

# Data Análisis con No Detectados

## Límites Reportados y Configuración

Ing. A.Otiniano

UNI

2024-06-17



Analytics AoZ

# Table of contents I

- 1 Breve Historia de la Estadística de Datos Censados
- 2 Límites de Reporte
- 3 Configuración de Límites
- 4 Data entre Límites
- 5 Sumario

# Objetivos

- Entender el concepto de **Survival Analysis**.

# Objetivos

- Entender el concepto de **Survival Analysis**.
- Conocer los diferentes **Límites de Reporte**.

# Objetivos

- Entender el concepto de **Survival Analysis**.
- Conocer los diferentes **Límites de Reporte**.
- Diferenciar los **Límites de Detección y Cuantificación**.

# No Detectados

¿Qué son los No detectados?

- Son *data real* - no debe ser eliminada.

# No Detectados

¿Qué son los No detectados?

- Son *data real* - no debe ser eliminada.
- Data censada - conocida como arriba o debajo del **threshold**.

# No Detectados

¿Qué son los No detectados?

- Son *data real* - no debe ser eliminada.
- Data censada - conocida como arriba o debajo del **threshold**.
- Valores censado por la derecha, la izquierda o por intervalos:

# No Detectados

¿Qué son los No detectados?

- Son *data real* - no debe ser eliminada.
- Data censada - conocida como arriba o debajo del **threshold**.
- Valores censado por la derecha, la izquierda o por intervalos:
  - censado-derecho:  $<1 <— 1$

# No Detectados

¿Qué son los No detectados?

- Son *data real* - no debe ser eliminada.
- Data censada - conocida como arriba o debajo del **threshold**.
- Valores censados por la derecha, la izquierda o por intervalos:
  - censado-derecho:  $<1 \leftarrow 1$
  - censado-izquierdo:  $>1 \rightarrow 1$

# No Detectados

¿Qué son los No detectados?

- Son *data real* - no debe ser eliminada.
- Data censada - conocida como arriba o debajo del **threshold**.
- Valores censados por la derecha, la izquierda o por intervalos:
  - censado-derecho:  $<1 \leftarrow 1$
  - censado-izquierdo:  $>1 \rightarrow 1$
  - **intervalo censado (0 a 1)**  $[0 \leftarrow \rightarrow 1]$ , más usado.

## Section 1

# Breve Historia de la Estadística de Datos Censados

# Trabajos Realizados

- Data censada usada desde los 1950s en estadística industrial y medicina.

# Trabajos Realizados

- Data censada usada desde los 1950s en estadística industrial y medicina.
- Los métodos han sido desarrollados para diferentes ramas en geociencias censored value.

# Trabajos Realizados

- Data censada usada desde los 1950s en estadística industrial y medicina.
- Los métodos han sido desarrollados para diferentes ramas en geociencias censored value.
- Es denominado **Survival Analysis** o **Reliability analysis**.

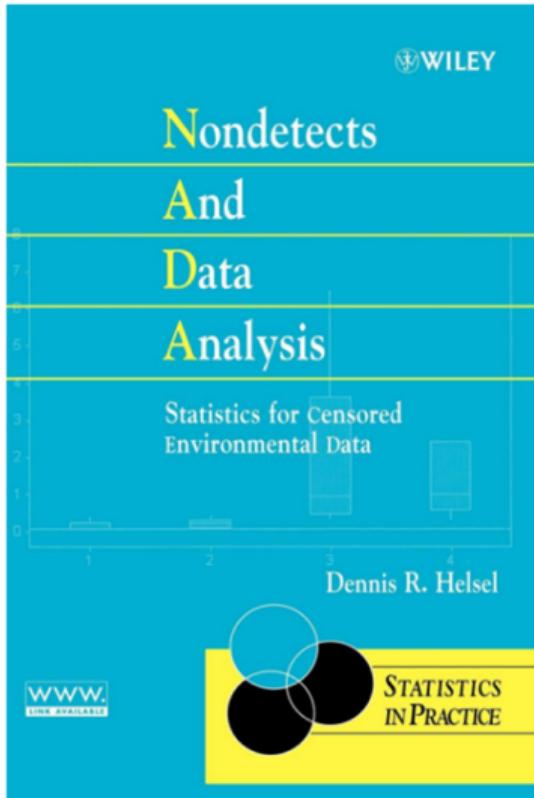
# Trabajos Realizados

- Data censada usada desde los 1950s en estadística industrial y medicina.
- Los métodos han sido desarrollados para diferentes ramas en geociencias censored value.
- Es denominado **Survival Analysis** o **Reliability analysis**.
- La información fue reconocida gracias a Hesel con el libro [Nondetects And Data Analysis 2005](#)

