

# STATISTICS AND ITS ROLE IN PSYCHOLOGICAL RESEARCH

**Siu L. Chow**

*Department of Psychology, University of Regina, Canada*

**Keywords:** associated probability, conditional probability, confidence-interval estimate, correlation, descriptive statistics, deviation score, effect size, inferential statistics, random sampling distribution, regression, standard deviation, standard error, statistical power, statistical significance, sum of squares, test statistic, Type I error, Type II error

## Contents

1. Introduction
  2. Descriptive Statistics
    - 2.1. Four Levels of Measurement
    - 2.2. Data—Raw and Derived
    - 2.3. Data Tabulation and Distributions
      - 2.3.1. Simple Frequency Distribution
      - 2.3.2. Derived Distributions
      - 2.3.3. Utilities of Various Distributions
    - 2.4. Succinct Description of Data
      - 2.4.1. The Shape of the Simple Frequency Distribution
      - 2.4.2. Measures of Central Tendency
      - 2.4.3. Degrees of Freedom (*df*)
      - 2.4.4. Measures of Dispersion
    - 2.5. Standardization
    - 2.6. Correlation and Regression
      - 2.6.1. Linear Correlation
      - 2.6.2. Non-Linearity
      - 2.6.3. Linear Regression
  3. Bridging Descriptive and Inferential Statistics
    - 3.1. Random Sampling
    - 3.2. Random Sampling Distribution of Means
    - 3.3. The Random Sampling Distribution of Differences
    - 3.4. Theoretical Properties of the Sampling Distribution of Differences
  4. Inferential Statistics
    - 4.1. Experimental Hypothesis versus Statistical Hypothesis
    - 4.2. The Implication of  $H_0$
    - 4.3. The Decision Rule—Criterion and Conditional Probability
    - 4.4. The Level of Significance
    - 4.5. The Meaning of Statistical Significance
    - 4.6. “ $H_0$  Is Never True” Revisited
  5. Effect Size and Statistical Power
    - 5.1. The Putative Anomaly of Significance Tests
    - 5.2. “Effect”—Statistical versus Substantive
    - 5.3. Statistical Power
    - 5.4. Reservations about Statistical Power
- Glossary

Bibliography  
Biographical Sketch

## Summary

As readers will have noticed, some everyday words are given technical meanings in statistical parlance (e.g. “mean,” “normal,” “significance,” “effect,” and “power”). It is necessary to resist the temptation of conflating their vernacular and technical meanings. A failure to do so may have a lot to do with the ready acceptance of the “effect size” and “power” arguments in recent years.

To recapitulate, statistics is used (i) to describe succinctly data in terms of the shape, central tendency, and dispersion of their simple frequency distribution, and (ii) to make decisions about the properties of the statistical populations on the basis of sample statistics. Statistical decisions are made with reference to a body of theoretical distributions: the distributions of various test statistics that are in turn derived from the appropriate sample statistics. In every case, the calculated test statistic is compared to the theoretical distribution, which is made up of an infinite number of tokens of the test statistic in question. Hence, the “in the long run” caveat should be made explicit in every probabilistic statement based on inferential statistics (e.g. “the result is significant at the 0.05 level in the long run”).

Despite the recent movement to discourage psychologists from conducting significance tests, significance tests can be (and ought to be) defended by (i) clarifying some concepts, (ii) examining the role of statistics in empirical research, and (iii) showing that the sampling distribution of the test statistic is both the bridge between descriptive and inferential statistics and the probability foundation of significance tests.

## 1. Introduction

Statistics, as a branch of applied mathematics, consists of univariate and multivariate procedures. Psychologists use univariate procedures when they measure only one variable; they use multivariate procedures when multiple variables are used (a) to ascertain the relationship between two or more variables, (b) to derive the test statistic, or (c) to extract factors (or latent variables). As multivariate statistics is introduced in *The Construction and Use of Psychological Tests and Measures*, this article is almost exclusively about univariate statistics. The exception is the topic of linear correlation and regression.

The distinction needs to be made before proceeding between the substantive population and the statistical population. Suppose that an experiment is carried out to study the effects of diet supplements on athletic performance. The substantive population consists of all athletes. The sample selected from the substantive population is divided into two sub-samples. The experimental sub-sample receives the prescribed diet supplements and the control sub-sample receives a placebo.

In this experimental context, the two groups are not samples of the substantive population, “all athletes.” Instead, they are samples of two statistical populations

defined by the experimental manipulation “athletes given diet supplements” and “athletes given the placebo.” In general terms, even if there is only one substantive population in an empirical study, there are as many statistical populations as there are data-collection conditions. This has the following five implications.

