

Part II: PCNA and Bootstrap Resampling

1. Partial Correlation Network Analysis

More Statistics tutorial at www.dumblittledoctor.com



PCNA: Generating a Path Diagram

When there is not a hypothesized diagram for a SEM analysis, we can generate a path diagram using partial correlation network analysis.

In 2006, Marrelec discussed the concept of detecting an underlying connectivity network in data, and the methods for analysis. He noted the importance of detection without hypothesized relationships, as SEM requires. In 2007, Marrelec et. al. published a work praising the use of **Partial Correlation Network Analysis (PCNA)** in conjunction with SEM.

Partial correlation analysis is a technique that allows us to investigate the relationship between two variables free of influence from other variables.

Consider two variables, X and Y. We want to know the correlation of X and Y while controlling for Z. The most intuitive way to understand partial correlation is to consider two regressions.

$$X = a_1 + b_1Z + \hat{\epsilon}_1$$

$$Y = a_2 + b_2Z + \hat{\epsilon}_2$$

$$\text{Corr}(X - \hat{a}_1 - \hat{b}_1Z, Y - \hat{a}_2 - \hat{b}_2Z) = \text{PartialCorr}(X, Y | Z)$$

Our PCNA Bootstrap Methodology

We have N variables, and we are interested to know which pairs have significant relationships when controlling for all other variables in the system. Additionally, we are interested to know which pairs' relationships is changed by the disease state of the measured tissue, for example.

For each pair of variables, i and j, we regress the two variables individually on all other variables in the system, and calculate the corresponding residuals. This creates two variables, R_i and R_j , representing the original variables free of the influence of all other variables in the system. Then we can evaluate their correlation.

$$\begin{aligned} Y_i &= \sum_{k \neq i, j} a_k Y_k + \hat{\epsilon}_i \\ Y_j &= \sum_{k \neq i, j} a_k Y_k + \hat{\epsilon}_j \end{aligned} \quad \Rightarrow \quad \begin{aligned} R_i &= Y_i - \hat{Y}_i \\ R_j &= Y_j - \hat{Y}_j \end{aligned}$$

$corr(R_i, R_j)$ is the partial correlation of the variables. However, this is just one number, so we cannot incorporate the influence of covariates into the significance test of this value. This is why we use a bootstrapping procedure.

More Statistics tutorial at www.dumblittledoctor.com

Part II: PCNA and Bootstrap Resampling

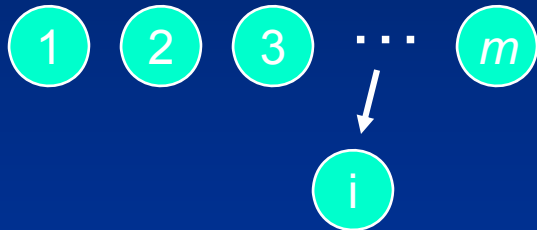
2. Bootstrap Resampling



Bootstrap Resampling

The idea behind bootstrapping is *resampling with replacement*.

We have our original sample of m subjects.



Select one of them at random, and then replace it before randomly selecting the next. Repeat this m times.

Now we have a sample of m subjects consisting of subjects from the original sample. However, some subjects may be repeated, and some subjects from the original sample may not be present in our resample.

Use each resample to calculate the partial correlation. Now we have a population of n measurements for each pair of variables. If we perform this analysis on our two datasets individually, we will have 1000 estimates of partial correlation for the normal tissue and 1000 estimates for the diseased tissue.

Bootstrap Resampling

We will let the significance of the relationships in the normal dataset represent the general significance of partial correlation among variables in the system.

