Stats 95

- **Kepler tried to record the paths of planets in the sky, Harvey to measure the flow of blood in the circulatory system, and chemists tried to produce pure gold knowing it was an element, though they failed at it. What was there approach to science, and what was the new perspective Pearson made, and what was the one crucial distinction he failed to make.** *Before Pearson, the things that science dealt with were real and palpable. Pearson proposed that the observable phenomena were only random reflections, what was real was the probability distribution. The real things of science were not things that we could observe but mathematical functions that described the randomness of what was observed. The four parameters of the distribution were what we were really interested in, but in the end we can never really determine them, we can only estimate them, and Pearson failed to understand that they could only be estimated.*

What was the nature of research like before experimental methods became established? *They were idiosyncratic to each scientist, lesser scientists would produce vast amounts of data would be accumulated but would not advance knowledge, such as in the inconclusive attempts to measure the speed of light. They described their conclusion and selectively presented data that demonstrated or was indicative of the point they wanted to make, without clear procedures to replicate and not with full transparency.*

-

- **In examining the fertility index and rainfall records at Rothamsted, what two main errors did he conclude about the indexes and how did he express results about the rainfall?** *He examined the competing indexes, and showed when reduced to their elemental algebra, they were all versions of the same formula. And he found that the amount of rainfall was greater than the type of fertilizer used, that the effects for fertilizer and rainfall had been confounded.*
-

- **What problems did Fisher have to tackle in the designing the experiment that could determine if the Lady could distinguish between milk-then-tea and tea-then-milk cups of tea?** *She could guess randomly with a 50% probability of being correct. She could also make an error is spite of being able to distinguish, getting 9 our of 10. How many cups to present, in what order, how much to instruct the lady, and determine beforehand the likelihood of possible outcomes.*

- **What did Fisher conclude scientists needed to start with to do an experiment, and what would a useful experiment do?** *Scientists need to start with a mathematical model of the outcome of the potential experiment, which is a set of equations in which some of the symbols stand for numbers that will be collected as data, and other symbols stand for the overall outcomes of the experiment. Then a useful experiment allows for estimation of those outcomes.*

"Before the middle of the eighteenth century there is little indication of a willingness of astronomers to combine observations; indeed there was sometimes an outright refusal to combine them. The idea that accuracy could be increased by combining measurements made under different conditions was slow to come. They feared that errors in one observation would contaminate others, that errors would multiply, not compensate." – Stigler, 1986

As long as you

"…randomize, randomize, randomize." -- Fisher

Sampling

- Random Sampling
  - Required for *inference* (applying the results from a small sample to larger population).
- Convenience Sampling
  - More biased, but less expensive
  - Requires clear definition of sample origins

Statistics: Given the information in your hand, what is in the pail?

Probability: Given the information in the pail, what is in your hand?

Hypothesis Testing Text

- Identify Population and **COMPARISON GROUP**, state assumptions
- Define the Null Hypothesis
- Define the Research Hypothesis or Alternative hypothesis
  - Define the Research and Control Group
  - Define the Dependent and Independent Variables
- State relevant characteristics of comparison distribution
- Determine critical cutoff values
- Calculate statistic
- **Reject or Fail to Reject the NULL Hypothesis**

Six Steps of Hypothesis Testing Guidelines

- Define the Independent Variables, including the Levels, and the Dependent Variables
- Decide which test you will use (identify the **COMPARISON DISTRIBUTION**) and check Assumptions.
- State Hypotheses (specify Means, Independent Variable, Units of Measurement, and Direction of results)
- Select Cut-offs
    1. One-tailed or Two-tailed (prefer Two-Tailed and use One-Tailed only when it does not make sense to use the Two-tailed).
    2. Pick Alpha (usually .05)
    3. Find cut-off using Table
- Calculate Statistic
- Decide

Hypothesis Testing: Drivers Age

- The average age for licensed drivers in a county is *m = 42.6, σ = 12 and the distribution is approximately normal. A county police officer was interested in whether the average age of those receiving parking tickets differed from that of the average age of the population. She obtained a sample of N = 25 drivers receiving parking tickets. The average age for these drivers was M = 40.5. Perform the six steps of hypothesis testing necessary to determine whether this group differs from the population of drivers in the county.*

Hypothesis Testing Responses: Drivers' Age

- Step 1.
  - Independent variable, Drivers Type (Population / With Tickets)
  - Dependent Variable, Age of Drivers.
- Step 2. Identify Population, all drivers in the county

