Análisis de Sobrevivencia

Tarea 2

Christian Badillo

Tabla de contenidos

1 Estimación de Parámetros

1.1 Distribución Exponencial

Se estima el modelo exponencial con la función `survreg` de la librería `survival`.

```
# Estimación de los parámetros.
desing <- Surv(time = survival_time, event = censoring)
exp_fit <- survreg(desing ~ 1,
                   dist = "exponential",
                   data = myData)
```

Tabla 1: Modelo de sobrevivencia con distribución exponencial

term	estimate	std.error	statistic	p.value
(Intercept)	5.950824	0.0393445	151.2493	0

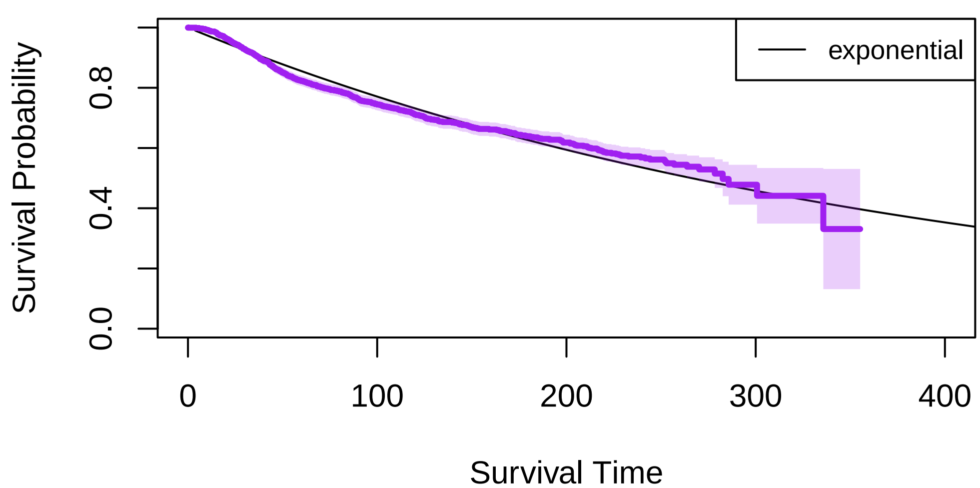

Figura 1: Curva de sobrevivencia estimada para el modelo exponencial.

En la tabla 1 se presentan los resultados de la estimación y en la figura 1 se presenta la función de supervivencia estimada y el estimador Kaplan-Meier, donde se observa que el modelo exponencial se ajusta bastante bien a la curva. El estimador de λ es 5.9508235 que es estadísticamente significativo con un *p-value* $< 2e - 16$.

1.2 Distribución Weibull.

```
# Estimación de los parámetros.
weibull_fit <- survreg(desing ~ 1,
                       dist = "weibull",
                       data = myData)
```

Tabla 2: Modelo de sobrevivencia con distribución Weibull

term	estimate	std.error	statistic	p.value
(Intercept)	5.9413967	0.0500655	118.6724434	0.0000000
Log(scale)	-0.0102885	0.0345763	-0.2975605	0.7660386

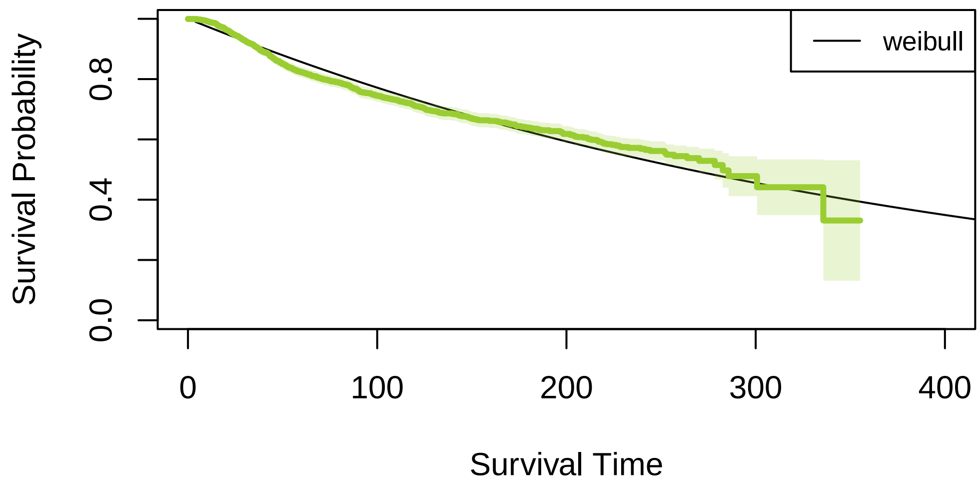

Figura 2: Curva de sobrevivencia estimada para el modelo Weibull.

