Package 'STAND'

September 25, 2015

Type Package

Title Statistical Analysis of Non-Detects

Version 2.0

Date 2015-09-10

Author E. L. Frome <fromeEL@ornl.gov> and D. P. Frome

Maintainer E. P. Adams <eric.adams@orau.org>

Description Provides functions for the analysis of
occupational and environmental data with non-detects. Maximum
likelihood (ML) methods for censored log-normal data and
non-parametric methods based on the product limit estimate (PLE)
for left censored data are used to calculate all of the
statistics recommended by the American Industrial Hygiene
Association (AIHA) for the complete data case. Functions for
the analysis of complete samples using exact methods are also
provided for the lognormal model. Revised from 2007-11-05
'survfit~1'.

License GPL (>= 2)

URL http://www.csm.ornl.gov/esh/statoed/

Depends survival

NeedsCompilation no

Repository CRAN

Date/Publication 2015-09-25 17:03:51

R topics documented:

About-STAND   *Statistical Analysis of Non-detect Data*

Description

Environmental exposure measurements are, in general, positive and may be subject to left censoring; i.e., the measured value is less than a "detection limit", and is referred to as a non-detect or "less than" value. This package calculates the censored data equivalent of a number of statistics that are used in the analysis of environmental data that do not contain non-detects, i.e. the usual complete data case.

Details

Package:	stand
Type:	Package
Version:	2.0
Date:	2015-09-10
License:	GPL version 2 or newer

In occupational monitoring, strategies for assessing workplace exposures typically focus on the mean exposure level or the probability that any measurement exceeds a limit. Parametric methods used to determine acceptable levels of exposure are based on a two parameter lognormal distribution. The mean exposure level, an upper percentile, the exceedance fraction, and confidence limits

for each of these statistics are calculated. Statistical methods for random samples (without non-detects) from the lognormal distribution are well known for each of these situations–see e.g., Lyles and Kupper (1996). In this package the maximum likelihood method for randomly left censored lognormal data is used to calculate these statistics, and graphical methods are provided to evaluate the lognormal assumption. Nonparametric methods based on the product limit estimate for left censored data are used to estimate the mean exposure level, and the upper confidence limit on an upper percentile (i.e., the upper tolerance limit) is obtained using a nonparametric approach.

The American Industrial Hygiene Association (AIHA) has published a consensus standard with two basic strategies for evaluating an exposure profile—see Mulhausen and Damiano(1998), Ignacio and Bullock (2006). Most of the AIHA methods are based on the assumptions that the exposure data does not contain non-detects, and that a lognormal distribution can be used to describe the data. Exposure monitoring results from a compliant workplace tend to contain a high percentage of non-detected results when the detection limit is close to the exposure limit, and in some situations, the lognormal assumption may not be reasonable. The function `IH.summary` calculates most of the statistics proposed by AIHA for exposure data with non-detects. All of the methods are described in the report Frome and Wambach (2005).

Note

Throughout this document and the online help files the greek letter γ is used to represent the confidence level for a one-sided confidence limit (default value 0.95). This is represented by `gam` or `gamma` in the argument list and value of functions that compute confidence limits.

References

Aitchison, J. and J. A. C. Brown (1969), *The Lognormal Distribution*, Cambridge, U.K., Cambridge University Press.

Akritas, M. G., T. F. Ruscitti, and G. S. Patil (1994), "Statistical Analysis of Censored Environmental Data," *Handbook of Statistics*, Vol. 12, G. P. Patil and C. R. Rao (eds), 221-242, Elsevier Science, New York.

American Conference of Governmental Industrial Hygienists (ACGIH) (2004), "Notice of Intended Change In: 2004 TLVs and BEIs," *ACGIH*, p. 60, Cincinnati, OH.

Burrows, G. L. (1963), "Statistical Tolerance Limits - What are They," *Applied Statistics*, 12, 133-144.

Armstrong, B. G. (1992), "Confidence Intervals for Arithmetic Means of Lognormally Distributed Exposures," *American Industrial Hygiene Association Journal*, 53(8), 481-485.

Chambers, J. M., W. S. Cleveland, B. Kleiner, and P. A. Tukey (1983), *Graphical Methods for Data Analysis*, Duxbury Press, Boston.

Clopper, C. J. and Pearson, E. S. (1934), "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial," *Biometrika*, 26, 404-413.

Cohen, A. C. (1991), *Truncated and Censored Samples*, Marcel Dekker, Inc., New York.

Crow, E. L. and K. Shimizu (1988), *Lognormal Distribution*, Marcel Decker, New York.

Cox, D. R. and D. V. Hinkley (1979), *Theoretical Statistics*, Chapman and Hall, New York.

Cox, D. R. and D. Oakes (1984), *Analysis of Survival Data*, Chapman and Hall, New York.

Department of Energy (December, 1999), "Chronic Beryllium Disease Prevention Program, Federal Register," *10 CFR Part 850*, volume 64, number 235, 68854-68914.

Fowlkes, E. B. (1979), "Some Methods for Studying the Mixture of Two Normal (Lognormal) Distributions," *Journal of the American Statistical Association*, 74, 561-575.

Frome, E. L. and Wambach, P. F. (2005), "Statistical Methods and Software for the Analysis of Occupational Exposure Data with Non-Detectable Values," *ORNL/TM-2005/52,Oak Ridge National Laboratory*, Oak Ridge, TN 37830. Available at: http://www.csm.ornl.gov/esh/aoed/ORNLTM2005-52.pdf

Hahn, G. J. and W. Q. Meeker (1991), *Statistical Intervals*, John Wiley and Sons, New York.

Hewett, P. and G. H. Ganser, (1997), "Simple Procedures for Calculating Confidence Intervals Around the Sample Mean and Exceedance Fraction Derived from Lognormally Distributed Data," *Applied Occupational and Environmental Hygiene*, 12(2), 132-147.

Helsel, D. (1990), "Less Than Obvious: Statistical Treatment of Date Below the Detection Limit," *Environmental Science and Technology*, 24(12), 1767-1774.

Hesel, D. R. and T. A. Cohn (1988), "Estimation of Descriptive Statistics for Multiply Censored Water Quality Data," *Water Resources Research*, 24, 1997-2004.

Johnson, N. L. and B. L. Welch (1940), "Application of the Non-Central t-Distribution," *Biometrika*, 31(3/4), 362-389.

Ignacio, J. S. and W. H. Bullock (2006), *A Strategy for Assessing and Managing Occupational Exposures*, Third Edition, AIHA Press, Fairfax, VA.

Kalbfleisch, J. D. and R. L. Prentice (1980), *The Statistical Analysis of Failure Time Data*, John Wiley and Sons, New York.

Kaplan, E. L. and Meir, P. (1958), "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, 457-481.

Land, C. E. (1972), "An Evaluation of Approximate Confidence Interval Estimation Methods for Lognormal Means," *Technometrics*, 14(1), 145-158.

Lyles R. H. and L. L. Kupper (1996), "On Strategies for Comparing Occupational Exposure Data to Limits," *American Industrial Hygiene Association Journal*, 57:6-15.

Meeker, W. Q. and L. A. Escobar (1998), *Statistical Methods for Reliability Data*, John Wiley and Sons, New York.

Moulton, L. H. and N. A. Halsey (1995), "A Mixture Model with Detection Limits for Regression Analysis of Antibody Response on Vaccine," *Biometrics*, 51, 1570-1578.

Mulhausen, J. R. and J. Damiano (1998), *A Strategy for Assessing and Managing Occupational Exposures*, Second Edition, AIHA Press, Fairfax, VA.

Neuman, M. C., P. M. Dixon, B. B. Looney, and J. E. Pinder (1989), "Estimating Mean and Variance for Environmental Samples with Below Detection Limit Observations," *Water Resources Bulletin*, 25, 905-916.

Ng, M. P. (2002), "A Modification of Peto's Nonparametric Estimation of Survival Curves for Interval-Censored Data," *Biometrics*, volume 58, number 2, pp. 439-442.

Odeh, R. E. and D. B. Owen (1980), *Tables for Normal Tolerance Limits, Sampling Plans, and Screening*, Marcel Deker, New York.

Peto, R. (1973), "Experimental Survival Curves for Interval-censored Data," *Applied Statistics*, volume 22, number 1, pp. 86-91.

Schmee, J., D. Gladstein, and W. Nelson (1985), "Confidence Limits for Parameters of a Normal Distribution From Singly Censored Samples, Using Maximum Likelihood," *Technometrics*, 27, 119-128.

Schmoyer, R. L., J. J. Beauchamp, C. C. Brandt and F. O. Hoffman, Jr. (1996), "Difficulties with the Lognormal Model in Mean Estimation and Testing," *Environmental and Ecological Statistics*, 3, 81-97.

Sommerville, P. N. (1958), "Tables for Obtaining Non-Parametric Confidence Limits," *Annals of Mathematical Statistics*, 29, 599-601.

Taylor, D. J., L. L. Kupper, S. M. Rappaport, and R. H. Lyles (2001), "A Mixture Model for Occupational Exposure Mean Testing with a Limit of Detection," *Biometrics*, 57, 681-688.

Tuggle, R. M. (1982), "Assessment of Occupational Exposure Using One-Sided Tolerance Limits," *American Industrial Hygiene Association Journal*, 43, 338-346.

Turnbull, B. W. (1976), "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data," *Journal of the Royal Statistical Society*, Series B (Methodological), 38(3), 290-295.

Venables, W. N. and B. D. Ripley (2002), *Modern Applied Statistics with S*, 4th edition. Springer-Verlag, New York.

Verrill, S. and R. A. Johnson (1998), "Tables and Large-Sample Distribution Theory for Censored-Data Correlation Statistics for Testing Normality," *Journal of the American Statistical Association*, 83(404), 1192-1197.

Waller, L. A., and B. W. Turnbull, (1992), "Probability Plotting with Censored Data," *The American Statistician*, 46(1), 5-12.

Examples

```
# Example 1 from Frome and Wambach (2005) ORNL/TM-2005/52
# NOTE THAT FUNCTIONS NAMES AND DETAILS HAVE BEEN REVISED IN THIS PACKAGE
# the results are the same. For example lnorm.ml() replaces mlndln().
data(SESdata)
mle<-lnorm.ml(SESdata)
unlist(mle[1:4])            # ML estimates mu sigma E(X) and sigma^2
unlist(mle[5:8])            # ML Estimates of standard errors
unlist(mle[9:13])           # additional  output from ORNL/TM-2005/52
IH.summary(SESdata,L=0.2)   #  All sumarry statistics for SESdata
#  lognormal q-q plot for SESdata Figure in ORNL/TM-2005/52
qq.lnorm(plend(SESdata),mle$mu,mle$sigma)
title("SESdata: Smelter-Elevated Surfaces")
```

---

aihand   *Industrial Hygiene Air Monitoring Data*

Description

Data from Mulhausen and Damiano Appendix V is used to illustrate data with "less-than" values (non-detects).

Usage

```
data(aihand)
```

Format

A data frame with 15 observations on the following 3 variables

xnd air monitoring data with 3 non-detects

det 0 if non-detect, 1 if detect

x air monitoring data (mg/m^3)

Details

The data in column 1 was obtained from the data in column 3 by assuming a limit of detection of 1.9. The original data in column 3 is used as an example in Appendices V, VI, and VII (see Tables V.1, V.5, V.6, VI.2) to illustrate methods of analysis when there are no non-detects.