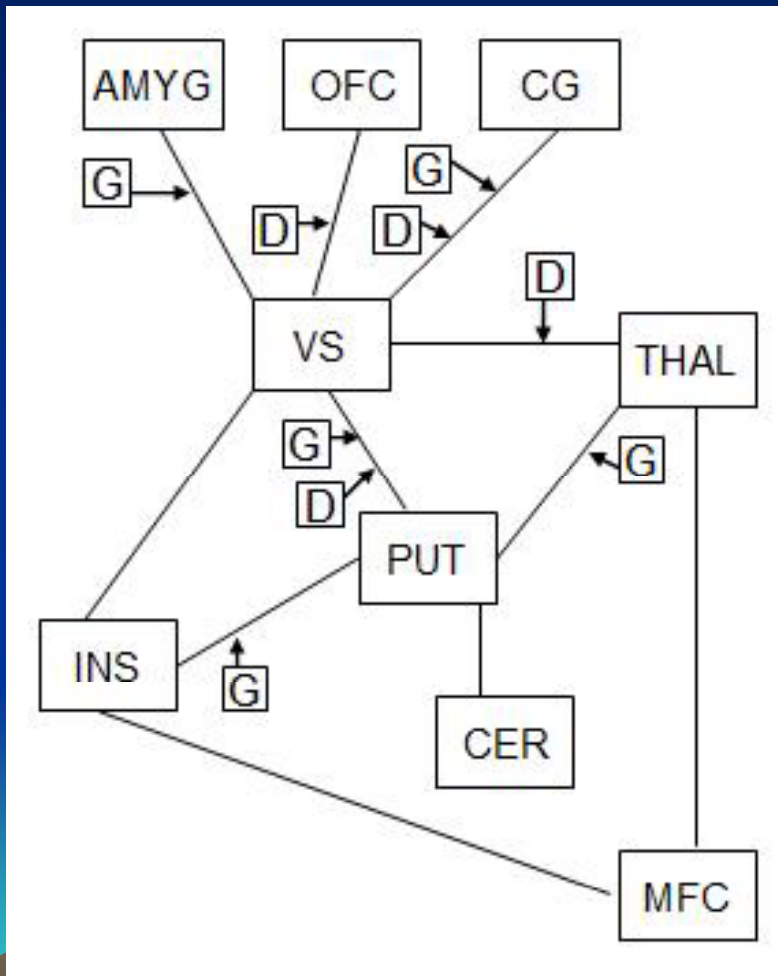
We can create a difference variable to estimate the difference of the partial correlation between the normal tissue and diseased tissue. The significance of the differences represents the influence of disease on the partial correlation between variables.

The results we must evaluate are two lists of partial correlations (those for the normal tissue, and those for the diseased tissue).

Normal	Diseased	Difference
V1	W1	$V1 - W1$
V2	W2	$V2 - W2$
V3	W3	$V3 - W3$
...

Sort the normal and difference variables. If 0 is contained in the middle 95% of the observations, then we would say the relationship or influence of disease is insignificant for this pair of variables. (This is called the percentile method).

Results



The results of the PCNA bootstrap in the brain data (four datasets; covariates: drug, group) example is shown at the left. **No arrows!**

At this point, we would ask the collaborating researcher for input on the directionality of each path. For paths not easily determined, we can implement one path in each direction.

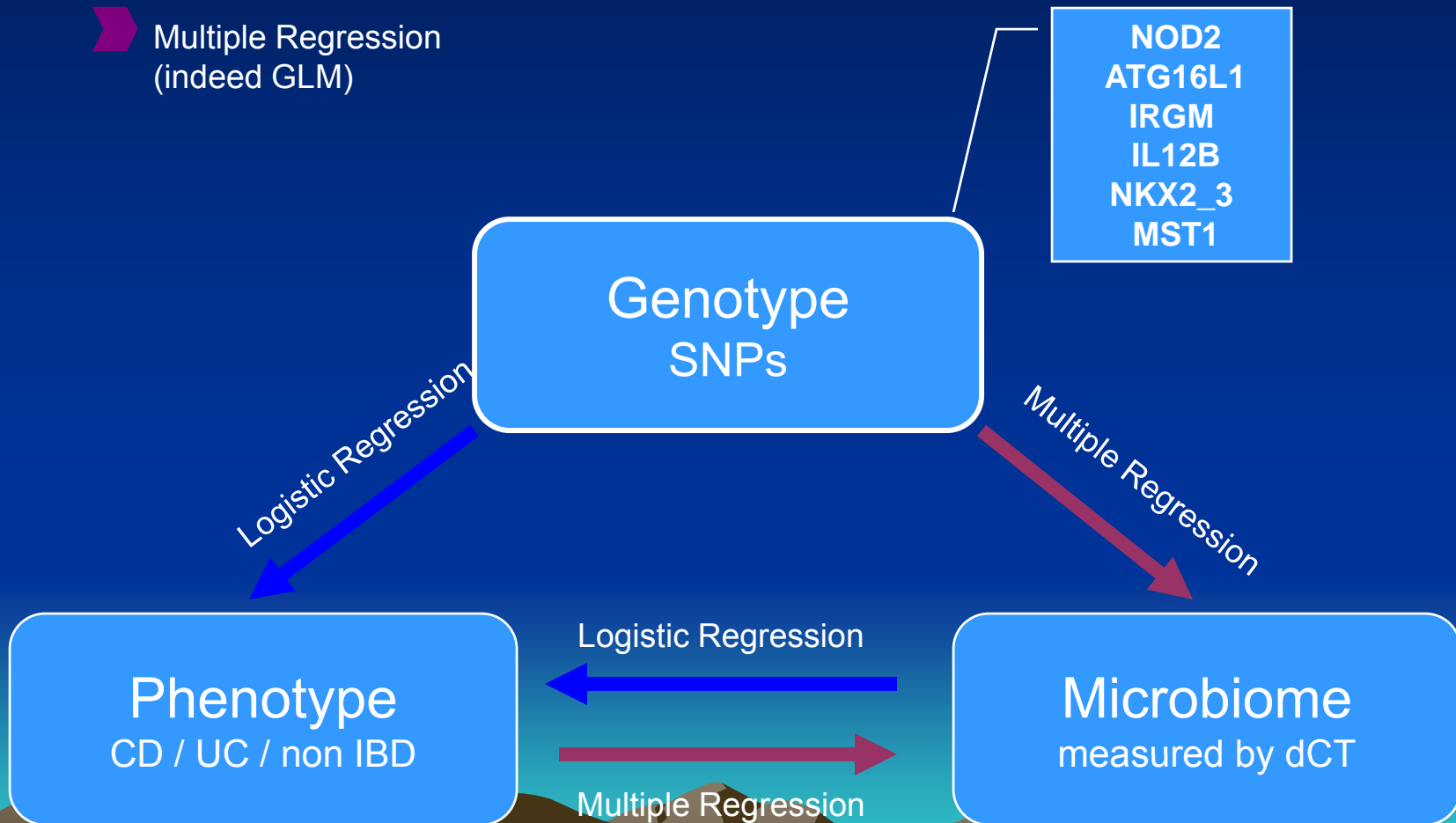
The results would be a hypothesized relationship that can be verified using structural equation modeling with an independent data set.

