

PC2 - Analisis de Supervivencia y Confiabilidad

Estos datos del Registro de Tumores de Connecticut presentan las cifras ajustadas a la edad de las incidencias de cáncer de piel de melanoma por cada 100.000 personas en Connecticut para los años de 1936 a 1972.

La base de datos melanoma contiene un total de 205 observaciones de 6 variables:

1. **no.** código del paciente.
2. **status.** el estado de los pacientes y el final del estudio (1=muere por melanoma maligno, 2=vive al final del estudio, 3=muere por otras causas).
3. **days.** tiempo de estudio después de la operación con melanoma maligno.
4. **ulc.** superficie del melanoma vista por un microscopio si muestra ulceración (1=Presente, 2=Ausencia).
5. **thick.** el grosor del tumor.
6. **sex.** género de los pacientes (1=Femenino, 2=Masculino).

- 1) Realice las pruebas Logrank, Wilcoxon y Tarone-Ware para la igualdad de las dos funciones de supervivencia estimadas para el grupo Masculino y Femenino.

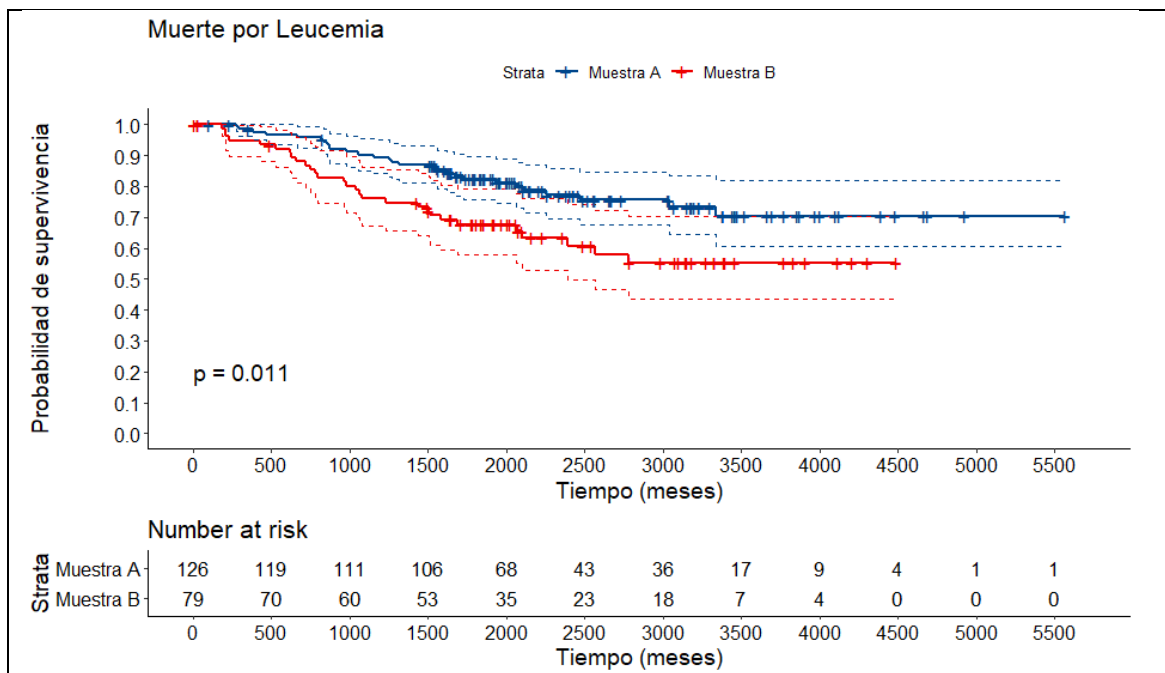
```
ggsurvplot(mod1,
  data=datos,
  surv.median.line = "hv", # Especificamos la mediana de la supervivencia
  conf.int = T, # Intervalo de confianza
  conf.int.style="step", # Tipo de visualización del IC
  pval.coord=c(0,0.20), # Coordenadas del p value
  risk.table = T, # Añadir tabla de eventos/ censurados
  pval=TRUE, # Añadir el valor p del método de comparación
  size=1, # Tamaño de línea
  palette = "lancet", # Paleta de colores
  #ggtheme = theme_hc(), # Tema de estructura
  xlab="Tiempo (meses)", # Nombre del eje X
  ylab="Probabilidad de supervivencia", # Nombre del eje Y
  title="Muerte por Leucemia", # Título
  legend.labs=c("Muestra A", "Muestra B"), # Nombre de categoría
  xlim=c(0,5700), # Escala en el eje X
  ylim=c(0,1.0), # Escala del eje Y (Siempre dejar en 1.0)
  break.y.by=c(0.1), # Particiones del eje Y
  break.x.by=c(500))+ # Particiones del eje X
  theme_survminer(base_size = 10, base_family = "Arial")

# Prueba de LogRank
LogRank<-survdif(Surv(days,status)~sex,data=datos, rho = 0);LogRank
# p= 0.008

# Prueba de Wilcoxon
Wilcoxon<-survdif(Surv(days,status)~sex,data=datos, rho = 1);Wilcoxon
# p= 0.008

# Prueba de tarone-ware
taerone<-survdif(Surv(days,status)~sex,data=datos, rho = 0.5);taerone
# p= 0.009
```