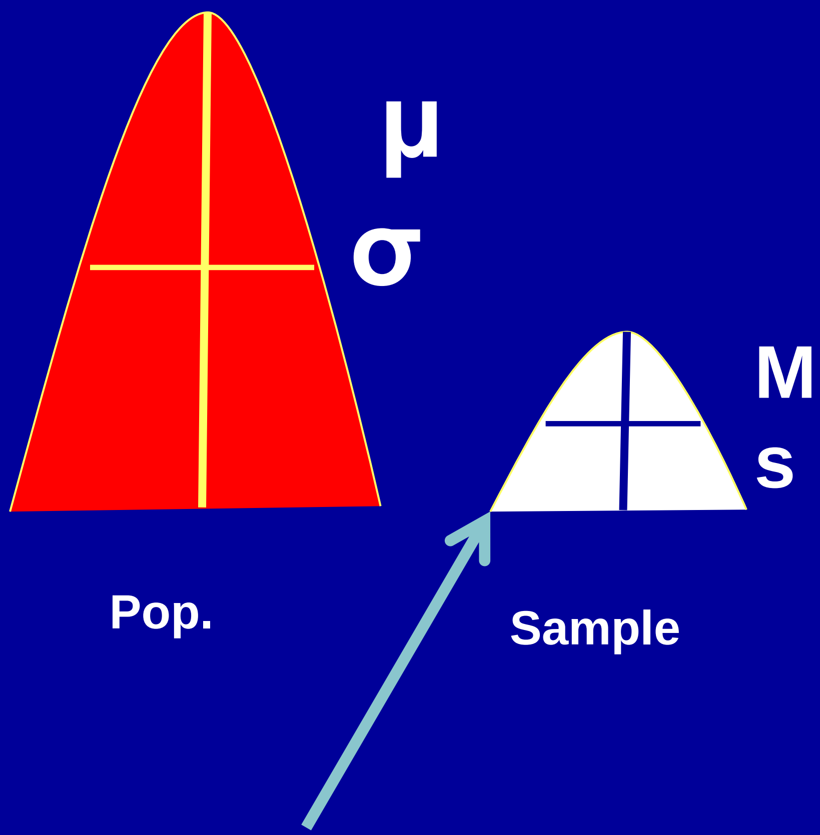
  Comparison Distribution (next slide)

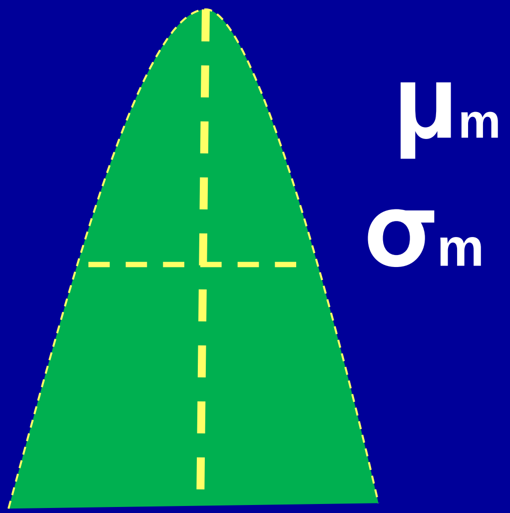
  Assumptions (which test to apply)

When To Use Z-test

- 1 Nominal Independent Variable
- 1 scale (continuous or interval) Dependent variable
- Known Population parameters for μ and σ
- Random selection
- Normally shaped population distribution

Comparison Distribution





Comparison Distribution:
Normal distribution of
DV (and IV)
with μ_m, σ_m
and same N

Think Like a Statistician!
Where did this sample come from?
Hint: Central Limit Theorem

Assumptions for Parametric Tests

- Dependent variable is a scale variable → interval or ratio
  - If the dependent variable <u>is ordinal or nominal, it is a non-parametric test</u>
- Participants are randomly selected
  - If there is no randomization, it is a non-parametric test (and likely a flawed experiment, or one limited in generalization)
- The shape of the population of interest is approximately normal
  - If the shape is not normal, it is a non-parametric test

Assumptions for Parametric Tests

TABLE 8-2. THE THREE ASSUMPTIONS FOR HYPOTHESIS TESTING

We must be aware of the assumptions for the hypothesis test that we choose, and we must be cautious in choosing to proceed with a hypothesis test even though our data may not meet all of the assumptions. Note that in addition to these three assumptions, for many hypothesis tests, including the *z* test, the independent variable must be nominal.