Current Work

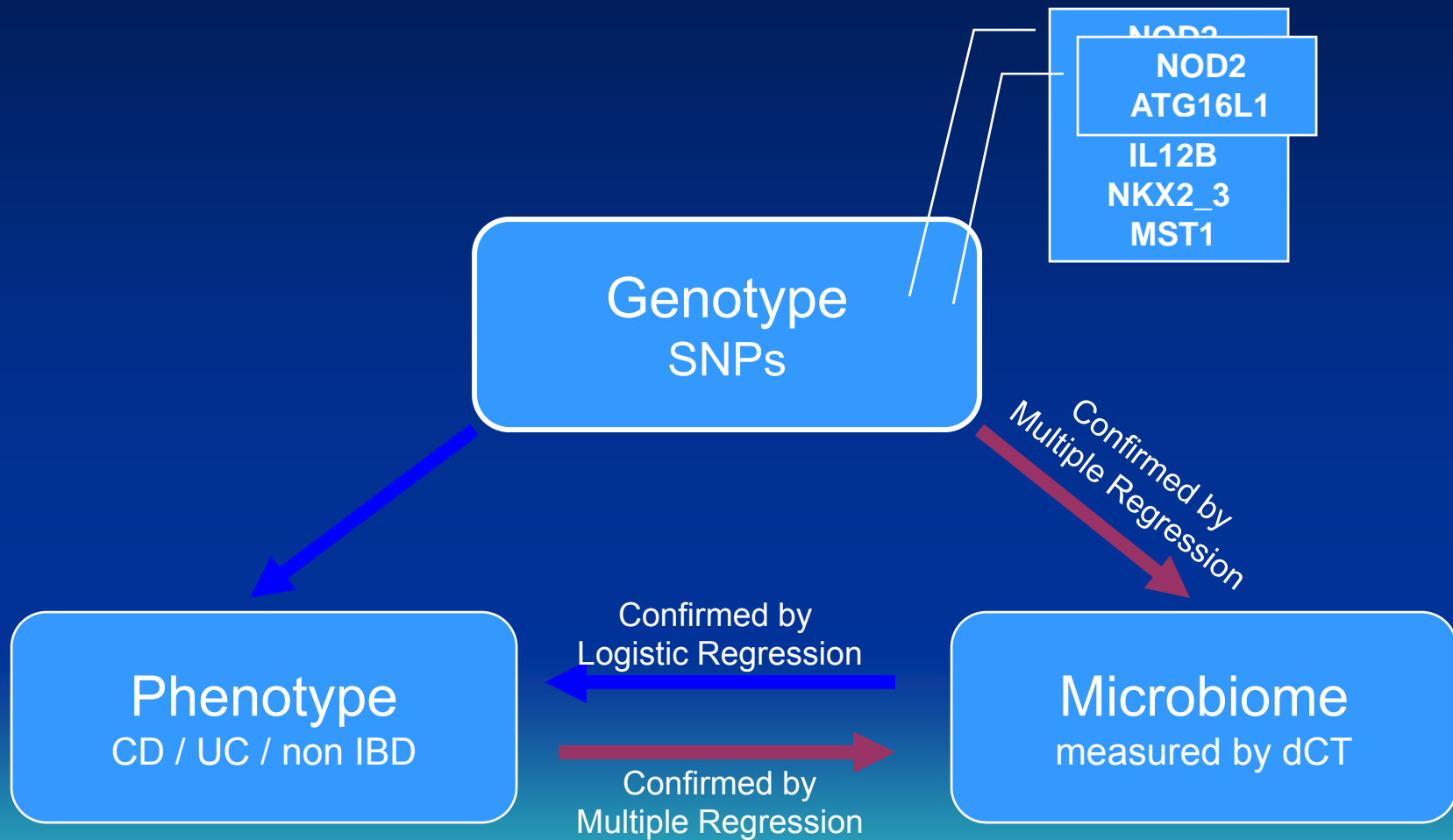
SEM with both continuous and categorical variables (shown next.)

