PROMOTING STATISTICAL LITERACY AND UNDERSTANDING OF YOUTH POPULATION DYNAMICS IN A NEW STATISTICS AND PROBABILITY COURSE FOR SENIOR HIGH SCHOOL

Enriqueta Reston, Jennefer Piramide and Carmelita Loquias
University of San Carlos, Philippines
edreston93@gmail.com

In response to the challenge of implementing a new Statistics and Probability course for senior high school in the Philippines, we developed curriculum resource materials in a course disk anchored on inquiry-based constructivist pedagogy using real data about the youth sector. Offical youth statistics and data from various surveys about the youth were used to provide the database for teaching-learning and assessment activities that build students' conceptual understanding of random variables, probability and sampling distributions, the Central Limit Theorem, and the inferential methods of parameter estimation and hypothesis testing. With the twin goal of promoting statistical literacy and understanding the dynamics of the youth population, the activities provide the learners the opportunity to raise their social consciousness on the vital role of the youth sector in society. This paper describes some of these activities and their databases.

INTRODUCTION

The past few years have witnessed tremendous changes in the way statistics is taught and recognized as a science that cuts across disciplines. Recent curriculum reforms in teaching statistics reflect the increasing interest in making statistics more meaningful to students by using real data that show their importance in various social contexts and in everyday life. Review of related literature in statistics education curriculum development reveal that over the past two decades, a number of curriculum development projects have been presented to address the need for reform in teaching statistics with the overarching goal of promoting and developing statistical literacy among different groups of students. The cases of South Africa in 2005 and New Zealand in 2008 are among those cited as examples of innovations in national statistics curricula that place statistical literacy at the core of the statistics component of school mathematics curriculum (Ridgway, Nicholson, & Mc Cusker, 2011). The curriculum reform in these two countries was characterized with sustained collaboration among academics, statistical offices and professional associations in the development of curriculum resource materials, technology integration and sustained teacher development (North and Scheiber, 2008; Forbes, 2014). Verschut and Bakker (2010) also described a new statistics curriculum for senior general secondary education in the Netherlands that focused on learning statistics through a data-based and problem-oriented approach. In this new curriculum, teachers are encouraged to let students work with real data sets and information technology. In order to achieve coherence in the intended curriculum, they recommend to: (1) build around central themes; (2) emphasize the purpose, and (3) make relationships between chance and data explicit.

Aside from the development of statistical literacy as the overarching goal in statistics education (Gal and Garfield, 1997, Rumsey, 2002), the goals of statistics education have also evolved with increasing emphasis on empowering citizens from all walks of life to function effectively in an information-laden society (European Commission, 1996; cited in Murray & Gal, 2002). It is recognized that statistical information produced by an increasing number of public agencies and private organizations play a vital role in enabling citizens to develop awareness and capability to react intelligently to various social, political, economic phenomena. Hence, the relevance of the social context is an important consideration in choosing real data for teaching statistical concepts and processes.

In the Philippines, the implementation of the *K to 12* Basic Education reform in 2012 expanded the basic education cycle with an additional two years of senior high school. Along with this reform is the addition of a new *Statistics and Probability* course as one of the core courses for Grades 11 and 12. It is the students' first stand-alone statistics and probability course apart from the statistics and probability topics which were integrated in the school mathematics curriculum from Grades 1 through 10 (Department of Education, 2013). This educational reform provides

In: J. Engel (Ed.), *Promoting understanding of statistics about society.*
Proceedings of the Roundtable Conference of the International Association of Statistics Education (IASE), July 2016, Berlin, Germany. iase-web.org/Conference_Proceedings.php

ample opportunities for designing a curriculum to adapt to changing contextual realities and emerging trends in teaching statistics as a discipline using real data with technology-mediated teaching-learning activities. This paper intends to contribute to the limited curriculum resource materials in the country for teaching *Statistics and Probability* within the national *K to 12* curriculum framework. Further, we endeavor to strike a balance in meeting global standards of statistics instruction with considerations of the local situation, needs and resources available. Applying the concept of "*glocalized*" education, the teaching-learning and assessment activities provide meaningful integration of local data sets about the Filipino youth and some youth statistics generated by international surveys as the platform for young people to learn statistical concepts and methods and make sense of the data involving the youth population to which they belong.

