

MODELANDO DATOS MEDIANTE DISTRIBUCIONES DE DAÑO ACUMULATIVO

VÍCTOR LEIVA

UNIVERSIDAD DE VALPARAÍSO, CHILE

victor.leiva@uv.cl, victor.leiva@yahoo.com

XXXVI JORNADAS NACIONALES DE ESTADÍSTICA
TEMUCO, CHILE, 6 AL 9 DE OCTUBRE DE 2009

En este trabajo también participaron:

MICHELLI BARROS

UNIVERSIDADE FEDERAL DE CAMPINA GRANDE, BRASIL

GILBERTO A. PAULA

UNIVERSIDADE DE SÃO PAULO, BRASIL

SAMUEL C. SAUNDERS

UNIVERSITY OF WASHINGTON, SEATTLE, USA

Contenidos del curso

- 1 PRELIMINARES
- 2 LA DISTRIBUCIÓN BIRNBAUM-SAUNDERS (BS)
- 3 LA DISTRIBUCIÓN BS GENERALIZADA (BSG)
- 4 DISTRIBUCIÓN SENH-NORMAL
- 5 MODELOS DE REGRESIÓN BS Y SU DIAGNÓSTICO

PRELIMINARES

Historia del modelo de Birnbaum & Saunders

La distribución Birnbaum-Saunders y su generalización han tenido la siguiente evolución:

- Davis (1952).
- Epstein & Sobel (1953, 1954a, 1954b, 1955).
- Birnbaum & Saunders (1958, 1968, 1969a, 1969b).
- Díaz & Leiva (2005).

Birnbaum & Saunders

Zygmund William Birnbaum (Bill) - Samuel C. Saunders (Sam)



(1903-2000)



Implementación computacional de los modelo BS y BSG

El lenguaje R desarrollado originalmente por Ihaka & Gentleman (1996) es un paquete de software de fuente abierta útil para análisis y gráficos estadísticos que puede ser obtenido sin costo desde <http://www.R-project.org>.

La distribución BS ha sido implementada por Leiva, Hernández & Riquelme (2006) en código R en un paquete llamado `bs` disponible desde CRAN (<http://CRAN.R-project.org/>).

La distribución BS generalizada ha sido implementada por Barros, Paula & Leiva (2009) en código R en el paquete `gbs` también disponible desde CRAN.

Detalles acerca de cómo descargar e instalar R y de los paquetes `bs` y `gbs` serán revisados durante el desarrollo de este curso.

Proceso de fatiga de materiales

La distribución Birnbaum-Saunders es también conocida como **distribución de vida de la fatiga**. El proceso de fatiga de materiales consta de las siguientes tres etapas:

- i) Comienzo de una fisura imperceptible (P_5 - P_{10} de la vida útil).
- ii) Crecimiento y propagación de la fisura producto del estrés cíclico y la tensión.
- iii) Ruptura o falla del material debido a fatiga (evento puntual que no ocupa tiempo).

Justificación física

Considere un espécimen de material expuesto a una secuencia de m cargas cíclicas, digamos $\{l_i; i = 1, \dots, m\}$. El esquema de carga es el siguiente:

$$\underbrace{l_1, l_2, \dots, l_m}_{\text{Ciclo 1}} \underbrace{l_{m+1}, l_{m+2}, \dots, l_{2m}}_{\text{Ciclo 2}} \dots \underbrace{l_{jm+1}, l_{jm+2}, \dots, l_{jm+m}}_{\text{Ciclo (j+1)}}$$

esto es, $l_{jm+i} = l_{km+i}$, con $j \neq k$.

Birnbaum & Saunders (1969a) supusieron que la carga es una función unimodal y continua en un intervalo. Esta carga estresa el material y lo debilita.

Supuestos estadísticos

Explicado el entorno físico, ahora se mencionan los supuestos estadísticos dados por Birnbaum & Saunders (1969a):

- (1) Un espécimen de un material está expuesto a cargas cíclicas que producen desgaste.
- (2) La ruptura (falla) ocurre cuando el tamaño de la grieta excede un umbral denotado por ω .
- (3) La secuencia de cargas son las mismas de un ciclo a otro.
- (4) El crecimiento de la grieta, X_i , debido a la carga l_i durante el ciclo j -ésimo es una v.a.
- (5) El tamaño total de la grieta, Y_j , debido al ciclo j -ésimo es también una v.a. con media μ y varianza σ^2 .
- (6) Las v.a. Y_i son independientes.

Supuestos estadísticos (cont.)

La extensión de la grieta debido al ciclo $(j + 1)$ -ésimo de carga es

$$Y_{j+1} = X_{jm+1} + X_{jm+2} + \cdots + X_{jm+m}, \quad j = 0, 1, 2, \dots$$

Bajo una aplicación repetitiva de n ciclos de cargas, el tamaño acumulado de la grieta al final del ciclo n -ésimo está dado por

$$Z_n = \sum_{i=1}^n Y_i. \quad (1)$$

Entonces, basados en Ecuación (1), los supuestos [▶ \(5\) y \(6\)](#) y el TLC, es posible establecer que Z_n converge en distribución a una normal con media $n\mu$ y varianza $n\sigma^2$, cuando $n \rightarrow \infty$.

Supuestos estadísticos (cont.)

Sea N el número de ciclos requeridos hasta la falla. La fda de N , basada en el teorema de probabilidad total, es

$$\begin{aligned}\mathbb{P}(N \leq n) &= \mathbb{P}(N \leq n, Z_n > \omega) + \mathbb{P}(N \leq n, Z_n \leq \omega) \\ &= \mathbb{P}(Z_n > \omega) + \mathbb{P}(N \leq n, Z_n \leq \omega),\end{aligned}$$

donde ω es el umbral dado en supuesto (2). Note que $\mathbb{P}(N \leq n, Z_n \leq \omega) > 0$, ya que $Z_n \sim N(n\mu, n\sigma^2)$, lo cual puede despreciarse, tal que $\mathbb{P}(N \leq n) \approx \mathbb{P}(Z_n > \omega)$. Así,

$$\mathbb{P}(N \leq n) \approx \mathbb{P}\left(\frac{Z_n - n\mu}{\sigma\sqrt{n}} > \frac{\omega - n\mu}{\sigma\sqrt{n}}\right) = \Phi\left(\frac{\sqrt{\omega\mu}}{\sigma} \left[\sqrt{\frac{n}{\omega/\mu}} - \sqrt{\frac{\omega/\mu}{n}}\right]\right), \quad (2)$$

donde $\Phi(\cdot)$ es la fda $N(0,1)$.

Génesis estadística (cont.)

Birnbaum & Saunders (1969a) usaron Ecuación (2) para definir una distribución estadística continua idealizando la v.a.d. N mediante la v.a.c. T , es decir, el número de ciclos hasta la falla fue reemplazado por el tiempo hasta que la falla ocurre. Así, tomando

$$\alpha = \frac{\sigma}{\sqrt{\omega\mu}}, \quad \beta = \frac{\omega}{\mu}, \quad \text{y} \quad a_t(\alpha, \beta) = \frac{1}{\alpha} \left[\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}} \right],$$

se obtiene $F_T(t) = \Phi(a_t(\alpha, \beta)); \quad t > 0, \alpha >, \beta > 0,$

la fda BS para el tiempo de vida de la fatiga con parámetros de forma α y de escala β . Esto significa que

$$Z = \frac{1}{\alpha} \left[\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}} \right] \sim N(0, 1).$$

VARIABLES DE VIDA

Por lo general, en la literatura relacionada a métodos y modelos estadísticos para datos de tiempos de vida se tiene que:

Variable de vida es una v.a. (T) continua y positiva ($T > 0$).

Una variable de vida puede representar, por ejemplo, las horas de funcionamiento de un motor. Sin embargo, una variable de vida también puede medir:

- Número de ciclos de carga de una batería hasta fallar.
- Cantidad de kilómetros recorridos por un carro hasta fallar.
- Resistencia de un material hasta su ruptura.
- Cantidad de agua caída.
- Concentración de un contaminante.
- Entre otras.

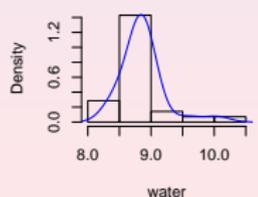
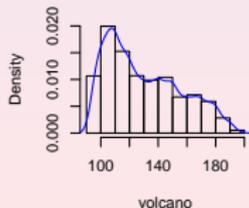
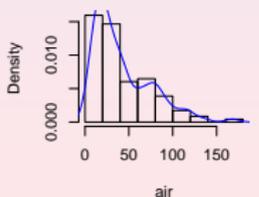
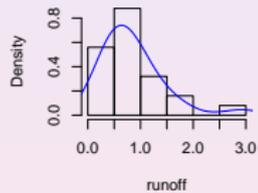
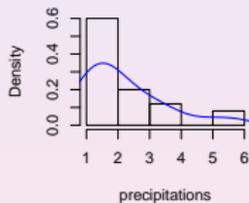
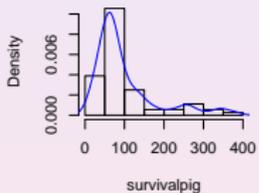
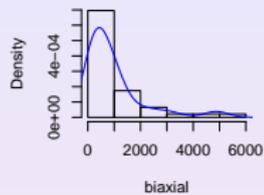
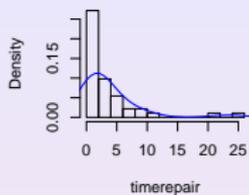
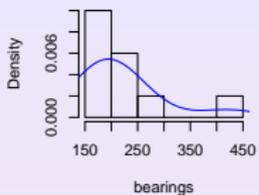
Datos de tiempos de vida

A continuación se discuten algunos conjuntos de datos no negativos implementados en los paquetes `bs` y `gbs` llamados:

- `psi21`, `psi26`, `psi31` (Birnbaum & Saunders, 1969b).
- `bearings` (McCool, 1974).
- `precipita`, `runoff` (Folks & Chikkara, 1978).
- `biaxial` (Rieck, 1989).
- `timerepair` (Hsieh, 1990).
- `survivalpig` (Kundu, Kannan & Balakrishnan, 2008).
- `air` (Leiva, Barros, Paula & Sanhueza, 2008).
- `water` (Leiva, Sanhueza & Angulo, 2009).

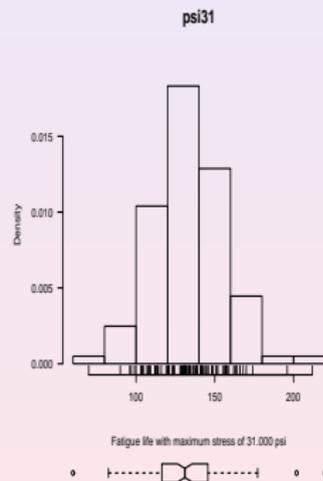
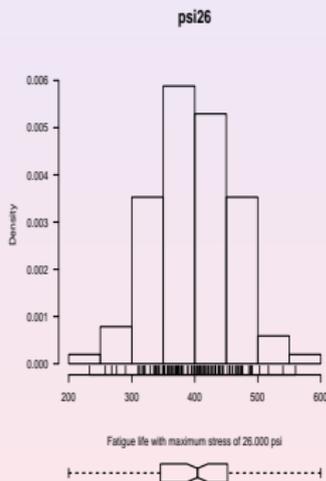
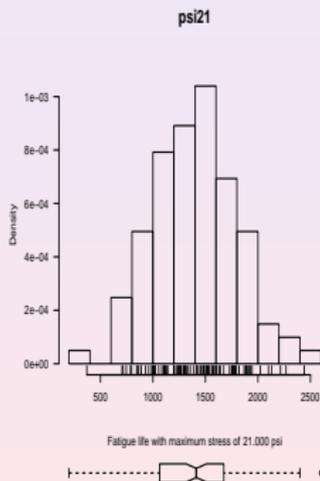
Datos de tiempos de vida

Histogramas y estimación kernel mediante R para algunos datos:



Datos de tiempos de vida

Histogramas producidos por el paquete gbs para datos de Birnbaum-Saunders (1969b):



Codigos de los gráficos

Gráfico 1:

```
require(bs)
data(timerepair)
par(mfrow = c(3, 3))
hist(timerepair, breaks = 10, probability = TRUE)
lifetime <- density(timerepair, bw = bcv)
lines(lifetime, col = 4)
```

Gráfico 2:

```
require(bs)
data(psi31)
histgbs(psi31, mainTitle = psi31,
        xLabel = Fatigue life with stress of 31 psi,
        colourHistogram = 0, colourBoxPlot = 0)
```

Distribución de vida

Una distribución de vida (life distribution) es un modelo de probabilidad asociado con una variable de vida; ver Marshall & Olkin (2007).

Una distribución de vida es en general:

- Asimétrica.
- Unimodal.
- Positivamente sesgada.
- De dos parámetros.
- Poseedora de momentos de cualquier orden.

Distribución de vida

Entre las distribuciones de vida más conocidas están:

- La distribución exponencial.
- La distribución lognormal.
- La distribución Weibull.
- La distribución gamma.
- La distribución Birnbaum-Saunders.
- La distribución Gaussiana inversa.

Seleccionando una distribución (cont.)

La mayoría de los modelos probabilísticos empleados para describir datos de tiempos de vida pueden elegirse con base en:

- I. Un argumento teórico (físico) para el mecanismo de falla.
- II. Un modelo que haya sido usado previamente con éxito.

Sin embargo, cualquiera sea la justificación para elegir una o varias leyes de probabilidad como candidatas, la distribución de vida que se escogerá para resolver el problema debe ser un modelo lógico que haya superado pruebas de bondad de ajuste y criterios de selección de modelos.

