cufunctions **Crib Sheet (rev. Mar 2024)** *Red Italics mean changeable*

In Excel, variable **name** at top of each column – with NO SPACES, hyphens, or special characters except underscore _
Highlight data (easy to do columns – use **skip/nrows** below if columns have other stuff), copy, go into RStudio.

Read Data into R: *Met* **= curead()**

              NEJM **= curead(skip=*S*, nrows=*N*)** to skip *S* lines, to read *N* lines of data (after variable names)

Always *check #variables, #observations in* **curead** *output. Proofread. To see if any missing,* **str(na.omit(*NEJM*))**

attach(*NEJM*) to avoid unnecessary typing like *NEJM*$*variable*   or   **data=*NEJM***
To switch to a second dataset, you need to **detach(*NEJM*)** before next **attach**, but only if variable names repeat
No need to **attach** for **curepmeas**, **culinreg**, **culogist**, **cucox**.
If you save workspace when you quit RStudio with **q("yes")***, you can later resume where you left off.*
This is especially convenient if your Excel data don't change, so you read into RStudio with **curead** *just once.*
Help available if you type **? cu***function-name*   http://biomath.net/cufunctions.html to list all **cufunctions**
To generate R commands for cufunctions with many arguments, http://biomath.net/resources/cufun-rgen.html may help.

Esc Hit the Esc key in the middle of typing a command if you want to start over (especially if you see a "**+**")
Summary: **str(*NEJM*)** for overview of dataset (what **curead** puts out)
            cutable1(*tgstudy*) for single variable,   **cutable1 (*NEJM*, brief=*T*)** for entire dataset briefly
            cutable1(*tgstudy*, *Diet*)   **cutable1(*NEJM*, *Diet*, compare=*T*)** to summarize, compare subgroups
            hist(*tcpre*, breaks=#) for histogram (optional parameter **breaks** to set # of cut points)
Correlations for some: rcorr(cbind(*tcpre,tgpre,hcpre*)) for Pearson parametric correlation
                  rcorr(cbind(*tcpre,tgpre,hcpre*), type="spearman") for Spearman rank correlation
Correlations for all: **cutable1(*NEJM*, docorr=*T*)**
To define new variables: ***NEJM$tcpmm = NEJM$tcpre/38.67***
                 ***NEJM$tgpcat* = cut(*NEJM$tgpre,* c(*0,75,100,150,200*))**
To select a subgroup with all variables: *nolean* **= subset(*Met*, *WTCAT* =="*overwt*" | *WTCAT* =="*obese*")**
To select a few variables with all subjects: *metfew* **= *Met*[, c("*TG*", "*WTCAT*", "*Metsyn*")]**

cufunctions for **Continuous Variables** *(http://biomath.net/resources/cufun-rgen.html may be of use)*:

