Basic Terms

Shane Hutton, Ph.D.

Vanderbilt University

Basic Terms

- Variable: any characteristic, behavior, or condition that can occur in differing amounts or kinds
- Independent variable (IV): a variable that we either manipulate or observe to determine its effects on the dependent variable; the “cause”
  - Other terms sometimes used for independent variable: factor, predictor, X variable
- Dependent variable (DV): the variable that is affected by changes in the independent variable; the “outcome”
  - Other terms sometimes used for dependent variable: criterion, outcome, Y variable
  - DV depends on IV, IV affects (or does not affect) DV

Basic Terms

- IV/DV examples
  - The older you are, the more difficulty you have learning a second language.
    - IV: age, DV: ability to learn second language
  - Having a bad relationship with your parents does not cause homosexuality.
    - IV: relationship with parents, DV: sexual orientation
  - Final grade in a course is related to class attendance.
    - IV: class attendance, DV: grade

Basic Terms

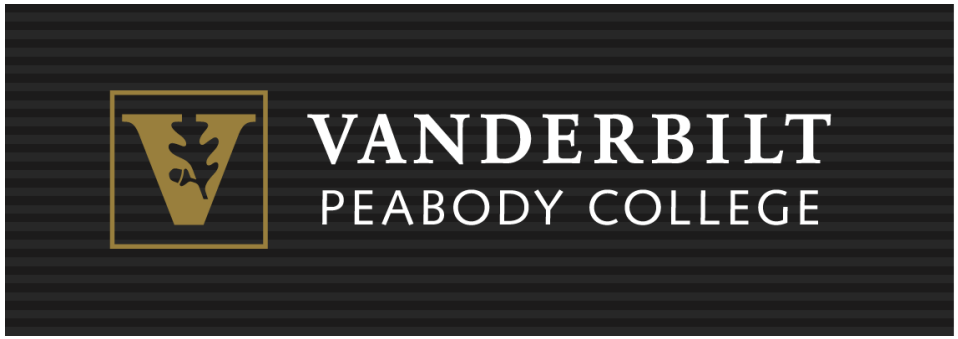
- Research methods:
  - Experimental: manipulate IV and measure DV
  - Correlational: no manipulation, measure both IV and DV as is
- Data: measurements collected in research; often scores, but may be ranks, categories, or types

Basic Terms

- Population: the entire group of interest
  - Example: college students
  - Need to carefully define the population
    - Students at Vanderbilt?
    - Students in Tennessee?
    - Students in all American colleges?
- Sample: subset of population; the group that you actually study and from which you collect data
  - The sample must be random and representative

Basic Terms

- Statistics: a set of mathematical procedures and techniques used to make sense of data
  - Descriptive statistics: used to organize, summarize, and simplify data
  - Inferential statistics: used to make generalizations about the population from the sample data


VANDERBILT
PEABODY COLLEGE

Classifying Variables

Shane Hutton, Ph.D.

Vanderbilt University

Continuous Variable

- Continuous variable/data: There are an infinite number of potential values between two adjacent values; they can take on a full range of values, including numbers out to several decimal places; fractional values are meaningful amounts of the variable.
  - Examples: age, height, GPA

Continuous Variable

- Important note: Individual scores from most psychological tests, inventories, and questionnaires are usually reported as a whole value. However, these scores are treated as continuous.
  - Examples: Beck Depression Inventory (scale of 0 – 63), Likert rating scales, such as from 1 to 7

Discrete Variable

- Discrete variable/data: Only includes whole values or categories with no fractional amounts between them.
  - Examples: marital status, letter grades, type of treatment

Quantitative/Qualitative

- Quantitative: varies by amount
  - Examples: number of errors, weight
- Qualitative: varies by form or class
  - Examples: seasons, letter grades

Classifying Variables

TABLE 1.3 A List of 20 Variables Showing How They Fit Into the Three Categories That Describe Them

