

Universidad de Granada

Escuela Internacional de Posgrado

Máster en Estadística Aplicada

Materia: Técnicas en Análisis de Supervivencia.

Alumno: Francisco Javier Márquez Rosales



**UNIVERSIDAD  
DE GRANADA**

## Tema 5: Modelos de Regresión:

### Ejercicios:

Diciembre, 2022

### Actividad:

Se considera la v.a. tiempo de desempleo. Se sabe que ciertos factores influyen de manera decisiva en el tiempo que un individuo permanece desempleado. Se considera una muestra de 40 individuos inscritos en en la oficina de empleo de una determinada localidad y se registra el tiempo transcurrido hasta que, bien dejan de estar desempleados porque encuentran trabajo (censura=1), bien dejan de aparecer en la lista de desempleados por alguna otra razón (censura=0). Asimismo, para cada individuo se registra información referente a su edad, sexo, sector profesional en el que desarrolla su actividad. Los datos se presentan en la tabla adjunta.

Ajustar un modelo de riesgos proporcionales con función de riesgo base constante y discutir los resultados.

Tiempo desempleo	Sexo	Edad	Censura	sector
36	MUJER	40	1	OTROS
25,32	MUJER	48	0	OTROS
46,2	MUJER	50	0	OTROS
10,32	MUJER	50	0	OTROS
30,72	MUJER	39	0	AGRICULTURA
9,6	MUJER	45	0	AGRICULTURA
32,54	MUJER	45	0	AGRICULTURA
60	MUJER	60	1	AGRICULTURA
8,23	MUJER	60	0	CONSTRUCCION
60	MUJER	60	1	CONSTRUCCION

(se muestran los primeros 10 registros)

### Solución:

En primer lugar hacemos la lectura de los datos y revisamos inicialmente su estructura y contenido.

```
empleo<-read.csv("empleo.csv",header = TRUE)
str(empleo)

## 'data.frame':    40 obs. of  5 variables:
##  $ Tiempo : num  60 60 46.2 36 32.5 ...
```

```
## $ Sexo : chr "MUJER" "MUJER" "MUJER" "MUJER" ...
## $ Edad : int 60 60 50 40 45 39 48 55 45 30 ...
## $ Censura: int 1 1 0 1 0 0 0 1 0 0 ...
## $ Sector : chr "AGRICULTURA" "CONSTRUCCION" "OTROS" "OTROS" ...

summary(empleo)

##      Tiempo      Sexo      Edad      Censura
## Min. : 1.660 Length:40 Min. :18.00 Min. :0.00
## 1st Qu.: 5.402 Class :character 1st Qu.:27.25 1st Qu.:0.00
## Median :11.620 Mode :character Median :45.00 Median :0.00
## Mean :19.283 Mean :39.95 Mean :0.25
## 3rd Qu.:30.180 3rd Qu.:50.00 3rd Qu.:0.25
## Max. :60.000 Max. :62.00 Max. :1.00

##      Sector
## Length:40
## Class :character
## Mode :character
```

Luego, hacemos la preparación de la data. Para esto, recodificamos la variable sexo y Sector de forma de poder incluirlas en el análisis como factores

```
# Creamos la nueva variable
empleo <- empleo %>%
  mutate(Sexo = as.factor(Sexo), "HOMBRE" = 1, "MUJER" = 2,
         Sector = as.factor(Sector), "AGRICULTURA" = 1, "CONSTRUCCION" = 2,
         "OTROS" = 4, "SERVICIOS" = 3)
```