Seleccionando una distribución (cont.)

Ejemplo de argumentos teóricos:

- “Desgaste de tipo fase”, con tiempos exponenciales para cada fase, justifica una distribución gamma o Erlang.
- Tipo “valores extremos” justifica el uso de una distribución Weibull.
- “Degradación multiplicativa” justifica el uso de una distribución lognormal.
- “Tiempos de primer paso” justifica el uso de la distribución Gaussiana inversa.
- “Fatiga” o “daño acumulativo” justifica la distribución BS.

Descriptores de distribuciones de vida

Una variable de vida T con fda $F_T(t)$ posee:

Densidad:
$$f_T(t) = \frac{d}{dt} F_T(t), t > 0.$$

Función de confiabilidad:
$$R_T(t) = P(T > t) = 1 - F_T(t), t > 0,$$

Tasa de fallas:
$$h_T(t) = \frac{f_T(t)}{R_T(t)}, t > 0, R_T(t) > 0,$$

Tasa de fallas promedio:
$$FRA(t) = \frac{1}{t} \int_0^t h_T(u) du, t > 0,$$

Confiabilidad condicionada:
$$R_T(t/x) = \frac{R_T(t+x)}{R_T(x)}, t > 0, x > 0, R_T(x) > 0.$$

Tiempo medio hasta la falla:
$$\mu = \mathbb{E}[T] = \int_0^\infty R_T(t) dt.$$

Vida media residual:
$$\mu_x = \mathbb{E}[T|T > x] = \frac{\int_x^\infty R_T(y) dy}{R_T(x)}, x > 0, R_T(x) > 0.$$

Tipos de datos de vida

Tipos muestreos:

- Completo o no censurado.
- Truncado.
- Censurado.
 - Tipo I o por tiempo.
 - Tipo II o por fracción de fallas.
 - Tipo III (mezcla de tipo I y tipo II).
 - Tipo IV o aleatoria.

Tipos de datos censurados

- Por la izquierda: $\mathbb{P}(T < t_i)$, para $i \in L$.
- Por intervalos: $\mathbb{P}(t_{i-1} < T < t_i)$, para $i \in I$.
- Por la derecha: $\mathbb{P}(T > t_i)$, para $i \in C$.
- Falla: $f_T(t_i)$, para $i \in D$.

Suponiendo observaciones independientes, la función de verosimilitud se construye de la siguiente forma:

$$L(\boldsymbol{\theta}) = \left[\prod_{i \in L} F(t_i) \right] \left[\prod_{i \in I} [F(t_i) - F(t_{i-1})] \right] \left[\prod_{i \in C} R(t_i) \right] \left[\prod_{i \in D} f_T(t_i) \right].$$

LA DISTRIBUCIÓN BIRNBAUM-SAUNDERS (BS)

Distribuciones de vida de la fatiga

En este curso, las distribuciones de vida se centrarán en aquellas que modelan vida para fatiga (fatigue life distributions).

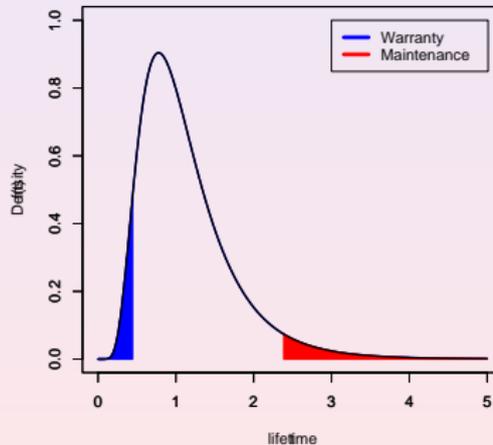
Cuatro “fatigue life distributions” que pueden ser generadas usando argumentos físicos son:

- La distribución BS.
- La distribución gamma.
- La distribución Gaussiana inversa.
- La distribución lognormal.

Estos modelos se ajustan bien en el centro de la distribución de vida, pero los extremos (colas) de la distribución ponen en duda la decisión de seleccionar alguna de ellas.

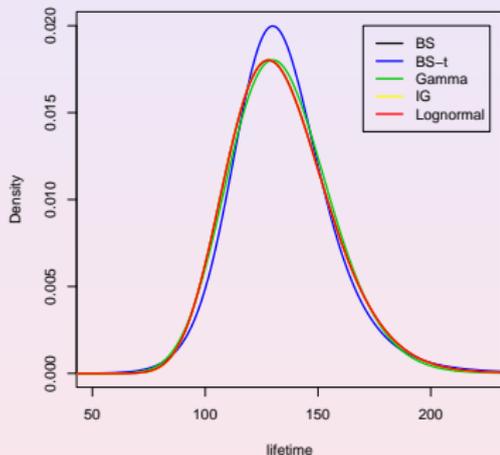
Seleccionando una distribución de vida de la fatiga

Los analistas generalmente están interesados en los percentiles bajos o altos de la distribución de vida (por ejemplo, debido a problemas de mantenimiento o de garantía, dosis letales o emergencia ambientales).

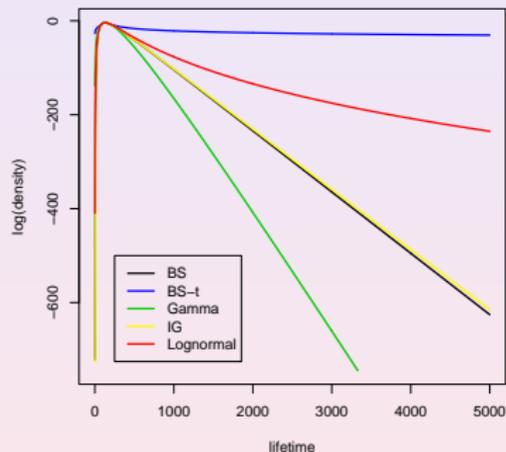


Seleccionando una distribución de vida de la fatiga

Figura 2.1 compara las distribuciones de vida de la fatiga BS, gamma, IG y LN mediante los datos [psi31](#).



(a) fdp



(b) log(fdp)

Figura 2.1: gráficos para los modelos indicados usando [psi31](#).

Motivación y nuevas aplicaciones del modelo BS

Medicina: varios tipos de cáncer y enfermedades cardíacas crónicas.

Calidad del aire: efecto acumulativo de contaminantes de la atmósfera.

Calidad del agua: efecto acumulativo de nutrientes vegetales.

Neurociencias: producción de spikes en neuronas.

Salud pública: Mortalidad humana asociada con diversos factores.

Forestación: incidencia del diámetro de un árbol en su mortalidad.

La distribución Birnbaum-Saunders

Existen dos forma de generar la distribución BS: uno físico (aproximado) y otro matemático (exacto).

El modelo BS corresponde a la distribución de la v.a.

$$T = \beta \left[\frac{\alpha Z}{2} + \sqrt{\frac{\alpha^2 Z^2}{2} + 1} \right]^2,$$

donde

$$Z = \frac{1}{\alpha} \left[\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}} \right] \sim N(0, 1).$$

$\alpha > 0$ es el parámetro de forma y $\beta > 0$ es el parámetro de escala y la mediana de la distribución. Esto es denotado por $T \sim BS(\alpha, \beta)$.

La distribución Birnbaum-Saunders

Densidad BS:

$$f_T(t) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{1}{2\alpha^2} \left[\frac{t}{\beta} + \frac{\beta}{t} - 2\right]\right) \frac{t^{-\frac{3}{2}} [t+\beta]}{2\alpha\sqrt{\beta}}, t > 0.$$

Función de distribución BS:

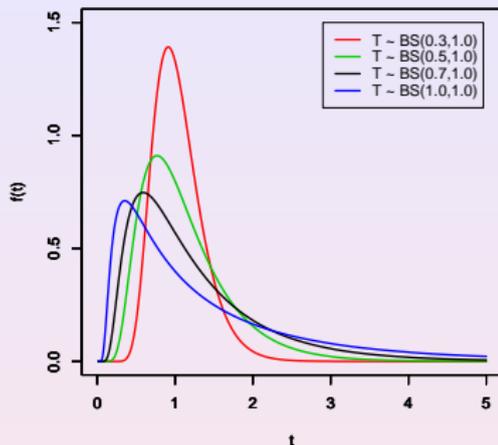
$$F_T(t) = \Phi\left(\frac{1}{\alpha} \left[\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}}\right]\right), t > 0.$$

Función cuantil BS:

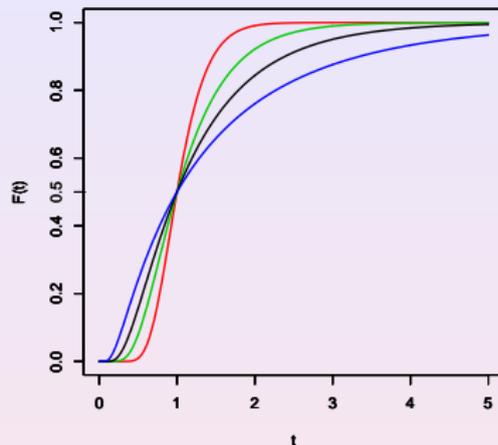
$$t(q) = \frac{\beta}{4} \left[\alpha z_q + \sqrt{\alpha^2 z_q^2 + 4} \right]^2$$

donde z_q es el cuantil q -ésimo $N(0, 1)$. Si $q = 0,5$, entonces $t(0,5) = \beta$, de este modo, el parámetro β es la mediana.

La distribución Birnbaum-Saunders



(a) fdp



(b) fda

Figura 2.1: gráficos de la fdp y fda de la distribución BS.

La distribución Birnbaum-Saunders

Propiedades BS:

- 1 $cT \sim \text{BS}(\alpha, c\beta)$, con $c > 0$.
- 2 $1/T \sim \text{BS}(\alpha, 1/\beta)$.

Propiedad 1 indica que la distribución BS pertenece a la familia de escala.

Propiedad 2 indica que la distribución BS está en la familia cerrada bajo recíprocos (Saunders, 1974).

Estas dos propiedades son claves para el modelamiento estadístico y la estimación de momentos basados en la distribución BS.

La distribución Birnbaum-Saunders

Momentos BS:

$$\mathbb{E}[T^n] = \beta^n \sum_{j=0}^n \binom{2n}{2j} \sum_{i=0}^j \binom{j}{i} \frac{(2n - 2j + 2i)!}{2^{n-j+i} (n - j + i)!} \left[\frac{\alpha}{2}\right]^{2n-2j+2i}.$$

Media y varianza BS:

$$\mathbb{E}[T] = \beta \left[1 + \frac{\alpha^2}{2}\right] \quad \text{y} \quad \text{Var}[T] = \beta^2 \alpha^2 \left[1 + \frac{5}{4}\alpha^2\right].$$

La distribución Birnbaum-Saunders

Confiabilidad y tasa de fallas BS:

$$R_T(t) = \Phi(-a_t(\alpha, \beta)) \quad \text{y}$$

$$h_T(t) = \frac{f_T(t)}{R_T(t)} = \frac{\phi(a_t(\alpha, \beta)) \frac{d}{dt} a_t(\alpha, \beta)}{\Phi(-a_t(\alpha, \beta))}.$$

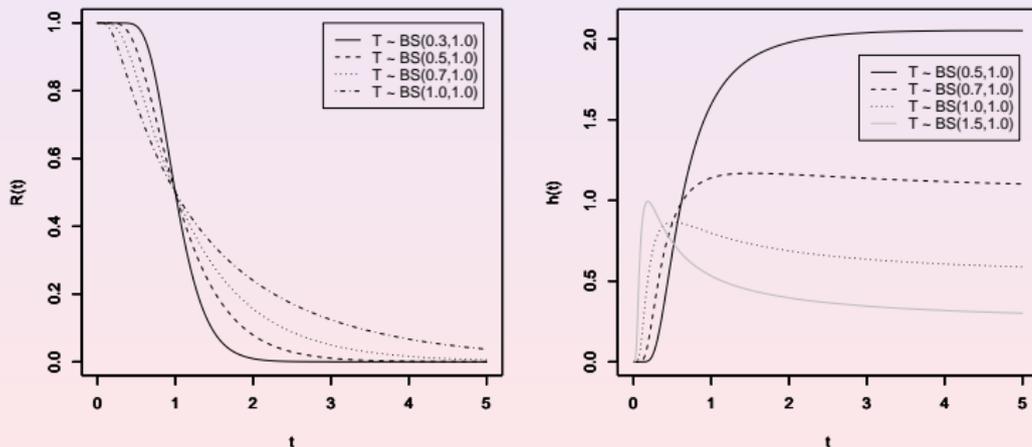


Figura 2.1: funciones de confiabilidad y riesgo de una distribución BS.