Variables	Continuous vs. Discrete	Qualitative vs. Quantitative	Scale of Measurement
Sex (male, female)	Discrete	Qualitative	Nominal
Seasons (spring, summer, fall, winter)	Discrete	Qualitative	Nominal
Number of dreams recalled	Discrete	Quantitative	Ratio
Number of errors	Discrete	Quantitative	Ratio
Duration of drug abuse (in years)	Continuous	Quantitative	Ratio
Ranking of favorite foods	Discrete	Quantitative	Ordinal
Ratings of satisfaction (1 to 7)	Discrete	Quantitative	Interval
Body type (slim, average, heavy)	Discrete	Qualitative	Nominal
Score (from 0% to 100%) on an exam	Continuous	Quantitative	Ratio
Number of students in your class	Discrete	Quantitative	Ratio
Temperature (degrees Fahrenheit)	Continuous	Quantitative	Interval
Time (in seconds) to memorize a list	Continuous	Quantitative	Ratio
The size of a reward (in grams)	Continuous	Quantitative	Ratio
Position standing in line	Discrete	Quantitative	Ordinal
Political affiliation (Republican, Democrat)	Discrete	Qualitative	Nominal
Type of distraction (auditory, visual)	Discrete	Qualitative	Nominal
A letter grade (A, B, C, D, F)	Discrete	Qualitative	Ordinal
Weight (in pounds) of an infant	Continuous	Quantitative	Ratio
A college student's SAT score	Discrete	Quantitative	Interval
Number of lever presses per minute	Discrete	Quantitative	Ratio

Scales of Measurement

- There are four scales of measurement
  - Nominal
  - Ordinal
  - Interval
  - Ratio

Nominal

- Nominal: names; they label observations but do not provide quantitative information
  - Examples: political affiliation (Democrat, Republican, Independent); geographical region (Northeast, Midwest, Southeast, Southwest)

Ordinal

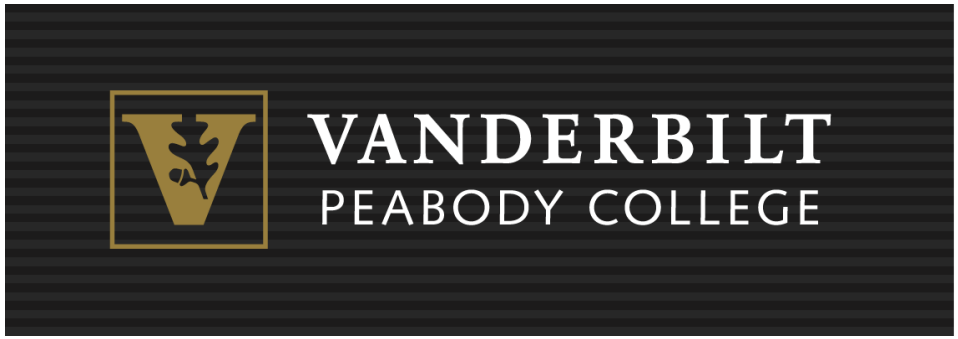
- Ordinal: rank order; measurements provide some information about relative amount or magnitude; indicate more or less but not exactly how much; the exact amount of difference between adjacent values cannot be assumed to be equal across the scale
  - Examples: order of finishing a task (1^{st}, 2^{nd}, 3^{rd}, etc.); socioeconomic status (low, middle, upper); spiciness of hot sauce (mild, medium, hot, extremely hot); attitude rating scale of 1 to 4

Interval

- Interval: a series of scores that are organized in an ordered sequence with equal intervals between values on the measurement scale; no meaningful zero point
  - Examples: temperature in Fahrenheit; SAT scores

Ratio

- Ratio: an interval scale with an meaningful zero point
  - The most common ratio variables are distance, size, time, and frequency of occurrence
  - Examples: dose of a medication in milligrams; reaction time; number of correct answers on a test; number of times a rat presses a lever to receive food
- Many statistical computer programs refer to both interval and ratio variables as scale variables because both interval observations and ratio observations are analyzed with the same statistical tests.


VANDERBILT
PEABODY COLLEGE

Question Review: Classifying Variables

Shane Hutton, Ph.D.

