Istituto Zooprofilattico Sperimentale delle Venezie

Statistical methods to estimate the assigned value in presence of multiple censored results

Marzia Mancin, Lisa Barco, Antonia Ricci

Marzia Mancin
Statistician

Food Safety Department- Risk Analysis and Public Health Surveillance
Istituto Zooprofilattico Sperimentale delle Venezie
mmancin@izsvenezie.it

Eurachem: Portoroz, 12 October 2017

What is a censored result?

A censored result is a value partially know because it is **below** (<, **left censoring**) or **above** (>, **right censoring**) a certain value

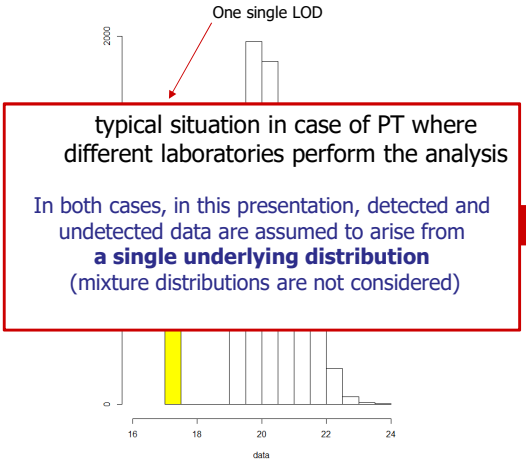
Left censored result

- Left censored result is a measurement below the field limit of detection (LOD) or quantification (LOQ) (<0.1, <10,…)
- Left censored values are usually associated with limitations of measurement and are commonly reported as results below measurement capacity of the available analytical equipment
- The true values of the left censored observations are considered to lie between zero and censor limit (CL)
- Chemical contaminants: example of typical context of censored results

One or multiple censoring level

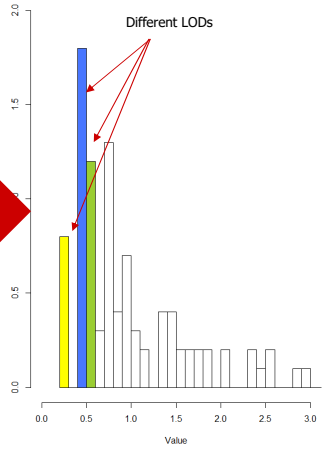
One censoring level: one single value of LOD (<0.01) in the analysed data set

Multiple censoring level: multiple LODs in the analysed data set (<0.01, <0.1, <0.3….)



typical situation in case of PT where different laboratories perform the analysis

In both cases, in this presentation, detected and undetected data are assumed to arise from **a single underlying distribution** (mixture distributions are not considered)



Annex E
(informative)

Illustrative Examples

INTERNATIONAL STANDARD **ISO 13528**

Statistical methods for use in proficiency testing by interlaboratory comparison

These examples are intended to illustrate the procedures specified in this Standard, so the reader can determine that their calculations are correct. Specific examples should not be considered to be recommendations for use in particular proficiency testing schemes.

More than 20% of censored data

E.1 Effect of censored values (section 5.5.3.3)

Table E.1 shows 23 results for a round of a proficiency testing scheme, of which 5 results are indicated as 'Less Than' some amount. The robust mean (x^*) and standard deviation (s^*) from Algorithm A are shown for 3 different calculations, where the '<' signs are discarded and data analysed as quantitative data; the results with '<' values are ignored; and where 0,5 times the result is inserted as an estimate of the quantitative result. In each scenario the results that would have been outside the acceptance limit are indicated with '#'. This assumes that the evaluation would be 'unacceptable' (action signal) for any result where the quantitative part is outside the $x^* \pm 3s^*$. The proficiency testing provider could have alternative rules for evaluating results with '<' or '>' signs.

Table E.1 — Sample dataset with truncated (<) results, and three options for accommodating results

Participant	Result	'<' ignored	'< ' deleted	0,5 x '<' value
A	<10	10	--	5
B	<10	10	--	5
C	12	12	12	12
D	19	19	19	19
E	<20	20	--	10
F	20	20	20	20
G	23	23	23	23

ISO 13528:2015

Participant	Result	'<' ignored	'< ' deleted	0,5 x '<' value
Y	45	45	45 #	45
Z	<50	50 #	--	25
Summary				
Number of Results	23	23	18	23
x^*		26,01	26,81	23,95
s^*		7,23	5,29	8,60

The choice of how to handle the "less than" samples has a significant effect on the robust mean and standard deviation, and on the performance evaluation. The proficiency testing provider is expected to determine an appropriate method.

$$u\left(x_{pt}\right) = 1,25 \times \frac{s^*}{\sqrt{p}}$$

1.884449	1.558581	2.24153

5.5.3.2 When consensus statistics are used, it may not be possible to evaluate performance if the number of censored values is large enough that a robust method is affected by the censoring. In circumstances where the number of censored results is sufficient to affect a robust method, then the results should be evaluated using statistical methods which allow unbiased estimation in the presence of censored data[22], or the results should not be evaluated. When in doubt about the effect of the procedure chosen, the proficiency testing provider should calculate summary statistics and performance evaluations with each of the alternative statistical procedures considered potentially applicable in the circumstances, and investigate the importance of any difference(s).

