Updates since publication of the 3rd edition of my book (2012) on robust methods

clnorm generates data from a contaminated lognormal distribution

G-AND-H DISTRIBUTION

ghtrim Computes trimmed mean of a g-and-h distribution when $g > 0$. (With $g=0$, trimmed mean is zero.)

rngh: another way of generating multivariate data with specified correlation

BINOMIAL:

acbinomciv2 Computes a confidence interval for $p$ using the Agresti-Coull method.

twobicipv adds a p-value when using Beal’s method

SIGN TEST:

For the functions signt and signtpv, setting AC=T, the Agresti-Coull method is used; otherwise Pratt’s method is used.

MEDIAN

hdpb computes a confidence interval for any quantile using the Harrell-Davis estimator. (In contrast, hdci uses a bootstrap estimate of the se.) Supplied mainly for convenience. This function is essentially the same a onesampb, only it defaults to using the Harrell-Davis estimator.
Also stored as qcipb

NORMALITY AND NUMBER OF OUTLIERS

BASED ON MAD-MEDIAN RULE, IS THE NUMBER OF OUTLIERS UNUSUAL FOR WHAT IS EXPECTED UNDER NORMALITY? USE THE FUNCTION

\texttt{normTmm}

\texttt{medhd2g}: For convenience, compare two independent groups using Harrell-Davis estimator by calling \texttt{pb2gen}

twoKlin Test for \(K\) independent groups using step-down Fisher p-value method plus Hochberg to control FWE Increased power compared to using robust T3, better power but no confidence intervals.

\texttt{ptests}: Fisher method. Have p-values based on \(N\) independent tests, test the hypothesis that all \(N\) tests are true.

EFFECT SIZE: \texttt{med.effect} computes a version of the explanatory measure of effect size based on medians and the percentage bend midvariance.

\texttt{D.akp.effect}: function(x,null.value=0,\(tr=.2\)) Computes the robust effect size for one-sample case using a simple modification of Algina, Keselman, Penfield Psyc Methods, 2005, 317-328

OUTLIERS AND USUAL MEAN-COVARIANCE MATRIX \texttt{wmean.cov} for multivariate data, computes mean and covariance matrix, used when detecting outliers via standard
Mahalanobis distance—demonstration purposes only So out(x,outfun=wmean.cov) will use Mahalanobis distance

**ANCOVA:**

ols2ci OLS version of reg2ci

ancov2COV does two covariates using all of the deepest covariate values. Nonparametric global test can have more power than ancovampG.

CLASSanc: does classic ANCOVA

cidmulv2 uses Hochberg, better power than cidmul

ancovaG: Generalizes the R function ancova so that any hypothesis testing method can be used to compare groups at specified design points.

ancovampG generalizes ancovamp so that any hypothesis testing method can be used to compare groups at specified design points.

ancdes: updated so that when using ancovampG, a broader range of covariate values can be used.

When using ancovampG, setting the argument DH=T, the deepest half of the covariate points will be used when comparing groups. The fraction of covariate points can be altered via the argument FRAC

ancovaWMW like ancova, only use Cliff modification of the Wilcoxon-Mann-Whitney
reg1way: For two or more independent groups, test the global hypothesis that all of the corresponding parameters are equal. That is, the intercepts are equal, all of the slopes for first predictor are equal, etc., as opposed to performing individual tests for the intercepts, etc. Can use any robust regression estimator, default is Theil–Sen

Also see ancGpar below: Designed for two groups

ancGLOB: like ancova, allows curvature, but tests the global hypothesis that $M(Y - X=x)$ is the same for two independent groups based on a collection of x values.

REGRESSION

regse: computes a bootstrap estimate of the squared standard errors of the estimator used. By default, Theil–Sen is used.

CURVATURE: rplot.res applies rplot and then plots res (dep var) vs the omitted variables.

regcits FOR convenience, defaults to multicore processor with C version of tsreg

reg1wayISO test equality of slopes only. That is, do not include the hypothesis that the intercepts are equal. For convenience, for the two group case, can use ancGpar.

reg1wayMC and reg1wayISOMC use a multicore processor.

regGmcp does all pairwise comparisons where the null for each pair is that all parameters
are equal.

FOR OLS, use ols1way; ols1wayISO compares slopes only.

ancGpar(x1,y1,x2,y2): Supplied for convenience. Same as reg1way only it handles two
groups only and data are not stored in list mode as assumed by reg1way

ancGparMC uses multicore processor.