Vanderbilt University

Question Review

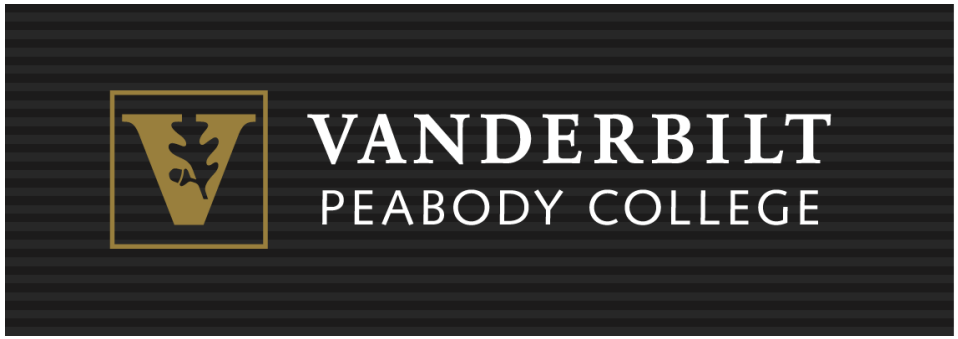
A researcher wants to know if crowding affects college students' performance on a math test. 100 Vanderbilt students are randomly selected. 50 students take a test in a 400-seat auditorium with students filling every seat in the room (crowded condition). The other 50 students take the same test in the auditorium with no other students present (uncrowded condition). Tests are scored for accuracy (possible range of scores is 0 – 100).

Question Review

- Crowding is the independent variable and it is discrete.
- Math performance is the dependent variable and it is continuous.

Question Review

- Number of hours worked per week in paid employment
  - Ratio
- College major
  - Nominal
- Perceived difficulty of college major rated on a scale from 1 to 5
  - Interval
- Employment status: unemployed, part-time, full-time
  - Ordinal


VANDERBILT
PEABODY COLLEGE

Constructing Frequency Distributions

Shane Hutton, Ph.D.

Vanderbilt University

Frequency Distributions

- A frequency distribution is an organization of data to show the number of times each score occurs in the sample.
- A frequency distribution may be displayed in a table or a graph.
- A frequency distribution may be ungrouped or grouped.

Frequency Distributions

1. Ungrouped frequency distribution:
   - List each individual score or category on the scale in order from lowest to highest or highest to lowest.
   - For each score or category, indicate the frequency, or the number of times it occurs in the data set.
2. Grouped frequency distribution
   - Scores or values are grouped into intervals.
   - For each interval, indicate the frequency or number of scores.

Ungrouped Frequency Distribution

Ungrouped frequency distribution

1. List each individual score.
2. Tally the number of times each score occurs in the data set.
3. List the frequency of each score

Example:

Suppose we asked students to respond to the question “How good are you at math? [Responses from 1(terrible) to 7 (excellent)]”

6	5	6	1	4	4	5	6	5
3	5	6	7	3	5	5	4	5
4	7	2	6	5	5	2	6	5
5	4	5	3	5	7	6	5	3

Ungrouped Frequency Distribution

Score	f
7	3
6	6
5	14
4	5
3	4
2	2
1	1

Adding More Information

- Sample size: N = $\sum f$ = the sum of all frequencies = the total number of scores in the sample
- Proportion: the relative frequency of a score in the data set = p = $\frac{f}{N}$
- Percentage: the relative frequency expressed as a percentage = proportion(100) = $\frac{f}{N}$(100)
- Cumulative frequency = cf = sum of frequencies at and below that category or value
- Cumulative percentage = $\frac{cf}{N}$(100) = % of scores at and below that category or value

Adding More Information

Score	f	p	%	cf	cf%
7	3	.09	9%	35	100%
6	6	.17	17%	32	91%
5	14	.40	40%	26	74%
4	5	.14	14%	12	34%
3	4	.11	11%	7	20%
2	2	.06	6%	3	9%
1	1	.03	3%	1	3%

Grouped Frequency Distribution

Grouped frequency distribution

- Frequencies are displayed for ranges of scores rather than for individual values

Example:

How much money (in cash) do you have with you right now?