Most popular statistical methods for censored results

- Substitution
- Kaplan Meier (KM): Non parametric method
- Robust regression on order statistics (ROS): semi-parametric method
- Maximum likelihood estimation (MLE): parametric method

Preliminary analysis of ISO 13528 data

- Identification of distinct LOD values

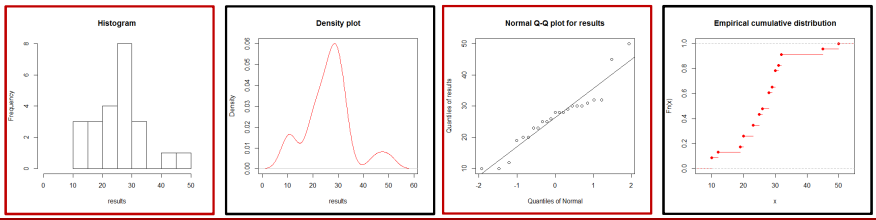
values			
10	20	30	50

4 different points of censoring (multicensored)

- Calculation of the proportion of undetected data in the dataset

n	n.cen	pct.cen	min	max
23.00	5.00	21.74	10.00	50.00

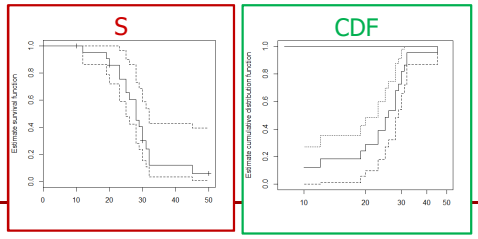
- Graphical display of the data (histogram, density plot, qqplot, cumulative distribution)



Kaplan Meier method (KM)

KM

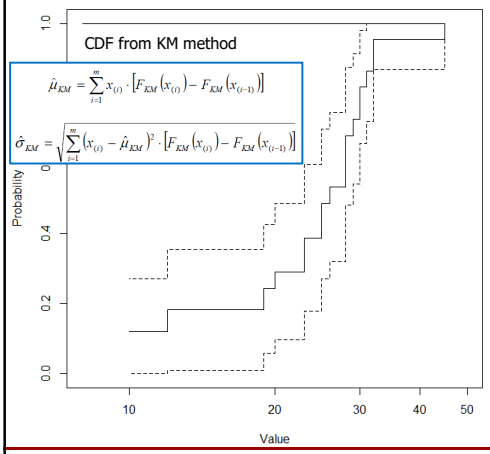
- KM is the historical standard non parametric technique for censored data
- KM assumes that all detected and undetected data arise from the same population but it does not require the assumption of a specific distribution to estimate summary statistics for multiple censored dataset
- KM was originally developed for estimating the mean of right-censored (>) survival data (S) in medical research
- KM estimates the cumulative distribution function (CDF) by accounting for the left-censored (<) observations at each concentration



S=1-CDF

KM: Empirical cumulative distribution function of ISO 13528 data

KM estimates **the approximate proportion of concentrations below each observed level by sorting the distinct sample values, although the exact concentration of undetected data is unknown**. The Kaplan Meier CDF allows to estimate the population mean and standard deviation



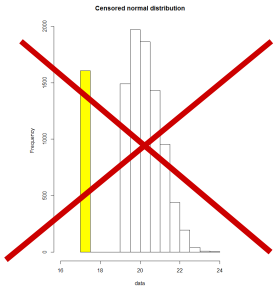
$$F_{KM}(x_{(i)}) = \Pr(X < x_{(i)}) = 1 - \Pr(X \geq x_{(i)}) \prod_{k=i}^{m} \left(1 - \frac{d_k}{n_k}\right)$$

obs	n.risk	n.event	prob
10	2	0	0.1212121
12	3	1	0.1212121
19	4	1	0.1818182
20	6	1	0.2424242
23	8	2	0.2909091
25	10	2	0.3878788
26	11	1	0.4848485
28	14	3	0.5333333
29	15	1	0.6787879
30	18	2	0.7272727
31	19	1	0.8181818
32	21	2	0.8636364
45	22	1	0.9545455
50	23	0	1.0000000

Kaplan Meier method (KM)

KM

- KM tends to be insensitive to outliers
- KM is recommended for smaller data sets (n<50) **with less than 50% censoring and multiple censoring level**
- KM **IS NOT recommended** for data sets **with only one censoring level** (smaller than the all uncensored values)

