Contents

1 Overview and Descriptive Statistics **2**
- 1.1 Populations and Samples 2
- 1.2 Pictorial and Tabular Methods in Descriptive Statistics 2
- 1.3 Measures of Location 3
- 1.4 Measures of Variability 3

2 Probability **4**
- 2.1 Sample Spaces and Events 4
- 2.2 Axioms, Interpretations, and Properties of Probability 4
- 2.3 Counting Techniques 5
- 2.4 Conditional Probability 5
- 2.5 Independence 6

3 Discrete Random Variables and Probability Distributions **7**
- 3.1 Random Variables 7
- 3.2 Probability Distributions for Discrete Random Variables 7
- 3.3 Expected Values of Discrete Random Variables 8
- 3.4 Moments and Moment Generating Functions 9
- 3.5 The Binomial Probability Distribution 9
- 3.6 Hypergeometric and Negative Binomial Distributions 9
- 3.7 The Poisson Probability Distribution 9

Chapter 1

Overview and Descriptive Statistics

1.1 Populations and Samples

1.2 Pictorial and Tabular Methods in Descriptive Statistics

Procedure 1 (Steps for constructing a stem-and-leaf display)**.**

1. Select one or more leading digits for the stem values. The trailing digits becomes the leaves.
2. List possible stem values in a vertical column.
3. Record the leaf for every observation beside the corresponding stem value.
4. Order the leaves from smallest to largest on each line.
5. Indicate the units for stems and ;eaves someplace in the display.

Procedure 2 (A histogram for counting data)**.** First, determine the frequency and relative frequency of each x value. Then mark possible x values on a horizontal scale. Above each calue, draw a rectangle whose height is the relative frequency (or alternatively, the frequency of that value).

Procedure 3 (A histogram for measurement data: equal class widths)**.** Determine the frequency and relative frequency for each class. Mark the class boundaries on a horizontal measurement axis. Above each class interval, draw a rectangle whose height is the corresponding relative frequence (or frequency).

Procedure 4 (A histogram for measurement data: unequal class widths)**.** After determining frequencies and relative frequencies, calculate the height of each rectangle using the formula

$$\text{rectangle height} = \frac{\text{relative frequency of the class}}{\text{class width}}$$

The resulting rectangle heights are usually called *densities*, and the vertical scale is the **density scale**. This prescription will also work when class widths are equal.

1.3 Measures of Location

Definition 1. The **sample mean** $\bar{x}$ of observations $x_1, x_2, \ldots, x_n$ is given by

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}. \tag{1.1}$$

The numerator of $\bar{x}$ can be written more informally as $\sum x_i$ where the summation is over all sample observations.

Definition 2. The **sample mean** is obtained by first ordering the n observations from smallest to largest (with any repeated values included so that every sample observation appears in the ordered list). Then,

$$\bar{x} = \begin{cases} \text{The single middle value if } n \text{ is odd} \\ \text{The average of the two middle values if } n \text{ is even} \end{cases}$$

1.4 Measures of Variability

Definition 3. The **sample variance**, denoted by s^2, is given by

$$s^2 = \frac{\left(\sum (x_i - \bar{x})\right)^2}{n-1} = \frac{S_{xx}}{n_1}$$

The **sample standard deviation**, denoted by s, is the (positive) square root of the variance:

$$s = \sqrt{s^2}$$

Proposition 1. *Let $x_1, x_2, \ldots, x_n$ be a sample and c be a constant.*

1. *If $y_1 = x_1 + c, y_2 = x_2 + c, \ldots, y_n = x_n + c$, then $s_y^2 = s_x^2$, and*
2. *If $y_1 = cx_1, \ldots, y_n = cx_n$, then $s_y^2 = c^2 s_x^2, sy = |c| s_x$,*

where s_x^2 is the sample variance of the x's and s_y^2 is the sample variance of the y's.

Definition 4. Order the n observations from smallest to largest and separate the samllest half from the largest half; the median $\bar{x}$ is included in both halves if n is odd. Then the **lower fourth** is the median of the smallest half and the **upper fourth** is the median of the largest half. A measure of spread that is resistant to outliers is the **fourth spread** f_s, given by

$$f_s = \text{upper fourth - lower fourth}$$

Definition 5. Any observation farthen than $1.5 f_s$ from the closts fourth is an **outlier**. An outlier is **extreme** if it is more than $3 f_s$ from the nearest fourth, and it is **mild** otherwise.