# Trabajos Realizados

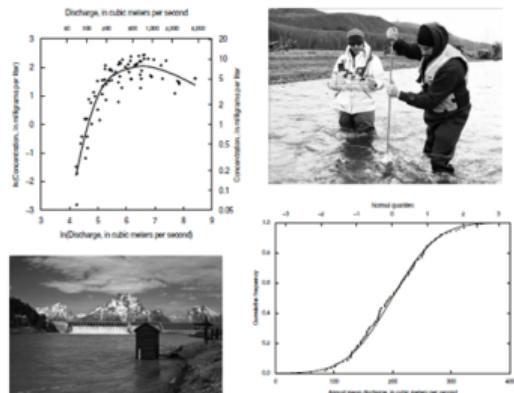
- Data censada usada desde los 1950s en estadística industrial y medicina.
- Los métodos han sido desarrollados para diferentes ramas en geociencias censored value.
- Es denominado **Survival Analysis** o **Reliability analysis**.
- La información fue reconocida gracias a Hesel con el libro [Nondetects And Data Analysis 2005](#)
- Existe información de su aplicabilidad en sedimentos de quebrada, aguas y suelos, en roca en menor proporción.

# Libros Importantes



## Statistical Methods in Water Resources

Chapter 3 of  
Section A, Statistical Analysis  
Book 4, Hydrologic Analysis and Interpretation



Techniques and Methods 4–A3  
Supersedes USGS Techniques of Water-Resources Investigations, book 4, chapter A3

U.S. Department of the Interior  
U.S. Geological Survey

**NADA2: Data Analysis for Censored Environmental Data**

Contains methods described by Dennis Helsel in his book "Statistics for Censored Environmental Data".

Version: 1.1.6  
Depends: R (≥ 3.6), EnvStats (≥ 2.4)  
Imports: grDevices, graphics, stats, utils, fitter, introdStat, Kendall, mkr, knitr, markdown,为人, car, alr4, rms  
Suggests: Published: 2024-02-26  
Author: Paul Julian [aut, cre], Dennis Helsel [aut, eph]  
Maintainer: Paul Julian <pauljulianphd@gmail.com>  
BugReports: <https://github.com/SwampThing/PaulNADA2/issues>  
License: MIT + file LICENSE  
URL: <https://github.com/SwampThing/PaulNADA2>  
NeedsCompilation: no  
Materials: [README](#) [NEWS](#)  
CRAN checks: [NADA2 results](#)

Documentation:

Reference manual: [NADA2.pdf](#)  
Vignettes: [DataAnalysis](#)

Downloads:

Package source: [NADA2\\_1.1.6.tar.gz](#)  
Windows binaries: r-devel: [NADA2\\_1.1.6.zip](#), r-release: [NADA2\\_1.1.6.zip](#)  
macOS binaries: r-release (arm64): [NADA2\\_1.1.6.zip](#), r-oldrel (arm64): [NADA2\\_1.1.6.zip](#)  
Old sources: [NADA2 archive](#)

Linking:

Please use the canonical form <https://CRAN.R-project.org/package=NADA2> for  
retrieving this package.

## Section 2

### Límites de Reporte

# Conceptos de Límites de Reporte

Es considerado un término **general**, calculado para una variedad de usos y formas. Existen dos tipos principales de límite de reporte:

- Límite de detección.

Desviación estándar asumida ser constante.

# Conceptos de Límites de Reporte

Es considerado un término **general**, calculado para una variedad de usos y formas. Existen dos tipos principales de límite de reporte:

- Límite de detección.
- Límite de cuantificación.

Desviación estándar asumida ser constante.

# Clases

- Límite de Detección (LD): Valores medidos sobre el **threshold** son improbable a resultar desde una concentración de cero real.

