

## Defining Our Terms: The Analog Grip

In our first lab, we mastered the machine. Now, we must master the mind. We are stepping away from the keyboard to practice the “Analog Grip.” Statistics is not just about computing; it is about defining the boundaries of the world before we measure it.

**Bill the Statypus says:** Finally. We have returned to the pinnacle of research technology: the Dixon Ticonderoga No. 2. A sharp pencil and a ...

**Sally the Statypus says:** (Gently interjecting) Thanks, Bill! I will take it from here. I’ll be your primary guide today. Before we ever touch a keyboard again, we have to know exactly *who* we are talking about and *what* we are measuring. Pick up your pencil. Let’s go!

### 1. Section 2.1: Foundations (The “Who”)

#### Action Item: Manual Reading

Read section 2.1 of `r.statypus.org` before moving forward.

#### Action Item: The Nesting Dolls

Navigate to the “Now it’s your turn!” block immediately following **Example 2.1** in the textbook. Record your answers to these questions below.

#### Statypus Insight: The Approximation Bridge

Remember how these units relate: an **individual** is a member of a **sample**, which is a subset of the **population**.

Because of this nesting relationship, the logic of inference follows: just as a *sample* approximates a *population*, a *statistic* (what we calculate from a sample) approximates a *parameter* (calculated from a population).

## 2. The Census Myth and the Statistical Reality

In a perfect world, we would perform a **Census**—a study where we measure every single individual in the entire population. If we did this, we would know the absolute truth. There would be no guessing.

**Bill the Statypus says:** A census is the pinnacle of architectural purity. It is a complete enumeration of the population. But unless you have an infinite budget, a frozen population, and a fleet of researchers who never sleep, a census is a logistical pipe dream. In the real world, the truth is often hidden behind the impossible scale of the population.

**Sally the Statypus says:** Exactly! People move, people graduate, and some people just don't answer their phones. Since we can't talk to everyone, we have to take a representative slice. Because we can't have the 'Whole,' we are forced to rely on 'Statistics.' (Pun intended, Bill!)

### Action Item: The Census Challenge

Think about our study of the **average cost to attend SU**. Even at a single university, why is a true census nearly impossible to achieve? Using your pencil, list three specific logistical reasons why we can't just get the "perfect" number from every single student.

1. \_\_\_\_\_
2. \_\_\_\_\_
3. \_\_\_\_\_

### Statypus Insight: The Statistician's Burden

Because a census is usually impossible, we are forced to use a **Sample** to estimate the truth of the **Population**.

This is the entire reason the field of Statistics exists. We take the information we *can* get (a statistic) and use it to make a high-level guess about the information we *can't* get (the parameter).

**Bill the Statypus says:** (Still chuckling) That was a good one, Sally!

### 3. Qualitative vs. Quantitative: Identity vs. Amount

#### Action Item: Manual Reading

Read Section 2.2 up until the start of 2.2.3 of `r.statypus.org` before moving forward.

Before we can do math, we have to know if math even makes sense for the variable. We divide our data into two broad categories:

- **Qualitative:** Variables that describe a quality or category. These are labels. Math (like subtraction) is meaningless here.
- **Quantitative:** Variables that describe a quantity or amount. Arithmetic tells a story with these numbers.

#### Statypus Insight: The ID Trap (Section 2.2.2)

Look at **Example 2.6** in the text. The variable `orig_id` is made of integers. However, if you subtract one baby's ID from another, does the result mean anything? No. Because the arithmetic is meaningless, ID numbers are **Qualitative**.

#### Action Item: The ID Test

Explain why the `intake date` in the *Asylum1849* dataset is considered **Quantitative** even though dates often look like labels. (Hint: What happens if you subtract one date from another?)

#### Reflection: Our Own Variability

List six variables that could be collected about you. Make sure you have three that are qualitative and three that are quantitative.

## 4. Discrete vs. Continuous: The Measured and the Counted

### Action Item: Manual Reading

Read the rest of Section 2.2 of [r.statypus.org](http://r.statypus.org) before moving forward.

Once we know a variable is Quantitative, we have to determine its “grain.”

- **Discrete:** Results of **counting**. There are “gaps” between the values (e.g., 1 egg, 2 eggs).
- **Continuous:** Results of **measuring**. These can take on any value in an interval (e.g., 1.427 inches).

**Bill the Statypus says:** In pure mathematics, a continuous variable has infinite precision. But look at my pencil. No matter how sharp I make this Ticonderoga, the mark it leaves has a physical width. It covers a range. Our research tools are rarely as infinite as the math.

### Statypus Insight: The 2.2.4 Recording Issue

In **Section 2.2.4**, we admit that the world is messy. Age is naturally **Continuous** (time flows and doesn’t ‘tick’ like a clock), but we almost always record it as a **Discrete** integer (“I am 21”). This is called **discretized data**.

### Action Item: The Precision Grid

Classify the following variables. In the “Why” column, note if the variable is counted, measured, or if it is a continuous variable being “discretized” (recorded as an integer).

Variable	Disc.	Cont.	Why?
Number of eggs in a nest			
Exact weight of a gold coin			
Height (recorded in cm)			
Number of pages in a book			
Total time spent waiting			

### Statypus Insight: The Table is Set

Now that we have agreed upon definitions, we can finally move on to actually thinking about doing statistics.