Ahora hacemos el ajuste de un primer modelo, plateamos el modelo solo con una variable predictora, en este caso la variable 'sexo'.

```
# Ajuste del modelo
modelo <- coxph(Surv(Tiempo, Censura) ~ Sexo, data = empleo)

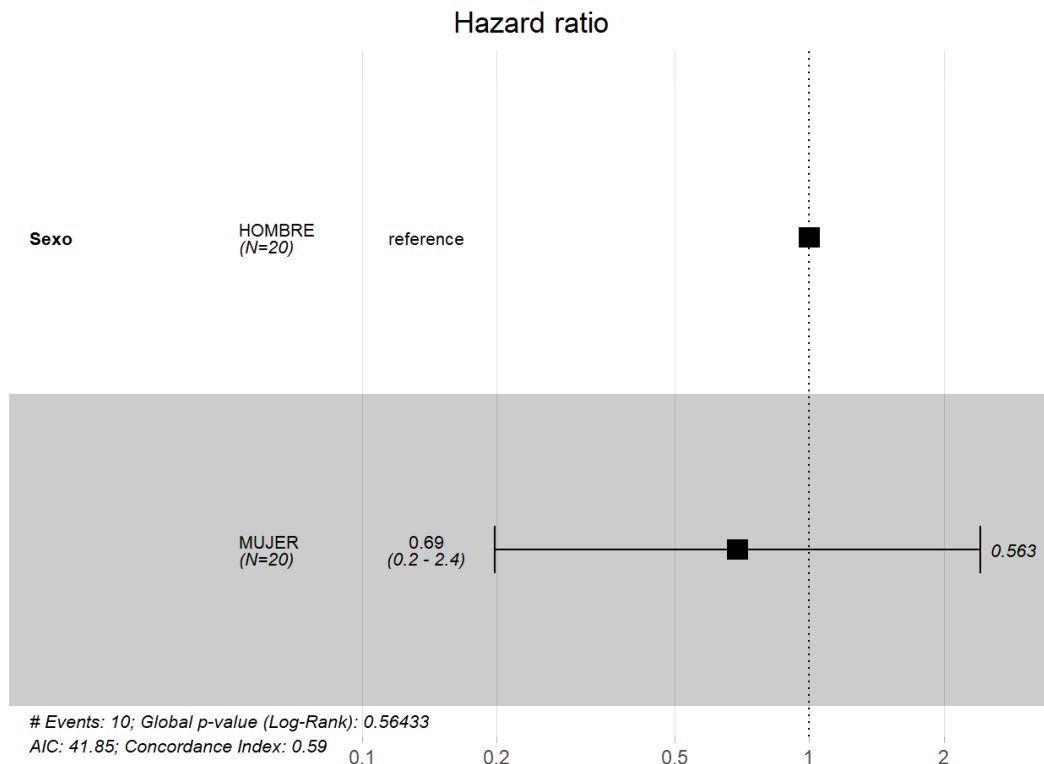
# Resumen del modelo ajustado
modelo

## Call:
## coxph(formula = Surv(Tiempo, Censura) ~ Sexo, data = empleo)
##
##      coef exp(coef) se(coef)      z      p
```

```
## Sexo -0.3690    0.6915    0.6384 -0.578 0.563
##
## Likelihood ratio test=0.33 on 1 df, p=0.5643
## n= 40, number of events= 10
```

El hazard rate asociado con el factor (columna  $\exp(\text{coef})$ ) es menor que 1, lo que indica que las mujeres tienen más posibilidad a no ser empleadas que los hombres. El test de bondad de ajuste del modelo (“likelihood ratio test”) no resulta significativo indicando que, el sexo no contribuye de forma determinante a explicar la supervivencia a la condición de desempleo de los sujetos del estudio.

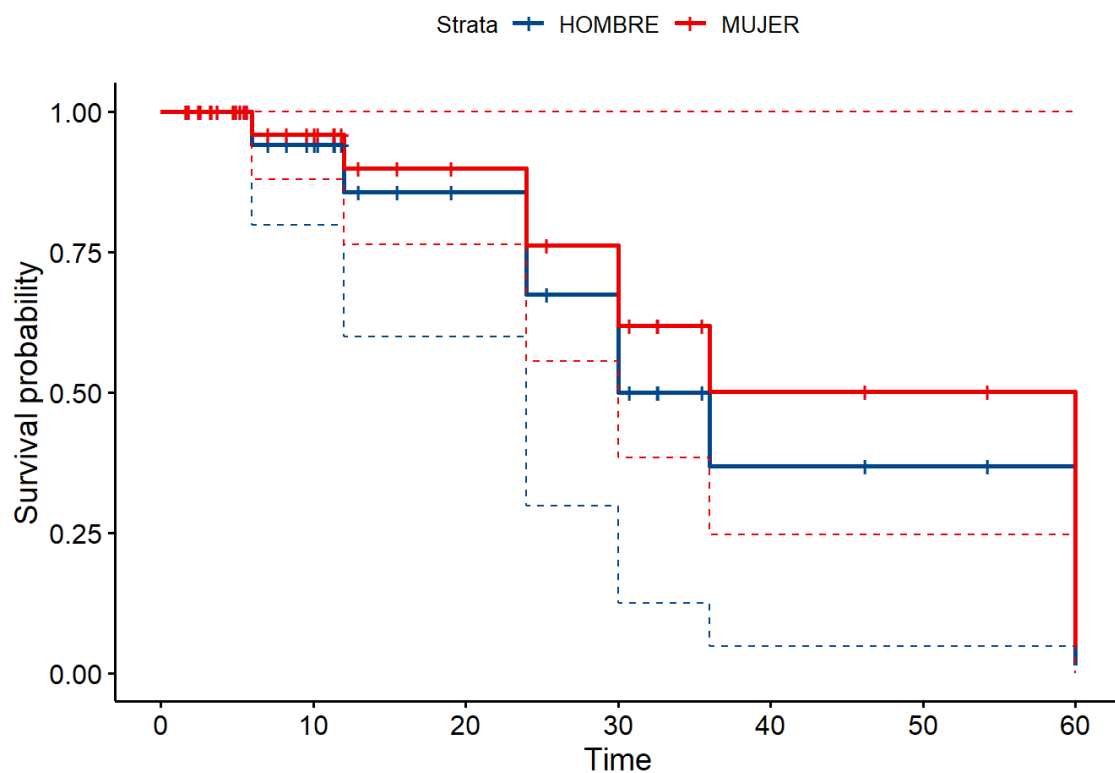
```
# Gráfico de la relevancia de los coeficientes
ggforest(modelo)
```