Get complete list of optional arguments with ? cu*function-name*
Scatter plot: **cuscatter(*tcstudy*, *tcpre*)**
Single Factor (one-way): cu1way(*tcchange*, *Diet*) for one-way anova + FisherLSD group comparisons + bar graph
            ,ebars=1 (default) for SD on graph, **2** for SE, **3** for Conf.Int, *or* **4** for nonparametric Dunn (IQR)
            ,dots=1 for data points as well; **,fill="grey" ="black"** etc for single color
            ,plot="box" for boxplot **="violin"** for violin plot **="no"** for no plot
            ,psignif=*P* for **C**ompact **L**etter **D**isplay on bars to indicate *p<P* (**0** for no CLD)
            ,g1order=c("*name1*","*name2*","*…*") for reordering the factor levels
            ,padj="bonferroni" for Bonferroni; **,padj="holm"** for Holm (default is **none**)
Two Factors (two-way): cu2way(*tcchange*, *Diet*, *sex*) [options as with **cu1way**; g2order for 2nd factor]
            ,interact=T (default) for ***Diet*sex*** interaction model; **=F** for no-interaction model
Anal. of Covariance: **cucov1way(*tcstudy*, *tcpre*, *Diet*)** for regression (no interaction), plot, contrasts
Ancova w. Interaction: **cucov1way(*tcstudy*, *tcpre*, *Diet*, c(*150,220*))** for regression, plot, contrasts at *150,220*
Ancova with 2nd factor: **cucov2way(*tcstudy*, *tcpre*, *Diet*, *sex*)** for regression, plot, contrasts
            ,interact=T (default) for Diet*sex interaction model; **=F** for no-interaction model
Ancova w. 2nd factor & Interaction w.covariate: **cucov2way(*tcstudy*, *tcpre*, *Diet*, *sex*, c(*150,220*))** for regression, plot, etc
Repeated Measures: **curepmeas(*delta*, "*TG*", "*Diet*", "*sex*",cov="age+sex*age", interact=*T*, sumdiff=*1*, mainx=*1*)**
            (1 line per subject, with a *TG* for each level of Diet: *TG.AAD TG.LowSat TG.Step1*)
            "*sex*" optional second fixed factor [other options as with **cu2way, cu1way** above]
            , cov= optional covariates besides the second fixed factor (in double quotes) as in **culinreg**
            , sumdiff=1 (default) or **=-1** for difference bar graph; **=0** for no difference bar graph
            , mainx=1 (default); **=2** to interchange the two factors in bar graph
Multiple Regression: **culinreg(*Met*, "*LDL*", "*TG+TC+HDL+LN_TG+BMI*")**
   display/dredge Options for **curepmeas** *and* **culinreg***:* **same as for culogist (Logistic Regression)**

cufunctions for **Categorical Variables** *(http://biomath.net/resources/cufun-rgen.html may be of use)*:
Get complete list of optional arguments with ? cu*function-name*
Chi-square/Fisher for 2x2s: **cu1way, cu2way, curepmeas** *as described above under continuous variables*
                        ,ordinal=c("…") to treat variable as ordinal;   **,scale="percent"** to show % instead of n
Kaplan-Meier: **cukm(*time*, *status*, *Treatment*)** for survival curves (default)
               ,kmtype="ci" for cumulative incidence curves
               ,pvalue=F to not show p-value on graph; **,censor=F** to not show censored points;
               ,scale="percent" for % instead of fraction; **,legend="bottom"/"left"/"right"/** instead of **"top"**

Logistic Regression: **culogist(*Met*, "*MetSyn*", "*HDL+LN_TG+BMI+GLUC+INS*")** to model odds
                     , logitlog="log" to model risk
  display Options: **, xs=c(*0.4*, *0.6*)** list of probability cut points for 2x2 tables (besides 0.5)
                     , printfit=T (default) to get (**T**) or not get (**F**) all fitted values/probabilities/risk-scores
                     , ndecfit=0 (default) to specify decimals in fitted value display (not for cucox)
                     , color="blue" (default **"red"**) for ROC curve color
  dredge Options: **, dodredge=T** (default) to get (**T**) or not get (**F**) all possible models
                     , usemod=*2* to use the usemod-th best found by dredge (default **1**)
                     , m.min=*2*, m.max=*3*   only models with at least 2 and at most 3 predictors
                     , fixed=c(*"INS", "BMI"*) only models that contain both INS and BMI
                     , subset=*!(HDL&LN_TG)* no models that contain both HDL and LN_TG

Cox Proportional Hazards:
               cucox(*coxdata*, "*TimeToEvent*", "*Outcome*", "*C.Index + scale(LVEF) + scale(BNP) + gender*")
        ,fun="pct"(default) to graph % survival; **="event"** for cumulative events; **="cumhaz"** for cumulative hazard
    display/dredge Options: **same as for culogist (Logistic Regression)**

