

Una introducción al análisis de sobrevivencia usando el software R (Parte 2)

Diego I. Gallardo Mateluna

Departamento de Matemática, Facultad de Ingeniería
Universidad de Atacama
Copiapó, Chile

Minicurso para la Sociedad Chilena de Estadística (SOCHE)

diego.gallardo@uda.cl

04 de Diciembre de 2020



Resumen

- Elementos de interés en el análisis de sobrevivencia.
- Censura.
- Estimadores no paramétricos de la función de sobrevivencia.
- Modelos paramétricos.
- Modelos de regresión en análisis de sobrevivencia.
- Inferencia.
- Residuos.
- Otros tópicos.



Modelos de regresión

Modelos de regresión en análisis de sobrevivencia



Modelos de regresión

(Kleinbaum and Klein, 2002)



Perro



Hombre

Una creencia popular indica que **un año canino** equivale a **7 años de humano**. En otras palabras, esto lo podemos expresar como

$$S_P(t) = S_H(7t).$$



Modelos de regresión

Modelos de tiempo de vida acelerado



Modelos de regresión

Modelos de tiempo de vida acelerado (TVA) (en inglés, AFT, accelerated failure time model).

Asumamos que estamos interesados en el **tiempo de vida** de personas de cierto país clasificados en un **grupo de control** y un grupo de **tratamiento** (digamos grupos 1 y 2).





Modelos de regresión

Modelos de tiempo de vida acelerado (TVA) (en inglés, AFT, accelerated failure time model).

Asumamos que estamos interesados en el **tiempo de vida** de personas de cierto país clasificados en un **grupo de control** y un grupo de **tratamiento** (digamos grupos 1 y 2).

Supongamos además que las **funciones de sobrevida** de ambos grupos son dadas por $S_1(\cdot)$ y $S_2(\cdot)$, respectivamente.





Modelos de regresión

Modelos de tiempo de vida acelerado (TVA) (en inglés, AFT, accelerated failure time model).

Asumamos que estamos interesados en el **tiempo de vida** de personas de cierto país clasificados en un **grupo de control** y un grupo de **tratamiento** (digamos grupos 1 y 2).

Supongamos además que las **funciones de sobrevivencia** de ambos grupos son dadas por $S_1(\cdot)$ y $S_2(\cdot)$, respectivamente.

Bajo el supuesto de **modelos de TVA**, tendríamos que

$$S_1(t) = S_2(\theta t), \quad t \geq 0.$$

En que θ es el **factor de aceleración**.



Modelos de regresión

Interpretación para θ .

- Si $\theta = 2$, entonces el tiempo para alcanzar cualquier percentil de tiempo en el grupo 2 es el **doble** que en el grupo 1.



Modelos de regresión

Interpretación para θ .

- Si $\theta = 2$, entonces el tiempo para alcanzar cualquier percentil de tiempo en el grupo 2 es el **doble** que en el grupo 1.
- Si $\theta = 0.5$, entonces el tiempo para alcanzar cualquier percentil de tiempo en el grupo 2 es la **mitad** que en el grupo 1.



Modelos de regresión

Interpretación para θ .

- Si $\theta = 2$, entonces el tiempo para alcanzar cualquier percentil de tiempo en el grupo 2 es el **doble** que en el grupo 1.
- Si $\theta = 0.5$, entonces el tiempo para alcanzar cualquier percentil de tiempo en el grupo 2 es **la mitad** que en el grupo 1.
- Si $\theta = 1$, entonces el tiempo para alcanzar cualquier percentil de tiempo en el grupo 2 es **el mismo** que en el grupo 1.



Modelos de regresión

Una forma general de considerar **modelos de TVA** incluyendo **covariables** es la familia de localización-escala para la variable $Y = \log T$.



Modelos de regresión

Una forma general de considerar **modelos de TVA** incluyendo **covariables** es la familia de localización-escala para la variable $Y = \log T$.

Suponga que disponemos de un conjunto de **covariables** $\mathbf{x} = (1, x_1, x_2, \dots, x_p)$, entonces

$$S(t | \mathbf{x}, \psi) = S_0 \left(\frac{y - \mu(\mathbf{x})}{\sigma} \right),$$

en que $y = \log(t)$, $\mu(\mathbf{x}) = \mathbf{x}^\top \boldsymbol{\beta} = \beta_0 + x_1 \beta_1 + x_2 \beta_2 + \dots + x_p \beta_p$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)^\top$ y $\psi = (\boldsymbol{\beta}^\top, \sigma)$.



Modelos de regresión

Una forma general de considerar **modelos de TVA** incluyendo **covariables** es la familia de localización-escala para la variable $Y = \log T$.

Suponga que disponemos de un conjunto de **covariables** $\mathbf{x} = (1, x_1, x_2, \dots, x_p)$, entonces

$$S(t | \mathbf{x}, \psi) = S_0\left(\frac{y - \mu(\mathbf{x})}{\sigma}\right),$$

en que $y = \log(t)$, $\mu(\mathbf{x}) = \mathbf{x}^\top \boldsymbol{\beta} = \beta_0 + x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \dots, \beta_p)^\top$ y $\psi = (\boldsymbol{\beta}^\top, \sigma)$.

Casos particulares:

- **VE**: $S_0(y) = \exp(-e^y)$. (Asociado al modelo **WEI**)
- **N**: $S_0(y) = \Phi(e^y)$. (Asociado al modelo **LN**)
- **L**: $S_0(y) = (1 + e^y)^{-1}$. (Asociado al modelo **LL**)



Modelos de regresión

Recordemos

Dist. Y	Dist. T	μ	σ
WEI (α, σ)	VE (μ, σ)	$\log \alpha$	$1/\gamma$
LN (μ, σ)	N (μ, σ^2)	-	-
LL (α, σ)	L (μ, σ)	$\log \alpha$	$1/\gamma$

De esta forma, para os modelos WEI e LL,

$$\mu(\mathbf{x}) = \beta_0 + x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p$$

es equivalente a

$$\alpha(\mathbf{x}) = \exp(\beta_0 + x_1\beta_1 + x_2\beta_2 + \dots + x_p\beta_p).$$



Modelos de regresión

Note que

$$\begin{aligned} S(t \mid \mathbf{x}, \psi) &= S_0\left(\frac{y - \mu(\mathbf{x})}{\sigma}\right) \\ &= S_0\left(\frac{\log t - \mu(\mathbf{x})}{\sigma}\right), \\ &= S_0\left(\frac{1}{\sigma} \log (\theta(\mathbf{x})t)\right), \end{aligned}$$

en que $\theta(\mathbf{x}) = \exp\{\mu(\mathbf{x})\}$.



Modelos de regresión

Note que

$$\begin{aligned}
 S(t \mid \mathbf{x}, \psi) &= S_0\left(\frac{y - \mu(\mathbf{x})}{\sigma}\right) \\
 &= S_0\left(\frac{\log t - \mu(\mathbf{x})}{\sigma}\right), \\
 &= S_0\left(\frac{1}{\sigma} \log (\theta(\mathbf{x})t)\right),
 \end{aligned}$$

en que $\theta(\mathbf{x}) = \exp\{\mu(\mathbf{x})\}$.

NOTA: para que el modelo sea de TVA, σ debe ser **constante** (es decir, no debe ser modelado).



Modelos de regresión

En relación a la **interpretación** de los coeficientes de regresión, supongamos que la j -ésima ($j = 1, \dots, p$) covariable tiene coeficiente β_j asociado:

- Si x_j es **binaria**, manteniendo todas las demás covariables **fijas**, el **factor de aceleración** para comparar el grupo asociado a $x_j = 1$ versus el grupo asociado a $x_j = 0$ es $\exp(\beta_j)$.



Modelos de regresión

En relación a la **interpretación** de los coeficientes de regresión, supongamos que la j -ésima ($j = 1, \dots, p$) covariable tiene coeficiente β_j asociado:

- Si x_j es **binaria**, manteniendo todas las demás covariables **fijas**, el **factor de aceleración** para comparar el grupo asociado a $x_j = 1$ versus el grupo asociado a $x_j = 0$ es $\exp(\beta_j)$.
- Si x_j es **continua**, manteniendo todas las demás covariables **fijas**, el **factor de aceleración** es $\exp(\beta_j)$ para cada unidad en que x_j es aumentada (vale también para x_j discreta).



Modelos de regresión

Ejemplo: Suponga que T representa el tiempo hasta que una cinta transportadora de minerales **falle** por un defecto mecánico. Suponga que disponemos de dos covariables.

- x_1 : **marca de la cinta** (0: CPS; 1: Cintaplus).
- x_2 : **peso de la carga** (en toneladas).



Modelos de regresión

Ejemplo: Suponga que T representa el tiempo hasta que una cinta transportadora de minerales **falle** por un defecto mecánico. Suponga que disponemos de dos covariables.

- x_1 : **marca de la cinta** (0: CPS; 1: Cintaplus).
- x_2 : **peso de la carga** (en toneladas).

Asuma que se usó el modelo VE para $Y = \log T$ y que

$$\mu(x_i) = 1.25 + 0.85x_{1i} - 0.15x_{2i},$$

es decir, $\beta_0 = 1.25$, $\beta_1 = 0.85$ y $\beta_2 = -0.15$.



Modelos de regresión

Interpretación de los coeficientes (basado en la mediana, pero es válido para cualquier percentil):



Modelos de regresión

Interpretación de los coeficientes (basado en la mediana, pero es válido para cualquier percentil):

- $\exp(\beta_1) = \exp(0.85) \approx 2.34$. Esto quiere decir que, fijado un peso de la carga, la mediana del tiempo de falla de las cintas transportadoras de minerales marca Cintaplus es 2.34 veces la mediana del mismo tiempo para la marca CPS.



Modelos de regresión

Interpretación de los coeficientes (basado en la mediana, pero es válido para cualquier percentil):

- $\exp(\beta_1) = \exp(0.85) \approx 2.34$. Esto quiere decir que, fijado un peso de la carga, la mediana del tiempo de falla de las cintas transportadoras de minerales marca Cintaplus es 2.34 veces la mediana del mismo tiempo para la marca CPS.
- $\exp(\beta_2) = \exp(-0.15) \approx 0.86$. Esto quiere decir que, fijada una marca de la cinta, por cada tonelada que aumenta el peso de la carga la mediana del tiempo de falla disminuirá un 14%.





