Statistical Hypothesis Testing

Ghostbusters

Ghostbusters

- How many cards should (the incredibly inappropriate) Professor Murray show each student to determine if they have psychic ability?

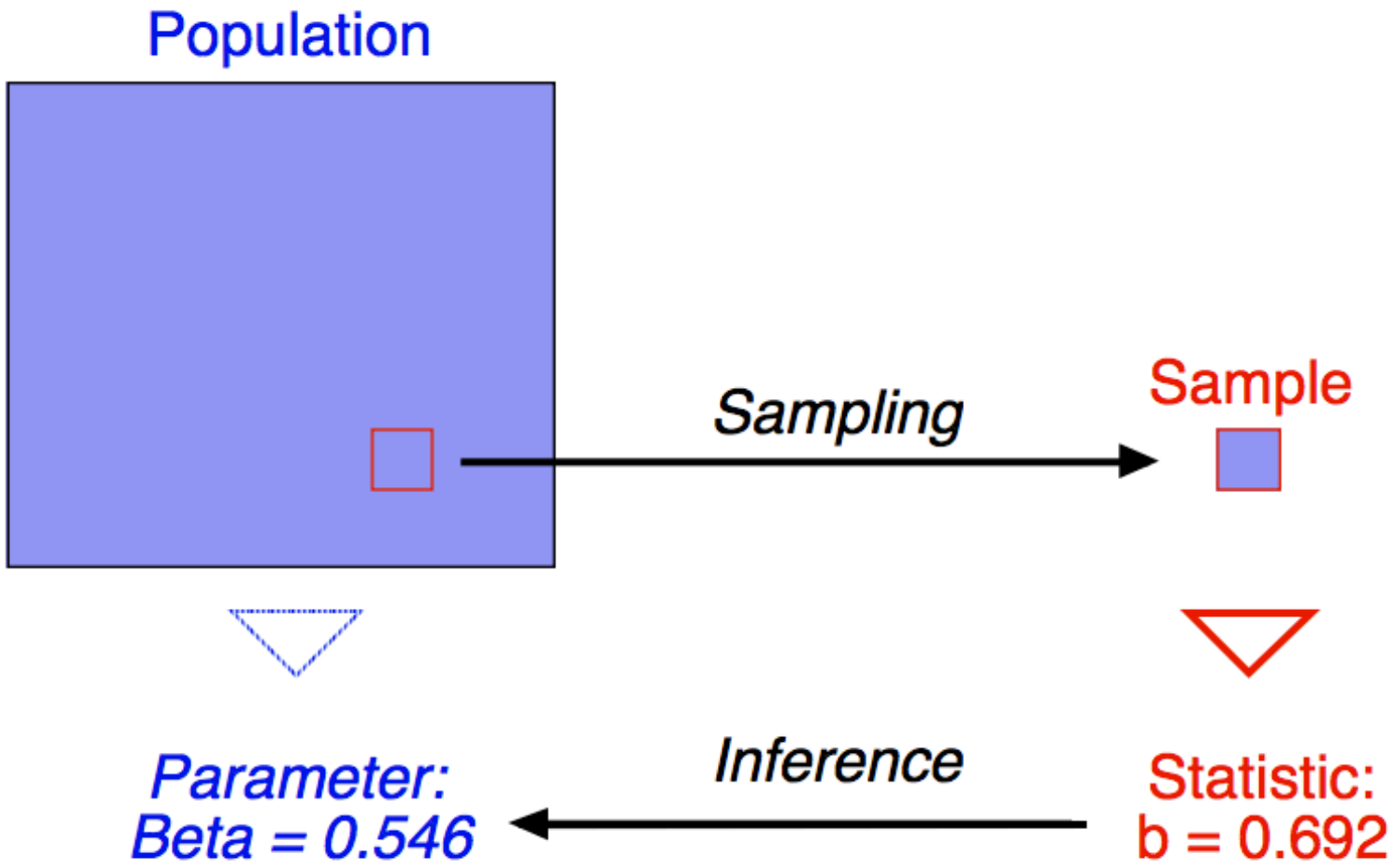
"The Lady Tasting Tea"

- British statistician Ronald Fisher (1890-1962) had a friend named Muriel Bristol, who claimed that she could distinguish, by taste alone, whether a cup of tea was poured with the milk first or the tea first.
- Fisher wanted to determine if she could or not.
- Suppose she is presented with 2 cups of tea and determines correctly for each one whether the tea or milk was added first. Should we believe in her abilities? What if she guesses 10 cups correctly? What if she guesses 8/10 cups correctly?

