EnvStats: the R Package for E-Sci

Armando Flores

11/06/2019

#### Introduction

“EnvStats” is an R Package created to provide statistical and graphical analyses of environmental data, with a focus on chemical concentrations and physical parameters in mandated environmental monitoring. Full of incredible functions, it contains detailed explanations of the environmental statistical methods used in the literature as well as regulatory guidance documents. Although designed for one purpose, “EnvStats” is a powerful set of tools for analysis of environmental data with an extensive user manual and example datasets.

#### Features and Basic Usage

* Functions for summary statistics and summary plots to compare distributions of groups side-by-side.
* Added probability distributions, as well as the ability to compute quantities associated with these distributions and generate random numbers from said distributions.
* Plot probability distributions to follow their change with the value of distribution parameter(s).
* Estimate distribution parameters and distribution quartiles, compute confidence intervals for probability distributions, including methods for lognormal and gamma distributions.
* Perform and plot the results of goodness-of-fit tests:
  + Observed and Fitted Distributions
  + Quantile-Quantile Plots
  + Results of the Shaprio-Wilk test, Kolmogorov-Smirnov test, etc.
* Functions for assessment of Box-Cox data transformations.
* Compute parametric and non-parametric prediction intervals, simultaneous prediction intervals, and tolerance intervals.
* Additional functions for hypothesis tests:
  + Nonparametric estimation and test for seasonal trends
  + Fisher’s one-sample randomization (permutation) test for location
  + Quantile test to detect a shift in the tail of one population relative to another
  + Two-sample linear rank tests
  + Test for serial correlation based on von Neumann rank test
* Perform calibration based on a machine signal to determine decision and detection limits and report estimated concentrations.
* Perform power and sample size computations and create companion plots for sampling designs based on confidence intervals, hypothesis tests, prediction intervals, and tolerance intervals.
* Handle singly and multiply censored (less-than-detection-limit) data.
  + Empirical CDF and Quantile-Quantile Plots
  + Parameter/Quantile Estimation and Confidence Intervals
  + Prediction and Tolerance Intervals
  + Goodness-of-Fit Tests
  + Optimal Box-Cox Transformations
  + Two-Sample Rank Tests
* Functions for performing Monte Carlo simulation and probabilistic risk assessment.
* Reproduce specific examples in EPA guidance documents by using built-in data sets from these documents and companion scripts.

#### Installation

##install.packages("EnvStats")   
# Installs the EnvStats package.  
library(EnvStats)   
# Pulls the installed package from the library into the working enviornment.

#### Examples

# Generate a purely random normal process, then use serialCorrelationTest  
# to test for the presence of correlation.  
# (Note: the call to set.seed allows you to reproduce this example.)  
set.seed(345)  
x <- rnorm(100)  
# Look at the data  
#-----------------  
dev.new()  
ts.plot(x)  
dev.new()  
acf(x)  
# Test for serial correlation  
#----------------------------  
serialCorrelationTest(x)

##   
## Rank von Neumann Test for Lag-1 Autocorrelation (Beta  
## Approximation)  
##   
## data: x  
## RVN = 1.9297, p-value = 0.7253  
## alternative hypothesis: true rho is not equal to 0  
## sample estimates:  
## rho   
## 0.02773737

#### Resources

An incredibly detailed document that exhaustingly goes over several functions, explanations, and examples.  
**EnvStats-manual.pdf**:<https://cran.r-project.org/package=EnvStats/EnvStats.pdf>

A book was written by the R Package author, Steven P. Millard. **“EnvStats: An R Package for Environmental Statistics”**:<https://www.springer.com/gp/book/9781461484554#aboutAuthors>