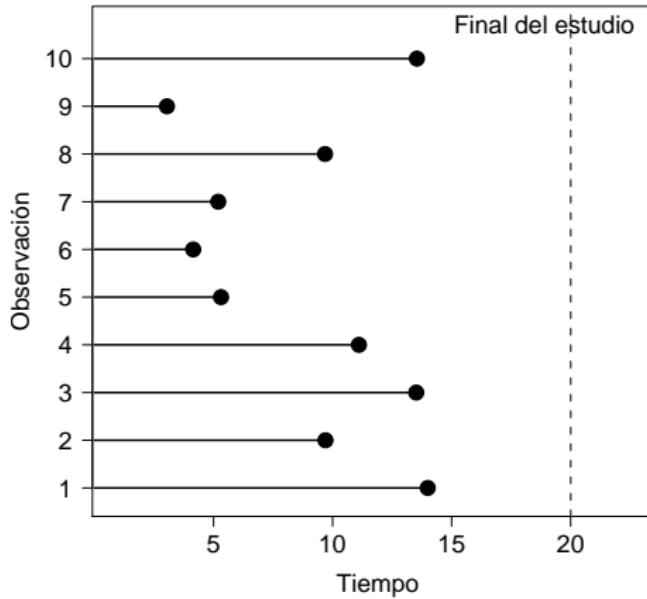
Inferencia

Inferencia



Tipos de censura

Caso a: **datos completos.**



Inferencia

Datos completos.

Supuestos:

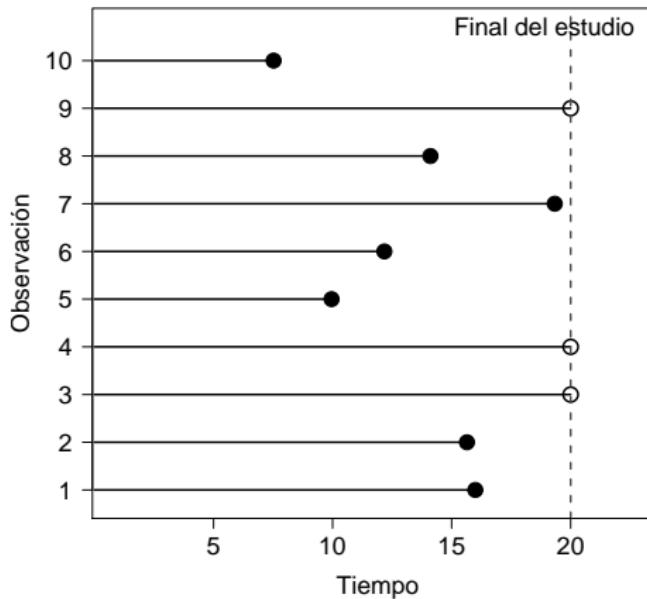
- **Independencia** de las observaciones.

$$\mathcal{L}(\theta) = \prod_{i=1}^n f(t_i; \theta).$$



Tipos de censura

Caso b: Censura tipo I.



Inferencia

Censura tipo I (El estudio será terminado después de un período establecido de tiempo).

Supuestos (Williams and Lagakos, 1977):

- **Independencia** de las observaciones.



Inferencia

Censura tipo I (El estudio será terminado después de un período establecido de tiempo).

Supuestos (Williams and Lagakos, 1977):

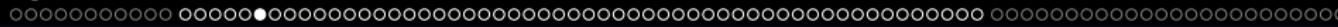
- **Independencia** de las observaciones.

En la censura tipo I, $C_1 = \dots = C_n = C$ son **determinísticos**. Es simple probar que

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &= \prod_{i=1}^n f(t_i; \boldsymbol{\theta})^{\delta_i} P(T_i > t_i; \boldsymbol{\theta})^{1-\delta_i} \\ &= \prod_{i:\delta_i=1} f(t_i; \boldsymbol{\theta}) \times [P(T_i > C; \boldsymbol{\theta})]^{n-r}. \\ &= \prod_{i:\delta_i=1} f(t_i; \boldsymbol{\theta}) \times [S(C; \boldsymbol{\theta})]^{n-r},\end{aligned}$$

en que $r = \sum_{i=1}^n \delta_i$ es la cantidad de **fallas** en la muestra.





Inferencia

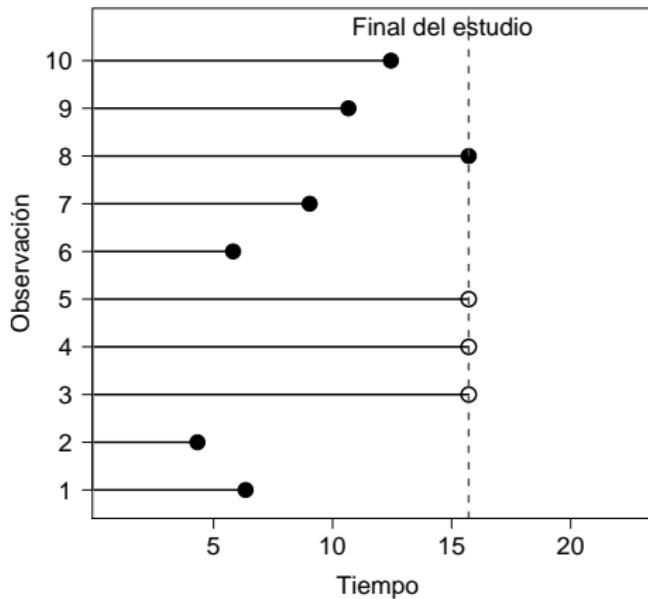
Observación: note que también es posible escribir esta verosimilitud como

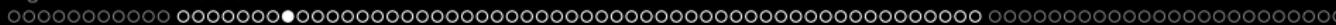
$$\mathcal{L}(\theta) = \prod_{i=1}^n f(t_i; \theta)^{\delta_i} S(t_i; \theta)^{1-\delta_i}.$$



Tipos de censura

Caso c: Censura tipo II.





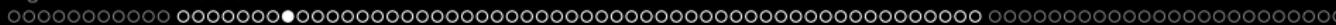
Inferencia

Censura tipo II (El estudio será terminado cuando ocurran r fallas).

Supuestos (Williams and Lagakos, 1977):

- **Independencia** de las observaciones.
- T_i y C_i son **independientes**.





Inferencia

Censura tipo II (El estudio será terminado cuando ocurran r fallas).

Supuestos (Williams and Lagakos, 1977):

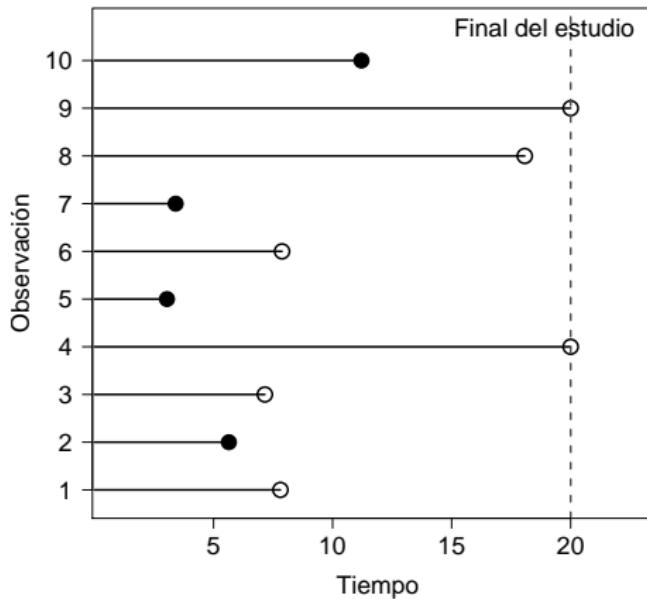
- **Independencia** de las observaciones.
- T_i y C_i son **independientes**.

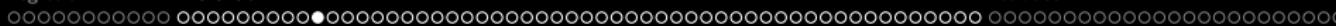
$$\mathcal{L}(\theta) = \frac{n!}{(n-r)!} \prod_{i=1}^r f(t_i; \theta)^{\delta_i} S(t_i; \theta)^{1-\delta_i}.$$



Tipos de censura

Caso d: Censura aleatoria.



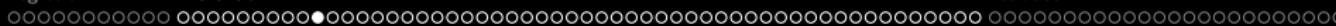


Inferencia

Censura aleatoria. Supuestos (Williams and Lagakos, 1977):

- **Independencia** de las observaciones.
- T_i y C_i son **independientes**.
- Las funciones de supervivencia de T_i y C_i son $S(\cdot; \theta)$ y $G(\cdot; \xi)$, respectivamente. (siendo $f(\cdot; \theta)$ y $g(\cdot; \xi)$ las funciones de densidad respectivas). $S(\cdot; \psi)$ y $G(\cdot; \xi)$ no comparten parámetros en común. Este supuesto es conocido en la literatura como **censura no informativa**.





Inferencia

Censura aleatoria. Supuestos (Williams and Lagakos, 1977):

- **Independencia** de las observaciones.
- T_i y C_i son **independientes**.
- Las funciones de supervivencia de T_i y C_i son $S(\cdot; \theta)$ y $G(\cdot; \xi)$, respectivamente. (siendo $f(\cdot; \theta)$ y $g(\cdot; \xi)$ las funciones de densidad respectivas). $S(\cdot; \psi)$ y $G(\cdot; \xi)$ no comparten parámetros en común. Este supuesto es conocido en la literatura como **censura no informativa**.

$$\mathcal{L}(\theta) = \prod_{i=1}^n g(t_i; \xi)^{\delta_i} G(t_i; \xi)^{1-\delta_i} \times \prod_{i=1}^n f(t_i; \theta)^{\delta_i} S(t_i; \theta)^{1-\delta_i}.$$



Inferencia

Conclusión: sobre censura a la derecha y censura tipo I, II o aleatoria, la función de verosimilitud es proporcional a

$$\mathcal{L}(\boldsymbol{\theta}) \propto \prod_{i=1}^n f(t_i; \boldsymbol{\theta})^{\delta_i} S(t_i; \boldsymbol{\theta})^{1-\delta_i}.$$



Inferencia

Conclusión: sobre censura a la derecha y censura tipo I, II o aleatoria, la función de verosimilitud es proporcional a

$$\mathcal{L}(\boldsymbol{\theta}) \propto \prod_{i=1}^n f(t_i; \boldsymbol{\theta})^{\delta_i} S(t_i; \boldsymbol{\theta})^{1-\delta_i}.$$

Equivalentemente,

$$\begin{aligned}\mathcal{L}(\boldsymbol{\theta}) &\propto \prod_{i=1}^n \left(\frac{f(t_i; \boldsymbol{\theta})}{S(t_i; \boldsymbol{\theta})} \right)^{\delta_i} S(t_i; \boldsymbol{\theta}) \\ &= \prod_{i=1}^n \lambda(t_i; \boldsymbol{\theta})^{\delta_i} S(t_i; \boldsymbol{\theta}).\end{aligned}$$





Inferencia

Inferencia estadística para modelos paramétricos puede ser realizada basada en las metodologías estándar de máxima verosimilitud.





Inferencia

Inferencia estadística para modelos paramétricos puede ser realizada basada en las metodologías estándar de máxima verosimilitud.

Sea

$$I(\psi) = -\frac{\partial^2 \log \mathcal{L}(\psi)}{\partial \psi \partial \psi^\top},$$

la matriz hessiana del modelo y

$$\mathcal{I}(\psi) = \mathbb{E}(I(\psi)),$$

la respectiva matriz de información de Fisher.





