

The Logic of the Gap: Mapping the Difference

1. The Observation: Where is the Gap?

In this block, we move from looking at a single population to looking at the **Difference** between two groups. Before we ask if a difference is “significant,” we must first visualize where it lives.

Lead Observer’s Big Question

If there were actually **NO difference** between these groups in the real world, how often would we see a gap this large just by “random chance”?

2. Independent Data: The Two-Hill Logic

The Setup: You measure 50 *Iris setosa* and 50 *Iris versicolor*.

- **The Relationship:** Row 1 of Group A has no connection to Row 1 of Group B. They are separate individuals.

Bill the Statypus says: “I couldn’t find a base R function that properly displayed these concepts, and honestly, I was too busy to actually write and debug a proper function from scratch. So, I just hacked together `DifferenceLogic.r` to show you the daylight between the hills. It’s not pretty, but it works.”

Task 1: The Independent Gap. Run the first section of the R script. Refer to the plot and sketch the two overlapping hills below. Mark the horizontal distance between their peaks as your “Observed Gap.”

Calculation: The Two-Hill Sketch

Sketch the Independent Hills here based on the R plot.

3. Paired Data: The Single-Hill Logic

The Setup: You measure the *exact same 50 plants* in the Morning and again in the Evening.

- **The Relationship:** Row 1 in the Morning is the **exact same plant** as Row 1 in the Evening.
- **The Mapping:** We subtract the pairs *first* ($d = x_{pm} - x_{am}$) to create a single “Hill of Change.”

Bill the Statypus says: “When data is paired, we don’t care about the two hills anymore. We only care about the one hill of differences. Run the second part of the script to see it.”

Task 2: The Paired Change. Run the second section of `DifferenceLogic.r`. Sketch the single distribution centered at 0. Mark where your calculated “Change” sits on the x-axis.

Calculation: The Hill of Change

Sketch the Paired Change distribution here based on the R plot.

