**EVIDENCE-BASED PRACTICE**

**SUZANNE M. MAHON, RN, DNSC, AOCN®, APNG**

**ASSOCIATE EDITOR**

**Reading a Research Article Part I: Types of Variables**

*Dana Oliver, MT (ASCP), MPH, and Suzanne M. Mahon, RN, DNSc, AOCN®, APNG*

Oncology nurses constantly are confronted with statistical data. Although most nurses receive some instruction on statistical process in basic preparation, and those who pursue graduate degrees receive even more, statis- tics and their interpretation remain a source of confusion and consternation for many. Nurses who choose to implement ﬁndings from a research paper need to have knowl- edge of statistical analysis to comprehend the ramiﬁcations of the ﬁndings and explain the data to patients in understandable terms. This article is the ﬁrst in a series that will examine speciﬁc issues and considerations in applied statistics to assist nurses as they read and implement the ﬁndings of research into clinical practice. This part speciﬁcally will address identiﬁcation of and issues related to the assessment of different types of variables. Future articles will describe statistical tests and issues including parametric statistics, nonparametric statistics, common statistical errors, and meta-analyses. An additional focus of this series is to acquaint readers with statistical terminology.

# Deﬁnitions

on symptom management, psychosocial adjustment, quality of life, and continuous quality improvement issues. Statistical tests can be used to describe research ﬁndings during all aspects of oncology care. This includes research related to prevention, early detection, diagnostic modalities, treatment choices, management of the side effects of disease and treatment, and a myriad of issues related to psychosocial adjustment throughout the cancer trajectory.

# Secondary Analysis

Once conclusions are formed, suggestions for treatment or nursing interventions are inferred from statistical summaries. Often, these summaries and data sets are not lim- ited to a single interpretation of the primary analysis objective. A secondary analysis of data also can be performed. In this case, the data are reanalyzed or combined with another data set to answer and assess ad- ditional questions and issues. Also, a meta- analysis may be undertaken. This type of retrospective study compares the results of a researcher’s previously published study with other published studies that have evaluated

value. One of the best ways to eliminate these problems is to consult a statistician before data collection begins, midway during data collec- tion, and at the end of data collection to ensure that the analysis is appropriate.

# Basic Considerations When Reading, Designing, or Evaluating a Research Study

Rules and assumptions are used to de- termine which of the many statistical tests would be most appropriate to use when pro- viding a summary for a researcher’s primary objective. These can be thought of as a series of questions that break an analysis into steps. The questions focus on assessment of the type of data, sample size, distribution of the data, and data variability.

# What Is the Data Type?

The ﬁrst issue to consider is the type of data being analyzed. Four groups or types of data exist: nominal, ordinal, interval, and ratio.

Statistical analysis is a process by which the efﬁcacy of the same new treatment and

data collected from or for a research project are organized to form meaningful conclu- sions. Statistics also provide a standardized means of summarizing conclusions across (or between) similar types of studies con- ducted at various sites.

# Primary Analysis

Statistics are used in the primary analysis of research data. For example, in the oncol- ogy nursing setting, studies frequently focus

are similar in design.

# Statistical Errors

Like any tool, statistics can be misused. On- cology nurses should consider and be aware of some of the more common scenarios in which this occurs. The most common situations are underpowered studies (i.e., the sample size is too small to draw reliable conclusions), selection of the wrong statistical test, and misinterpretation of the signiﬁcance of the p

*Dana Oliver, MT (ASCP), MPH, is a biostat- istician in Cancer Center Operations, and Suzanne M. Mahon, RN, DNSc, AOCN®, APNG, is an assistant clinical professor in the Division of Hematology and Oncology, both at Saint Louis University in Missouri. (Mention of speciﬁc products and opinions related to those products do not indicate or imply endorsement by the* Clinical Journal of Oncology Nursing *or the Oncology Nursing Society.)*

Digital Object Identiﬁer: 10.1188/05.CJON.110-112

A nominal variable or data type includes a variable that can be counted, such as gen- der, race, tumor type, occupation, smokers versus nonsmokers, presence or absence of toxicities, or children versus adults.

An ordinal variable describes or deﬁnes subgroups in an ordered fashion. Ordinal data types frequently are used in nursing research. A common presentation is a Likert scale or an assessment tool that stratiﬁes pain intensity on a scale of 1–10. This type of vari- able can be used to describe the severity of many different symptoms, such as nausea and vomiting or fatigue. Ordinal data also might include age groups, such as infants, children,

absolute size of the sample and whether a power analysis was calculated.