THE THREE ASSUMPTIONS	BREAKING THE ASSUMPTIONS
1. Dependent variable is measured on an interval scale.	Usually OK if the data are not clearly nominal or ordinal.
2. Participants are randomly selected.	OK if we are cautious about generalizing.
3. Population distribution is approximately normal.	OK if the sample includes at least 30 scores.

Assumptions Exercise

1. A set of researchers interested in the health of professional hockey players randomly selected a sample of 40 hockey players from the National Hockey League and asked them to report the number of work-related injuries they incurred over the past month. Would it be appropriate for the researchers to use a parametric hypothesis test? Evaluate each of the assumptions for parametric tests to explain your answer.
2. A New York Times article published on April 24, 2007, reported the research of Dr. Giorgio Vallortigara. In this study, the researchers investigated whether a dog's tail wags in a preferred direction in response to positive as opposed to negative stimuli. The researchers answered this question by recruiting 30 dogs that were family pets. Filming the dog's tail from above, they allowed each dog to view (through a slat in its cage) its owner, an unfamiliar human, a cat (positive stimuli), and an unfamiliar, dominant dog (a negative stimulus). Would it be appropriate for the researchers to use a parametric hypothesis test? Evaluate each of the assumptions for parametric tests to explain your answer.

Assumptions Answers

- **It would be appropriate** for the researchers to perform a parametric test. The assumption that the dependent variable is measured on an interval scale is met. Having players record the number of times they have work-related injuries is actually a discrete, ratio variable, so it definitely meets the assumption of equal intervals. The assumption of random selection is met. Given that the entire population can be identified (the National Hockey League), and given that the researchers said they randomly selected from this population, the assumption of random selection has been met. It is difficult to know whether the distribution of injuries in the population is normal, but given that the researchers have a sample size greater than 30, the sampling distribution of the mean will likely be normal.
- Based on the study description, **it would not be appropriate** for Dr. Vallortigara and his colleagues to use a parametric test. First, the assumption that the dependent variable is measured on an interval scale is violated. The dependent variable is the direction in which the dog wags its tail. This is a nominal variable. Second, it is highly unlikely that the sample of dogs was randomly selected from the population of all dogs that are family pets, as it would be impossible to identify the entire population to begin with. It is unclear what kind of distribution to expect with tail wagging, but the sampling distribution of the mean may be normally distributed, given that there are 30 dogs in the sample.

Hypothesis Testing Responses: Drivers' Age

- Step 3: Null Hypothesis: The average age of drivers who receive parking tickets is no different from that of the population of drivers. H_0: $\mu_1 = \mu_2$
- Alternate Hypothesis: The average age of drivers who receive parking tickets will differ from that of the population of drivers. H_1: $\mu_1 \neq \mu_2$

Hypothesis Testing Responses: Drivers' Age

- Step 4: The critical values, using a *p level of .05 and a two-tailed test, are –1.96 and +1.96.*
- Step 5: Calculate the statistic with a *mean of 42.6 and a standard error of 2.4* .
- *Step 5. Calculate Statistic*

$$z = \frac{\mathrm{M} - \mu_m}{\sigma_m} = \frac{40.5 - 42.6}{2.4} = -0.875$$

- Step 6: Our calculated test statistic does not exceed our critical value. Therefore, we fail to reject the null hypothesis. We have failed to conclude that the age of drivers receiving parking tickets is different from the average age of drivers in the population. Our results are inconclusive and additional research is required.

$$\sigma_m = \frac{\sigma}{\sqrt{N}} = \frac{12}{\sqrt{25}} = \frac{12}{5} = 2.4$$

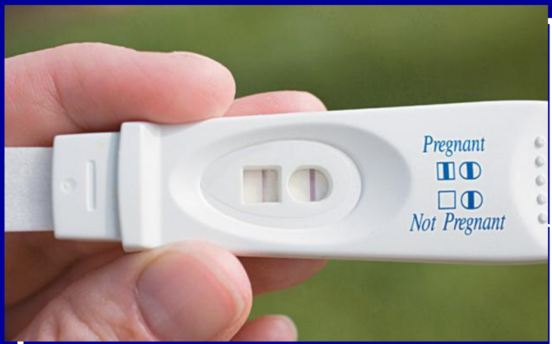
How to Write Results

- Statistical test, Variables, N, statistic, p value.
  - E.g., a Z-statistic of Ticketed Drivers, N= 29, $z = -.875$, $p > .05$ *or* $p = .1895$ (SPSS will give you this number).
  - E.g., a Z-statistic for MFTP scores, N = 97, z = -.33, p> .05 or p = .4678. (SPSS will give you this number)

