

3.2 El modelo Weibull

3.2.1 Muestras completas

Si se tiene una muestra aleatoria X_1, \dots, X_n sin censura de una población Weibull, $X_i \sim \text{WEI}(\theta, \beta)$, la función de verosimilitud se puede expresar como:

$$L(\theta, \beta) = \prod_{i=1}^n \frac{\beta}{\theta^\beta} x_i^{\beta-1} \exp \left\{ - \left(\frac{x_i}{\theta} \right)^\beta \right\},$$

y la función log-verosimilitud es:

$$l(\theta, \beta) = n \log \beta - \beta n \log \theta + (\beta - 1) \sum_{i=1}^n \log x_i - \sum_{i=1}^n \left(\frac{x_i}{\theta} \right)^\beta.$$

En este caso, los EMV de θ y β se obtienen al resolver el siguiente sistema de ecuaciones:

$$\begin{cases} \frac{\partial l(\theta, \beta)}{\partial \theta} = 0 \\ \frac{\partial l(\theta, \beta)}{\partial \beta} = 0 \end{cases},$$

donde

$$\frac{\partial l(\theta, \beta)}{\partial \theta} = -\frac{n\beta}{\theta} + \frac{\beta}{\theta^{\beta+1}} \sum_{i=1}^n x_i^\beta$$

y

$$\frac{\partial l(\theta, \beta)}{\partial \beta} = \frac{n}{\beta} - n \log \theta + \sum_{i=1}^n \log x_i - \sum_{i=1}^n \left(\frac{x_i}{\theta} \right)^\beta \log \left(\frac{x_i}{\theta} \right),$$

Este sistema de ecuaciones se reduce a

$$\frac{\sum_{i=1}^n x_i^{\hat{\beta}} \log x_i}{\sum_{i=1}^n x_i^{\hat{\beta}}} - \frac{1}{\hat{\beta}} = \frac{\sum_{i=1}^n \log x_i}{n}$$

y

$$\hat{\theta} = \left(\frac{\sum_{i=1}^n x_i^{\hat{\beta}}}{n} \right)^{1/\hat{\beta}},$$

donde $\hat{\theta}$ y $\hat{\beta}$ son los EMV.

El sistema de ecuaciones no se puede resolver en forma cerrada. Como en la solución es única, es posible usar un método numérico para aproximar a $\hat{\theta}$ y $\hat{\beta}$. El procedimiento Newton-Raphson, por ejemplo, puede usarse para resolver una ecuación $g(\hat{\beta}) = 0$ usando aproximaciones sucesivas $\hat{\beta}_j$, donde $\hat{\beta}_{j+1} = \hat{\beta}_j - g(\hat{\beta}_j)/g'(\hat{\beta}_j)$. Muchas otras técnicas pueden ser implementadas usando una computadora.

Para obtener la matriz información es necesario encontrar el Hessiano de $l(\theta, \beta)$, para lo cual

$$\frac{\partial^2 l(\theta, \beta)}{\partial \theta^2} = \frac{n\beta}{\theta^2} - \frac{\beta(\beta+1)}{\theta^{\beta+2}} \sum_{i=1}^n X_i^\beta,$$

$$\frac{\partial^2 l(\theta, \beta)}{\partial \beta^2} = -\frac{n}{\beta^2} - \sum_{i=1}^n \left(\frac{X_i}{\theta}\right)^\beta \log\left(\frac{X_i}{\theta}\right)^2,$$

y

$$\frac{\partial^2 l(\theta, \beta)}{\partial \theta \beta} = -\frac{n}{\theta} + \frac{1}{\theta} \sum_{i=1}^n \left(\frac{X_i}{\theta}\right)^\beta + \frac{\beta}{\theta} \sum_{i=1}^n \left(\frac{X_i}{\theta}\right)^\beta \log\left(\frac{X_i}{\theta}\right).$$

La matriz información para θ y β es entonces

$$\mathbf{I}(\theta, \beta) = \mathbf{E} \begin{bmatrix} -\frac{\partial^2 l(\theta, \beta)}{\partial \theta^2} & -\frac{\partial^2 l(\theta, \beta)}{\partial \theta \beta} \\ -\frac{\partial^2 l(\theta, \beta)}{\partial \theta \beta} & -\frac{\partial^2 l(\theta, \beta)}{\partial \beta^2} \end{bmatrix}$$