Inferencia

Bajo las condiciones de regularidad, el estimador de máxima verosimilitud de ψ , digamos $\hat{\psi}$, se tiene que

$$\sqrt{n}(\hat{\psi} - \psi) \sim N(\mathbf{0}, \mathcal{I}(\psi)), \quad n \rightarrow +\infty.$$



Inferencia

Bajo las condiciones de regularidad, el estimador de máxima verosimilitud de ψ , digamos $\hat{\psi}$, se tiene que

$$\sqrt{n}(\hat{\psi} - \psi) \sim N(\mathbf{0}, \mathcal{I}(\psi)), \quad n \rightarrow +\infty.$$

Otros procedimientos estándar también pueden ser realizados, tales como:

- Test de hipótesis (razón de verosimilitudes, score, Wald, gradiente).
- Intervalos de confianza.





Inferencia

Bajo las condiciones de regularidad, el estimador de máxima verosimilitud de ψ , digamos $\hat{\psi}$, se tiene que

$$\sqrt{n}(\hat{\psi} - \psi) \sim N(\mathbf{0}, \mathcal{I}(\psi)), \quad n \rightarrow +\infty.$$

Otros procedimientos estándar también pueden ser realizados, tales como:

- Test de hipótesis (razón de verosimilitudes, score, Wald, gradiente).
- Intervalos de confianza.

En muchos casos $\mathcal{I}(\psi)$ es complicada de calcular (o simplemente no se puede calcular), por lo que es usual utilizar $I(\psi)$ en su lugar.





Inferencia

Ejemplo: Para el modelo Weibull tenemos que

$$\lambda(t \mid \mathbf{x}, \psi) = \frac{\gamma}{\alpha(\mathbf{x})} t^{\gamma-1}$$

$$S(t \mid \mathbf{x}, \psi) = \exp \left\{ - \left(\frac{t}{\alpha(\mathbf{x})} \right)^{\gamma} \right\},$$

where $\alpha(\mathbf{x}) = \exp \{ \mathbf{x}^\top \boldsymbol{\beta} \}$



Inferencia

Por lo que La función de verosimilitud para el modelo Weibull (excepto por una constante que no depende de ψ) es dada por

$$\begin{aligned} \mathcal{L}(\psi) &\propto \prod_{i=1}^n \left\{ \frac{\gamma}{\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}} z_i^{\gamma-1} \right\}^{\delta_i} \exp\left\{ -\left(\frac{z_i}{\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}} \right)^\gamma \right\} \\ &\propto \gamma^r \exp \left\{ \exp \left(-\sum_{i=1}^n \delta_i \mathbf{x}_i^\top \boldsymbol{\beta} \right) + (\gamma - 1) \sum_{i=1}^n \log(z_i) \right. \\ &\quad \left. - \sum_{i=1}^n \left(\frac{z_i}{\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}} \right)^\gamma \right\}. \end{aligned}$$



Inferencia

Por lo que La función de verosimilitud para el modelo Weibull (excepto por una constante que no depende de ψ) es dada por

$$\begin{aligned} \mathcal{L}(\psi) &\propto \prod_{i=1}^n \left\{ \frac{\gamma}{\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}} z_i^{\gamma-1} \right\}^{\delta_i} \exp\left\{ -\left(\frac{z_i}{\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}} \right)^\gamma \right\} \\ &\propto \gamma^r \exp \left\{ \exp \left(-\sum_{i=1}^n \delta_i \mathbf{x}_i^\top \boldsymbol{\beta} \right) + (\gamma - 1) \sum_{i=1}^n \log(z_i) \right. \\ &\quad \left. - \sum_{i=1}^n \left(\frac{z_i}{\exp\{\mathbf{x}_i^\top \boldsymbol{\beta}\}} \right)^\gamma \right\}. \end{aligned}$$





Inferencia

Ejemplo: Así, la función de log-verosimilitud es dada por

$$\begin{aligned}\log \mathcal{L}(\psi) = & r \log(\gamma) + \exp \left(- \sum_{i=1}^n \delta_i \mathbf{x}_i^\top \boldsymbol{\beta} \right) + (\gamma - 1) \sum_{i=1}^n \log(z_i) \\ & - \sum_{i=1}^n \left(\frac{z_i}{\exp \{ \mathbf{x}_i^\top \boldsymbol{\beta} \}} \right)^\gamma.\end{aligned}$$



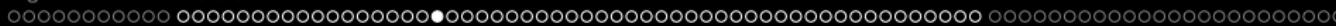


Inferencia

Para ajustar estos modelos en R, podemos usar la función `survreg` del paquete `survival`.

```
survreg(formula, data, weights, subset,  
na.action, dist="weibull", init=NULL, scale=0,  
control,parms=NULL,model=FALSE, x=FALSE,  
y=TRUE, robust=FALSE, cluster, score=FALSE, ...)
```





Inferencia

Para ajustar estos modelos en R, podemos usar la función `survreg` del paquete `survival`.

```
survreg(formula, data, weights, subset,  
        na.action, dist="weibull", init=NULL, scale=0,  
        control, parms=NULL, model=FALSE, x=FALSE,  
        y=TRUE, robust=FALSE, cluster, score=FALSE, ...)
```

formula: `Surv(time, delta) ~ cov1+...+covp`

dist: weibull, exponential, lognormal, loglogistic

scale: valor opcional fijado para la escala. Si es ≤ 0 , entonces la escala es estimada a partir de los datos.





Estimadores NP de la función de sobrevida

Por ejemplo, para el conjunto de datos de cáncer de pulmón y el modelo Weibull tenemos

```
> m.weibull=survreg(Surv(t2, d2) ~ ph.karno+age+sex,dist='weibull')  
> summary(m.weibull)
```

Call:

```
survreg(formula = Surv(t2, d2) ~ ph.karno + age + as.factor(sex2),  
        dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	-0.20408	0.64133	-0.32	0.7503
ph.karno	0.00926	0.00446	2.08	0.0379
age	-0.00891	0.00711	-1.25	0.2100
as.factor(sex2)2	0.37019	0.12796	2.89	0.0038
Log(scale)	-0.28085	0.06171	-4.55	5.3e-06

Scale= 0.755

Weibull distribution

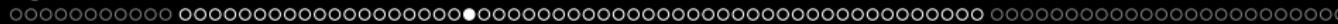
Loglik(model)= -171 Loglik(intercept only)= -179.8

Chisq= 17.59 on 3 degrees of freedom, p= 0.00053

Number of Newton-Raphson Iterations: 5

n= 227





Inferencia

Note que el modelo ajustado es

$$S(t \mid \mathbf{x}_i, \psi) = \exp \left\{ - \left(\frac{t}{\alpha(\mathbf{x}_i)} \right)^{1/0.755} \right\},$$

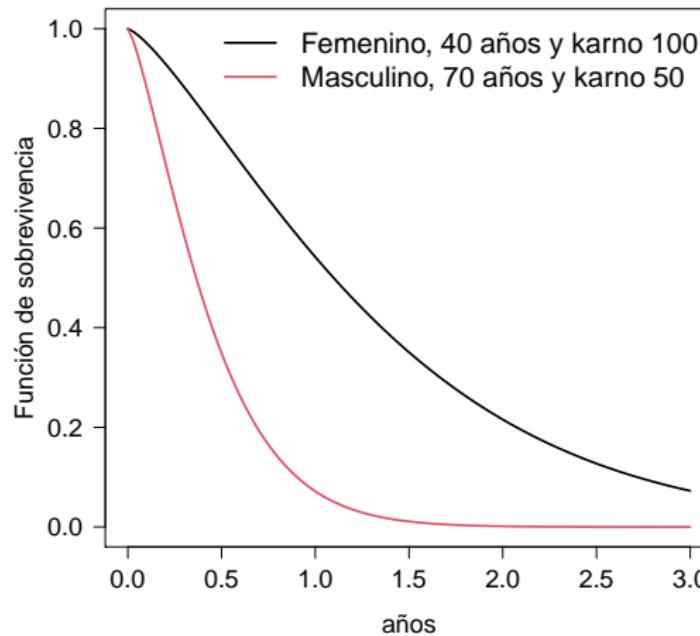
en que

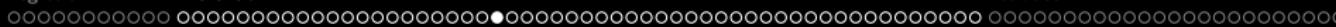
$$\alpha(\mathbf{x}_i) = \exp(-0.2041 + 0.0093 \times \text{ph.karno}_i - 0.0089 \times \text{age}_i + 0.3702 \times \text{sex}_i).$$



Inferencia

Función de sobrevivencia estimada para diferentes perfiles.



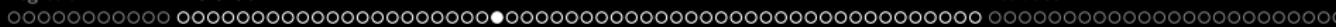


Inferencia

Interpretando los coeficientes

- $\exp(\beta_1) = \exp(0.0093) \approx 1.009$: para una edad y sexo fijados, por cada punto que aumenta el puntaje ph.karno la mediana del **tiempo de muerte** para pacientes con cáncer de pulmón **aumenta en 0.9%**.



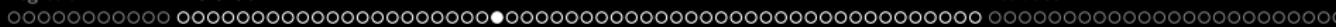


Inferencia

Interpretando los coeficientes

- $\exp(\beta_1) = \exp(0.0093) \approx 1.009$: para una edad y sexo fijados, por cada punto que aumenta el puntaje ph.karno la mediana del **tiempo de muerte** para pacientes con cáncer de pulmón **aumenta en 0.9%**.
- $\exp(\beta_2) = \exp(-0.0089) \approx 0.991$: para un puntaje ph.karno y sexo fijados, por cada año que aumenta la edad la mediana del **tiempo de muerte** para pacientes con cáncer de pulmón **disminuye en 0.9%**.





Inferencia

Interpretando los coeficientes

- $\exp(\beta_1) = \exp(0.0093) \approx 1.009$: para una edad y sexo fijados, por cada punto que aumenta el puntaje ph.karno la mediana del **tiempo de muerte** para pacientes con cáncer de pulmón **aumenta en 0.9%**.
- $\exp(\beta_2) = \exp(-0.0089) \approx 0.991$: para un puntaje ph.karno y sexo fijados, por cada año que aumenta la edad la mediana del **tiempo de muerte** para pacientes con cáncer de pulmón **disminuye en 0.9%**.
- $\exp(\beta_3) = \exp(0.3702) \approx 1.448$: para un puntaje ph.karno y edad fijados, la mediana del **tiempo de muerte** para pacientes con cáncer de pulmón de mujeres **aumenta en 44.8%** en relación a hombres.

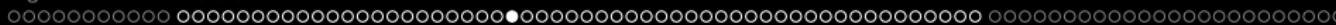




Inferencia

¿Qué pasa si no usamos la información de que algunas observaciones son **censuradas**?





Inferencia

¿Qué pasa si no usamos la información de que algunas observaciones son censuradas?

(Malas) alternativas:

- Usar todos los tiempos **como si fuesen fallas.**