La distribución Birnbaum-Saunders

Estimación de ML BS:

$$\hat{\alpha} = \sqrt{\frac{S}{\hat{\beta}} + \frac{\hat{\beta}}{R}} - 2 \quad \text{y} \quad \hat{\beta}^2 - \hat{\beta} [k(\hat{\beta}) + 2R] + R [k(\hat{\beta}) + S] = 0.$$

Estimación de momentos BS:

$$\tilde{\alpha} = \sqrt{2 \left[\sqrt{\frac{S}{R}} - 1 \right]} \quad \text{y} \quad \tilde{\beta} = \sqrt{SR},$$

donde

$$S = \frac{1}{n} \sum_{i=1}^n t_i, \quad R = \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{t_i}} \quad \text{y} \quad k(x) = \frac{1}{\frac{1}{n} \sum_{i=1}^n \frac{1}{x+t_i}}.$$

La distribución Birnbaum-Saunders

Estimación de MC BS:

$$F_T(t) = \Phi \left(\frac{1}{\alpha} \left[\sqrt{\frac{t}{\beta}} - \sqrt{\frac{\beta}{t}} \right] \right) \Rightarrow \underbrace{t}_{y} = \underbrace{\beta}_{b} + \underbrace{\alpha\sqrt{\beta}}_a \underbrace{\sqrt{t}\Phi^{-1}(F_T(t))}_x,$$

donde Φ^{-1} es fda inversa $N(0, 1)$. Así, ya que $y \approx b + ax$, graficando $x_i = \bar{p}_i$ vs. $y_i = t_i$, se tiene aproximadamente una línea recta de datos BS, y $\bar{F}(t_i) = [i - 0.3]/[n + 0.4]$. Bondad de ajuste puede basarse en el R^2 de regresión cuando se ajusta $[\bar{p}_i, t_i]$, para $i = 1, \dots, n$. Finalmente, α y β pueden estimarse usando el método de mínimos cuadrados mediante

$$\bar{\alpha} = \frac{\bar{b}}{\sqrt{\bar{a}}} \quad \text{y} \quad \bar{\beta} = \bar{a}.$$

Descarga e instalación del software R y del paquete bs

The R Project for Statistical Computing - Windows Internet Explorer

http://www.r-project.org/

File Edit View Favorites Tools Help

Convert Select

The R Project for Statistical Computing

The R Project for Statistical Computing



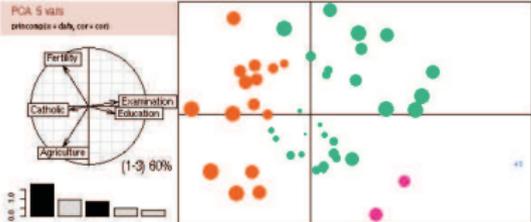
About R
[What is R?](#)
[Contributors](#)
[Screenshots](#)
[What's new?](#)

Download
[CRAN](#)

R Project
[Foundation](#)
[Members & Donors](#)
[Mailing Lists](#)
[Bug Tracking](#)
[Developer Page](#)
[Conferences](#)
[Search](#)

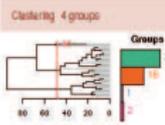
Documentation
[Manuals](#)
[FAQs](#)
[Newsletter](#)
[Wiki](#)
[Books](#)
[Certification](#)

PCA 5 vars
 phtconstr + daly, cor + conr

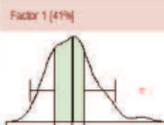


(1-3) 80%

Clustering 4 groups



Factor 1 (41%)



Factor 3 (19%)



Getting Started:

- R is a free software environment for statistical computing and graphics. It compiles and runs on a wide variety of UNIX platforms, Windows and MacOS. To [download R](#), please choose your preferred [CRAN mirror](#).
- If you have questions about R like how to download and install the software, or what the license terms are, please read our [answers to frequently asked](#)

http://cran.r-project.org/mirrors.html

Internet 100%

7:50 PM

Descarga e instalación del software R y del paquete bs

The R Project for Statistical Computing - Windows Internet Explorer

http://www.r-project.org/

File Edit View Favorites Tools Help

Convert Select

The R Project for Statistical Computing

CRAN Mirrors

The Comprehensive R Archive Network is available at the following URLs, please choose a location close to you. Some statistics on the status of the mirrors can be found [here](#).

Argentina	http://mirror.cricvt.edu.ar/r/	CONICET, Mendoza
Australia	http://cran.ms.unimelb.edu.au/	University of Melbourne
Austria	http://cran.at-r-project.org/	Wirtschaftsuniversität Wien
Belgium	http://www.freestatistics.org/cran/	K.U. Leuven Association
Brazil	http://cran.br-r-project.org/	Universidade Federal do Paraná
	http://cran.fiocruz.br/	Oswaldo Cruz Foundation, Rio de Janeiro
	http://www.sps.fmvz.usp.br/CRAN/	University of Sao Paulo, Sao Paulo
	http://brieger.esalq.usp.br/CRAN/	University of Sao Paulo, Piracicaba
Canada	http://cran.star.sfu.ca/	Simon Fraser University, Burnaby
	http://probability.ca/cran/	University of Toronto
Chile	http://drichlet.mat.puc.cl/	Pontificia Universidad Católica de Chile, Santiago
China	http://mirrors.geoexpat.com/cran/	GeoExpat.Com
Croatia		

Start

Internet Explorer slides Adobe WinEdt Tinn-R Yap Microsoft Excel CoreDRAW R1-GS-view 7:51 PM

Descarga e instalación del software R y del paquete bs

The screenshot shows a Windows Internet Explorer browser window displaying the Comprehensive R Archive Network (CRAN) website. The address bar shows the URL <http://www.ups.fmz.usp.br/CRAN/>. The page title is "The Comprehensive R Archive Network".

The main content area is titled "Frequently used pages" and contains two main sections:

Download and Install R

Precompiled binary distributions of the base system and contributed packages, **Windows and Mac** users most likely want one of these versions of R:

- [Linux](#)
- [MacOS X](#)
- [Windows](#)

Source Code for all Platforms

Windows and Mac users most likely want the precompiled binaries listed in the upper box, not the source code. The sources have to be compiled before you can use them. If you do not know what this means, you probably do not want to do it!

- **The latest release** (2008-12-22): [R-2.8.1.tar.gz](#) (read [what's new](#) in the latest version).
- Sources of [R alpha and beta releases](#) (daily snapshots, created only in time periods before a planned release).
- Daily snapshots of current patched and development versions are [available here](#). Please read about [new features and bug fixes](#) before filing corresponding feature requests or bug reports.
- Source code of older versions of R is [available here](#).
- Contributed extension [packages](#)

The left sidebar contains navigation links for CRAN, Mirrors, What's new?, Task Views, Search, About R, R Homepage, Software, R Sources, R Binaries, Packages, Other, Documentation, Manuals, FAQs, Contributed, and Newsletter.

The taskbar at the bottom shows several open applications: Internet Explorer, slides, Adobe..., WinEdit..., Tinn-R, Yab - [course], Microsoft E..., CoreDRAW..., R1 - GS-view, and the system clock shows 7:52 PM.

Descarga e instalación del software R y del paquete bs

The screenshot shows a web browser window titled "The Comprehensive R Archive Network - Windows Internet Explorer". The address bar shows the URL <http://www.sps.fhmz.usp.br/CRAN/>. The page content is titled "R for Windows" and includes the following text:

This directory contains binaries for a base distribution and packages to run on i386/x64 Windows.

Note: CRAN does not have Windows systems and cannot check these binaries for viruses. Use the normal precautions with downloaded executables.

Subdirectories:

- [base](#) Binaries for base distribution (managed by Duncan Murdoch)
- [contrib](#) Binaries of contributed packages (managed by Uwe Ligges)

Please do not submit binaries to CRAN. Package developers might want to contact Duncan Murdoch or Uwe Ligges directly in case of questions / suggestions related to Windows binaries.

You may also want to read the [R FAQ](#) and [R for Windows FAQ](#).

Last modified: April 4, 2004, by Friedrich Leisch

The left sidebar contains navigation links for CRAN, Mirrors, What's new?, Task Views, Search, About R, R Homepage, Software, R Sources, R Binaries, Packages, Other, and Documentation (Manuals, FAQs, Contributed, Newsletter).

The browser's taskbar at the bottom shows several open applications: Internet Explorer, slides, Adobe..., WinEdt - [...], Tinn-R, Yab - [course], Microsoft E..., CoreDRAW..., R1 - GS-view, and a system tray with the time 7:34 PM.

Descarga e instalación del software R y del paquete bs

The screenshot shows a Windows Internet Explorer browser window displaying the 'R-2.8.1 for Windows' download page from the Comprehensive R Archive Network. The browser's address bar shows the URL <http://www.vps.fmvz.usp.br/CRAN/>. The page content includes the R logo, navigation links for CRAN (Mirrors, What's new?, Task Views, Search, About R, R Homepage), Software (R Sources, R Binaries, Packages, Other), and Documentation (Manuals, FAQs, Contributed, Newsletter). A prominent section titled 'Download R 2.8.1 for Windows (34 megabytes)' is visible, along with a progress indicator showing '0% of R-2.8.1-win32.exe from www.vps.fmvz.usp.br'. A 'File Download - Security Warning' dialog box is open in the foreground, asking 'Do you want to run or save this file?' for 'R-2.8.1-win32.exe'. The dialog box provides details: Name: R-2.8.1-win32.exe, Type: Application, 31.2MB, From: www.vps.fmvz.usp.br. It includes 'Run', 'Save', and 'Cancel' buttons. A warning icon and text state: 'While files from the Internet can be useful, this file type can potentially harm your computer. If you do not trust the source, do not run or save this software. [View details here.](#)' The Windows taskbar at the bottom shows the Start button, several open applications (Internet Explorer, slides, Adobe, WinEdt, Ten-4, Yap - [course], Microsoft E...), and the system tray with the date and time (7:55 PM).

Descarga e instalación del software R y del paquete bs

The screenshot shows a Windows Internet Explorer browser window displaying the CRAN website. The address bar shows the URL <http://www.sps.fernand.usp.br/CRAN/>. The page title is "The Comprehensive R Archive Network - Windows Internet Explorer". The main content area is titled "Contributed Packages" and contains the following sections:

- Installation of Packages**: A paragraph explaining that users should type `help("INSTALL")` or `help("install.packages")` in R for installation instructions. It also mentions the manual [R Installation and Administration](#).
- CRAN Task Views**: A paragraph stating that these allow browsing packages by topic and provide tools for automatic installation. It notes that currently, 23 views are available.
- Daily Package Check Results**: A paragraph stating that all packages are tested regularly on machines running [Debian GNU Linux](#). It also mentions checks for MacOS X and Windows.
- Writing Your Own Packages**: A paragraph mentioning the manual [Writing R Extensions](#).
- Available Bundles and Packages**: A paragraph stating that the CRAN package repository features 1678 objects, including 1671 packages and 7 bundles containing 26 packages, for a total of 1697 available packages.

At the bottom of the page, there is a navigation bar with the letters [A](#) through [Z](#). The browser's taskbar at the bottom shows the Start button, several open applications (Internet Explorer, slides, Adobe, WinEdt, Tim-R, Yip - [course], Microsoft E..., CoreDRAW..., R1 - GS-view), and the system clock showing 7:55 PM on 10/2/2009.

Descarga e instalación del software R y del paquete bs

The screenshot shows a Windows Internet Explorer browser window displaying the Comprehensive R Archive Network (CRAN) website. The address bar shows the URL <http://www.r-project.org/CRAN/>. The page features a navigation menu on the left with categories like CRAN, Mirrors, What's new?, Task Views, Search, About R, R Homepage, Software, Documentation, Manuals, FAQs, Contributed, and Newsletter. The main content area lists several R packages, with the package 'bs' highlighted in bold. The description for 'bs' is: "Package for Generalized Birnbaum-Saunders Distributions". Other packages listed include 'gamlss.cens', 'gamlss.clust', 'gamlss.mv', 'gamlss.nl', 'gamlss.tx', 'gap', 'gbet', 'gbm', 'gcExplorer', 'gcl', 'gclus', 'gcreec', 'gdata', 'gee', 'geepack', 'geiger', 'genalg', 'gene2pathway', 'genetics', 'geoR', 'geoRglm', 'geomandata', 'geometry', 'geonames', and 'geozoo'.

Descarga e instalación del software R y del paquete bs

The Comprehensive R Archive Network - Windows Internet Explorer

http://www.r-project.org/CRAN/

File Edit View Favorites Tools Help

Convert Select

The Comprehensive R Archive Network



gbs: Package for Generalized Birnbaum-Saunders Distributions

A collection of utilities for analyzing censored and uncensored data from generalized Birnbaum-Saunders distributions

Version: 1.0
 Depends: R (≥ 1.8.0), nlme
 Date: 2008-05-05
 Author: Michelli Barros, Victor Leiva and Gilberto A. Paula
 Maintainer: Victor Leiva <victor.leiva@uv.cl>
 License: GPL
 URL: <http://www.r-project.org>, <http://www.dew.cl/leiva>
 In views: [Distributions](#), [Survival](#)
 CRAN checks: [gbs results](#)

Downloads:

Package source: [gbs_1.0.tar.gz](#)
 Mac OS X binary: [gbs_1.0.tgz](#)
 Windows binary: [gbs_1.0.zip](#)
 Reference manual: [gbs.pdf](#)

CRAN
[Mirrors](#)
[What's new?](#)
[Task Views](#)
[Search](#)

About R
[R Homepage](#)

Software
[R Sources](#)
[R Binaries](#)
[Packages](#)
[Other](#)