Como esta matriz está en términos de los parámetros desconocidos θ y β , es posible estimarla usando los EMV con

$$\tilde{\mathbf{I}}(\hat{\theta}, \hat{\beta}) = \begin{bmatrix} -\frac{\partial^2 l}{\partial \theta^2} \Big|_{\theta=\hat{\theta}} & -\frac{\partial^2 l}{\partial \theta \beta} \Big|_{\theta=\hat{\theta}, \beta=\hat{\beta}} \\ -\frac{\partial^2 l}{\partial \theta \beta} \Big|_{\theta=\hat{\theta}, \beta=\hat{\beta}} & -\frac{\partial^2 l}{\partial \beta^2} \Big|_{\beta=\hat{\beta}} \end{bmatrix}.$$

De esta forma, la inversa de esta matriz es un estimador de la matriz de covarianzas

de $(\hat{\theta}, \hat{\beta})^T$:

$$\tilde{\mathbf{I}}^{-1}(\hat{\theta}, \hat{\beta}) = \begin{bmatrix} \widetilde{\text{Var}}[\hat{\theta}] & \widetilde{\text{Cov}}[\hat{\theta}, \hat{\beta}] \\ \widetilde{\text{Cov}}[\hat{\theta}, \hat{\beta}] & \widetilde{\text{Var}}[\hat{\beta}] \end{bmatrix}.$$

Usando los resultados asintóticos de máxima verosimilitud se tiene la siguiente aproximación

$$(\hat{\theta}, \hat{\beta})^T \sim \text{NBV} \left((\theta, \beta)^T, \tilde{\mathbf{I}}^{-1}(\hat{\theta}, \hat{\beta}) \right),$$

donde NBV denota la distribución normal bivariada. por lo que

$$\hat{\theta} \sim \text{N}(\theta, \widetilde{\text{Var}}[\hat{\theta}]) \quad \text{aproximadamente,}$$

y

$$\hat{\beta} \sim \text{N}(\beta, \widetilde{\text{Var}}[\hat{\beta}]) \quad \text{aproximadamente.}$$

3.2.2 Muestras con censura

Supóngase que se tiene una muestra de tamaño n tomada de una distribución Weibull con parámetro de escala $\lambda^{1/\gamma}$ y parámetro de forma γ , $\text{WEI}(1/[\lambda^{1/\gamma}], \gamma)$. Supóngase también que hay r muertes de los n individuos y $n - r$ tiempos de supervivencia con censura. La función de densidad y la función de supervivencia están dadas por

$$f(t) = \lambda \gamma t^{\gamma-1} \exp(-\lambda t^\gamma), \quad S(t) = \exp(-\lambda t^\gamma).$$

Esta es una reparametrización de la Weibull; sin embargo, los resultados en esta sección son equivalentes.

La función de verosimilitud se puede expresar como:

$$\begin{aligned} L(\lambda, \gamma) &= \prod_{i=1}^n \{f(t_i; \lambda, \gamma)\}^{\delta_i} \{S(t_i; \lambda, \gamma)\}^{1-\delta_i} \\ &= \prod_{i=1}^n \{\lambda \gamma t_i^{\gamma-1} \exp(-\lambda t_i^\gamma)\}^{\delta_i} \{\exp(-\lambda t_i^\gamma)\}^{1-\delta_i}. \end{aligned}$$

La función log-verosimilitud correspondiente está dada por

$$l(\lambda, \gamma) = \log L(\lambda, \gamma) = \sum_{i=1}^n \delta_i \log(\lambda\gamma) + (\gamma - 1) \sum_{i=1}^n \delta_i \log t_i - \lambda \sum_{i=1}^n t_i^\gamma,$$

y como $\sum_{i=1}^n \delta_i = r$, la función verosimilitud es

$$l(\lambda, \gamma) = r \log(\lambda\gamma) + (\gamma - 1) \sum_{i=1}^n \delta_i \log t_i - \lambda \sum_{i=1}^n t_i^\gamma.$$