.05	6	30	0	65	4.75	50	0	100	.17	15	210
2	175	42	65	2	200	0	10	0	15	20	30
12	0	30	45	7	10	80	12	28	.75	100	33
1	14	5	23.45								

Grouped Frequency Distribution

Steps for constructing a grouped frequency distribution

- Step 1: Construct the interval

  a. Determine the interval width

  - Depends in part on the nature of the data, variable
  - Usually want about 10 intervals
  - Calculate estimated interval width = (highest score – lowest score)/(# of intervals)

Grouped Frequency Distribution

Steps for constructing a grouped frequency distribution

- Step 1: Construct the interval (cont.)

  b. Write the lower and upper limits for each interval

  - Often, lowest score is the lower limit of the first interval
  - Upper limit = lower limit + width – 1 unit
  - Make sure every interval is equal in width
  - List enough intervals to include all scores
  - Don’t start or end with an empty interval (f = 0)

- Step 2: List the frequencies, record the number of scores in each interval

Grouped Frequency Distribution

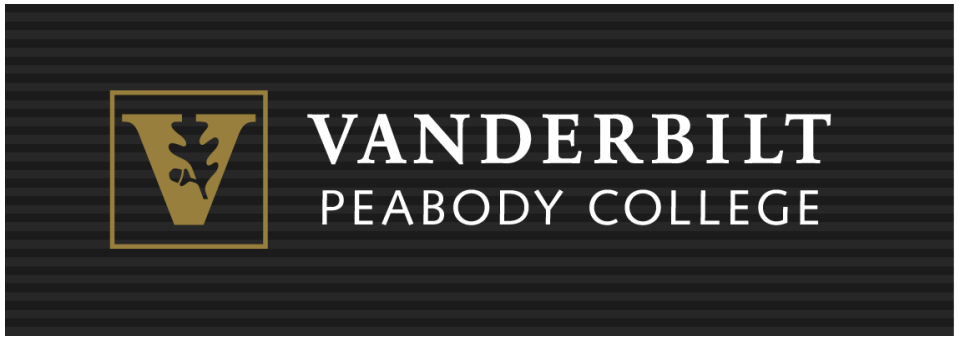
Example:

How much money (in cash) do you have with you right now?

.05	6	30	0	65	4.75	50	0	100	.17	15	210
2	175	42	65	2	200	0	10	0	15	20	30
12	0	30	45	7	10	80	12	28	.75	100	33
1	14	5	23.45								

Grouped Frequency Distribution

Class interval	Exact limits	f	%	cf	c%
200-219	199.5-219.5	2	5%	40	100%
180-199	179.5-199.49	0	0%	38	95%
160-179	159.5-179.5	1	2.5%	38	95%
140-159	139.5-159.5	0	0%	37	92.5%
120-139	119.5-139.5	0	0%	37	92.5%
100-119	99.5-119.5	2	5%	37	92.5%
80-99	79.5-99.5	1	2.5%	35	87.5%
60-79	59.5-79.5	2	5%	34	85%
40-59	39.5-59.5	3	7.5%	32	80%
20-39	19.5-39.5	7	17.5%	29	72.5%
0-19	0-19.5	22	55%	22	55%


VANDERBILT
PEABODY COLLEGE

Graphing Frequency Distributions

Shane Hutton, Ph.D.

Vanderbilt University

Graphing Frequency Distributions

- Features of graphs of frequency distributions
  - Vertical axis (Y-axis) is used to display frequency
  - Horizontal axis (X-axis) is used to display scores/categories of the variable
  - Each axis is labeled appropriately

