Asignacion 2 Demografia

Keyla Zúñiga Martínez

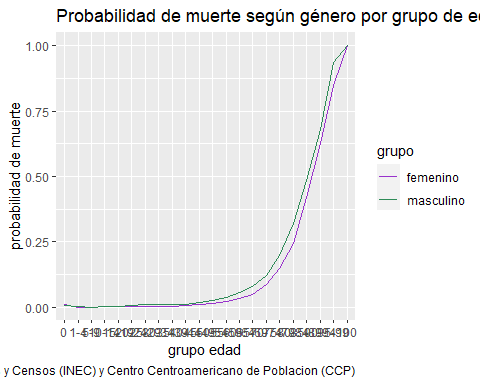
2023-06-19

library(readxl)  
library(ggplot2)  
demo2 <- read\_excel("C:/Users/User/Desktop/demo/demo2.xlsx",   
 sheet = "Hoja2")

## New names:  
## • `` -> `...4`  
## • `` -> `...5`

# Pregunta 2

demo2$edad = as.character(demo2$edad)  
demo2$edad = factor(demo2$edad, levels = unique(demo2$edad))  
  
ggplot(demo2, aes(x= edad, y= prob, group= grupo, colour=grupo))+   
 geom\_line()+  
 xlab("grupo edad")+ ylab("probabilidad de muerte")+  
 labs(title = "Probabilidad de muerte según género por grupo de edad", caption = "Fuente: Instituto Nacional de Estadisticas y Censos (INEC) y Centro Centroamericano de Poblacion (CCP)") +  
 scale\_color\_manual(values = c("femenino" = "#9932CC", "masculino" = "#2E8B57")) +  
 theme\_get()

 La probabilidad de muerte de los hombres empieza a cambiar en las edades a partir de los 45 en adelante, esta tiende a ser mayor, una de las razones es porque la tasa de mortalidad masculina es más alta a comparacion de las mujeres, por ende las probabilidades de morir tambien tienden a ser más altas.

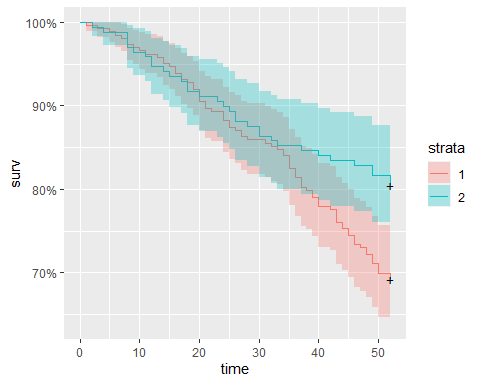
load("C:/Users/User/Downloads/reincidencia.Rdata")  
  
library(survival)  
library(ggfortify)

## Warning: package 'ggfortify' was built under R version 4.3.1

educa <- ifelse(reincidencia$educ %in% c(2, 3), 1, 2)  
  
km = survfit(Surv(week, arrest)~educa, stype = 1, data = reincidencia)  
summary(km)

## Call: survfit(formula = Surv(week, arrest) ~ educa, data = reincidencia,   
## stype = 1)  
##   
## educa=1   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 1 263 1 0.996 0.00380 0.989 1.000  
## 3 262 1 0.992 0.00536 0.982 1.000  
## 5 261 1 0.989 0.00655 0.976 1.000  
## 6 260 1 0.985 0.00755 0.970 1.000  
## 7 259 1 0.981 0.00842 0.965 0.998  
## 8 258 2 0.973 0.00993 0.954 0.993  
## 9 256 1 0.970 0.01059 0.949 0.991  
## 10 255 1 0.966 0.01121 0.944 0.988  
## 11 254 1 0.962 0.01179 0.939 0.985  
## 13 253 1 0.958 0.01234 0.934 0.983  
## 14 252 2 0.951 0.01337 0.925 0.977  
## 15 250 1 0.947 0.01384 0.920 0.974  
## 16 249 2 0.939 0.01474 0.911 0.969  
## 17 247 2 0.932 0.01557 0.902 0.963  
## 18 245 1 0.928 0.01596 0.897 0.960  
## 19 244 2 0.920 0.01671 0.888 0.954  
## 20 242 4 0.905 0.01809 0.870 0.941  
## 21 238 2 0.897 0.01872 0.861 0.935  
## 22 236 1 0.894 0.01902 0.857 0.932  
## 24 235 3 0.882 0.01988 0.844 0.922  
## 25 232 2 0.875 0.02043 0.835 0.915  
## 26 230 1 0.871 0.02069 0.831 0.912  
## 27 229 2 0.863 0.02119 0.823 0.906  
## 28 227 1 0.859 0.02144 0.818 0.902  
## 31 226 1 0.856 0.02168 0.814 0.899  
## 32 225 1 0.852 0.02191 0.810 0.896  
## 33 224 1 0.848 0.02214 0.806 0.892  
## 34 223 2 0.840 0.02259 0.797 0.886  
## 35 221 4 0.825 0.02342 0.780 0.872  
## 36 217 3 0.814 0.02401 0.768 0.862  
## 37 214 3 0.802 0.02456 0.756 0.852  
## 38 211 1 0.798 0.02474 0.751 0.848  
## 39 210 2 0.791 0.02508 0.743 0.842  
## 40 208 3 0.779 0.02557 0.731 0.831  
## 42 205 1 0.776 0.02572 0.727 0.828  
## 43 204 4 0.760 0.02632 0.711 0.814  
## 44 200 2 0.753 0.02660 0.702 0.807  
## 45 198 2 0.745 0.02687 0.694 0.800  
## 46 196 3 0.734 0.02725 0.682 0.789  
## 47 193 1 0.730 0.02737 0.678 0.786  
## 48 192 2 0.722 0.02761 0.670 0.779  
## 49 190 3 0.711 0.02795 0.658 0.768  
## 50 187 3 0.700 0.02827 0.646 0.757  
## 52 184 2 0.692 0.02847 0.638 0.750  
##   
## educa=2   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 2 169 1 0.994 0.00590 0.983 1.000  
## 4 168 1 0.988 0.00832 0.972 1.000  
## 8 167 3 0.970 0.01303 0.945 0.996  
## 9 164 1 0.964 0.01423 0.937 0.993  
## 11 163 1 0.959 0.01533 0.929 0.989  
## 12 162 2 0.947 0.01727 0.913 0.981  
## 14 160 1 0.941 0.01815 0.906 0.977  
## 15 159 1 0.935 0.01898 0.898 0.973  
## 17 158 1 0.929 0.01976 0.891 0.969  
## 18 157 2 0.917 0.02120 0.877 0.960  
## 20 155 1 0.911 0.02188 0.869 0.955  
## 23 154 1 0.905 0.02252 0.862 0.951  
## 24 153 1 0.899 0.02314 0.855 0.946  
## 25 152 1 0.893 0.02373 0.848 0.941  
## 26 151 2 0.882 0.02485 0.834 0.932  
## 28 149 1 0.876 0.02538 0.827 0.927  
## 30 148 2 0.864 0.02638 0.814 0.917  
## 32 146 1 0.858 0.02685 0.807 0.912  
## 33 145 1 0.852 0.02731 0.800 0.907  
## 37 144 1 0.846 0.02775 0.793 0.902  
## 40 143 1 0.840 0.02818 0.787 0.897  
## 42 142 1 0.834 0.02860 0.780 0.892  
## 46 141 1 0.828 0.02900 0.773 0.887  
## 49 140 2 0.817 0.02977 0.760 0.877  
## 52 138 2 0.805 0.03049 0.747 0.867