Source

Table V.2 page 244 in Mulhausen and Damiano (1998) and Table IV.3 page 349 in Ignacio and Bullock (2006)

References

Ignacio, J. S. and W. H. Bullock (2006), *A Strategy for Assessing and Managing Occupational Exposures*, Third Edition, AIHA Press, Fairfax, VA.

Mulhausen, J. R. and J. Damiano (1998), *A Strategy for Assessing and Managing Occupational Exposures*, Second Edition, AIHA Press, Fairfax, VA.

Examples

```
data(aihand) # Table V.2 Mulhausen and Damiano (1998)

# calculate summary statistics for non-detect data
ndt <- IH.summary(aihand,L=5)
x <- aihand$x ;  det <- rep(1,length(x))
aiha<-data.frame(x,det) #  complete data
# calculate summary statistics for complete data
cdt <- IH.summary(aiha,L=5)
# output results in a table
round(cbind(cdt,ndt),3)

#  rm(aiha,aihand,cdt,det,ndt)
# add results for exact method for complete data case
# for Xp and the exceedance fraction
```

---

beTWA   *TWA Beryllium Exposure Data*

Description

As part of a chronic disease prevention program the DOE adopted an 8-hour time-weighted average (TWA) occupational exposure limit (OEL) value of 0.2 $\mu g/m^3$ proposed by the American Conference of Government Industrial Hygienists (*DOE 10 CRF Part 850* and *ACGIH* 2004).

Usage

```
data(beTWA)
```

Format

A data frame with 280 observations on the following 2 variables:

twa 8-hour TWA Beryllium exposure $\mu g/m^3$
det 0 if non-detect; 1 if detect

Details

The beTWA data set is the results of 280 personal 8-hour TWA beryllium exposure readings at a DOE facility. This data contains 175 non-detects that range in value from 0.005 to 0.1 $\mu g/m^3$. A detailed description and analysis of this data is given as Example 2 in Section 4 of Frome and Wambach (2005).

References

Frome, E. L. and Wambach, P. F. (2005), "Statistical Methods and Software for the Analysis of Occupational Exposure Data with Non-Detectable Values," *ORNL/TM-2005/52,Oak Ridge National Laboratory*, Oak Ridge , TN 37830 Available at: http://www.csm.ornl.gov/esh/aoed/ORNLTM2005-52.pdf

Examples

```
data(beTWA)
##  calculate all of the summary statistics described in
##  Example 2 in Section 4 of Frome and Wambach (2005)
round( IH.summary(beTWA,L=0.2), 3)
```

---

cansdata   *Container Data Used To Evaluate Beryllium Surface Contamination*

Description

Surface wipe samples obtained from containers that are used to ship beryllium components.

Usage

```
data(cansdata)
```

Format

A data frame with 120 observations on the following 4 variables:

x Surface wipe sample $\mu g/100cm^2$

det 1 if detect, 0 if non-detect

strata a factor with levels A and B

sample a factor with levels 1 and 2

Details

In a scoping survey, the investigator decides to divide the survey unit into two strata: A, used recently, and B, not used for several years. The specified limit that is used to determine if the survey unit is contaminated is $L = 0.2\mu g/100cm^2$. An initial sample of n = 30 was obtained from each stratum (sample = 1). The initial survey produced discrepant results that were hard to interpret. A second sample of n = 30 surface wipe samples was obtained from strata A and B. Results below the limit of quantification are reported as non-detects.

References

Frome, E. L. and Wambach, P. F. (2005), " Statistical Methods and Software for the Analysis of Occupational Exposure Data with Non-Detectable Values," *ORNL/TM-2005/52,Oak Ridge National Laboratory*, Oak Ridge, TN 37830. Available at: http://www.csm.ornl.gov/esh/aoed/ORNLTM2005-52.pdf

Examples

```
data(cansdata)
#     subset container data into stratum A and stratum B
sa60 <- cansdata[ cansdata$st=="A",] ; Ia <- "Be Samples From  Stratum A"
sb60 <- cansdata[ cansdata$st=="B",] ; Ib <- "Be Samples From  Stratum B"
mle.sa60 <- unlist(lnorm.ml(sa60))  # MLEs for stratum A
mle.sb60 <- unlist(lnorm.ml(sb60))
#     print MLE for stratum A and B
round( data.frame(mle.sa60,mle.sb60),3)
#
# Q-Q plot for each stratum
par( mfcol=c(1,2) )
qq.lnorm(plend(sa60),mle.sa60[1,2],ylim=c(0.01,1.2),xlim=c(-0.5,2.5),main=Ia )
qq.lnorm(plend(sb60),mle.sb60[1,2],ylim=c(0.01,1.2),xlim=c(-0.5,2.5),main=Ib )
#   list all summary statistics by Strata
round(IH.summary(cansdata,L=0.2,bcol=3),4)
```

---

efclnp   *Nonparametric Confidence Limits for the Exceedance Fraction*

Description

When the distribution function for the X's is not specified a nonparametric approach can be used to estimate the exceedance fraction $FL = Pr[X > L]$ the proportion of measurements that exceed the limit L.

Usage

```
efclnp(dd,gam = 0.95,L)
```

Arguments

dd	An n by 2 matrix or data frame with x (exposure) variable in column 1, and det = 0 for non-detect or 1 for detect in column 2
gam	one-sided confidence level γ. Default is 0.95
L	L is specified limit for the exceedance fraction; e.g., the occupational exposure limit

Details

Given a random sample of size n the number y of nonconforming observations (i.e., y = number of X's that exceed the limit L) is described using the binomial distribution. The point estimate of FL is $fnp = y/n$ and confidence limits are obtained using the method of Clopper and Pearson (1934) (Hahn and Meeker, 1991) and the R documentation for base R function `binom.test`.

Value

A LIST with components:

fnp	nonparametric estimate of exceedance fraction (as percent)
fnp.LCL	is the 100*γ% lower confidence limit for fnp
fnp.UCL	is the 100*γ% upper confidence limit for fnp
L	is specified limit for the exceedance fraction(e.g. OEL)
gam	one-sided confidence level γ. Default is 0.95

Assumptions

All non-detects < L

Note

The estimates of the exceedance fraction and CL's are in percentage units

Author(s)

E. L. Frome

References

Clopper, C. J. and E. S. Pearson (1934), "The Use of Confidence or Fiducial Limits Illustrated in the Case of the Binomial," *Biometrika*, 26, 404-413.

Hahn, G. J. and W. Q. Meeker (1991), *Statistical Intervals*, John Wiley and Sons, New York.

See Also

efraction.ml

Examples

```
#   calculate nonparametric estimate
#   for Example 2 in ORNLTM2005
data(beTWA)
unlist(efclnp(beTWA,L=0.2))
# calculate ML estimates of exceedance fraction and CLs
unlist(efraction.ml(beTWA,L=0.2))
```

---

efraction.exact *Exceedance Fraction and Exact Confidence Limits*

Description

Calculate estimate of the exceedance fraction $FL = Pr[X > L]$ and exact confidence limits for random sample from normal/lognormal distribution. This function should only be used for complete samples.

Usage

```
efraction.exact(x, gam = 0.95, L=NA ,logx=TRUE,wpnt=FALSE)
```

Arguments

x	vector of data values
gam	one-sided confidence level γ. Default is 0.95
L	L is specified limit for the exceedance fraction; e.g., the occupational exposure limit
logx	If TRUE, sample is from lognormal, else normal. Default is TRUE
wpnt	if TRUE, show warning from pnt. Default is $FALSE$

Details

The exceedance fraction represent the proportion of the X's that exceed a given limit Lp. The null hypothesis of interest is $Ho : F \geq Fo = 1-p$; i.e., Fo is the maximum proportion of the population that can exceed the limit Lp. The null hypothesis is rejected if the $100\gamma\%$ UCL for FL is less than Fo , indicating that the exposure profile is acceptable. The type I error rate for this test is less than or equal to $\alpha = 1 - \gamma$.

Value

A LIST with components:

f	estimate of exceedance fraction for lognormal distribution as %
fe.LCL	100*γ% exact lower confidence limit % units
fe.UCL	100*γ% exact upper confidence limit % units
L	L is specified limit for the exceedance fraction, e.g. the occupational exposure limit
gam	one-sided confidence level γ. Default is 0.95
Logx	If TRUE, sample is from lognormal, else normal. Default is TRUE

Note

(fe.LCL, fe.UCL) is an approximate $100(2\gamma - 1)$ percent confidence interval for F. The R function `uniroot` is used to find the noncentrality parameter of noncentral t distribution to calculate CL's for $U = (L - \mu)/\sigma$ where F = pnorm(U). In some versions of R this may cause a warning message. See R bug report RP 9171 full precision was not achieved in 'pnt'. This warning message may occur in `uniroot` calls to `pt` and does not effect the precision of the final result

Author(s)

E. L. Frome

References

Johnson, N. L. and B. L. Welch (1940), "Application of the Non-Central t-Distribution," *Biometrika*, 31(3/4), 362-389.

Lyles, R. H. and L. L. Kupper (1996), "On strategies for comparing occupational exposure data to limits," *American Industrial Hygiene Association Journal*. 57:6-15.

Frome, E. L. and Wambach, P. F. (2005), "Statistical Methods and Software for the Analysis of Occupational Exposure Data with Non-Detectable Values," *ORNL/TM-2005/52,Oak Ridge National Laboratory*, Oak Ridge, TN 37830. Available at: http://www.csm.ornl.gov/esh/aoed/ORNLTM2005-52.pdf

Ignacio, J. S. and W. H. Bullock (2006), *A Strategy for Assesing and Managing Occupational Exposures*, Third Edition, AIHA Press, Fairfax, VA.

Mulhausen, J. R. and J. Damiano (1998), *A Strategy for Assessing and Managing Occupational Exposures*, Second Edition, AIHA Press, Fairfax, VA.

See Also

Help files for efraction.ml,efclnp, percentile.exact, efraction.exact, aihand

Examples

```
# calculate exceedance fraction and exact CLs for Example data
# Appendix  Mulhausen and Damiano(1998) Ignacion and Bullock (2006
data(aihand)
x <- aihand$x ;  det <- rep(1,length(x))
aiha<-data.frame(x,det) #  complete data
unlist(efraction.exact(x,gam=0.95,L=5) ) #  exact CLs
unlist(efraction.ml(aiha,gam=0.95,L=5))  #  ML CLs
unlist(efclnp(aiha,L=5))                 #  nonparametric CLs
```

---

efraction.ml	*Calculate ML Estimate of Exceedance Fraction and Confidence Limits*

Description

Calculate the ML estimate of the exceedance fraction $F = Pr[X > L]$ and "large sample" confidence limits for lognormal data with non-detects.