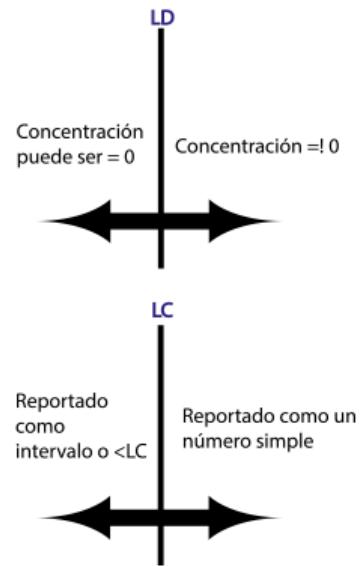


Figure 2: LD & LC

# Clases

- Límite de Detección (LD): Valores medidos sobre el **threshold** son improbable a resultar desde una concentración de cero real.
- Límite de Cuatificación (LC): **Threshold** sobre cual valores numéricos únicos (más que un intervalo o  $<LC$ ) son reportados.

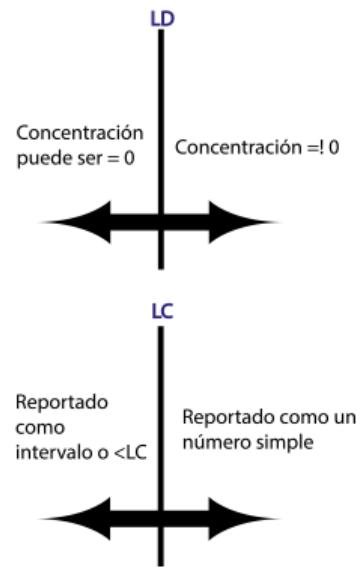
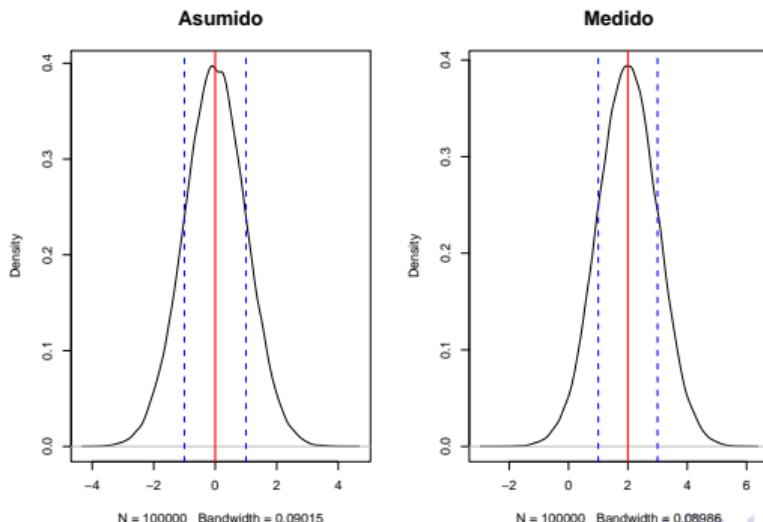


Figure 2: LD & LC

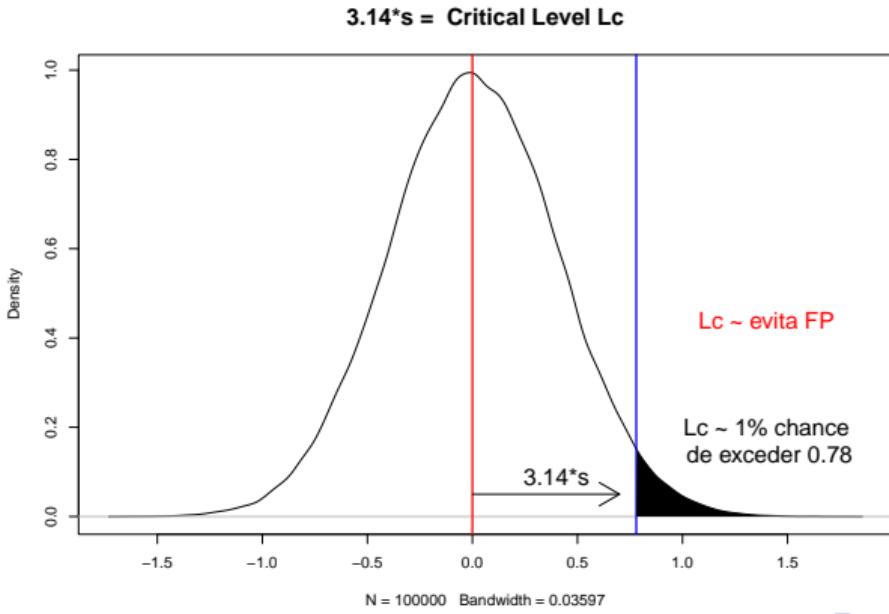
# Configurando el Límite de Detección

No es sencillo medir una señal de 0, por este motivo usamos un ruido alrededor de una solución estándar (**muy baja!**) = desviacion std ( $s$ ), por ende asumimos el ruido alrededor de 0. Actualmente, se toma un modelo de regresión para calcular  $s$  en función de la concentración.