PC2 - Analisis de Supervivencia y Confiabilidad



- 2) Empleando el Software R para grupo femenino, masculino y ambos encuentre: a) La función de densidad Exponencial o Weibull (gráfico) b) La función de distribución acumulada (gráfico) c) La función de supervivencia (gráfico) d) La función de riesgo (gráfico) e) Función de riesgo acumulado (gráfico) f) Interpretar cada gráfica.

```
# a) La función de densidad Exponencial o Weibull

weib <- flexsurvreg(Surv(days, status) ~ 1, data = datos, dist = "weibull");weib
#Ajustando a una dist. Weibull
summary(weib)

expo <- flexsurvreg(Surv(days, status) ~ 1, data = datos, dist = "exp") #Ajuste
exponencial
expo
estimate<-data.frame(expo$res)
theta<-1/estimate$est
paste("La media del tiempo =",theta)

# HIPOTESIS
# H0: Tiempos de estudio después de la operación con melanoma maligno se ajusta a una
distribución Exponencial
# H1: Tiempos de estudio después de la operación con melanoma maligno se ajusta a una
distribución Weibull

Razon_ver<- -2*expo$loglik+2*weib$loglik ;Razon_ver #pvalor
chi_c<-qchisq(0.95,1);chi_c
# Como Razon_ver (0.4503844) menor a Chi_c (3.841459) no se rechaza la H0,
# la distribución que mejor ajuste al grupo tratado es la exponencial.

# Gráfica de la función densidad Exponencial
f_t<-((1/theta)*exp(-datos$days/theta))
windows()
fig1<-plot(datos$days,f_t, main="a) Función Exponencial",
xlab="Tiempo ", ylab="Probabilidad",col="52", type="o")
# La función de densidad es creciente. a mayor tiempo mayor será su
# probabilidad para cada paciente

#####
# b) La función de distribución acumulada

F_t<-1-exp(-datos$days/theta)
```