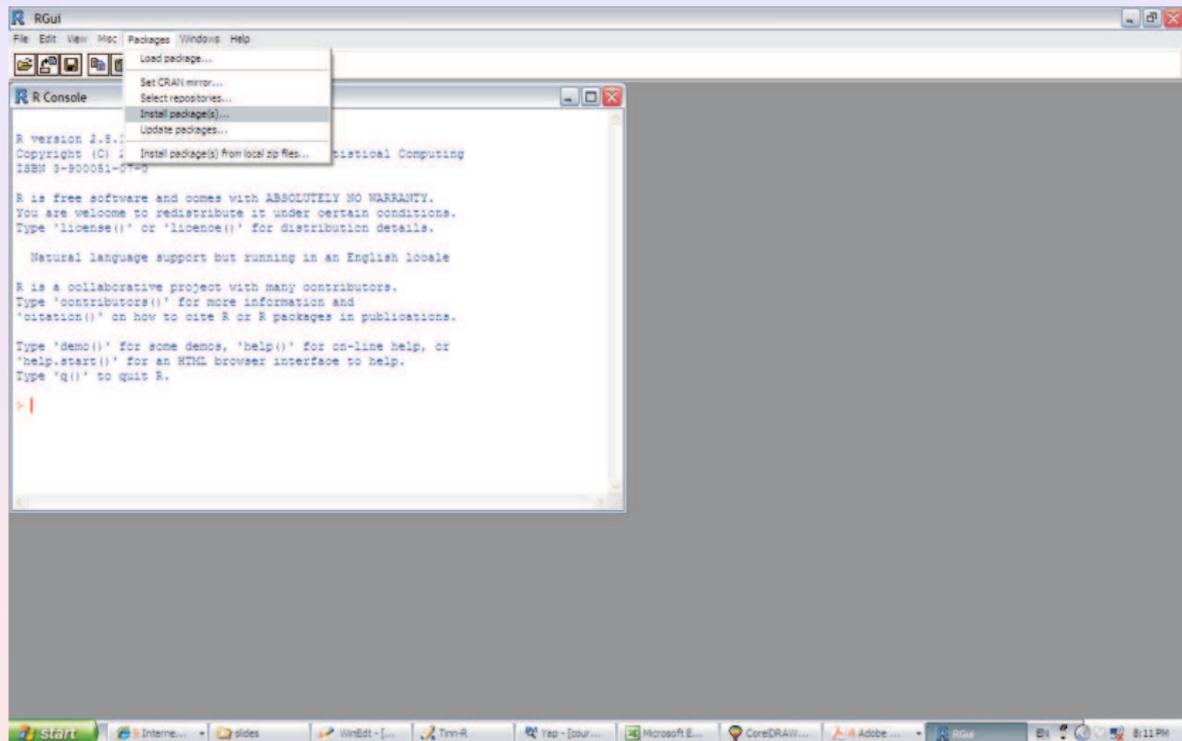
Documentation
[Manuals](#)
[FAQs](#)
[Contributed](#)
[Newsletter](#)

Done

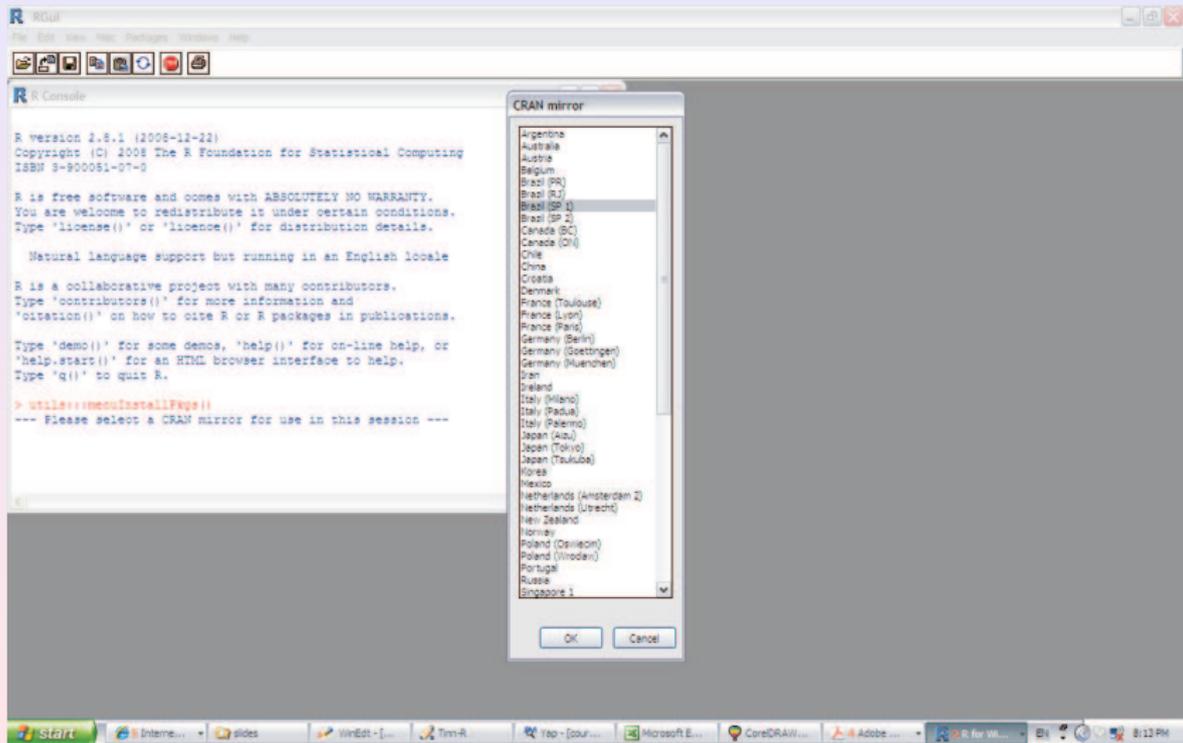
Internet 100%

Start Internet... slides Adobe... WinEdt - [...] Tinn-R YAP - [course] Microsoft Excel CoreDRAW... R1 - GSview 7:57 PM

Descarga e instalación del software R y del paquete bs



Descarga e instalación del software R y del paquete bs



Descarga e instalación del software R y del paquete bs

The screenshot shows the R GUI interface. The R Console window displays the following text:

```
R version 2.8.1 (2008-12-22)
Copyright (C) 2008 The R Foundation for Statistical Computing
ISBN 3-900051-07-0

R is free software and comes with ABSOLUTELY NO WARRANTY.
You are welcome to redistribute it under certain conditions.
Type 'license()' or 'licence()' for distribution details.

Natural language support but running in an English locale

R is a collaborative project with many contributors.
Type 'contributors()' for more information and
'citation()' on how to cite R or R packages in publications.

Type 'demo()' for some demos, 'help()' for on-line help, or
'help.start()' for an HTML browser interface to help.
Type 'q()' to quit R.

> chooseCRANmirror()
> utils:::menuInstallPage()
```

The Packages dialog box is open, showing a list of installed and available packages. The list includes:

- bs
- FFCDIHS
- fronpar
- Frading
- fr
- FunCluster
- FunRoots
- FunNet
- FunNet
- FuzzyPCR
- FuzzyOP
- FuzzyRaniTests
- FunRegime
- g.data
- GIDBI
- gefit
- gem
- gemar
- GMBoost
- gemisc
- gemisc.cens
- gemisc.dist
- gemisc.mix
- gemisc.ml
- gemisc.r
- gfo
- gfoe
- gfoe
- gfm
- gse
- g:Explorer
- gc
- gcus
- gconvac
- gdata
- GDD
- gee
- gesack
- geiger
- GMABEL

The dialog box has "OK" and "Cancel" buttons at the bottom.

El paquete bs

El paquete bs contiene las funciones probabilísticas:

```
> require(bs)
> dbs(3, alpha = 0.5, beta = 1.0, log = FALSE)
[1] 0.02133878
> pbs(1, alpha = 0.5, beta = 1.0)
[1] 0.5
> qbs(0.5, alpha = 0.5, beta = 2.5)
[1] 2.5
> rbs1(6, alpha = 0.5, beta = 1.0)
[1] 0.6138 1.1142 0.8761 1.2089 1.8756 1.9979
```

descriptores de distribuciones de vida, 3 generadores de números aleatorios, 3 métodos de estimación (ML, momentos y mínimos cuadrados) y varios métodos de bondad de ajuste (PP, QQ, KS y cartas de ajuste basadas en momentos) y criterios de selección de modelos basados en pérdida de información (AIC, BIC y HQ).

Funciones del paquete bs

La instrucción `> simulBsMle(100,0.5,1.0)` simula 3 muestras de $n = 100$ desde $T \sim BS(\alpha = 0,5, \beta = 1,0)$, una para cada generador implementado, calcula los estimadores ML de α y β , y lleva a cabo bondad de ajuste para cada muestra. La siguiente table resumen los resultados de este estudio.

	MLE(alpha)	MLE(beta)	KS	KS p-value
Sample1	0.66456	1.06143	0.06701	0.7603
Sample2	0.43454	0.93341	0.09310	0.3514
Sample3	0.60549	0.86124	0.08475	0.4692)

Funciones del paquete bs (cont.)

La instrucción `estimates <- est1bs(psi31)` calcula el estimadores ML de α y β y usa el estimador media-media como valor de partida para β obteniendo en este caso:

```
> betaStartingValue
[1] 131.8193
> alpha
[1] 0.1703847
> beta
[1] 131.8188
> converge
[1] TRUE
> iteration
[1] 2
```

El comando `indicatorsbs()` calcula además de los estimadores ML de α y β , los estimadores del CV, CS y CC del modelo BS.

Funciones del paquete bs (cont.)

La instrucción: `gmbps(psi31)` estima α y β y construye a gráfico tipo PP, donde $\hat{\alpha} = 0,1686$, $\hat{\beta} = 131,9224$ y $R^2 = 0,9775$. Figura 4 muestra este gráfico para `psi31`.

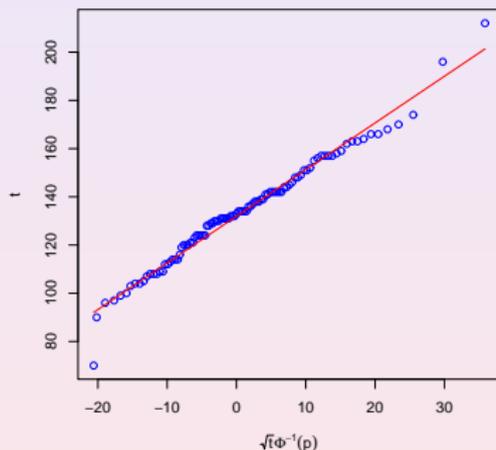


Figura 2.1: gráfico tipo PP for la BS basado en psi31.

Funciones del paquete bs (cont.)

12 muestras fueron generadas para $\alpha = 0,2(0,2)2,4$, las cuales pueden representar, por ejemplo, datos diarios para cada mes durante un año. Para cada muestra, estimaciones de los coeficiente de sesgo y curtosis, β_1 y β_2 , respectivamente, fueron obtenidas y guardadas en los vectores x e y . Las instrucciones:

```
plot(x, y,  
      xlab = expression(beta*1)  
      ylab = expression(beta*2), col = 4)  
fitbetabs(0.2, 2)  
fitbetabs(2.4, 3)
```

permiten obtener cartas de ajuste que son mostradas en Figura 3.1.

Funciones del paquete bs (cont.)

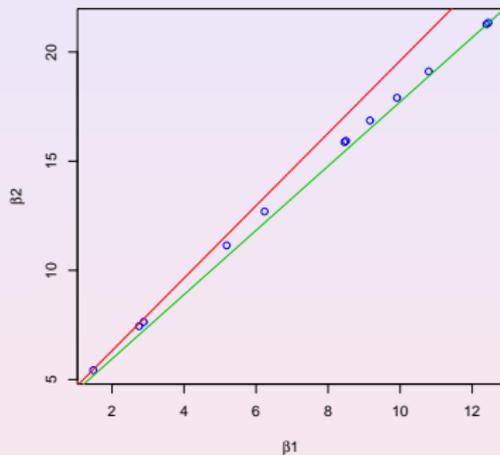


Figura 2.1: carta β_1 - β_2 for datos simulados.

LA DISTRIBUCIÓN BIRNBAUM-SAUNDERS GENERALIZADA (BSG)

Una motivación para la BSG

Para resolver el problema de ajuste en los percentiles bajos y altos de la distribución de vida es necesario considerar modelos de vida más flexibles (p.e. en curtosis) que los tradicionales.

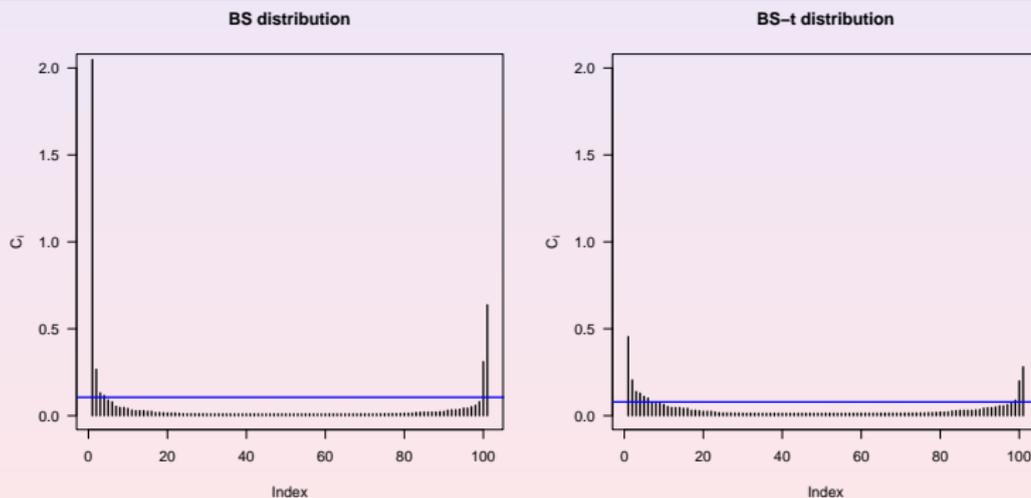


Figura 3.1: gráfico de influencia de observaciones atípicas.

