<http://www.tamuc.edu/academics/graduateSchool/documents/thesis--dissertation-services/>

Script for Summary of Propensity Score Matching in Education

Slide 1- Introduction. Hi, my name is Jerry Perez and today I’ll be talking about Propensity Score Matching in Education. This summary is based on an article by Lane F. C., To Y. M., Shelley K., and Henson R. K. The article is titled: “An Illustrative Example of Propensity Score Matching with Education Research.

Slide 2 - Introduction to Propensity Score Matching. Let’s get started with an introduction to Propensity score matching.

* Propensity Score Matching also abbreviated as PSM is a quasi-experimental technique endorsed by the U.S. Department of Education to control for covariates such as self-selection bias and non-random assignment.
* PSM is a statistical matching technique designed to estimate the effect of a treatment, policy, or other intervention by accounting for the covariates that are also related to the treatment.
* PSM can control for non-group differences and then apply the results to the matched groups.
* A covariant that PSM can control for is self-selection. Self-selection is also known as the volunteer effect and is when a participant is motivated to take part in the research study.

Slide 3 - Self-selection and random assignment. Self selection can have a big impact on a research study, so it is a good idea to have a clear understanding of what it is. Self-selection is also known as the volunteer effect.

* An example of self-selection is when a student volunteers to answer a question in class.
* Typically, a student will raise her hand to answer a question when she is confident she knows the answer.
* In research, typically participants will self-select to be in a study for personal reasons. Time away from work, extra credit on grades or recognition are some personal reasons for self-selection.
* Random assignment is when all participants have an equal chance of being selected to a test or non-test group.
* Self-selection makes random assignment impossible and can produce non-equivalent groups.

Slide 4 - Self-selection, non-random assignment, and experimenter bias

* Self-selection is one confounding variable; others include: non-random assignment, experimenter bias, small sample size, etc. which can produce non-equivalent groups.
* Non-random assignment is when participants with certain characteristics are selected. For example, students in a teacher’s class are selected because they are convenient. Medical research often uses non-random assignment in order to study a specific condition or treatment.
* Experimenter bias is when the researcher knowingly or unknowingly influences the participants’ responses. For example, a researcher may state, “most people found the book enjoyable. Did you like it?”
* PSM minimizes self-selection, non-random assignment, and experimenter bias by matching participants through covariates.

Slide 5 - How PSM and SPSS Work Together

* SPSS has built in PSM formulas that allow for calculation and comparison of covariates.
* SPSS requires all relevant covariates be included in the study - gender, age, and socio-economic status.

* PSM requires large sample sizes because many participants may be excluded due to poor matching.
* PSM is useful in educational, medical and economic research using non-random samples.

Slide 6 - Using SPSS to Generate PSM Data

* The following slides provide step-by-step instructions on how to use SPSS to generate PSM data.
* SPSS version 22 was used in the slides, but the steps also apply to SPSS 20 and newer.

Slide 7 – To begin, click on Start button on your computer, All Programs and your version of SPSS. This demonstration is using SPSS 22, but is also applicable for versions 20 and 21.

Slide 8 – Once you open SPSS, click on File, Open and Data. The article does provide a sample data set that can be used for practice purposes. This illustration does use part of the sample data set.

Slide 9 – The data set we will be using is titled Propensity and is located on the local hard drive. Once you’ve located the data set, select it and click on open.

Slide 10 – Here is a screenshot of what the Propensity data set looks like. I would like to point out the first column titled ID. This is the ID assigned to each of the participants. The second column is the current group assignment for each participant. The other columns have other variables included in the data set.

Slide 11 – Once we have the data set open, go to the ribbon and select Analyze. Then select Regression and Binary Logistic. These steps can be taken in Data View, which is what we are currently seeing, or Variable View.

Slide 12 – After selecting Logistic Regression, this window opens. We can select the Dependent variable as well as the covariates. The Dependent variable selected is Hospital. Meaning, we want to know if the participants went to the hospital based on the covariates selected. The covariates selected are: income, race, smoker and age. After making our selections, we click on Save on the top right.

Slide 13 – In the Logistics Regression Save window we will check Probabilities as well as Include the covariance matrix. After making those selections, click on continue.

Slide 14 – After running our analysis, we now have a new variable named PRE\_1. This is Predicted probabilities. This is visible in the Variable View.

Slide 15 – Here the Predicted probabilities column has been moved to the left close to the ID and the Group columns. Remember, I pointed out the ID and Group columns? You can see the IDs are now ordered by Predicted probabilities. We can also now match the Predicted probabilities IDs into different groups. This type of ordering would be helpful in a medical study where we are looking for participants that have similar characteristics. Selecting participants and placing them into different treatment groups is a manual process

Slide 16 – Now that we have a variable named Pre\_1 that includes certain variables, we can return to the ribbon and select Analyze. Then select Regression and Binary Logistic. This time we will select Treatment group as the dependent variable because we are placing the participants in a treatment group. For covariates we select Pre\_1 and diabetic and click ok.

Slide 17 – Here are some of the results I want to bring to your attention. On the Classification Table, notice that 59 out of the 100 participants in our study were included after we matched them for suitability. PSM excluded 41 participants from the study.

The Variables in the Equitation Table shows a Sig. of .074 for the Pre\_1 participants and .612 for the diabetic participants.

Slide 18 – In Summary, we covered:

* Introduction to Propensity Score Matching
* Self-selection
* Self-selection, non-random assignment, and experimenter bias
* How PSM and SPSS Work Together
* Using SPSS for Propensity Score Matching
* Screenshots of steps to take in SPSS to get a propensity score
* Article is available for review at:
* Lane F. C., To Y. M., Shelley K., Henson R. K. (2012) An Illustrative Example of Propensity Score Matching with Education Research *Career and Technical Education Research, 37(3), pp. 187-212*
* <http://eds.b.ebscohost.com/ehost/pdfviewer/pdfviewer?sid=27b00001-3877-4bf1-97b9-35044384f2d8%40sessionmgr110&vid=0&hid=112>