Season of Birth & Later Outcomes: Old Questions, New Answers, Bigger Data

Stephen M. Salisbury, Computational Social Science REU Student Researcher
Dr. Kasey Buckles, Brian and Jeannelle Brady Associate Professor of Economics
Summer 2016, University of Notre Dame, South Bend, IN USA

Previous Study Provides Framework
In July 2013, Dr. Kasey Buckles and Dr. Daniel Hungerman published a study in The Review of Economics and Statistics using data from live birth certificates and the U.S. Census. In this study, they discovered that “women giving birth in the winter look different from other women: they are younger, less educated, and less likely to be married” (Buckles & Hungerman, 2013, p. 711). While there was a significant amount of previous research showing the link between season of birth and later outcomes, this study showed that the characteristics of those giving birth varied by season as well. The challenge now is to see if similar results are found when the same variables are considered using a data set that includes many different countries from around the globe.

International Data Set Provides New Data for Comparison
My job this summer has been to help Dr. Buckles assess an international data set to determine if it had the needed information to expand the domestic study that she and Dr. Hungerman did into one that looks at similar factors but from a much broader, global scale. In addition, I looked for other possible sources of demographic data that could be merged with the initial data sets to provide the tools for more comprehensive analysis.

TIMSS & PIRLS Lead the Way
Located at Boston College’s Lynch School of Education, IEA’s TIMSS & PIRLS International Study Center conducts regular international comparative assessments of student achievement in mathematics and science (TIMSS) and in reading (PIRLS) in more than 60 countries. TIMSS (the Trends in International Mathematics and Science Study) and PIRLS (the Progress in International Reading Literacy Study) together comprise the core cycle of studies for IEA – the International Association for the Evaluation of Educational Achievement.

Looking for New Answers to Old Questions Using Bigger Data
- **Old Question**: Is there a connection between season of birth and later outcomes?
- **New Answer**: The correlations between season of birth and family demographics are significant and can significantly influence later outcomes. The new working hypothesis is that the relationship between season of birth and family background varies systematically across countries.
- **Bigger Data**: A new data set was needed that provided information from multiple countries and regions but that had variables similar to the ones used in the domestic research.

Three Steps to Building a Bigger Data Set
1. **Find the Data**:
   - We needed an international data set that had the three most important variables: Parental SES, Month of Birth, and Academic Outcomes. The TIMMS & PIRLS surveys appeared to be the answer.

2. **Compile the Data**:
   - Once we confirmed that these were the files we wanted to use, we had to compile the data in a way that would work for us. This involved building 2 data sets nearly 3.0 GB in size. I used SPSS, Stata, and Excel to make this happen. We now have cross sectional data from 81 countries using 3 different surveys dating back 20 years.

3. **Add more Data**:
   - Once we had compiled the TIMMS & PIRLS data sets into their aggregate form, Dr. Buckles asked me to find even more data to use for consideration in their analysis. I was able to access the CIA Factbook and the National Climatic Data Center to begin compiling geographic, economic, demographic, and climatic data for each of the 81 countries in our aggregate data set.

Next Steps in the Research Process
Clean the Data: This will involve recoding the variables to fit the needs of the new study.
Analyze the Data: Once all of the variables have been coded and reorganized in a format that meets the needs of the researchers, correlation analysis can begin.

Conclusion
The process of data analysis is multi-layered. As a Social Scientist, we often don’t have the resources to compile our own data, so being able to effectively utilize data that has been gathered by other researchers is a necessity in today’s academic environment. Following these steps and learning how to program data analytic software for your purposes, exhibits the ability to properly evaluate data and data sources, and being able to present data in a clear understandable way are tools that, if cultivated, can lead to a long and prosperous career in academic research.

References

Acknowledgements
Thank you to the National Science Foundation, the University of Notre Dame and its Center for Research Computing, Dr. Paul Brenner, Dr. Kasey Buckles, Kallie O’Connell, Caitllyn Smith, and Julie Vecchio of the Center for Digital Scholarship at the University of Notre Dame.