Distribuciones de contornos elípticos

La familia de **distribuciones contornos elípticos** es una clase de distribuciones multivariadas cuyas densidades tienen la misma forma elíptica que la distribución normal. Sin embargo, ellas también incluyen distribuciones cuyas colas son más y menos pesadas que la distribución normal. Además, la distribución normal es un caso particular de esta familia.

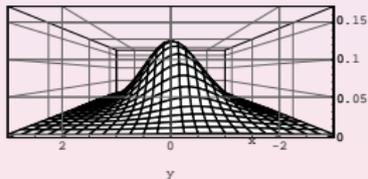


Figura 3.1: distribución normal bivariada.

Distribuciones de contornos elípticos

La familia de **distribuciones contornos elípticos** es una clase de distribuciones multivariadas cuyas densidades tienen la misma forma elíptica que la distribución normal. Sin embargo, ellas también incluyen distribuciones cuyas colas son más y menos pesadas que la distribución normal. Además, la distribución normal es un caso particular de esta familia.

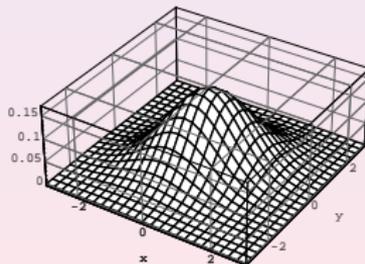


Figura 3.1: distribución normal bivariada.

Distribuciones de contornos elípticos

La familia de **distribuciones contornos elípticos** es una clase de distribuciones multivariadas cuyas densidades tienen la misma forma elíptica que la distribución normal. Sin embargo, ellas también incluyen distribuciones cuyas colas son más y menos pesadas que la distribución normal. Además, la distribución normal es un caso particular de esta familia.

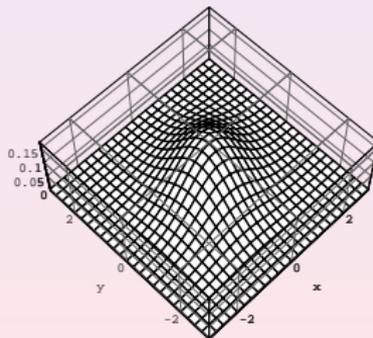


Figura 3.1: distribución normal bivariada.

Distribuciones de contornos elípticos

La familia de **distribuciones contornos elípticos** es una clase de distribuciones multivariadas cuyas densidades tienen la misma forma elíptica que la distribución normal. Sin embargo, ellas también incluyen distribuciones cuyas colas son más y menos pesadas que la distribución normal. Además, la distribución normal es un caso particular de esta familia.

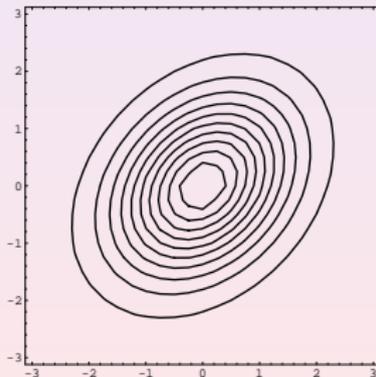


Figura 3.1: gráfico de contornos de una distribución normal bivariada.

Distribuciones de contornos elípticos

Para el caso unidimensional, las distribuciones elípticas corresponden a todas las distribuciones simétricas en \mathbb{R} , las que se caracterizan por sus parámetros:

- De posición, μ ,
- De escala, σ^2 , y
- Una función generadora de densidades o kernel, g .

La notación usada en este caso es: $X \sim \text{EC}(\mu, \sigma^2; g)$.

En este curso se consideran distribuciones esféricas en \mathbb{R} :

$$Z = \frac{X - \mu}{\sigma} \sim \text{EC}(0, 1; g) \equiv S(g).$$

Distribuciones simétricas

Sea $Z \sim S(g)$. Entonces,

Densidad simétrica

$$f_Z(z) = c g(z^2); \quad z \in \mathbb{R}, \quad (3)$$

donde $g(\cdot)$ es el kernel de la fdp de Z y c es una constante de normalización, tal que $\int_{-\infty}^{+\infty} g(z^2) dz = 1/c$.

Función de distribución simétrica

$$F_Z(z) = \mathbb{P}(Z \leq z) = 1 - F_Z(-z), \quad z \in \mathbb{R}.$$

la cual es útil para calcular los percentiles de la distribución de Z , obtener números aleatorios y realizar pruebas de bondad de ajuste.

Distribuciones simétricas

Tabla 4.1: kernel (g) y constantes de normalización (c) de $Z \sim S(g)$.

Distribución	Notación	c	$g = g(u), u > 0$
Tipo Kotz (KT)	$KT(q, r, s)$	$\frac{s r^{2q-1}}{\Gamma\left(\frac{2q-1}{2s}\right)}$	$u^{q-1} \exp(-ru^s), q > \frac{1}{2}, r, s > 0$
Pearson VII (PVII)	$PVII(q, r)$	$\frac{\Gamma(q)}{[r\pi]^{\frac{1}{2}} \Gamma(q-\frac{1}{2})}$	$\left[1 + \frac{u}{r}\right]^{-q}; q > \frac{1}{2}, r > 0$
Logística III (LIII)	$LIII(q)$	$\frac{\Gamma(2q)}{[\Gamma(q)]^2}$	$\frac{[\exp(\sqrt{u})]^q}{[1+\exp(\sqrt{u})]^{2q}}; q > 0$

Algunos casos particulares de las distribuciones especificadas en Tabla 4.1 son:

- La distribución logística es $LIII(q = 1)$.
- La distribución normal, $N(0, 1)$, es $KT(q = 1, r = \frac{1}{2}, s = 1)$.
- La distribución exponencial potencia, $EP(\nu)$, es $KT(q = 1, r = \frac{1}{2}, s = \nu)$, siendo $\nu > 0$ un parámetro de forma.
- La distribución t -Student con ν grados de libertad, $t(\nu)$, es $PVII(q = \frac{\nu+1}{2}, r = \nu)$, siendo $\nu > 0$ un parámetro de forma.

Distribución Birnbaum-Saunders generalizada

Basados en

$$Z = \frac{1}{\alpha} \left[\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}} \right] \sim N(0, 1).$$

Díaz-García y Leiva (2005) supusieron que

$$Z = \frac{1}{\alpha} \left[\sqrt{\frac{T}{\beta}} - \sqrt{\frac{\beta}{T}} \right] \sim S(g),$$

donde g es la función generadora de densidades de una distribución simétrica en \mathbb{R} . Con lo cual se consigue que

$$T = \beta \left[\frac{\alpha Z}{2} + \sqrt{\frac{\alpha^2 Z^2}{2} + 1} \right]^2 \sim \text{BSG}(\alpha, \beta; g), \quad \alpha > 0, \quad \beta > 0.$$

Densidad BSG

Sea $T \sim \text{BSG}(\alpha, \beta; g)$. Entonces,

Densidad BSG:

$$f_T(t) = c g \left(\frac{1}{\alpha^2} \left[\frac{t}{\beta} + \frac{\beta}{t} - 2 \right] \right) \frac{t^{-3/2} [t + \beta]}{2\alpha\beta^{1/2}}, \quad t > 0.$$

donde c y $g(\cdot)$ son la constante de normalización y kernel de la distribución $Z \sim S(g)$.

Observación:

La distribución BSG obtenida a partir de una distribución normal estándar coincide con la distribución Birnbaum-Saunders clásica.

Casos especiales BSG

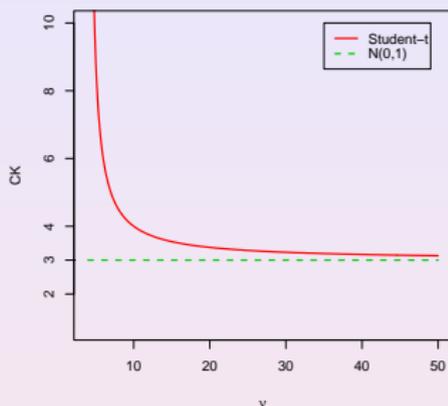
La distribución BSG puede ser obtenida desde los kernels:

- Cauchy.
- Laplace o exponencial doble.
- Logística.
- t -Student.
- Exponencial potencia.
- Normal (BS clásica).

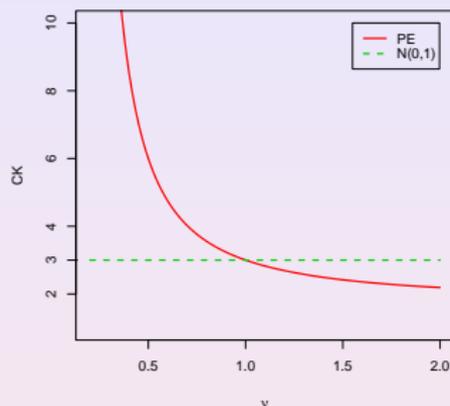
Motivación:

Esta extensión consigue hacer la curtosis de la distribución BS clásica flexible.

Curtosis BSG flexible



(a) t-Student



(b) Exponencial potencia

Figura 4.5: coeficiente de curtosis de las distribuciones t -Student y EP v /s normal estándar.

Dos casos BSG particulares

Densidad de una distribución BSG (caso t -Student):

$$f_T(t) = \frac{\Gamma\left(\frac{\nu+1}{2}\right)}{(\nu\pi)^{\frac{1}{2}}\Gamma\left(\frac{\nu}{2}\right)} \left[1 + \frac{1}{\nu\alpha^2} \left[\frac{t}{\beta} + \frac{\beta}{t} - 2\right]\right]^{-\frac{\nu+1}{2}} \frac{t^{-\frac{3}{2}}[t + \beta]}{2\alpha\beta^{\frac{1}{2}}}$$

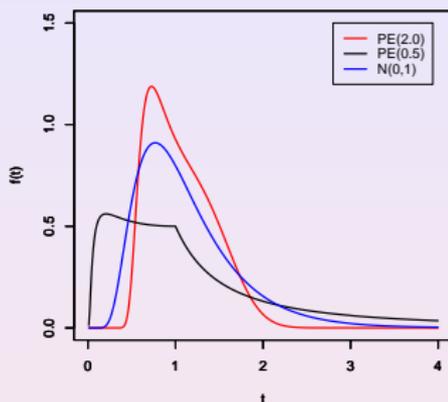
con $t > 0, \alpha > 0, \beta > 0$ y $\nu > 0$.

Densidad de una distribución BSG (caso EP):

$$f_T(t) = \frac{\nu}{2^{\frac{1}{2\nu}}\Gamma\left(\frac{1}{2\nu}\right)} \exp\left(-\frac{1}{2\alpha^{2\nu}} \left[\frac{t}{\beta} + \frac{\beta}{t} - 2\right]^\nu\right) \frac{t^{-\frac{3}{2}}[t + \beta]}{2\alpha\beta^{\frac{1}{2}}}$$

con $t > 0, \alpha > 0, \beta > 0$ y $\nu > 0$.

Dos casos BSG particulares



(a) EP

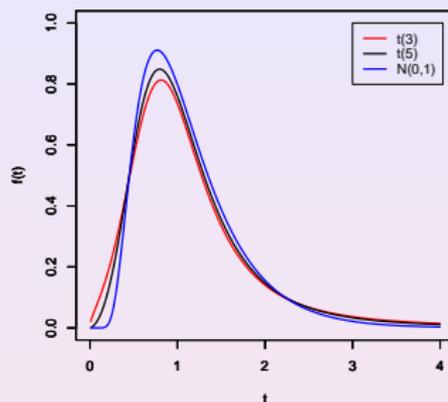
(b) t -Student

Figura 4.6: gráficos de densidades de la distribución BSG para $\alpha = 0,5$ y $\beta = 1,0$, obtenidas desde distribuciones EP y t -Student para los valores que se indican.

Propiedades de la distribución BSG

Sea $T \sim \text{BSG}(\alpha, \beta; g)$. Entonces,

Propiedades BSG:

- 1 $Z = [\sqrt{T/\beta} - \sqrt{T/\beta}]/\alpha \sim S(g)$.
- 2 $cT \sim \text{BSG}(\alpha, c\beta; g)$, con $c > 0$.
- 3 $1/T \sim \text{BSG}(\alpha, 1/\beta; g)$.
- 4 La fda de T es $F_T(t) = F_Z(a_t(\alpha, \beta))$, donde $F_Z(\cdot)$ es la fda de $Z \sim S(g)$ y $a_t(\alpha, \beta) = [\sqrt{t/\beta} - \sqrt{\beta/t}]/\alpha$.
- 5 La función cuantil de T es $t(q) = [\beta/4] [\alpha z_q + \sqrt{\alpha^2 z_q^2 + 4}]^2$, donde z_q es el cuantil q -ésimo de $Z \sim S(g)$. Aquí β es también la mediana .

Inferencia en la distribución BSG

Sea $T \sim \text{BSG}(\alpha, \beta; g)$. Entonces,

La estimación de los parámetros α y β de la distribución BSG se puede llevar a cabo mediante el método verosimilitud máxima usando métodos numéricos.

Para una cierta clase de distribuciones BSG, el algoritmo EM puede ser implementado.

Los parámetros adicionales de la distribución BSG (diferentes a α y β) pueden fijarse o ser obtenidos mediante alguna información proporcionada desde los datos.