Usage

```
efraction.ml(dd, gam = 0.95, L = 5, dat = TRUE)
```

Arguments

dd	if dat is TRUE dd is an n by 2 matrix or data frame with x in column 1 det in column 2
gam	one-sided confidence level γ. Default is 0.95
L	L is specified limit for the exceedance fraction; e.g., the occupational exposure limit
dat	if dat is FALSE, then dd is a list from lnorm.ml. Default is TRUE

Details

The exceedance fraction FL represent the proportion of the X's that exceed a given limit Lp. The null hypothesis of interest is $Ho : FL \geq Fo = 1 - p$; i.e., Fo is the maximum proportion of the population that can exceed the limit Lp. The ML point estimate of FL is $f = 1 - N(v)$ where $v = [log(L) - \mu]/\sigma$, and N(v) is the standard normal distribution function. The large sample $100\gamma\%$ LCL for $V = [log(L) - \mu]/\sigma$ is LCLv $= v - t(\gamma, m - 1)var(v)^{1/2}$, where

$$var(v) = p1^2 var(\mu) + p2^2 var(\sigma) + 2p1p2cov(\mu, \sigma)$$

, and p1 and p2 are partial derivatives of v with respect to μ and σ. The $100\gamma\%$ UCL for FL is $UF(L, \gamma) = 1 - N(LCLv)$. The $100\gamma\%$ LCL for FL is $LF(L, \gamma) = 1 - N(UCLv)$, where $UCLv = u + t(\gamma, m - 1)var(v)^{1/2}$. The null hypothesis $Ho : FL = 1 - p$ is rejected if the $100\gamma\%$ UCL for FL is less than Fo, indicating that the exposure profile is acceptable. The large sample ML estimates of the exceedance fraction and $100\gamma\%$ confidence limits for lognormal data are calculated using the output from lnorm.ml.

Value

A LIST with components:

f	is the ML estimate of exceedance fraction for lognormal distribution
f.LCL	is the 100*γ% lower confidence limit for f
f.UCL	is the 100*γ% upper confidence limit for f
L	L is specified limit for the exceedance fraction; e.g., the occupational exposure limit
gam	one-sided confidence level γ. Default is 0.95

Note

(f.LCL, f.UCL) is an $100(2\gamma - 1)$ percent confidence interval for F

Author(s)

E. L. Frome

References

Frome, E. L. and Wambach, P. F. (2005), "Statistical Methods and Software for the Analysis of Occupational Exposure Data with Non-Detectable Values," *ORNL/TM-2005/52,Oak Ridge National Laboratory*, Oak Ridge, TN 37830. Available at: `http://www.csm.ornl.gov/esh/aoed/ORNLTM2005-52.pdf`

See Also

`lnorm.ml`,`percentile.ml`

Examples

```
# calculate ML estimate of exceedance fraction and CLs for Example 2 in ORNLTM2005-52
data(beTWA)
unlist(efraction.ml(beTWA,L=0.2))
#  calculate nonparametric CLs
unlist(efclnp(beTWA,L=0.2))
```

---

filmbadge   *Quarterly Film Badge Data*

Description

Example of quarterly film badge data with non-detects

Usage

```
data(filmbadge)
```

Format

A data frame with 28 observations on the following 6 variables.

dlow lower end of annual dose

dhigh upper end of annual

Q1 dose for quarter 1

Q2 dose for quarter 2

Q3 dose for quarter 3

Q4 dose for quarter 4

Details

The product limit estimate (PLE) of the distribution function F(x) was first proposed by Kaplan and Meier (1958) for right-censored data, and Schmoyer et. al. (1996) defined the PLE for situations in which left censored data occurs. Both left censoring and right censoring are special cases of the general PLE (Peto,1973; Turnbull, 1976). A non-detect or left censored dose occurs when the dose is less than a detection limit. For a non-detect it is only known that the dose does not exceed the limit of detection(LOD). To obtain an estimate of the annual dose distribution F(x) from quarterly doses the general PLE is required since the annual doses will be "interval censored" if at least two of the quarterly doses are non-detects. Consider, for example, a worker with quarterly dose of 0, 50, 0, and 100 mrem. The quarterly interval doses are (0,30), (50,50), (0,30), and (100,100) assuming an LOD of 30 mrem. The annual dose is obtained by adding the lower and upper bounds of the quarterly doses and is equal to (150,210) for the example, i.e., it is only known that the dose is between 150 and 210.

Note

The dose unit is mrem (100 mrem= 10mSv). The LOD is 30 mrem. This is a representative sample of quarterly external dose records for 28 workers at the Y-12 facility in Oak Ridge (see *ORAUT-OTIB-0044, Rev. 01-A* and *ORAUT-OTIB-0064*) and is used here to illustrate the calculation of the PLE for interval censored data. The R function survreg can be used to obtain ML estimates of the parameters for the lognormal model and the covariance matrix that is needed for CLs for the exceedance fraction and 95th percentile.

References

Kaplan, E. L. and P. Meir (1958), "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, 457-481.

Ng, M. P. (2002), "A Modification of Peto's Nonparametric Estimation of Survival Curves for Interval-Censored Data," *Biometrics*, volume 58, number 2, pp. 439-442.

ORAUT (Oak Ridge Associated Universities Team), 2005c, "Historical Evaluation of the Film Badge Dosimetry Program at the Y-12 Facility in Oak Ridge, Tennessee: Part 1 - Gamma Radiation", *ORAUT-OTIB-0044, Rev. 01-A (was ORAUT-RPRT-0032, Rev. 00)*, Oak Ridge, Tennessee.

ORAUT (Oak Ridge Associated Universities Team), 2007, "External Coworker Dosimetry Data for the Y-12 National Security Complex". *ORAUT-OTIB-0064* (Under Revision).

Peto, R. (1973), "Experimental Survival Curves for Interval-censored Data," *Applied Statistics*, volume 22, number 1, pp. 86-91.

Schmoyer, R. L., J. J. Beauchamp, C. C. Brandt and F. O. Hoffman, Jr. (1996), "Difficulties with the Lognormal Model in Mean Estimation and Testing," *Environmental and Ecological Statistics*, 3, 81-97.

Turnbull, B. W. (1976), "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data," *Journal of the Royal Statistical Society*, Series B (Methodological), 38(3), 290-295.

Examples

```
data(filmbadge)
head(filmbadge)  #  LOOK AT FIRST FIVE RECORDS
```

```
#   USE icfit() TO CALCULATE PLE FOR INTERVAL CENSORED DATA
par( mfrow=c(1,2) )
out <- icfit(filmbadge$dlow,filmbadge$dhigh)
#   PLOT EXCEEDANCE S(x) vs x USING icplot()
tp <- "PLE of Exceedance for Filmbadge Data"
icplot(out$surv, out$time,XLAB="Dose",YLAB="Exceedance Probability",main=tp,cex.main=.8)
#   USE pleicf() TO CALCULATE PLE FOR filmbadge DATA
ple.fb <- pleicf(filmbadge[,1:2],FALSE)
#   USE qq.lnorm  FOR LOGNORMAL Q-Q PLOT FOR INTERVAL CENSORED DATA
tmp <- qq.lnorm(ple.fb)
GM <-round(exp(tmp$par[1])); GSD <- round(exp(tmp$par[2]),2)
tp<-paste("Lognormal Q-Q plot for Filmbadge\n  Data GM= ",GM,"GSD= ",GSD)
title(tp,cex.main=0.8) # title for q-q plot with graphical parameter estimates
#  RESULTS FROM  pleicf()
round(ple.fb,3)
#
```

---

icfit	*Calculates the Self-Consistent Estimate of Survival from Interval Censored Data*

Description

This function calculates the self-consistent estimate of survival for interval censored data. (i.e., the nonparametric maximum likelihood estimate that generalizes the Kaplan-Meier estimate to interval censored data). The censoring is such that if the i*th* observation fails at x, we only observe that $L[i] < x \le R[i]$. Data may be entered with "exact" values, i.e., $L[i] = x = R[i]$. In that case the $L[i]$ is changed internally to $L[i]*$ which is the next lower of any of the observed endpoints (unless $R[i] = 0$ then an error results).

Usage

```
icfit(L, R, initp = NA, minerror = 1e-06, maxcount = 1000)
```

Arguments

L	a vector of the left endpoints of the interval
R	a vector of the right endpoints of the interval
initp	a vector with an initial estimate of the density function. This vector should sum to 1 and have a length equal to the number of unique values of L and R combined. If initp = NA (default) then an initial value is estimated from the data.
minerror	The minimum error for convergence purposes. The EM algorithm stops when error < minerror, where error is the maximum of the reduced gradients (see Gentleman and Geyer, 1994). Default = 1e-06.
maxcount	the maximum number of iterations. Default is 10000.

Details

The algorithm is basically an EM-algorithm applied to interval censored data (see Turnbull, 1976); however, first there is a primary reduction (See Aragon and Eberly, 1992). Convergence is defined when the maximum reduced gradient is less than minerror, and the Kuhn-Tucker conditions are approximately met, otherwise a warning will result. (see Gentleman and Geyer, 1994). There may be other faster algorithms, but they are more complicated (see Aragon and Eberly, 1992). The output for time is sort(unique(c(0,L,R,Inf))) without the Inf. The output for p keeps the value related to Inf so that p may be inserted into initp for another run. The outputs for p and surv act as if the jumps in the survival curve happen at the largest of the possible times (see Gentleman and Geyer, 1994, Table 2, for a more accurate way to present p).

Value

Returns a list with the following elements:

u	a vector of Lagrange multipliers. If there are any negative values of u the Kuhn-Tucker conditions for convergence are not met. If this happens a warning will result.
error	this is the maximum of the reduced gradients. If convergence is correct then error < minerror and all values of u are nonnegative, otherwise a warning results.
count	number of iterations of the self-consistent algorithm (i.e., EM-algorithm)
time	a vector of times (see details)
p	a vector of probabilities, all except the last values are associated with the time vector above, i.e., $p[i] = Prob(X = time[i])$. The last value is associated with time==Inf. (see details).
surv	a vector of survival values associated with the time vector above, i.e., $surv[i] = Prob(X > time[i])$

Note

The functions icfit, icplot, and ictest and documentation for these functions are from Michael P. Fay. You are free to distribute these functions to whomever is interested. They come with no warranty however.

Author(s)

Michael P. Fay

References

Aragon, J. and Eberly, D. (1992), "On Convergence of Convex Minorant Algorithms for Distribution Estimation with Interval-Censored Data," *Journal of Computational and Graphical Statistics.* 1: 129-140.

Gentleman, R. and Geyer, C. J. (1994), "Maximum Likelihood for Interval Censored Data: Consistency and Computation," *Biometrika*, 81, 618-623.

Turnbull, B. W. (1976), "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data," *Journal of the Royal Statistical Society*, Series B,(Methodological), 38(3), 290-295.

See Also

[plekm](#), [qq.lnorm](#)

Examples

```
# Calculate and plot a Kaplan-Meier type curve for interval censored data.
# This is S(x) = 1 - F(x) and is the sample estimate of the probability
# of exceeding x.  The filmbadge data is used as an example.
data(filmbadge)
out <- icfit(filmbadge$dlow,filmbadge$dhigh)
icplot(out$surv, out$time,XLAB="Dose",YLAB="Exceedance Probability")
```

---

icplot   *Plots Survival Functions*

Description

This function takes a vector of survival values and a vector of time values and either plots a survival function or adds the survival lines to an existing plot.