"What Cold Possibly Go Worng?": Type I and Type II Errors

	Reality	
Perception	YES *Research H1 True*	NO *Research H1 False*
YES *Research H1 True*	Hit	False Alarm (Type I) *False Positive*
NO *Research H1 False*	Miss (Type II) *False Negative*	Correct Rejection

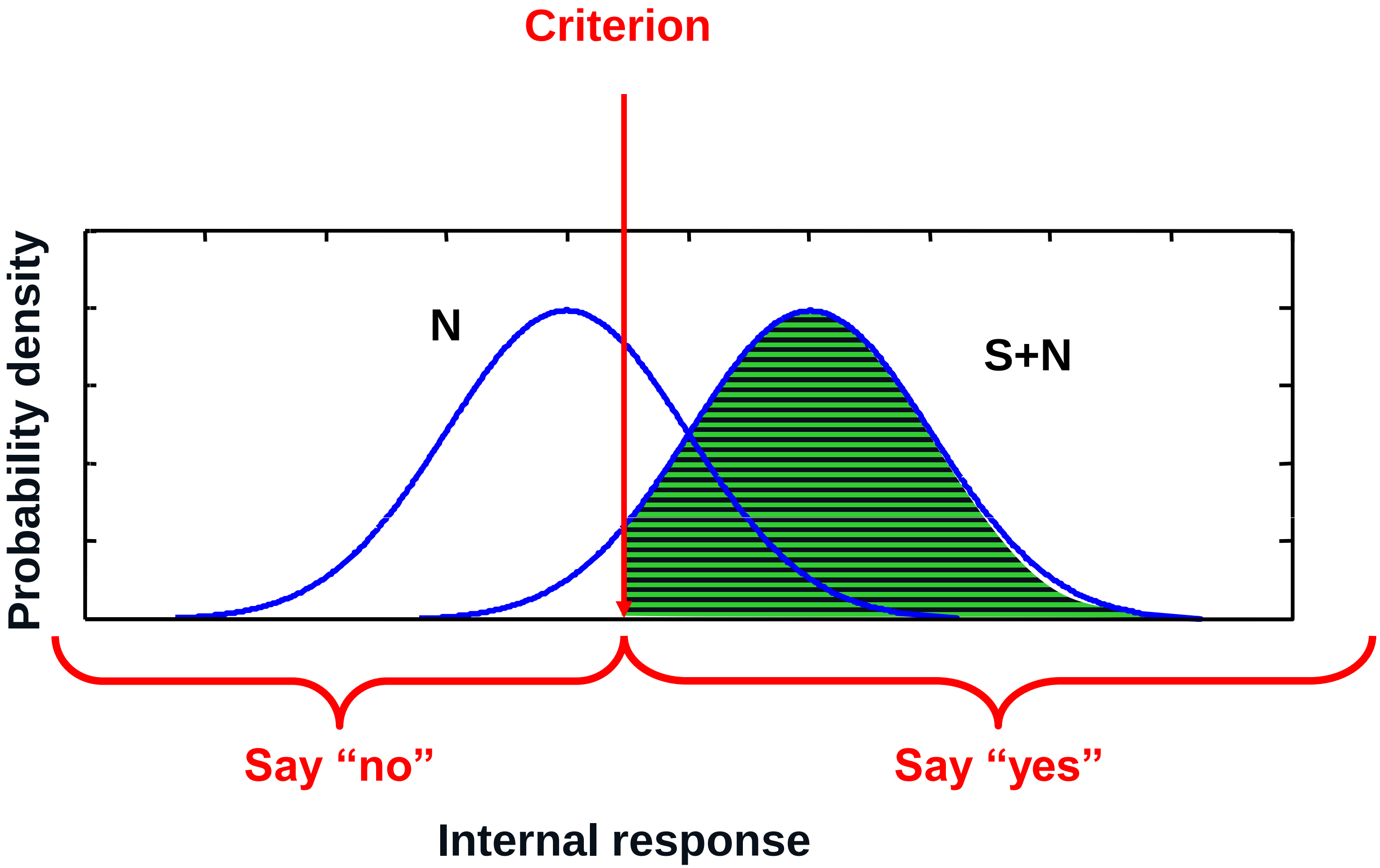
“What Cold Possibly Go Worng?”: Type I and Type II Errors



