

Supplemental Chapter - Hypothesis Testing

Mallory Barnes

Hypothesis Testing

Hypothesis testing is a way to decide if a certain statement about a population might be true based on sample data.

Example:

Let's say we want to know if the average pollution in a set of water samples is above the legal limit. Or if young deer in a region are, on average, healthy.

Step 1: Define Your Hypotheses

First, we need to define two hypotheses: the **research hypothesis** and the **null hypothesis**.

- **Research Hypothesis (H_a):** This is what we aim to support. **Remember, we can never “prove” H_a , only fail to reject H_0 .** It can take a few forms based on the question:
 - H_a : average pollution > legal limit (pollution is too high)
 - H_a : average pollution < legal limit (pollution is too low)
 - H_a : average pollution \neq legal limit (pollution is just different)
- **Null Hypothesis (H_0):** This is the default or ‘no change’ scenario. It’s opposite to the research hypothesis.
 - H_0 : average pollution = legal limit (for the first H_a)
 - H_0 : average pollution = legal limit (for the second H_a)
 - H_0 : average pollution = legal limit (for the third H_a)

Step 2: Choose Your Test Statistic

Based on the data, we'll compute a **test statistic**. This number will help us decide which hypothesis seems more likely.

Step 3: Determine the Rejection Region

Before running the test, we decide on a **rejection region**. If our test statistic falls in this region, we'll reject the null hypothesis.

Step 4: Check Assumptions

Before drawing conclusions, ensure that the test's conditions and assumptions are satisfied.

Step 5: Draw Conclusions

Finally, based on the test statistic and the rejection region, decide whether to reject the null hypothesis.

Errors in Hypothesis Testing

Sometimes, even with the best methods, we make incorrect decisions.

- **Type I Error (α):** This happens when we mistakenly reject the true null hypothesis. Imagine wrongly accusing someone innocent. Typically, α is set at 0.05 (5%).
- **Type II Error (β):** Here, we mistakenly accept a false null hypothesis. Think of it as letting a guilty person go free.

Decision	If the null hypothesis is True	If the null hypothesis is False
Reject H_0	Type I error (prob = α)	Correct (prob = $1 - \beta$)
Fail to reject H_0	Correct (prob = $1 - \alpha$)	Type II error (prob = β)

Key Takeaway: As α gets smaller, β gets bigger, and vice-versa.

One Tail or Two?

Consider our pollution scenario, where the historic pollution level was 10.0 ppb. We can set our hypotheses as:

- H_0 : pollution = 10.0 ppb. This is a two-tailed test. We want to know if the pollution is either significantly higher or lower than 10 ppb.
 - H_0 : pollution < 10.0 ppb. A one-tailed test. We're checking if pollution might be significantly above 10 ppb.
 - H_0 : pollution > 10.0 ppb. Another one-tailed test. We're seeing if pollution might be significantly below 10 ppb.
-

Deciphering Significance with P-values

The p-value gives us an idea of how strange our data would appear if the null hypothesis were indeed accurate.

- **One-Tailed Test:** The p-value shows the likelihood of observing an average as extreme as our sample's if the null hypothesis stands.
- **Two-Tailed Test:** This p-value represents the odds of spotting an average as different from the null value as our sample's.

Rule of Thumb: If the p-value is less than α , we opt to reject the null hypothesis.