```
> mw.weibull=survreg(Surv(t2, rep(1,length(t2))) ~  
ph.karno+age+as.factor(sex2),dist='weibull')
```





Inferencia

¿Qué pasa si no usamos la información de que algunas observaciones son **censuradas**?

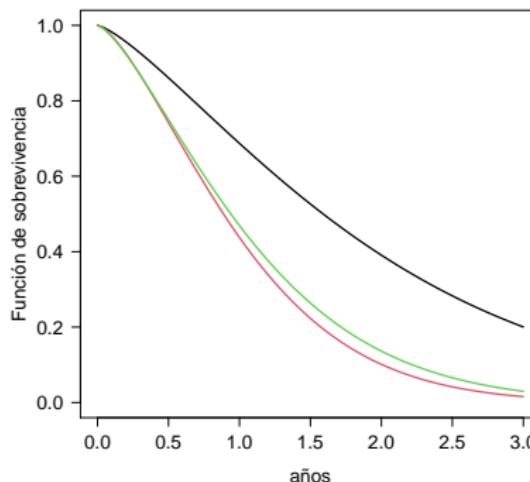
(Malas) alternativas:

- Usar todos los tiempos **como si fuesen fallas**.
`> mw.weibull=survreg(Surv(t2, rep(1,length(t2))) ~ ph.karno+age+as.factor(sex2),dist='weibull')`
- Usar **sólo** los tiempos que fueron **fallas**.
`> ii=which(d2==1)`
`> mw2.weibull=survreg(Surv(t2[ii], d2[ii]) ~ ph.karno[ii]+age[ii]+as.factor(sex)[ii],dist='weibull')`



Inferencia

Función de sobrevivencia **estimada** para Femenino, 40 años y karno 100.



Negro: Uso correcto de datos censurados.

Rojo: Usando todos los datos como si fuesen fallas.

Verde: Usando sólo las fallas.





Inferencia

Para testear **hipótesis simples** asociadas a las covariables, podemos usar el test de **Wald** o bien, el test de **razón de verosimilitud**.



Inferencia

Para testear las hipótesis

$$H_0 : \beta_{\text{age}} = 0 \text{ versus } H_1 : \beta_{\text{age}} \neq 0,$$

el test de **Wald** requiere la estadística observada

$$SW = \frac{\hat{\beta}_{\text{age}}}{\text{se}(\hat{\beta}_{\text{age}})} = -\frac{0.0089}{0.0071} = -1.25$$

con un p-valor asociado de 0.2100.

La ventaja es que este resultado es presentado directamente con la función `summary()`.





Estimadores NP de la función de sobrevivencia

```
> m.weibull=survreg(Surv(time, delta) ~ ph.karno+age+sex,dist='weibull')  
> summary(m.weibull)
```

Call:

```
survreg(formula = Surv(t2, d2) ~ ph.karno + age + as.factor(sex2),  
        dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	-0.20408	0.64133	-0.32	0.7503
ph.karno	0.00926	0.00446	2.08	0.0379
age	-0.00891	0.00711	-1.25	0.2100
as.factor(sex2)2	0.37019	0.12796	2.89	0.0038
Log(scale)	-0.28085	0.06171	-4.55	5.3e-06

Scale= 0.755

Weibull distribution

Loglik(model)= -171 Loglik(intercept only)= -179.8

Chisq= 17.59 on 3 degrees of freedom, p= 0.00053

Number of Newton-Raphson Iterations: 5

n= 227





Inferencia

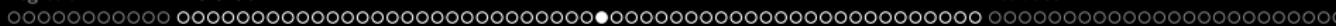
Ejemplo: para testear las hipótesis

$$H_0 : \beta_{\text{age}} = 0 \text{ versus } H_1 : \beta_{\text{age}} \neq 0,$$

el test de **RV** requiere la estadística $SLR = -2 \left\{ \log \mathcal{L}(\hat{\psi}_0) - \log \mathcal{L}(\hat{\psi}) \right\}$.

Nos falta la log-verosimilitud del **modelo reducido!**





Estimadores NP de la función de sobrevida

Para el modelo reducido tenemos que

```
> m0.weibull=survreg(Surv(time, delta) ~ ph.karno+sex,dist='weibull')
> summary(m0.weibull)
```

Call:

```
survreg(formula = Surv(t2, d2) ~ ph.karno + as.factor(sex2),
        dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	-0.87582	0.35745	-2.45	0.0143
ph.karno	0.01063	0.00436	2.44	0.0147
as.factor(sex2)2	0.37576	0.12815	2.93	0.0034
Log(scale)	-0.28033	0.06168	-4.55	5.5e-06

Scale= 0.756

Weibull distribution

Loglik(model)= -171.8 Loglik(intercept only)= -179.8
Chisq= 15.99 on 2 degrees of freedom, p= 0.00034

Number of Newton-Raphson Iterations: 5

n= 227





Inferencia

Para testear las hipótesis

$$H_0 : \beta_{\text{age}} = 0 \text{ versus } H_1 : \beta_{\text{age}} \neq 0,$$

el test de RV requiere la estadística

$$SLR = -2 \left\{ \log \mathcal{L}(\hat{\psi}_0) - \log \mathcal{L}(\hat{\psi}) \right\} = -2 \{-171.8 - (-171.0)\} = 1.6$$

con un p-valor asociado de 0.2059.

```
> pchisq(1.603829, df=1, lower.tail=FALSE)  
[1] 0.2053614
```





Inferencia

Note que también podemos estar interesados en testear hipótesis asociadas al parámetro de forma. Para el modelo Weibull, las hipótesis

$$H_0 : \log \sigma = 0 \text{ versus } H_1 : \log \sigma \neq 0,$$

implican testear el modelo **exponencial** (H_0) versus el modelo **Weibull** (H_1).

Nuevamente es posible usar el test de **Wald** y el test de **RV** fácilmente.



Estimadores NP de la función de sobrevida

Volviendo al ejemplo de cáncer de pulmón del modelo **Weibull completo**

```
> m.weibull=survreg(Surv(time, delta) ~ ph.karno+age+sex,dist='weibull')
> summary(m.weibull)
```

Call:

```
survreg(formula = Surv(t2, d2) ~ ph.karno + age + as.factor(sex2),
        dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	-0.20408	0.64133	-0.32	0.7503
ph.karno	0.00926	0.00446	2.08	0.0379
age	-0.00891	0.00711	-1.25	0.2100
as.factor(sex2)2	0.37019	0.12796	2.89	0.0038
Log(scale)	-0.28085	0.06171	-4.55	5.3e-06

Scale= 0.755

Weibull distribution

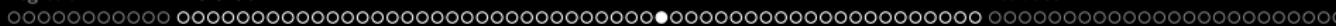
Loglik(model)= -171 Loglik(intercept only)= -179.8

Chisq= 17.59 on 3 degrees of freedom, p= 0.00053

Number of Newton-Raphson Iterations: 5

n= 227





Estimadores NP de la función de sobrevida

Volviendo al ejemplo de cáncer de pulmón del modelo **Weibull completo**

```
> m.weibull=survreg(Surv(time, delta) ~ ph.karno+age+sex,dist='weibull')
> summary(m.weibull)
```

Call:

```
survreg(formula = Surv(t2, d2) ~ ph.karno + age + as.factor(sex2),
        dist = "weibull")
```

	Value	Std. Error	z	p
(Intercept)	-0.20408	0.64133	-0.32	0.7503
ph.karno	0.00926	0.00446	2.08	0.0379
age	-0.00891	0.00711	-1.25	0.2100
as.factor(sex2)2	0.37019	0.12796	2.89	0.0038
Log(scale)	-0.28085	0.06171	-4.55	5.3e-06

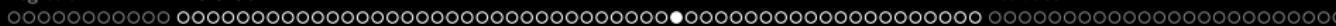
Scale= 0.755

Weibull distribution

```
Loglik(model)= -171    Loglik(intercept only)= -179.8
Chisq= 17.59 on 3 degrees of freedom, p= 0.00053
Number of Newton-Raphson Iterations: 5
n= 227
```

Es decir, el test de **Wald** p-valor asociado de < 0.0001.





Estimadores NP de la función de sobrevida

Para el modelo reducido (que en este caso sería el modelo exponencial) tenemos que

```
> m.exp=survreg(Surv(t2, d2) ~ ph.karno+age+as.factor(sex2),dist='exponential')
> summary(m.exp)
```

Call:

```
survreg(formula = Surv(t2, d2) ~ ph.karno + age + as.factor(sex2),
         dist = "exponential")
```

	Value	Std. Error	z	p
(Intercept)	-0.35736	0.84396	-0.42	0.6720
ph.karno	0.01277	0.00596	2.14	0.0321
age	-0.01147	0.00928	-1.24	0.2168
as.factor(sex2)2	0.47085	0.16728	2.81	0.0049

Scale fixed at 1

Exponential distribution

Loglik(model)= -180 Loglik(intercept only)= -188.3

Chisq= 16.68 on 3 degrees of freedom, p= 0.00082

Number of Newton-Raphson Iterations: 4

n= 227





Inferencia

La estadística de **RV** es

$$SLR = -2 \left\{ \log \mathcal{L}(\hat{\psi}_0) - \log \mathcal{L}(\hat{\psi}) \right\} = -2 \{-180.0 - (-171.0)\} = 18.0$$

con un p-valor asociado <0.0001.

```
> pchisq(18.0246, df=1, lower.tail=FALSE)
[1] 2.180687e-05
```

En ambos casos rechazamos la hipótesis nula (es decir, preferimos el modelo Weibull que el modelo exponencial).





Inferencia

Pensemos ahora en hipótesis compuestas de la forma

$$H_0 : R\psi = r \quad \text{versus} \quad H_1 : R\psi \neq r$$

en que R es una matriz de $q \times (p + 2)$ acorde a las q hipótesis que queremos expresar sobre los parámetros.





Inferencia

Por ejemplo, para el conjunto de datos de cáncer de pulmón
 $\psi = (\beta_{\text{Intercepto}}, \beta_{\text{ph.karno}}, \beta_{\text{age}}, \beta_{\text{sex}}, \log(\sigma))^\top$ y suponga que estamos interesados en **testear** las hipótesis

$$H_0^{(1)} : \beta_{\text{age}} = 0 \quad \text{versus} \quad H_1^{(1)} : \beta_{\text{age}} \neq 0$$

$$H_0^{(2)} : \beta_{\text{age}} = \beta_{\text{sex}} \quad \text{versus} \quad H_1^{(2)} : \beta_{\text{age}} \neq \beta_{\text{sex}}.$$





Inferencia

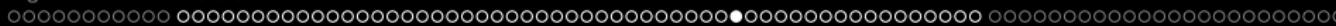
Por ejemplo, para el conjunto de datos de cáncer de pulmón
 $\psi = (\beta_{\text{Intercepto}}, \beta_{\text{ph.karno}}, \beta_{\text{age}}, \beta_{\text{sex}}, \log(\sigma))^{\top}$ y suponga que estamos interesados en **testear** las hipótesis

$$\begin{aligned} H_0^{(1)} : \beta_{\text{age}} &= 0 & \text{versus} & \quad H_1^{(1)} : \beta_{\text{age}} \neq 0 \\ H_0^{(2)} : \beta_{\text{age}} &= \beta_{\text{sex}} & \text{versus} & \quad H_1^{(2)} : \beta_{\text{age}} \neq \beta_{\text{sex}}. \end{aligned}$$

Entonces para este problema en particular,

$$R = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \end{pmatrix} \text{ y } r = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$





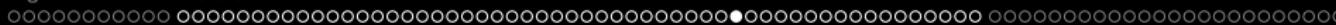
Inferencia

Para el problema general, el test de **Wald** usa la estadística

$$SLR = (\hat{R}\psi - r)^\top \left[R \times I(\hat{\psi}) \times R^\top \right]^{-1} (\hat{R}\psi - r).$$

Sobre H_0 , $SLR \sim \chi^2_{(q)}$.