For OLS, use olsJ2

ancpar: For two independent groups, compare Y hat at specified design points. One
or more covariates is allowed. It fits a robust parametric linear model. By default, uses
Theil–Sen. Covariate values are chosen as done by the function ancova if none are specified.
Basically, a robust analog of the Johnson-Neyman method. This function uses the functions
ancts and anctsmcp.

ancJN ANCOVA, robust analog of the Johnson–Neyman method: compare predicted Y
values at specified points based on a parametric fit. Same as ancts but this name better
reflects the goal.

ancJNmcp same as anctsmcp, but changed name to better reflect goal

MCP for comparing Y hat values among J independent groups.

reg2difplot: plot the difference in predicted Y values as a function of points in pts.

ANCOVA: INDEPENDENT GROUPS
ANCova: DEPENDENT GROUPS

ancovaV2: like ancova, but use a bootstrap and resample from the joint distribution for each group. All indications are that this can increase power compared to ancova.

Dancova: Same as ancova (nearest points using running interval smoother) but designed for two dependent groups. Trimmed means are used via function yuend or trimci.

Dancovapb: Percentile bootstrap analog of Dancova

DancovaV2: Dependent group analog of ancovaV2. Power advantage.

library(’devtools’) install github( ’WRScpp’, ’mrxiaohe’)

CORRELATION:

compare robust correlations, overlapping case: twoDcorR

Non-overlapping case twoDNOV

scor.C Uses outpro.C or outmgv.C

so corb(x,y,corfun=scor.C) fast ci using skipped correlation. Assumes C++ routines have been installed.

scorci.C Use cpp version of scor

scorciMC
Compute a confidence interval for the skipped correlation using a bootstrap method that allows heteroscedasticity. Access to a multicore processor is assumed.

scorci: same as scorciMC, only does not use a multicore processor.

qcor computes measure of association based on fit obtained via the quantile regression estimator.

ANOVA:

t1way, t2way and t3way:

With version 18, updated t1way, t2way and t3way. They now contain arguments IV1, IV2 and IV3, respectively that, if supplied, are taken to be the independent variable. That is, the group levels for each factor. If specified, x is now taken to be a vector (the dependent variable).

aov2depth is for main effects of FACTOR A in 2 by K ANOVA. Useful in various settings where the pattern of the location estimates is important. See for example ancGLOB

Qanova: one-way design, test global hypothesis that all quantiles are equal, performs well with tied values

For comparing slopes, see REGRESSION below

PLOTS and Explanatory power:

g5plot: Plots the distribution for up to five variables using a kernel density estimator.
s2plot plots two discrete distributions, generalizes splot

WITHIN GROUPS ANOVA USING AFFINE EQUIVARIANT ESTIMATOR:

rmdzeroG and rmdzeroGMC NOTE: Not so good when using affine equivariant estimator.

q2by2 2-by-2 ANOVA all main effects and interactions using quantiles q=c(…)

INTERACTIONS:

On 12/31/14 a bug in the function adrun was corrected that impacts the function adtest.

ANCOVA: ancdifplot Plot the difference between the estimates of the dependent variables plus a confidence band.

ancdifplotj-function(x1,y1,x2,y2,fr1=1,fr2=1,tr=.2,alpha=.05,pr=TRUE,xout=FALSE,outfun=out,LP=TRUE,
nmin=8,scat=TRUE,xlab='X',ylab='Y',report=FALSE,...)

rplot has been updated: For one or two predictors, when plotting setting argument LP=FALSE results in the usual plot. LP=TRUE, the regression surface is further smoothed via lplot

rplotCI. For one predictor, same as rplot but includes a 1-alpha confidence band.

rplotpbCI Like rplotCI, only use percentile bootstrap

lplotPV tests the hypothesis that explanatory power, based on LOESS, is zero. It returns a p-value. It is sensitive to heteroscedasticity, so it tests the hypothesis of no association.
lplot.pred Predicted values based on fit returned by loess

qregplots Plots one or more parametric quantile regression lines.

qhdsm does a smooth for one or more quantiles using hd, running interval followed by LOESS

difQplot Plot that provides perspective on the degree a distribution is symmetric about zero. This function plots the sum of q and 1-q quantiles. If the distributions are symmetric the plot should be approximately a horizontal line. If in addition the median of the difference scores is zero, the horizontal line will intersection the y-axis at zero.

difQpci. Like difQplot, only use fewer quantiles and compute p-values.

qhdplotsm Plots smooths of quantile regression lines for one or more quantiles using rplotsm with Harrell–Davis estimator (bootstrap bagging)

q indicates the quantiles to be used.