Los estimadores de máxima verosimilitud de λ y γ se encuentran al diferenciar $l(\lambda, \gamma)$ con respecto a λ y γ a la vez, al igualar a cero las dos derivadas, y al encontrar la solución del sistema de ecuaciones $\hat{\lambda}$ y $\hat{\gamma}$. Las ecuaciones del sistema resultante son

$$\frac{r}{\hat{\lambda}} - \sum_{i=1}^n t_i^{\hat{\gamma}} = 0,$$

y

$$\frac{r}{\hat{\gamma}} + \sum_{i=1}^n \delta_i \log t_i - \hat{\lambda} \sum_{i=1}^n t_i^{\hat{\gamma}} \log t_i = 0.$$

Por lo tanto

$$\hat{\lambda} = r / \sum_{i=1}^n t_i^{\hat{\gamma}},$$

y al sustituir $\hat{\lambda}$ en la primera ecuación del sistema se tiene que

$$\frac{r}{\hat{\gamma}} + \sum_{i=1}^n \delta_i \log t_i - \frac{r}{\sum_{i=1}^n t_i^{\hat{\gamma}}} \sum_{i=1}^n t_i^{\hat{\gamma}} \log t_i = 0.$$

Esta es una ecuación no lineal la cual solo puede resolverse usando un procedimiento iterativo tal como Newton-Raphson.

Una vez que se encuentran los estimadores de λ y γ , se pueden estimar percentiles del tiempo de supervivencia. El p -ésimo percentil de la distribución es

$$\hat{t}(p) = \left[\frac{1}{\hat{\lambda}} \log \left(\frac{100}{100 - p} \right) \right]^{1/\hat{\gamma}},$$

y entonces el estimador de la mediana del tiempo de supervivencia está dado por

$$\hat{t}(50) = \left[\frac{1}{\hat{\lambda}} \log 2 \right]^{1/\hat{\gamma}}.$$

El error estándar del p -ésimo percentil estimado se puede obtener mas fácilmente al encontrar primero la varianza de $\log \hat{t}(p)$ y al usar el siguiente resultado:

$$\text{Var}[g(\hat{\theta}_1, \hat{\theta}_2)] \approx \left(\frac{\partial g}{\partial \hat{\theta}_1} \right)^2 \text{Var}[\hat{\theta}_1] + \left(\frac{\partial g}{\partial \hat{\theta}_2} \right)^2 \text{Var}[\hat{\theta}_2] + 2 \left(\frac{\partial g}{\partial \hat{\theta}_1} \frac{\partial g}{\partial \hat{\theta}_2} \right) \text{Cov}[\hat{\theta}_1, \hat{\theta}_2]. \quad (3.3)$$

Aquí se tiene que,

$$\log \hat{t}(p) = \frac{1}{\hat{\gamma}} \log \left\{ \hat{\gamma}^{-1} \log \left(\frac{100}{100-p} \right) \right\},$$

y entonces

$$\log \hat{t}(p) = \frac{1}{\hat{\gamma}} \left\{ c_p - \log \hat{\lambda} \right\},$$

donde

$$c_p = \log \log \left(\frac{100}{100-p} \right).$$

Usando el resultado de la Ecuación (3.3), se tiene que

$$\begin{aligned} \text{Var}[\log \hat{t}(p)] &\approx \left(\frac{\partial \log \hat{t}(p)}{\partial \hat{\lambda}} \right)^2 \text{Var}[\hat{\lambda}] + \left(\frac{\partial \log \hat{t}(p)}{\partial \hat{\gamma}} \right)^2 \text{Var}[\hat{\gamma}] \\ &\quad + 2 \frac{\partial \log \hat{t}(p)}{\partial \hat{\lambda}} \frac{\partial \log \hat{t}(p)}{\partial \hat{\gamma}} \text{Cov}[\hat{\lambda}, \hat{\gamma}]. \end{aligned}$$

