

Comments on Statistical Issues in March 2014

Commentary

Yong Gyu Park

Department of Biostatistics, The Catholic University of Korea College of Medicine, Seoul, Korea

In this section, we explain the statistical methods for analyzing the Korean National Health and Nutrition Examination Survey (KNHANES) data, which appeared in the articles titled, “Coffee consumption and bone mineral density in Korean premenopausal women”, by Choi et al.¹⁾ and “The characteristics of false respondents on a self-reported smoking survey of Korean women: Korean National Health and Nutrition Examination Survey, 2008”, by Lee et al.²⁾ published in January 2014.

STATISTICAL METHODS FOR ANALYZING KNHANES DATA

The KNHANES is a nationwide cross-sectional survey which has been conducted by the Korea Centers for Disease Control and Prevention since 1998, is designed to accurately assess national health and nutrition levels, and consists of a health interview, health examination, and nutritional assessment. A complex, stratified, multistage cluster sampling design with proportional allocation was used for the selected household units that participate in the survey.

Numbers of researchers and articles using the KNHANES data have rapidly increased in recent years; however, there are still many mistakes in the statistical analysis methods. Typical examples of such mistakes are 1) ignoring sample design and 2) fallacious presentation of the study results.

As stated above, KNHANES data are obtained by a complex, stratified, multistage cluster sample design; thus, the data should be analyzed using proper weights. ‘Proper weights’ means that each observation in KNHANES data is obtained by a different sampling probability. On the other hand, the most well known statistical methods assume that each observation is obtained

by simple random sampling, and thus all observations have the same sampling probability (weight). Therefore, if we attempt to analyze KNHANES data using conventional statistical methods, we obtain seriously biased results.

There are many statistical programs such as SAS, SPSS, R, SUDAAN, and STATA, which could be used to analyze KNHANES data. In SAS, we can analyze the following:

PROC SURVEYMEANS (mean analysis)
 PROC SURVEYFREQ (proportion analysis; chi-square test)
 PROC SURVEYREG (regression analysis; t-test, analysis of variance, regression)
 PROC SURVEYLOGISTIC (logistic analysis)
 PROC SURVEYPHREG (Cox regression)

In SPSS, we can analyze the following using Complex Sampling:

Frequency analysis
 Descriptive statistics
 Cross tabulation
 Proportions
 General linear model
 Ordinal regression
 Cox regression

The analysis results of KNHANES data are usually presented as weighted mean \pm standard error of mean (SEM) or weighted proportion (SE). The reason for providing standard error instead of standard deviation is attributed to the fact that standard deviation only describes variation of sample data. On the other hand, standard error provides the precision of estimate (weighted mean/weight proportion) of the national population, which is

entirely pertinent to the aims of KNHANES.

We present a well-turned expression of 'statistical analyses' in one of the KNHANES data articles.³⁾

"SAS ver. 9.2 (SAS Institute Inc., Cary, NC, USA) survey procedure was used for statistical analysis, using KNHANES sampling weights to acquire nationally representative estimates. The analysis was adjusted for survey year to minimize the variations between survey years. The data in this study are presented as the mean \pm SE or proportion (SE) for continuous or categorical variables, respectively. ... Multivariable logistic regression analyses were applied to examine the association between insulin resistance and periodontitis. The odds ratios of periodontitis were calculated using the insulin-sensitive group as the reference. Calculations were made, adjusting for survey year, age, educational level, house-hold income, smoking status, alcohol consumption, exercise, use of floss, use of interproximal toothbrush and brushing teeth before bed. A P-value < 0.05 was considered statistically significant."

CONFLICT OF INTEREST

No potential conflict of interest relevant to this article was reported.

REFERENCES

1. Choi EJ, Kim KH, Koh YJ, Lee JS, Lee DR, Park SM. Coffee consumption and bone mineral density in Korean premenopausal women. *Korean J Fam Med* 2014;35:12-8.
2. Lee DR, Kim HS, Lee J. The characteristics of false respondents on a self-reported smoking survey of Korean women: Korean National Health and Nutrition Examination Survey, 2008. *Korean J Fam Med* 2014;35:28-34.
3. Lim SG, Han K, Kim HA, Pyo SW, Cho YS, Kim KS, et al. Association between insulin resistance and periodontitis in Korean adults. *J Clin Periodontol* 2014;41:121-30.