat  
cat("Estimated Median (Exponential Model):", median\_exp, "years.\n")

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

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

## Call: survfit(formula = Surv(intake\_age, menopause\_age, menopause) ~   
## race, data = df)  
##   
## race=White non-Hispanic   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 47.3 152 1 0.993 0.00656 0.981 1.000  
## 47.8 167 1 0.987 0.00881 0.970 1.000  
## 47.9 166 1 0.982 0.01058 0.961 1.000  
## 48.2 176 1 0.976 0.01190 0.953 1.000  
## 48.6 182 1 0.971 0.01298 0.945 0.996  
## 48.7 181 1 0.965 0.01398 0.938 0.993  
## 48.8 183 1 0.960 0.01486 0.931 0.990  
## 49.0 183 1 0.955 0.01568 0.924 0.986  
## 49.2 179 1 0.949 0.01647 0.918 0.982  
## 49.2 177 1 0.944 0.01723 0.911 0.978  
## 49.3 176 1 0.939 0.01795 0.904 0.974  
## 49.3 175 1 0.933 0.01863 0.897 0.971  
## 49.4 170 1 0.928 0.01931 0.891 0.966  
## 49.5 168 1 0.922 0.01997 0.884 0.962  
## 49.7 162 1 0.917 0.02064 0.877 0.958  
## 50.0 150 1 0.910 0.02139 0.869 0.953  
## 50.1 148 1 0.904 0.02211 0.862 0.949  
## 50.2 143 1 0.898 0.02285 0.854 0.944  
## 50.3 136 1 0.891 0.02361 0.846 0.939  
## 50.4 132 1 0.885 0.02438 0.838 0.934  
## 50.4 130 1 0.878 0.02512 0.830 0.928  
## 50.7 118 1 0.870 0.02599 0.821 0.923  
## 50.9 111 1 0.863 0.02691 0.811 0.917  
## 51.4 98 1 0.854 0.02804 0.801 0.911  
## 51.5 94 1 0.845 0.02918 0.789 0.904  
## 51.5 92 1 0.835 0.03027 0.778 0.897  
## 51.6 92 1 0.826 0.03127 0.767 0.890  
## 51.6 89 1 0.817 0.03227 0.756 0.883  
## 51.9 81 1 0.807 0.03341 0.744 0.875  
## 52.2 73 1 0.796 0.03473 0.731 0.867  
## 52.5 70 1 0.785 0.03605 0.717 0.859  
## 52.5 67 1 0.773 0.03737 0.703 0.850  
## 52.7 66 1 0.761 0.03859 0.689 0.841  
## 52.8 66 1 0.750 0.03969 0.676 0.832  
## 52.9 62 1 0.738 0.04085 0.662 0.822  
## 53.0 58 1 0.725 0.04208 0.647 0.812  
## 53.1 54 1 0.711 0.04339 0.631 0.802  
## 53.2 53 1 0.698 0.04460 0.616 0.791  
## 53.2 51 1 0.684 0.04578 0.600 0.780  
## 53.2 50 1 0.671 0.04686 0.585 0.769  
## 53.3 48 1 0.657 0.04792 0.569 0.758  
## 53.5 45 1 0.642 0.04903 0.553 0.746  
## 53.6 41 1 0.626 0.05027 0.535 0.733  
## 53.7 39 1 0.610 0.05149 0.517 0.720  
## 53.9 32 1 0.591 0.05329 0.496 0.706  
## 54.0 31 1 0.572 0.05488 0.474 0.691  
## 54.9 24 1 0.548 0.05754 0.446 0.674  
## 55.0 22 1 0.523 0.06008 0.418 0.655  
## 55.0 21 1 0.498 0.06218 0.390 0.637  
## 55.7 15 1 0.465 0.06632 0.352 0.615  
## 55.8 14 1 0.432 0.06941 0.315 0.592  
## 56.3 14 1 0.401 0.07098 0.284 0.567  
## 56.4 13 1 0.370 0.07192 0.253 0.542  
## 56.5 11 1 0.337 0.07283 0.220 0.514  
## 56.6 10 1 0.303 0.07292 0.189 0.486  
## 56.7 9 1 0.269 0.07217 0.159 0.455  
##   
## race=Black non-Hispanic   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 47.5 15 1 0.933 0.0644 0.8153 1.000  
## 49.0 17 1 0.878 0.0807 0.7337 1.000  
## 50.1 17 1 0.827 0.0910 0.6663 1.000  
## 50.5 18 1 0.781 0.0968 0.6123 0.996  
## 51.1 15 1 0.729 0.1034 0.5518 0.963  
## 51.3 14 1 0.677 0.1084 0.4944 0.926  
## 52.8 10 1 0.609 0.1168 0.4183 0.887  
## 53.0 9 1 0.541 0.1218 0.3483 0.841  
## 53.4 6 1 0.451 0.1307 0.2557 0.796  
## 53.9 6 1 0.376 0.1288 0.1921 0.736  
## 54.4 3 1 0.251 0.1336 0.0882 0.712  
## 55.7 2 1 0.125 0.1110 0.0221 0.711  
## 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 16 1 0.938 0.0605 0.826 1.000  
## 48.0 19 1 0.888 0.0748 0.753 1.000  
## 50.3 21 1 0.846 0.0823 0.699 1.000  
## 52.1 13 1 0.781 0.0984 0.610 1.000  
## 53.4 8 1 0.683 0.1255 0.477 0.979  
## 53.9 4 1 0.512 0.1753 0.262 1.000