Inferencia

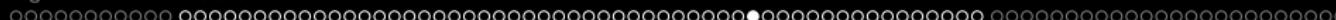
Para el problema general, el test de **Wald** usa la estadística

$$SLR = (\hat{R}\psi - r)^\top \left[R \times I(\hat{\psi}) \times R^\top \right]^{-1} (\hat{R}\psi - r).$$

Sobre H_0 , $SLR \sim \chi^2_{(q)}$.

Ya no es tan inmediato testear esta hipótesis (en el sentido que el software no proporciona directamente este resultado), pero recordando que $I(\hat{\psi})$ es la matriz de **varianzas-covarianzas** estimadas de $\hat{\psi}$, podemos calcular “a mano” esta estadística.





Inferencia

La matriz de **varianzas-covarianzas** puede ser fácilmente extraída usando la función vcov como se ejemplifica a continuación

```
> round(vcov(m.weibull),6)
```

	(Intercept)	ph.karno	age	as.factor(sex2)2	Log(scale)
(Intercept)	0.411307	-0.002062	-0.003787	-0.005928	-0.002436
ph.karno	-0.002062	0.000020	0.000007	-0.000011	0.000046
age	-0.003787	0.000007	0.000051	0.000022	-0.000029
as.factor(sex2)2	-0.005928	-0.000011	0.000022	0.016373	0.001243
Log(scale)	-0.002436	0.000046	-0.000029	0.001243	0.003808



Inferencia

Además, los **coeficientes de regresión** y el parámetro de **escala** pueden ser obtenidos mediante

```
> coef(m.weibull)
  (Intercept)      ph.karno          age  as.factor(sex2)2
-0.204079169    0.009264172   -0.008910583    0.370187051
> m.weibull$scale
[1] 0.7551445
```





Inferencia

Esto permite calcular fácilmente la estadística de **Wald** y su respectivo p-valor. Para el ejemplo, tenemos que

```
> psi=c(coef(m.weibull),log(m.weibull$scale))
> I.phi=vcov(m.weibull)
> R=matrix(c(0,0,1,0,0,0,0,1,-1,0),nrow=2,ncol=5,byrow=T)
> r=matrix(0,nrow=2)
> SLR=t(R%*%psi-r)%*%solve(R%*%I.phi%*%t(R))%*%(R%*%psi-r)
> p.value=pchisq(SLR,df=2,lower.tail=FALSE)
> SLR
[1] 10.12547
> p.value
[1] 0.006328214
```

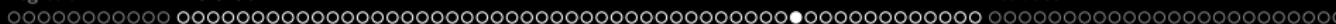
Es decir, con cualquier valor usual de significación rechazamos la hipótesis nula.



Inferencia

Otras funciones útiles para aplicar directamente en un objeto `survreg`. Para selección de modelos, es útil calcular el criterio de Akaike (Akaike, 1974; **AIC**) y el criterio de Schwarz (Schwarz, 1978; **BIC**).





Inferencia

Otras funciones útiles para aplicar directamente en un objeto `survreg`. Para selección de modelos, es útil calcular el criterio de Akaike (Akaike, 1974; **AIC**) y el criterio de Schwarz (Schwarz, 1978; **BIC**).

```
> m.weibull=survreg(Surv(t2, d2) ~ ph.karno+age+as.factor(sex2),  
  dist='weibull')  
> AIC(m.weibull)  
[1] 351.9927  
> BIC(m.weibull)  
[1] 369.1174
```



Inferencia

Un resumen para los modelos WEI, LN y LL, usando las 3 covariables (ph.karno, edad y sex2) en el conjunto de datos de cáncer de pulmón.

Modelo	AIC	BIC
WEI	351.9927	369.1174
LN	373.4434	390.5681
LL	354.2376	371.3624





Inferencia

Un resumen para los modelos WEI, LN y LL, usando las 3 covariables (ph.karno, edad y sex2) en el conjunto de datos de cáncer de pulmón.

Modelo	AIC	BIC
WEI	351.9927	369.1174
LN	373.4434	390.5681
LL	354.2376	371.3624

Basado en ambos criterios, el modelo Weibull es que el que provee un mejor ajuste al conjunto de datos de pulmón (entre los 3 modelos comparados).





Inferencia

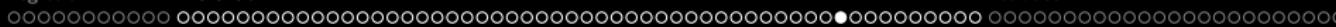
También podemos usar la función que permite seleccionar covariables

```
> m.weibull=survreg(Surv(t2, d2) ~ ph.karno+age+as.factor(sex2),  
  dist='weibull')  
> step(m.weibull)  
Start: AIC=351.99  
Surv(t2, d2) ~ ph.karno + age + as.factor(sex2)
```

	Df	AIC
- age	1	351.60
<none>		351.99
- ph.karno	1	354.29
- as.factor(sex2)	1	359.00

...





Inferencia

...

Step: AIC=351.6

Surv(t2, d2) ~ ph.karno + as.factor(sex2)

	Df	AIC
<none>		351.60
- ph.karno	1	355.51
- as.factor(sex2)	1	358.87

Call:

```
survreg(formula = Surv(t2, d2) ~ ph.karno + as.factor(sex2),  
        dist = "weibull")
```

Coefficients:

	ph.karno	as.factor(sex2)2
(Intercept)	0.01062524	0.37576225

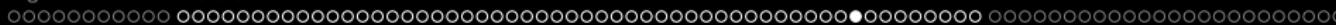
Scale= 0.7555318

Loglik(model)= -171.8 Loglik(intercept only)= -179.8

Chisq= 15.99 on 2 degrees of freedom, p= 0.000337

n= 227





Inferencia

Un resumen de las covariables seleccionadas para los modelos WEI, LN y LL se muestra a continuación.

modelo	covariables seleccionadas	AIC
WEI	ph.karno+sex2	351.60
LN	ph.karno+age+sex2	373.44
LL	ph.karno+sex2	353.55

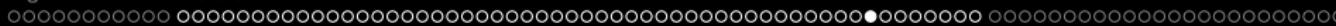
Note que el modelo WEI sigue siendo el preferido bajo el criterio AIC luego de hacer la selección de covariables.



Inferencia

Una alternativa bastante atractiva para ajustar modelos a datos **censurados** viene dada a través del paquete `gamlss` (generalized additive models for location scale and shape).





Inferencia

Una alternativa bastante atractiva para ajustar modelos a datos **censurados** viene dada a través del paquete `gamlss` (generalized additive models for location scale and shape).

El paquete `gamlss.dist` incluye una serie de distribuciones que pueden ser usadas en el contexto de datos censurados. Entre otras, tenemos: Box-Cox Cole and Green (**BCCG**), gamma (**GA**), familia de distribución gamma (**GAF**), inversa gamma (**IGAMMA**), inversa gaussian (**IG**), generalized gamma (**GG**), generalized inverse gamma (**GIG**), tres parametrizaciones alternativas de la Weibull (**WEI**, **WEI2**, **WEI3**), entre otras.



Inferencia

Por ejemplo, la función **GA** asociado al modelo gamma con la siguiente parametrización

Details

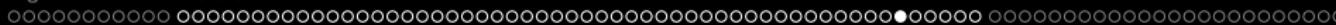
The specific parameterization of the gamma distribution used in GA is

$$f(y|\mu, \sigma) = \frac{y^{(1/\sigma^2)-1} \exp[-y/(\sigma^2\mu)]}{(\sigma^2\mu)^{(1/\sigma^2)}\Gamma(1/\sigma^2)}$$

for $y > 0$, $\mu > 0$ and $\sigma > 0$.

Es decir, $\mathbb{E}(Y) = \mu$ y $\text{Var}(Y) = \mu^2\sigma^2$.



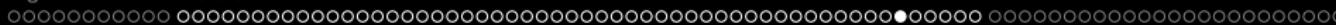


Inferencia

De esta forma, podemos ajustar este modelo incluyendo **covariables** en μ (la **media** en este caso) usando

```
> require(gamlss); require(gamlss.dist)
> require(gamlss.cens)
> m.ga=gamlss(Surv(t2, d2) ~ ph.karno+age+as.factor(sex2), family=cens(GA))
```





Inferencia

De esta forma, podemos ajustar este modelo incluyendo **covariables** en μ (la **media** en este caso) usando

```
> require(gamlss); require(gamlss.dist)
> require(gamlss.cens)
> m.ga=gamlss(Surv(t2, d2) ~ ph.karno+age+as.factor(sex2), family=cens(GA))
```

NOTA: para este modelo la ligación por defecto para μ y para σ es el logaritmo. Chequear con el manual del paquete cuál es la ligación para cada parámetro.

Por otra parte, por defecto la especificación usual de las covariables (como aparece arriba) es asumida para μ .





Inferencia

```
> m.ga
```

```
Family: c("GArc", "right censored Gamma")
```

```
Fitting method: RS()
```

```
Call: gamlss(formula = Surv(t2, d2) ~ ph.karno + age + as.factor(sex2),  
family = cens(GA))
```

Mu Coefficients:

	ph.karno	age	as.factor(sex2)
(Intercept)	0.011427	-0.009726	0.403649
-0.409178			

Sigma Coefficients:

(Intercept)
-0.2183

Degrees of Freedom for the fit: 5 Residual Deg. of Freedom 222

Global Deviance: 341.451

AIC: 351.451

SBC: 368.575





Inferencia

Un resumen con los errores estándar de cada parámetro puede ser obtenido usando

```
> summary(m.ga)
```

```
...
```

```
Mu link function: log
```

```
Mu Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.409178	0.582928	-0.702	0.483452
ph.karno	0.011427	0.004432	2.578	0.010567 *
age	-0.009726	0.006053	-1.607	0.109512
as.factor(sex2)2	0.403649	0.109943	3.671	0.000302 ***

```
---
```

```
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
```

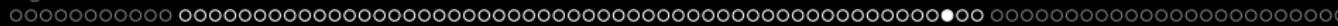
```
Sigma link function: log
```

```
Sigma Coefficients:
```

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-0.21825	0.04282	-5.097	7.3e-07 ***

```
---
```

```
Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1
```

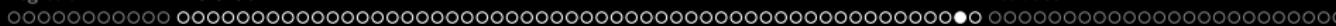


Inferencia

Si quisieramos incluir covariables en sigma también, lo podemos como sigue

```
> m.ga2=gamlss(Surv(t2, d2) ~ ph.karno+age+as.factor(sex2),  
sigma.fo=~ph.karno+age+as.factor(sex2), family=cens(GA))
```





Inferencia

```
> m.ga2
```

```
Family: c("GArc", "right censored Gamma")
```

```
Fitting method: RS()
```

```
Call: gamm4(formula = Surv(t2, d2) ~ ph.karno + age + as.factor(sex2),  
sigma.formula = ~ph.karno + age + as.factor(sex2), family = cens(GA))
```

Mu Coefficients:

	ph.karno	age	as.factor(sex2)
(Intercept)	0.013027	-0.003293	0.331798

Sigma Coefficients:

	ph.karno	age	as.factor(sex2)
(Intercept)	-0.002736	0.013879	-0.101953

Degrees of Freedom for the fit: 8 Residual Deg. of Freedom 219

Global Deviance: 331.444

AIC: 347.444

SBC: 374.843



Inferencia

Nota: es posible usar la función **step** con un objeto ajustado con la función **gamlss**. Sin embargo, la selección de covariables sólo se realizará para el parámetro μ .