# El Nivel Crítico

Se considera que  $3.14 * s > 0$  debería tener un chance de solo **1%** de realmente originar una concentración de 0, asumiendo una distribución normal.



# Concepto primario

Una señal de verdadero cero es improbable a ser medido sobre este **threshold** (critical level). Evita falso positivo, pero no falso negativo.

**Table 1.** Quantities derivable from a confusion matrix in Bayesian estimators for common loss functions

		Estimate		$\Sigma$
		Positive	Negative	
Truth	Positive	$TP$	$FN$	$\hat{N}_+ = TP + FN$
	Negative	$FP$	$TN$	$\hat{N}_- = FP + TN$
Total	$N_+ = TP + FP$	$N_- = FN + TN$	$N = TP + FP + FN + TN$	

Table Quantities

Figure 3: Conceptos de FP y FN

## Section 3

# Configuración de Límites

# Límite de Detección (Ld)

- Para evitar los falsos negativos se configura un límite más alto, Límite de Detección (Ld).

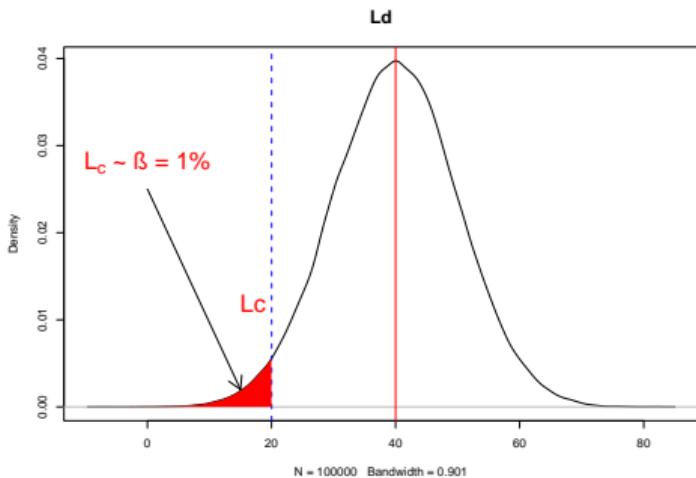


Figure 4: Límite de Detección

# Límite de Detección (Ld)

- Para evitar los falsos negativos se configura un límite más alto, Límite de Detección (Ld).
- Una concentración verdadera a un límite más alto que Ld tiene 1% de chance de ser medida y reportada como  $< L_c$ . Esto todavía tiene un 50% de chance de ser medido y reportado como  $< L_d$ ,

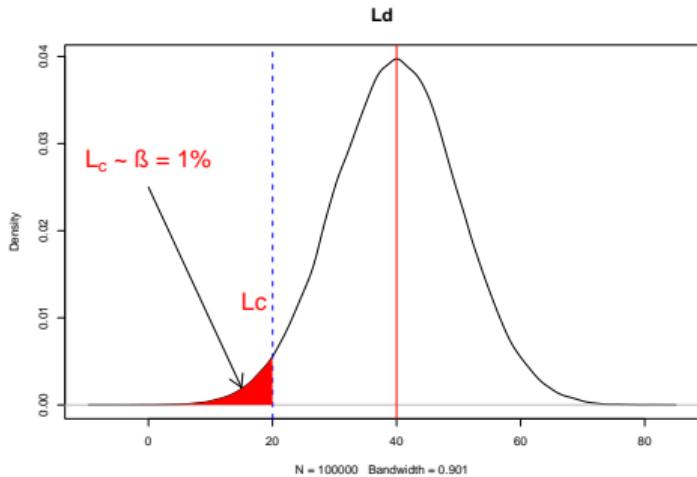


Figure 4: Límite de Detección

# Límite de Detección (Ld)

Práctica estándar:  $\alpha = \beta = 1\%$ ,  $Ld = 2*Lc$ .

