Math 140
Introductory Statistics

Professor Silvia Fernández

Chapter 6
Based on the book *Statistics in Action*
by A. Watkins, R. Scheaffer, and G. Cobb.

6.1 Probability Distribution from Data

We have three ways of specifying a population:

- 1. List of all (individual) units
- 2. Frequency Table (p. 68)
- 3. Relative Frequency or
  Proportion Table (p. 359)

How can we calculate mean and SD on each?

List of all units

Number	Type	Value x	x - μ
1	Penny	1 ¢	-3
2	Penny	1 ¢	-3
3	Penny	1 ¢	-3
4	Penny	1 ¢	-3
5	Penny	1 ¢	-3
6	Nickel	5 ¢	1
7	Nickel	5 ¢	1
8	Nickel	5 ¢	1
9	Dime	10 ¢	6
10	Dime	10 ¢	6
	Total = 10 coins	Sum = 40 cents	

$$\mu = \text{population mean} = \frac{\sum x}{n}$$

$$\mu = \frac{1+1+1+1+1+5+5+5+10+10}{10} = 4$$

$$\sigma_n = \text{SD} = \sqrt{\frac{\sum (x-\mu)^2}{n}}$$

$$\sigma_n = \sqrt{\frac{9+9+9+9+9+1+1+1+36+36}{10}} =$$

$$\sigma_n = \sqrt{\frac{120}{10}} = \sqrt{12} \approx 3.4641$$

Frequency Tables

Type	Value x	Frequency f	$x \cdot f$
Penny	1 ¢	5	5
Nickel	5 ¢	3	15
Dime	10 ¢	2	20
Sum		10	40

$$n = \sum f = 10$$

$$\mu = \text{population mean} = \frac{\sum x \cdot f}{n}$$

$$\mu = \frac{5 + 15 + 20}{10} = 4$$

$$\sigma_n = \text{SD} = \sqrt{\frac{\sum (x - \mu)^2 \cdot f}{n}}$$

$$\sigma_n = \sqrt{\frac{9 \cdot 5 + 1 \cdot 3 + 36 \cdot 2}{10}} =$$

$$\sigma_n = \sqrt{\frac{120}{10}} = \sqrt{12} \approx 3.4641$$

Relative Frequency or Proportion Table

Type	Value x	Proportion of coins $P(x)$	$x \cdot P(x)$
Penny	1 ¢	0.5	0.5
Nickel	5 ¢	0.3	1.5
Dime	10 ¢	0.2	2.0
Sum		1	4.0

$$\mu = \text{population mean} = \sum x \cdot P(x)$$

$$\mu = 0.5 + 1.5 + 2.0 = 4$$

$$\sigma_n = \text{SD} = \sqrt{\sum (x-\mu)^2 \cdot P(x)}$$

$$\sigma_n = \sqrt{9 \cdot (0.5) + 1 \cdot (0.3) + 36 \cdot (0.2)} =$$

$$\sigma_n = \sqrt{4.5 + 0.3 + 7.2} = \sqrt{12} \approx 3.4641$$

Summary of Mean/SD

List of all units

$$\mu = \frac{\sum x}{n}$$

$$\sigma_n = \sqrt{\frac{\sum (x-\mu)^2}{n}}$$

Frequency Table

$$\mu = \frac{\sum x \cdot f}{n}$$

$$\sigma_n = \sqrt{\frac{\sum (x-\mu)^2 \cdot f}{n}}$$

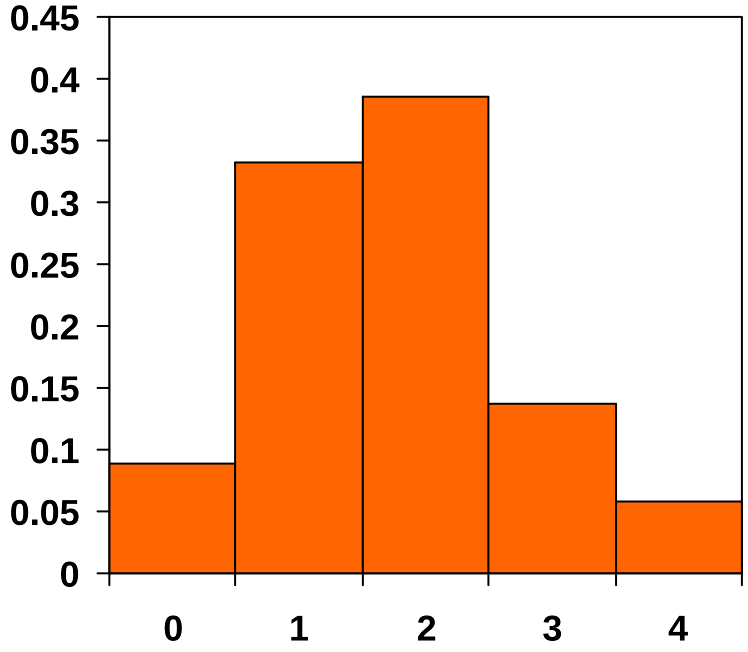
Relative Frequency (or Proportion) Table

$$\mu = \sum x \cdot P(x)$$

$$\sigma_n = \sqrt{\sum (x-\mu)^2 \cdot P(x)}$$

Example (page 359)

Number of Motor Vehicles (per household), x	Proportion of households, $P(x)$
0	0.088
1	0.332
2	0.385
3	0.137
4	0.058



The number of motor vehicles per household. The “4” represents “4 or more,” but the proportion of households with more than four vehicles is very small.

