Please consider providing feedback to Bob Obenchain on his JMP scripts for "Local Control" and "Artificial LTD Distributions"

Bob Obenchain, PhD, FASA, Principal Consultant
Risk Benefit Statistics LLC
13212 Griffin Run, Carmel, IN 46033-8835
(317) 580-0144, wizbob@att.net

Getting Started:

The distribution archive includes an icon (LC.BMP) if you want to add "LocalControl" and "Artificial LTD Simulation" items to your JMP menu bar. (Alternatively, use one of the many icons provided by JMP …such as the "Magnifying Glass.") Simply select "Edit-Customize-Menus" within an active JMP session. Personally, I use the fifth and sixth slots on the "Analyze" menu for these items, and I keep all of my LC files in a folder with the following path and name: "C:\Program Files\SAS\JMP\8\Local Control".

Otherwise, each time that you want to use the "LocalControl" script, you will need to first open the *.JSL file (from wherever you decided to store it) and then click on the JMP "Run Scrip" icon. The script will then start up in single-step "Debug Mode" unless you edit the "LocalControl.JSL" file (using the JMP script editor) by deleting its first line of /*debug step*/.

Display Captions? Augment Captions with Audio?

If you want the script to display on-screen prompt dialogs, you will need to edit the "LocalControl.jsl" file as follows.

// If showCaps is nonzero, show captions. Default is showCaps = 0.
showCaps = **1**;

// If isSpoken is nonzero, shown captions will also be spoken on systems that support audio
isSpoken = 1;

If you want to see on-screen prompt dialogs, but do NOT want to hear them "spoken," set… isSpoken = **0**;

Fundamental Notation:

y = observed outcome variable(s)
t = observed treatment assignment (usually non-random)
x = observed baseline covariate(s)

Phases of a "Local Control" Analysis using the JMP Script:

[1] Launch the script and open a JMP data table.

[2] Select 1 or 2 ***y***-outcome variables, a binary treatment ***t***-variable, and 1 or more numeric ***x***-covariates measured at baseline by, first, highlighting their variable names and, then, by adding them to 1 of the three groupings (***y***, ***t*** or ***x***) within the pop-up "Select Columns" dialog box. Next, before pressing the OK button, be sure to punch the radio button for either Homoskedastic or Heteroskedastic within-cluster-and-treatment-group variability.

[3] Once patients have been hierarchically clustered in ***x***-space, you will be prompted to specify a desired number of clusters in a pop-up dialog box. The initial default number of clusters will usually be 50, but a range of values from only 1 to as many as one-half of the total number of patients is allowed. One usually starts by selecting either the default number of clusters or 1 cluster.

[4] Once three different values for "Number of Clusters" have been specified, the script will start displaying and updating an outcome "Sensitivity Analysis" graphic. This plot shows 3 summary statistics of the LTD distribution versus the number of clusters on a log scale. The 3 statistics plotted are the estimated **mean** of the LTD distribution, the mean plus 2 sigma-hat **upper limit**, and the mean minus 2 sigma-hat **lower limit**. Explore as many different values for number of clusters as you wish. Calculations consume more and more time and memory as the number of clusters increases. There is no real need to explore the very largest numbers of clusters if (a) the upper limit to lower limit band is getting much wider and/or (b) the Number of "Informative Clusters" (NCinf) displayed in the Sensitivity Analysis table (data_SA.JMP) is not increasing significantly.

[5] When you have finished exploring alternatives for "Number of Clusters," press the "Cancel" button on the pop-up dialog box. The Sensitivity Analysis table (data_SA.JMP) will then usually be displayed on top of all the others. If not, locate this table within the bottom section of the JMP "Window" menu. The final SA graphic(s) will already have been saved to the open JMP Journal file, but one (or both) can be regenerated from the script(s) embedded within the "data_SA.JMP" table.

[6] Again, using the bottom section of the JMP "Window" menu, you will usually want to more carefully examine the implied LTD distribution for one (or more) specific Number of Clusters requested (NCreq.) Locate the corresponding "LC_NCreq.JMP" table and either run the "LTDdist[1|2]" script embedded within it (to display the LTD distribution in a normal Q-Q plot

as well as a weighted histogram and to interactively kernel-smooth this distribution) or else generate other sorts of analyses or graphics using the JMP menus.

[7] To explore the question of whether the X-covariates used to form clusters are also predictive of the LTD distribution itself, run the "LTDjoin" script embedded within a LC_NCreq.JMP" table. Your original data table now has additional variables (columns) with names of the form "C_NCreq" that contain the cluster number for each patient for each different value of NCreq explored. The "LTDjoin" script uses JMP table joining capability to match patient level data to within cluster LTD estimates using the "C_NCreg " column of the revised input table with the corresponding "C_NCreq" column of the "LC_NCreg.JMP" table. The JMP "Analyze-Fit Model" or "Analyze- Modeling-Partition" menus (for regression models or regression "trees" respectively) are typically used to try to predict the LTD distribution using X-covariates.

[8] After displaying outcome LTD sensitivity over a range of alternative numbers-of-clusters, the user will usually wish to focus on visualizing details of the NN/LTD distribution for some specific number-of-clusters. The "Local Control" script attempts to make the "more" or "most" relevant patient comparisons in the specified X-space for a given requested number-of-clusters. "Artificial LTD Distribution" script allows its user to compare and contrast the relevant NN/LTD distribution with the corresponding distribution from random clusters (ignoring the specified X variables.) This allows the user to literally see how much treatment-selection-bias (imbalance) has been "detected" using the specified X-space clustering metric, clustering algorithm and number-of-clusters. If the NN/LTD and Artificial LTD distributions are not clearly "different," then no meaningful "adjustment" for X-variables has occurred.