Un paquete R llamado gbs

Barros, Paula & Leiva (2009) desarrollaron el paquete gbs disponible desde CRAN (<http://CRAN.R-project.org/>), el cual contiene las funciones probabilísticas:

```
> require(gbs)

> dgb(3, alpha = 1.0, beta = 1.0, nu = 1.0, kernel = normal)
[1] 0.07883672

> pgb(.75, alpha = 1.0, beta = 1.0, nu = 1.0, kernel = normal)
[1] 0.386415

> qgb(0.5, alpha = 1.0, beta = 1.0, nu = 1.0, kernel = normal)
[1] 1
```

indicadores de tiempos de vida, 1 generador de números aleatorios, 1 método de estimación para datos no censurados y otro para datos censurados y métodos de bondad de ajuste y diagnóstico de influencia.

El paquete gbs (cont.)

Este paquete tiene 4 funciones para estimar los parámetros de distribuciones BSG, los cuales son:

BS: `mlebs()` o `mlegbs()`,

BS-lap y BS-logis: `mlegbs()` y

BS- t_ν : `mlebstNuFixed()` o `mlebst()`.

El paquete gbs (cont.)

Un ejemplo del uso de este paquete basado en datos no censurados `psi31` es:

(a) Análisis exploratorio de datos. La instrucción:

```
descriptiveSummary(psi31)
```

nos permite obtener Tabla 3.2 (la moda de los datos puede también ser calculada usando este comando).

Table 3.2: estadística descriptiva para `psi31`.

Mediana	Media	D.E	CV	CS	CC	Rango	Min.	Max.	n
133.00	133.73	22.36	16.72 %	0.33	0.97	142	70	212	101

El paquete gbs (cont.)

La instrucción:

```
histgbs(psi31,  
        mainTitle = "",  
        xLabel = "Lifetimes with maximum stress of 31.000 psi",  
        yLabel = "Density",  
        yRange = c(0, 0.02),  
        kernel = "t",  
        colourHistogram = 4,  
        colourBoxPlot = 4,  
        density = T)
```

produce simultáneamente un box-plot y un histograma para `psi31` y ajusta la fdp BSG correspondiente. (El box-plot puede suprimirse con la instrucción `boxPlot = FALSE`.) Figura 4.3 muestra estos gráficos.

El paquete gbs (cont.)

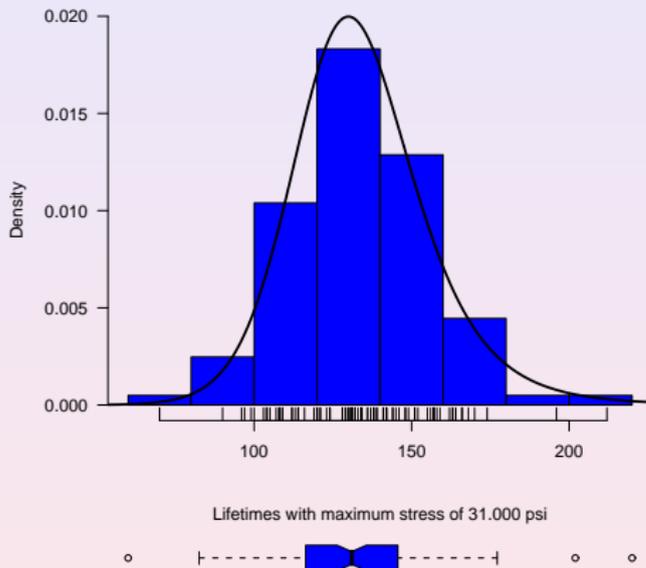


Figura 4.6: histograma y box-plot para ψ_{31} .

El paquete gbs (cont.)

(b) **Estimación.** Las instrucciones:

```
estimates    <- mlebst(psi31)
estimates
alpha        <- estimates$alphaEstimate
beta         <- estimates$betaEstimate
nu           <- estimates$nuOptimal
loglikelihood <- estimates$logLikelihood
```

permiten elegir automáticamente el valor de ν que maximiza la función de verosimilitud y calcular los EMV de α y β del modelo BS- t_ν . Estos resultados son guardados en la variable `estimates`.

El paquete gbs (cont.)

Obtaining:

`$alphaEstimate`

[1] 0.1474941

`$betaEstimate`

[1] 132.4927

`$nuOptimal`

[1] 8

`$logLikelihood`

[1] 455.6816

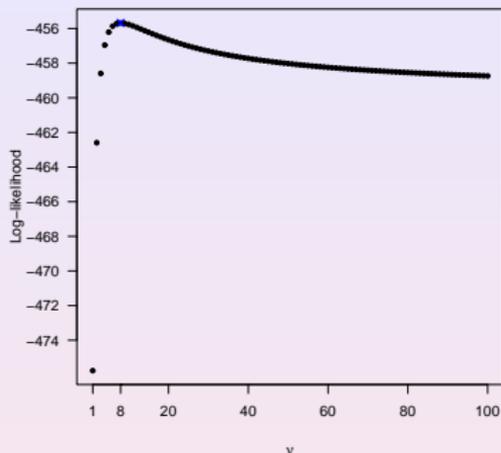


Figura 4.6: función de log-verosimilitud para el modelo $BS-t_\nu$ versus ν basado en [psi31](#).

El paquete gbs (cont.)

(c) **Diagnóstico.** La instrucción:

```
diagnosticsgbs(psi31, kernel = "normal", method = "conformal")
title(main = "BS distribution")
```

produce uno de los gráficos de Figura 3.5, el cual muestra la robustez de los EMV de los parámetros del modelo BS- t_8 .

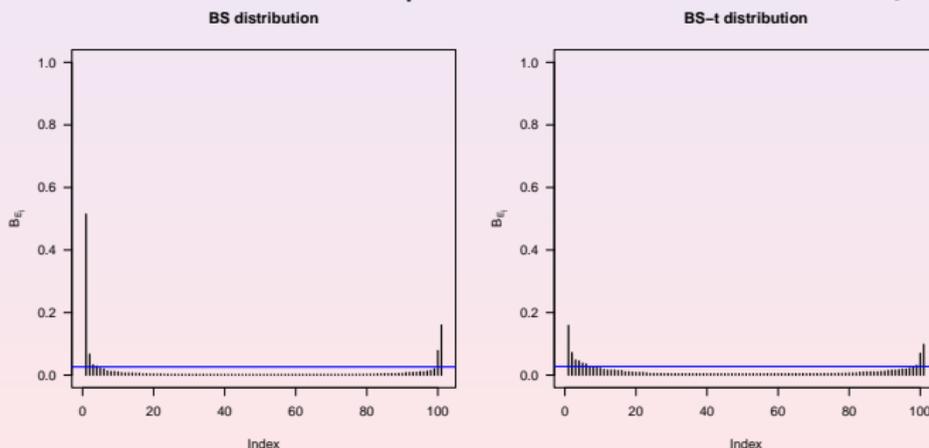


Figure 3.5: gráfico índice de influencia local conforme.

El paquete gbs (cont.)

Las instrucciones:

```
rcgbs(psi31, casesRemoved = 1, kernel = "t")
rcgbs(psi31, casesRemoved = 1, kernel = "normal")
```

permiten calcular los cambios relativos (CR) en porcentaje de cada parámetro estimado definido como

$$CR_{\theta_j} = \left| \frac{\hat{\theta}_j - \hat{\theta}_{j(I)}}{\hat{\theta}_j} \right| \times 100\%, \quad j = 1, 2,$$

donde $\hat{\theta}_{j(I)}$ denota los EMV de θ_j , con $\theta_1 = \alpha$ y $\theta_2 = \beta$, después de eliminar el conjunto I de casos. Tabla 3.3 presenta los resultados de los códigos anteriores.

El paquete gbs (cont.)

Table 3.3: CR para los parámetros de algunos models BSG para `psi31`.

Caso(s) removido	BS- t_8		BS	
	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\alpha}$	$\hat{\beta}$
{1}	5.86	0.16	6.95	0.67
{2}	2.22	0.30	2.09	0.39
{100}	2.18	0.27	2.30	0.40
{101}	0.64	0.36	3.57	0.49
{1, 2, 100, 101}	3.77	0.31	16.13	0.17

El paquete gbs (cont.)

(d) **Estimación con datos censurados.** Las instrucciones:

```
pc <- 20
status <- c(rep(1, floor((length(psi31) - pc) * (length(psi31)) / 100)),
            rep(0, (floor(pc * (length(psi31)) / 100))))
mlegbsc(psi31, status)
```

permite calcular los EMV de los parámetros α y β de las distribuciones BS- t_ν con datos censurados. Tabla 3.4 presenta estos resultados.

Table 3.4: EMV de α y β y prueba KS para los modelos indicados.

Modelo	$\hat{\alpha}$	$\hat{\beta}$	KS (valor-p)
BS- t_2	0.134	132.143	0.083 (p = 0,492)
BS- t_4	0.161	133.062	0.058 (p = 0,884)
BS- t_8	0.179	133.860	0.084 (p = 0,481)
BS- t_{20}	0.193	134.422	0.101 (p = 0,254)
BS- t_{100}	0.202	134.708	0.110 (p = 0,173)
BS	0.205	134.771	0.112 (p = 0,156)

El paquete gbs (cont.)

(e) **Bondad de ajuste con datos censurados.** Las instrucciones:

```
nuFixed = 4
ksgbsc(psi31, status,
       kernel = "t",
       graph = TRUE,
       xLabel = "Lifetimes with maximum stress of 31.000 psi",
       yLabel = "Cumulative distribution function")
```

llevan a cabo bondad de ajuste basada en la prueba KS y proporcionan una figura que muestra la coherencia entre fda empírica y la fda $BS-t_4$ teórica. Tabla 3.4 y Figura 3.6 presentan estos resultados.

El paquete gbs (cont.)

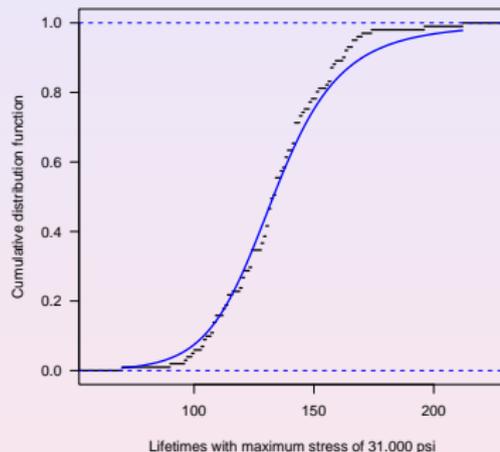


Figure 3.6: fda empírica vs. fda $BS-t_4$ teórica para $psi31c20$.

DISTRIBUCIÓN SENH-NORMAL

Distribución senh-normal (SN)

Rieck & Nedelman (1989) desarrollaron la distribución senh-normal.

Características:

- Simétrica.
- Presenta grados de curtosis mayores y menores que el modelo normal.
- Admite unimodalidad y bimodalidad.
- Tiene como caso particular la distribución log-Birnbaum-Saunders (log-BS).

Distribución sinh-normal

La distribución SN se obtiene mediante la transformación

$$Y = \gamma + \sigma \operatorname{arcsinh} \left(\frac{\alpha Z}{2} \right),$$

donde $Z \sim N(0, 1)$, $\alpha > 0$ es el parámetro de forma, $\gamma \in \mathbb{R}$ es el parámetro de posición y $\sigma > 0$ es el parámetro de escala. En este caso, la notación $Y \sim \operatorname{SN}(\alpha, \gamma, \sigma)$ es usada. La fdp de Y es

$$f_Y(y) = \phi \left(\frac{2}{\alpha} \sinh \left(\frac{y - \gamma}{\sigma} \right) \right) \frac{2 \cosh([y - \gamma]/\sigma)}{\alpha \sigma}, \quad y \in \mathbb{R}.$$

Distribución sinh-normal

- Función de distribución SN

$$F_Y(y) = \Phi \left(\frac{2}{\alpha} \sinh \left(\frac{y - \gamma}{\sigma} \right) \right), \quad y \in \mathbb{R}.$$

- Función cuantil SN

$$y(q) = F_Y^{-1}(q) = \gamma + \sigma \operatorname{arcsinh} \left(\frac{\alpha z_q}{2} \right),$$

donde z_q es el cuantil de orden q de la v.a. $Z \sim N(0, 1)$.

- Función de confiabilidad

$$R_Y(y) = \Phi \left(-\frac{2}{\alpha} \sinh \left(\frac{y - \gamma}{\sigma} \right) \right).$$

Distribución sinh-normal

- Tasa de fallas

$$h_Y(y) = \frac{2\phi\left(\frac{2}{\alpha} \sinh\left(\frac{y-\gamma}{\sigma}\right)\right) \cosh\left(\frac{y-\gamma}{\sigma}\right)}{\alpha\sigma\Phi\left(-\frac{2}{\alpha} \sinh\left(\frac{y-\gamma}{\sigma}\right)\right)}$$

- Función generadora de momentos

$$m(s) = \exp(\gamma s) \left[\frac{K_a(\delta^{-2}) + K_b(\delta^{-2})}{2K_{\frac{1}{2}}(\delta^{-2})} \right],$$

donde $a = \frac{\sigma s + 1}{2}$, $b = \frac{\sigma s - 1}{2}$ y $K(\cdot)$ es la función de Bessel modificada de tercer tipo definida por

$$K_\lambda(v) = \frac{1}{2} \left[\frac{v}{2} \right]^\lambda \int_0^\infty u^{-\lambda-1} \exp\left(-u - \frac{v^2}{4u}\right) du.$$

Distribución senh-normal

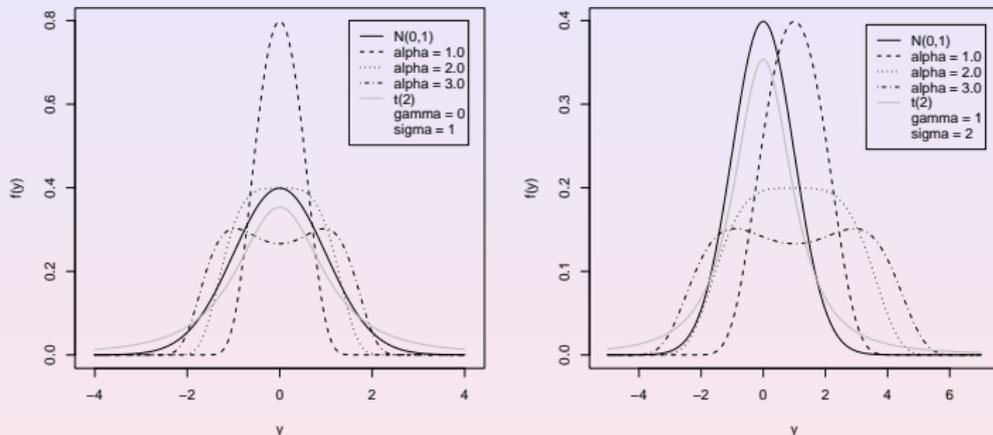


Figure 3.6: gráficos de densidades de la distribución SN para los valores indicados.

