INTRODUCTORY STATISTICS WITH DEFERRED UNI VARIATE FOR LONG TERM ACADEMIC DEGREE PROGRAM

Lita Pertiwi, Febry Utami, Noor Laila, Aulia Dini, Nuha Gunawan
STIS 53 Social Division, Indonesia
STIS 56 Computer Division, Indonesia
STIS 53 Economic Division, Indonesia
STIS 53 Economic Division, Indonesia
STIS 53 Social Division, Indonesia
febry.utami.id@gmail.com

There are educations such as four-year academic degree statistics. These thought may not be directly applicable to short-term statistical education such as four-month non-degree course. In official statistics, multi variate is used to start long-term statistical education. Supporting infrastructure necessary includes: (1) bi variate stratification of universe, (2) two-way dummy table for census publication, (3) complete enumeration or census questionnaire, (4) Custom made tabulation software. Multi variate always uses actual census publication and questionnaire. This way, firstly, introductory statistics avoids any dichotomy of cluster sampling and stratified sampling. Secondly, introductory statistics initiates with available options in the universe for stratification. Thirdly, it starts with non-physical contiguity (spatial boundary). Lastly, it establishes given probability P(B/A=k & C=j) instead of P(B).

INTRODUCTION

Using a lot of mathematical notation and other assumption, it makes understanding statistics difficult. It leads to misinterpretation or bias from using statistics in some research (Zidek, 1986). On the other hand, statistics is believed as a key to research and project related research (Weldon & Tham, 1990). Teaching multivariate is supposed to include sticking to the conceptual, rather than mathematical by non-parametric methods Anderson (2002).

Furthermore, integrating the computer and computing on teaching statistic is also important (Zidek, 1986) . This implementation is presupposed the availability of appropriate statistical packages and adequate computing statistics (Keung & Soon, 1990).

BI VARIATE STRATIFICATION OF UNIVERSE

The primary topic in teaching introductory statistics may not be examining the universe (or the population). Many times universe already has some kind of stratification, introductory statistics may choose which stratum after stratum to be taught, not obligatory to cover all strata Biehler (1986). A variate can be introduced to form a simultaneous stratification to arrive at bi variate stratification. It is not compulsory to teach any topic using all strata but which stratum after stratum instead. That is if the intention is to create simultaneous stratification with implicit (available) stratification. Example of implicit (available) stratification is regional minimum wage [headline of Kompas Indonesia daily newspaper date 02 November 2017].

START WITH GIVEN PROBABILITY P(B/A=K & C=J) INSTEAD OF P(B)

Based on research by Clayson (1990), it is suggested to introduce statistics to college business student with reversing the topics start with introducing multivariate analysis first. He used interrelated data and elaborates with some techniques such as multiple regression, cluster analysis, multidimensional scaling, questionnaire design and contingency table.

Introductory statistics may choose which stratum after stratum to be taught and implicitly stratum A is represented by a value of (A=k). Any discussion about B within stratum A such as P(B) is actually P(B/A=k). Other variate for example C can be introduced to form a simultaneous stratification to arrive at bi variate stratification. Therefore P(B) is actually P(B/A=k & C=j). Conditional probability can compute P(B/A=k & C=j) with for example multiple regression (James Clayson, 1990). In earlier times, simple probability is taught in the condition of nonexistent computer. Nowadays student can one way or another access a computer or smart phone having

regression. Encourage teaching multi variate with conditional probability P(B/A=k & C=j), specifically tri variate and bi variate.

TWO WAY DUMMY TABLE FOR CENSUS PUBLICATION
        The simplest dummy table can be a row vector with one column subtotal or a column vector with last row as subtotal. This vector can be considered as empirical conditional probability such as P(B/A=k) with stratum A subtotal. Since the data is from census the vector leads to marginal empirical distribution. For completeness all strata can be shown as a matrix namely two-way dummy table. The remaining energy, remaining resources, remaining time, and remaining human capacity alternately can be devoted to create two-way table or three-way table. Do not create a hundred one-way dummy tables; instead create twenty three-way dummy tables or at the minimum two-way dummy tables, even though not covering all strata. This is also has been recommended by Clayson (1990), to create contingency tables.

CUSTOM MADE TABULATION SOFTWARE
        Keung & Soon (1990) gave a recommendation to college business students to use a non-mathematical data driven for teaching multivariate. One way to do that is by using a proper package and a computing facility on big and undifferentiated business data.
        Other than minimum wage there are some implicit physical stratifications in the universe such as residential area, four types of street class, six levels of power line wattage. These stratifications are well known for ease and speedy custom made in – house software developer. Source code for the universe may not be ready for distribution instantly. At least source code for a particular stratum including simple coherency check can be made open to general audience before the release of tabulation result. It is recommended to use at least more than one computer language the same raw data and for the same dummy table. Further different style of programming is welcomed such as procedural programming and object oriented programming. This is also has been recommended by Zidek (1986), to integrate computer with statistics curriculum.