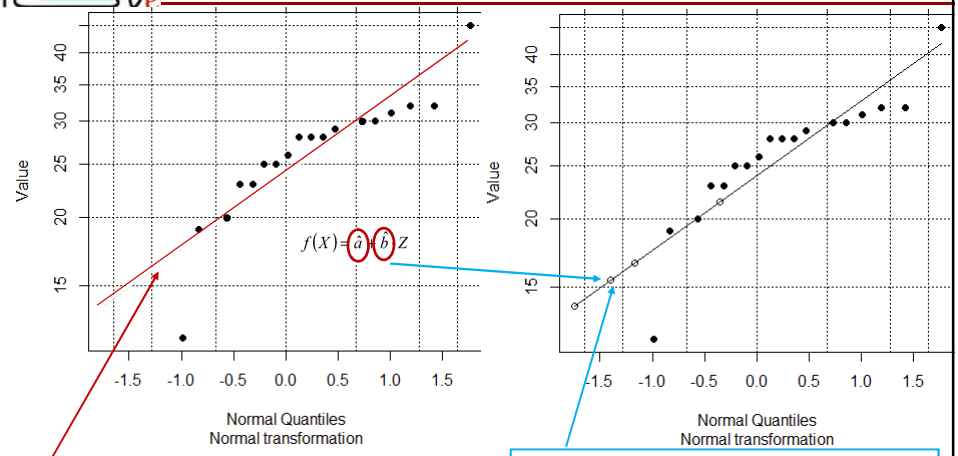


Robust regression on order statistics (ROS)

ROS

- ROS is a semiparametric method for censored data: it assumes an underlying parametric distribution **for the uncensored values**
- ROS method is based on a simple linear regression model using ordered detected values and distributional (normal or log normal) quantiles to estimate the concentration of the censored values
- ROS is a probability plotting and regression procedure that imputes the censored data using the estimate parameters of a linear regression model of uncensored observed values vs their normal quantiles (or log normal quantile)

ROS: Censored probability plot of ISO 13528 data



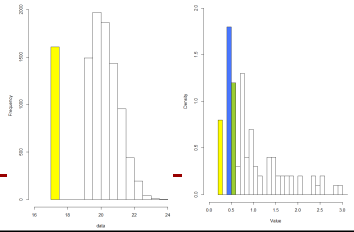
ROS estimates the parameters of a linear regression model of uncensored observed values vs their normal quantiles

ROS imputes the censored data using the estimate parameters (a and b) and estimate the overall sample mean and sample standard deviation

Robust regression on order statistics (ROS)

ROS

- The required assumption is that the **response variable is a linear function of the normal (lognormal) quantiles**
- The imputated values **are only used collectively to estimate summary statistics and they are not considered estimates for specific samples**
- ROS is recommended for **large (n>50)** data sets **with less than 50% censoring and multiple censoring level**
- ROS is recommended for **small (n<50)** data sets **with less than 80% censoring and multiple censoring level**
- ROS **can be used** for data sets **with only one censoring level**



Maximum likelihood estimation (MLE)

MLE

- MLE is a parametric technique and the best approach from a methodological perspective
- Uncensored and censored data are assumed to follow a given (known) statistical distribution such as Normal, Log-normal, Weibull, Gamma…
- The sample parameters are those estimates that **maximise the likelihood function** after the definition of a parametric distribution for the data
- MLE uses **uncesored observations, the proportion of censored observations and a distributional assumption** to compute estimates of summary statistics

Standard Likelihood function

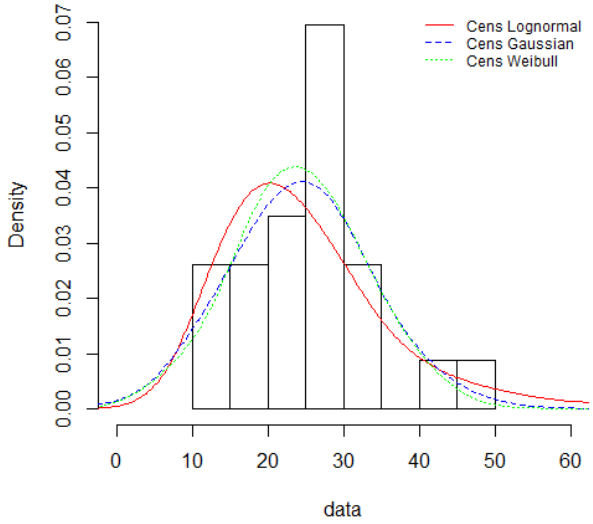
$$L(\vartheta) = \prod_{i=1}^{n} f(y_i; \vartheta)$$

Likelihood function with left censored data

$$L(\vartheta) = \prod_{i=1}^{n} (f(y_i; \vartheta))^{\delta_i} (F(y_i; \vartheta))^{1-\delta_i}$$

$f(y_i; \vartheta)$= Density function
$F(y_i; \vartheta)$= Cumulative distribution function
δ_i = Censoring indicator

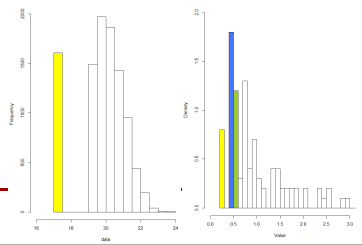
Density plot comparison of parametric distributions considering the ISO 13528 data



Maximum likelihood estimation (MLE)

MLE

- If the assumption regarding the data distribution is realistic, this method is accurate
- MLE is recommended for **large (n>50)** data sets **with less than 80% censoring and multiple censoring level**
- For some authors, MLE is suggested also for **small (n<50)** data sets **with less than 50% censoring and multiple censoring level**
- MLE **can be used** for data sets **with only one censoring level**



Which are the results obtained applying these different methods to ISO 13528 data?