	REALITY of PREGNANCY	
TEST RESULTS	**YES**	**NO**
“PREGNANT” (Reject Null)	HIT (Pregnant & “+” on test)	FALSE ALARM (Type I) Also called False Positive (Not Pregnant & “+”)
“NOT PREGNANT” (Fail to reject the Null)	Miss (Type II) Also called False Negative (Pregnant & “-”)	Correct Rejection (Not Pregnant & “-”)

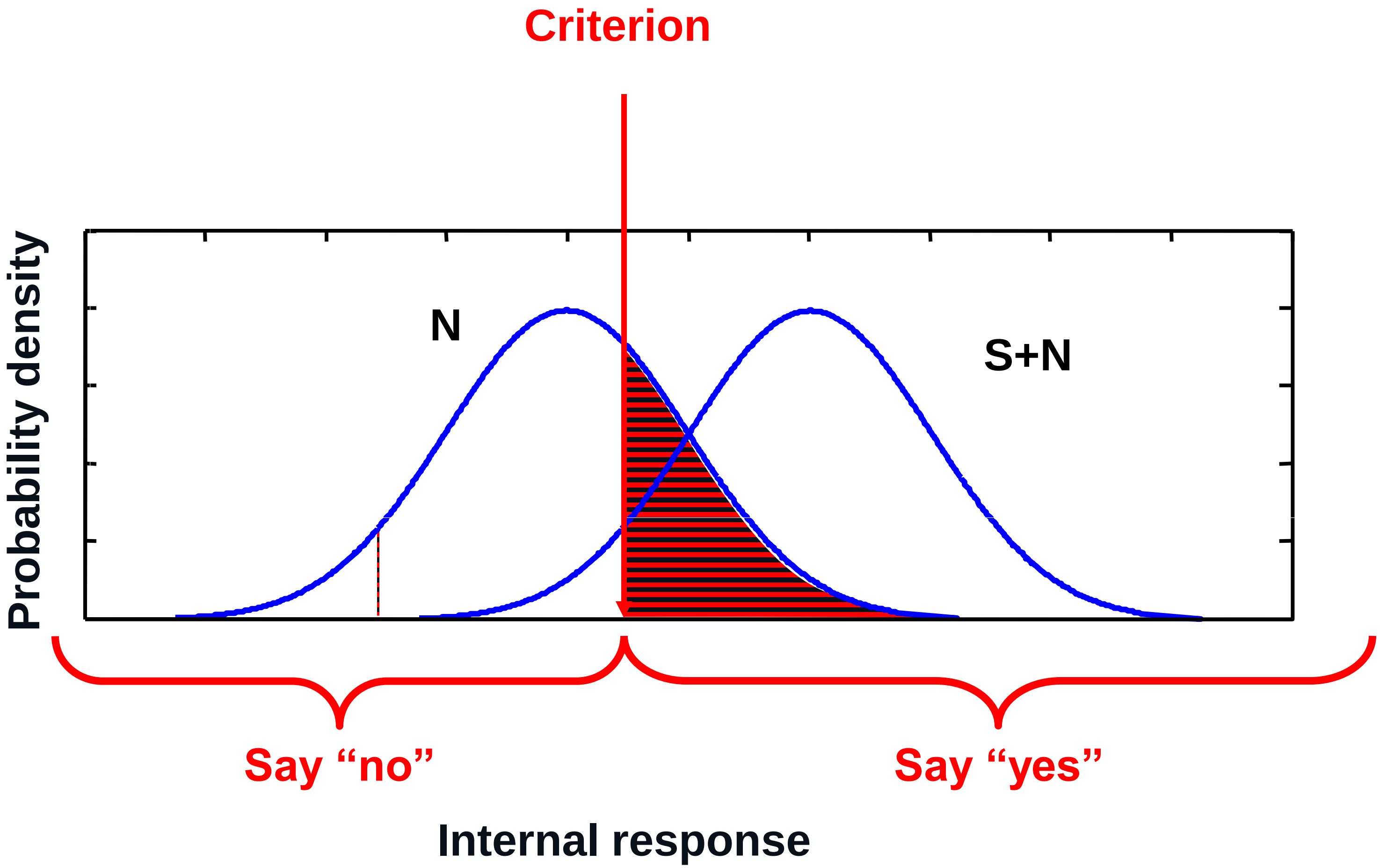
Hits

(response “yes” on signal trial)