Usage

```
icplot(surv, time = as.numeric(names(surv)), xrange = NA, lines.only = FALSE,
XLAB = "Time", YLAB = "Probability", LTY = 1, ...)
```

Arguments

surv   a vector of survival values.

time   a vector of times. These are related to the vector of survival by $Prob(X > time[i]) = surv[i]$. If this vector is not given then the default is to use as.numeric(names(surv)).

xrange   the range of the x values. The default is c(0,max(time[time<Inf])). This value may need to be supplied when plotting more than one survival curve on a single plot (see lines.only).

lines.only   a logical value; default = FALSE. If lines.only = FALSE the function draws a new set of axes. If lines.only = TRUE the function adds lines to an existing plot. To print 2 lines on one plot, call the function twice, the first time with lines.only = FALSE and xrange = range(c(t1,t2)) where t1 and t2 are the (finite) times for the two survival curves, the second time with lines.only = TRUE.

XLAB   a character string denoting the x label. Default = "Time".

YLAB   a character string denoting the y label. Default = "Probability".

LTY   an integer denoting the line type (lty value).

...   additional plotting parameters (except xlab and ylab).

Details

See `icplot` details. This may not be the most accurate way to present the data. See Betensky, Lindsey, Ryan, and Wand (1999, p. 238) for an alternative method.

Value

Returns a plot or adds a line to an existing plot.

Note

The functions `icfit`, `icplot`, and `ictest` and documentation for these functions are from Michael P. Fay. You are free to distribute these functions to whomever is interested. They come with no warrantee however.

Author(s)

Michael P. Fay

References

Betensky, R. A., Lindsey, J. C., Ryan, L. M., and Wand, M. P. (1999), "Local EM Estimation of the Hazard Function for Interval-Censored Data," *Biometrics*, 55: 238-245.

See Also

`icfit`, `ictest`

Examples

```
# Plot two Survival curves on one plot.
#  need data set for this example
#   s1<- icfit(left[treatment==1],right[treatment==1])
#   s2<- icfit(left[treatment==2],right[treatment==2])
#   icplot(s1$surv,s1$time,xrange=range(c(s1$time,s2$time)))
#   icplot(s2$surv,s2$time,lines.only=TRUE,LTY=2)
```

---

ictest   *Performs Tests for Interval Censored Data*

Description

This function performs several different tests for interval censored data. It has 3 different models that generalize either the Wilcoxon rank sum test (model = "PO") or the logrank test (model = "GPH" or model = "Sun"). Each model may be one of 2 types, either an asymptotic permutation test or a score test.

The censoring is such that if the i*th* observation fails at x, we only observe that $L[i] < x \leq R[i]$. Data may be entered with "exact" values, i.e., $L[i] = x = R[i]$. In that case the $L[i]$ is changed

internally to $L * [i]$ which is the next lower of any of the observed endpoints (unless $R[i] = 0$ then an error results).

The function requires a previously calculated survival curve (see [icfit](#)).

Usage

```
ictest(L, R, S, group, model, type = "permutation", fuzz , output.scores)
```

Arguments

L	a vector of left endpoints of the interval. We assume each $L[i] \geq 0$.
R	a vector of right endpoints of the interval. Exact values may be entered as L[i] == R[i] they are changed internally.
S	a vector of survival values calculated from all the data (i.e., ignoring group membership) (see time below), typically output from icfit.
group	a vector denoting the group for which the test is desired. If group is a factor or character then a k-sample test is performed, where k is the number of unique values of group. If group is numeric then a "correlation" type test is performed. If there are only two groups, both methods give the same results.
model	a character vector with three possible values describing the model: model = "GPH" (default) gives the grouped proportional hazards model. This generalizes a logrank test. model = "Sun" gives a Logistic model. This generalizes another form of the logrank test. model = "PO" gives a proportional odds model. This generalizes the Wilcoxon rank sum test. (see details).
type	a character vector with two possible values, "permutation" or "score" (see details)
fuzz	a small numeric value. Because we need to determine places in the survival curve where there are no changes, and the machine may have rounding error, we use this. (Default = 1e-12)
output.scores	a logical value. output.scores = TRUE outputs the scores in the output list. Default is output.scores = FALSE

Details

The 3 models are compared in depth in Fay (1999). For censored data two common likelihoods are the marginal likelihood or the ranks and the likelihood with nuisance parameters for the baseline survival. Here we use the latter likelihood (as in Finkelstein, 1986, Fay, 1996, and Sun, 1996). It is difficult to create proper score tests for this likelihood because the number of nuisance parameters typically grows with the sample size and often many of the parameters are equal at the nonparametric MLE, i.e., they are on the boundary of the parameter space. One way around this is to perform a permutation test on the scores (Fay, 1996). A second way around (the boundary problem at least) it is to redefine the interval points so that no boundary problems exist (see Fay, 1996). These are the two methods used here.

We present two generalizations of the logrank test. The method of Sun (1996) is more difficult to calculate and has no theoretical advantages of which I am aware. The grouped proportional hazards

model of Finkelstein (1996) is recommended. Note that when `icfit` and `ictest` are used on right-censored data, because the method of estimating variance is different, even Sun's method does not produce exactly the standard logrank test results.

There are some typos in Appendix II of Fay (1999). See the S code for the corrections.

Value

Returns a list with the following elements:

scores	only returned if output.scores = T. This is a vector the same length as L and R, containing the scores used in the permutation test.
U	The efficient score vector. When group is a factor or character vector then each element of U has the interpretation as the weighted sum of "observed" minus "expected" deaths for the group element defined by the label of U. Thus negative values indicate better than average survival (see Fay, 1999).
chisq.value	Chi-square value of the test
df	Degrees of freedom for the chi-square test.
pvalue	p-value from the chi-square value.
test	a character vector of length one, either "2-sample","correlation" or "k-sample" where k in the number of unique group values.
model	same as model input, either "GPH","Sun" or "PO"
type	same as type input, either "permutation" or "score"

Note

The functions [icfit](), [icplot](), and [ictest]() and documentation for these functions are from Michael P. Fay. You are free to distribute these functions to whomever is interested. They come with no warranty however.

Author(s)

Michael P. Fay

References

Fay, M. P. (1996), "Rank Invariant Tests for Interval Censored Data Under the Grouped Continuous Model," *Biometrics*, 52: 811-822.

Fay, M. P. (1999), "Comparing Several Score Tests for Interval Censored Data," *Statistics in Medicine*, 18: 273-285.

Finkelstein, D. M. (1986), "A Proportional Hazards Model for Interval Censored Failure Time Data," *Biometrics*, 42: 845-854.

Sun, J. (1996), "A Non-parametric Test for Interval Censored Failure Time Data With Applications to AIDS Studies," *Statistics in Medicine*, 15: 1387-1395.

See Also

[icfit](),[icplot]()

Examples

```
## Perform a logrank-type test using the observed information variance.
## need data set for this example
#   out<-icfit(left,right)
#  ictest(left,right,out$surv,group,out$time,model = "GPH",type = "score")
#
## Perform a Wilcoxon rank sum-type test using asymptotic permutation variance.
#
# ictest(left,right,out$surv,group,out$time, model = "PO",type = "permutation")
```

IH.summary   *Summary Statistic for Samples With Non-detects*

Description

Summary statistic described by The American Industrial Hygiene Association (AIHA) for occupational exposure data are calculated for samples with non-detects (aka left censored data). Parametric estimates are based on a lognormal model using maximum likelihood (ML). Nonparametric methods are based on the product limit estimate (PLE) for left censored data.

Usage

```
IH.summary(dd,L, p = 0.95, gam = 0.95,bcol=NA)
```

Arguments

dd	An n by 2 matrix or data frame with x (exposure) variable in column 1, and det = 0 for non-detect or 1 for detect in column 2
L	L is specified limit for the exceedance fraction; e.g., the occupational exposure limit
p	p is probability for Xp the 100p*th* percentile. Default is 0.95
gam	one-sided confidence level γ. Default is 0.95
bcol	Column number that contains a BY variable. This column must contain a factor and the value of each of the summary statistics is calculated for each level of the factor. Default NA

Details

Regulatory and advisory criteria for evaluating the adequacy of occupational exposure controls are generally expressed as limits that are not to be exceeded in a work shift or shorter time-period if the agent is acutely hazardous. Exposure monitoring results above the limit require minimal interpretation and should trigger immediate corrective action. Demonstrating compliance with a limit is more difficult. AIHA has published a consensus standard with two basic strategies for evaluating an exposure profile—see Mulhausen and Damiano(1998), Ignacio and Bullock (2006). The first approach is based on the mean of the exposure distribution, and the second approach

considers the "upper tail" of the exposure profile. Statistical methods for estimating the mean, an upper percentile of the distribution, the exceedance fraction, and the uncertainty in each of these parameters are provided by this package. Most of the AIHA methods are based on the assumptions that the exposure data does not contain non-detects, and that a lognormal distribution can be used to describe the data. Exposure monitoring results from a compliant workplace tend to contain a high percentage of non-detected results when the detection limit is close to the exposure limit, and in some situations, the lognormal assumption may not be reasonable. All of these methods are described in a companion report by Frome and Wambach (2005).

Value

A data.frame with column names based on levels of the BY variable and row names:

mu	ML estimate of mean of y=log(x)
se.mu	Estimate of standard error of mu
sigma	ML estimate of sigma
se.sigma	Estimate of standard error of sigma
GM	MLE of geometric mean
GSD	MLE of geometric standard deviation
EX	MLE of E(X) the (arithmetic) mean
EX-LCL	Lower Confidence Limit for E(X)
EX-UCL	Upper Confidence Limit for E(X)
KM-mean	Kaplan-Meier(KM) Estimate of E(X)
KM-LCL	KM Lower Confidence Limit for E(X)
KM-UCL	KM Upper Confidence Limit for E(X)
KM-se	Standard Error of KM-mean
obs.Xp	Estimate of Xp from PLE
Xp	ML estimate of Xp the pth percentile
Xp.LCL	MLE of LX(p,gam) the LCL for Xp
Xp.UCL	MLE of UX(p,gam) the UCL for Xp
zL	MLE of the Z value for limit L
NpUTL	Nonparametric estimate of the UTL $p - \gamma$
Maximum	Largest value in the data set
NonDet	percent of X's that are left censored, i.e., non-detects
n	number of observations in the data set
Rsq	Square of correlation for the quantile-quantile (q-q) plot
m	number X's greater than the LOD
f	MLE of exceedance fraction F for limit L
f.LCL	LCf(L,gam) MLE of LCL for F
F.UCL	UCf(L,gam) MLE of UCL for F
fnp	Nonparametric estimate of F for limit L

`fnp.LCL`	Nonparametric estimate of LCL for F
`fnp.UCL`	Nonparametric estimate of UCL for F
`m2log(L)`	-2 times the log-likelihood function
`L`	L is specified limit for the exceedance fraction; e.g., the occupational exposure limit
`p`	percentile for UTL p-γ
`gam`	one-sided confidence level γ. Default is 0.95

Author(s)

E. L. Frome

References

Frome, E. L. and Wambach, P. F. (2005), "Statistical Methods and Software for the Analysis of Occupational Exposure Data with Non-Detectable Values," *ORNL/TM-2005/52,Oak Ridge National Laboratory*, Oak Ridge, TN 37830. Available at: http://www.csm.ornl.gov/esh/aoed/ORNLTM2005-52.pdf

Ignacio, J. S. and W. H. Bullock (2006), *A Strategy for Assesing and Managing Occupational Exposures*, Third Edition, AIHA Press, Fairfax, VA.

Mulhausen, J. R. and J. Damiano (1998), *A Strategy for Assesing and Managing Occupational Exposures*, Second Edition, AIHA Press, Fairfax, VA.