COURSE BACKGROUND

Statistics and Probability is a core subject in the new *K to 12* Senior High School (SHS) program in the Philippines which has been implemented beginning School Year 2016-2017. It will be taken by Grades 11 or 12 students with a required time allotment of 80 hours in a semester. The pre-requisite knowledge for this course includes topics taken under the *Statistics and Probability* learning domain which is part of a spiral mathematics curriculum from Grades 1 through 10. For the statistics component, these topics include simple data representations using various kinds of tables and graphs, descriptive summary measures and frequency distributions of data. For the probability component, the topics include understanding the language of probability of events, statistical experiments, possible outcomes in an experiment and calculating simple probabilities. This new *Statistics and Probability* course covers six basic topics; namely: (1) Random Variables and Probability Distributions, (2) The Normal Distribution, (3) Sampling Distribution and the Central Limit Theorem, (4) Estimation of Parameters, (5) Hypothesis Testing, and (6) Correlation and Regression (Department of Education, 2013).

Further, in line with the development of statistical literacy as the overarching goal of statistics education, the course seeks to develop this important skill in students to enable them to read, organize and make sense of data that they encounter in their daily lives from various sources. Thus, with the goal of developing statistical literacy among the students and in keeping with the goals of the *K to 12* curriculum reform in the Philippines, this project was initiated to make the teaching of statistics more meaningful and relevant by using real data in context, and to promote understanding on the social dimensions of the youth sector population.

YOUTH STATISTICS AND DATABASES

The youth generally comprise a considerable sector of a nation's population. The definition of youth varies across institutions and countries. In this paper, we adopt the definition set by Republic Act 8044 in the Philippines, otherwise known as the "*Youth in Nation-building Act*" which defines the *youth* in the Philippines as those who are 15 to 30 years old. Adopting this official definition, the number of youth in the Philippines totals 20.7 million or constituting nearly one-third of the country's total population in 1995 Census and this has increased to 27.4 million in the 2010 Census of the Philippines (Philippine Statistical Authority, 2010)

To promote understanding of the youth sector along with their vital role and impact in society, we use youth statistics to design a *Statistics and Probability* course for senior high school students in the Philippines, aged between 16 and 18 years. The integration of youth data in the design of teaching-learning activities for the course makes use of data from the Philippine Statistical Authority (PSA), http://nap.psa.gov.ph/i-stats/. The Interactive Statistical Database (*i-stats*) is one of the services offered by the PSA on the Internet. The *i-stats* is not an open database but a subscription service for viewing and downloading datasets. We have subscribed to this database with the permission to use selected variables for our course development project. Data on selected variables concerning the Filipino youth from the national census and other sample surveys about the youth are used. Aside from data from PSA, selected variables from other youth surveys such as the 2011 Project Youth Survey conducted by Social Weather Station (SWS), http://www.sws.org.ph/, will be used to illustrate some basic statistical concepts. Since basketball is a popular sport among young people in the Philippines, and data sets related to such sport is of great interest among young people, data on basketball players in the Philippine Basketball

Association, (http://www.interaksyon.cpm/interaktv/2013-pba-pre-draft-measurements) will be also used to illustrate concepts and methods in correlation and regression analyses. Further, to provide an international perspective of the youth population and their impact on society, youth statistics from the following databases were also used in this course development project.

- Minnesota Population Center, Integreated Public Use Microdata Series (IPUMS), http://international.ipums.org/international
- European Commission, Eurostat Databases on Youth Statistics, http://ec.europa.eu/eurostat/data/database

Variables of interest in these databases include youth population demographics, education and training, employment, health, social inclusion, participation, volunteering and youth in the digital world. By providing students the opportunity to work on these real data about the youth population through activities that build their understanding of statistical concepts and help develop their statistical literacy skills, we also promote in them a deeper understanding of their characteristics and vital role in society and national development.

COURSE DESIGN ELEMENTS AND PRINCIPLES

The activity resource book is a purposeful collection of relevant teaching-learning-assessment activities which are constructively aligned with the content and performance standards, target learning competencies, and assessment in the SHS Statistics and Probability curriculum guide. The teaching-learning and assessment activities are intended to promote statistical literacy and conceptual understanding while at the same time, promote students' social consciousness and understanding of the youth's population dynamics and vital role in society. The activities follow inductive teaching and learning approaches where instruction begins with a scenario - a set of observations or data set to interpret, a case study to analyze, or a real data-based problem to solve instead of beginning with general theories and principles, and eventually getting to applications (Prince and Felder, 2006). Real data sets are stored in *Microsoft Excel* spreadsheets. The design of the material adheres to the principles of constructive alignment, curricular connection and relevance, "glocalization" and use of social context.