Graphical Review

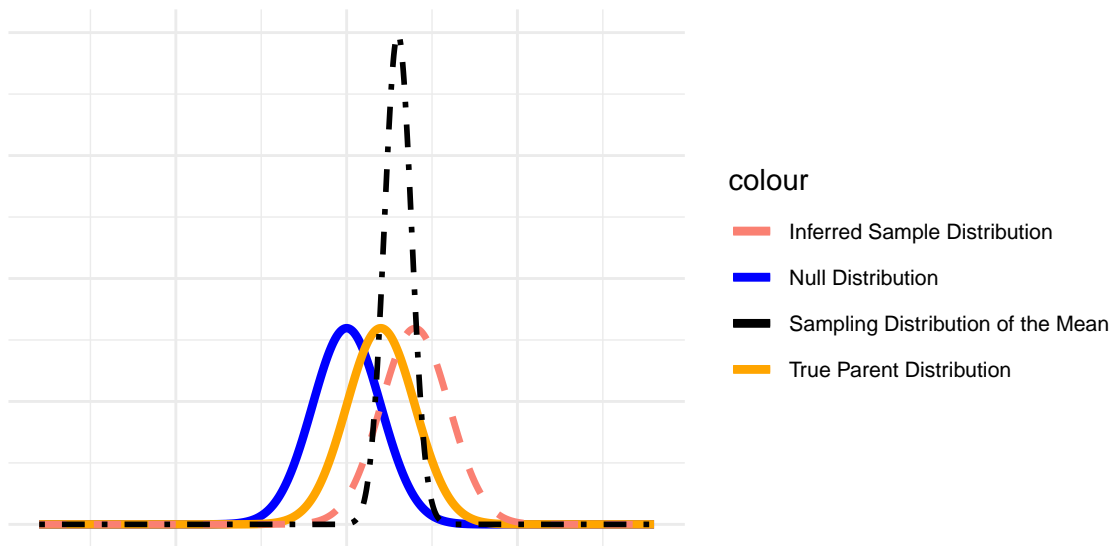
Key Players in Hypothesis Testing Visualization

Sure thing, I've revised the Markdown code to ensure consistent terminology throughout the document. Here's the revised version:

Key Players in Hypothesis Testing Visualization

We define and visualize the core components essential to understanding the graphical representations of hypothesis testing:

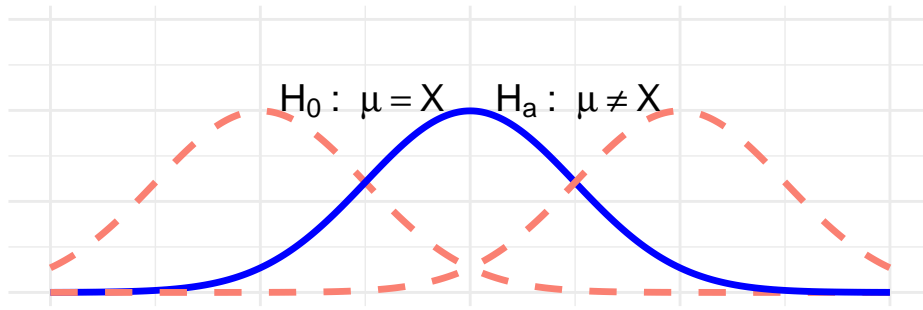
1. **Null Distribution** - The hypothesized parent distribution under the assumption that the null hypothesis H_0 is true.
2. **Inferred Sample Distribution** - The inferred parent distribution of our sample, often associated with the alternate hypothesis H_a .
3. **True Parent Distribution** - The genuine distribution from which our sample originates.
4. **Sampling Distribution of the Sample Mean** - Represents the means of various samples drawn from the parent distribution.



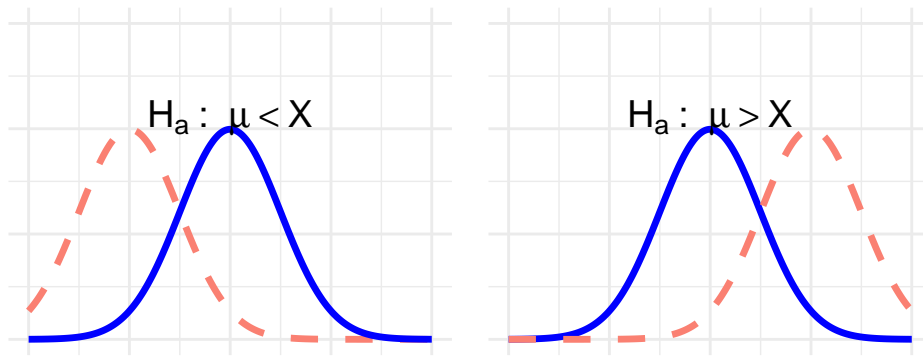
Using the above, we can easily identify and understand the different components in the hypothesis testing figures.

With this in place, you should be able to refer back to this section as a quick reference when going through the subsequent detailed graphs.