See complete list of references at `About-STAND`

See Also

See Also `lnorm.ml`, `efraction.ml`, `percentile.ml`, `kmms`

Examples

```
# Analysis for cansdata  Example 1 from ORNLTM2005-52
data(cansdata)
Allcans<- round(IH.summary(cansdata,L=0.2,bcol=NA),3)
# Example using cansdata with By variable
cansout <- round(IH.summary(cansdata,L=0.2,bcol=3),3)
#  combine out from both analysis
cbind(Allcans,cansout)
```

---

kmms *Kaplan-Meier (KM) Mean and Standard Error*

Description

Kaplan- Meier Estimate of Mean and Standard Error of the Mean for Left Censored Data

Usage

```
kmms(dd, gam = 0.95)
```

Arguments

dd	An n by 2 matrix or data frame with x (exposure) variable in column 1, and det= 0 for non-detect or 1 for detect in column 2
gam	one-sided confidence level γ. Default is 0.95

Details

The product limit estimate (PLE) of the cumulative distribution function was first proposed by Kaplan and Meier (1958) for right censored data. Turnbull (1976) provides a more general treatment of nonparametric estimation of the distribution function for arbitrary censoring. For randomly left censored data, the PLE is defined by Schmoyer et al. (1996)–see `plend`.

The mean of the PLE is a censoring-adjusted point estimate of E(X) the mean of X. An approximate standard error of the PLE mean can be obtained using the method of Kaplan and Meier (1958), and the $100\gamma\%$ UCL is $KM.mean + t(\gamma - 1, m - 1)sp$, where sp is the Kaplan-Meier standard error of the mean adjusted by the factor $m/(m-1)$, where m is the number of detects in the sample. When there is no censoring this reduces to the second approximate method described by Land (1972).

Value

A LIST with components:

KM.mean	Kaplan- Meier(KM) estimate of mean E(X)
KM.LCL	KM estimate of lower confidence limit
KM.UCL	KM estimate of upper confidence limit
KM.se	estimate of standard error of KM-mean
gamma	one-sided confidence level γ. Default 0.95

Note

Error in KM.se corrected on 12 June 2007. KM standard error is adjusted by multiplying by sqrt(m/(m-1)) where m is number of detected values. Error occurred if there were ties in detected values by calculating the number of unique detected values. For example, for beTWA sqrt(m/(m-1)) is 1.004796 . Due to error 1.008032 was used. The sqrt(m/(m-1)) will always be smaller after correction, depending on value of m and the number of ties. See the example.

Author(s)

E. L. Frome

References

Kaplan, E. L. and Meier, P. (1958), "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, 457-481.

Schmoyer, R. L., J. J. Beauchamp, C. C. Brandt and F. O. Hoffman, Jr. (1996), "Difficulties with the Lognormal Model in Mean Estimation and Testing," *Environmental and Ecological Statistics*, 3, 81-97.

See Also

plend, plekm

Examples

```
# results for beTWA data using kmms in stand Ver 1.1 with error
#    KM.mean      KM.LCL      KM.UCL      KM.se       gamma
# 0.018626709 0.014085780 0.023167637 0.002720092 0.950000000
#
data(beTWA) # Use data from Example 2 in ORNLTM2002-51
unlist(kmms(beTWA))
```

---

lnorm.ml   *ML Estimation for Lognormal Data with Non-detects*

Description

When an exposure measurement may be less than a detection limit closed form and exact methods have not been developed for the lognormal model. The maximum likelihood (ML) principle is used to develop an algorithm for parameter estimation, and to obtain large sample equivalents of confidence limits for the mean exposure level, the 100p*th* percentile, and the exceedance fraction. For a detailed discussion of assumptions, properties, and computational issues related to ML estimation see Cox and Hinkley (1979) and Cohen (1991).

Usage

```
lnorm.ml(dd)
```

Arguments

dd   An n by 2 matrix or data frame with
                x (exposure) variable in column 1, and
                det= 0 for non-detect or 1 for detect in Column 2

Details

For notational convenience the m detected values $x[i]$ are listed first followed by the $nx[i]$ indicating non-detects, so that the data are $x[i], i = 1, \ldots, m, nx[i]i = m + 1, \ldots, n$. If $nx[i]$ is the same for each non-detect, this is referred to as a left singly censored sample (Type I censoring) and nx is the limit of detection(LOD). If the $nx[i]$ are different, this is known as randomly (or progressively) left-censored data[see Cohen(1991) and Schmoyer et al (1996)]. In some situations a value of 0 is recorded when the exposure measurement is less than the LOD. In this situation, the value of $nx[i]$ is the LOD indicating that x is in the interval $(0, nx[i])$. The probability density function for lognormal distribution is

$$g(x; \mu, \sigma) = exp[-(log(x) - \mu)^2/(2\sigma^2)]/[\sigma x \sqrt{(2\Pi)}]$$

where $y = log(x)$ is normally distributed with mean μ and standard deviation σ [Atkinson and Brown (1969)]. The geometric mean of X is $GM = exp(\mu)$ and the geometric standard deviation is $GSD = exp(\sigma)$. Strom and Stansberry (2000) provide a summary of these and other relationships for lognormal parameters. Assuming the data are a random sample from a lognormal distribution, the log of the likelihood function for the unknown parameters μ and σ given the data is

$$L(\mu, \sigma) = \sum log[g(x; \mu, \sigma)] + \sum log[G(nx; \mu, \sigma)],$$

where $G(x; \mu, \sigma)$ is the lognormal distribution function, i.e., $G(nx; \mu, \sigma)$ is the probability that $x \leq nx$. The first summation is over $i = 1, \ldots, m$, and the second is over $i = m + 1, \ldots, n$.

To test that the mean of $X > L$, $Ho : E(X) > L$ at the $\alpha = 1 - \gamma$ significance level a one-sided upper $100\gamma\%$ confidence limit can be used. One method for calculating this UCL is to use the censored data equivalent of Cox's direct method; i.e., calculate the ML estimate of $\phi = \mu + [1/2]\sigma^2$, and $var(\phi) = var(\mu + [1/2]\sigma^2)$ where

$$var(\phi) = var(\mu) + [1/4]var(\sigma^2) + cov(\mu, \sigma^2).$$

The ML estimator of E(X) is $exp(\phi)$, the $100\gamma\%$ LCL for E(X) is exp$[\phi - tvar(\phi)]$, and the $100\gamma\%$ UCL for E(x) is $exp[\phi + tvar(\phi)]$, where $t = t(\gamma, m - 1)$. The resulting confidence interval (LCL, UCL) has confidence level $100(2\gamma - 1)\%$. An equivalent procedure is to estimate $\phi = \mu + [1/2]\sigma^2$ and its standard error directly, i.e., by maximizing the log-likelihood with parameters $\mu + [1/2]\sigma^2$ and σ^2. ML estimates of $\mu, \sigma, \phi, \sigma^2$, estimates of their standard errors, and covariance terms are calculated.

Value

A list with components:

mu	ML estimate of μ
sigma	ML estimate of σ
logEX	ML estimate of log of E(X)
SigmaSq	ML estimate of σ^2
se.mu	ML estimate of standard error of μ
se.sigma	ML estimate of standard error of σ
se.logEX	ML estimate of standard error of log of E(X)

se.Sigmasq	ML estimate of standard error of σ^2
cov.musig	ML estimate of cov(μ,σ)
m	number of detects
n	number of observations in the data set
m2log(L)	-2 times the log-likelihood function
convergence	convergence indicator from optim

Note

Local function ndln is called by optim for mu and sigma and local function ndln2 is called by optim for logEX and Sigmasq.

Author(s)

E. L. Frome

References

Cohen, A. C. (1991), *Truncated and Censored Samples*, Marcel Decker, New York

Cox, D. R. and D. V. Hinkley (1979), *Theoretical Statistics*, Chapman and Hall, New York.

Frome, E. L. and Wambach, P. F. (2005), "Statistical Methods and Software for the Analysis of Occupational Exposure Data with Non-Detectable Values," *ORNL/TM-2005/52,Oak Ridge National Laboratory*, Oak Ridge, TN 37830. Available at: http://www.csm.ornl.gov/esh/aoed/ORNLTM2005-52.pdf

See Also

optim, efraction.ml, percentile.ml

Examples

```
# Calculate MLE for Example 2 in ORNLTM2005-52
data(beTWA)
mle.TWA<- unlist(lnorm.ml(beTWA)) # ML for Be monitoring data
mle.TWA[1:4]  #  ML estimates of parameters
mle.TWA[5:8]  #  Standard errors of ML estimates
mle.TWA[9:13] #  additional results from lnorm.ml
```

| npower.lnorm | *Sample Size and Power For Lognormal Distribution* |

Description

Find either the sample size or power for complete sample from lognormal distribution

Usage

```
npower.lnorm(n=NA,power=NA,fstar=1,p=0.95,gamma=0.95)
```

Arguments

n	sample size
power	power of the test = 1 - β
fstar	true percent of X's $\geq$ limit L
p	probability for Xp the 100p*th* percentile. Default is 0.95
gamma	one-sided confidence level γ. Default is 0.95

Details

Find either the sample size n or the power of the test for specified values of fstar, p, and gamma. Either n is missing or power is missing.

The null hypothesis of interest is $Ho : F \geq Fo = 1 - p$; i.e., Fo is the maximum proportion of the population that can exceed the limit Lp. The null hypothesis is rejected if the $100\gamma\%$ UCL for F is less than Fo , indicating that the exposure profile is acceptable. For the complete data case this is equivalent to testing the null hypothesis $Ho : Xp \geq Lp$ at the $\alpha = (1 - \gamma)$ significance level. See [efraction.exact](), [percentile.exact]() and Section 2.3 of Frome and Wambach(2005) for further details.

Value

A vector with components:

n	sample size
power	power of the test = 1 -β
fstar	true percent of X's $\geq$ limit L
p	probability for Xp the 100p*th* percentile. Default is 0.95
gamma	one-sided confidence level γ. Default is 0.95

Note

The R function uniroot is used to find a parameter of the non-central t distribution. In some versions of R this may cause a warning message. See R bug report RP 9171 full precision was not achieved in 'pnt'. This warning message may occur in uniroot calls to pt and does not effect the precision of the final result

Author(s)

E.L. Frome

References

Johnson, N. L. and B. L. Welch (1940), "Application of the Non-Central t-Distribution," *Biometrika*, 31(3/4), 362-389.

Lyles R. H. and L. L. Kupper (1996), "On strategies for comparing occupational exposure data to limits," *American Industrial Hygiene Association Journal*, 57:6-15.

Frome, E. L. and Wambach, P. F. (2005), "Statistical Methods and Software for the Analysis of Occupational Exposure Data with Non-Detectable Values," *ORNL/TM-2005/52,Oak Ridge National Laboratory*, Oak Ridge, TN 37830. Available at: http://www.csm.ornl.gov/esh/aoed/ORNLTM2005-52.pdf

Ignacio, J. S. and W. H. Bullock (2006), *A Strategy for Assesing and Managing Occupational Exposures*, Third Edition, AIHA Press, Fairfax, VA.

Mulhausen, J. R. and J. Damiano (1998), *A Strategy for Assessing and Managing Occupational Exposures*, Second Edition, AIHA Press, Fairfax, VA.