The course is divided into two modules. Module 1, entitled *Understanding the Foundations of Statistical Inference through Youth Demographics*, comprise activities that use national census and sample survey data to build conceptual understanding on such topics as random variables, probability distributions, sampling distributions and the Central Limit Theorem which lay the foundation of inference. Sample activities are described below.

Sample Activity 1. Randomness is all around you.

In this activity, students will explore the variability in selected variables related to the youth sector in Philippine households and define discrete and continuous random variables from survey data. Using the 2010 Census of the Philippines, they will simulate taking a random sample of 100 households from one geographical unit of the census data. Whether or not a household selected has a youth member is a random phenomenon. From this random phenomenon, we define random variable X as follows: Let X = the number of youth members in the household selected. If the results show that $X = 0, 1, 2, 3, \ldots, 10$ youth members in the households, for instance, students will construct the relative frequency distribution table for all the possible values of X as a springboard for introducing the concept of probability distributions. From the random variable X, the activity will provide students through guided inquiry the opportunity to define other possible random variables from the data.

Sample Activity 2: Understanding sampling distributions through responses in a youth survey

This activity aims to build students' conceptual understanding of sampling distributions and the Central Limit Theorem. Using the variable *Age* in the 2011 SWS Youth Survey, students will perform simulations for drawing simple random samples of size 10 repeatedly to introduce the concept of sampling distribution of a statistic. Sampling distribution is a topic that provides an important foundation for understanding the methods of statistical inference. Several studies in statistics education literature have documented that many students have struggled with the notion

of the sampling distribution and such concept along with probability distributions and the Central Limit Theorem were found difficult to understand (Lunsford, Rowell & Goodson-Espy, 2006). In the Philippines, an evaluation study on locally available textbooks in college introductory statistics commissioned by the Philippine Statistical Association revealed that this topic is skipped and there is a disconnect between the topic on probability and the methods of inference (David & Maligalig, 2006).

From the 2011 SWS Youth Survey, we define a new population distribution of ages of the youth respondents in the whereby students select simple random samples of size 10. Sampling is done repeatedly with replacement and students record their values of the ages and calculate the sample mean and standard deviation. The data set will be available in Microsoft *Excel* for computer-based simulation of simple random sampling. However, using paper cut-out of printed values of the ages may serve as alternative when classes do not have technology access. Students will record the distribution of sample mean values and observe their variation by constructing a relative frequency distribution table and a histogram showing the sampling distributions of the means ages. Through guided inquiry, students will realize that the sample mean is a random variable with a corresponding probability distribution. The activity will serve as the springboard for introducing the concept of sampling distribution of the mean. Students will repeat the procedure for increased sample sizes of 20, 30 and 40 and explore the shape (symmetry), center (mean) and spread (standard deviation) of the distribution. They will verify that although the population distributions have distinctly different shapes, the distribution of the sample means approaches the normal distribution as the number of samples increase. The activity will further lead to the introduction of the Central Limit Theorem.

The second module of activities is entitled *Youth Statistics as the Tool for Social Inquiry and Statistical Investigations.* This module comprises activities which use youth data from national and international comparative surveys to provide the context for inquiry and statistical investigations and illustrate the methods of statistical inference; namely: Estimation of Parameters and Tests of Hypotheses. Two sample activities are described below.

Sample Activity 3.Youth Statistics for Parameter Estimation and Hypothesis-testing

In this activity, youth statistics will be used as tools to provide the context for inquiry and statistical investigations. Students will apply knowledge of parameter estimation by constructing confidence intervals on sample survey results and perform tests of hypothesis to establish statistical significance of differences and relationships in selected variables on the youth. This activity uses the same data set as in Activity 2. A box containing 100 values, printed on index cards, of the ages of the youth respondents referred to in Activity 2 will be used along with some graphing papers. The data is treated as a population with a known mean and a standard deviation. Students will take turn in randomly drawing 10 cards from the box and return them after recording the values. Then, the mean and standard deviation for each sample will be calculated. Students will record collectively the distribution of sample mean values and construct a dot plot. This will be used as springboard for introducing the concept of "margin of error for estimating a population mean". To test students' understanding on margin of error, the activity also provides several other situations where they have to determine a reasonable estimate of the margin of error and think of how would the estimate vary if the sample size is increased to 20, 25, 30 or even 50. Students will repeat the procedure using *Microsoft Excel* or any similar spreadsheet for larger sample sizes (e.g. 25, 30, and 50) in order to justify their claims. The results from this simulation activity are used to define margin of error and suggest a method or formula for estimating the margin of error from a single sample of size n with a standard deviation s.

