Stats Happening – Spring 2011

1. Be careful when log-transforming count data

2. Odds Ratio versus Relative Risk

3. Multiple Imputation in multilevel settings with WinMICE

4. Longitudinal MANOVA in SAS

5. Assessing confounding effects in Stata

6. Spatial Analysis using SAM (Spatial Analysis in Macroecology)

7. Revolution-R software

8. Editing plots in SAS

_____

1. Be careful when log-transforming count data

Count data are common to many disciplines, and are often log-transformed to satisfy parametric assumptions in statistical analysis. If a single observation has a count of zero, the entire data set needs to be adjusted (e.g. adding the value 1 to every observation) to adjust for having an undefined value for the log of zero. In a recent article published in *Methods in Ecology & Evolution*, O'Hara and Kotze recommend that count data should be analyzed by models based on Poisson and Negative Binomial distributions rather than log-transformed data. Their study was based on simulated data that demonstrated these methods performed better than log-transformed data, except when the dispersion was small and the mean counts were large.
O'Hara, R.B. & Kotze, D.J. (2010) Do not log-transform count data. *Methods in Ecology and Evolution*, 1, 118-122.

2. Odds Ratio versus Relative Risk

Both the odds ratio (OR) and the relative risk (RR) are used extensively in studies involving logistic regression. The odds are defined by the likelihood of an event occurring divided by the likelihood of that event not occurring, or $P(A)/(1-P(A))$. Therefore, the OR is defined as the ratio

of the odds of an event in one group versus the odds of an event in another group, or (P(A)/(1-P(A))/(P(B)/(1-P(B)). The relative risk (RR) is simply the ratio of the probability of an event in one group versus the probability of an event in another group, or P(A)/P(B). Often the OR will be misinterpreted as the RR. In cases where the rate of event occurrence is rare, the OR will be a close approximation of the RR. However, when the event occurrence is more common (>10%), there will be a greater disparity between the two values. Logistic regression is used to estimate adjusted OR's when controlling for other variables. Similarly, Log-binomial models can be used to estimate RR's in the presence of covariates. Both can easily be modeled in any software package using a generalized linear model procedure with a binomial distribution specification and either a logit link function to obtain OR's or a log link function to obtain RR's.

For further information see: http://www.ats.ucla.edu/stat/sas/faq/relative_risk.htm

3. Multiple Imputation in multilevel settings with WinMICE

Missing data problems can arise with multilevel data. Unfortunately imputation methods designed for single-level data are often not appropriate in these situations. WinMICE is a free, easy-to-use standalone program in Windows OS, that implements imputation on the linear mixed model. The WinMICE software especially addresses the challenge of missing data in a multilevel setting, when data are clustered within classes. A nested Gibbs sampler is used to estimate the parameters of the multilevel model for individual variables. In the current version, only linear models and a maximum of two levels can be handled.

For further information see: http://web.inter.nl.net/users/S.van.Buuren/mi/hmtl/mice.htm

4. Longitudinal MANOVA in SAS

There is a growing interest in the analysis of data where subjects are measured on several occasions with regards to a collection of response variables. Often researchers are interested in studying the joint evolution of these response variables over time. There are three alternatives methods to modeling multivariate longitudinal data: the multivariate repeated measurement model with a Kronecker product covariance structure (PROC MIXED), the random coefficients mixed model (PROC MIXED) and the structural equation modeling approach (PROC CALIS). These three different approaches can easily be modeled in SAS using the Mixed or Calis procedures.

For more information see: http://www2.sas.com/proceedings/sugi31/187-31.pdf

5. Assessing confounding effects in Stata

There are two practical tools implemented in Stata to help researchers better understand the presence and direction of possible confounding effects in their data. These are two postestimation commands that can be used after most common estimation commands such as linear regression, logistic regression, conditional logistic regression, Poisson regression and Cox proportional hazards models. One is the all-possible-estimates method (*confall*) and the other is the change-in-estimates method (*chest*). The all-possible-estimates method allows the user to

inspect possible estimates with many different adjustments. The change-in-estimates method selects variables in a stepwise fashion.

Wang Z, (2007), Two postestimation commands for assessing confounding effects in epidemiological studies. *The Stata Journal,* 7, 2, 183-196.

6. Spatial Analysis using SAM (Spatial Analysis in Macroecology)

Regardless of your field of study (despite what the name seems to indicate), if you use spatial statistical analyses methods in your research you might benefit from the capabilities offered by SAM. SAM is a powerful, user-friendly, free Microsoft Windows program containing not only a GIS mapping module but also an extensive package of tools for spatial statistical analysis.

Detailed information can be found in: Rangel, T.F.L.V.B, Diniz-Filho, J.A.F and Bini, L.M. (2006) Towards an Integrated Computational Tool for Spatial Analysis in Marcoecology and Biogeography. Global Ecology and Biogeography, 15:321-327.

To download SAM and for additional information see: http://www.ecoevol.ufg.br/sam/

7. Revolution-R software

An enterprise version of the R statistical package from Revolution Analytics is available for free to users affiliated with academic institutions. The software runs standard R code in a more efficient and user-friendly environment. Advantages over standard R include automatic updates to the latest stable version of R, an enhanced Graphical User Interface, increased capacity and speed for analyzing large data sets (terabyte-class data), and a more efficient use of multi-core processors and multiprocessor workstations.

More information is available at http://www.revolutionanalytics.com/.

8. Editing plots in SAS

Performing an analysis in SAS and need publication quality plots? There is now an easier way in SAS to customize your graphs. Creation of editable graphs in SAS can be enabled either temporarily (during a given SAS session), or permanently. Graphs created in SAS while this feature is enabled can later be edited using SAS GRAPH ODS graphics editor. To enable this feature temporarily, first select the Results window, and then enter *sgedit on* in the command line above the Results window. To enable this feature permanently, open the Registry Editor by entering *regedit* in the command line. Select *SAS_REGISTRY > ODS > GUI > RESULTS*. Double-click on *Statistical Graphics Editor*, and change *Value Data* to *On*.