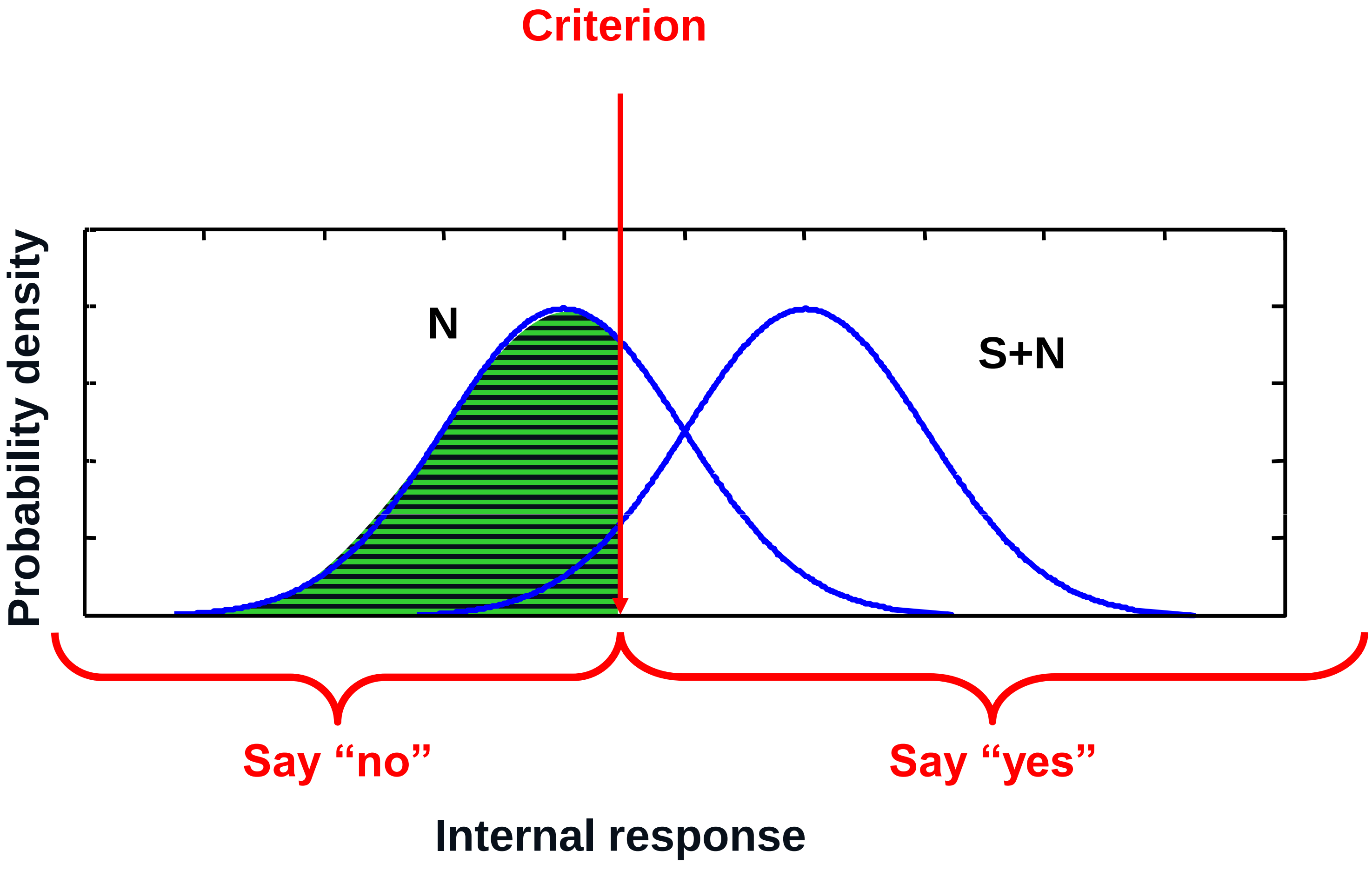


False Alarms

(response "yes" on no-signal trial)