First, statistics deal with methodologically defined statistical populations. Second, statistical conclusions are about data in their capacity to represent the statistical populations, not about substantive issues. Third, apart from very exceptional cases, research data (however numerous) are treated as sample data. Fourth, testing the statistical hypothesis is not corroborating the substantive theory. Fifth, data owe their substantive meanings to the theoretical foundation of the research (for the three embedding conditional syllogisms, see *Experimentation in Psychology—Rationale, Concepts, and Issues*).

Henceforth, “population” and “sample” refer to statistical population and statistical sample, respectively. A parameter is a property of the population, whereas a statistic is a characteristic of the sample. A test statistic (e.g. the student-*t*) is an index derived from the sample statistic. The test statistic is used to make a statistical decision about the population.

In terms of utility, statistics is divided into descriptive and inferential statistics. Psychologists use descriptive statistics to describe research data succinctly. The sample statistic (e.g. the sample mean,  $\bar{X}$ ) thus obtained is used to derive the test statistic (e.g. the student-*t*) that features in inferential statistics. This is made possible by virtue of the “random sampling distribution” of the sample statistic. Inferential statistics consists of procedures used for (a) drawing conclusions about a population parameter on the basis of a sample statistic, and (b) testing statistical hypotheses.

## **2. Descriptive Statistics**

To measure something is to assign numerical values to observations according to some well-defined rules. The rules give rise to data at four levels: categorical, ordinal, interval, or ratio. A preliminary step in statistical analysis is to organize the data in terms of the research design. Psychologists use descriptive statistics to transform and describe succinctly their data in either tabular or graphical form. These procedures provide the summary indices used in further analyses.

### **2.1. Four Levels of Measurement**

Using numbers to designate or categorize observation units is measurement at the nominal or categorical level. An example is the number on the bus that signifies its route. Apart from counting, nominal data are amenable to no other statistical procedure. An example of ordinal data is the result of ranking or rating research participants in terms of some quality (e.g. their enthusiasm). The interval between two successive ranks (or ratings) is indeterminate. Consequently, the difference between any two consecutive ranks (e.g. Ranks 1 and 2) may not be the same as that between another pair of consecutive ranks (e.g. Ranks 2 and 3).

Temperature is an example of the interval-scale measurement. The size of two successive intervals is constant. For example, the difference between 20°C and 30°C is the same as that between 10°C and 20°C. However, owing to the fact that 0°C does not mean the complete absence of heat (i.e. there is no absolute zero in the Celsius scale), it is not possible to say that 30°C is twice as warm as 15°C.

In addition to having a constant difference between two successive intervals, it is possible to make a definite statement about the ratio between two distances by virtue of the fact that 0 m means no distance. Hence, a distance of 4 km is twice as far as 2 km because of the absolute zero in the variable, *distance*. Measurements that have properties like those of *distance* are ratio data.

**2.2. Data—Raw and Derived**

Suppose that subjects are given 60 minutes to solve as many anagram problems as possible. The scores thus obtained are raw scores when they are not changed numerically in any way. In a slightly different data collection situation, the subjects may be allowed as much time as they need. Their data may be converted into the average number of problems solved in a 30-minute period or the average amount of time required to solve a problem. That is, derived data may be obtained by applying an appropriate arithmetic operation to the raw scores so as to render more meaningful the research data.

**2.3. Data Tabulation and Distributions**

Data organization is guided by considering the best way (i) to describe the entire set of data without enumerating them individually, (ii) to compare any score to the rest of the scores, (iii) to determine the probability of obtaining a score with a particular value, (iv) to ascertain the probability of obtaining a score within or outside a specified range of values, (v) to represent the data graphically, and (vi) to describe the graphical representation thus obtained.

**2.3.1. Simple Frequency Distribution**

The entries in panel 1 of Table 1 represent the performance of 25 individuals. This method of presentation becomes impracticable if scores are more numerous. Moreover, it is not conducive to carrying out the six objectives just mentioned. Hence, the data are described in a more useful way by (a) identifying the various distinct scores (the “Score” row in panel 2), and (b) counting the number of times each score occurs (i.e. the “Frequency” row in panel 2). This way of representing the data is the tabular “simple frequency distribution” (or “frequency distribution” for short).