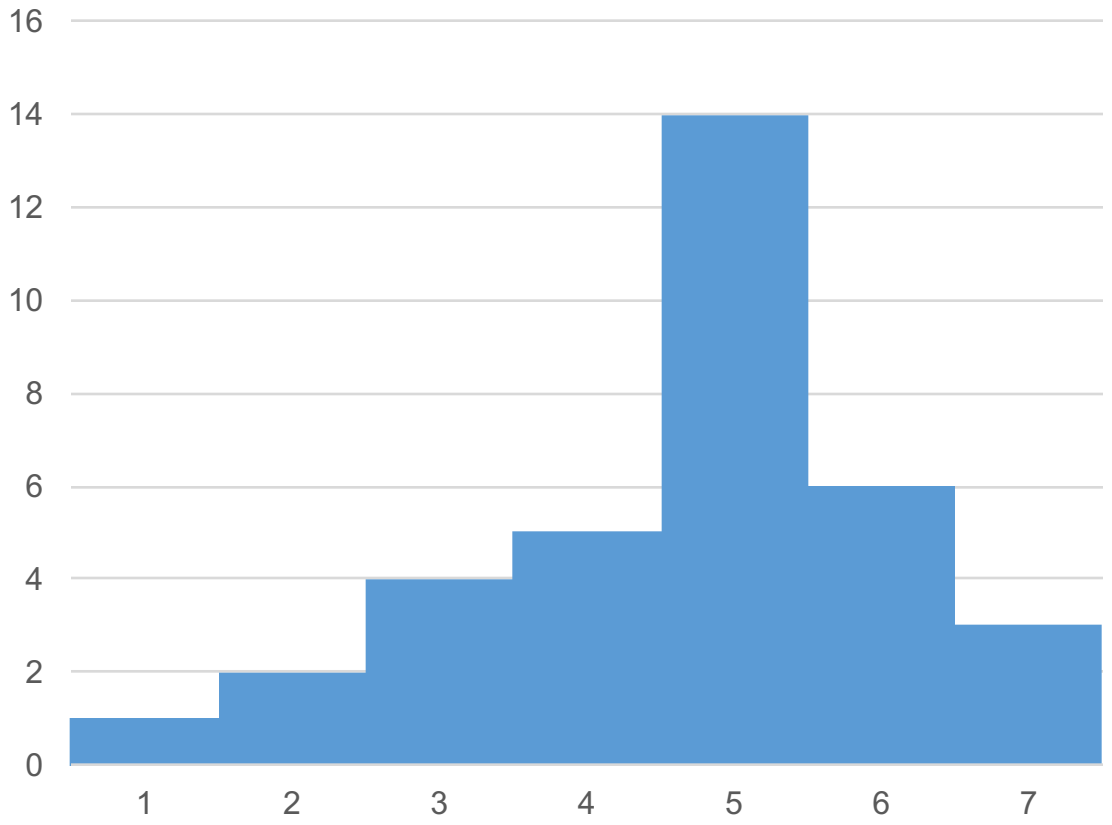
Graphing Frequency Distributions

- Types of graphs for continuous data
  - Frequency histogram: vertical bar centered over each score or score interval on X axis and adjacent bars touch
  - Frequency polygon: point centered over each score or score interval on X axis and points are connected by straight lines

Graphing Frequency Distributions

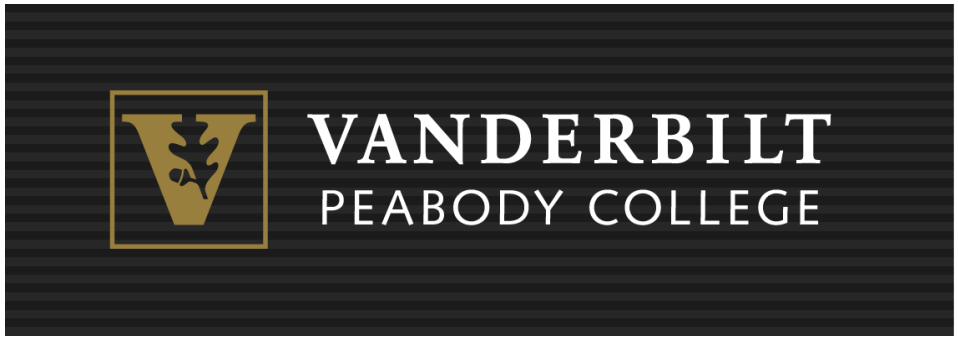
Example: Self-rated math ability

Score	f
7	3
6	6
5	14
4	5
3	4
2	2
1	1



Graphing Frequency Distributions

Class interval	f
200-219	2
180-199	0
160-179	1
140-159	0
120-139	0
100-119	2
80-99	1
60-79	2
40-59	3
20-39	7
0-19	22


VANDERBILT
PEABODY COLLEGE

Describing the Shape of a Frequency Distribution

Shane Hutton, Ph.D.