```
> m.ga2=gamlss(Surv(time, delta) ~ ph.karno+age+sex,  
  sigma.fo=~ph.karno+age+sex,family=cens(GA))  
> step(m.ga2)
```

(Se omitirá la salida de esta función)



Residuos

Análisis de residuos.



Residuos

Tipos de residuos:

- Residuos de Cox-Snell.
- Residuos estandarizados.
- Residuos de devianza.
- Residuos aleatorizados de cuantil.



Residuos

Los **residuos de Cox-Snell** (Cox and Snell, 1968) son definidos como

$$r_i^C = \Lambda(t_i | x_i, \hat{\psi}).$$

Es decir, la función de riesgo acumulada estimada en cada tiempo observado.



Residuos

Los **residuos de Cox-Snell** (Cox and Snell, 1968) son definidos como

$$r_i^C = \Lambda(t_i | x_i, \hat{\psi}).$$

Es decir, la función de riesgo acumulada estimada en cada tiempo observado.

Si el modelo fuese adecuado, $r_1^C, r_2^C, \dots, r_n^C$ deberían comportarse como una muestra aleatoria censurada del modelo exponencial de media 1, digamos $E(1)$.



Residuos

Esto puede ser chequeado mediante las técnicas que vimos anteriormente.

- Podemos construir el **estimador de KM** basado en $(r_1^C, \delta_1), (r_2^C, \delta_2), \dots, (r_n^C, \delta_n)$ con sus respectivos IC y superponer la curva de la función $S(t) = e^{-t}$, que corresponde a la función de sobrevivencia del modelo E(1).



Residuos

Esto puede ser chequeado mediante las técnicas que vimos anteriormente.

- Podemos construir el **estimador de KM** basado en $(r_1^C, \delta_1), (r_2^C, \delta_2), \dots, (r_n^C, \delta_n)$ con sus respectivos IC y superponer la curva de la función $S(t) = e^{-t}$, que corresponde a la función de sobrevivencia del modelo E(1).
- Podemos construir el gráfico de r_i^C versus $\hat{\Lambda}(r_i^C)$, $i = 1, \dots, n$. Si el modelo es apropiado, deberíamos observar una **recta de倾inación 1** que pase por el **origen** pues $\Lambda(t) = t$ para el modelo E(1).



Residuos

¿Cómo calcular estos residuos en R? Infelizmente no hay una forma directa de calcularlos, pero es simple usando la **definición**.

```
> alpha.est<-exp(m.weibull$linear.predictors)
> sigma.est<-m.weibull$scale
> rC.w<--log(pweibull(t2,shape=alpha.est,scale=sigma.est,lower.tail=FALSE))
```

NOTA: para el modelo **LN** y **LL**, basta cambiar pweibull por plnorm o pllogis (paquete actuar).



Residuos

Los **residuos estandarizados** (también conocidos como residuos normalizados) son definidos por

$$r_i^P = \frac{y_i - \hat{\mu}(\mathbf{x})}{\hat{\sigma}} = \frac{\log t_i - \mathbf{x}^\top \hat{\boldsymbol{\beta}}}{\hat{\sigma}}.$$

Residuos

Los **residuos estandarizados** (también conocidos como residuos normalizados) son definidos por

$$r_i^P = \frac{y_i - \hat{\mu}(\mathbf{x})}{\hat{\sigma}} = \frac{\log t_i - \mathbf{x}^\top \hat{\beta}}{\hat{\sigma}}.$$

Si el modelo fuese apropiado, entonces $r_1^P, r_2^P, \dots, r_n^P$ debe ser una muestra censurada del respectivo **modelo estándar**. Es decir, **VE**(0, 1), **N**(0, 1) y **L**(0, 1) para el modelo **WEI**, **LN** y **LL**, respectivamente.

Técnicas similares a las discutidas anteriormente pueden ser usadas para validar esta hipótesis.



Residuos

Los **residuos estandarizados** (también conocidos como residuos normalizados) son definidos por

$$r_i^P = \frac{y_i - \hat{\mu}(\mathbf{x})}{\hat{\sigma}} = \frac{\log t_i - \mathbf{x}^\top \hat{\beta}}{\hat{\sigma}}.$$

Si el modelo fuese apropiado, entonces $r_1^P, r_2^P, \dots, r_n^P$ debe ser una muestra censurada del respectivo **modelo estándar**. Es decir, **VE(0, 1)**, **N(0, 1)** y **L(0, 1)** para el modelo **WEI**, **LN** y **LL**, respectivamente.

Técnicas similares a las discutidas anteriormente pueden ser usadas para validar esta hipótesis.

Nota: Hacer el análisis de residuos basado en los residuos de **Cox-Snell** es equivalente a hacer el análisis basado en los residuos estandarizados.



Residuos

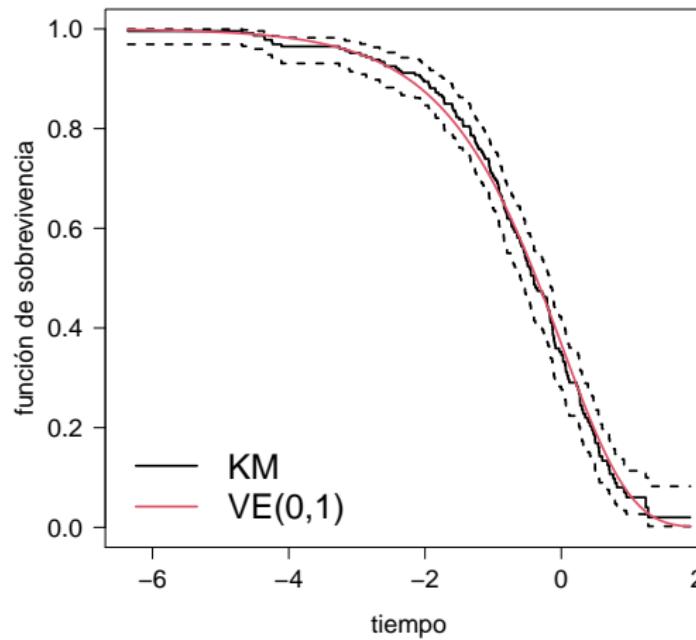
Estos residuos pueden ser calculados usando la función `residuals` sobre un objeto `psm` del paquete `rms`

```
> require(rms)
> m2.weibull=psm(Surv(t2, d2) ~ ph.karno+age+as.factor(sex2),dist="weibull")
> rP.w=residuals(m2.weibull, type="censored.normalized")
> KM.rP.w=survfit(rP.w ~ 1, conf.type="log-log")
> plot(KM.rP.w, ...)
> curve(exp(-exp(x)), ...)
```



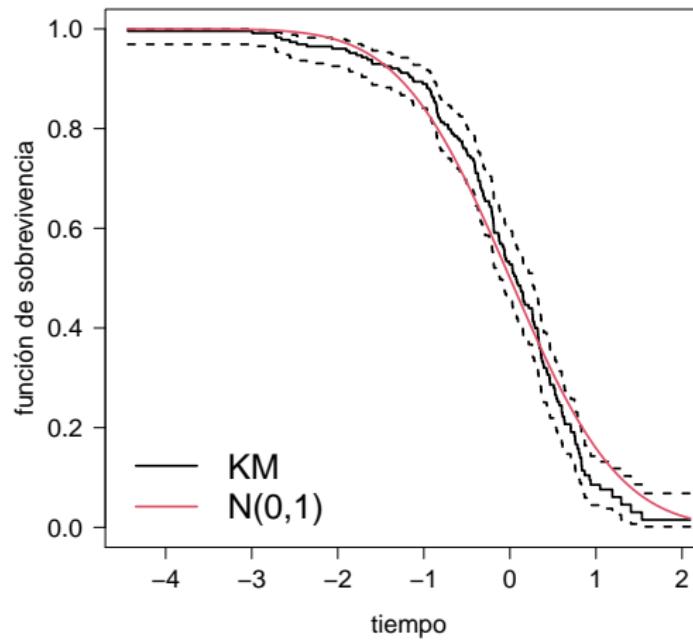
Residuos

Para el modelo WEI en los datos de cáncer de pulmón tenemos los siguientes residuos estandarizados



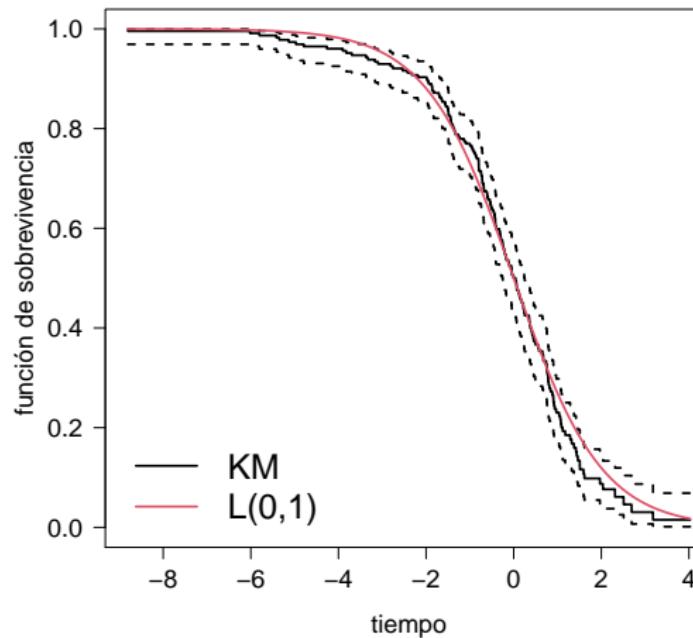
Residuos

Para el modelo LN en los datos de cáncer de pulmón tenemos los siguientes residuos estandarizados



Residuos

Para el modelo LL en los datos de cáncer de pulmón tenemos los siguientes residuos estandarizados



Residuos

Los **residuos de devianza** (también conocidos en la literatura como residuos de devianza de martingal) son definidos como

$$r_i^D = \text{signo}(r_i^M) \left[-2 \left(r_i^M + \delta_i \log(\delta_i - r_i^M) \right) \right]^{1/2}$$

en que $r_i^M = \delta_i - r_i^C$ son los denominados **residuos de martingal**.



Residuos

Los **residuos de devianza** (también conocidos en la literatura como residuos de devianza de martingal) son definidos como

$$r_i^D = \text{signo}(r_i^M) \left[-2 \left(r_i^M + \delta_i \log(\delta_i - r_i^M) \right) \right]^{1/2}$$

en que $r_i^M = \delta_i - r_i^C$ son los denominados **residuos de martingal**.

Si el modelo fuese apropiado, entonces $r_1^D, r_2^D, \dots, r_n^D$ deberían comportarse como una muestra aleatoria del modelo $\mathbf{N}(0, 1)$.



Residuos

Aprovechando que los residuos de Cox-Snell ya fueron calculados, podemos calcular los residuos de Martingal y luego los de devianza