Pathway Assumptions and Statistical Models

- Logistic Regression
- Multiple Regression (indeed GLM)



Confirmed Pathway Model



Selected References

- [1] Bollen, K.A., *Structural Equations with Latent Variables*. 1989: John Wiley and Sons, Inc.
- [2] Blalock, H.M., *Causal Inferences in Nonexperimental Research*. 1964: University of North Carolina Press.
- [3] Duncan, O.D., *Path analysis: Sociological examples*. American Journal of Sociology, 1966. **72**: p. 1-16.
- [4] Wolfle, L.M., *The Introduction of Path Analysis to the Social Sciences, and Some Emergent Themes: An Annotated Bibliography*. Structural Equation Modeling, 2003. **10**(1): p. 1-34.
- [5] Joreskog, K. and D. Sorbom, *LISREL 8: User's Reference Guide*. 1996: Scientific Software International, Inc.
- [6] Jaccard, J. and C.K. Wan, *LISREL Approaches to Interaction Effects in Multiple Regression*. Quantitative Applications in the Social Sciences. 1996: Sage Publications.
- [7] Byrne, B.M., *Structural Equation Modeling with EQS: Basic Concepts, Applications, and Programming*. Second ed. 2006: Lawrence Erlbaum Associates.
- [8] Coenders, G., J.M. Batista-Foguet, and W.E. Saris, *Simple, Efficient and Distribution-free Approach to Interaction Effects in Complex Structural Equation Models*. Quality & Quantity, 2008. **42**(3): p. 369-396.
- [9] Cronbach, L.J. and N. Webb, *Between class and within class effects in a reported aptitude x treatment interaction: a reanalysis of a study by G. L. Anderson*. Journal of Educational Psychology, 1979. **67**: p. 717-724.
- [10] Muthen, B., *Means and covariance structure analysis of hierarchical data*. UCLA Statistics series. Vol. #62. 1990.
- [11] Marcoulides and Schumacker, eds. *Advanced Structural Equation Modeling*. 1996.
- [12] Scientific Software International. *Multilevel structural equation modeling*. 2005-2008 [cited 2008; LISREL implementation of multilevel SEM]. Available from: <http://www.ssicentral.com/lisrel/techdocs/Session12.pdf>.
- [13] Garson GD. *Statnotes: Partial Least Squares*. 1998, 2008, 2009 [cited 2009; Introduction to Partial Least Squares SEM and Regression]. Available from: <http://faculty.chass.ncsu.edu/garson/PA765/pls.htm>.
- [14] Hoyle RH. *Statistical Strategies for Small Sample Research*. 1999: Sage Publications.
- [15] Joreskog KG and Wold H, *The ML and PLS Techniques for Modeling with Latent Variables*, in *Systems under indirect observation: causality, structure, prediction*, K.G. Joreskog and H. Wold, Editors. 1982.
- [16] Tenenhaus M, Vinzi VE, Chatelin Y-M, Lauro C. *PLS path modeling*. Computational Statistics & Data Analysis, 2005. **48**: p. 159-205.
- [17] Wold H, ed. *The Fix-Point Approach to Interdependent Systems*. 1981, North-Holland Publishing Company.
- [18] Wold H. *Estimation of Principal Components and Related Models by Iterative Least Squares*. Multivariate analysis; proceedings of an international symposium held in Dayton, Ohio, June 14-19, 1965, 1966: p. 391-420.
- [19] Wold H, *PLS for multivariate linear modeling*, in *QSAR: Chemometric methods in molecular design: Methods and principles in medicinal chemistry*, H.v.d. Waterbeemd, Editor. 1994.
- [20] Wold H, *Partial Least Squares*, in *Encyclopedia of Statistical Sciences*. 2006, John Wiley & Sons, Inc.