autoplot(km)



survdiff(Surv(week, arrest)~educ, data = reincidencia)

## Call:  
## survdiff(formula = Surv(week, arrest) ~ educ, data = reincidencia)  
##   
## N Observed Expected (O-E)^2/E (O-E)^2/V  
## educ=2 24 4 6.81 1.157 1.24  
## educ=3 239 77 61.54 3.884 8.49  
## educ=4 119 27 31.59 0.668 0.93  
## educ=5 39 5 10.78 3.102 3.45  
## educ=6 11 1 3.28 1.581 1.64  
##   
## Chisq= 10.5 on 4 degrees of freedom, p= 0.03

quantile(km, quantiles= c(0.25, 0.5, 0.75))

## $quantile  
## 25 50 75  
## educa=1 45 NA NA  
## educa=2 NA NA NA  
##   
## $lower  
## 25 50 75  
## educa=1 39 NA NA  
## educa=2 52 NA NA  
##   
## $upper  
## 25 50 75  
## educa=1 NA NA NA  
## educa=2 NA NA NA

Una de las razones por la que el calculo de la mediana de “NA” es que mas de la mitad de la muestra no pasaron por el evento en el tiempo de interes, entoces, no se puede calcular un punto medio de sobrevivencia de las observaciones

Con respecto a las curvas, en las edades menores a 35 años no se logra ver una diferencia de la sobrevivencia de acuerdo a la educacion, sin embargo, se observa que entre más avance la edad de la poblacion las diferencias en la supervivencia por educacion tienden a ser mas visibles, es decir, la educacion 4,5 y 6 tienen a sobrevivir más.

# pregunta 4

library(readxl)  
base <- read\_excel("C:/Users/User/Downloads/asignacion2 (1).xlsx")

base$evento = base$terminoobs  
base$evento = as.numeric(base$evento)  
  
base$evento[is.na(base$evento)] = 0  
base$evento = ifelse(base$evento == 0, 0, 1)  
  
base$fecha = base$terminoobs  
  
diferencia1 = base$terminoobs - base$entradaobs  
diferencia = base$entrevista-base$entradaobs  
  
base$fecha = ifelse(base$evento == 0, diferencia, diferencia1)  
  
observacion.km = survfit(Surv(fecha,evento)~1, stype = 1, data = base)  
summary(observacion.km)

## Call: survfit(formula = Surv(fecha, evento) ~ 1, data = base, stype = 1)  
##   
## time n.risk n.event survival std.err lower 95% CI upper 95% CI  
## 8 21 1 0.952 0.0465 0.8655 1.000  
## 13 18 1 0.899 0.0676 0.7763 1.000  
## 20 17 1 0.847 0.0818 0.7006 1.000  
## 21 16 1 0.794 0.0922 0.6321 0.997  
## 35 13 1 0.733 0.1034 0.5556 0.966  
## 37 12 1 0.672 0.1113 0.4853 0.929  
## 40 10 1 0.604 0.1187 0.4113 0.888  
## 58 9 1 0.537 0.1231 0.3429 0.842  
## 73 7 1 0.460 0.1272 0.2680 0.791  
## 97 4 1 0.345 0.1380 0.1578 0.756  
## 102 3 1 0.230 0.1315 0.0752 0.705  
## 108 2 1 0.115 0.1046 0.0194 0.684

autoplot(observacion.km)