Como podemos ver se usa la categoría “Hombre” como referencia para la estimación de efectos del modelo. El hazard ratio asociado con las Mujeres se sitúa en 0.69. Por tanto, ser mujer reduce la posibilidad de ser empleado por un factor de 0,69 o equivalentemente un 31% ( $100 \times (1 - 0.69)$ ). Concluimos que ser mujer no es determinante para la obtención de empleo.

Obtenemos el gráfico de supervivencia e incluimos la opción del intervalo de confianza

```
# Gráfico con intervalos de confianza
sex.df <- with(empleo, data.frame(Sexo = c("HOMBRE", "MUJER")))
ggsurvplot(survfit(modelo,newdata = sex.df), data = empleo,
            palette = "lancet",
            conf.int = TRUE,
            conf.int.style = "step",
            legend.labs=c("HOMBRE", "MUJER"))
```



La forma del gráfico respalda la hipótesis de que las mujeres tienen mayor supervivencia en la condición de desempleo que los hombres.

Ahora evaluaremos un modelo ampliado incluyendo las variables sexo, sector y edad covariantes de la data de empleo.

```
# Ajuste del modelo
```

```
modelo2 <- coxph(Surv(Tiempo, Censura) ~ Sector+Sexo+Edad, data = empleo)
```

```
# Resumen del modelo ajustado
```

```
modelo2
```

```
## Call:
```

```
## coxph(formula = Surv(Tiempo, Censura) ~ Sector + Sexo + Edad,
```

```
##      data = empleo)
```

```
##
```

```
##           coef exp(coef) se(coef)      z      p
```

```
## Sector  0.60698   1.83489  0.51972  1.168 0.243
```

```
## Sexo   -0.22349   0.79972  0.68715 -0.325 0.745
```

```
## Edad   -0.01966   0.98053  0.02759 -0.712 0.476
```

```
##
```

```
## Likelihood ratio test=3.93 on 3 df, p=0.2689
```

```
## n= 40, number of events= 10
```

```
summary(modelo2)
```

```
## Call:
```

```
## coxph(formula = Surv(Tiempo, Censura) ~ Sector + Sexo + Edad,
```

```
##      data = empleo2)
```

```
##
```

```
##      n= 40, number of events= 10
```

```
##
```

```
##           coef exp(coef) se(coef)      z Pr(>|z|)
```

```
## Sector  0.60698   1.83489  0.51972  1.168  0.243
```

```
## Sexo   -0.22349   0.79972  0.68715 -0.325  0.745
```

```
## Edad   -0.01966   0.98053  0.02759 -0.712  0.476
```

```
##
```

```
##           exp(coef) exp(-coef) lower .95 upper .95
```

```
## Sector    1.8349      0.545    0.6626    5.082
```

```
## Sexo      0.7997      1.250    0.2080    3.075
```

```
## Edad      0.9805      1.020    0.9289    1.035
```

```
##
```

```
## Concordance= 0.766 (se = 0.078 )
```

```
## Likelihood ratio test= 3.93 on 3 df, p=0.3
```

```
## Wald test              = 3.56 on 3 df, p=0.3
```

```
## Score (logrank) test = 4.11 on 3 df, p=0.3
```

Los valores  $p$  resultantes de los contrastes de significación de los coeficientes del modelo indican que ninguna de las variables consideradas, sexo, edad y sector resulta significativa para la contribución a la supervivencia en condición de desempleo de los individuos.