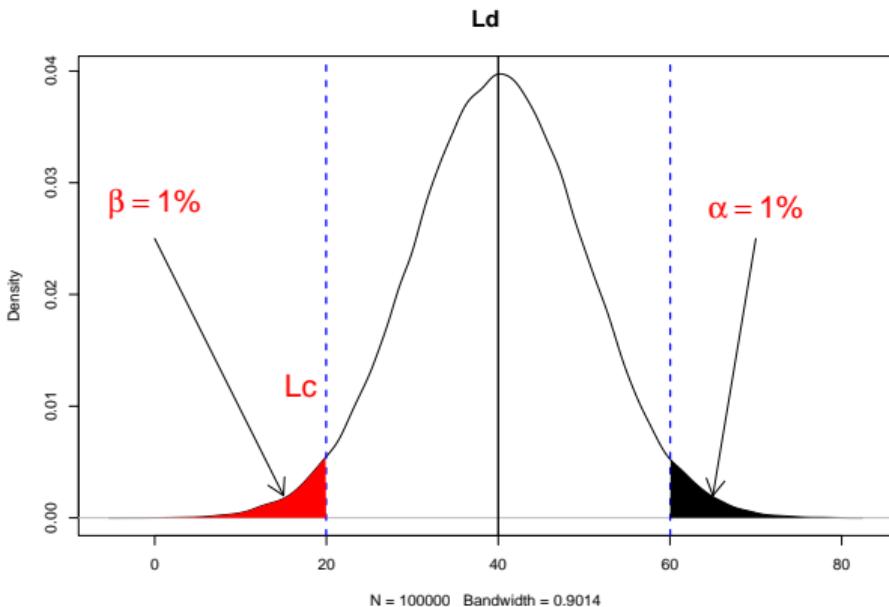


Figure 5: Límite de Detección

## Concepto primario

Una concentración verdadera en level crítico y puede ser 0, es improbable a ser medido sobre el límite de detección. Esto no evita falsos negativos para valores por encima del Ld ser llamados  $< Ld$ .

# Límite de Cuantificación (Ld)

- Es más alto que el **threshold** que el Ld.

# Límite de Cuantificación (Ld)

- Es más alto que el **threshold** que el Ld.
- Sobre el Lc un número reportado como un simple número es confiable.

# Límite de Cuantificación (Ld)

- Es más alto que el **threshold** que el Ld.
- Sobre el Lc un número reportado como un simple número es confiable.
- Usualmente basado en una señal de radio de ruido.

# Configuración del Límite de Cuantificación (Ld)

- Límite de cuantificación es seteado a varios valores, usualmente es igual a 10 veces el ruido ( $\text{signal/noise} = 10$ ).

$$10 * s = 3.18 * (3.14 * s) = 3.18 * Lc$$

Esto es  $1.6 * Ld$ . Usualmente redondeado a  $2 * Ld$ . La definición más común es  $2 * Ld$ .

# Configuración del Límite de Cuantificación (Ld)

- Límite de cuantificación es seteado a varios valores, usualmente es igual a 10 veces el ruido ( $\text{signal/noise} = 10$ ).

$$10 * s = 3.18 * (3.14 * s) = 3.18 * Lc$$

Esto es  $1.6 * Ld$ . Usualmente redondeado a  $2 * Ld$ . La definición más común es  $2 * Ld$ .

- Otros multiplos son usados  $5 * Ld$ , etc. Eso es dependiendo del laboratorio y los métodos o pautas que usan para su elección.

# Concepto primario

Sobre ese nivel la señal es suficientemente grande mayor que el ruido para reportar como un simple número confiable.

## Notaciones usadas

Entre el Ld y el Lc existen observaciones las cuales no son cero, pero aún son muy variables para ser reportadas como un número simple. Laboratorios tiene diferentes formas de reportarlo como:

- \* "valores J", "valores E" y "data remarcada".

Un valor remarcado (4J) no puede ser verdadero certamente abajo o encima de otro (5J). Demasiado ruido.

## Section 4

# Data entre Límites

# Data entre Límites

- Es acordado que la data sobre el  $L_c$  es reporta como un simple número.