Propensity Matching: **require(MatchIt)** *once in a session, then*
        matched.pro = **match.data(matchit(*tertiary* ~ *score*, data=*propen*, method="exact", ratio=1))**
To use matched dataset, add **,wtnam="weights"**: **culogist(*matched.pro*, "*outcome*", "*tertiary*", wtnam="weights")**
cucox(*matched.pro*, "*TimeToEvent*", "*Outcome*", "*C.Index + scale(LVEF) + scale(BNP) + gender*", wtnam="weights")

A Few Other R Functions to Know:
One Group:
1-sample t-test: **t.test(*NEJM*$*tcpre*, mu=*180*)** to compare baseline with 180 mg/dL
Paired t-test: **t.test(*MM*$*tcpre*, *MM*$*tcstudy*, paired=T)** same as **t.test(*MM*$*tcchange*)**
Wilcoxon signed rank: **wilcox.test(*MM*$*tcpre*, *MM*$*tcstudy*, paired=T)** same as **wilcox.test(*MM*$*tcchange*)**
Binomial test of observed proportion: **binom.test(*#in one group*, *#total*, *expected proportion*)**
Two Groups:
Unpaired t-test: **t.test(*tcchange* ~ *Diet*, var.equal=T, data=*AM*)**
Wilcoxon rank sum: **wilcox.test(*tcchange* ~ *Diet*, data=*AM*)**

To clean out workspace: **remove(list=ls())**, *then* **ls()** *to confirm workspace is empty, then* **Quit** *saving workspace.*
Exporting R Output:
Console output: Highlight text, copy to clipboard, paste elsewhere
Graphs: Export/Copy to Clipboard, paste elsewhere; Editable Graphs in Windows: choose **Metafile**
Editable Graphs on Mac, add argument: **,emf=T** *to create* **cu*function-name*.emf; =name.emf** *(to change name)*
      "Open-Office" Presentation: Insert/Picture/from File; Right-click/Break; Save as PowerPoint