See Also

Help files for efraction.ml,percentile.ml, efclnp,aihand

Examples

```
#                           EXAMPLE 1
#    Table VII.1 Mulhausen and Damiano (1998) adapted from
#    Table II in Lyles and Kupper (1996) JAIHA vol 57 6-15 Table II
#    Sample Size Needed When Using UTL(95,95) to Show 95% Confidence
#    that the 95th Percentile is below the OEL (Power = 0.8)
rx<-c(1.5,2,2.5,3)
sdx<- sqrt(c(0.5,1,1.5,2,2.5,3))
tabn<-matrix(0,4,6)
for ( i in 1:4) {
  for (j in 1:6) {
fstar<- 100*(1 -pnorm( log(rx[i])/sdx[j] + qnorm(0.95) ))
tabn[i,j]<- npower.lnorm(NA,0.8,fstar,p=0.95,gamma=0.95)[1]
}
}
cn<- paste("GSD = ",round(exp(sdx),2),sep="" )
dimnames(tabn)<-list( round(1/rx,2),cn)
rm(cn,rx,sdx)
tabn
#                           EXAMPLE 2
top<-"Power For Sample Size n = 20 for p=0.95 gamma=0.95"
fstar <- seq(0.2,4.8,0.1)
pow <- rep(1,length(fstar))
for (i in 1 : length(fstar)) {
pow[i]<-npower.lnorm(20,NA,fstar[i],p=0.95,gamma=0.95)[2]
}
```

```
plot(fstar,pow,xlim=c(0,5),ylim=c(0,1),main=top,
xlab="fstar = True Percent of Xs > L(Specified Limit )",ylab="Power")
```

| nptl | *Nonparametric Upper Tolerance Limit* |

Description

Given a random sample of size n from a continuous distribution, then, with a confidence level of at least γ, at least 100p percent of the population will be below the k*th* largest value in the sample.

Usage

```
nptl(n , p = 0.95, gam = 0.95)
```

Arguments

n	the sample size
p	probability for Xp the 100p*th* percentile. Default is 0.95
gam	one-sided confidence level γ. Default is 0.95

Value

k	index of the order statistic

Note

The maximum non-detect must be less than the k*th* largest value.

Author(s)

E. L. Frome

References

Sommerville, P. N. (1958), "Tables for Obtaining Non-Parametric Confidence Limits," *Annals of Mathematical Statistics*, 29, 599-601.

Examples

```
data(beTWA)
k<- nptl(length(beTWA[,1]))
rev(sort(beTWA[,1]))[k]
```

---

percentile.exact	*Estimate of Xp and Exact Confidence Limits for Normal/Lognormal*

Description

Calculate estimate of Xp the 100*p percentile of the normal/lognormal distribution, and the lower and upper 100*γ% exact confidence limits. The resulting interval (Xp.LCL,Xp.UCL) is an approximate $100 * (2\gamma - 1)$ percent confidence interval for Xp the 100*p percentile. This function should only be used for complete samples.

Usage

```
percentile.exact(x, p = 0.95, gam = 0.95,logx=TRUE,wpnt=FALSE)
```

Arguments

x	vector of positive data values
p	probability for Xp the 100p*th* percentile. Default is 0.95
gam	one-sided confidence level γ. Default 0.95
logx	If TRUE, sample is from lognormal, else normal. Default is TRUE
wpnt	if TRUE, show warning from pnt. Default is $FALSE$

Details

A point estimate of Xp, the 100p*th* percentile of a normal/lognormal distribution is calculated. Exact confidence limits for Xp are calculated using the quantile function of the non-central t distribution. The exact UCL is $m + K * s$, where m is the sample mean, s is the sample standard deviation, and the $K factor$ depends on n, p, and γ. The exact LCL is $m + K' * s$. The local function kf calculates K and K' using the quantile function of the non-central t distribution qt.

The null hypothesis $Ho : Xp \geq Lp$ is rejected at the $\alpha = (1 - \gamma)$ significance level if the 100γ% UCL for Xp is less than the specified limit Lp (indicating the exposure profile is acceptable).

Value

A LIST with components:

Xp	estimate of the p*th* percentile of the distribution
Xpe.LCL	$100 * \gamma$% exact lower confidence limit for Xp
Xpe.UCL	$100 * \gamma$% exact upper confidence limit for Xp
p	probability for Xp the 100p*th* percentile. Default 0.95
gam	one-sided confidence level γ. Default is 0.95
Logx	If TRUE, sample is from lognormal, else normal. Default is TRUE
n	sample size
Ku	the K factor used to calculate the exact UCL
Kl	the K' factor used to calculate the exact LCL

Note

The UCL is also referred to as an upper tolerance limit, i.e., if p = 0.95 and $\gamma = 0.99$ then Xpe.UCL is the exact UTL 95% - 99%.

Author(s)

E. L. Frome

References

Burrows, G. L. (1963), "Statistical Tolerance Limits - What are They," *Applied Statistics*, 12, 133-144.

Johnson, N. L. and B. L. Welch (1940), "Application of the Non-Central t-Distribution," *Biometrika*, 31(3/4), 362-389.

Lyles R. H. and L. L. Kupper (1996), "On Strategies for Comparing Occupational Exposure Data to Limits," *American Industrial Hygiene Association Journal*, 57:6-15.

Tuggle, R. M. (1982), "Assessment of Occupational Exposure Using One-Sided Tolerance Limits," *American Industrial Hygiene Association Journal*, 43, 338-346.

Frome, E. L. and Wambach, P. F. (2005), "Statistical Methods and Software for the Analysis of Occupational Exposure Data with Non-Detectable Values," *ORNL/TM-2005/52,Oak Ridge National Laboratory*, Oak Ridge, TN 37830. Available at: http://www.csm.ornl.gov/esh/aoed/ORNLTM2005-52.pdf

Ignacio, J. S. and W. H. Bullock (2006), *A Strategy for Assesing and Managing Occupational Exposures*, Third Edition, AIHA Press, Fairfax, VA.

Mulhausen, J. R. and J. Damiano (1998), *A Strategy for Assessing and Managing Occupational Exposures*, Second Edition, AIHA Press, Fairfax, VA.

See Also

Help files for `percentile.ml`, `efraction.exact`, `aihand`

Examples

```
#                    EXAMPLE 1
# calculate 95th percentile and exact CLs for Example data
# Appendix  Mulhausen and Damiano (1998)
data(aihand)
x <- aihand$x ;  det <- rep(1,length(x))
aiha <- data.frame(x,det) #  complete data
unlist(percentile.exact(x,gam=0.95,p=0.95) )[1:5]  #  exact CLs
unlist(percentile.ml(aiha,gam=0.95,p=0.95))   #  ML CLs
#                    EXAMPLE 2
#  Ignacio and Bullock (2006) Mulhausen and Damiano (1998)
#  Calculate TABLE VII.3 (page 272) Factor for One-Sided Tolerance
#  Limits for Normal Distribution (Abridged Version)
#  Same as Table III Burrows(1963) Panel 3 Page 138
nn <- c(seq(3,25),seq(30,50,5))
pv <-c(0.75,0.9,0.95,0.99,0.999)
```

```
tab <- matrix(0,length(nn),length(pv))
  for( k in (1:length(nn) ) ){
  xx <- seq(1,nn[k])
  for(j in (1:length(pv))) {
  tab[k,j ]<- percentile.exact(xx,pv[j],gam=0.95,FALSE)$Ku
}}
dimnames(tab)<-(list(nn,pv)) ; rm(nn,pv,xx)
round(tab,3)
#
#                      EXAMPLE 3
#  Calculate TABLE I One Sided Tolerance Factor K'
#  Tuggle(1982) Page 339 (Abridged Version)
nn <- c(seq(3,20),50,50000000)
pv <-c(0.9,0.95,0.99)
tab <- matrix(0,length(nn),length(pv))
  for( k in (1:length(nn) ) ){
  xx <- seq(1,nn[k])
  for(j in (1:length(pv))) {
  tab[k,j ]<- percentile.exact(xx,pv[j],gam=0.95,FALSE)$Kl
}}
dimnames(tab)<-(list(nn,pv)) ; rm(nn,pv,xx)
round(tab,3)
```

---

percentile.ml   *Calculate ML Estimate of Xp and Confidence Limits*

Description

Calculate the ML estimate of Xp the 100p*th* percentile of the lognormal distribution, and the lower and upper $100 * \gamma\%$ confidence limits LX(p,γ) and UX(p,γ). The upper confidence limit is used to test the null hypothesis that the exposure profile is "unacceptable". If $UX(p,\gamma) < L$ the null hypothesis is rejected and workplace is considered "safe" or the object/area is not contaminated. The Type I error is $\leq \alpha = 1 - \gamma$. The resulting interval (LX,UX) is an approximate $100 * (2\gamma - 1)$ percent confidence interval for Xp.

Usage

```
percentile.ml(dd, p = 0.95, gam = 0.95, dat = TRUE)
```

Arguments

dd	An n by 2 matrix or data frame with x (exposure) variable in column 1, and det= 0 for non-detect or 1 for detect in column 2
p	is probability for Xp the 100p*th* percentile. Default is 0.95
gam	one-sided confidence level γ. Default is 0.95
dat	if dat is FALSE then dd is a list from `lnorm.ml`. Default is TRUE

Details

The point estimate of $Yp = log(Xp)$ is $\mu + z\sigma$ where μ and σ are ML estimates and z is qnorm(p). The variance of the estimate is

$$var(\mu + z\sigma) = var(\mu) + Z^2 pvar(\sigma) + 2zcov(\mu, \sigma)$$

The $100\gamma\%$ LCL and UCL for Xp are

$$LX(p, \gamma) = exp[Yp - t(\gamma, (m-1))var(Yp)^{1/2}],$$

$$UX(p, \gamma) = exp[Yp + t(\gamma, (m-1))var(Yp)^{1/2}].$$

The ML estimates of $var(\mu)$, $var(\sigma)$, and $cov(\mu, \sigma)$ are obtained from the ML variance-covariance matrix using `lnorm.ml`. The null hypothesis $Ho : Xp \geq Lp$ is rejected at the $\alpha = (1 - \gamma)$ significance level if the $100\gamma\%$ UCL for Xp < Lp (indicating the exposure profile is acceptable).

Value

A LIST with components:

Xp	ML estimate of the p*th* percentile of lognormal distribution
Xp.LCL	$100 * \gamma\%$ lower confidence limit for Xp
Xp.UCL	$100 * \gamma\%$ upper confidence limit for Xp
p	probability for Xp the 100p*th* percentile. Default 0.95
gam	one-sided confidence level γ. Default is 0.95

Note

The UCL is also referred to as an upper tolerance limit(UTL), i.e., if p = 0.95 and gam = 0.99 then Xp.UCL is the UTL-95%-99%.

Author(s)

E. L. Frome

References

Cohen, A. C. (1991), *Truncated and Censored Samples*, Marcel Decker, New York

Cox, D. R. and D. V. Hinkley (1979), *Theoretical Statistics*, Chapman and Hall, New York.