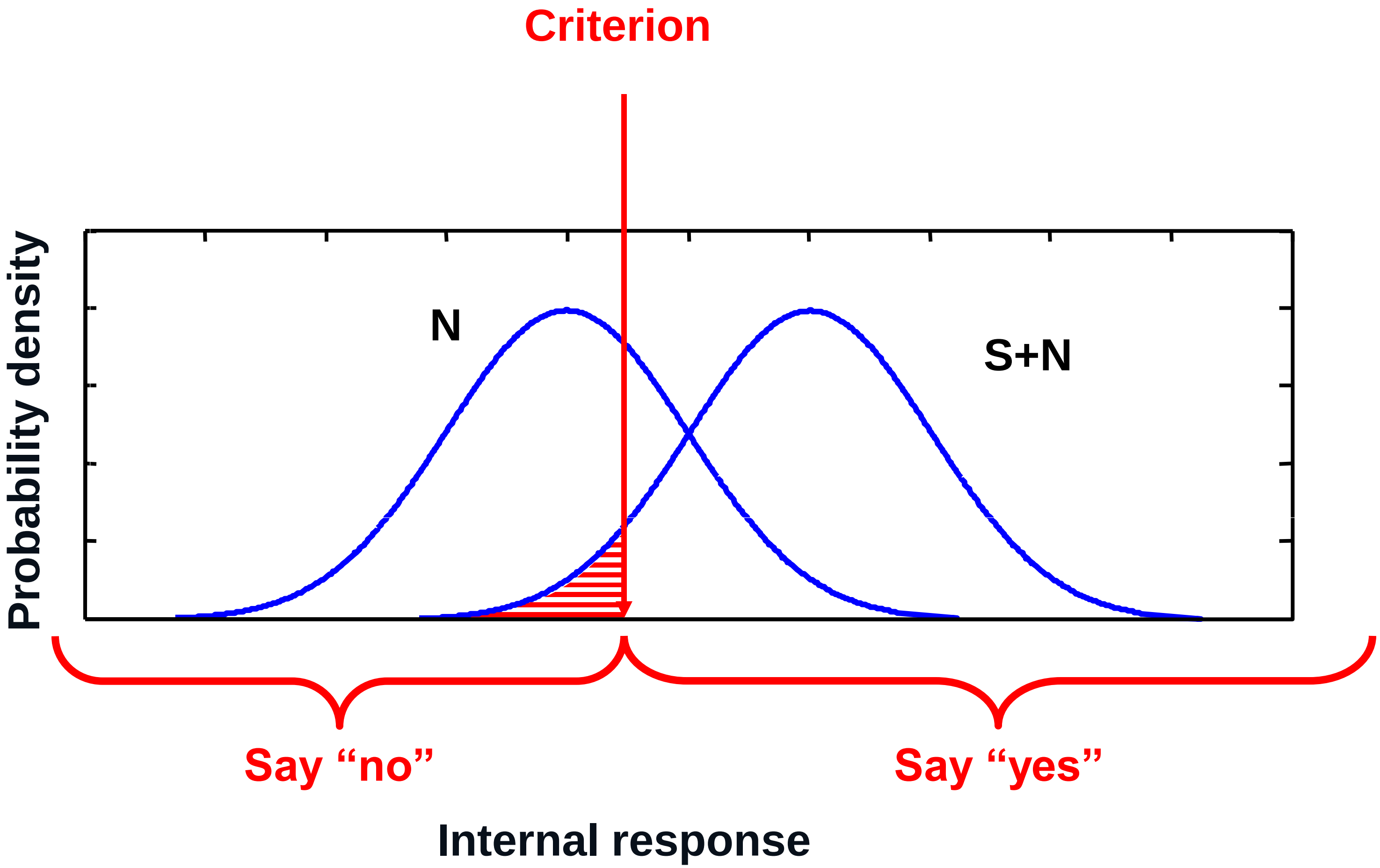


Correct rejects
(response "no" on no-signal trial)