Summary results: Mean, standard deviation, uncertainty

- Number of results<50
- % of censored results:21.74

→ All statistical methods can be used: **KM, ROS and MLE**

	Sub. with CL	Sub. with 0.5*CL	KM	ROS	Lognormal	Weibull	Gaussian
mean	26.3478	23.7391	24.7878	24.88	24.7487	24.3015	24.2091
u_m	1.9627	1.9653	1.7770	1.8813	1.8910	1.8868	1.9838
sd	9.4128	9.4254	8.52	7.8949	10.9744	8.5676	9.0233
u_{sd}	1.6219	1.5122	1.8214	1.9141	2.4632	1.7868	2.0439

*Uncertainty values are obtained using bootstrap method with N=100,000 bootstrap replicates (pakage boot of R software)
*Pakage NADA of R software provides KM, ROS and MLE method (MLE only for normal and log-normal distribution)
*Packages Survreg of R software provides MLE methos for censored data for every distribution (weibull, gamma, normal…)

Which is the best statistical method?

rMSE: root mean square error

rMSE is considered a measure of overall imprecision or overall accuracy
rMSE is obtained from the combination of **bias** and **precision**

Bias= difference between the mean of 100,000 parameter estimates and the true value

$$Bias = (\bar{x} - \vartheta)$$

Precision=variance of the 100,000 parameter estimates

$$Precision = \frac{1}{G-1}\sum_{g=1}^{G}(x_g - \bar{x})^2$$

$$rMSE = \sqrt{(\bar{x} - \vartheta)^2 + \frac{1}{G-1}\sum_{g=1}^{G}(x_g - \bar{x})^2}$$

Precision of methods: rMSE%

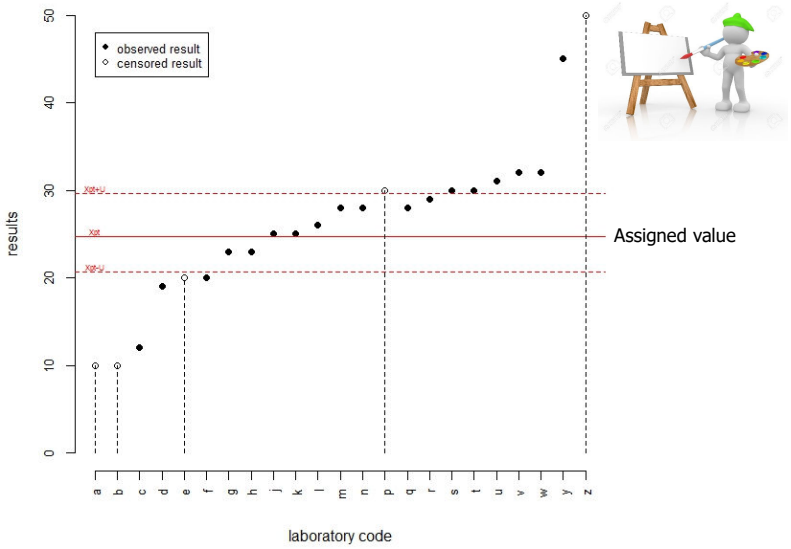
	KM	ROS	Lognormal	Weibull	Gaussian
mean	24.787	24.88	24.748	24.301	24.209
u_m	1.823	1.975	1.857	1.910	2.117
sd	8.522	7.894	10.974	8.567	9.023
u_{sd}	1.778	1.884	2.398	1.726	1.885
rMSE%	7.655	7.756	7.525	7.848	8.708

$$rMSE\% = \frac{rMES}{\vartheta}\%$$

IF rMSE is not available? Use common sense

- Are information or assumptions about the data distribution available? MLE
- Aren't? KM or ROS

Representation of participants' results



Conclusions

- The aim of statistical methods for censored data is to estimate the assigned value, standard deviation and uncertaity of the assigned value, taking into account that a percentage of data is not detected
- No method **MUST BE** used to estimate the values of censored data but only to obtain summary statistics
- No single method is unequivocally more suitable across different scenarios (one or multiple censoring, % of censored data, knowledge about the data distribution)
- No method is recommended for data sets with more than 80% censoring (independently of the sample size)
- Statistical competence has to be integrated with qualified knowledge of the type of data to analyse.

Remark of a statistician

Statistic science provides mathematical solutions, as in the critical case of a high number of censored data (50-80%).

However, I believe that in the context of PTs, the PT provider must evaluate if **he wants** to obtain an assigned value from participants results to evaluate them, when **the majority of the participants have not a value to evaluate**!!

THANK YOU FOR YOUR ATTENTION

"All models are wrong, but some models are good approximations of the truth"

Substitution methods

- Method widely used in different disciplines because it is easy to implement
- Replacement of censored observation with a value between 0 and censor limit
- Common values for substitution are 0, 0.5, square root of the censoring limit and the censoring limit itself
- Any single value between 0 and censoring limit is arguably as good as another (it is not a rule or a preferable value or real theoretical justification): to use the uniform distribution with range [0; CL] to replace the censored data

BUT

- The simple substitution produces biased estimates of summary statistics that are dependent on the value being substituted especially when **the number of censored data is not negligible (>15%)**
- When the censored limit is chosen as value of substitution, an **over-estimation of the mean and an under estimation of the variability** will occur

Why to have an interest in censored results?