[Source: U.S. Census Bureau, American Community Survey, 2004, factfinder.census.gov.]

How to Sample

We have three ways of specifying a population:

- 1. List of all (individual) units
- 2. Frequency Table (p. 68)
- 3. Relative Frequency or
  Proportion Table (p. 359)

How can we get (or simulate) a sample from these distributions?

How to Sample

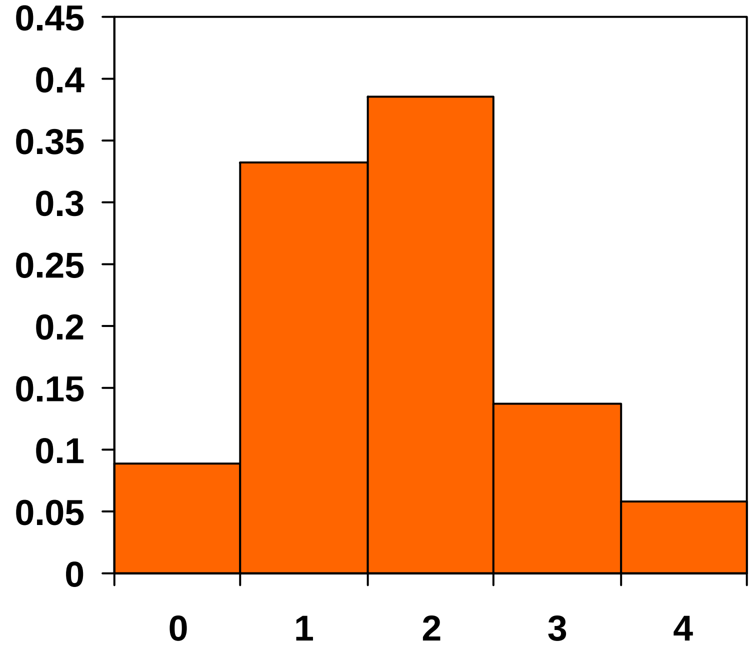
- List of all Units and Frequency Tables.
  - 1. Make a numbered list of all units. If using a frequency table remember to repeat each value x according to its **multiplicity**.
  - 2. Draw as many numbers at random as needed.
- Relative Frequency (or Proportion) Tables
  - 1. Assign a number to each proportion in the table, using as many digits as the decimal precision of the proportions.
  - 2. Select as many of these numbers at random as needed.
  - 3. Use the assignation to realize the characteristics of the values in the sample.

Ways to generate random numbers

- By using **`rand`** or **`randint`** in your calculator (Preferred way)
- By using a string or a table of random digits (page 828)
- By writing numbers in slips of paper, mixing, and drawing at random.

Example (page 359)

Number of Motor Vehicles (per household), x	Proportion of households, $P(x)$
0	0.088
1	0.332
2	0.385
3	0.137
4	0.058



The number of motor vehicles per household. The “4” represents “4 or more,” but the proportion of households with more than four vehicles is very small.

[Source: U.S. Census Bureau, American Community Survey, 2004, factfinder.census.gov.]

Example (part 1)

Number of Motor Vehicles (per household), x	Proportion of households, $P(x)$	Random Numbers Representing this Category
0	0.088	$001\text{-}088$
1	0.332	$089\text{-}420$
2	0.385	$421\text{-}805$
3	0.137	$806\text{-}942$
4	0.058	$943\text{-}999,000$

420 = 88 + 332

805 = 420 + 385

942 = 805 + 137

1000 = 942 + 58

Note: If you are using a random-digit table then you should use 000 as the equivalent of 1000

Example (part 2)

Number of Motor Vehicles (per household), x	Proportion of households, $P(x)$	Random Numbers Representing this Category
0	0.088	$001\text{-}088$
1	0.332	$089\text{-}420$
2	0.385	$421\text{-}805$
3	0.137	$806\text{-}942$
4	0.058	$943\text{-}999$ 000

- Suppose we have the following string of random digits (already separated by groups of 3)

391 | 545 | 177 | 981 | 016 | 845 | 248

- The corresponding number of vehicles per household in this sample are:

Random Number	Vehicles in household
391	1
545	2
177	1
981	4

Sampling with or without replacement

- Without replacement.
  - Usual choice in practice.
  - Do not return slips of paper to the box.
  - Disregard repetitions when using tables of random digits or random numbers from a calculator.

Note: The two possibilities yield almost the same results as long as the sample size is small compared to the population size.

- With replacement
  - Seldom used in practice.
  - However it is much easier to do calculations in this case. (see 5.3, 5.4)
  - Do return slips of paper to box.
  - Allow repetitions of numbers when using tables of random digits or random numbers from a calculator.