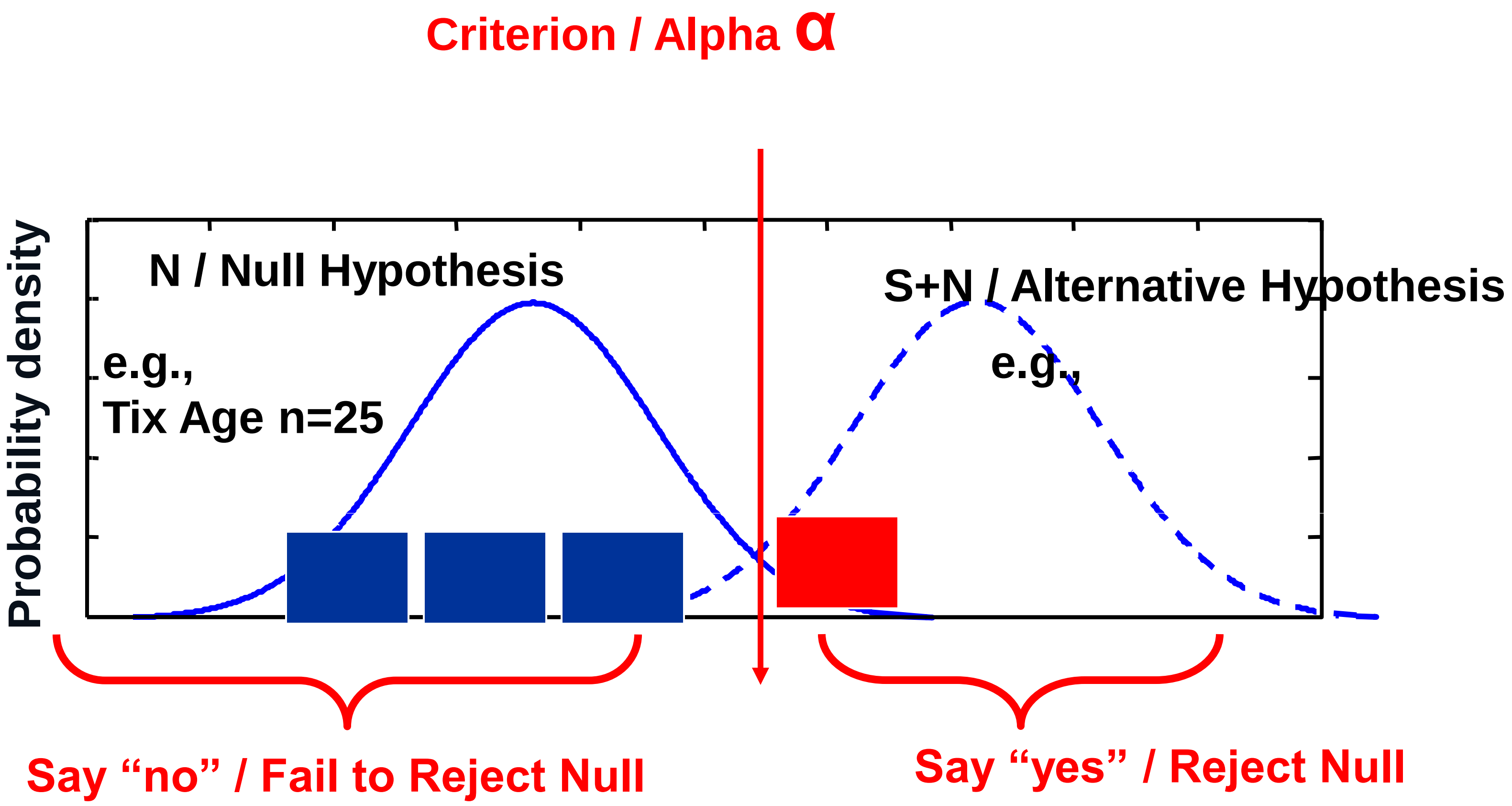


Misses

(response “no” on signal trial)



Hypothesis Testing & Signal Detection Theory



2.3 The Science of Explanation

- Experiment—2 critical features
- (1) Manipulation
  - independent variable
  - dependent variable—measured
  - Control Group Condition (or Variable)
  - Experimental Group Condition (or Variable)

- (2) Randomization

  - controls for a 3rd variable (you know exists but are not interested in)
  - versus self-selection

Variation in IV Causes Variation in DV

1. *Cause → Effect: whenever IV occurs, outcome DV should result.*

   Safe sex intervention ➔ Condoms in Park

2. *Cause absent → Effect absent*

   No SS intervention ➔ no condoms

3. *Cause variation → Effect variation*

   More or better interventions ➔ more condoms in park

Experimental & Control Groups

- Experimental Condition: Cause is valid
  - E.g., drug, alcohol
- Control Condition: cause is invalid
  - Placebo, juice

- Essence of experiment is to control conditions beforehand

The Science of Observation

- Validity—able to draw accurate inferences
  - construct validity: e.g., describing what intelligence is and is not, "construct" refers to the "theory"
  - predictive validity: over time you find X predicts Y
- Reliability—same result each time?
  - Test/Re-Test
  - Parallel
  - Inter-Item

Statistical Significance

- A finding is statistically significant if the data differ from what we would expect <u>from chance alone</u>, if there were, in fact, no actual difference.
- They may not be significant in the sense of big, important differences, but they occurred with a probability below the critical cutoff value, usually Alpha = .05 (z-stat = ±1.96 two-tailed or 1.64 one-tailed)
- **<u>Reject or Fail to Reject the NULL Hypothesis</u>**