Activity 2 can be extended to hypothesis testing on the difference between means. Using the same variable "Age" of youth respondents in the 2011 Youth Survey, students will randomly select two sets of 50 values, one set corresponding to the ages of male youths while the other set represents the ages of female youths. Presented in a table, students will compare the two distributions of ages. They will determine if the respondents differ significantly in terms of their ages. They will compute the mean and the standard deviation of the ages of the randomly selected male and female youths. Based on the results, the students will confer if indeed the two groups differ in terms of their mean ages. The teacher will benchmark on this difference to lead the class

to understanding the need for hypothesis testing. Use the situation as scaffolding for explicit steps of instructions. Otherwise, proceed to performing the hypothesis test using the claim "the two groups differ significantly in terms of their mean ages". The results from the calculations are used to formulate a decision as to when a significant difference can be established from two different sets of measure, especially the mean between groups and how important the confidence level is when stating the conclusion. Other similar situations would be presented to evaluate the depth of students' understanding on hypothesis testing on mean difference.

Sample Activity 4. Heights Versus Weights

With a sample of size of 30, taken from the data set on heights and weights of players in the Philippine Basketball Association (PBA) draftees in 2013 (complete data set taken from http://www.interaksyon.cpm/interaktv/2013-pba-pre-draft-measurements), students will identify the independent and dependent variables, explain their choice and manually create a scatter plot of the raw data and draw the line that seems to "fit" the data. Then they calculate some required summation of values to complete a table of values necessary to manually solve for the Pearson correlation coefficient, whose value they need to interpret with reference to the data set used. Using *Microsoft Excel,* the students will also do correlation analysis of the same sample data and that of the complete data set, and compare results. The same data set will be used to discuss the concept of regression analysis as an extension of this activity.

Aside from these activities, an assessment scheme will be included in the course disk, as well as bridging learning opportunities, and extension and enhancement activities that include web-based links to relevant sources. The assessment scheme will include rubrics as a scoring guide for the teachers and the students as well. Through the rubrics, the students will understand what and how much is expected of them and by what criteria their work will be evaluated.

INITIAL ASSESSMENTS AND FUTURE DIRECTIONS

The use of youth statistics from both local and international databases will provide a rich ground for approaching the teaching and learning of statistics from a cross-disciplinary social science perspective. Since these students primarily belong to the youth sector of the population, using youth statistics is expected to generate understanding and appreciations of the youth sector to which they belong and of important role they play in society.

As with most innovations in curriculum that require technology integration, there are merits and limitations, promises and pitfalls. In an initial needs assessment survey for secondary mathematics teachers in Metro Cebu participated by 92 teachers in 17 randomly selected high schools, the results pointed out several issues of concern in teaching *Statistics and Probability* with technology integration; namely: (1) access to a school computer laboratory for statistics classes, (2) teachers' capacity for technology integration; and (3) teachers' pedagogical content knowledge on the topics upon which these activities are anchored (Reston, 2015). While many schools in the Philippines including public high schools have now at least one computer laboratory, these laboratories are generally for computer classes and not dedicated for the teaching of statistics. For schools which have more technology-based resources, statistics teachers may have to make special arrangement for the use of computer laboratory. As for teachers' capacity for teaching lessons that appropriately combine content, teaching approaches and technology, the teachers rated themselves in a scale from 1 to 7 with an average of 5.06. A series of 8 three-hour workshop sessions conducted last April to May 2016 for incoming senior high school teachers also indicated that teachers' knowledge of random variables and sampling distributions need to be strengthened first before they can make sense of data in social contexts in facilitating these proposed activities with their students. While these proposed course materials will provide a springboard for developing statistical literacy skills and competencies among young people to meet the demands of 21st century society, the practicality of implementing them in technology-enabled statistics classrooms will have to be addressed foremost. One alternative we proposed to teachers is the use of the FLIPPED classroom model where students explore selected activities and data sets in the course disk with some directions and guiding questions before class time so that during class time the focus is on discussion and follow-up activities that connect students' knowledge of the lesson with the data and statistical concepts and methods.