Goggle

- Suppose new search engine Goggle recently announced the following information.
- "On a random sample of queries, our new search engine obtains 56% accuracy, as opposed to Google, which is only 54% accurate."
- Should we all switch to Goggle?

Populations and samples



Statistical inference

- **Parameter estimation**: infer the *value* of a population parameter based on a sample statistic.
  - What is an accurate estimate of student height on the Rhodes campus?
- **Hypothesis testing**: infer the *answer to a question* about a population parameter based on a sample statistic.
  - Are students on the Rhodes campus taller than those at Sewanee?
  - Are male students taller than female students?

Null and alternative hypotheses

- Statistical hypotheses are sets of statements about population parameters.
  - Null hypothesis (H_0)
  - Alternative hypothesis (H_1)
- Statistical hypotheses are closely related to, but different from, research hypotheses.
- A common null hypothesis is "no effect" --- complete randomness --- but a null hypothesis can also specify an effect of a particular magnitude.

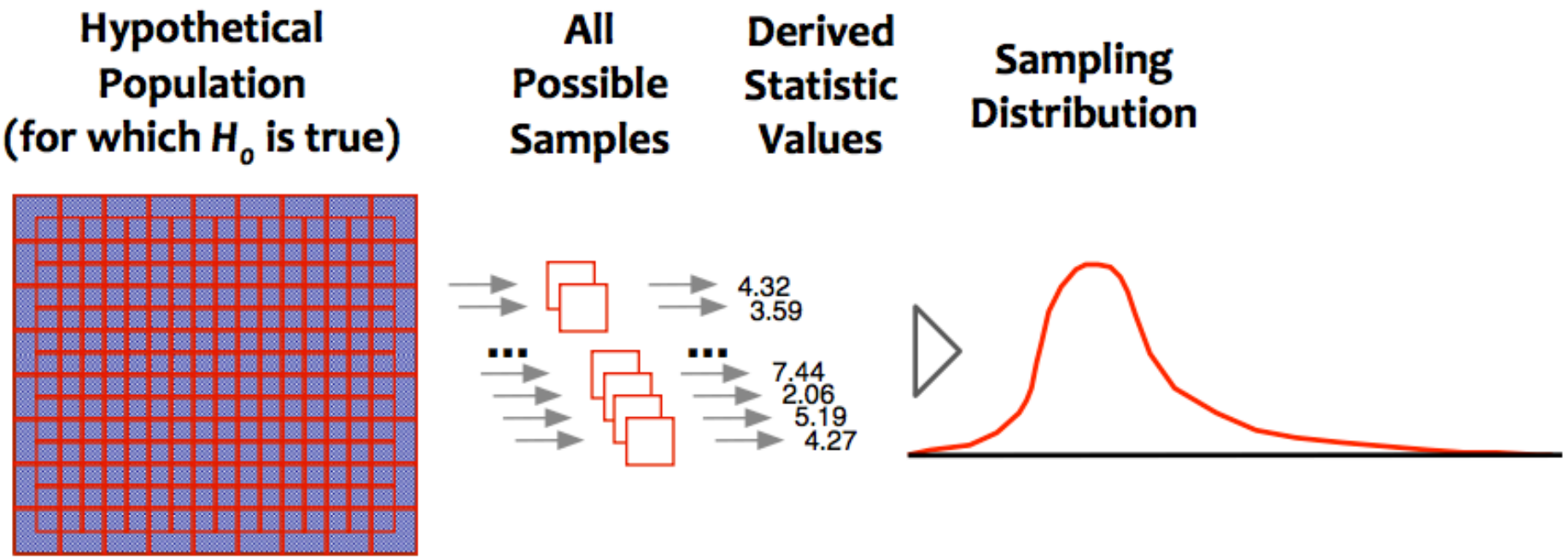
Null and alternative hypotheses

- Statistical hypotheses are statements about population parameters.
  - Null hypothesis (H_0)
  - Alternative hypothesis (H_1)
- What are some possible null/alt hypotheses for these questions?
  - Ghostbusters
  - Lady tasting tea
  - Student heights

