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.

Clarifying Alpha, P-value, and Confidence Level

Alpha (α)

Alpha (α) is the significance level of a statistical test, and it quantifies the risk of committing a Type I error. A Type I error happens when we incorrectly reject a true null hypothesis. The standard value for alpha is often set at 0.05, implying a 5% chance of making a Type I error. In other words, we are willing to accept a 5% risk of concluding that a difference exists when there is no actual difference.

P-value

The p-value is another crucial concept in hypothesis testing. It represents the probability of observing the obtained results, or something more extreme, assuming that the null hypothesis is true. A small p-value (usually 0.05) suggests that the observed data is inconsistent with the null hypothesis, and thus, you have evidence to reject it.

Confidence Level

The confidence level is related but distinct from alpha and p-value. While alpha quantifies the risk of a Type I error, the confidence level indicates how confident we are in our statistical estimates. The confidence level is calculated as the complement of alpha:

Confidence Level = 1 - α

For example, if is 0.05, the confidence level would be (1 - 0.05 = 0.95) or 95%. This means we are 95% confident that our results fall within a specific range.

Bringing It All Together

- **Alpha (α)**: Risk of Type I error (usually 5%)
- **P-value**: Probability of observed data given the null is true
- **Confidence Level**: Confidence in the range of our estimates (usually 95%)

Understanding the relationship and differences between these three concepts is crucial for accurate and meaningful interpretation of statistical tests.

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 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 - β)
Fail to reject H_0	Correct (prob = 1 - α)	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:

- $\mathbf{H_0}$: 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.
- $\mathbf{H_0}$: pollution $<$ 10.0 ppb. A one-tailed test. We're checking if pollution might be significantly above 10 ppb.
- $\mathbf{H_0}$: 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

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. **True Parent Distribution** - The actual distribution from which our sample originates.
3. **Inferred Parent Distribution** - The parent distribution inferred from our sample data. This is what we conceptualize as the distribution of H_a.
4. **Sampling Distribution of the Sample Mean** - Represents the distribution of sample means if we were to draw multiple samples from the parent distribution. This is crucial for making inferences about the **Inferred Parent Distribution**.

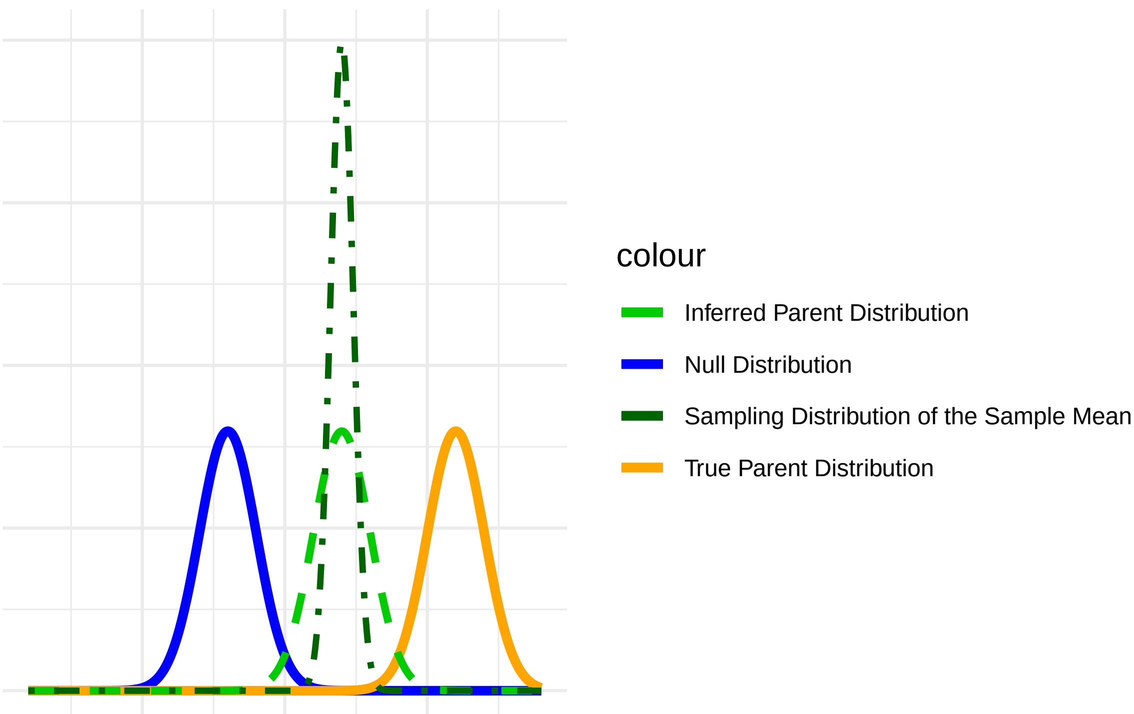

Figure 1: This figure shows the 4 important distributions that we need to think about with regard to hypothesis testing.

Using the above, we can hopefully 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.