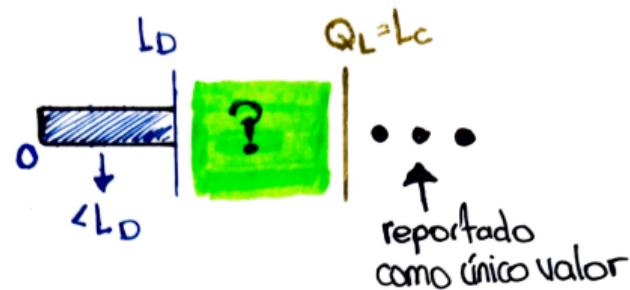


Figure 6: Data entre límites

# Data entre Límites

- Es acordado que la data sobre el  $L_c$  es reportado como un simple número.
- La mayoría acuerda que debajo del  $L_d$  es reportado como desde 0 hasta el  $L_d$  (intervalo) o  $<L_d$ .

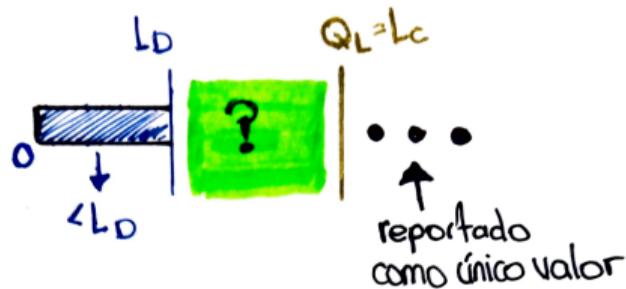


Figure 6: Data entre límites

# Data entre Límites

- Es acordado que la data sobre el  $L_c$  es reportada como un simple número.
- La mayoría acuerda que debajo del  $L_d$  es reportado como desde 0 hasta el  $L_d$  (intervalo) o  $< L_d$ .
- No existe una convención sobre que hacer con data entre el  $L_d$  y  $L_c$ . Existen tres opciones validas.

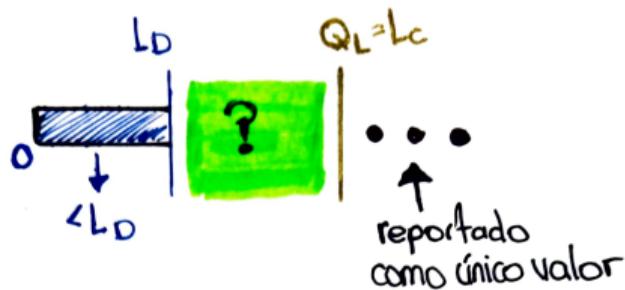
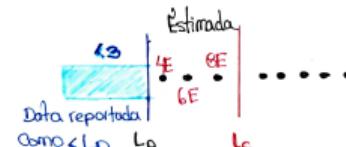
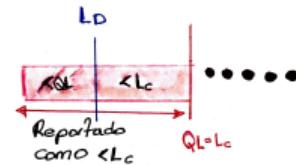


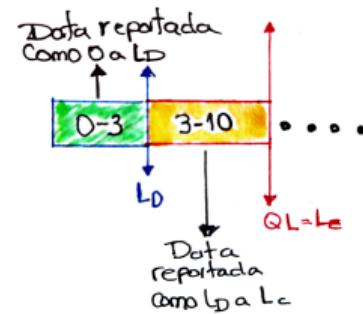
Figure 6: Data entre límites

# Data entre Límites

- Tres métodos son los más usados para el límite de reporte entre los cuales están:

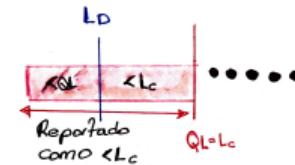


$4E < GE < BE (?)$

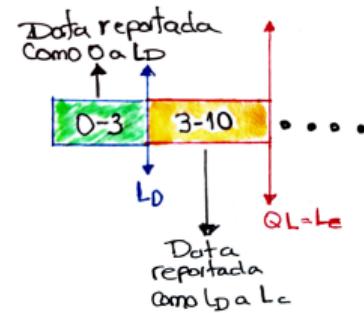


# Data entre Límites

- Tres métodos son los más usados para el límite de reporte entre los cuales están:
  - Usa el límite de cuantificación ( $L_c$ ) como el límite de reporte para todos los valores medidos debajo del límite de cuantificación ( $L_c$ ).



$4E < GE < BE (?)$



# Data entre Límites

- Tres métodos son los más usados para el límite de reporte entre los cuales están:
  - Usa el límite de cuantificación ( $L_c$ ) como el límite de reporte para todos los valores medidos debajo del límite de cuantificación ( $L_c$ ).
  - Use el límite de detección ( $L_d$ ) como el límite de reporte.