Vanderbilt University

Describing Frequency Distributions

- Symmetry
  - Symmetrical: each half of the distribution is the mirror image of the other half

Describing Frequency Distributions

- Skewness: asymmetrical with a few extreme values on one side of the distribution
  - Positively skewed: many relatively low values with a few extremely high values

Describing Frequency Distributions

- Skewness: asymmetrical with a few extreme values on one side of the distribution
  - Negatively skewed: many relatively high values with a few extremely low values

Describing Frequency Distributions

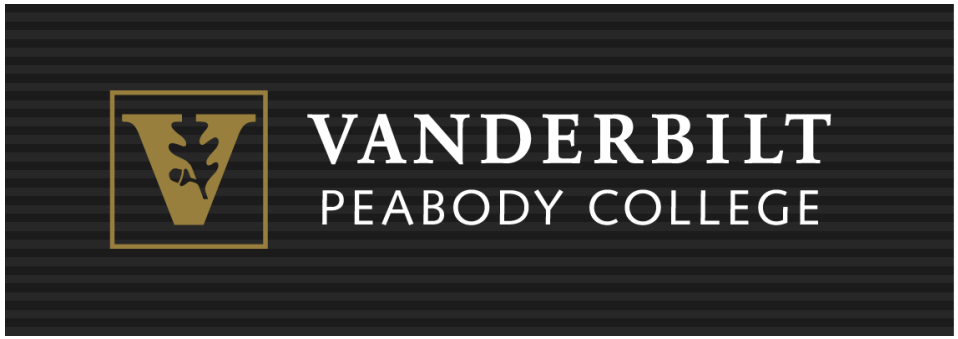
- Bimodal: having two distinct peaks

Describing Frequency Distributions

- Kurtosis: flatness or peakedness of the distribution
  - Platykurtic: flat

Describing Frequency Distributions

- Kurtosis: flatness or peakedness of the distribution
  - Leptokurtic: sharply peaked


VANDERBILT
PEABODY COLLEGE

Graphs for Discrete Data

Shane Hutton, Ph.D.

Vanderbilt University

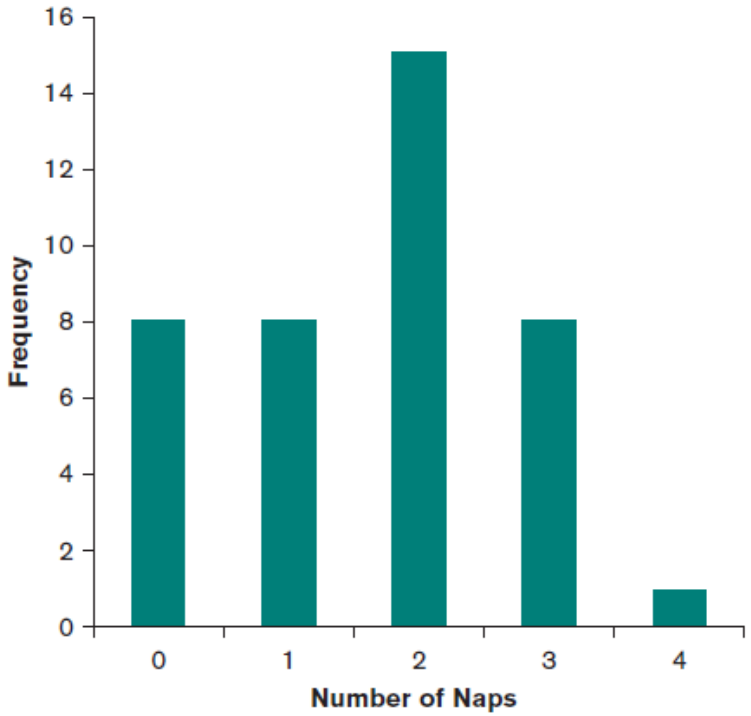
Graphs for Discrete Data

- There are two types of graphs routinely used to display discrete or categorical data
  - Bar chart
  - Pie chart