Hypothesis Testing Concepts

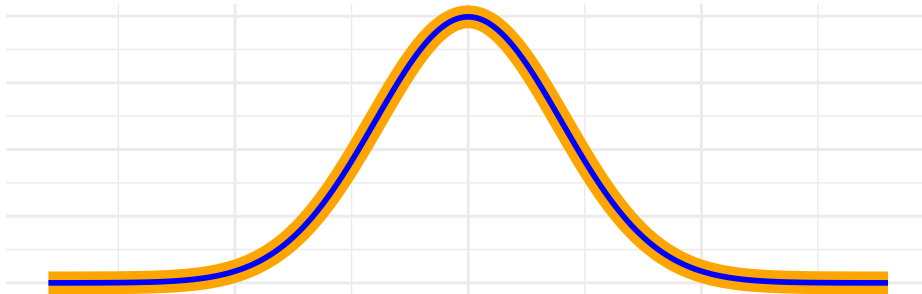


For a two-tailed alternative, we are interested in the possibility that a sample comes from a parent distribution that may have a lower or higher location than the null.

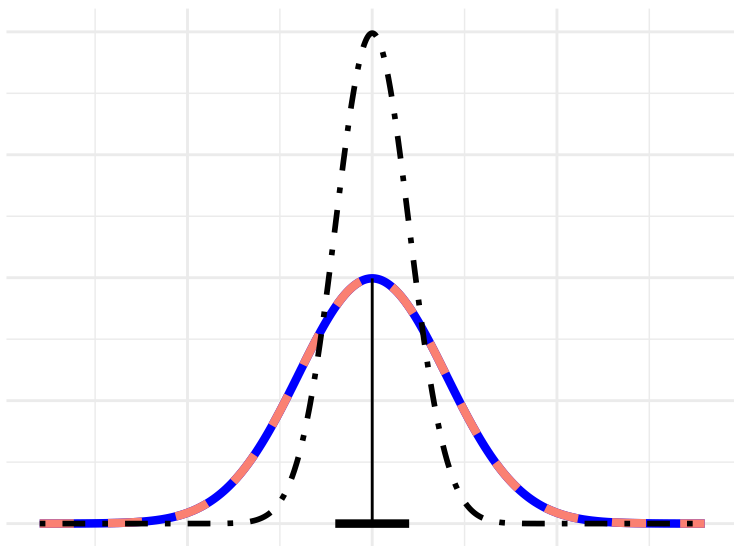


For a one-tailed alternative, we are interested in the possibility that a sample comes from a parent distribution that is either at a lower or higher location than the null, but not both.

$$H_0: \mu = X \quad \mu \geq X \quad \mu \leq X$$



In a "perfect" world in which the null hypothesis is true, the sample's parent distribution (solid, orange) is exactly the same parent distribution described by the null hypothesis (solid, blue).



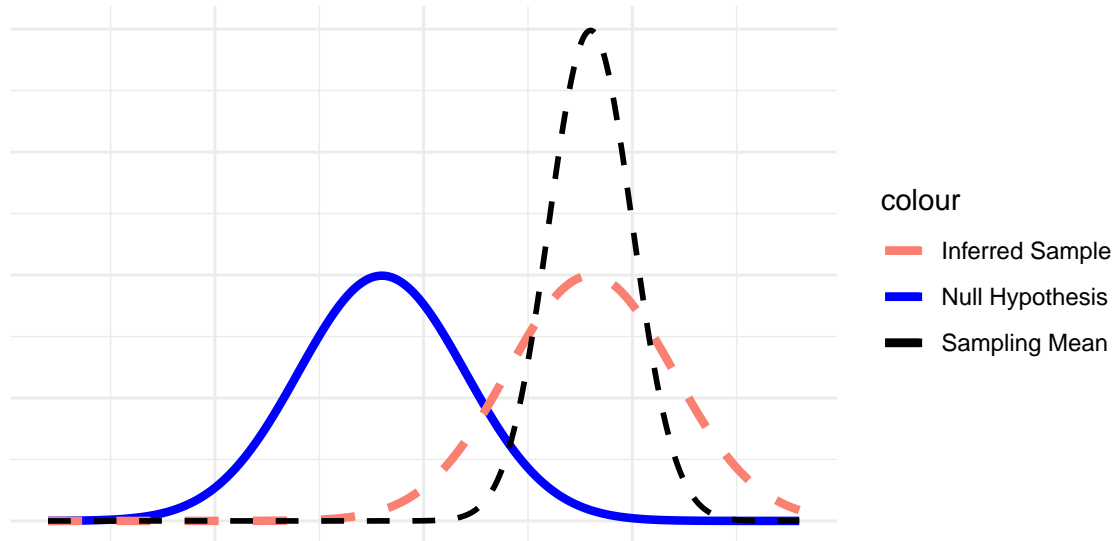
We never know the true parent distribution of the sample – we infer it from the sample. Here, the tall dash-dotted line shows the sampling distribution of the mean, from which we infer the parent distribution (salmon, dashed).