- [21] Volkow, N.D., et al., *Expectation Enhances the Regional Brain Metabolic and the Reinforcing Effects of Stimulants in Cocaine Abusers*. The Journal of Neuroscience, 2003. **23**(36): p. 11461-11468.
- [22] Crowder, M.J. and D.J. Hand, *Analysis of Repeated Measures*. 1990: Chapman & Hall.
- [23] Cudeck, R., S. du Toit, and D. Sorbom, eds. *Structural Equation Modeling: Present and Future, A Festschrift in honor of Karl Joreskog*. 2001, Scientific Software International.
- [24] Blackwell, E., C.F.M.D. Leon, and G.E. Miller, *Applying Mixed Regression Models to the Analysis of Repeated-Measures Data in Psychosomatic Medicine*. Psychosomatic Medicine, 2006. **68**: p. 870-878.
- [25] Garson, G.D. *Statnotes: Topics in Multivariate Analysis, Structural Equation Modeling*. 1998, 2008 [cited 2008; Introduction to SEM]. Available from: <http://faculty.chass.ncsu.edu/garson/PA765/partialr.htm>.
- [26] Krzanowski, W.J. and F.H.C. Marriott, *Multivariate Analysis Part 2: Classification, covariance structures and repeated measurements*. 1995, New York: Halsted Press.
- [27] Rovine, M.J. and P.C.M. Molenaar, *A LISREL Model for the Analysis of Repeated Measures With a Patterned Covariance Matrix*. Structural Equation Modeling, 1998. **5**(4): p. 318-343.
- [28] Vittinghoff, E., et al., *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*. 2005: Springer.
- [29] Diggle, P.J., et al., *Analysis of Longitudinal Data*. Second ed. Oxford Statistical Science Series. 2002.
- [30] Monahan, J.F., *Numerical Methods of Statistics*. 2001: Cambridge University Press.
- [31] Gill, P.E., W. Murray, and M.H. Wright, *Practical Optimization*. 1981, London: Academic Press, Inc.
- [32] Stanghellini, E. and N. Wermuth, *On the identification of path analysis models with one hidden variable*. Biometrika, 2005. **92**(2): p. 337-350.
- [33] Dunn, G., B. Everitt, and A. Pickles, *Modelling Covariances and Latent Variables Using EQS*. 1993: Chapman & Hall.
- [34] Skrondal, A. and S. Rabe-Hesketh, *Generalized Latent Variable Modeling: Multilevel, Longitudinal, and Structural Equation Models*. 2004: Chapman & Hall/CRC.
- [35] Marrelec, G., et al., *Using partial correlation to enhance structural equation modeling of functional MRI data*. Magnetic Resonance Imaging, 2007. **25**: p. 1181-1189.
- [36] Garson, G.D. *Statnotes: Partial Correlation*. 1996, 2008 [cited 2008; Introduction to Partial Correlation]. Available from: <http://faculty.chass.ncsu.edu/garson/PA765/partialr.htm>.
- [37] Rees, D.G., *Foundations of Statistics*. 1987: CRC Press.
- [38] Lunneborg, C.E. and J.P. Tousignant, *Efron's Bootstrap with Application to the Repeated Measures Design*. Multivariate Behavioral Research, 1985. **20**: p. 161-178.
- [39] Berkovits, I., G.R. Hancock, and J. Nevitt, *Bootstrap Resampling Approaches for Repeated Measure Designs: Relative Robustness to Sphericity and Normality Violations*. Educational and Psychological Measurement, 2000. **60**: p. 877-892.
- [40] Seco, G.V., et al., *A Comparison of the Bootstrap-F, Improved General Approximation, and Brown-Forsythe Multivariate Approaches in a Mixed Repeated Measures Design*. Educational and Psychological Measurement, 2006. **66**(35-62).
- [41] Fisher, R.A., *The distribution of the partial correlation coefficient*. Metron, 1924. **3-4**: p. 329-332.
- [42] Marrelec, G., P. Bellec, and H. Benali, *Exploring Large-Scale Brain Networks in Functional MRI*. 2006.