Part 1: The Notation Audit — r vs. ρ

*We use different symbols to distinguish between what we **calculate** from a small group and what we **infer** about the entire world.*

- **Point Estimate (r):** The correlation of our 96 toy cars.
- **Parameter (ρ):** The "true" correlation of the entire population.

In your own words, why is it dangerous to assume that our calculated r is exactly equal to the population ρ?

Part 2: The Shuffle Test

To investigate ρ, we will use a specific R function to simulate alternate realities.

1. **The Setup:** Run the code below and record the 95% Confidence Interval for ρ.

Coding Corner: Toolbox Initialization

```
load(url("https://statypus.org/files/StatypusCh12.RData"))
rhoConfInt(ToyCars$Height, ToyCars$Distance)
```

95% CI for ρ: [_____ , _____]

2. **Visualizing the Shuffle:** Sketch the histogram produced by R below (the default 10,000 shuffles). Mark the location of your original sample r on the horizontal axis.

Part 3: The "Low-Tech" Logic of Power

The process you just ran is a brute-force method called **Bootstrapping**. We don't rely on complex theoretical formulas or calculus; we simply use the computer's power to simulate thousands of alternate realities by "shuffling" our existing data.

Think of it as the ultimate "low-tech" solution: if we want to know how much our r might change in a different sample, we just have the computer build 10,000 new samples using the 96 cars we already have. In this view, our original r is just one of the 10,000 possible samples the computer found. By **picking 96 cars at random** (with replacement) over and over, we see how much the value of r naturally moves during the "shuffle."

1. **The Low-Power Test:** What happens if the computer isn't working as hard? Run the command again, but force R to only do 50 shuffles.

   Coding Corner: Simulation Control

   ```
   rhoConfInt(ToyCars$Height, ToyCars$Distance, size = 50)
   ```

2. **Sketching the "Bumpy" Reality:** Sketch the new 50-shuffle histogram in the space below.

3. **The Shuffle Shift:** How has the **shape** changed compared to your 10,000-shuffle sketch on Page 1? What does this tell you about using a small number of simulations?

Part 4: Reflection — The Proxy Population

1. **The Skew of Reality:** When the data was collected, the author and his children started to get a bit impatient and ran fewer trials as the height increased. This created a higher density of cars at lower heights. Look at your bootstrap histograms. How does that physical impatience during data collection show up in the shape of your computer-generated results?

2. **The Mechanism of the Shuffle:** Bootstrapping relies on *resampling with replacement*. If we took 10,000 samples but **did not** use replacement—meaning we just re-shuffled the same 96 cars without ever duplicating a data point—would the r value change at all from sample to sample?

Reality Check: Making an Educated Guess

Look at your original 10,000-shuffle Confidence Interval. If someone claimed there was actually **zero** relationship between Height and Distance ($\rho = 0$), take a guess based *only* on your bootstrap results: do you think we have enough evidence to say they are wrong? Why or why not?