COMPLETE ENUMERATION OR CENSUS QUESTIONNAIRE
        Teaching statistics on long term academic program can be adapted from some statistics training which organized by Bureau of Census. Example statistics training in U.S. Bureau of the Census is using case study approach on the process of planning a survey or census (Shoemaker, Bryson, Brown, & Solomon, 1986).
        Census questionnaire has only a handful of substance questions (and filter questions) which may lead to fast response from respondent and hopefully short time tabulation. When teaching data collection and tabulation discourage sampling but first encourage census. For instance census tabulation covering three or two strata, it is not required to cover all strata at once. Sampling in general requires bi variate theoretical frequency distribution which often resorts to bi variate normal. Census does not that much affected by more difficult concept of tri variate theoretical frequency distribution. Census may avoid early discussion on theoretical frequency distribution of sampling error (Anderson, 2002).

INTRODUCE NON-PHYSICAL CONTIGUITY (SPATIAL BOUNDARY)
        Several high end official regional minimum wages are probably found in Southern Tangerang (shared land border with Jakarta), Sidoarjo (shared land border with Surabaya), Badung (shared land border with Denpasar). Physically they are all far apart but based on regional minimum wages they are regarded as neighbors or contiguous by non-physical boundary. One of the well-known examples in the world is direct flight. A direct flight between Singapore and Manila indicates that Singapore and Manila are contiguous Jakarta and Manila are not considered as neighbors because there is not any direct flight, although flight distance between Jakarta and Manila is comparable to that of Singapore and Manila. This may help in archipelago with many regions do not shared any land border.

CONCLUSION

Two-way dummy tables will hopefully ease teaching univariate from two different sources of raw data such as iterative proportional fitting [raking].

Bivariate spatial also considered to be taught. Not only considered physical boundary but also non-physical boundary such as direct flight. In bivariate setting, it seem more intuitive to use concept related to the distance in attribute space between a point and the points that correspond to its geographic neighbors. Local Geary c is used in bivariate/multivariate spatial.

REFERENCES

American Statistical Association. (2012, June 1). Traits of a Successful Statistician. STATtr@k. Retrieved from http://stattrak.amstat.org/2012/06/01/successfulstatistician/#

Anderson, M.J. (2002). Teaching Multivariate Statistics to Ecologist and the Design of Ecological Experiments to Statisticians: Lesson from Both Sides. Proceedings of 6th International Conference on Teaching Statistics (ICOTS 6), [pp.1-6] Cape Town, South Africa.

Anselin, Luc. 2017. A Local Indicator of Multivariate Spatial Association: Extending Geary's c. Chicago: Center for Spatial Data Science University of Chicago.

Biehler, Rolf. (1986). Exploratory data Analysis and The Secondary Stochastics Curriculum. Proceedings of 2nd International Conference on Teaching Statistics (ICOTS 2), [pp.79-85] Victoria, Canada.

Clayson, James. (1990). Teaching Multivariate Analysis to Business Students. Proceedings of 3rd International Conference on Teaching Statistics (ICOTS 3), [pp.332-337] Dunedin, New Zealand.

Daza, Sebastian. (2012, August 25). Raking weights with R. Retrieved from http://sebadaza.com/survey/2012/08/25/raking/

Indonesian Law Number 38 year 2004 regarding Road. (2017, November 12). Kepmen Menteri PU No.58 tahun 2012 tentang penetapan Kelas Jalan Jawa dan Sumatera. Retrieved from http://preservasi-pupera.info/prv/wp-content/uploads/2016/07/Kepmen-Menteri-PU-No.58-tahun-2012-tentang-penetapan-Kelas-Jalan-Jawa-dan-Sumatera.pdf

Kompas Indonesia daily newspaper. (2017, November 2). Relokasi Usaha Menjadi Pilihan Pelaku Industri. Kompas Indonesia, Head line page.

Shoemaker, H. H., Jr., Bryson, K.R., Brown, P. (Tim). , & Solomon, L. (1986). Academic versus Applied Training in Statistics Approaches Used by the U.S. Bureau of the Census. Proceedings of 2nd International Conference on Teaching Statistics (ICOTS 2), [pp.249-253] Victoria, Canada.

Weldon, K. L., Tham P. (1990). What is Basic Statistics? Lesson from a Canadian-Indonesian Project. Proceedings of 3rd International Conference on Teaching Statistics (ICOTS 3), Dunedin, New Zealand.

Wong, W.K., Soon, T.W. (1990). Teaching Multivariate Analysis to Business-Major Students. Proceedings of 3rd International Conference on Teaching Statistics (ICOTS 3), [pp.343-349] Dunedin, New Zealand.

Zidek, J.V. (1986). Statistication: The Quest for a Curriculum. Proceedings of 2nd International Conference on Teaching Statistics (ICOTS 2), [pp.1-17] Victoria, Canada