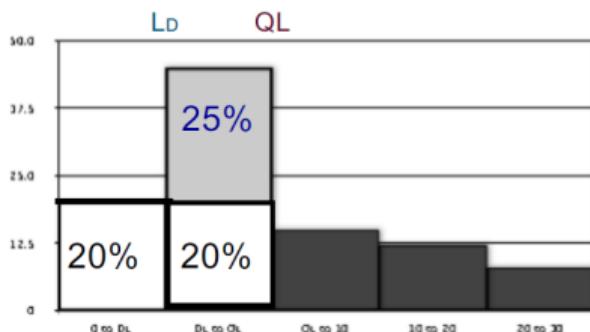
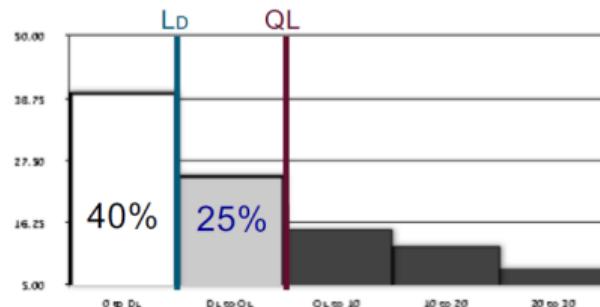
# Data entre Límites

- Tres métodos son los más usados para el límite de reporte entre los cuales están:
  - Usa el límite de cuantificación ( $L_c$ ) como el límite de reporte para todos los valores medidos debajo del límite de cuantificación ( $L_c$ ).
  - Use el límite de detección ( $L_d$ ) como el límite de reporte.
  - Usar métodos de intervalo censoreado.



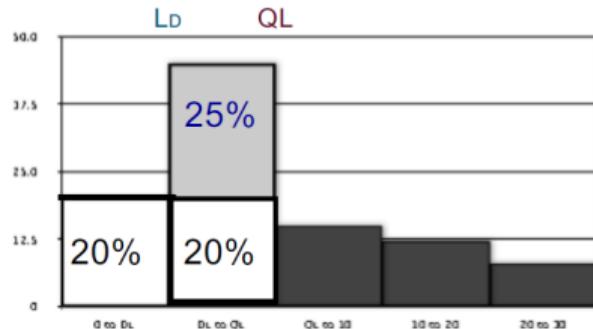
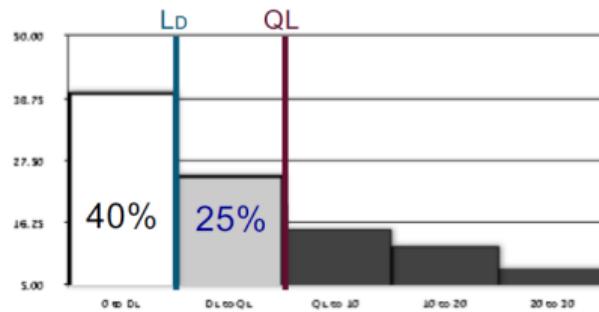
# Censado Interno

- **Evitar data con censado interno**, produce estimacion desviada de hacia arriba de la media e forma incorrecta:



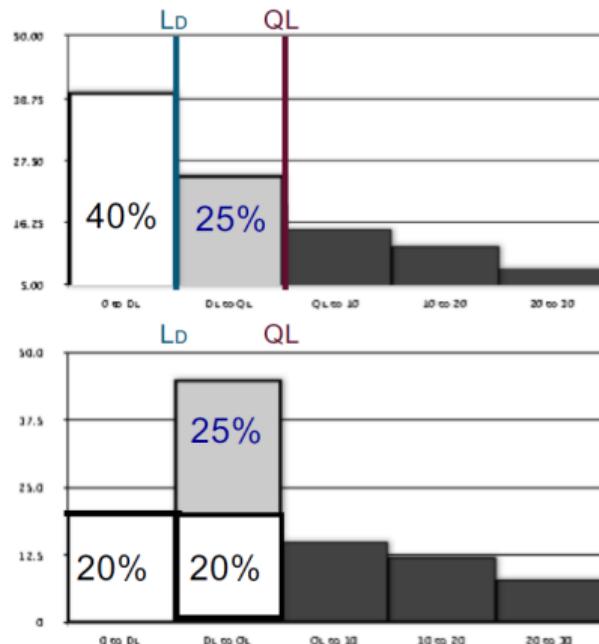
# Censado Interno

- **Evitar data con censado interno**, produce estimacion desviada de hacia arriba de la media e forma incorrecta:
  - Data como medida.