EXAMPLE: qhdplotsm(x,y,q=c(.2,.5,.8)) will plot three smooths corresponding to the .2, .5 and .8 quantile regression lines.

qhdsm Also plots smooths of quantile regression lines for one or more quantiles but it smooths the initial fit via lplot when the argument LP=T.

qhdsm2g plots quantile smooth via qhdsm for two groups.

ols.pred.ci
plot the ols regression line and a 1-alpha confidence interval for the predicted values

reg2plot: Plot two regression lines.

regp2plot: For two predictors, create scatterplot and plot regression surface. By default, use Theil–Sen

reg2g.p2plot: For two groups, and two predictors, create scatterplot and plot regression surfaces. By default, use Theil–Sen

reg2difplot: For two groups and two covariates, Fit a regression model to both groups assuming have two predictors. Get predicted Y values based on points in pts. By default, use pts=x1

x1 a matrix containing two predictors x2 a matrix containing two predictors

Compute differences in predicted values and plot the results as a function of the points in the argument pts

COMPARING GROUPS (FOR COMPARING REGRESSION MODELS, SEE REGRESSION.)

trimcimul

For J dependent groups, perform trimci on each column

M1M2: When comparing two groups, might want to compare the groups using two or more dependent variables. For example, imagine that matrices x1 and x2 have 6 columns
containing dependent variables of interest. Goal: compare groups using data in col 1 of x1 and x2, do same for col 2 data, etc. M1M2(x1,x2,yuen,alpha=.05) will compare the groups for each of the 6 dependent variables using Yuen’s test by default.

ddepv2 is an updated version of ddep that does a better job of handling heteroscedasticity among the marginal distributions. This new version uses a weighted estimate of the grand mean, the weights being based on bootstrap estimates of the squared standard errors.

ddeptr is a modification of ddep designed specifically for trimmed means. Unlike ddep, a weighted estimate of the grand mean is used. This helps deal with the heteroscedasticity among the marginal distributions.

qcomhd Compare quantiles of independent groups using Harrell-Davis estimator and percentile bootstrap method. Works very well with tied values, particularly when comparing quantiles close to 0 or 1—much better than the shift function.

cbmhd: two independent groups, compare quantiles using generalization of Wilcoxon-Mann-Whitney qwmwhd: uses a range of quantiles.

Dqcomhd Compare marginal quantiles of dependent groups using HD, tied values OK

Missing values: can do complete case analysis or use all available data.

Dqdif Compare two dependent groups by comparing the q and 1-q quantiles of the difference scores

q should be < .5 if the groups do not differ, then the difference scores should be symmetric
about zero. In particular, the sum of q and 1-q quantiles should be zero.

difQpci: applies Dqdif to a collection of q values.

Dependent groups Plot that provides perspective on the degree a distribution is symmetric about zero. This function plots the sum of q and 1-q quantiles. A 1-alpha confidence interval for the sum is indicated by a + If the distributions are symmetric the plot should be approximately a horizontal line. If in addition the median of the difference scores is zero, the horizontal line will intersect the y-axis at zero.

Similar to difQplot, only plots fewer quantiles by default and returns p-values for each quantile indicated by the argument q.

Dependent groups: yuendv2 is like yuend only it also returns a measure of effect size.

bsqrm is a test statistic for comparing two independent groups using an M-estimator, derived by F. Ozdemir bsrcmbt computes a critical value based on the test statistic computed with bsqrm. The method performs very well in simulations, in terms of Type I errors

bwmarpb: does between by within multiple comparisons using marginal measures of dependent groups and a percentile bootstrap method

disc2com for discrete data with small sample space: A global test of P(X=x)=P(Y=x) for all x.

binband For two independent groups, compare cell probabilities of a multinomial distri-
PREDICTION ERROR

rplotCV uses a running interval with leave-one-out cross validation.

SMpre uses a running interval with leave-one-out cross validation for all possible models, number of predictors is assumed to be $p \leq 5$.