- Statistician at the Food Safety Department- Risk Analysis and Public Health Unit
- Involved in risk assessment related to the consumption of food matrices contaminated by a given chemical substance
- A high number of left censored data at different levels of censoring
- To investigate the statistical methods to treat the left censored data

To share the knowledge, applicable in the context of PTs, where different laboratories with different LOD take part (Example E1 Annex E, ISO 13528:2015)

Data from ISO 13528:2015 (Example E1)

	1	2	3	4	5	6	7	....	16	17	18	19	20	21	22	23
results	<10	<10	12	19	20	<20	23	....	<30	30	30	31	32	32	45	<50
results	10	10	12	19	20	20	23	.....	30	30	30	31	32	32	45	50
censored	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	.....	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE

Indicator variable for censored data

- Kaplan Meier (KM)
- Robust regression on order statistics (ROS)
- Maximum likelihood estimation (MLE):

Data from ISO 13528:2015

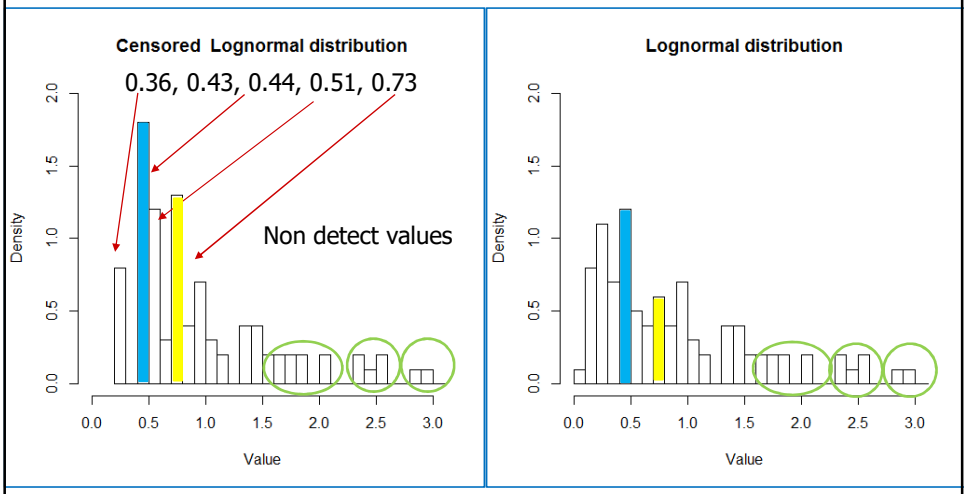
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23
results	<10	<10	12	19	20	<20	23	23	25	25	26	28	28	28	29	<30	30	30	31	32	32	45	<50
results	10	10	12	19	20	20	23	23	25	25	26	28	28	28	29	30	30	30	31	32	32	45	50
censored	TRUE	TRUE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE	FALSE	FALSE	FALSE	FALSE	FALSE	FALSE	TRUE

4 different points of censura (multicensored)

5 censored results

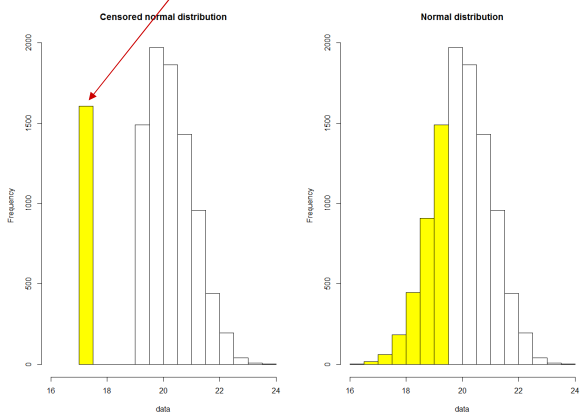
value	observed	N censored
10	0	2
20	1	1
30	2	1
50	0	1
		5

Simulated distribution



One or multiple censoring levels

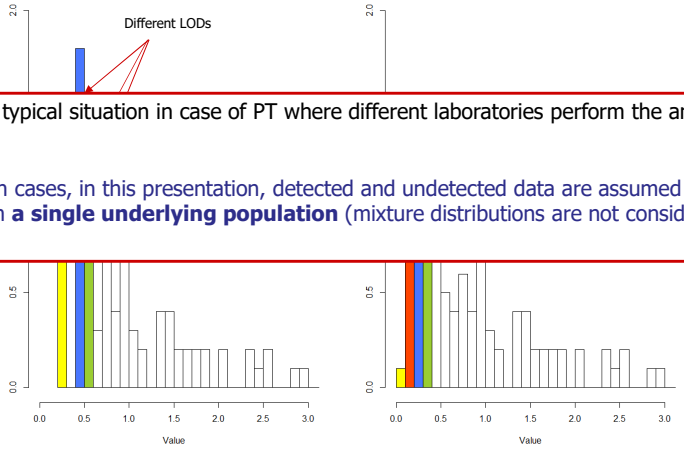
- **One censoring level**: one single value of LOD (<0.01) in the analysed data set



- **Multiple censoring level**: multiple LOD in the analysed data set (<0.01, <0.1, <0.3….)