ACKNOWLEDGMENT

The participation of Enriqueta Reston at the IASE 2016 Roundtable Conference was enabled/supported by a joint grant from the World Bank Trust Fund for Statistical Capacity Building, International Statistical Institute (ISI) and the International Association for Statistics Education (IASE).

REFERENCES

David, I. & Maligalig, D. (2006). Are we Teaching Statistics Correctly to our Youth? *The Philippine Statistician*, 55(3-4), pp. 1-28.

Department of Education (2013). The *K to 12* Basic Education Program. Pasig City: Department of Education, www.deped.gov.ph

Forbes, S. (2014). The Coming of Age of Statistics Education in New Zealand, and its Influence Internationally. *Journal of Statistics Education. 22 (2),* www.amstat.org/publications/jse/v22n2/forbes.pdf

Gal, I. &Garfield, J. (1997). Curricular Goals and Assessment Challenges in Statistics Education, In: I. Gal & J. B. Garfield (Eds.), *The Assessment Challenge in Statistics Education*, Amsterdam, The Netherlands: The International Statistical Institute.

Lunsford, M. L., Rowell, G. & Goodson-Espy, T. (2006). Classroom Research: Assessment of Student Understanding of Sampling Distributions of Means and the Central Limit Theorem in Post-Calculus Probability and Statistics Classes. *Journal of Statistics Education. 14 (3)* www.Amstat.Org/Publications/Jse/V14n3/Lunsford.Html

Murray, S. & Gal, I. (2002) Preparing for Diversity in Statistical Literacy: Institutional and Educational Implications, *Proceedings of the Sixth International Conference on Teaching Statistics (ICOTS6).*

National Census of the Philippines (2010). *Philippine Statistical Authority* (PSA), http://nap.psa.gov.ph/i-stats/.

North, D. & Scheiber, J (2008). Introducing Statistics at School Level in South Africa: The Crucial Role played by the National Statistics Office in Training In-service Teachers. In: C. Batanero, G. Burrill, C. Reading & A. Rossman. *Joint ICMI/IASE Study: Teaching Statistics in School Mathematics. Challenges for Teaching and Teacher Education.* Proceedings of the ICMI Study 18 and 2008 IASE Round Table Conference.

Prince, M. J. & Felder, R. M. (2006). Inductive teaching and learning methods: Definitions, comparisons, and research bases. *Journal of Engineering Education, 95 (2),* 123-138.

Philippine Congress (1995). Republic Act 8044: Youth in Nation-Building Act. http://www1.umn.edu/humanrts/research/Philippines/RA 208044.YouthInNation-BuildingAct.pdf

Reston, E. (2015). Teachers' Knowledge Bases and Pedagogical Approaches for Implementing the *K To 12* Mathematics Curriculum in the Philippines**.** *Proceedings of the Eighth International Conference on Science, Mathematics & Technology Education*, Jakarta, Indonesia: Universitas Negeri Jakarta.

Ridgway, J. , Nicholson, J. & Mc Cusker, S. (2011). Developing Statistical Literacy in Students and Teachers. In C. Batanero, G. Burrill & C. Reading (eds.), *Teaching Statistics in School Mathematics- Challenges for Teaching and Teacher Education*. A Joint ICMI/IASE Study, Springer Science + Business Media B.V. 2011.

Rumsey, D. (2002). Statistical Literacy as a Goal for Introductory Statistics Courses, *Journal of Statistics Education*, 10(3). www.amstat.org/publications/jse/v10n3/rumsey2.html

Verschut, A. & Bakker, A. (2010) Towards Evaluation Criteria for Coherence of a Data-based Statistics Curriculum. In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics (ICOTS8, July, 2010), Ljubljana, Slovenia*. Voorburg, The Netherlands: International Statistical Institute. www.stat.auckland.ac.nz/~iase/publications.php

Watkins, A., Bargagliotti, A, & Franklin, C. (2014). Simulation of the Sampling Distribution of the Mean Can Mislead. *Journal of Statistics Education*. 22, (3) www.amstat.org/publications/jse/v22n3/watkins.pdf.