Frome, E. L. and Wambach, P. F. (2005), "Statistical Methods and Software for the Analysis of Occupational Exposure Data with Non-Detectable Values," *ORNL/TM-2005/52,Oak Ridge National Laboratory*, Oak Ridge, TN 37830. Available at: http://www.csm.ornl.gov/esh/aoed/ORNLTM2005-52.pdf

See Also

Help files for `lnorm.ml`,`efraction.ml`

Examples

```
data(beTWA)
# calculate ML estimate of 95th percentile and CLs for Example 2 in ORNLTM2005-52
unlist(percentile.ml(beTWA,0.95,0.95))
```

---

percentile.ple	*Calculate Nonparametric Estimate of Xp and Confidence Limits*

Description

Find Xp, the 100p*th* percentile, and the $100\gamma\%$ nonparametric confidence limits from PLE of F(x).

Usage

```
percentile.ple(dd, p = 0.95, gam = 0.95, interp = TRUE)
```

Arguments

dd	An n by 2 matrix or data frame with x (exposure) variable in column 1, and det= 0 for non-detect or 1 for detect in column 2
p	Find x such that the PLE of F(x) = p. Default 0.95
gam	one-sided confidence level γ. Default is 0.95
interp	if interp is TRUE use linear interpolation. Default TRUE

Details

Find Xp the 100p*th* percentile and confidence limits from the PLE of F(x) – see plekm for additional details. If interp is TRUE use linear interpolation; otherwise, the upper confidence limit (UCL) for Xp, UX(p,γ), is the smallest value of x such that the LCL for F(x) is $\geq$ p, the lower confidence limit (LCL), LX(p,γ), is the largest value of x such that the UCL for F(x) is $\leq$ p.

Value

A list with components:

Xp	PLE of the 100p*th* percentile
LXp	PLE of LX(p,γ) the $100 * \gamma\%$ LCL for Xp
UXp	PLE of UX(p,γ) the $100 * \gamma\%$ UCL for Xp
p	probability for Xp the 100p*th* percentile
gam	one-sided confidence level γ. Default is 0.95

Author(s)

E. L. Frome

References

Frome, E. L. and Wambach, P. F. (2005), "Statistical Methods and Software for the Analysis of Occupational Exposure Data with Non-Detectable Values," *ORNL/TM-2005/52,Oak Ridge National Laboratory*, Oak Ridge, TN 37830. Available at: http://www.csm.ornl.gov/esh/aoed/ORNLTM2005-52.pdf

See Also

See Also as `plekm` and `SESdata`

Examples

```
# use data from example 2 in ORNL/TM-2005/52 to calculate
# 95 percent UCL for 95th percentile
data(beTWA)
unlist(percentile.ple(beTWA))
unlist(percentile.ml(beTWA)) # compare ML estimates
```

---

ple.plot   *Plot PLE With Confidence Limits*

Description

Plot the product limit estimate (PLE) of F(x) and $100(2\gamma - 1)\%$ two-sided confidence limits (CLs) for left censored data. A horizontal line corresponding to the Xp = 100p*th* percentile is added to the plot and the nonparametric confidence limits for Xp are displayed in the title.

Usage

```
ple.plot(dd, gam = 0.95, p = 0.95, xlow = 0, xh = NA, ylow = 0, yh = 1,...)
```

Arguments

dd	An n by 2 matrix or data frame with x (exposure) variable in column 1, and det = 0 for non-detect or 1 for detect in column 2
gam	one-sided confidence level γ. Default is 0.95
p	probability for Xp the 100p*th* percentile. Default is 0.95
xlow	minimum value on x axis. Default = 0
xh	maximum value on the x axis. Default = maximum value of x
ylow	minimum value on y axis. Default = 0
yh	maximum value on the y axis. Default = 1
...	Additional parameters to plot

Value

Data frame with columns

a	value of j*th* detect (ordered)
ple	PLE of F(x) at a
stder	standard error of F(x) at a
lower	lower CL for PLE at a
upper	upper CL for PLE at a
n	number of detects or non-detects $\geq$ a
r	number of detects equal to a

Note

If the solid horizontal line does not intersect the lower CL for the PLE, then the upper CL for Xp UX(p,γ) is not defined.

Author(s)

E. L. Frome

See Also

See Also [plekm](#plekm)

Examples

```
data(beTWA)
par( mfrow=c(1,2) )
ple.plot(beTWA)  #  plot the PLE of F(x) for the beTWA data
ple.plot(beTWA,ylow=0.8) #  plot the upper right tail
# Lognormal ML estimates of 95th percentile and CLs
unlist(percentile.ml(beTWA))
# PLE   estimates of 95th percentile and CLs
unlist(percentile.ple(beTWA))
#
```

---

pleicf   *Product Limit Estimate for Interval Censored Data*

Description

Compute Product Limit Estimate (PLE) of F(x) for interval censored data (i.e., the nonparametric maximum likelihood estimate that generalizes the Kaplan-Meier estimate to interval censored data).

Usage

```
pleicf(dd,nondet=TRUE,mine=1e-06,maxc=10000,eps=1e-14)
```

Arguments

dd	n by 2 matrix or data frame (see note below)
nondet	if TRUE, dd is left censored data
mine	minimum error for convergence in icfit. Default = 1e-06.
maxc	maximum number of iterations. Default is 10000.
eps	adjustment factor described by Ng. Default is 1e-14.

Details

This function is a driver function for [icfit](#) that uses an EM-algorithm applied to interval censored data (see Turnbull, 1976).

Value

Data frame with columns

a	value of j*th* uncensored value (ordered)
ple	PLE of F(x) at a
surv	$1 - F()$, i.e the "survival" or "exceedance" function
prob	prob[X = x] from icfit
n	sample size

Note

If nondet is TRUE column 1 of dd is the data value and column 2 is 1 if a detect and 0 otherwise. If nondet is FALSE dd contains the left and right endpoints required by [icfit](#).

Author(s)

E. L. Frome

References

Fay, M. P. (1999), "Comparing Several Score Tests for Interval Censored Data," *Statistics in Medicine*,18:273-85. (Corr: 1999, Vol 19, p.2681).

Ng, M. P. (2002), "A Modification of Peto's Nonparametric Estimation of Survival Curves for Interval-Censored Data," *Biometrics*, 58, 439-442.

Turnbull, B. W. (1976), "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data," *Journal of the Royal Statistical Society*, Series B (Methodological), 38(3), 290-295.

See Also

[icfit](#), [plend](#), [plekm](#)

Examples

```
# PLE for interval censored filmbadge data
data(filmbadge)
ple.fb<-pleicf(filmbadge[,1:2],FALSE) # PLE for input to qq.lnorm
tmp <- qq.lnorm(ple.fb) ; GM<-round(exp(tmp$par[1]));GSD<-round(exp(tmp$par[2]),2)
tp<-paste("Lognormal Q-Q plot for Filmbadge Data GM= ",GM,"GSD= ",GSD)
title(tp) # title for q-q plot with graphical parameter estimates
```

---

plekm	*Product Limit Estimate for Non-detects Using Kaplan-Meier*

Description

Compute Product Limit Estimate (PLE) of F(x) and Confidence Limits for data with non-detects (left censored data).

Usage

```
plekm(dd,gam)
```

Arguments

dd	An n by 2 matrix or data frame with x (exposure) variable in column 1, and det= 0 for non-detect or 1 for detect in column 2
gam	one-sided confidence level γ. Default is 0.95

Details

R function survreg is used to calculate Kaplan-Meier estimate of S(z), where z = k - x and k is greater than the largest x value. This technique of "reversing the data" to convert left censored data to right censored data was first suggested by Nelson (1972). conf.type = "plain" is required in survreg for correct CLs. The value of S(z) is then used to calculate F(x). Note that if γ = 0.95 the 90% two-sided CLs are calculated.

Value

Data frame with columns

a	is the value of j*th* detect (ordered)
ple	is PLE of F(x) at a
stder	standard error of F(x) at a
lower	lower CL for PLE at a
upper	upper CL for PLE at a
n	number of detects or non-detects $\leq$ a
d	number of detects equal to a

Note

In survival analysis S(x) = 1 - F(x) is the survival function i.e., S(x) = P [X > x]. In environmental and occupational situations S(x) is the "exceedance" function, i.e., S(x) = is the proportion of X values that exceed x. The PLE is the sample estimate of F(x), i.e., the proportion of values in the sample that are less than x.

Author(s)

E. L. Frome

References

Nelson, W.(1972), "Theory and Application of Hazard Plotting for Censored Failure Data", *Technometrics*, 14, 945-66

Frome, E. L. and Wambach, P.F. (2005) "Statistical Methods and Software for the Analysis of Occupational Exposure Data with Non-Detectable Values", *ORNL/TM-2005/52,Oak Ridge National Laboratory*, Oak Ridge, TN 37830. Available at: `http://www.csm.ornl.gov/esh/aoed/ORNLTM2005-52.pdf`

Schmoyer, R. L., J. J. Beauchamp, C. C. Brandt and F. O. Hoffman, Jr. (1996), "Difficulties with the Lognormal Model in Mean Estimation and Testing," *Environmental and Ecological Statistics*, 3, 81-97.

See Also

`plend`, `pleicf`

Examples

```
data(SESdata) ## use SESdata data set Example 1 from ORNLTM-2005/52
pkm<- plekm(SESdata)
qq.lnorm(pkm) #  lognormal q-q plot based on PLE
round(pkm,3)
```

| plend | *Compute Product Limit Estimate for Non-detects* |

Description

Compute Product Limit Estimate(PLE) of F(x) for positive data with non-detects (left censored data)

Usage

```
plend(dd)
```

Arguments

dd   An n by 2 matrix or data frame with
               x (exposure) variable in column 1, and
               det = 0 for non-detect or 1 for detect in column 2

Details

The product limit estimate (PLE) of the cumulative distribution function was first proposed by Kaplan and Meier (1958) for right censored data. Turnbull (1976) provides a more general treatment of nonparametric estimation of the distribution function for arbitrary censoring. For randomly left censored data, the PLE is defined as follows [Schmoyer et al. (1996)]. Let $a[1] < \ldots < a[m]$ be the m distinct values at which detects occur, r[j] is the number of detects at a[j], and n[j] is the sum of non-detects and detects that are less than or equal to a[j]. Then the PLE is defined to be 0 for $0 \leq x \leq a0$, where a0 is a[1] or the value of the detection limit for the smallest non-detect if it is less than a[1]. For $a0 \leq x < a[m]$ the PLE is $F[j] = \prod (n[j] - -r[j])/n[j]$, where the product is over all $a[j] > x$, and the PLE is 1 for $x \geq a[m]$. When there are only detects this reduces to the usual definition of the empirical cumulative distribution function.

Value

Data frame with columns

a(j)   value of j*th* detect (ordered)

ple(j)   PLE of F(x) at a(j)

n(j)   number of detects or non-detects $\leq$ a(j)

r(j)   number of detects equal to a(j)

surv(j)   1 - ple(j) is PLE of S(x)

Note

In survival analysis $S(x) = 1 - F(x)$ is the survival function i.e., $S(x) = P[X > x]$. In environmental and occupational situations $1 - F(x)$ is the "exceedance" function, i.e., $C(x) = 1 - F(x) = P[X > x]$.

Author(s)

E. L. Frome

References

Frome, E. L. and Wambach, P. F. (2005), "Statistical Methods and Software for the Analysis of Occupational Exposure Data with Non-Detectable Values," *ORNL/TM-2005/52,Oak Ridge National Laboratory*, Oak Ridge, TN 37830. Available at: http://www.csm.ornl.gov/esh/aoed/ORNLTM2005-52.pdf

Kaplan, E. L. and Meier, P. (1958), "Nonparametric Estimation from Incomplete Observations," *Journal of the American Statistical Association*, 457-481.