[9] To document your work, you could "name" and "save" the JMP journal, say, in Word *.DOC format. Similarly, you may wish to "re-name" and "save" the "data_SA.JMP" table plus one or more of the "LC_NCreq.JMP" tables.

You may also wish to delete any uninteresting or redundant columns within any generated "joinDt.JMP" data tables before you "re-name" and "save" them.

Simple Numerical Example (only 1 X-variable):

Data from : Appleton DR, French JM, Vanderpump p M PJ. "Ignoring a Covariate: An Example of Simpson's Paradox" ***Amer. Statist.*** 1996; 50: 340-341.

JMP dataset: AFV1314.JMP
Y-outcome: fatal (1 => subject failed to survive for 20 years)
Trtm variable: smoke (1 => smoker at the time of the initial survey
X-covariate: age_ decade

Note: Variable "age_decade" takes on only seven distinct values from 20 to 80. This coding essentially limits the number of "meaningful" clusters to only 7 …even though JMP can "force" more clusters than these.

Non-randomized "Lindner Center" Study Example, 2 Y-variables and 7 X-variables:

Data from: Kereiakes DJ, Obenchain RL, Barber BL, et al. "Abciximab provides cost effective survival advantage in high volume interventional practice." ***Am Heart J*** 2000; 140: 603-610.

JMP dataset: Lindner.JMP (996 patients, 10 variables)

Y-outcomes: Column 1 Name = lifeexpt (11.6 years if survive for 6 months; else 0 years)
Column 2 Name = cardbill (cardiac related total expenses within 6 months)

Trtm variable: Column 3 Name = abcix (1 => usual PCI care augmented with planned or emergency use of abciximab; 0 => usual PCI care alone.)

X-covariate: Column 4 Name = stent (1 => yes)
Column 5 Name = height (centimeters)
Column 6 Name = female (1 => yes)
Column 7 Name = diabetic (1 => yes)
Column 8 Name = acutemi (1 => yes)
Column 9 Name = ejecfrac (percentage; higher implies less blockage)
Column 10 Name = ves1proc (number of vessels involved in first PCI)

Simulated datasets designed to be much like but larger than the "Lindner Center" Example:

JMP dataset: LSIM10K.JMP (10,325 patients, 10 variables)
JMP dataset: LSIM5K.JMP (5,162 patients, 10 variables)

NOTE: LSIM5K.JMP m ay either be m erged w ith LSIM10K.JMP or used as a "hold-out" sample of data for validation of predictions made using the LSIM10K.JMP data.

Datasets with built-in "Bubble Plot" Scripts: Dynamically visualize results from LC analyses as of AFV1314.JMP and LSIM10K.JMP as clusters become small, compact and numerous…

JMP dataset: AFVbubble.JMP
JMP dataset: LsimBubl.JMP

General References:

Bang H, Robins JM. Doubly robust estimation in missing data and causal inference models. ***Biometrics*** 2005; 61(4): 962–973. ("Inverse Probability Weighting" is sufficient for unbiased estimation.)

Box GEP. "Robustness is the Strategy of Scientific Model Building." ***Robustness in Statistics.*** R.L. Launer and G.N. Wilkinson, eds. New York: Academic Press, 1979. (See quote on page 202.)

Cochran WG. "The Effectiveness of Adjustment by Subclassification in Removing Bias in Observational Studies." ***Biometrics*** 1968; 24: 205-213.

Hansen BB. "Full matching in an observational study of coaching for the SAT." ***JASA*** 2004; 99: 609-618.

Fraley C, Raftery AE. "Model based clustering, discriminant analysis and density estimation." ***JASA*** 2002; 97: 611-631.

Imbens GW, Angrist JD. "Identification and Estimation of Local Average Treatment Effects." ***Econometrica*** 1994; 62: 467-475.

Kaufman L, Rousseeuw PJ. ***Finding Groups in Data: An Introduction to Cluster Analysis***. New York: John Wiley and Sons, 1990.

McClellan M, McNeil BJ, Newhouse JP. "Does More In tensive Treatm ent of Myocardial Inf arction in the Elderly Reduce Mortality?: Analysis Using Instrumental Variables." ***JAMA*** 1994; 272: 859-866.

Obenchain RL. "Unsupervised Propensity Scoring: NN and IV Plots." ***2004 Proceedings of the ASA***. <<UPSnnivp.PDF>>

Obenchain RL. "TIX Simulations: Treatment Interactions with X-covariates. Lilly White Paper. <<TIXsimul.PDF>>

Obenchain RL. "Unsupervised and Supervised Propensity Scoring in R: the USPS package" September 2006. http://members.iquest.net/~softrx/index.htm. <<USPS_1.01-1.zip or USPS_1.01-1.tar.gz>>

Obenchain RL. "The Local Control Approach using JMP." Chapter 5 of ***Analysis of Observational Healthcare Data Using SAS®.*** Faries D, Leon AC, Maria Haro J, Obenchain RL, eds. Cary, NC: SAS Press. To appear September 2009.

Pocock SJ, Assmann SE, Enos LE, Kasten LE. Subgroup analysis, covariate adjustment and baseline comparisons in clinical trial reporting: current practice and problems. ***Stat Med*** 2002; 21: 2917–2930.

Rosenbaum PR, Rubin RB. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." ***Biometrika*** 1983; 70: 41-55. (See Theorem 2, page 44, on "Balancing Scores.")

Rubin DB. "Bias Reduction using Mahalanobis Metric Matching." ***Biometrics*** 1980; 36: 293-298.

Wainer H, Gessaroli M, Verdi M. Finding what is not there through the unfortunate binning of results: the Mendel effect. ["Visual Revelations" section.] ***Chance*** 2006; 19: 49–52.