Escape from omnishambles in statistics: back to the basics

Sangseok Lee

Department of Anesthesiology and Pain Medicine, Sanggye Paik Hospital, Inje University College of Medicine, Seoul, Korea

In 1859, Charles Darwin (1809–1882) wrote that the accumulation of minor variations among individuals could result in evolution. In his great work on evolution, he first introduced the statistical concept of continuous variation. Subsequently, Walter Weldon (1860–1906), Sir Francis Galton (1822–1911), and Karl Pearson (1857–1936) developed several statistical methods to measure biological variation. That was the beginning of modern biostatistics.

Many clinical symptoms and therapeutic results that we routinely experience in clinical practice show variation by chance. The image used to quantify the extent of such variation is referred to as a probability distribution. This is a pattern that shows the observed results collectively for each element of the same group or repeatedly observed results for specific groups. The distribution may be a normal, binomial, geometric, a Poisson distribution, or other forms. It is not an exaggeration to say that almost all biostatistics are based on the normal distribution, as biological phenomena, measurements of human psychological characteristics, and sociological variables are usually distributed normally. To determine the shape of the normal distribution, we need only know the mean and variance. Calculating probability is also easy because the fraction of the distribution under the total area of the z-value in the normal distribution curve is the same, regardless of the shape of the normal distribution curve. In addition, although a population may not follow a normal distribution, with a sufficiently large random sample, a normal distribution can be assumed in almost all cases due to

the central limit theorem. For example, a binary variable following a discrete distribution could be assumed to follow a normal distribution if the sample size were increased.

Most medical journals require statistical analysis to summarize the data obtained for a sample and to generalize the results. As medicine has become more complex, statistical analysis has also become increasingly complex and diverse. In the era of evidence-based medicine, the importance of statistical analysis is being emphasized. An error in statistical analysis techniques means that the conclusions of a medical paper are wrong. However, few researchers, peer-reviewers, and even journal editors are well versed in statistical analysis, and many medical papers are published without correcting statistical errors.

In 1966, Schor and Karten [1] reported on the application of statistics and the status of statistical errors after reviewing 295 papers published in 10 major medical journals, including *JAMA*. They discovered that typical statistical errors were the misapplication of nonparametric and parametric tests, the failure to apply corrections, and disregard for statistical independence. Only 28% of the reviewed papers were evaluated as statistically acceptable, while 67% were statistically defective and 5% were unacceptable. This research was an opportunity to reach consensus on the need to monitor the adequacy of statistics used in papers published in medical journals. Other research has reported similar results. For example, the famous report by Gore et al. [2] concluded that about 52% of the evaluated papers contained various statistical errors. The most common errors included inappropriate representation, presentation of scatter indicators, and inappropriate application of Student's t-test and the chi-square test. It is possible that these reports underestimated the percentage of errors, as many papers omit or dissimulate data, rendering post-verification impossible.

In a similar study in Korea, Lee and Ahn [3] analyzed 382 papers published from January 1980 to December 1989 in the *Journal of the Korean Medical Association*. They found one or more errors in the statistical analysis in 97.6% (290/297) of the papers using statistical methods. The two main analytical errors

Corresponding author: Sangseok Lee, M.D.
Department of Anesthesiology and Pain Medicine, Sanggye Paik Hospital, Inje University College of Medicine, 1342, Dongil-ro, Nowon-gu, Seoul 01757, Korea
Tel: 82-2-950-1171, Fax: 82-2-950-1323
E-mail: s2248@paik.ac.kr
ORCID: <http://orcid.org/0000-0001-7023-3668>

Korean J Anesthesiol 2015 October 68(5): 431-433
<http://dx.doi.org/10.4097/kjae.2015.68.5.431>

were 1) errors of omission and 2) errors of commission. The former included an incomplete description of basic data (19.2%), statistical tests performed yet not defined (58.2%), and incomplete description of the power or confidence interval (91.9%); the latter included inadequate descriptions of measures of central tendency or dispersion (27.9%), incorrect analysis (48.1%), multiplicity on hypothesis testing (65.2%), and unwarranted conclusions (52.2%).

What causes such high rates of statistical error in the medical literature? It is inadequate statistical methodology training for potential authors of medical papers. Without understanding the basic concepts of statistics, one cannot read the medical literature critically and it is impossible to obtain correct results. Statistics is the primary means for acquiring medical knowledge in medical research. Friedman and Phillips assessed the understanding of the concepts of correlation coefficient and confidence interval in 684 residents working in the pediatrics department of major general hospitals in the United States and found that only 17.3% understood the significance of correlation coefficients and 61.3% understood confidence intervals [4]. They concluded that it is difficult to read a medical article correctly, because the understanding of key medical biostatistics terms was so low. Therefore, the idea of starting a training program in statistics has emerged. The results of research in Korea are no different. Systematic medical statistics education is needed in training courses. To motivate trainees, I carefully propose that the medical statistics should be introduced in the professional training course as well professional examination.

