

**Y650 Topical Seminar in Inquiry Methodology
Data Analysis and Programming Using SAS
(Section no. 29479, Fall 2011)
Fridays 9:30 - 12:15 p.m.
Education 2015
Joanne Peng (856-8337, peng@indiana.edu)**

Course Description

1. This seminar will expose students to the versatility and power of SAS® for data analysis and file management. Specifically, students will learn and master techniques such as, verifying and recoding data, repetitive and conditional data processing, merging and updating data files. In addition, methods for verifying statistical assumptions (normality and randomness), for treating missing data, for making inferences based on categorical data (contingency table analysis and logistic regression modeling) will be introduced using SAS commands.
2. The seminar format will consist of lectures, discussions, computer exercises, and student presentations. Students are expected to perform in-depth analyses of an empirical data set to address one or more research questions in an applied field and write up the results and interpretations in journal manuscript styles. The paper should demonstrate student's ability to analyze data, knowledge about statistical methods and their interpretations, and proficiency in professional writing.
3. The primary prerequisite to this course is Y502 or an equivalent that covered descriptive and inferential statistics including the ordinary least squares regression models. An aptitude for statistical reasoning is beneficial. Deficiencies in particular areas will be remedied in tutorial sessions.
4. The primary texts are SAS Online Documents for Version 9.2 and Peng (2009), *Data Analysis using SAS®* by Sage. References include, but are not limited to the following--
 - a) Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics* (5th ed.). Needham Heights, MA: Allyn & Bacon. Website at http://wps.ablongman.com/ab_tabachnick_multistats_5/.
 - b) O'Connell, A. A., Goldstein, J., Peng, C.-Y. J. & Rogers, H. J. (2007). Multilevel logistic models for dichotomous and ordinal data. In A. A. O'Connell and B. McCoach (Ed.), *Multilevel analysis of educational data*. Greenwich, CT: Information Age Publishing Inc.
 - c) Peng, C.-Y. J., Harwell, M., Liou, S.-M., & Ehman, L. H. (2006). Advances in missing data methods and implications for educational research. In S. Sawilowsky (Ed.), *Real data analysis* (pp. 31-78). Greenwich, CT: Information Age Publishing, Inc.
 - d) Peng, C.-Y. J., & Zhu, J. (2007). Comparison of two approaches for handling missing covariates in logistic regression. *Educational and Psychological Measurement*.
 - e) Peng, C.-Y. J., So, T.-S. H., Stage, F. K., & St. John, E. P. (2002). The use and interpretation of logistic regression in higher education journals: 1988-1999, *Research in Higher Education*, 43(3), 259-293.
 - f) Peng, C.-Y. J., & Nichols, R. (2003). Using multinomial logistic model to predict adolescents' behavioral risk. *Journal of Modern Applied Statistical Methods*, 2(1), 177-188.
 - g) SAS Institute's *Logistic Regression: Examples using the SAS system* (SAS Institute, 1995).
 - h) Hosmer and Lemeshow's *Applied Logistic Regression* (John Wiley and Sons, 2001).
 - i) Kleinbaum's *Logistic Regression* (Springer, 2001).
 - j) Menard's *Applied Logistic Regression Analysis* (Sage Publication #106, 1992).

Schedule

<u>Week</u>	<u>Topic</u>	<u>Text and/or Ref</u>
1	Introduction to this course and SAS software	Ch. 1-3 from Peng's book
2	SAS data sets	Ch. 4 and Appendices A & B from Peng's book
3	Verifying data Debugging SAS programs	Ch. 5 Ch. 6 from Peng's book
4	Recoding data via SAS functions Selecting variables or observations from a SAS data set	Ch. 7 Ch. 18 from Peng's book
5	Repetitive and conditional data processing	Ch. 19
6	Structuring SAS data sets	Ch. 20 from Peng's book
7	Descriptive data analysis	Ch. 8
8	Effect size, confidence interval estimation, and test of normality	Ch. 9 from Peng's book
9-10	Missing data methods	SAS Online Documents V9.2 References (c) and (d)
11-12	Categorical data analysis	Ch. 13-14 from Peng's book
13	[Thanksgiving recess, no class]	
14	Fitting simple and multiple logistic regression models Model building strategies and methods	Ch. 17 from Peng's book References (e), (g), (h), (i), (j)
15	SAS programming with matrix language MACRO and simulations	SAS Online Documents V9.2 SAS Online Documents V9.2
16 (final's)	Student presentations	

Note: This class schedule is subject to change without prior notice.