Basic Statistics for the Exotic Animal Practitioner

Michelle A. Giuffrida, VMD, MSCE, DACVS

KEYWORDS

• Evidence-based practice • Evidence-based medicine • Clinical epidemiology
• Biostatistics

KEY POINTS

- Descriptive statistics are used to organize and summarize data obtained from study samples, whereas inferential statistics are used to generalize sample data to larger populations.
- The main framework for inferential statistics in medicine is hypothesis testing, which uses *P* values to judge whether observed results are likely to occur due to chance alone.
- The specific statistical methods depend on the distribution of data, but the general principles, interpretation, and limitations are consistent across hypothesis testing methods.
- Confidence intervals provide estimates of population parameters and associations along with a measure of how precise the estimate is likely to be.

Evidence-based practice (EBP) depends on clinicians' access to best research evidence, as determined through the process of critical appraisal. In veterinary medicine, the work of critical appraisal is still largely the responsibility of clinicians themselves. Clinical epidemiology guidelines such as those outlined earlier in this issue can help practitioners identify the strengths and weaknesses of different research methods. However, many studies' conclusions are based on statistical inference, so a working knowledge of basic biomedical statistics is also extremely helpful to the EBP practitioner.

The purpose of most clinical research is to explore whether causal associations exist between exposures (such as treatments, environment factors, or patient characteristic) and outcomes (such as the development, progression, or resolution of disease). Because it would be impractical to try achieve this by studying every single animal with a given exposure or outcome, most clinical research relies on data obtained from samples of subjects drawn from and presumed to represent the larger populations of similar animals. This setup means that a research study needs to do

Disclosure Statement: The author has nothing to disclose.
Small Animal Surgery, Department of Surgical and Radiological Sciences, University of California, Davis School of Veterinary Medicine, One Shields Avenue, Davis, CA 95616, USA
E-mail address: magiuffrida@ucdavis.edu

Vet Clin Exot Anim 20 (2017) 947–959
http://dx.doi.org/10.1016/j.cvex.2017.04.007 **vetexotic.theclinics.com**
1094-9194/17/Published by Elsevier Inc.

2 different things: (1) describe the data obtained from the immediate sample of animals in a useful and organized way, and (2) draw inferences from the data about cause-effect associations in the overall population. Descriptive and inferential statistics are used to accomplish these respective aims.

DESCRIPTIVE STATISTICS

Descriptive statistics refer to methods of organizing, summarizing, and displaying data. It would be hard for readers to understand what the data are showing if articles presented raw research data, especially if there are many animals, data points, or measurements. Descriptive statistics are thus very important for communicating the results of most clinical research studies, with the exception of case reports. In addition to reporting results, descriptive statistics are also used by researchers to identify data errors and to determine which inferential statistical methods are appropriate.

Exposures, outcomes, and other measured data points are referred to as variables. Variables are broadly categorized for biostatistics as being categorical or continuous (**Table 1**); these labels refer to characteristics of the data itself and help determine the appropriate type of descriptive and inferential statistics for each variable.

The distribution of a variable refers to what the data look like if a graph were made showing all the possible measurement values and how many times each value occurred in the data set. Data that have a bell-shaped curve follows the normal (Gaussian) distribution; data that take a non-bell-shaped form have a nonnormal distribution, of which there are many types[1] (**Fig. 1**). Continuous data can have either a normal or nonnormal distribution; categorical data alway take a nonnormal distribution. Although graphs are very helpful ways to summarize data, it would not be practical for journal articles to include graphs of every single variable. Instead, the distributions are summarized using measures of central tendency and variability, which are called summary statistics.

Central Tendency: Numbers that Describe the Middle of the Data Set

- Mean: the arithmetical average of the data values
  - Distorted by outliers (extreme data points)
  - Commonly used to describe continuous and ordinal data
- Median: the value above and below which half the data points lies
  - The data value at the 50th percentile
  - Not distorted by outliers
  - Commonly used to describe continuous and ordinal data
- Mode: the most frequently obtained values
  - Not distorted by outliers
  - Used to describe categorical data, or nonnormal continuous data cluster that are around 2 different points (bimodal)

Variability: Spread of the Data Around the Central Value

- Range: interval between the lowest and highest values obtained
  - Simple, only considers extreme values
  - Typically used to describe variability around the median or mode
- Interquartile range (IQR): interval between 25th and 75th percentile values
  - An estimate of spread around the median; the middle 50% of data points relative to the median
  - Recommended to describe variability of nonnormal data[2]