Recently, the *Korean Journal of Anesthesiology (KJA)* initiated its publication of **STATISTICAL ROUND** to support training in basic medical statistics for potential authors and reviewers of medical papers, by providing useful information and interesting articles. The *KJA* hopes that this effort will help readers and reviewers.

In next issue of the *KJA*, Kim [5] reported on the t-test in great detail in a reader-friendly manner. Because it covers everything from basic concepts through actual examples, the paper should help to establish concepts. The t-test is the most commonly used statistical analytic method in medical papers. The t-statistic was introduced in 1908 by the English statistician William Sealy Gosset (1876–1937), who was working for the Guinness Brewery in Dublin, Ireland. He published under the pen name ‘Student’, and developed Student’s t-distribution. He contrived the t-test as a cheap way to monitor the quality of stout. The t-test can be used to compare two group means, but it is also widely applied incorrectly to compare multiple groups (i.e., in all of the pairwise comparisons or comparing more than one intervention with the control condition or comparing the conditions at different times following an intervention). Williams et al. [6] reported

that more than half of all the articles in the *American Journal of Physiology* applied unpaired or paired t-tests. Of these, about 17% used the t-test incorrectly for the purpose of multiple comparisons, because they did not modify the test with a correction method (e.g., Bonferroni). Glantz [7] reported an analysis of the use of t-tests in the journal *Circulation* and found that 27% of 142 original articles used the t-test incorrectly to compare three or more groups. This misuse has led to wrong conclusions in some papers. Such errors increase the chance of reporting that some drug or therapy was effective without any supporting evidence.

Because it is virtually impossible to enroll the entire population in a study, we usually study a portion of the population as a sample. To understand the characteristics of the distribution from the observed samples, it is important to use statistics suitable for the distribution. However, many researchers fail to determine the distribution of the observed data in a clinical study. As a priority, they should draw a scatterplot of the observed data. It goes without saying that it is necessary to calculate an appropriate sample size to provide a basis for the statistical power.

Data in medical research are sometimes skewed and should be analyzed using different statistical tests in a manner appropriate to their distribution. The majority of medical papers rarely use the normality test, which shows how skewed their data are. Normality can be assessed graphically (e.g., comparing a histogram of the sample data to a normal probability curve) or statistically (e.g., the Shapiro–Wilk or Kolmogorov–Smirnov test). If data are not distributed normally, either non-parametric analysis should be applied or the data should be transformed to a normal distribution (e.g., log-transformation). Mathematically, data transformation is relatively simple; however, the interpretation of the results can be difficult.

Some researchers think that it is possible to derive objective findings using advanced statistical methods, without regard to the quality of the data. Of course, it is important to select the appropriate statistical analysis to achieve the research goal in scientific research. Equally important, however, is developing the proper research plan before data collection and analysis. If the data are not collected using correct processes, the quality of the data and reliability of the results will be low, even using the most advanced statistical analysis. It is easy to find good references about the statistical and methodological considerations for medical research in papers published in the *KJA* [8].

Statistics is not a magic wand. We are now lost in statistics and need to start again from the basics. Researchers should learn the basic principles of statistics and try to obtain good qualitative data. They should apply appropriate statistical techniques after identifying the nature of their data. Finally, they should analyze the results correctly and draw the right conclusions.

References

1. Schor S, Karten I. Statistical evaluation of medical journal manuscripts. *JAMA* 1966; 195: 1123-8.
2. Gore SM, Jones IG, Rytter EC. Misuse of statistical methods: critical assessment of articles in *BMJ* from January to March 1976. *Br Med J* 1977; 1: 85-7.
3. Lee HK, Ahn YO. An assessment of methodological and statistical validity of medical articles published in Korea, from 1980 to 1989. *Korean J Med Educ* 1991; 3: 52-69.
4. Friedman SB, Phillips S. What's the difference? Pediatric residents and their inaccurate concepts regarding statistics. *Pediatrics* 1981; 68: 644-6.
5. Kim TG. T-test as parametric statistic. *Korean J Anesthesiol* 2015 [in press]
6. Williams JL, Hathaway CA, Kloster KL, Layne BH. Low power, type II errors, and other statistical problems in recent cardiovascular research. *Am J Physiol* 1997; 273: H487-93.
7. Glantz SA. Biostatistics: how to detect, correct, and prevent errors in the medical literature. *Circulation* 1980; 61: 1-7.
8. Lee S, Kang H. Statistical and methodological considerations for reporting RCTs in medical literature. *Korean J Anesthesiol* 2015; 68: 106-15.