km\_plot <- ggsurvplot(km\_fit\_race, data = df, conf.int = TRUE,   
 ggtheme = theme\_classic(), xlab="Menopause Age (years)",   
 ylab="Survival Probability", xlim=c(44,60),  
 legend.title="Race", legend.labs=c("White", "Black", "Other"))  
km\_plot$plot +  
 scale\_x\_continuous(breaks = seq(44, 60, by = 1))



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

## 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

#race not significant

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

## Call:  
## coxph(formula = Surv(intake\_age, menopause\_age, menopause) ~   
## race + education, data = df)  
##   
## n= 380, number of events= 75   
##   
## coef exp(coef) se(coef) z Pr(>|z|)   
## raceBlack non-Hispanic 0.917290 2.502501 0.334301 2.744 0.00607 \*\*  
## raceOther Ethnicity -0.053662 0.947753 0.433777 -0.124 0.90155   
## educationCollege Graduate -0.656311 0.518761 0.320016 -2.051 0.04028 \*   
## educationSome College 0.003105 1.003110 0.308536 0.010 0.99197   
## educationHigh School or less -0.662156 0.515738 0.407687 -1.624 0.10434   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## exp(coef) exp(-coef) lower .95 upper .95  
## raceBlack non-Hispanic 2.5025 0.3996 1.2996 4.8187  
## raceOther Ethnicity 0.9478 1.0551 0.4050 2.2178  
## educationCollege Graduate 0.5188 1.9277 0.2771 0.9713  
## educationSome College 1.0031 0.9969 0.5479 1.8364  
## educationHigh School or less 0.5157 1.9390 0.2320 1.1467  
##   
## Concordance= 0.585 (se = 0.038 )  
## Likelihood ratio test= 11.98 on 5 df, p=0.03  
## Wald test = 12.63 on 5 df, p=0.03  
## Score (logrank) test = 13.05 on 5 df, p=0.02

# (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(intake\_age, menopause\_age, menopause) ~ race + education, data = dd)  
  
summary(cox\_model\_ref\_other)