**Panel 1: A complete enumeration of all the scores**

15	14	14	13	13	13	12	12	12	12
11	11	11	11	11	10	10	10	10	9
9	9	8	8	7					

**Panel 2: The simple frequency distribution**

Score	15	14	13	12	11	10	9	8	7
Frequency	1	2	3	4	5	4	3	2	1

**Panel 3: Distributions derived from the simple frequency distribution**

1	2	3	4	5	6
Score value	Frequency	Cumulative frequency	Cumulative percentage	Relative frequency	Cumulative relative frequency
15	1	25	100	0.04	1.00
14	2	24	96	0.08	0.96
13	3	22	88	0.12	0.88
12	4	19	76	0.16	0.76
11	5	15	60	0.20	0.60
10	4	10	40	0.16	0.40
9	3	6	24	0.12	0.24
8	2	3	12	0.08	0.12
7	1	1	4	0.04	0.04
	Total = 25				

Table 1. Various ways of tabulating data

**2.3.2. Derived Distributions**

The frequency distributions tabulated in panel 2 of Table 1 have been represented in columns 1 and 2 of panel 3. This is used to derive other useful distributions: (a) the cumulative percentage distribution (column 3), (b) the cumulative percentage (column 4), (c) the relative frequency (probability) distribution (column 4), and (d) the cumulative probability distribution (column 6).

Cumulative frequencies are obtained by answering the question “How many scores equal or are smaller than  $X$ ?” where  $X$  assumes every value in ascending order of numerical magnitude. For example, when  $X$  is 8, the answer is 3 (i.e. the sum of 1 plus 2) because there is one occurrence of 7 and two occurrences of 8. A cumulative percentage is obtained when 100 multiply a cumulative relative frequency.

A score’s frequency is transformed into its corresponding relative frequency when the total number of scores divides the frequency. As relative frequency is probability, the entries in column 5 are the respective probabilities of occurrence of the scores. Relative frequencies may be cumulated in the same way as are the frequencies. The results are the cumulative probabilities.

### 2.3.3. Utilities of Various Distributions

Psychologists derive various distributions from the simple frequency distribution to answer different questions. For example, the simple frequency distribution is used to determine the shape of the distribution (see *Section 2.4.1. The Shape of the Simple Frequency Distribution*). The cumulative percentage distribution makes it easy to determine the standing of a score relative to the rest of the scores. For example, it can be seen from column 3 in panel 3 of Table 1 that 22 out of 25 scores have a value equal to or smaller than 13. Similarly, column 4 shows that a score of 13 equals, or is better than, 88% of the scores (see column 5).

The relative frequencies make it easy to determine readily what probability or proportion of times a particular score may occur (e.g. the probability of getting a score of 12 is 0.16 from column 5). Likewise, it is easily seen that the probability of getting a score between 9 and 12, inclusive, is 0.64 (i.e.  $0.12 + 0.16 + 0.20 + 0.16$ ). The cumulative probability distribution in column 6 is used to answer the following questions:

- (a) What is the probability of getting a score whose value is  $X$  or larger?
- (b) What is the probability of getting a score whose value is  $X$  or smaller?
- (c) What are  $X_1$  and  $X_2$  such that they include 95% of all scores?

The probability in (a) or (b) is the associated probability of  $X$ . In like manner, psychologists answer questions about the associated probability of the test statistic with a cumulative probability distribution at a higher level of abstraction (see *Section 3.2. Random Sampling Distribution of Means*). The ability to do so is the very ability required in making statistical decisions about chance influences or using many of the statistical tables.

## 2.4. Succinct Description of Data

Research data are described succinctly by reporting three properties of their simple frequency distribution: its shape, central tendency, and dispersion (or variability).

### 2.4.1. The Shape of the Simple Frequency Distribution

The shape of the simple frequency distribution depicted by columns 1 and 2 in panel 3 of Table 1 is seen when the frequency distribution is represented graphically in the form of a histogram (Figure 1a) or a polygon (Figure 1b). Columns 1 and 6 jointly depict the cumulative probability distribution whose shape is shown in Figure 1c. In all cases, the score-values are shown on the X or horizontal axis, whereas the frequency of occurrence of a score-value is represented the Y or vertical axis in Figures 1a and 1b; the vertical axis of Figure 1c represents the cumulative probability.

A frequency distribution may be normal or non-normal in shape. The characterization “normal” in this context does not have any clinical connotation. It refers to the properties of being symmetrical and looking like a bell, as well as having two tails that extend to positive and negative infinities without touching the X axis. Any distribution that does not have these features is a non-normal distribution.