# Are Data Normally Distributed?

Ideally, data should be distributed normal- ly to use parametric statistics for analysis. However, researchers usually do not report whether a study has normal distribution of data. To some extent, readers must trust that researchers considered the distribution of data and ultimately selected the proper statistical tools to analyze data.

Demographic data often are reported in research papers. If a comparison is being made among groups, readers often may ex- amine demographic tables and observe simi- larities and dissimilarities among groups. Ideally, the distributions should be very similar to prevent bias and draw meaning- ful conclusions. Coward (2003) reported pilot data on the impact of a support group intervention during treatment for the initial diagnosis of breast cancer. That study will be used to demonstrate assessment of different types of variables and statistical methods in

teenagers, young adults, middle adults, and

**TABLE 1. STUDY PARTICIPANT CHARACTERISTICS BY GROUP**

older adults. Tumor types can be arranged as

ordinal data by stage groupings.

The third and fourth types of data are re- ferred to as interval and ratio level variables

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | **—** |  |  | **—** |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |  |  |  |  |
|  |

**EXPERIMENTAL COMPARISON**

**GROUP (N = 22) GROUP (N = 17)**

and frequently are lumped together into one category. This is because the differ- ence between the two seldom is important in statistical analysis (Norusis, 2004). The appropriate statistical test used is the same

**VARIABLE**

**Age (years)**a\* **Education (years) Months since diagnosis**

**X SD**

46.1 7.1

17.5 3.4

2.9 1.9

**X SD**

51.8 11.4

16.1 3.4

3.7 3.0

whether the data are interval or ratio. Many researchers think of these two data types more simply as continuous variables. Com-

**Religion**

**N % N %**

mon examples include actual age in years, pack-years of tobacco use, days since che- motherapy treatment, laboratory values, or years of employment as an oncology nurse. In summary, nominal and ordinal vari- ables are referred to as categorical variables, and interval and ratio variables are referred to as continuous variables. The type(s) of information or variable being studied and collected in a trial is one of the criteria used to determine which speciﬁc statistical test a researcher will be required to use when as- sessing and drawing conclusions from data

collected for a trial.

# Is the Sample Size Adequate?

Assessment of the adequacy of sample size may be one of the most challenging aspects of reading and evaluating a research study. When the number of subjects accrued to a clinical trial is small (approximately less than 30 subjects), then the power of the study might be too low to reliably summarize the information with the traditionally used meth- ods of statistical analysis (Pett, 1997). The power of a study is measured on a scale rang- ing from 0%–100%. The “rule of thumb” is to have about 80% power for a full study. Pilot studies do not require such a large sample and frequently are conducted to obtain data to conduct power analyses. In practice, most

Protestant

Catholic Jewish None

**Race**

Caucasian African American Hispanic

Asian

**Treatment**

Mastectomy and reconstructionb\* Mastectomy

Lumpectomy Radiation therapy Chemotherapy Hormone therapyb\*

**Financial status** Quite secure Comfortable Okay

Marginal Poor

**Physical health status**

Very good Good

Some disability, but doing okay

**Living arrangement**

Alone

Spouse or partner Spouse and childrenb\* Other

a t test

b chi-square

\*p < 0.05

13 59

3 14

1 5

5 23

20 91

0 0

1 5

1 5

12 55

16 73

6 27

6 27

16 73

8 36

3 14

9 41

7 32

1 5

2 9

4 18

11 50

7 32

5 23

3 14

13 59

1 5

14 82

1 6

0 0

2 12

15 88

1 6

1 6

0 0

3 18

10 59

6 35

6 35

11 65

1 6

1 6

8 47

7 41

0 0

1 6

5 29

7 41

5 29

6 35

5 29

3 18

3 18

researchers must consult a statistician to have a power analysis calculated before initiating a study. When reading a study, consider the

*Note.* From “Facilitation of Self-Transcendence in a Breast Cancer Support Group: II,” by D.D. Coward, 2003, *Oncology Nursing Forum, 30,* p. 295. Copyright 2003 by the Oncology Nursing Society. Reprinted with permission.