One or multiple censoring level

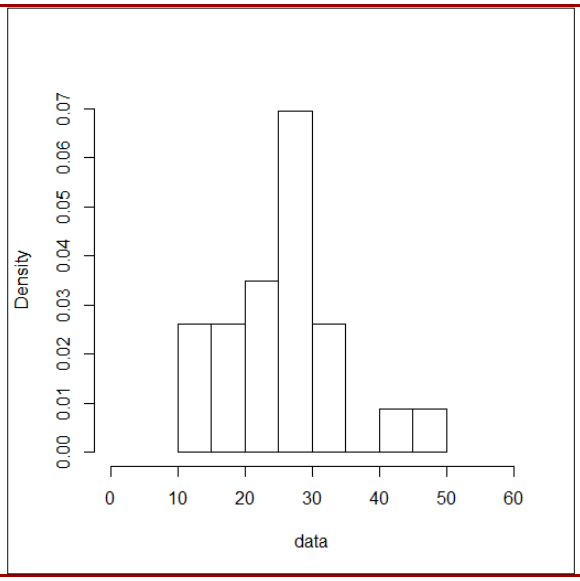
- **One censoring level**: one single value of LOD (<0.01) in the analysed data set
- **Multiple censoring level**: multiple LODs in the analysed data set (<0.01, <0.1, <0.3….)



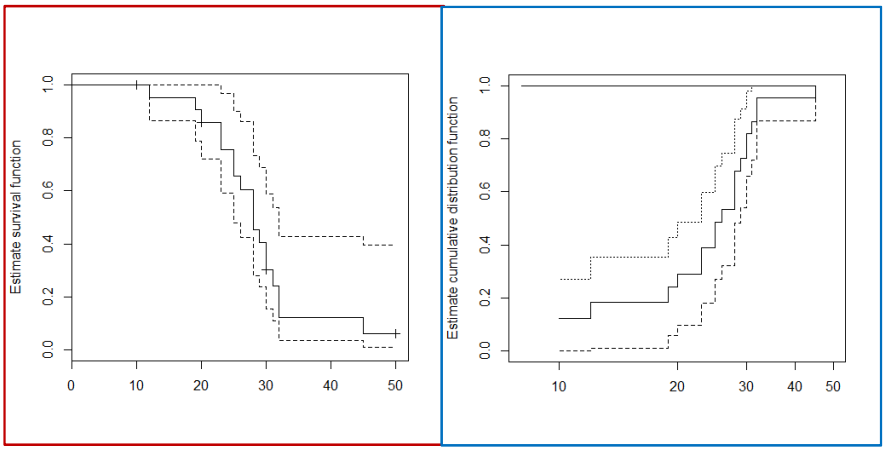
typical situation in case of PT where different laboratories perform the analysis

In both cases, in this presentation, detected and undetected data are assumed to arise from **a single underlying population** (mixture distributions are not considered)

Histogram of data distribution



Statistical methods for censored results

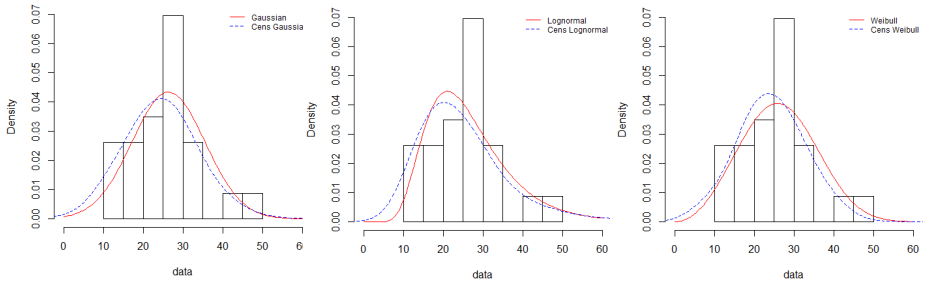


- KM estimates the cumulative distribution function (CDF) by accounting for the censored observations at each concentration

Statistical methods

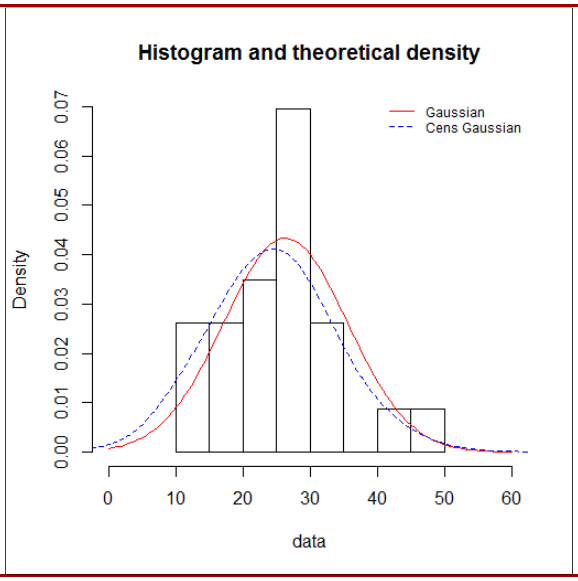
- Parametric method:
  - the data follow some known distribution, such as the Normal, Log-normal, Weibull, Gamma…
  - This assumption includes data both above and below the censor limit
  - Estimates are obtained using maximum likelihood methods (MLE), where the uncesored observations, the proportion of censored observations and a distribution are considered
  - If the assumption regarding the data distribution is realistic, this method is accurated
- Non parametric method:
  - Non parametric methods are a good alternative analysis method when the assumption about the distribution for parametric methods are not meet
  - These methods focus on the ordering of the data rather than on its distribution
  - The assumptions for this method are fewer, and it handles very etreme values better than parametric method
- Semi parametric method
  - They combine some of the best features of parametric estimation with some of the best features of non parametric estimation