## Call:  
## coxph(formula = Surv(intake\_age, menopause\_age, menopause) ~   
## race + education, data = dd)  
##   
## n= 380, number of events= 75   
##   
## coef exp(coef) se(coef) z Pr(>|z|)   
## raceWhite non-Hispanic 0.053662 1.055128 0.433777 0.124 0.9015   
## raceBlack non-Hispanic 0.970952 2.640458 0.504297 1.925 0.0542 .  
## educationCollege Graduate -0.656311 0.518761 0.320016 -2.051 0.0403 \*  
## educationSome College 0.003105 1.003110 0.308536 0.010 0.9920   
## educationHigh School or less -0.662156 0.515738 0.407687 -1.624 0.1043   
## ---  
## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1  
##   
## exp(coef) exp(-coef) lower .95 upper .95  
## raceWhite non-Hispanic 1.0551 0.9478 0.4509 2.4691  
## raceBlack non-Hispanic 2.6405 0.3787 0.9827 7.0948  
## educationCollege Graduate 0.5188 1.9277 0.2771 0.9713  
## educationSome College 1.0031 0.9969 0.5479 1.8364  
## educationHigh School or less 0.5157 1.9390 0.2320 1.1467  
##   
## Concordance= 0.585 (se = 0.038 )  
## Likelihood ratio test= 11.98 on 5 df, p=0.03  
## Wald test = 12.63 on 5 df, p=0.03  
## Score (logrank) test = 13.05 on 5 df, p=0.02

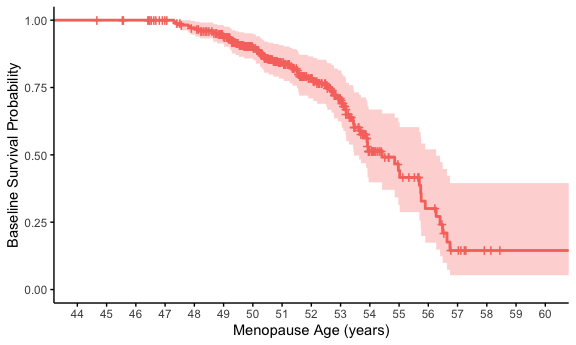
# Extract coefficients and standard errors  
#coef\_black <- coef(cox\_model\_age)["raceBlack non-Hispanic"]  
#coef\_other <- coef(cox\_model\_age)["raceOther Ethnicity"]  
#SE\_black <- summary(cox\_model\_age)$coef["raceBlack non-Hispanic", #"se(coef)"]  
#SE\_other <- summary(cox\_model\_age)$coef["raceOther 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")  
#cat("95% Confidence Interval: (", lower\_CI, ",", upper\_CI, ")\n")  
#  
#vcov\_matrix <- vcov(cox\_model\_age)  
#vcov\_matrix["raceBlack non-Hispanic", "raceOther Ethnicity"]  
  