```
> rM.w=delta-rC.w  
> rM.ln=delta-rC.ln  
> rM.ll=delta-rC.ll  
  
> rD.w=sign(rM.w)*sqrt(-2*(rM.w+delta*log(delta-rM.w)))  
> rD.ln=sign(rM.ln)*sqrt(-2*(rM.ln+delta*log(delta-rM.ln)))  
> rD.ll=sign(rM.ll)*sqrt(-2*(rM.ll+delta*log(delta-rM.ll)))
```

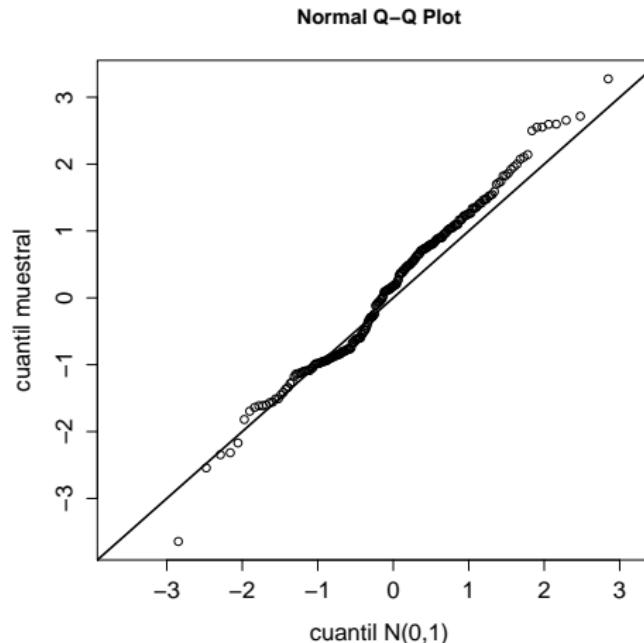




Residuos

Podemos hacer un **qq-plot** de los residuos de devianza de cada modelo.
Caso WEI

`qqnorm(rD.w,...)`

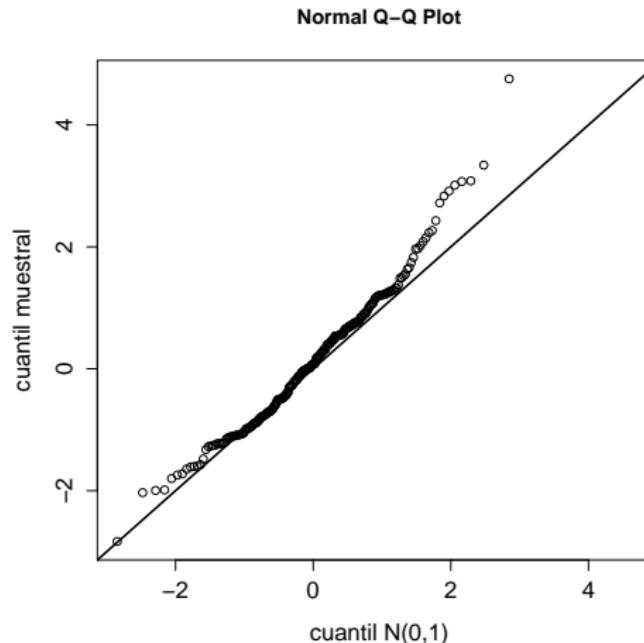




Residuos

Podemos hacer un **qq-plot** de los residuos de devianza de cada modelo.
Caso LN

```
qqnorm(rD.ln,...)
```

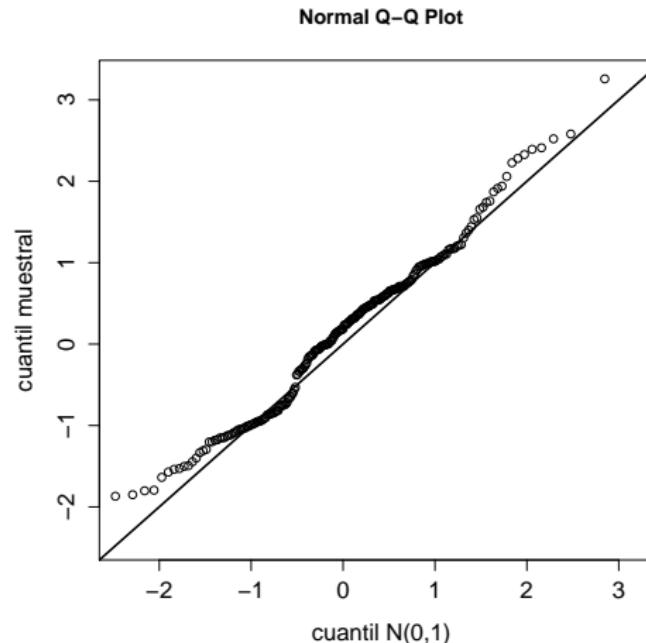




Residuos

Podemos hacer un **qq-plot** de los residuos de devianza de cada modelo.
Caso LL

```
qnorm(rD.ll,...)
```



Residuos

La ventaja de los **residuos de devianza** es que podemos aplicar test de hipótesis para verificar el supuesto de normalidad de los residuos, dado que el resultado teórico es que los residuos se comportan como una muestra aleatoria (no censurada) del modelo $N(0, 1)$.



Residuos

La ventaja de los **residuos de devianza** es que podemos aplicar test de hipótesis para verificar el supuesto de normalidad de los residuos, dado que el resultado teórico es que los residuos se comportan como una muestra aleatoria (no censurada) del modelo $N(0, 1)$.

Esto se puede validar a través de diversos **tests de normalidad**:
Kolmogorov-Smirnov, Shapiro-Wilks, Anderson-Darling,
Cramér-Von-Mises, test de asimetría (Shapiro, Wilk and Chen (1968)),
test de curtosis (Shapiro, Wilk and Chen (1968)), Jarque-Bera, Agostino,
etc.



Residuos

```
> ks.test(rD.w, "pnorm")
One-sample Kolmogorov-Smirnov test

data: rD.w
D = 0.12011, p-value = 0.00286
alternative hypothesis: two-sided

> shapiro.test(rD.w)
Shapiro-Wilk normality test

data: rD.w
W = 0.98847, p-value = 0.0653
```





Residuos

```
> require(nortest)
> ad.test(rD.w)

Anderson-Darling normality test

data: rD.w
A = 1.1163, p-value = 0.006237

> cvm.test(rD.w)

Cramer-von Mises normality test

data: rD.w
W = 0.19818, p-value = 0.005607
```





Residuos

```
> require(normtest)
> kurtosis.norm.test(rD.w)
    Kurtosis test for normality

data: rD.w
T = 2.8206, p-value = 0.5785

> skewness.norm.test(rD.w)
    Skewness test for normality

data: rD.w
T = 0.0069342, p-value = 0.962

> jb.norm.test(rD.w)
    Jarque-Bera test for normality

data: rD.w
JB = 0.30619, p-value = 0.856
```



Residuos

```
> require(moments)
> agostino.test(rD.w)
  D'Agostino skewness test

data: rD.w
skew = 0.0069342, z = 0.0439602, p-value = 0.9649
alternative hypothesis: data have a skewness
```



Residuos

Un resumen de los tests de normalidad para los 3 modelos ajustados en el conjunto de datos de cáncer de pulmón.

	WEI	LN	LL
KS	0.0029	0.0915	0.0540
SW	0.0653	0.0015	0.0234
AD	0.0062	0.0183	0.0044
CVM	0.0056	0.0863	0.0081
curtosis	0.5595	0.0210	0.6010
asimetria	0.9665	0.0010	0.2920
JB	0.8465	0.0020	0.4595
AG	0.9649	0.0007	0.2792



Residuos

Los **residuos de cuantil aleatorizados** (Dunn and Smyth, 1996) para datos continuos son definidos como

$$r_i^{CA} = \Phi^{-1} \left(F(t_i | \mathbf{x}_i, \hat{\psi}) \right).$$



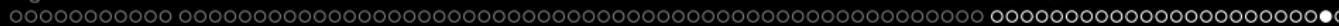
Residuos

Los **residuos de cuantil aleatorizados** (Dunn and Smyth, 1996) para datos continuos son definidos como

$$r_i^{CA} = \Phi^{-1} \left(F(t_i | \mathbf{x}_i, \hat{\psi}) \right).$$

Si el modelo fuese apropiado, entonces $r_1^{CA}, r_2^{CA}, \dots, r_n^{CA}$ deberían comportarse como una muestra aleatoria del modelo $N(0, 1)$.

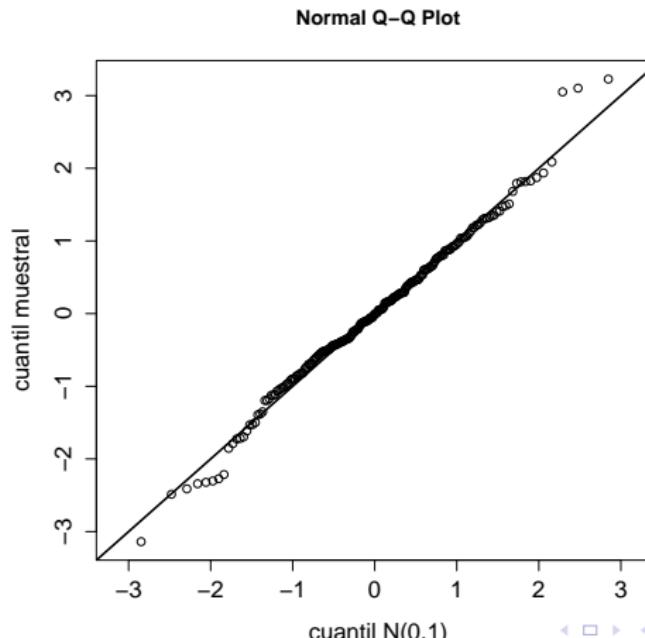




Residuos

Para un objeto `gamlss`, estos residuos son simples de calcular.