Distribución senh-normal

Rieck & Nedelman (1991) probaron que si $T \sim \text{BS}(\alpha, \beta)$, entonces

$$Y = \log(T) \sim \text{SN}(\alpha, \gamma, \sigma = 2),$$

donde $\gamma = \log(\beta)$. Así,

$$M_Y(s) = \mathbb{E}[\exp(sY)] = \mathbb{E}[T^s].$$

De esta manera, los momentos de cualquier orden de la distribución Birnbaum-Saunders pueden ser calculados a partir de la función generadora de momentos de la distribución SN.

MODELOS DE REGRESIÓN BS

Modelo de regresión BS con datos no censurados

Características:

Sean T_1, T_2, \dots, T_n v.a. independientes tal que $T_i \sim \text{BS}(\alpha_i, \beta_i)$, para $i = 1, 2, \dots, n$. Admitamos que la distribución de cada v.a. T_i depende de un vector de p variables explicativas

$\mathbf{x}_i^\top = [x_{i1}, x_{i2}, \dots, x_{ip}]$, en otras palabras, asuma que

- ① $\beta_i = \exp(\mathbf{x}_i^\top \boldsymbol{\eta})$, para $i = 1, 2, \dots, n$, donde $\boldsymbol{\eta} = [\eta_1, \eta_2, \dots, \eta_p]^\top$ es un vector de parámetros desconocidos;
- ② El parámetro de forma α_i no depende del vector de covariables \mathbf{x}_i^\top , esto es, $\alpha_i = \alpha$, para $i = 1, 2, \dots, n$.

Modelo de regresión BS con datos no censurados

Los supuestos anteriores combinados con la propiedad de proporcionalidad de la distribución BS (familia de escala) implican

$$T_i = \exp(\mathbf{x}_i^\top \boldsymbol{\eta}) \delta_i \sim \text{BS}(\alpha, \exp(\mathbf{x}_i^\top \boldsymbol{\eta})),$$

donde $\delta_i \sim \text{BS}(\alpha, 1)$, para $i = 1, 2, \dots, n$. Considerando $y_i = \log(T_i)$, se obtiene

$$y_i = \mathbf{x}_i^\top \boldsymbol{\eta} + \log(\delta_i),$$

para $i = 1, 2, \dots, n$. Esto quiere decir que estamos en presencia de un modelo de regresión log-lineal para la distribución BS cuya componente aleatoria ϵ_i es igual a $\log(\delta_i)$.

Ahora, usando la relación entre las distribuciones BS y SN, se concluye que

$$\epsilon_i = \log(\delta_i) \sim \text{SN}(\alpha, 0, 2), \quad i = 1, 2, \dots, n.$$

Modelo de regresión BS con datos no censurados

Modelo de regresión

$$y_i = \mathbf{x}_i^\top \boldsymbol{\eta} + \epsilon_i, \quad (4)$$

donde y_i son las respuestas observadas para la i -ésima unidad experimental, $\boldsymbol{\eta} = [\eta_1, \eta_2, \dots, \eta_p]^\top$ es un vector de parámetros desconocidos a ser estimado, $\mathbf{x}_i^\top = [x_{i1}, x_{i2}, \dots, x_{ip}]^\top$ son valores observados de p covariables y $\epsilon_i \sim \text{SN}(\alpha, 0, 2)$ es el término del error del modelo, para $i = 1, \dots, n$.

Este modelo se denomina regresión log-Birnbaum-Saunders y estamos interesados en estimar el vector $\boldsymbol{\theta} = [\boldsymbol{\eta}^\top, \alpha]^\top$.

Modelo de regresión BS con datos no censurados

El logaritmo de la función de verosimilitud del modelo (4) para $\boldsymbol{\theta} = [\alpha, \boldsymbol{\eta}^\top]^\top$ posee la siguiente forma:

$$\begin{aligned} \ell(\boldsymbol{\theta}) &= \sum_{i=1}^n \log \left(\frac{1}{\alpha \sqrt{2\pi}} \cosh \left(\frac{y_i - \gamma_i}{2} \right) \exp \left(-\frac{2}{\alpha^2} \left[\sinh \left(\frac{y_i - \gamma_i}{2} \right) \right]^2 \right) \right) \\ &\propto \sum_{i=1}^n \log(\xi_{i1}) - \frac{1}{2} \sum_{i=1}^n \xi_{i2}^2, \end{aligned}$$

donde

$$\xi_{i1} = \frac{2}{\alpha} \cosh \left(\frac{y_i - \gamma_i}{2} \right), \quad \xi_{i2} = \frac{2}{\alpha} \sinh \left(\frac{y_i - \gamma_i}{2} \right), \quad (5)$$

y $\gamma_i = \mathbf{x}_i^\top \boldsymbol{\eta}$, para $i = 1, 2, \dots, n$.

Modelo de regresión BS con datos no censurados

Las funciones score para α y η_j , para $j = 1, \dots, p$, están dadas, respectivamente, por

$$\dot{\mathbf{L}}_{\alpha} = \frac{1}{\alpha} \sum_{i=1}^n [\xi_{i2}^2 - 1]$$

y

$$\dot{\mathbf{L}}_{\eta_j} = \sum_{i=1}^n \left[\frac{x_{ij}}{\alpha^2} \sinh(y_i - \gamma_i) - \frac{x_{ij}}{2} \tanh\left(\frac{y_i - \gamma_i}{2}\right) \right], \quad j = 1, \dots, p.$$

Modelo de regresión BS con datos no censurados

La inferencia asintótica para el vector de parámetros $\boldsymbol{\theta} = [\boldsymbol{\eta}^\top, \alpha]^\top$ se puede basar en la normalidad asintótica el EMV $\hat{\boldsymbol{\theta}}$, dada por

$$\hat{\boldsymbol{\theta}} \underset{\sim}{\sim} N_{p+1}(\boldsymbol{\theta}, \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}),$$

donde $\boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}$ es la matriz de varianzas-covarianzas de $\hat{\boldsymbol{\theta}}$ que puede ser aproximada por la matriz de información observada $-\ddot{\mathbf{L}}^{-1}$ y obtenida a partir de

$$\ddot{\mathbf{L}} = \begin{bmatrix} \ddot{\mathbf{L}}_{\eta\eta} & \ddot{\mathbf{L}}_{\eta\alpha} \\ \ddot{\mathbf{L}}_{\alpha\eta} & \ddot{\mathbf{L}}_{\alpha\alpha} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top \mathbf{V} \mathbf{X} & \mathbf{X}^\top \mathbf{k} \\ \mathbf{k}^\top \mathbf{X} & \frac{1}{\alpha^2} \sum_{i=1}^n [1 - 3\xi_{i2}^2] \end{bmatrix}, \quad (6)$$

Modelo de regresión BS con datos no censurados

Donde $\mathbf{V} = \text{diag}\{v_1(\boldsymbol{\theta}), \dots, v_n(\boldsymbol{\theta})\}$ e $\mathbf{k} = [k_1, k_2, \dots, k_n]^\top$, con

$$v_i(\boldsymbol{\theta}) = \frac{1}{4} \left[\text{sech} \left(\frac{y_i - \gamma_i}{2} \right) \right]^2 - \frac{1}{\alpha^2} \cosh(y_i - \gamma_i), \quad (7)$$

y

$$k_i = -\frac{2}{\alpha^3} \sinh(y_i - \gamma_i), \quad i = 1, 2, \dots, n.$$

Una región de confianza del $100 \times [1 - \varrho] \%$ aproximada para el parámetro $\boldsymbol{\theta}$ puede ser obtenida mediante

$$\mathcal{R} \equiv \left\{ \boldsymbol{\theta} \in \mathbb{R}^{p+1} : [\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}]^\top \boldsymbol{\Sigma}_{\hat{\boldsymbol{\theta}}}^{-1} [\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}] \leq \chi_{1-\varrho}^2(p+1) \right\},$$

donde $\chi_{1-\varrho}^2(p+1)$ denota el $(1 - \varrho)$ -ésimo percentil de la distribución chi-cuadrado con $p+1$ grados de libertad.

Modelo de regresión BS con datos censurados

Considere le modelo de regresión log-BS dado por

$$y_i = \mathbf{x}_i^\top \boldsymbol{\eta} + \epsilon_i, \quad (8)$$

donde y_i es le logaritmo del tiempo de sobrevivencia o del tiempo de censura para a i -ésima unidad experimental,

$\boldsymbol{\eta} = [\eta_1, \eta_2, \dots, \eta_p]^\top$ es el vector de parámetros desconocidos a ser estimados, $\mathbf{x}_i^\top = (x_{i1}, x_{i2}, \dots, x_{ip},)$ son observaciones de p variables explicativas y $\epsilon_i \sim \text{SN}(\alpha, 0, 2)$.

Suponga una censura no informativa y que los tiempos de sobrevivencia y de censura son independientes. Denote por D y C el conjunto de individuos para los cuales y_i es el logaritmo del tiempo de sobrevivencia o el logaritmo del tiempo de censura, respectivamente.

Modelo de regresión BS con datos censurados

El logaritmo de la función de verosimilitud total del modelo (8) para $\boldsymbol{\theta} = [\alpha, \boldsymbol{\eta}^\top]^\top$ posee la siguiente forma:

$$\ell(\boldsymbol{\theta}) \propto \sum_{i \in D} \left[\log(\xi_{i1}) - \frac{\xi_{i2}^2}{2} \right] + \sum_{i \in C} \log(\Phi(-\xi_{i2})), \quad (9)$$

donde ξ_{i1} , ξ_{i2} y γ_i , para $i = 1, 2, \dots, n$, están dadas en la Ecuación (5).

Las funciones score para α y η_j , para $j = 1, \dots, p$, son, respectivamente,

$$\dot{L}_\alpha = \frac{1}{\alpha} \left[\sum_{i \in D} [\xi_{i2}^2 - 1] + \sum_{i \in C} \xi_{i2} h(\xi_{i2}) \right]$$

Modelo de regresión BS con datos censurados

$$\dot{\mathbf{L}}_{\eta_j} = \sum_{i \in D} \left[\frac{x_{ij}}{\alpha^2} \sinh(y_i - \gamma_i) - \frac{x_{ij}}{2} \tanh\left(\frac{y_i - \gamma_i}{2}\right) \right] + \sum_{i \in C} \frac{x_{ij} \xi_{i1} h(\xi_{i2})}{2},$$

donde $h(\xi_{i2}) = \frac{\phi(\xi_{i2})}{1 - \Phi(\xi_{i2})}$, con $\phi(\cdot)$ y $\Phi(\cdot)$ denotando la fdp y la fda $N(0, 1)$, respectivamente.

A matriz de información de Fisher observada es

$$\ddot{\mathbf{L}} = \begin{bmatrix} \ddot{\mathbf{L}}_{\eta\eta} & \ddot{\mathbf{L}}_{\eta\alpha} \\ \ddot{\mathbf{L}}_{\alpha\eta} & \mathbf{L}_{\alpha\alpha} \end{bmatrix} = \begin{bmatrix} \mathbf{X}^\top \mathbf{V} \mathbf{X} & \mathbf{X}^\top \mathbf{k} \\ \mathbf{k}^\top \mathbf{X} & \text{tr}(\mathbf{G}) \end{bmatrix}, \quad (10)$$

Modelo de regresión BS con datos censurados

Donde $\mathbf{V} = \text{diag}\{v_1(\boldsymbol{\theta}), v_2(\boldsymbol{\theta}), \dots, v_n(\boldsymbol{\theta})\}$,

$\mathbf{K} = [k_1(\boldsymbol{\theta}), k_1(\boldsymbol{\theta}), \dots, k_n(\boldsymbol{\theta})]^\top$ e $\mathbf{G} = \text{diag}\{g_1(\boldsymbol{\theta}), \dots, g_n(\boldsymbol{\theta})\}$,

con

$$v_i(\boldsymbol{\theta}) = \begin{cases} \frac{1}{4} \left[\text{sech} \left(\frac{y_i - \gamma_i}{2} \right) \right]^2 - \frac{1}{\alpha^2} \cosh(y_i - \gamma_i), & i \in D; \\ -\frac{1}{4} \xi_{i2} h(\xi_{i2}) - \frac{1}{4} \xi_{i1}^2 h'(\xi_{i2}), & i \in C; \end{cases} \quad (11)$$

$$k_i(\boldsymbol{\theta}) = \begin{cases} -\frac{2}{\alpha^3} \sinh(y_i - \gamma_i), & i \in D; \\ -\frac{1}{2\alpha} \xi_{i1} h(\xi_{i2}) - \frac{1}{\alpha^3} \sinh(y_i - \gamma_i) h'(\xi_{i2}), & i \in C; \end{cases}$$

Modelo de regresión BS con datos censurados

y

$$g_i(\boldsymbol{\theta}) = \begin{cases} \frac{1}{\alpha^2} - \frac{3\xi_{i2}^2}{\alpha^2}, & i \in D; \\ -\frac{2}{\alpha^2}\xi_{i2} h(\xi_{i2}) - \frac{1}{\alpha^2}\xi_{i2}^2 h'(\xi_{i2}), & i \in C; \end{cases}$$

Influencia local

El método de influencia local tiene como objetivo principal evaluar cambios en los resultados del análisis cuando pequeñas perturbaciones son incorporadas al modelo y/o a los datos. Si esas perturbaciones causaran efectos desproporcionales, esto puede ser indicio de que el modelo está mal ajustado o que puedan existir violaciones serias de los supuestos en que se basa el modelo.