DATA MANAGEMENT

idmatchv2. Identify matching ID’s from two files and match the corresponding variables. If there are missing values, have the option of including them.

mch2num converts character data, stored in a matrix, to numeric

rmblo(x,y) Removes only bad leverage points and returns the remaining data.

regunstack Like unstack but returns ind and dep variables in list mode. Convenient when using olsmcp and related functions.

args(regunstack) function (x, grp, xcols, ycol) grp indicates column of x containing group id xcols indicates columns of x containing independent variable, etc.

Psychometrics:

cofalpha

does coefficient alpha via the R package psych library(psych)
Rcoefalpha returns a robust analog of alpha.

REGRESSION INTERACTION

olshc4.inter convenient for testing $H_0$ no interaction using usual product term and OLS

regei.inter convenient for testing $H_0$ no interaction using usual product term using robust estimator.

REGRESSION

Rfit has been added; it uses the R package Rfit that computes a rank-based estimate of the regression parameters. Caution: the hypothesis testing method assumes homoscedasticity and tied values among the dependent variable can result in poor power.

regYci

Compute confidence interval for the typical value of Y, given X, based on the Theil–Sen estimator

plotPV = T will plot p-values for all the points.

tsregF faster version of tsreg: it breaks when estimates change by less than argument tol=.0001

tshdreg: Theil-Sen, but uses Harrell-Davis estimator rather than the usual sample median. Also, the intercept is taken to be the median of the residuals.
As previously indicated under ANCOVA: reg1way and reg1wayMC TEST $H_0$: all regression parameters among $J$ independent groups are equal. That is, equal intercepts, equal slope coefficients for each of $p$ predictors. Allows both types of heteroscedasticity.

For OLS, use ols1way $J \geq 2$ groups or olsJ2, two groups only.

reg2ciMC: same as reg2ci, but uses a multicore processor

reg1mcp For $J$ independent groups, do all pairwise comparisons of the intercept, then the slopes for the first independent variable, then the second independent variable, etc. MC=T will use a multicore processor if available. Uses Theil-Sen by default.

regYhat: for convenience, estimate $Y$ given $x$ based on a specified regression model.

For MCP when comparing slopes (and intercept) of $J$ independent groups, when using OLS, use the R function

olsmcp: It seems best among methods considered so far. It does pairwise comparisons of intercepts follows by first slope, etc. Allows both types of heteroscedasticity.

For pairwise comparisons of regression models (testing the hypothesis that simultaneously the corresponding intercepts and slopes among $J$ independent groups are identical, use

olsLmcp. same as olsJmcp

For pair of groups, it tests the global hypothesis that all of the corresponding parameters are equal. That is the regression models are identical. FWE controlled by Hochberg method.
olsWmcp All pairwise comparisons for each parameter among J independent groups

stsreg and tstsreg have been generalized to allow more than one covariate

tssnmreg has been added. It is like tstsreg but uses snmreg when checking for regression
outliers; it can have lower execution time than tstsreg.

COMPARE regression parameters of two dependent groups.

DregG: defaults to tsreg GLOBAL test.

DregMC: multicore processor version of DregG

DregGOLS OLS version

Dancols: Compare the OLS regression lines of two dependent (within) groups at specified
design points. Analog of Johnson–Neyman method.

Dancts: Compare robust regression lines of two dependent (within) groups at specified
design points. Analog of Johnson–Neyman method. Use bootstrap estimate of se of Y hat

Danctspb and DanctspcMC Like Dancts, only use bootstrap samples where points are
resample, similar to regci

difreg: For two dependent groups, test hypothesis that the parameters are equal. A test
is performed for the intercepts, the first slope, etc. difregMC uses multicore processor.

difregOLS is OLS version
Dancovamp: like ancovamp, only designed for dependent groups.

Quantile Regression

For discrete dependent variable, instances arise where qreg fails to run. Try Qreg if this happens.

C++ versions of some of the functions are now available thanks to Xiao He.
They are stored in the file Rallfun_C.txt on my web page.
They include tsreg_C, tshdreg_C, stsreg_C, tstsreg_C, mgvout_C, outpro_C, fdepthv2_C and so on.

#First install two packages:
install.packages("Rcpp")
install.packages("RcppArmadillo")
library(RcppArmadillo)

#Import the source code "Rallfun_C.txt", which includes functions that call the C++ functions:
source(file.choose())  #choose Rallfun_C.txt

dynamically load the compiled C++ code:
dyn.load(file.choose())  #Choose Cfile_so.txt

EASIER:
library('devtools')

install_github('WRScpp', 'mrxiaohe')

library(WRScpp)