El ajuste del modelo Weibull y la curva de sobrevivencia estimada se presentan en la tabla 2 y figura 2 respectivamente. El intercepto estimado es muy similar al del modelo exponencial pero cabe destacar que el parámetro de escala es muy cercano a 1 ($e^{\log \alpha} = 0.9897642$) y de hecho no es significativamente distinto a 0 en escala logarítmica, por lo cual podemos decir el modelo Weibull se ajusta bien, pero se reduce a un modelo exponencial.

2 Estimaciones con variable explicatoria

Tabla 3: Modelo de sobrevivencia exponencial con variable explicatoria

term	estimate	std.error	statistic	p.value
(Intercept)	6.2408305	0.0696733	89.57277	0e+00
Type.of.Breast.SurgeryMastectomy	-0.4548707	0.0847027	-5.37020	1e-07

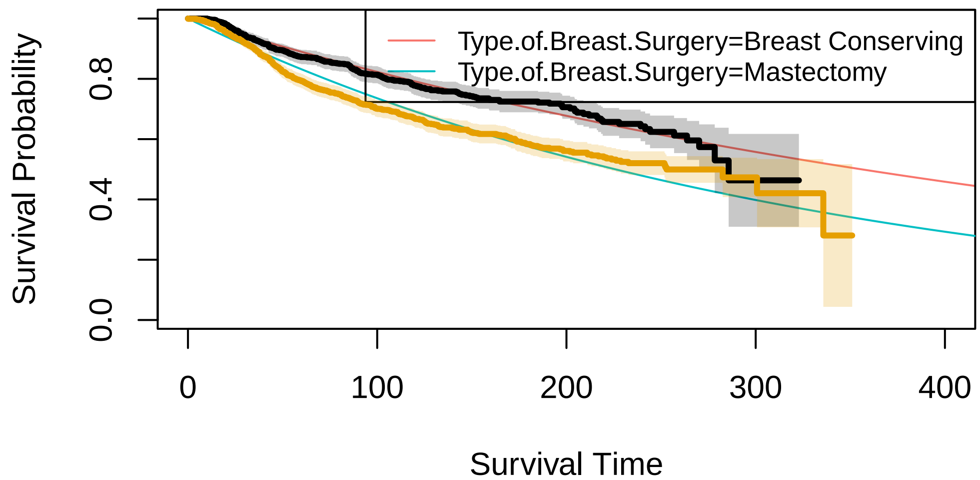

Figura 3: Curvas de sobrevivencia estimadas (exponencial).

Tabla 4: Modelo de sobrevivencia weibull con variable explicatoria

term	estimate	std.error	statistic	p.value
(Intercept)	6.2161963	0.0789339	78.7519630	0.0000000
Type.of.Breast.SurgeryMastectomy	-0.4466316	0.0839892	-5.3177269	0.0000001
Log(scale)	-0.0212468	0.0347070	-0.6121751	0.5404219

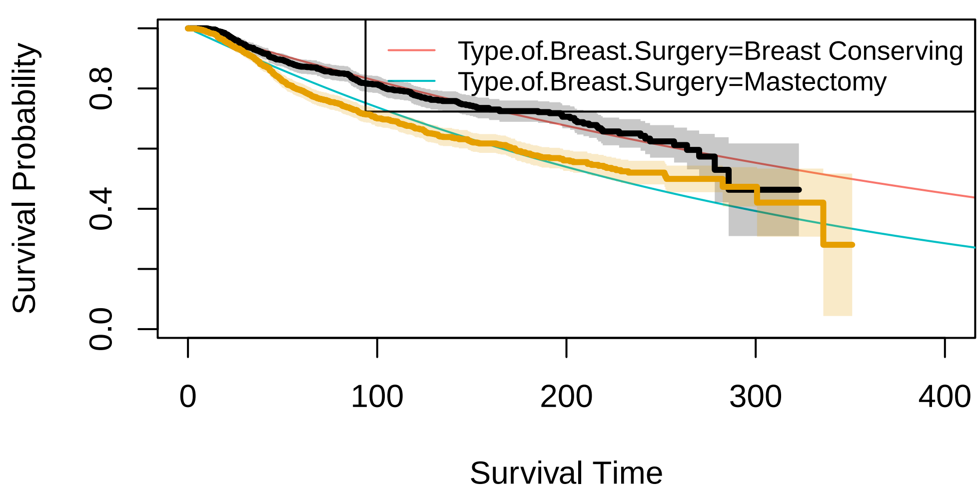

Figura 4: Curvas de sobrevivencia estimadas (weibull).