Bar Chart

FIGURE 2.11 A Bar Chart

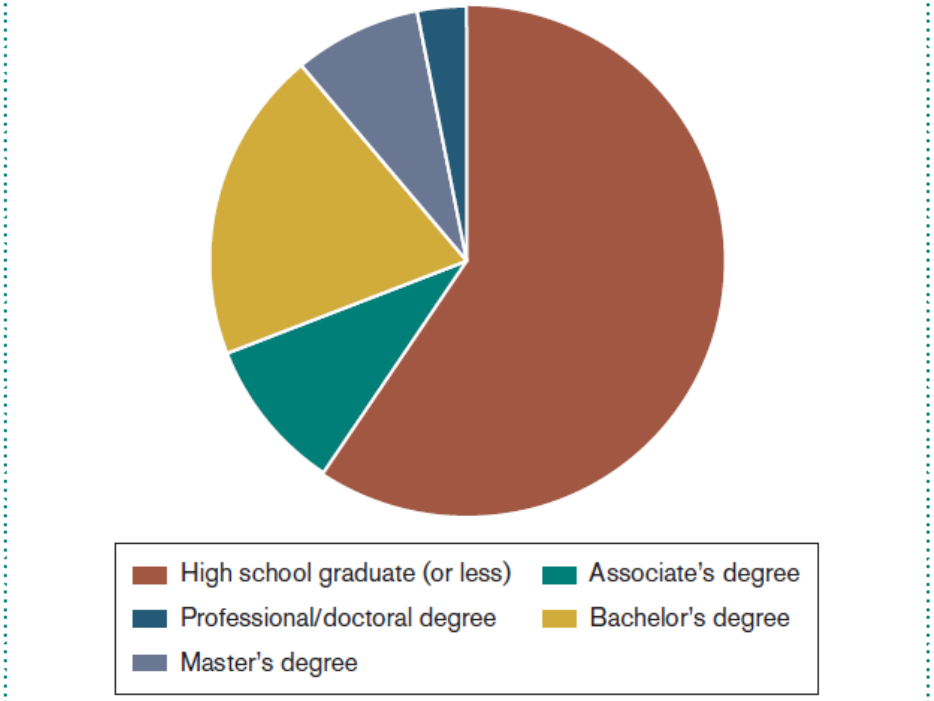
Number of Naps	$f(x)$
0	8
1	8
2	15
3	8
4	1



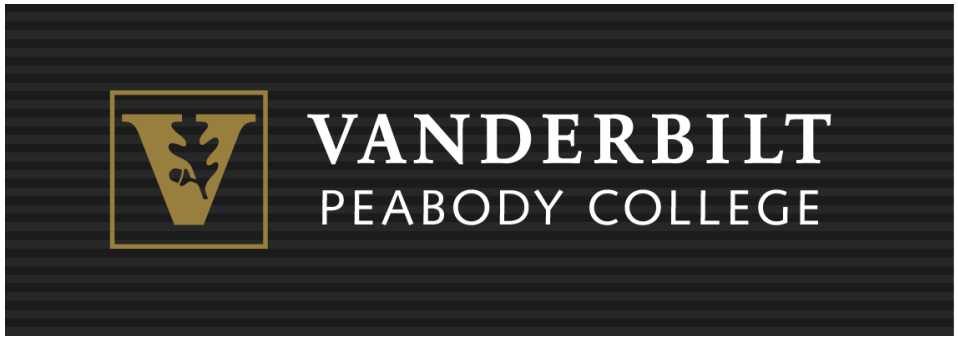
A frequency table (left) and a bar chart (right) summarizing the average number of naps per day that mothers give their children, who are younger than age 3.

Pie Chart

FIGURE 2.12 A Pie Chart for the Distribution of Educational Attainment of the Population 25 Years and Older in the United States, 2012



Source: Table created based on data from the U.S. Census Bureau, Current Population Survey, 2012 Annual Social and Economic Supplement.


VANDERBILT
PEABODY COLLEGE