Density plot of ISO 13528 data: Gaussian, Lognormal and Weibull distributions

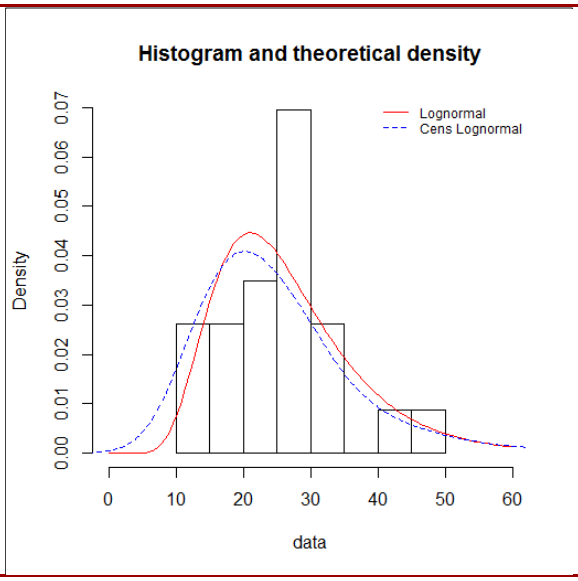


A shift of the distribution to the left is evident if the information about the censoring is taken into account

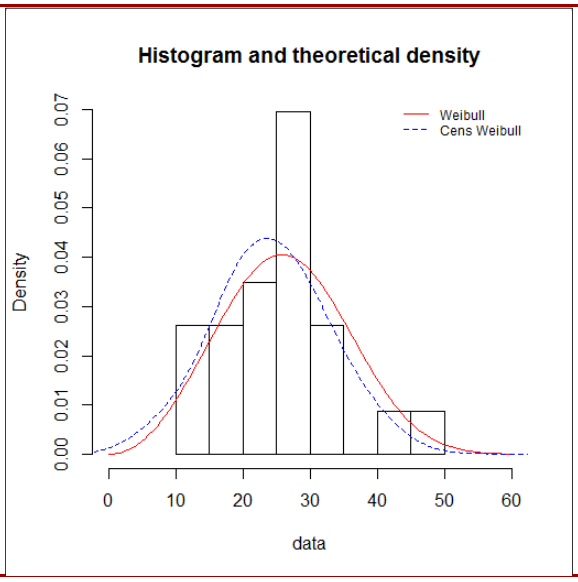
Density plot of Gaussian distribution



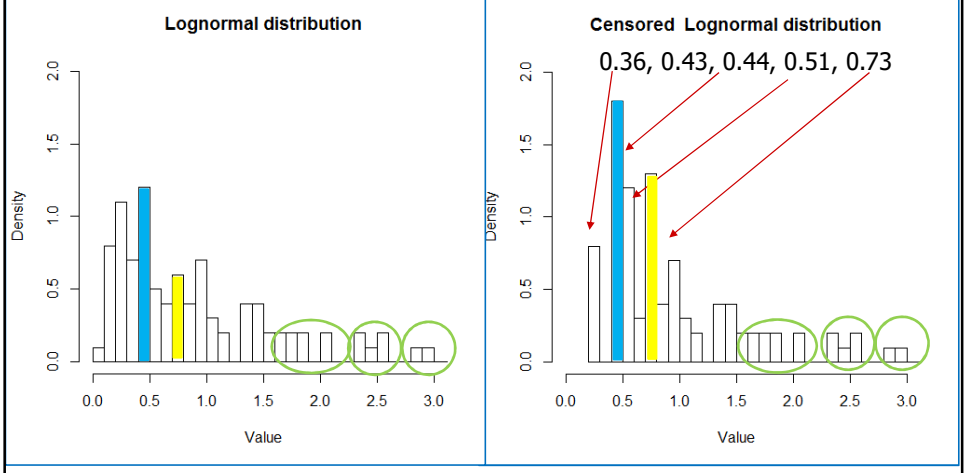
Density plot of Lognormal distribution



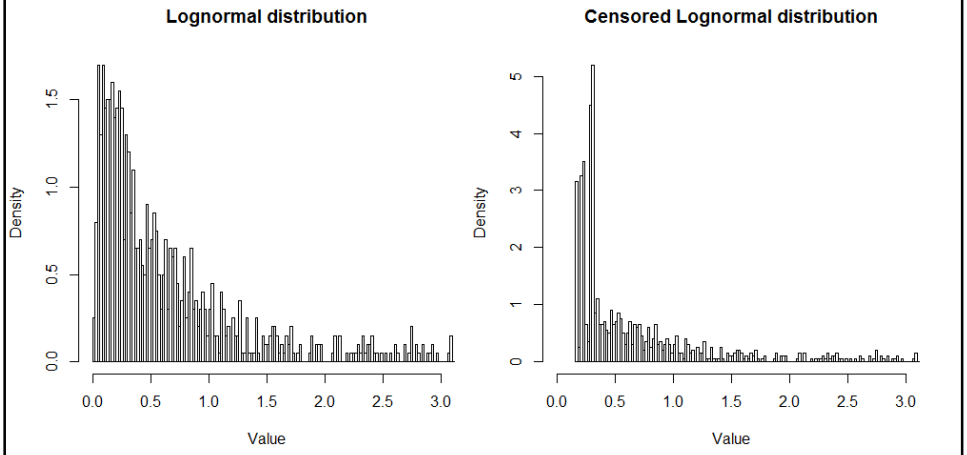
Density plot of Weibull distribution



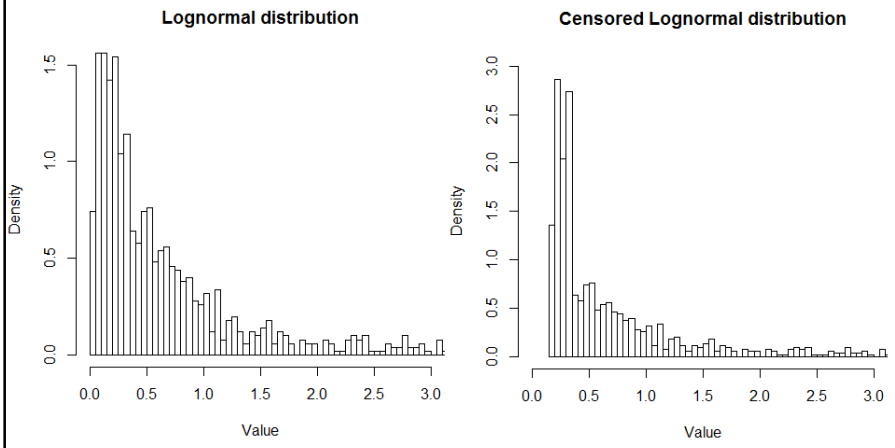
Simulated distribution

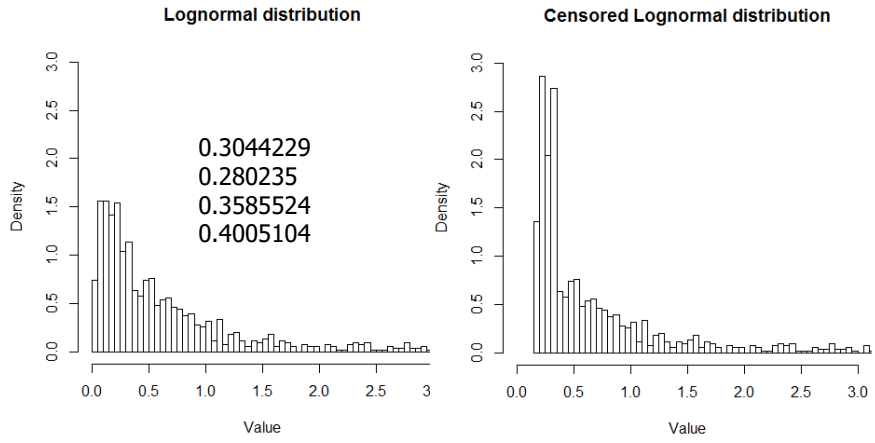


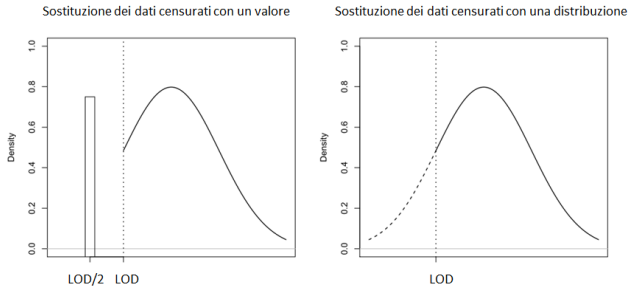
MLE: Density plot of parametric distributions



MLE: Density plot of parametric distributions







Robust regression on order statistics (ROS) is a semi-parametric method that can be used to estimate means and other statistics with censored data. Unlike Kaplan-Meier, ROS internally assumes that the underlying population is approximately normal or lognormalA dataset that is not normally distributed (symmetric bell-shaped curve) but that can be transformed using a natural logarithm so that the data set can be evaluated using a normal-theory test (Unified Guidance).. However, the assumption is directly applied to only the censored measurements and not to the full data set (hence the term 'semi-parametric'). In particular, ROS plots the detected values on a probability plot (with a regular or log-transformed axis) and calculates a linear regression line in order to approximate the parameters of the underlying (assumed) distribution. This fitted distribution is then utilized to generate imputed estimates for each of the censored measurements, which are then combined with the known (detected) values to summary statistics of interest (e.g., mean, variance). The method is labeled 'robust' because the detected measurements are used 'as is' to make estimates, rather than simply using the fitted distributional parameters from the probability plot
http://www.itrcweb.org/gsmc-1/Content/GW%20Stats/5%20Methods%20in%20indiv%20Topics/5%207%20Nondetects.htm