```
> m.ga=gamlss(Surv(t2, d2) ~ ph.karno+age+as.factor(sex2), family=cens(GA))
> rCA.ga=residual(m.ga)
> qqnorm(rCA.ga, ...)
```



Inferencia

Otros tópicos que usualmente se ven en análisis de sobrevivencia.



Otros tópicos

- **Modelo de riesgos proporcionales de Cox.** (Cox, 1972).

Se caracteriza por definir la **función de riesgo** del modelo en la forma

$$\lambda(t \mid \mathbf{x}_i) = \lambda_0(t) \exp(\mathbf{x}_i^\top \boldsymbol{\beta}), \quad i = 1, \dots, n,$$

es decir, una estructura **semi-paramétrica** (no paramétrica para $\lambda_0(t)$ y paramétrica para $\exp(\mathbf{x}^\top \boldsymbol{\beta})$).



Otros tópicos

- **Riesgos competitivos:** Cuando se está interesado en **múltiples eventos de interés**, pero sólo puede ser observado uno de esos eventos en cada individuo.



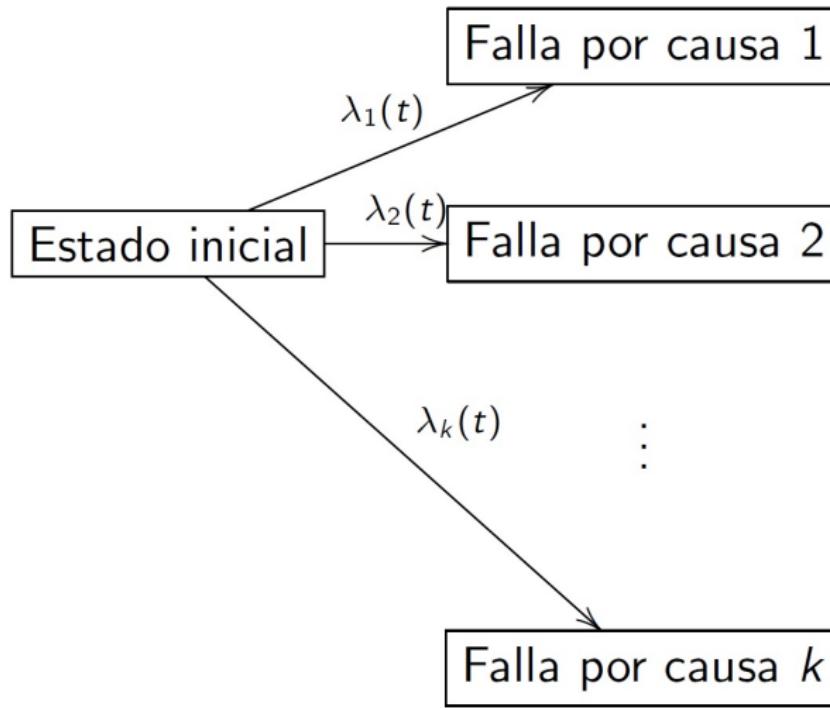
Otros tópicos

- **Riesgos competitivos:** Cuando se está interesado en **múltiples eventos de interés**, pero sólo puede ser observado uno de esos eventos en cada individuo.

Ejemplo: Se desea saber el tiempo en que cierto tipo de camiones mineros dejan de funcionar por fallas no atribuibles a sus conductores. Las fallas pueden ser producidas por problemas mecánicos, problemas eléctricos o bien, por problemas hidráulicos.



Otros tópicos



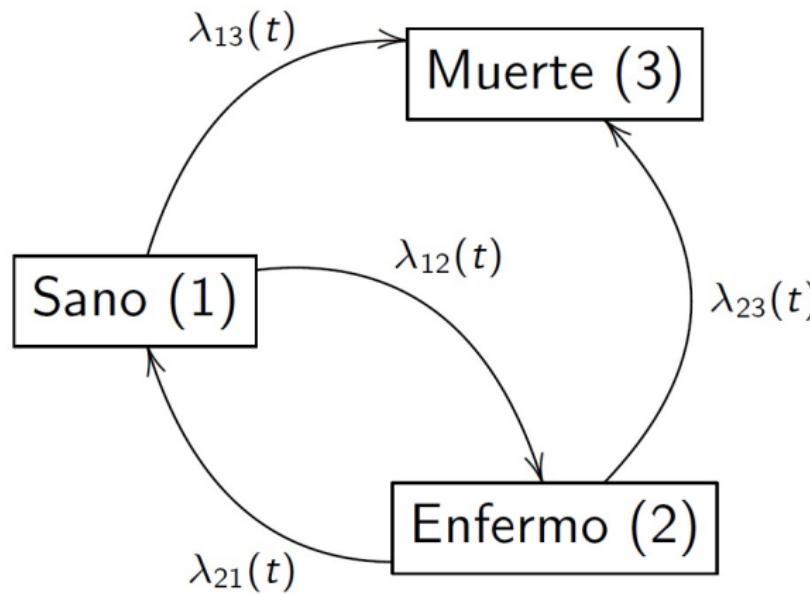
Otros tópicos

- **Modelos multi-estados:** Cuando se está interesado en **múltiples eventos de interés**, pero pueden ser observados uno o más de esos eventos en cada individuo.



Otros tópicos

- **Modelos multi-estados:** Cuando se está interesado en **múltiples eventos de interés**, pero pueden ser observados uno o más de esos eventos en cada individuo.



Otros tópicos

- **Fragilidad** (efectos aleatorios): Cuando las observaciones están agrupadas en clusters (digamos, J clusters).

$$\lambda_{ij}(t \mid w_j; \mathbf{x}_{ij}) = w_j \lambda_0(t) \exp(\mathbf{x}_{ij}^\top \boldsymbol{\beta}), \quad i = 1, \dots, n_j; j = 1, \dots, J,$$

en que w_1, \dots, w_J representa los términos de fragilidad y consideramos que el cluster j posee n_j observaciones.



Otros tópicos

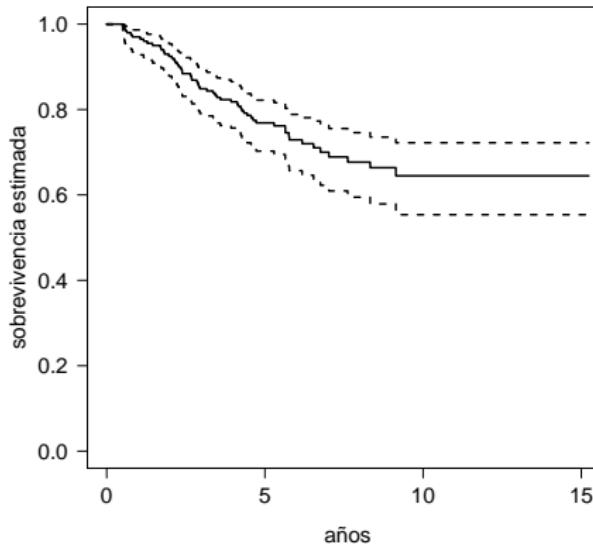
- **Fracción de cura:** Cuando existe en la población una proporción de individuos que nunca experimentarán el evento de interés.



Otros tópicos

- **Fracción de cura:** Cuando existe en la población una proporción de individuos que nunca experimentarán el evento de interés.

Ejemplo: Pacientes con cáncer de melanoma.



Otros tópicos

En este caso, una alternativa es usar el modelo de mixtura (Berkson and Gage, 1952).

$$S_{pop}(t; \theta) = 1 - \pi + \pi S(t; \lambda),$$

en que $1 - \pi \in (0, 1)$ representa la proporción de individuos curados y $S(t; \lambda)$ representa la función de sobrevivencia de los individuos susceptibles (no curados).



Otros tópicos

En este caso, una alternativa es usar el modelo de mixtura (Berkson and Gage, 1952).

$$S_{pop}(t; \theta) = 1 - \pi + \pi S(t; \lambda),$$

en que $1 - \pi \in (0, 1)$ representa la proporción de individuos curados y $S(t; \lambda)$ representa la función de sobrevivencia de los individuos susceptibles (no curados).

Note que

$$\lim_{t \rightarrow +\infty} S_{pop}(t) = 1 - \pi + \pi \lim_{t \rightarrow +\infty} S(t; \lambda) = 1 - \pi > 0,$$

pues se asume que $S(t; \lambda)$ es propia.



Referencias



Referencias Bibliográficas

-  Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, **19**, 716-723.
-  Berkson, J., Gage, R.P. (1952). Survival Curve for Cancer Patients Following Treatment. *Journal of the American Statistical Association*, **47**, 501-515.
-  Cox, D., Snell, E. (1968) A general definition of residuals. *Journal of the Royal Statistical Society, Series B*, **30**, 248-275.
-  Dunn, P.K., Smyth, G.K. (1996). Randomized quantile residuals. *Journal of Computational and Graphical Statistics*, **5**, 236-244.
-  Harrell Jr, F.E. (2020). rms: Regression Modeling Strategies. R package version 6.0-1. <https://CRAN.R-project.org/package=rms>
-  Hosmer, D.W., Lemeshow, S. (1999). *Applied Survival Analysis*. John Wiley and Sons, New York.
-  Kalbfleish, J.D, Prentice, R.L. (1980). *The Statistical Analysis of Failure Time Data*. John Wiley and Sons, New York.



Referencias Bibliográficas

-  Kaplan, E.L., Meier, P. (1958). The Statistical Analysis of Failure Time Data. John Wiley and Sons, New York.
-  Klein, J.P., Moeschberger, M.L. (1997). Survival Analysis: Techniques for Censored and Truncated Data. Springer-Verlag, New York.
-  Kleinbaum, D.G., Klein, M. (2012). Survival analysis: A self-learning text. New York: Springer. Third edition.
-  Nelson, W. (1972). Theory and Applications of Hazard Plotting for Censored Failure Data. Technometrics, **14**, 945-965.
-  Rigby R.A., Stasinopoulos D.M. (2005). Generalized additive models for location, scale and shape,(with discussion), Appl. Statist., **54**, **3**, 507-554.
-  Schwarz, G.E. (1978) Estimating the dimension of a model. Annals of Statistics, **6**, 461-464.



Referencias Bibliográficas

-  Stasinopoulos, M., Rigby, R. (2020). *gamlss.dist*: Distributions for Generalized Additive Models for Location Scale and Shape. R package version 5.1-6.
<https://CRAN.R-project.org/package=gamlss.dist>
-  Stasinopoulos, M., Rigbym B., Mortan, N. (2018). *gamlss.cens*: Fitting an Interval Response Variable Using *gamlss.family* Distributions. R package version 5.0-1.
<https://CRAN.R-project.org/package=gamlss.cens>
-  Therneau T (2020). A Package for Survival Analysis in R. R package version 3.1-12. URL: <https://CRAN.R-project.org/package=survival>.
-  Williams, J.S., Lagakos, S.W. (1977) Models for censored survival analysis: Constant-sum and variable-sum models. *Biometrika*, **64**, 215-224.
-  Yee, T.H. (2010). The VGAM Package for Categorical Data Analysis. *Journal of Statistical Software*, **32**(10), 1-34. URL <http://www.jstatsoft.org/v32/i10/>



Muchas gracias!

