Hypothesis Testing: Decision Making under Uncertainty

Hypotheses: For our purposes a **hypothesis** is a testable statement about reality.

Hypothesis Test: A **hypothesis test** is a procedure for deciding between two or more competing hypotheses using data.

Null and Alternative Hypotheses: In this course we will be concerned with testing two mutually exclusive and exhaustive hypotheses:

$\mathbf{H_1}$: alternative or research hypothesis

1. Hypothesis for which we seek evidence
2. Initially assumed false

$\mathbf{H_0}$: null hypothesis

1. Initially assumed true
2. Contradictory to H_1

By exhaustive we mean that between them H_0 and H_1 cover ("exhaust") all possibilities.

Proof by Contradiction: Our hypothesis testing procedure uses proof by contradiction: we seek to prove the research hypothesis H_1 by contradicting H_0 using data.

Exercise 1. Give an example of proof by contradiction.

Exercise 2: Suppose a friend has a nickel, either one-headed or two-headed. She won't show us the nickel but she will flip it repeatedly and give us the outcomes. Suppose we want to prove the coin is two-headed. Answer the following:

1. Identify the alternative and null hypotheses, H_1 and H_0.
2. In terms of the outcomes from repeatedly flipping the coin, what must occur for you to reject H_0 and thereby conclude H_1 is true?
3. Suppose your friend gets 10 heads in a row. Does this result contradict H_0?
4. Suppose your friend gets 20 heads in a row. Does this result contradict H_0?

Problem: Proof by Contradiction Using Random Data. Note that no matter how many heads in a row your friend reports, H_0 is not contradicted. Why? Because we can always reconcile such a result with H_0, coin is one-headed, by appealing to chance. Something really unlikely - but not impossible - occurred. This being the case, how can we ever prove H_1 is true?

Solution: Proof by Contradiction Using Random Data. When doing proof by contradiction using random data, we proceed as follows. Although we can usually reconcile any outcome with H_0 using chance, at some point this reconciliation is too implausible so we reject H_0 and conclude H_1 is true.

Problem: Need to Measure Plausibility of H_0/Data Reconciliation via Chance? In order to implement the above solution/procedure, we need an objective, *quantitative* measure of the plausibility of reconciling H_0 with the data using chance.

Solution: P-value Measures Plausibility of H_0/Data Reconciliation via Chance. The p-value, a probability, quantifies the plausibility of reconciling the data and H_0 by appealing to chance. The smaller the p-value, the smaller this plausibility and thus the stronger the evidence against H_0 in favor of H_1.

P-value: Probability, **assuming H_0 true**, of repeating experiment and getting data at least as contradictory to H_0 (and favoring H_1) as the actual data.

Exercise 3: For the coin-tossing scenario of exercise 2, do the following:

1. Suppose your friend gets 10 heads in a row. What is the p-value for this data?
2. Suppose your friend gets 20 heads in a row. What is the p-value for this data?

When to Reject H_0? Since the p-value measures the plausibility of reconciling the data with H_0 using chance we see that

1. the smaller the p-value the smaller the plausibility of the chance explanation;
2. the smaller the p-value the greater the evidence **against** H_0; and
3. the smaller the p-value the greater the evidence **for** H_1.

Thus **when the p-value is low we reject H_0!** But how low/small must the p-value be for us to reject H_0 and assert H_1 is true? In order to answer this question, we must consider the errors we can commit doing hypothesis testing.

Hypothesis Testing Errors: Type I and Type II Errors

Hypothesis Testing Errors: Because hypothesis testing is a form of induction, it is an error-prone process. The following table sets out the various possibilities:

Outcome	H_0 True	H_0 False
Fail to Reject H_0	Correct	Type II Error
Reject H_0	Type I Error	Correct

Note that there are two possible errors we can make, depending on whether H_0 is true or not:

> **Type I Error:** Rejecting H_0 when H_0 is true
> **Type II error:** Failing to reject H_0 when H_0 is false

We have names for the probabilities of committing these errors:

> α**:** Probability of committing a type I error
> β**:** Probability of committing a type II error

What α to Use: Since α and β are error probabilities, we want to test our hypotheses in such a way that $\alpha = \beta = 0$. Unfortunately, this is not possible. In almost all cases we face a trade-off when attempting to minimize these error probabilities. For a fixed amount of data (sample size n) if we decrease α, we increase β. Likewise, if we decrease β, we increase α. The only way to simultaneously decrease α and β is to increase the amount of information (sample size n) or to use a better hypothesis testing method. Since acquiring data is usually expensive, we can never fully escape the trade-off between the two error probabilities. Thus we determine a suitable trade-off using the Neyman-Pearson approach: we weigh the costs associated with the two types of error and determine the maximum type I error probability we can tolerate. We then conduct the hypothesis test so that the probability of a type I error does not exceed this maximum value, which is called the *significance level*, or α, of the test. The rationale of the Neyman-Pearson approach is this: By using the largest type I error probability we can tolerate, we keep the probability of a type II error as low as possible. In other words, using a smaller α than necessary will cause β to be larger than necessary.

Conducting α-level Hypothesis Test: Once we have decided what level of significance or α to use, how do we test our hypotheses so as to achieve α? We use the following decision rule:

> **Hypothesis Testing Decision Rule:** If the p-value $\leq \alpha$, reject H_0; otherwise, fail to reject H_0.

Hypothesis Testing Outline: Summarizing the preceding, we arrive at the following hypothesis testing procedure:

1. Identify H_0 and H_1
2. Determine the significance level α
3. Acquire suitable data
4. Compute the p-value
5. If the p-value $\leq \alpha$ reject H_0 and assert H_1 true; otherwise, fail to reject H_0.