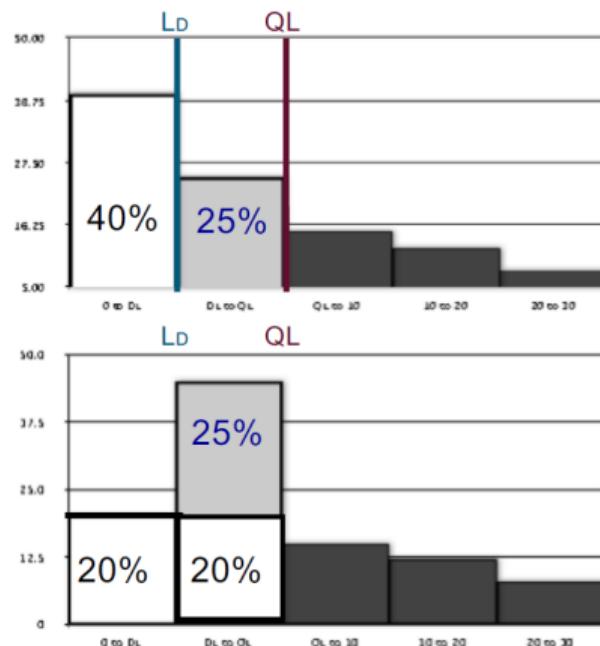
# Censado Interno

- **Evitar data con censado interno**, produce estimacion desviada de hacia arriba de la media e forma incorrecta:
  - Data como medida.
  - Reportada como medida  $<L_d$  como  $<L_c$ .



# Censado Interno

- **Evitar data con censado interno**, produce estimacion desviada de hacia arriba de la media e forma incorrecta:
  - Data como medida.
  - Reportada como medida  $<L_d$  como  $<L_c$ .
  - Data reportada sesgada.



## Section 5

### Sumario

# Sumario de Límites

- Límite de detección (Ld): debajo podría ser concentración cero.

# Sumario de Límites

- Límite de detección (Ld): debajo podría ser concentración cero.
- Límite de cuantificación (Ld): debajo esta la data con mucho ruido para ser reportado como un simple número.

# Sumario de Límites

- Límite de detección (Ld): debajo podría ser concentración cero.
- Límite de cuantificación (Ld): debajo esta la data con mucho ruido para ser reportado como un simple número.
- Data puede ser propiamente censada:

# Sumario de Límites

- Límite de detección (Ld): debajo podría ser concentración cero.
- Límite de cuantificación (Ld): debajo esta la data con mucho ruido para ser reportado como un simple número.
- Data puede ser propiamente censada:
  - Al límite de cuantificación (Lc).

# Sumario de Límites

- Límite de detección (Ld): debajo podría ser concentración cero.
- Límite de cuantificación (Ld): debajo esta la data con mucho ruido para ser reportado como un simple número.
- Data puede ser propiamente censada:
  - Al límite de cuantificación (Lc).
  - Al límite de detección (Ld).

# Sumario de Límites

- Límite de detección (Ld): debajo podría ser concentración cero.
- Límite de cuantificación (Ld): debajo esta la data con mucho ruido para ser reportado como un simple número.
- Data puede ser propiamente censada:
  - Al límite de cuantificación (Lc).
  - Al límite de detección (Ld).
  - Dentro de dos intervalos (0 a Ld) o (Ld a Lc).

# Sumario de Límites

- Límite de detección ( $L_d$ ): debajo podría ser concentración cero.
- Límite de cuantificación ( $L_d$ ): debajo esta la data con mucho ruido para ser reportado como un simple número.
- Data puede ser propiamente censada:
  - Al límite de cuantificación ( $L_c$ ).
  - Al límite de detección ( $L_d$ ).
  - Dentro de dos intervaloes (0 a  $L_d$ ) o ( $L_d$  a  $L_c$ ).
- No usar censado interno, o la data será sesgada hacia arriba con una incorrecta *cummulative distribution function (cdf)*. Si es usado por el laboratorio, re-censorear toda la data al  $L_c$ .