Chapter 2

Probability

2.1 Sample Spaces and Events

Definition 6. The **sample space** of an experiment, denoted by $\mathcal{S}$, is the set of all possible outcomes of that experiment.

Definition 7. An **event** is any collection (subset) of outcomes contained in the sample space $\mathcal{S}$. An event is said to be **simple** if it consists of exactly one outcome and **compound** if it consists of more than one outcome.

Definition 8. Some relations from set theory:

1. The **union** of two events A and B, denoted by $A \cup B$ and read "A or B" is the event consisting of all outcomes that are *either in A or in B or in both events* (so that the union includes outcomes for which both A and B occur as well as the outcomes for which exactly one occurs) - that is, all outcomes in at least one of the events.

2. The **intersection** of two events A and B, denoted by $A \cap B$ and read " *A and B* " is the event consisting of all outcomes that are in *both A and B*.

3. The **complement** of an event A, denoted by A', is the set of all outcomes in $\mathcal{S}$ that are not contained in A.

Definition 9. When A and B have no outcomes in common, they are said to be **disjoint** or **mutually exclusive** events. Mathematicians write this compactly as $A \cup B = \emptyset$ where $\emptyset$ denotes the event consisting of no outcomes whatsoever (the "null" or "empty" event).

2.2 Axioms, Interpretations, and Properties of Probability

Definition 10. Basic properties of probability:

1. For any event $A, P(A) \geq 0$.

2. $P(S) = 1$.

3. If $A_1, A_2, A_3, \cdots$ is an infinite collection of disjoint events, then

$$P(A_1 \cup A_2 \cup \cdots) = \sum_{i=1}^{\infty} P(A_i).$$

Proposition 2. *$P(\emptyset) = 0$ where $\emptyset$ is the null event. This in turn implies that the property contained in Axiom 3 is valid for a finite collection of events.*[1]

Proposition 3. *For any event $A, P(A) = 1 - P(A')$.*[2]

Proposition 4. *For any event $A, P(A) \leq 1$.*

Proposition 5. *For any events A and B,*

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B).^3 \\ P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) \\ &\quad - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \end{aligned}$$

2.3 Counting Techniques

Proposition 6. *If the first element or object of an ordered pair can be selected in n_1 ways, and for each of these n_1 ways the second element of the pair can be selected in n_2 ways, then the number of pairs is $n_1 n_2$.*

Proposition 7 (Product rule for K-tuples)**.** *Suppose a set consists of ordered collections of k elements (k-tuples) and that there are n_1 possible choices for the first element; for each choice of the first element, there are n_2 possible choices of the second element;. . . ; for each possible of the first $k - 1$ elements, there are n_k choices for kth element. Then there are $n_1 n_2 \cdot \cdots \cdot n_k$ possible k-tuples.*

Definition 11. Any ordered sequence k objects taken from a set of n distinct objects is called a **permutation** of size k of the objects. The numbers of permutations of size k that can be constructed from the n objects is denoted $P_{k,n}$.

Definition 12. For any positive integer m, $m!$ is read "m factorial" and is defined by $m! = m(m-1)(m-2) \ldots (2)(1)$. Also, $0! = 1$.

Definition 13. Given a set of n distinct objects, any unordered subset of size k of the objects is colled a **combination**. The number of combinations of size k that can be formed from n will be denoted by $\binom{n}{k}$. This notation is more common in probability than $C_{k,n}$, which would be analogous to notation for permutations.

2.4 Conditional Probability

Definition 14. For any two events A and B with $P(B) > 0$, the **conditional probability of A given B that has occured** is defined by

$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

[1]Proof page 57.
[2]Proof page 60.

Proposition 8 (The multiplication rule). $P(A \cap B) = P(A \mid B) \cdot P(B)$

Proposition 9 (The law of total probability). *Let $A_1, \ldots, A_k$ be mutually exclusive and exhaustive events. Then for any other event B.*

$$\begin{aligned} P(B) &= P(B \mid A_1) \cdot P(A_1) + \cdots + P(B \mid A_k) \cdot P(A_k) \\ &= \sum_{i=1}^{k} P(B \mid A_i) P(A_i) \end{aligned}$$

Proposition 10 (Bayes' Theorem). *Let $A_1, \ldots, A_k$ be a collection of mutually exclusive and exhaustive events with $P(A_i) > 0$ for $i = 0, \ldots, k$. Then for any other event B, for which $P(B) > 0$*

$$P(A_j \mid B) = \frac{P(A_j \cap B)}{P(B)} = \frac{P(B \mid A_j) P(A_j)}{\sum_{i=1}^{k} P(B \mid A_i) P(A_i)} \qquad j = 0, \ldots, k$$

2.5 Independence

Definition 15. Two events A and B are **independent** if $P(A \mid B) = P(A)$ and are **dependend** otherwise.