  
# (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  
## 47.3 182 1 0.994 0.00606 0.9822 1.000  
## 47.3 182 1 0.988 0.00857 0.9713 1.000  
## 47.5 190 1 0.982 0.01039 0.9621 1.000  
## 47.8 202 1 0.977 0.01175 0.9541 1.000  
## 47.9 202 1 0.972 0.01297 0.9465 0.997  
## 48.0 208 1 0.966 0.01406 0.9393 0.994  
## 48.2 213 1 0.961 0.01503 0.9325 0.991  
## 48.6 220 1 0.957 0.01590 0.9260 0.988  
## 48.7 219 1 0.952 0.01673 0.9196 0.985  
## 48.8 222 1 0.947 0.01750 0.9135 0.982  
## 49.0 224 1 0.942 0.01824 0.9074 0.979  
## 49.0 221 1 0.938 0.01897 0.9013 0.976  
## 49.2 216 1 0.933 0.01970 0.8951 0.972  
## 49.2 214 1 0.928 0.02043 0.8889 0.969  
## 49.3 213 1 0.923 0.02113 0.8827 0.966  
## 49.3 212 1 0.918 0.02181 0.8766 0.962  
## 49.4 205 1 0.913 0.02251 0.8704 0.959  
## 49.5 202 1 0.908 0.02318 0.8641 0.955  
## 49.7 194 1 0.903 0.02388 0.8577 0.951  
## 50.0 187 1 0.898 0.02457 0.8512 0.948  
## 50.1 187 1 0.893 0.02525 0.8449 0.944  
## 50.1 186 1 0.888 0.02591 0.8385 0.940  
## 50.2 180 1 0.883 0.02658 0.8320 0.936  
## 50.3 179 1 0.877 0.02724 0.8256 0.932  
## 50.3 171 1 0.872 0.02794 0.8189 0.928  
## 50.4 167 1 0.866 0.02864 0.8120 0.924  
## 50.4 165 1 0.861 0.02934 0.8052 0.920  
## 50.5 159 1 0.855 0.03009 0.7981 0.916  
## 50.7 152 1 0.849 0.03087 0.7907 0.912  
## 50.9 144 1 0.843 0.03171 0.7829 0.907  
## 51.1 133 1 0.836 0.03263 0.7745 0.903  
## 51.3 126 1 0.829 0.03364 0.7655 0.898  
## 51.4 125 1 0.822 0.03461 0.7567 0.892  
## 51.5 119 1 0.814 0.03563 0.7474 0.887  
## 51.5 117 1 0.807 0.03661 0.7383 0.882  
## 51.6 117 1 0.800 0.03757 0.7292 0.877  
## 51.6 114 1 0.792 0.03855 0.7200 0.871  
## 51.9 103 1 0.784 0.03961 0.7100 0.865  
## 52.1 100 1 0.776 0.04065 0.6999 0.860  
## 52.2 94 1 0.767 0.04172 0.6895 0.853  
## 52.5 90 1 0.758 0.04284 0.6786 0.847  
## 52.5 87 1 0.749 0.04400 0.6674 0.840  
## 52.7 87 1 0.740 0.04511 0.6565 0.834  
## 52.8 87 1 0.731 0.04616 0.6459 0.827  
## 52.8 86 1 0.722 0.04718 0.6351 0.821  
## 52.9 82 1 0.713 0.04822 0.6240 0.814  
## 53.0 77 1 0.703 0.04936 0.6123 0.806  
## 53.0 75 1 0.693 0.05048 0.6004 0.799  
## 53.1 72 1 0.682 0.05159 0.5883 0.791  
## 53.2 69 1 0.672 0.05273 0.5758 0.783  
## 53.2 67 1 0.661 0.05384 0.5633 0.775  
## 53.2 66 1 0.650 0.05491 0.5508 0.767  
## 53.3 63 1 0.639 0.05604 0.5380 0.759  
## 53.4 59 1 0.627 0.05727 0.5242 0.750  
## 53.4 58 1 0.615 0.05843 0.5106 0.741  
## 53.5 57 1 0.603 0.05953 0.4970 0.732  
## 53.6 53 1 0.591 0.06069 0.4831 0.723  
## 53.7 50 1 0.578 0.06183 0.4689 0.713  
## 53.9 43 1 0.564 0.06348 0.4519 0.703  
## 53.9 40 1 0.548 0.06511 0.4344 0.692  
## 53.9 39 1 0.533 0.06667 0.4171 0.681  
## 54.0 37 1 0.516 0.06824 0.3980 0.668  
## 54.4 27 1 0.492 0.07141 0.3701 0.654  
## 54.9 26 1 0.467 0.07427 0.3420 0.638  
## 55.0 24 1 0.442 0.07677 0.3141 0.621  
## 55.0 23 1 0.416 0.07874 0.2873 0.603  
## 55.7 18 1 0.387 0.08143 0.2562 0.584  
## 55.7 17 1 0.357 0.08325 0.2260 0.564  
## 55.8 16 1 0.328 0.08402 0.1987 0.542  
## 55.9 16 1 0.301 0.08420 0.1737 0.521  
## 56.3 15 1 0.271 0.08370 0.1481 0.497  
## 56.4 14 1 0.242 0.08274 0.1235 0.473  
## 56.5 12 1 0.210 0.08098 0.0987 0.447  
## 56.6 10 1 0.176 0.07859 0.0734 0.422  
## 56.7 9 1 0.145 0.07417 0.0532 0.395

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

## Estimated median menopause age (Kaplan-Meier): 54.42026 years.

p <- ggsurvplot(base\_surv, conf.inf = T, data=df,xlim=c(44,60),ggtheme = theme\_classic(), xlab="Menopause Age (years)", ylab="Baseline Survival Probability", legend = "none")  
p$plot +  
 scale\_x\_continuous(breaks = seq(44, 60, by = 1))



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

## chisq df p  
## race 1.869 2 0.39  
## education 0.924 3 0.82  
## GLOBAL 3.165 5 0.67