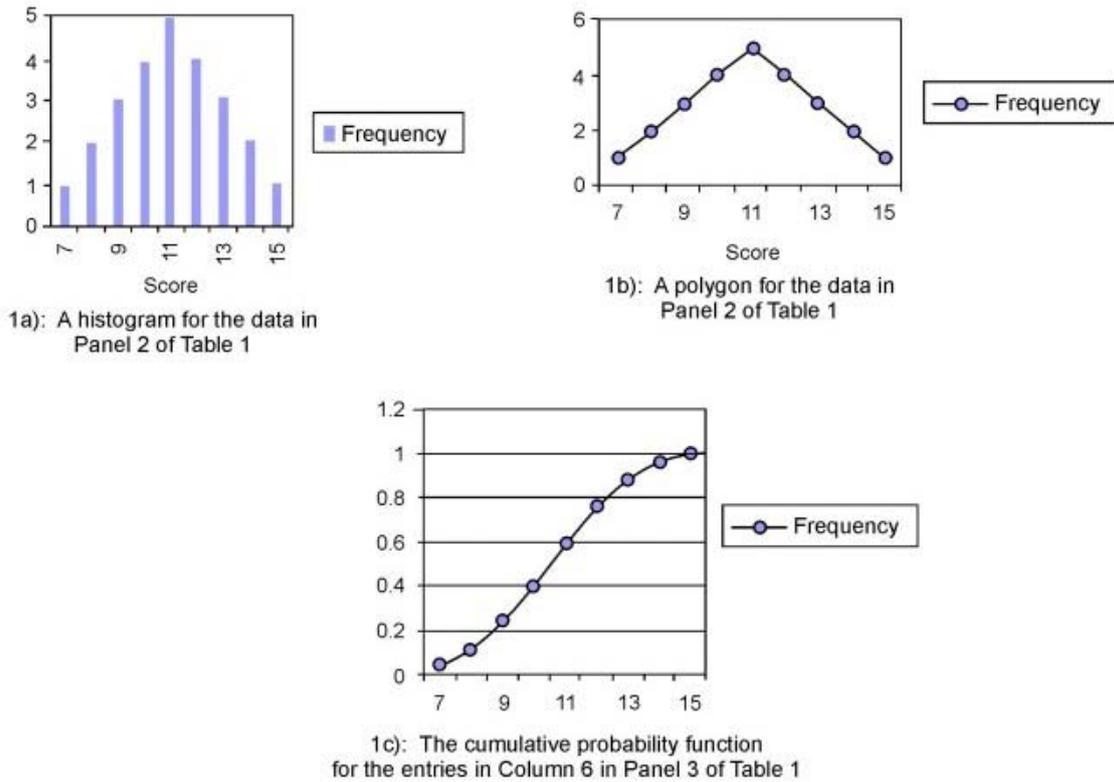


Figure 1. Graphical representations of the simple frequency (a & b) and cumulative probability distributions (c)

TO ACCESS ALL THE 25 PAGES OF THIS CHAPTER,  
 Visit: <http://www.eolss.net/Eolss-sampleAllChapter.aspx>

**Bibliography**

Chow S.L. (1998). A précis of “Statistical Significance: Rationale, Validity and Utility.” *Behavioral and Brain Sciences* **21**, 169–194. [This open-peer review consists of a précis of Chow’s (1996) *Statistical Significance: Rationale, Validity and Utility*; 39 commentaries on the book, and Chow’s reply to the commentaries. The set is a good place to review the arguments for and against significance tests.]

Cohen J. (1987). *Statistical Power Analysis for the Behavioral Sciences*, rev. edn. New York: Academic Press. [This book provides much of the impetus for the adoption of power analysis.]

Meehl P.E. (1967). Theory testing in psychology and physics: A methodological paradox. *Philosophy of Science* **34**, 103–115. [This is one of the first articles that remind psychologists not to conflate the statistical hypothesis with the substantive hypothesis. By implication, readers will learn that making the statistical hypothesis is not corroborating the substantive hypothesis.]

Siegel S. (1956). *Nonparametric Statistics for the Behavioral Sciences*, 312 pp. New York: McGraw-Hill. [This book is a classic introduction to nonparametric procedures.]

Wilkinson L. and Task Force on Statistical Inference, APA Board of Scientific Affairs. (1999). Statistical methods in psychology journals: Guidelines and explanations. *American Psychologist* **54**(8), 594–604. [This report is an advocacy document, in which a set of research agenda is set forth. Despite the Task Force's disavowal, much of what the report says is about research methods in general, and psychologists' putative departure from good research practice in particular. The report sets in high relief the importance of the philosophical and methodological issues.]

Winer B.J. (1962). *Statistical Principles in Experimental Design*, 672 pp. New York: McGraw-Hill. [This book is a classic introduction to analysis of variance.]

### **Biographical Sketch**

**Siu L. Chow** obtained his B.A. Honours from the University of Adelaide, Sydney, and Ph.D. from the University of Toronto, Canada. He is currently professor of psychology at the University of Regina, Canada.