Schmoyer, R. L., J. J. Beauchamp, C. C. Brandt and F. O. Hoffman, Jr. (1996), "Difficulties with the Lognormal Model in Mean Estimation and Testing," *Environmental and Ecological Statistics*, 3, 81-97.

Turnbull, B. W. (1976), "The Empirical Distribution Function with Arbitrarily Grouped, Censored and Truncated Data," *Journal of the Royal Statistical Society*, Series B (Methodological), 38(3), 290-295.

See Also

[plekm](#), [pleicf](#), [qq.lnorm](#)

Examples

```
data(SESdata) #  use SESdata data set Example 1 from ORNLTM-2005/52
pnd<- plend(SESdata)
Ia<-"Q-Q plot For SESdata "
qq.lnorm(pnd,main=Ia) #  lognormal q-q plot based on PLE
pnd
```

---

qq.lnorm	*Quantile-Quantile Plot for Censored Lognormal Data*

Description

qq.lnorm produces a lognormal quantile-quantile (q-q) plot based on the product limit estimate (PLE) of the cumulative distribution function (CDF) F(x) for censored data. A line is added to the plot.

Usage

```
qq.lnorm(pl, mu, sigma, aveple = TRUE,...)
```

Arguments

pl	A data frame with the data(x) in column 1 and PLE in column 2
mu	estimate of the log scale mean
sigma	estimate of log scale standard deviation
aveple	if TRUE, calculate plotting positions by averaging
...	Additional parameters to plot

Details

The PLE is used to determine the plotting positions on the horizontal axis for the censored data version of a theoretical q-q plot for the lognormal distribution. Waller and Turnbull (1992) provide a good overview of q-q plots and other graphical methods for censored data. The lognormal q-q plot is obtained by plotting detected values $a[j]$(on log scale) versus $H[p(j)]$ where $H(p)$ is the inverse of the distribution function of the standard normal distribution. If the largest data value is not censored then the PLE is 1 and H(1) is off scale. The "plotting positions" $p[j]$ are determined from the PLE of F(x) by multiplying each estimate by $n/(n+1)$, or by averaging adjacent values–see Meeker and Escobar (1998, Chap 6)]. In complete data case without ties the first approach is equivalent to replacing the sample CDF j/n with $j/(n+1)$, and for the second approach the plotting positions are equal to $(j-.5)/n$. If the lognormal distribution is a close approximation to the empirical distribution, the points on the plot will fall near a straight line. An objective evaluation of this is obtained by calculating Rsq the square of the correlation coefficient associated with the plot.

A line is added to the plot based on the values of mu and sigma. If either of these is missing mu and sigma are estimated by linear regression of $log(y)$ on $H[p(j)]$.

Value

A list with components

x	The x coordinates of the points that were plotted
y	The y coordinates of the points that were plotted
pp	The adjusted probabilities use to determine x
par	The values of mu, sigma, and Rsq

Note

Helsel and Cohen (1988) consider alternative procedures that can be used for calculating plotting positions for left censored data. Waller and Turnbull (1992) describe a modification of the Kaplan-Meier estimator that can be used for right censored data and note that for the purpose of assessing goodness of fit the choice of plotting positions makes little qualitative difference in the appearance of any particular plot. The two options in this function can be used for any type of censoring.

Author(s)

E. L. Frome

References

Fay, M. P. (1999), "Comparing Several Score Tests for Interval Censored Data," *Statistics in Medicine*, 1999; 18:273-85. (Corr: 1999, Vol 19, p.2681).

Frome, E. L. and Wambach, P. F. (2005), "Statistical Methods and Software for the Analysis of Occupational Exposure Data with Non-Detectable Values," *ORNL/TM-2005/52,Oak Ridge National Laboratory*, Oak Ridge, TN 37830. Available at: http://www.csm.ornl.gov/esh/aoed/ORNLTM2005-52.pdf

Hesel, D. R. and T. A. Cohn (1988), "Estimation of Descriptive Statistics for Multiply Censored Water Quality Data," *Water Resources Research*, 24, 1997-2004.

Meeker, W. Q. and L. A. Escobar (1998), *Statistical Methods for Reliability Data*, John Wiley and Sons, New York.

Ny, M. P. (2002), "A Modification of Peto's Nonparametric Estimation of Survival Curves for Interval-Censored Data," *Biometrics*, 58, 439-442.

Waller, L. A. and B. W. Turnbull (1992), "Probability Plotting with Censored Data," *The American Statistician*, 46(1), 5-12.

See Also

[plekm](#), [plend](#), [pleicf](#)

Examples

```
data(SESdata) #  use SESdata data set Example 1 from ORNLTM-2005/52
pnd<- plend(SESdata)
qq.lnorm(pnd) #  lognormal q-q plot based on PLE
Ia <- "Q-Q plot For SESdata "
qqout <- qq.lnorm(pnd,main=Ia) #  lognormal q-q plot based on PLE
qqout
```

---

readss	*Read Analyze Data From ASCII File*

Description

Read data from fn.txt (space delimited text file) or fn.csv (comma delimited text file) and calculate all summary statistics using [IH.summary](#) . Output results to an ASCII text file fnout.csv in CSV format

Usage

```
readss(fn,L,bcol=NA,rto=5,pstat=NA,reverse=FALSE,p=0.95,gam=0.95,comma=FALSE)
```

Arguments

fn	name of input data file in double quotes without the .txt or .csv extension
L	L is specified limit for the exceedance fraction; e.g., the occupational exposure limit
bcol	Column that contains the BY variable–see details. Default NA
rto	Round values to rto. Default = 5
pstat	Select a subset of statistics calculated by [IH.summary](#).Dafault All
reverse	If reverse is TRUE reverse rows and columns in output file. Default=FALSE
p	probability for Xp the 100p*th* percentile. Default is 0.95
gam	one-sided confidence level γ. Default is 0.95
comma	if TRUE,the input file is in csv format with column names. Default is FALSE

Details

Read data from a tab or comma delimited text file in the current folder/directory. The first column must contain measurements (observed X values). The second column is an indicator variable with 1 for a detected value and 0 for a non-detect. Additional columns can contain factors that can be used to define a BY variable. The first record in the file must contain valid R names. Valid names may contain letters (case sensitive), numbers, period, and underscore and should start with a letter (no spaces). This file would most likely be obtained from an Excel spread sheet using the file "Save As" option, with file Save as type:
Text(Tab delimited)(*.txt) or CSV(Comma delimited)(*.csv).

Value

Returns invisible data.frame from file fn.txt

Column 1	value of measurement
Column 2	indicator variable ; 1 for detect 0 for non-detect
Column 3	. . . additional variables

Side effects

Summary statistics calculated by `IH.summary` are computed for each subset of data as defined by the levels of the BY variable. A data frame with row names from IH.summary (or subset based on value of `pstat`) and column names defined by the values of the BY variable is output as an ASCII text file in CSV format fnout.csv in the working folder. If `reverse` is TRUE the rows and columns are reversed.

Note

For information about `factor` see R help file `factor` Each level of the BY variable must have at least two non-detects for this function. If this is not the case an error message is printed to the R console and the levels of the BY variable with less than 3 non-detects are printed.

Author(s)

E. L. Frome

References

see the help file for `lnorm.ml`, `efclnp efraction.ml`, `percentile.ml`, `kmms`

See Also

`About-STAND` for more details and a complete reference list

Examples

```
# to demonstrate the use of readss add a new factor grp to the cansdata
# this factor with four levels (A_1 A_2 B_1 B_2) combines strata and sample
data(cansdata)
grp <- paste(cansdata$strata,cansdata$sample,sep="_")
temp <- data.frame(cansdata,grp) # add four level factor grp to cansdata

#    the next line is NOT executable  use CUT AND PASTE
#    sink("demoread.txt") ; print(temp) ; sink()

# The preceding line writes temp to a text file demoread.txt in the current folder
# This file would normally be created by another program, e.g. Excel
#   now use readss() to read this space delimited text file and calculate
#   all of the summary statistics for each level of grp and output
#   the results to a new text file demoreadout.csv in the current folder

#     rdemo <- readss("demoread",L=0.2,bcol=5)

# rdemo is the R data frame that was used to calculate results in demoreadout.csv
# to see same results rounded to three places in R console use
# round( IH.summary(rdemo,L=0.2,bcol=5), 3)

#  To select a subset of statistics from IH.summary first define the subset
#  psel<-c("Xp.obs","Xp","Xp.UCL","f","f.UCL","Rsq","m","n")
#  entering the following command will overwrite demoreadout.csv
#  with rows and columns reversed and the subset of statistics as columns
#  and the results will be rounded to 4 places
#  rdemo <- readss("demoread",L=0.2,bcol=5,rto=4,pstat=psel,rev=TRUE)
#
#  to see same results rounded to three places in R console use
#  t(round( IH.summary(rdemo,L=0.2,bcol=5)[psel,], 3))
```

SESdata   *Samples from Elevated Surfaces of a Smelter*

Description

The Department of Energy (DOE) Chronic Beryllium Disease Prevention Program is concerned with monitoring objects (e.g., equipment, buildings) for beryllium contamination and workers for exposure to beryllium in the workplace.

The SESdata is the results of a survey to evaluate possible beryllium contamination based on 31 surface wipe samples from elevated surfaces (SES) of a smelter at a DOE facility. For equipment that is being evaluated for release to the public, or for non beryllium use, the DOE has established a release limit for removable beryllium contamination of $0.2\mu g/100cm^2$.

Usage

```
data(SESdata)
```

Format

A data frame with 31 observations on the following 2 variables

x beryllium $\mu g/100cm^2$

detect 0 if non-detect; 1 if detected

Details

Statistics of interest are the exceedance fraction and the 95th percentile. The exceedance fraction is an estimate of the percentage of surface area that is expected to exceed the release limit Lp = 0.2 $\mu g/100cm^2$ with p = 0.95. Both the point estimate and the UCL for F exceed Fo = 100 (1-p) = 5%, indicating that the equipment is not acceptable. In fact, at the 95 confidence level at least 19.5% of the surface area exceeds the release limit.

A more detailed description and analysis of this data is given as Example 1 in Section 4 of Frome and Wambach (2005)

References

Frome, E. L. and Wambach, P. F. (2005), "Statistical Methods and Software for the Analysis of Occupational Exposure Data with Non-Detectable Values," *ORNL/TM-2005/52,Oak Ridge National Laboratory*, Oak Ridge, TN 37830. Available at: http://www.csm.ornl.gov/esh/aoed/ORNLTM2005-52.pdf

Examples

```
data(SESdata)
mle.ses <- unlist(lnorm.ml(SESdata)) # ML for SESdata
print(mle.ses[1:4])  #  ML estimates of parameters
print(mle.ses[5:8])  #  Standard errors of ML estimates
#  Next line produces a lognormal q-q plot with ML line
qq.lnorm(plend(SESdata),mle.ses[1],mle.ses[2])
title("Lognormal Q-Q plot For SESdata  Example 1 in ORNLTM2005-52")
unlist(efraction.ml(SESdata,gam=0.95,L=0.2))   #  MLE of exceedance fraction and CLs
unlist(percentile.ml(SESdata,p=0.95,gam=0.95)) #  MLE of 95 percentile and CLs
```

Index