Se ajustaron los modelos pero tomando como variable explicatoria el tipo de cirugía (mastectomía total o parcial). Para el caso del modelo exponencial (véase tabla 3) el tiempo de supervivencia medio tomando en cuenta el tipo de cirugia de mastectomía parcial es de 513.2846278 meses que es mayor al de la mastectomía total de 325.6945089 meses, representando un 36.54% menor de tiempo de supervivencia esperado cuando se tiene una cirugía de mastectomía total. En la figura 3 se presentan las curvas de supervivencia para ambos tipos de cirugía estimadas y su estimador Kaplan-Meier.

Las conclusiones usando el modelo weibull son prácticamente las mismas siendo los tiempos de sobrevivencia medio de 500.7947081 y 320.3982181 para la mastectomía parcial y total respectivamente. El tiempo de sobrevivencia reducido para la cirugía parcial es de 36.02%. Además cabe destacar que el parámetro de escala sigue siendo no significativamente distinto de 1, por lo cual se puede concluir que el modelo weibull se reduce a un modelo exponencial. Además en ambos modelos el regresor es significativo, por lo que se concluye que el tipo de cirugía afecta la sobrevivencia de los pacientes.

3 Comapración de Modelos

Para la elección de modelos se puede utilizar la prueba de razón de verosimilitud (LRT, por sus siglas en inglés), en la cual se ajusta el modelo bajo una restricción en su espacio paramétrico y el modelo sin la restricción.

3.1 Prueba de razón de verosimilitud (LTR)

Dado un es modelo estadístico con un espacio parámetrico Θ. Una hipótesis nula sobre el espacio parámetrico se define como $\Theta_0 \subset \Theta$ igualmente se puede definir una hipótesis alternativa como el complmento del conjunto de la hipótesis nula $\Theta - \Theta_0$. Entonces, la prubea de razón de verosimilitud para la hipótesis nula $H_0 = \theta \in \Theta_0$ se define como:

$$\Lambda_{LR} = -2 \ln \frac{\sup_{\theta \in \Theta_0} \mathcal{L}(x_1, \ldots, x_n|\theta)}{\sup_{\theta \in \Theta} \mathcal{L}(x_1, \ldots, x_n|\theta)}$$

Usando propiedades del logaritmo, podemos expresar el cociente como una resta.

$$\Lambda_{LR} = -2 \left[\ell\left(\theta_0\right) - \ell(\hat{\theta}) \right]$$

El estadístico obtenido sigue asíntoticamente una distribución $\chi^2_{p_1 - p_0}$ donde p_1 es el número de parámetros en el modelo sin restricciones y p_0 el número de parámetros en el modelo restringido.

En nuestro caso para probar si una distribución Weibull es mejor que la exponencial, podemos usar el hecho de que la exponencial es un caso especial de la Weibull cuando su parámetro de forma (α) es 1. La función de densidad para esta distribución es:

$$pdf(t; \alpha, \lambda) = \lambda \alpha (\lambda t)^{\alpha - 1} e^{-(\lambda t)^{\alpha}}$$

Usando la función de log-verosimilitud de la distribución Weibull y como hipótesis nula $H_0 : \alpha = 1$ e hipótesis alternativa $H_1 : \alpha \neq 1$, podemos escribir la prueba LR como:

$$\Lambda_{LR} = -2 \left[(-n \ln \hat{\lambda} + \sum \hat{\lambda} t_i) - (n \ln \hat{\alpha} - n \ln \hat{\lambda} + (\hat{\alpha} - 1) \sum \ln t_i + \sum \hat{\lambda} t_i) \right]$$

Y la distribución asintótica de nuestro estadístico es $\Lambda_{LR} \sim \chi^2_1$.