In this even more perfect world, that parent distribution is the same as the parent distribution described by the null hypothesis and we have taken a perfectly representative sample, so all 3 curves line up perfectly on the same mean. The thick, short, flat black line is the confidence interval for the sample mean.

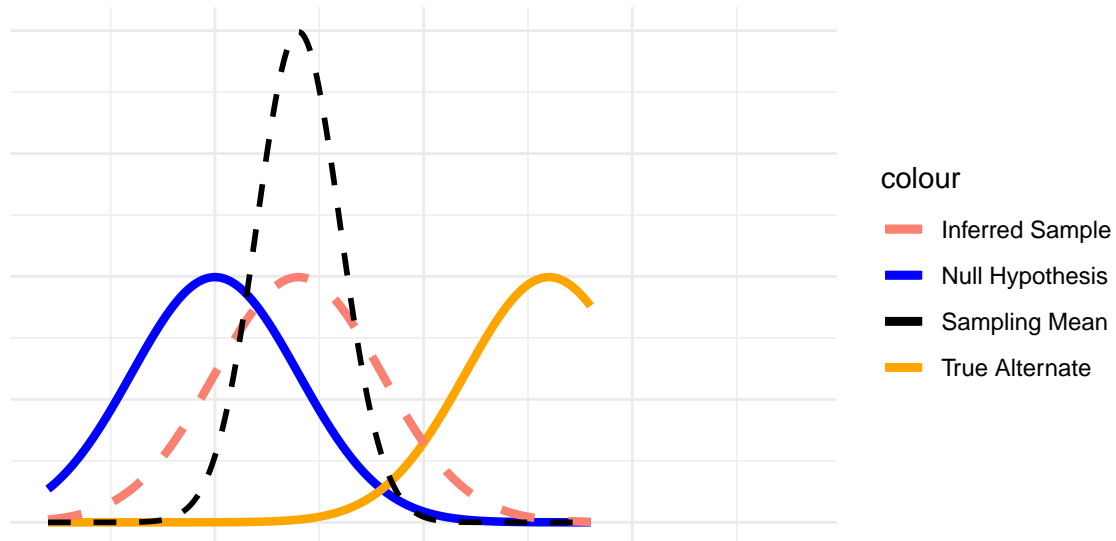


In an imperfect but convenient world, the sample is not a perfect representation of the parent

population, but is fairly close. The sample mean is close to hypothesized mean, and (in the 2-tailed case) the confidence interval for the sample mean “catches” the mean of the null hypothesis (thin solid line). A hypothesis test will correctly determine that there is not a significant difference between the sample mean and the mean of the null hypothesis.



In an imperfect and inconvenient world, the random sample is, by chance, sufficiently imperfect that the apparent (inferred) parent distribution is far from the true parent distribution and (in the 2-tailed case) the confidence interval for the sample mean no longer “catches” the mean of the null hypothesis. A hypothesis test will now find a significant difference between the sample mean and the mean of the null hypothesis. **This is a type I error.**



In another imperfect and inconvenient world, the sample (dashed lines) really is drawn from the alternative distribution (the sample's true parent distribution; orange), but is unrepresentative of its parent and similar to the null (solid black line). The confidence interval (in the 2-tailed case) of the sample "catches" the mean of the null hypothesis although it is far from the mean of the true parent of the sample. A hypothesis test will find no significant difference between the sample mean and the mean of the null hypothesis. **This is a type II error.**