Proposition 11. *A and B are independent if and only if*

$$P(A \cap B) = P(A) \cdot P(B)$$

Definition 16. Events $A_1, \ldots, A_n$ are **mutually independent** if for every $k (k = 2, 3, \ldots, n)$ and every subset of indices $i_1, i_2, \ldots, i_k$,

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1}) \cdot P(A_{i_2}) \cdot \cdots \cdot P(A_{i_k}) \tag{2.1}$$

Chapter 3

Discrete Random Variables and Probability Distributions

3.1 Random Variables

Definition 17. For a given sample space $\mathcal{S}$ of some experiment, a **random variable (rv)** if any rule that associates a number with each outcome in $\mathcal{S}$. In mathematical language, a random variable is a function whose domain is the sample space and whose range is the set of real numbers.

Definition 18. Any random variable whose only possible values are 0 and 1 is called a **Bernoulli random variable**.

Definition 19. A **discrete** random variable is an rv whose possible values either constitute a finite set or else can be listed in an infinite sequence in which there is a first element, a second element, and so on. A random variable is **continuous** if *both* the following apply:

1. Its set of possible values consists either of all numbers in a single interval on the number line (possibly infinite in extent, e.g., from $-\infty$ to ∞) or all numbers in a disjoint union of such intervals (e.g., $[0, 10] \cup [20, 30]$).
2. No possible value of the variable has positive probability, that is, $P(X = c) = 0$ for any possible value c.

3.2 Probability Distributions for Discrete Random Variables

Definition 20. The **probability distribution** or **probability mass function** (pmf) of a discrete rv is defined for every number x by $p(x) = P(X = x) = P(\text{all} s \in \mathcal{S} : X(s) = x.$[1]

[1] $P(X = x)$ is read "the probability that the rv X assumes the value x". For example, $P(X = 2)$ denotes the probability that the resulting X value is 2.

Definition 21. Suppose $p(x)$ depends on a qunatity that can be assigned any one a number of possible values, with each different value determining a different probability distribution. Such a quantity is called a **parameter** of the distribution. The collection of all probability distributions for different values of the parameter is called a **family** of probability distributions.

Definition 22. The **cumulative distribution function** (cdf) $F(x)$ of a discrete rv X with pmf $p(x)$ is defined for every number x by

$$F(X) = P(X \leq x) = \sum_{y:y \leq x} p(y)$$

For any number $x, F(x)$ is the probability that the observed value of X will be at most x.

Proposition 12. *For any two numbers a and b with $a \leq b$,*

$$P(a \leq X \leq b) = F(b) - F(a-)$$

where $F(a-)$ represents the maximum of $F(x)$ values to the left of a. Equivalently, if a is the limit of values of x approaching from the left, then $F(a-)$ is the limiting value of $F(x)$. In particular, if the only possible values are integers and if a and b are integers, then

$$\begin{aligned} P(a \leq X \leq b) &= P(X = a \text{ or } a+1 \text{ or } \dots \text{ or } b) \\ &= F(b) - F(a-1) \end{aligned}$$

Taking $a = b$ yields $P(X = a) = F(a) - F(a-1)$ in this case.

3.3 Expected Values of Discrete Random Variables

Definition 23. Let X be a discrete rv with set of possible values D and pmf $p(x)$. The **expected value** or **mean value** of X, denoted by $E(X)$ or μ_X, is

$$E(X) = \mu_X = \sum_{x \in D} x \cdot p(x)$$

This expected value will exist provided that $\sum_{x \in D} |x| \cdot p(x) < \infty$

Proposition 13. *If the rv X has a set of possible values D and pmf $p(x)$, then the expected value of any function $h(X)$, denoted by $E[h(X)]$ or $\mu_{h(X)}$, is computed by*

$$E[h(X)] = \sum_D h(x) \cdot p(x)$$

assuming that $\sum_D |h(x)| \cdot p(x)$ is finite.

Proposition 14.

$$E(aX + b) = a \cdot E(X) + b$$

(Or, using alternative notation, $\mu_{aX+b} = a \cdot \mu_X + b$)

3.4 Moments and Moment Generating Functions

3.5 The Binomial Probability Distribution

3.6 Hypergeometric and Negative Binomial Distributions

3.7 The Poisson Probability Distribution