Null and alternative hypotheses

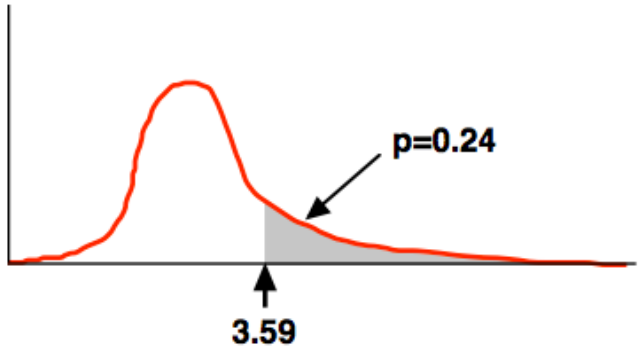
- Our goal is to use experimental data to determine if the data we have seen is ***very unlikely*** to have occurred if H_0 were true.

Sampling distributions



Statistical significance

- A value of a statistic is statistically significant if it (or a more extreme value) is unlikely to occur under the null hypothesis.



- α = P(reject H_0 | H_0 is true) = P(Type I error)

The meaning of a hypothesis test

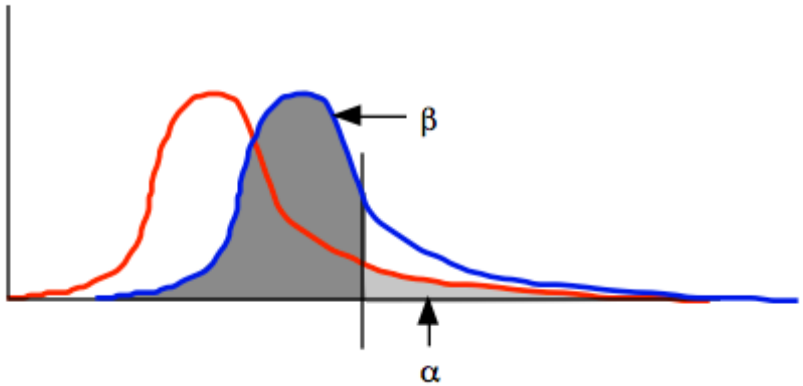
- A hypothesis test does not tell you the probability of H_0 or H_1 being true.
- Instead, it tells you the probability of a particular result (or a more extreme one) given that H_0 is true.
- This is very important!

Statistical vs substantive significance

- Statistical significance can be seen as a necessary, but not a sufficient, condition for importance.
- Effects can be
  - very small but statistically significant
  - very large but not statistically significant
- Hypothesis tests guard against one type of error only: mistakenly thinking your results imply H_1 when they are very probable under H_0.

Statistical power

- Lack of statistical significance does not necessarily imply that H0 is true.
- Test could have low statistical power



- $\beta = P(\text{Accept } H_0 \mid H_0 \text{ is false}) = P(\text{Type II error})$

Why do we do this?

- Wouldn't it be better to calculate what we really want to know?
  - Presumably, the probability of H1 being true given the data?
- This is often a much harder question to answer, because we usually need to know the entire space of hypotheses, which is often unknowable.

Hypothesis testing strategy

- Formulate null (and alternative hypotheses)
  - Example H_0: $\mu_A = \mu_B$ H_1: μ_A != μ_B
- Gather data
- Calculate a sample statistic (e.g., $\bar{x}$)
- Estimate the sampling distribution for that statistic given H_0.
- Use the sampling distribution to calculate the probability of obtaining the sample statistic (or a more extreme statistic) given H_0.
- If the probability is low, reject H_0 in favor of H_1.

The lady tasting tea

- How good was Muriel Bristol?

- This is called ***Fisher's exact test*** or a permutation test.

Ghostbusters

- Suppose you test for ESP using Bill Murray's method with 20 playing cards. The student guesses the suit on each one (hearts, diamonds, spades, clubs). Suppose they get 9 right. Do they have ESP?

- This is called a ***binomial test***.

Kinds of tests

- Discrete vs continuous
- One-sample tests
  - Used when a sample statistic is compared to the population from a hypothesis.
  - "Did this year's 141 students do better on the test than average?"
- Two-sample tests
  - Used when comparing two samples to see if they come from different distributions (must be independent samples).
  - "Did my 10am 141 class do better on the test than the 11am 141 class?"
- Paired tests
  - Similar to two-sample tests, but used when the two samples are not independent.
  - "Did my 10am 141 class do better on the first exam or the second exam?"