PC2 - Analisis de Supervivencia y Confiabilidad

```

windows()
fig2<-plot(datos$days,F_t, main="b) Función de distribución acumulada",
          xlab="Tiempo de falla", ylab="Probabilidad de fallar antes de t",
          col="52",type="o")

# Función de distr. acumulada es creciente, cuando más aumente el tiempo
# mayor es la probabilidad de morir antes de un tiempo T determinado.

#####
# c) La función de sobrevivencia

S_t<-exp(-datos$days/theta)
windows()
fig3<-plot(datos$days,S_t, main="c) Función de confiabilidad",
          xlab="Tiempo de falla", ylab="Probabilidad sobrevivir al tiempo t",
          col="52",type="o")

# La probabilidad de sobrevivir antes de un tiempo T determinado es decreciente.
# mientras más aumenta el tiempo T menos probabilidad de sobrevivir al tiempo T tiene

#####
# d) La función de riesgo

alfa<-1
h_t<-(alfa/theta)*(datos$days/theta)^(alfa-1)
windows()
fig4<-plot(datos$days,h_t, main="Función de riesgo Exponencial",
          xlab="Tiempo de falla", ylab="Tasa de riesgo al tiempo t",
          col="52",type="o")

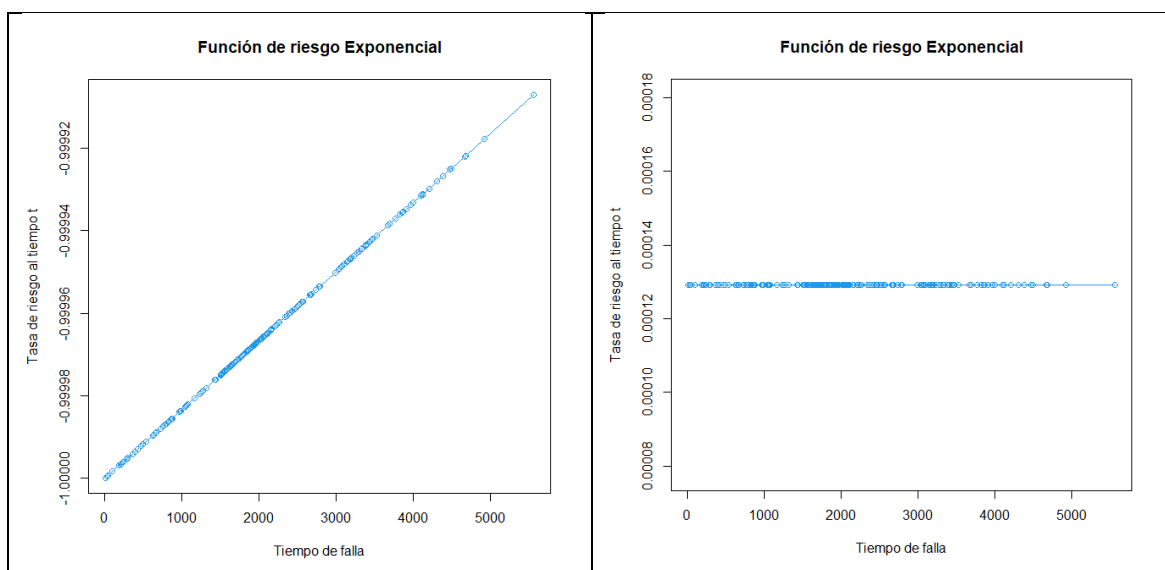
# Debido a que los datos se asemejan a un modelo paramétrico exponencial
# su función de densidad de riesgo va a ser constante para cada observación

#####
# e) Función de riesgo acumulado

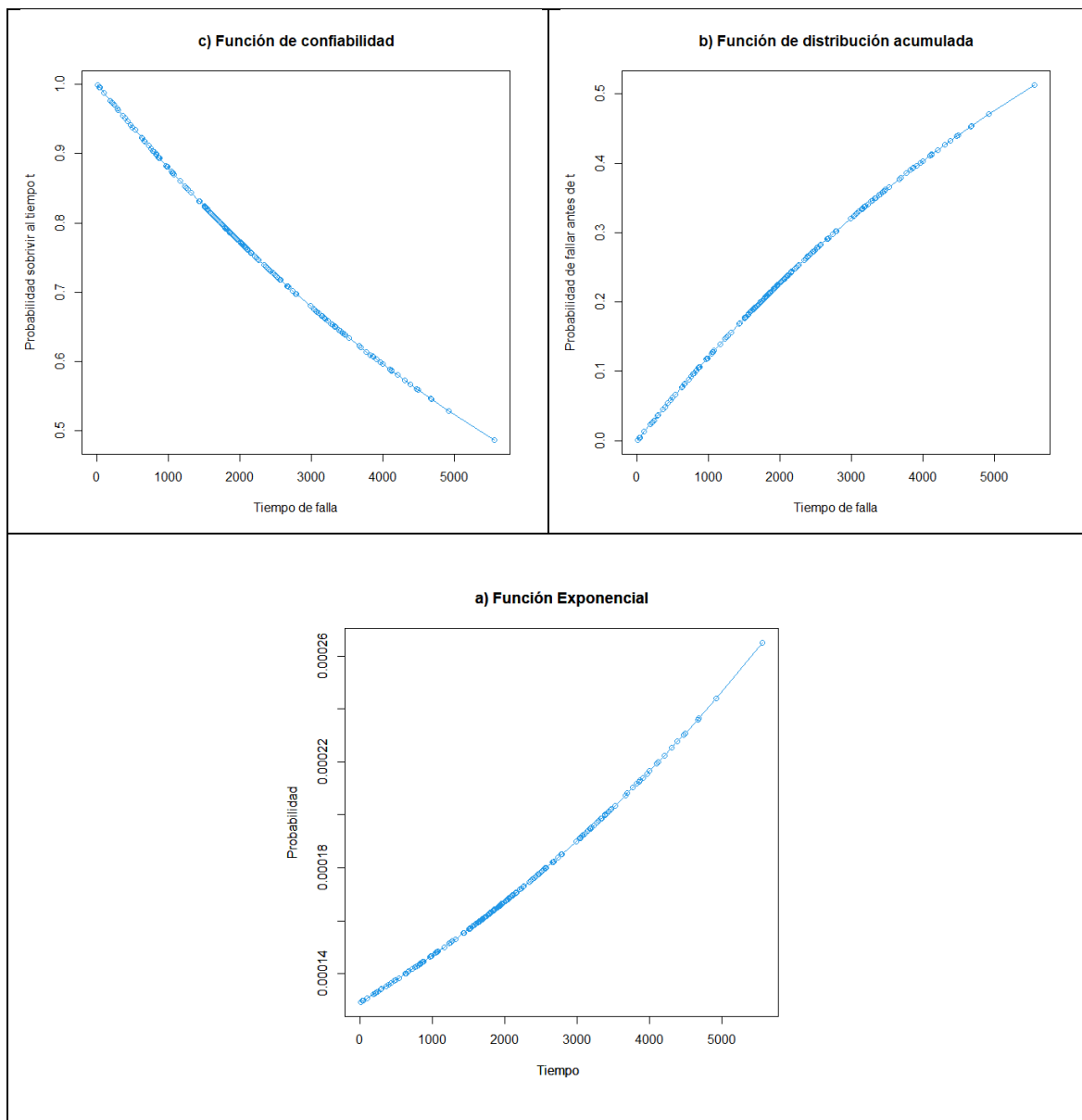
h_t<-(alfa/theta)*(datos$days/theta-theta) #función de riesgo integrada con alfa=1
windows()
fig4<-plot(datos$days,h_t, main="Función de riesgo Exponencial",
          xlab="Tiempo de falla", ylab="Tasa de riesgo al tiempo t",
          col="52",type="o")

# Como estamos observando un gráfico de función de riesgo acumulativo
# mientras aumente el tiempo también aumentara las tasas de riesgo para
# los pacientes con melanoma.