When installing **RStudio for the first time**, *install needed add-on packages and .Rprofile:*
reqpac = c("e1071","dunn.test","ggplot2","gmodels","ROCR","Hmisc","MuMIn","MatchIt","car","survminer","devEMF")
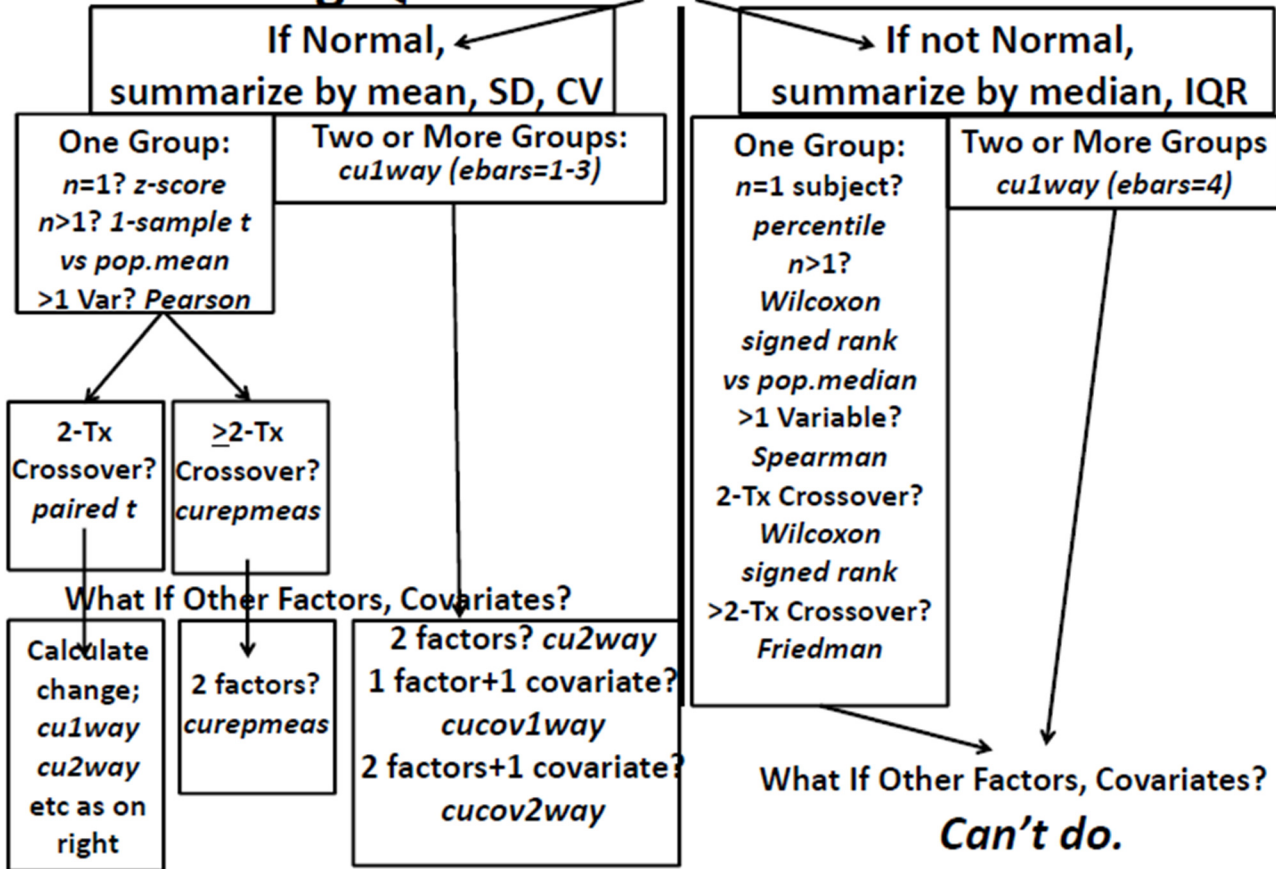**install.packages(reqpac, dependencies=T, type="binary", quiet=T, verbose=F)**
download.file("http://biomath.net/resources/.Rprofile", ".Rprofile", quiet=T)

Installing or Updating our cufunctions package:
install.packages("http://biomath.net/resources/cufunctions.tar.gz", repos=NULL, type="source")
[For the new installation to take effect, you have to exit RStudio and relaunch. Same for updating other packages.]

Asking Questions of Continuous Data

If Normal, summarize by mean, SD, CV	If not Normal, summarize by median, IQR

One Group:
n=1? *z-score*
n>1? *1-sample t*
vs pop.mean
>1 Var? *Pearson*

Two or More Groups:
cu1way (ebars=1-3)

One Group:
n=1 subject?
percentile
n>1?
Wilcoxon
signed rank
vs pop.median
>1 Variable?
Spearman
2-Tx Crossover?
Wilcoxon
signed rank
>2-Tx Crossover?
Friedman

Two or More Groups
cu1way (ebars=4)

2-Tx Crossover?
paired t

≥2-Tx Crossover?
curepmeas

What If Other Factors, Covariates?

Calculate change;
cu1way
cu2way
etc as on right

2 factors?
curepmeas

2 factors? *cu2way*
1 factor+1 covariate?
cucov1way
2 factors+1 covariate?
cucov2way

What If Other Factors, Covariates?

Can't do.

Extending 2x2 Analysis
biomath.net/stat/, cu1way,cu2way,curepmeas

	Control	Case
Risk Factor Low		
Risk Factor High		

What If >1 Factor?

Time to Caseness?
Censored Data?

Logistic Regression
culogist

Kaplan-Meier Analysis
cukm

If >1 factor AND time to caseness
Cox Proportional Hazards Modeling
cucox

http:/www.biomath.net/resources/
Sekhar Ramakrishnan sekhar@caa.columbia.edu   Steve Holleran sfh2@columbia.edu