La derivadas de $\log \hat{t}(p)$ con respecto a λ y γ son

$$\begin{aligned} \frac{\partial \log \hat{t}(p)}{\partial \hat{\lambda}} &= -\frac{1}{\hat{\lambda} \hat{\gamma}}, \\ \frac{\partial \log \hat{t}(p)}{\partial \hat{\gamma}} &= -\frac{c_p - \log \hat{\lambda}}{\hat{\gamma}^2}, \end{aligned}$$

y entonces

$$\text{Var}[\log \hat{t}(p)] \approx \frac{1}{\hat{\lambda}^2 \hat{\gamma}^2} \text{Var}[\hat{\lambda}] + \frac{(c_p - \log \hat{\lambda})^2}{\hat{\gamma}^4} \text{Var}[\hat{\gamma}] + \frac{2(c_p - \log \hat{\lambda})}{\hat{\lambda} \hat{\gamma}^3} \text{Cov}[\hat{\lambda}, \hat{\gamma}].$$

La varianza de $\hat{t}(p)$ se encuentra con la aproximación

$$\text{Var}[g(X)] \approx \left[\frac{dg(X)}{dX} \right]^2 \text{Var}[X],$$

por lo que

$$\text{Var}[\hat{t}(p)] \approx [\hat{t}(p)]^2 \text{Var}[\log \hat{t}(p)].$$

y entonces

$$\begin{aligned} \text{s.e.}[\hat{t}(p)] &= \frac{\hat{t}(p)}{\hat{\lambda}\hat{\gamma}^2} \left\{ \hat{\gamma}^2 \text{Var}[\hat{\lambda}] + \hat{\lambda}^2 (c_p - \log \hat{\lambda})^2 \text{Var}[\hat{\gamma}] \right. \\ &\quad \left. + 2\hat{\lambda}\hat{\gamma} (c_p - \log \hat{\lambda}) \text{Cov}[\hat{\lambda}, \hat{\gamma}] \right\}^{1/2}. \end{aligned}$$

Un intervalo de confianza de $100(1 - \alpha)\%$ para el p -ésimo percentil se puede encontrar al manipular con operaciones válidas los siguientes límites de confianza para $\log t(p)$

$$\log \hat{t}(p) \pm z_{1-\alpha/2} \text{s.e.}[\log \hat{t}(p)],$$

por lo cual los límites son

$$\hat{t}(p) \exp\{\pm z_{1-\alpha/2} \text{s.e.}[\log \hat{t}(p)]\},$$

Ejemplo 3.3. En el lenguaje de programación R, la función `survreg` ajusta al modelo Weibull $\text{WEI}(\exp \beta_0, 1/\alpha)$, de manera tal que la función de supervivencia es

$$S(t) = \exp \left\{ - \left(\frac{t}{\exp \beta_0} \right)^{1/\alpha} \right\}.$$

A continuación se ilustra el ajuste de los datos DIU presentado en la Tabla 2.3

```
diu <- data.frame(tiempo=c(10, 13, 18, 19, 23, 30, 36, 38, 54,
                          56, 59, 75, 93, 97, 104, 107, 107, 107),
                 estatus=c(1, 0, 0, 1, 0, 1, 1, 0, 0,
                          0, 1, 1, 1, 1, 0, 1, 0, 0))
```

```
library(survival)
```

```
# Encuentra el Modelo Weibull para los datos en diu
diu.wei <- survreg(Surv(tiempo, estatus)~1, data=diu,
                  dist='weibull')
```

En este caso, el signo `~1` en la fórmula indica a R que la población es homogénea. Algunos de los atributos del objeto `diu.wei` se pueden ver a continuación:

```
> #los EMV de BETA0 y log(alfa) son:
> diu.wei$icoef
(Intercept)  Log(scale)
  4.5915178  -0.5166504
> #la matriz de covarianzas correspondiente es:
> diu.wei$var
              (Intercept) Log(scale)
(Intercept)  0.04176813  0.01297235
Log(scale)   0.01297235  0.07541491
```

De esta forma, $\hat{\beta}_0 = 4.5915178$, $\text{s.e.}[\hat{\beta}_0] = \sqrt{0.04176813} = 0.2043725$, $\log \hat{\alpha} = -0.5166504$ y $\text{s.e.}[\log \hat{\alpha}] = \sqrt{0.07541491} = 0.2746178$. Un intervalo de 95% de confianza para $\log \alpha$ tiene límites $\log \hat{\alpha} \pm 1.96\text{s.e.}[\log \hat{\alpha}]$, por lo que el intervalo es $(-1.054901, 0.02160041)$. Como el intervalo incluye a cero, es pausable que el modelo exponencial se ajuste tan bien como el Weibull.