CLINICAL JOURNAL OF ONCOLOGY NURSING • VOLUME 9, NUMBER 1 • EVIDENCE-BASED PRACTICE **111**

nursing research. For example, if a reader examines the nominal demographic data (re- ligion, race, treatment, and living arrange- ment), ordinal demographic data (ﬁnancial

occurs, the data probably are skewed. For example, in Table 1, for the variable months since diagnosis in the comparison group, the standard deviation is 3.0, which is very

Data type

status and physical health status), and con- tinuous demographic variables (age, years

close to the mean of 3.7. This might suggest that the data are not normally distributed,

Categorical

Continuous

of education, and months since diagnosis) from the demographic table from Coward’s study on transcendence in breast cancer as

or skewed. However, most of the other standard deviations in the table are not very large when compared to the means, which

Nominal

Ordinal

Interval

Ratio

shown in Table 1, he or she will see that the experimental and comparison groups are similar in most variables. For the ordinal variables of ﬁnancial and physical health

status, the distributions are similar between

suggests that, overall, the data probably are distributed normally.

# What Is the Variability

NONPARAMETRIC

Skewed distribution

PARAMETRIC

Sample size

the experimental and control groups. This is illustrated in Table 1 when observing the physical health status variable. The percent- age of participants with some disability in the experimental group is not statistically signiﬁcantly different from the comparison group (32% versus 29%). When looking at the distribution of variables, a reader often can examine demographic data to look for large differences between groups to ensure that the study is not biased toward one group. In the Coward study, for the variable of treatment (a nominal variable), some of the groups were not similar. For the vari- able of hormone therapy, the experimental group was much larger than the comparison group (36% versus 6%). Comparing these two groups may be too difﬁcult because of the difference in the sizes of the two groups and the relatively small sample size. Overall, however, the distribution of the character- istics of the participants in the two groups in the study is fairly similar. This type of assessment can be applied to categorical types of data.

Assessment of similarity among groups of continuous variables is performed by ob- serving differences of the mean or median. The continuous variables in Table 1 include age, years of education, and number of months since diagnosis. The mean number of years of education and months since diagnosis are similar, whereas the mean age is dissimilar.

Another assessment that a reader can make quickly when looking at tables of comparative data is to consider whether the standard deviation approaches or exceeds the mean. This assessment can be applied to continuous types of data. When this

# in the Data?

Measures of variability attempt to describe or quantify the spread or range of observa- tions (Norusis, 2004). Range is the simplest measure of variability and considers the difference between the largest and smallest values. Standard deviation is a measure of how the mean or average represents the data (Field, 2000). In addition to using the standard deviation or range to assess the distribution or skewness of the data, a reader looks at these measures to determine whether the spread of the data between two groups is similar. As a general rule, the variance (standard devia- tion) of one group should not be more than twice that of the other group (Pett, 1997). In examining the data from Table 1, in particular months since diagnosis in the experimental group, multiplying the standard deviation of 1.9 by 2 equals 3.8. When comparing this number (3.8) with the standard deviation of the comparison group (3.0), 3.0 is not greater than double that of the experimental group. This estimation suggests that no signiﬁcant difference exists between the variability of the two groups being studied.

# Using the Data Type to Make Choices About Statistical Tests

After answering the questions about data type, power, distribution of data, and vari- ability of data, a researcher can make choices about which statistical test would be appro- priate to use to describe the data. Statistical tests can be divided primarily into two types: parametric tests, which are used to evaluate continuous level information, and nonpara- metric tests, which are used to evaluate cat-

Unequal variance

This ﬁgure provides an overview of how under- standing the data type (categorical or continu- ous) ultimately guides the process of selecting the appropriate statistical test.

**FIGURE 1. ALGORITHM FOR STATISTICAL TEST DECISION MAKING**

egorical information. However, a third group of statistical tests exists. They are methods that can provide a summary by mixing a categorical type of data with a continuous type of data. Figure 1 provides a schema that suggests how answering the four questions about variables guides the selection of statis- tical tests. The next article in this series will describe the proper selection of parametric statistical tests and the alternatives when the previously cited rules are not met.

**Author Contact:** Dana Oliver, MT (ASCP), MPH, can be reached at oliverda@slu.edu, with copy to editor at CJONeditor@jsobel

.com.

# References

Coward, D.D. (2003). Facilitation of self-tran- scendence in a breast cancer support group: II. *Oncology Nursing Forum, 30,* 291–300*.*

Field, A. (2000). *Discovering statistics using SPSS for Windows*. Thousand Oaks, CA: Sage.

Norusis, M.J. (2004). *SPSS 12.0 guide to data analysis*. Upper Saddle River, NJ: Prentice Hall.

Pett, M.A. (1997). *Nonparametric statistics in health care research.* Thousand Oaks, CA: Sage.

Copyright of Clinical Journal of Oncology Nursing is the property of Oncology Nursing Society and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.