```



PC2 - Analisis de Supervivencia y Confiabilidad

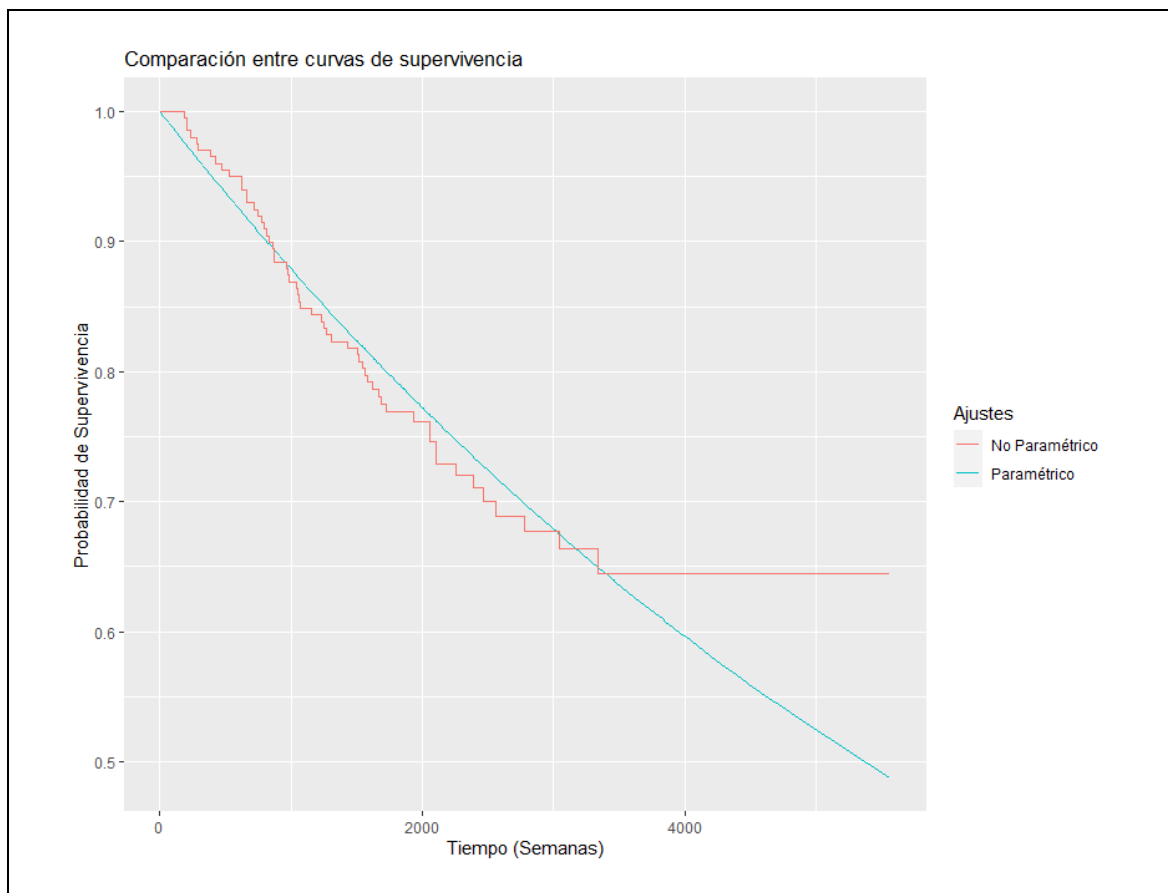


- 3) Compare la gráfica de función de supervivencia paramétrica y no paramétrica para grupo femenino, masculino y ambos.

```
flexgg<-data.frame(summary(expo,type="survival"))
km <- survfit(Surv(datos$days, status) ~ 1, data = datos)
kmgg<-data.frame(time=km$time,surv=km$surv)

windows()
ggplot() + geom_line(aes(time, est, col = "Paramétrico"), data =
flexgg) +
geom_step(aes(time,surv, col = "No Paramétrico"), data = kmgg) +
labs(x = "Tiempo (Semanas)",y = "Probabilidad de Supervivencia",
col = "Ajustes", title = "Comparación entre curvas de
supervivencia")
```

PC2 - Analisis de Supervivencia y Confiabilidad



4) Encuentre el mejor modelo de riesgos proporcionales; interprete los coeficientes.

```
survival::coxph(formula=Surv(days, status)~sex, data=datos)
Surv(datos$days, datos$status) # signo+ =Censurados
m1<-survival::coxph(formula=Surv(days, status)~sex, data=datos)
summary(m1)

summary(m1)$concordance
#           C           se(C)
# 0.5902217 0.0339631

m2<-survival::coxph(formula=Surv(days, status)~sex+ulc+thick, data=datos)
summary(m2)
#           coef exp(coef) se(coef)      z Pr(>|z|)
# sex    0.4594907 1.5832675 0.2667580 1.723  0.08498 .
# ulc    1.1668079 3.2117240 0.3114615 3.746  0.00018 ***
# thick  0.0011345 1.0011351 0.0003794 2.990  0.00279 **

# Las mujeres tienen 58.33% mas probabilidad de morir que los hombres por melanoma
# La ulc:(1-0.3114)= 0.6886 hay un 68.86 % menos riesgo de morir de la mujeres
# Por cada unidad de aumento del puntaje thick hay la misma
# probabilidad de morir por melanoma para las mujeres

m3<-survival::coxph(formula=Surv(days, status)~sex+thick, data=datos)
summary(m3)
summary(m2)$concordance; summary(m3)$concordance
# Ho: No hay diferencia entre el m2 y m3
# Ha: Hay diferencia entre el m2 y m3
anova(m3,m2)
# Existen diferencias entre el modelo m2 y m3
# BIC: Aquél modelo que tenga menor valor de BIC es el mejor
# modelo
BIC(m2); BIC(m3)
```

PC2 - Analisis de Supervivencia y Confiabilidad

```
*** El modelo 2 (solo ulc y thick) es el mejor

m4<-survival::coxph(formula=Surv(days, status)~sex+ulc, data=datos)
summary(m4)
summary(m2)$concordance; summary(m4)$concordance
# Ho: No hay diferencia entre el m3 y m4
# Ha: Hay diferencia entre el m3 y m4
anova(m2,m4)
#existen diferencias entre el modelo m2 y m4
BIC(m2); BIC(m4)

*** El modelo 2 es mejor

anova(m1,m2)
#existen diferencias entre el modelo m1 y m2
BIC(m1); BIC(m2) # El modelo 2 es mejor (ulc)
summary(m2)$concordance

# Concordance = 0.763 (se = 0.0336 )

# > Concordancia del modelo con los datos
# En este caso concuerdan en un 76.3%; es decir,
# El 76.3% de las personas han sido predichas en forma correcta por el
# modelo.con un 3.36% de desviación estándar
```