1.2 The Classical Scientific Method and Statistical Inference

“The whole of science is nothing more than a refinement of everyday thinking.”

- Albert Einstein

Population of units

Random Variable $X$

$Hypothesis$ (about $X$)

**EXPERIMENT**

“What actually happens this time, regardless of hypothesis.”

**THEORY**

“What ideally must follow, if hypothesis is true.”

Random Sample (empirical data)

$n = \# \text{ observations}$

- $x_1$
- $x_2$  \ldots  $x_n$
- $x_3$

**Mathematical Theorem** (formal proof)

Proof: If $Hypothesis$ (about $X$), then $Conclusion$ (about $X$).

QED

**Decision:** Accept or Reject $Hypothesis$

Analysis: Observed vs. Expected, under $Hypothesis$

“Is the difference statistically significant? Or just due to random, chance variation alone?”
Example:

Population of individuals

Hypothesis: “The prevalence (proportion) of a certain disease is 10%.”

Decision: 
Reject Hypothesis

Based on our sample, the prevalence of this disease in the population is significantly higher than 10%, around 12%.

Theory
“What ideally must follow, if hypothesis is true.”

Experiment
“What actually happens this time, regardless of hypothesis.”

Random Sample
(empirical data)

\[ n = 2500 \text{ individuals} \]

- Yes/No
- Yes/No
- Yes/No

Mathematical Theorem
(formal proof)

Suppose random variable \( X = \# \text{ Yes} = 300 \), i.e., estimated prevalence = \( \frac{300}{2500} = 0.12 \), or 12%.

If Hypothesis of 10% prevalence is true, then the “expected value” of \( X \) would be 250 out of a random sample of 2500.

Moreover, under these conditions, it can (and later will) be mathematically proved that the probability of obtaining a sample result that is as, or more, extreme than 12%, is only .00043 (the “p-value”), or less than one-twentieth of one percent. EXTREMELY RARE!!! Thus, our sample evidence is indeed statistically significant; it tends to strongly refute the original Hypothesis.