Influencia local

Cook (1986) propuso estudiar el comportamiento de la función

$$LD(\boldsymbol{\omega}) = 2 \left[\ell(\hat{\boldsymbol{\theta}}) - \ell(\hat{\boldsymbol{\theta}}_{\boldsymbol{\omega}}) \right], \quad (12)$$

en una vecindad del vector sin perturbar, $\boldsymbol{\omega}_0$.

EL procedimiento consiste en seleccionar una dirección unitaria, \boldsymbol{l} , tal que $\|\boldsymbol{l}\| = 1$, y entonces considerar el gráfico de $LD(\boldsymbol{\omega}_0 + a\boldsymbol{l})$ versus a , con $a \in \mathbb{R}$. Ese gráfico es denominado la línea proyectada. Cada línea proyectada puede caracterizarse mediante la curvatura normal $C_{\boldsymbol{l}}(\boldsymbol{\theta})$ en torno de $a = 0$.

Una sugerencia es considerar la dirección $\boldsymbol{l}_{\text{máx}}$ correspondiendo la curvatura máxima $C_{\boldsymbol{l}_{\text{máx}}}(\boldsymbol{\theta})$.

Influencia local

Cook (1986) mostró que la curvatura normal puede ser expresada en forma geral como

$$C_l(\boldsymbol{\theta}) = 2 \left| \mathbf{l}^\top \boldsymbol{\Delta}^\top \ddot{\mathbf{L}}^{-1} \boldsymbol{\Delta} \mathbf{l} \right|, \quad (13)$$

donde $\boldsymbol{\Delta}$ es una matriz $(p + q) \times n$ que depende del esquema de perturbación usado, cuyos elementos son

$$\Delta_{ji} = \frac{\partial^2 \ell(\boldsymbol{\theta} | \boldsymbol{\omega})}{\partial \theta_j \partial \omega_i}, \quad i = 1, 2, \dots, q, \quad j = 1, 2, \dots, p + 1,$$

con todas las cantidades siendo evaluadas en $\boldsymbol{\omega} = \boldsymbol{\omega}_0$ y $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}$. Entonces, $C_{l_{\text{máx}}}$ es le mayor autovalor de la matriz

$$\mathbf{B} = \boldsymbol{\Delta}^\top \ddot{\mathbf{L}}^{-1} \boldsymbol{\Delta}$$

y $l_{\text{máx}}$ es el autovector correspondiente.

Análisis de residuos

El residuo componente de desvío se define por

$$r_{DC_i} = \text{sign}(y_i - \hat{\gamma}_i) \sqrt{2 [\ell_i(\hat{\boldsymbol{\theta}}_S) - \ell_i(\hat{\boldsymbol{\theta}})]}, \quad i = 1, 2, \dots, n,$$

Davison y Gigli (1989) definieron el residuo componente de desvío para datos censurados como

$$r_{DC_i} = \text{sign}(y_i - \hat{\gamma}_i) \sqrt{-2 \log \left(\hat{R}_Y(y_i) \right)}, \quad i = 1, 2, \dots, n, \quad (14)$$

donde $\hat{R}_Y(y_i)$ es el EMV de la función de confiabilidad.

Análisis de residuos

Considerando α fijo o conocido y $\alpha < 2$, se tiene que el residuo componente de desvío para el modelo de regresión log-Birnbaum-Saunders es

$$r_{\text{DC}_i} = \begin{cases} \text{sign}(y_i - \hat{\gamma}_i) \sqrt{-2 \log \left(\cosh \left(\frac{y_i - \hat{\gamma}_i}{2} \right) \right) + \frac{2}{\hat{\alpha}^2} \left[\sinh \left(\frac{y_i - \hat{\gamma}_i}{2} \right) \right]^2}, & i \in D \\ \text{sign}(y_i - \hat{\gamma}_i) \sqrt{-2 \log \left(\Phi \left(-\frac{2}{\hat{\alpha}} \sinh \left(\frac{y_i - \hat{\gamma}_i}{2} \right) \right) \right)}, & i \in C. \end{cases}$$

Análisis de residuos

- Therneau, Grambsch y Fleming (1990) introdujeron el residuo componente de desvío en procesos de conteo usando básicamente residuos martingala.
- Los residuos martingala son asimétricos y asumen un valor máximo en $+1$ y un valor mínimo en $-\infty$.
- Los residuos martingala pueden expresarse como $r_{M_i} = \delta_i + \log[\hat{R}(y_i)]$, con $\delta_i = 0$ indicando si la observación es censurada y $\delta_i = 1$ si no lo es.

Análisis de residuos

- El residuo martingala para modelos de regresión log-BS asume la forma

$$r_{M_i} = \delta_i + \log \left\{ 1 - \Phi \left[\frac{2}{\hat{\alpha}} \sinh \left(\frac{y_i - \hat{\gamma}_i}{2} \right) \right] \right\}.$$

- El residuo componente de desvío propuesto por Therneau, Grambsch y Fleming (1990) es una transformación del residuo martingala con la finalidad de atenuar la asimetría.

Ilustración 1

Considere los datos de fatiga biaxial reportados por Brown y Miller (1978), donde la respuesta, N , es el número de ciclos hasta que ocurre la falla y la variable independiente, W , es el esfuerzo por ciclo (medida en MJ/m^3). Se estudiaron 46 observaciones para establecer un modelo adecuado de predicción, las que se presentan en el Cuadro 5.

Ilustración 1

W	N	W	N	W	N	W	N
11,5	3280	24,0	804	40,1	750	60,3	283
13,0	5046	24,6	1093	40,1	316	60,5	212
14,3	1563	25,2	1125	43,0	456	62,1	327
15,6	4707	25,5	884	44,1	552	62,8	373
16,0	977	26,3	1300	46,5	355	66,5	125
17,3	2834	27,9	852	47,3	242	67,0	187
19,3	2266	28,3	580	48,7	190	67,1	135
21,1	2208	28,4	1066	52,9	127	67,9	245
21,5	1040	28,6	1114	56,6	185	68,8	137
22,6	700	30,9	386	59,9	255	75,4	200
22,6	1583	31,9	745	60,2	195	100,5	190
24,0	482	34,5	736				

Table 3.4: Datos de fatiga biaxial de Brown y Miller (1978).

Ilustración 1

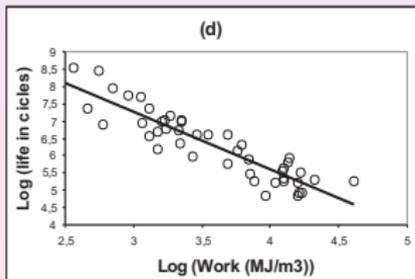
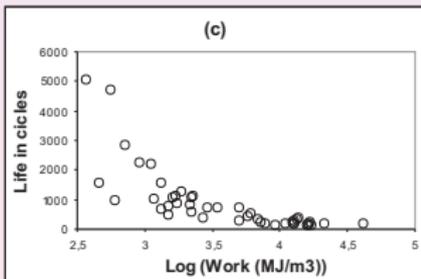
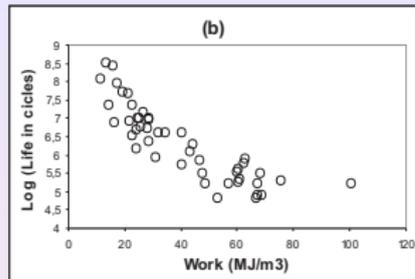
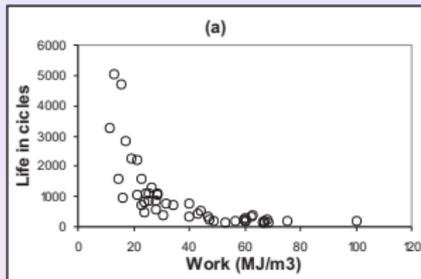


Figure 3.6: Diagramas de dispersión de W versus N (a), W versus $\log(N)$ (b), $\log(W)$ versus N (c) y $\log(W)$ versus $\log(N)$ (d).

Ilustración 1

- $\mu_{\log(N_i)} = \beta_0 + \beta_1 \log(W_i)$, $i = 1, 2, \dots, 46$, donde $\mu_{\log(N_i)}$ es la media del logaritmo de N_i .
- EMV de $\beta = [\beta_0, \beta_1]^T$ son (con sus errores estándar entre paréntesis)

$$\hat{\beta}_0 = 12,289(0,406) \quad \text{y} \quad \hat{\beta}_1 = -1,673(0,113).$$

- Los coeficientes β_0 y β_1 son estadísticamente significativos, con valores-p de 4,17416054E-31 y 9,92861E-19, respectivamente.
- Ecuación de predicción

$$\widehat{\mu_{\log(N_i)}} = 12,289 - 1,673 \log W.$$

Ilustración 1

Table 3.4: Cambios relativos en el vector $\hat{\eta}$ después de eliminar las observaciones indicadas.

Caso eliminado	$\hat{\eta}_0$	CR	p-value	$\hat{\eta}_1$	CR	Valor-ps
Ninguno	12.280	0.000	< 0,001	-1.671	0.000	< 0,001
2	12.123	0.013	< 0,001	-1.630	0.025	< 0,001
3	12.406	0.010	< 0,001	-1.703	0.019	< 0,001
4	12.085	0.016	< 0,001	-1.620	0.031	< 0,001
5	12.468	0.015	< 0,001	-1.719	0.029	< 0,001
12	12.379	0.008	< 0,001	-1.694	0.014	< 0,001
46	12.471	0.016	< 0,001	1.729	0.035	< 0,001

References

-  Barros, M., Paula, G.A., Leiva, V. (2009) An R implementation for generalized Birnbaum-Saunders distributions. *Comp. Stat. Data Anal.* , 53:1511-1528.
-  Birnbaum, Z. W., Saunders, S. C. (1969a) A new family of life distributions. *J. Appl. Prob.*6:319-327.
-  Birnbaum, Z. W., Saunders, S. C. (1969b) Estimation in a new family of life distributions. *J. Appl. Prob.*6:319-327.
-  Cysneiros, A., Cribari-Neto, F., Araujo, C.G.J. (2008) On Birnbaum-Saunders inference. *Comp. Stat. Data Anal.* 52:4939-4950.
-  Díaz, J.A., Leiva, V. (2005) A new family of life distributions based on the contoured elliptically distributions. *J. Stat. Plan. Infer.*128:445-457.
-  Fang, K. T., Kotz, S., Ng, K. W. (1990) *Symmetric, Multivariate and Related Distributions*. Chapman, London.

References

-  Ihaka, R., Gentleman, R. (1996) R: A language for data analysis and graphics. *J. Comp. Graph. Stat.*5: 299-314.
-  Kundu, D., Kannan, N., Balakrishnan, N. (2008) On the hazard function of Birnbaum-Saunders distribution and associated inference. *Comp. Stat. Data Anal.* 52:2692-2702.
-  Leiva, V., Hernández, H., Riquelme, M. (2006) A new package for the Birnbaum-Saunders distribution. *R News* 6:35-40.
-  Leiva, V., Riquelme, M., Balakrishnan, N., Sanhueza, A. 2008. Lifetime Analysis based on the Generalized Birnbaum-Saunders Distribution. *Comp. Stat. Data Anal.* 52:2079-2097.
-  Leiva, V., Barros, M., Paula, G.A., Sanhueza, A. (2008) Generalized Birnbaum-Saunders distributions applied to air pollutant concentration. *Environmetrics* 19:235-249.
-  Leiva, V., Sanhueza, A., Angulo, J.M. (2009) A length-biased version of the Birnbaum-Saunders distribution with application in water quality. *Stoch Environ. Res. Risk Assessm.* 23:299-307.

References

-  Lemonte, A., Cribari-Neto, F., Vasconcellos, K.L.P. (2007) Improved statistical inference for the two-parameter Birnbaum-Saunders distribution. *Comp. Stat. Data Anal.* 51:4656-4681.
-  Lemonte, A., Simas, A. Cribari-Neto, F. (2008) Bootstrap-based improved estimators for the two-parameter Birnbaum-Saunders distribution. *J. Stat. Comp. Simul.* 78:37-49.
-  Marshall, A.W., Olkin, I. (2007) *Life Distributions*. Springer, New York.
-  Sanhueza, A., Leiva, V., Balakrishnan, N., (2008) The generalized Birnbaum-Saunders distribution and its theory, methodology and application. *Comm. Stat. Theor. Meth.* 37:645-670.
-  Saunders, S.C. (2007) *Reliability, Life Testing and Prediction of Services Lives*. Springer, New York.

FIN

MUCHAS GRACIAS