Hypothesis Testing

Setting Up Hypothesis
Summary : Logic of Hypothesis Testing

Example 1 : James Bond and Martini Taste Test

Statisticians illustrate hypothesis testing through examples like R. Fisher's lady tasting tea story. A similar example is James Bond's Belief that martinis should be shaken, not stirred.

- An experiment is designed where Bond is given 16 taste tests and must identify if a martini was shaken or stirred
- The selection method is randomized by flipping a fair coin
- Bond correctly identifies 13 out of 16 drinks.

The question is : Was Bond just luck, or does he truly distinguish shaken from stirred martinis?

to test his, the Binomial distribution is used to calculate the probability of guessing correctly 13 or more times out of 16 if he is just guessing.

- Probability = 0.0106 (1.06%)
- Since this probability is very low, it suggests that Bond's performance is unlikely due to chance alone.
- This does not prove that Bond can distinguish between martinis, but it provides strong evidence that he can.

Date / /
Page No.

Example 2 : Physicians' Reactions to patient weight
A study examined whether physicians spend less time with obese patients.

- Randomly selected physicians were given identical patient charts, except:
  - Half saw an obese patient's chart
  - Half saw an average weight patient's chart
- 33 physicians viewed average-weight patient charts.
- 38 physicians viewed obese patient charts.

Findings:

- Mean time spent with obese patients : 24.7 minutes
- Mean time spent with average-weight patients : 31.4 minutes
- Difference : 6.7 minutes

The study asks: Is this difference due to physicians' bias, or just random chance?

- Random assignment does not guarantee that all other factors (eg physician's age, experience) are perfectly balanced.
- To test whether the observed 6.7 minute difference is due to chance, statistical methods calculate the probability
- Probability = 0.0057 (0.57%)
- Since this probability is very low, it suggests that physicians' decisions were influenced by patient weight rather than random chance

Date / /
Page No.

Summary: Understanding probability values in Hypothesis Testing
Key Concept: Probability Value (P-value).
The probability value (p-value) represents the probability of obtaining a certain outcome assuming the null hypothesis is true. It does not represent the probability that the null hypothesis itself is true.

Key takeaways
1 The p-value tells us the probability of an outcome given a hypothesis (not the probability that the hypothesis is true)
2 If the p-value is very low, we have evidence against the null hypothesis, but we do not compute the probality of the null hypothesis being false
3 Bayesian statistics allows for computing the probability of an hypothesis, but it requires prior knowledge of the probability before data collection, making it difficult in many situations.

Summary: Understanding the Null Hypothesis in Hypothesis Testing
Definition of the Null Hypothesis (H_0).
• The null hypothesis states that any observed effect in a study is due to chance rather than a real difference or relationship.
• It is typically the opposite of the researcher's hypothesis and serves as the baseline assumption

Examples of the Null Hypothesis
1 Physicians' Reactions Study
• Researchers tested whether physicians spent less time with obese patients.

Date / /
Page No.

- The null hypothesis (H_0): The mean time spent with obese patients is equal to the mean time spent with average-weight patients.

$$\mu_{obese} = \mu_{average}$$

- The alternative hypothesis (H_1): The time spent is not equal (or specifically less for obese patients)

$$\mu_{obese} < \mu_{average}$$

- Since the observed difference (6.7 minutes) was unlikely due to chance, the researchers rejected the null hypothesis and concluded that physicians spend less time with obese patients.

2 Correlation Between High school and college Grades

- The null hypothesis: There is no relationship between high school and college grades

$$\rho = 0$$

- If the correlation in the sample is strong and statistically significant, the null hypothesis is rejected, suggesting a real relationship exists in the population.

3 Coin flip experiment

- If testing whether someone can predict coin flips better than chance, the null hypothesis states that their accuracy is just random guessing $\pi = 0.5$

- If their accuracy significantly exceeds 50%, the null hypothesis is rejected, suggesting they may have real predictive ability.

Date / /
Page No.

Summary : Type I and Type II Errors in Hypothesis Testing

Understanding Type I and Type II Errors

In hypothesis testing, there are two possible errors when making decisions about the null hypothesis (H_0):

1. Type I Error (False Positive) : Rejecting a true null hypothesis.
- Example : In the Physicians' Reactions study researchers concluded that physicians spend less time with obese patients
- However, if no real difference exists and the observed difference was due to chance, rejecting H_0 would be a Type I error
- The probability of a Type I error is called α (alpha), also known as the significance level (commonly set at 0.05 or 0.01)
- A lower α reduces Type I errors, but increases the risk of Type II errors

2. Type II Error (False Negative) : Failing to reject a false null hypothesis.
- Example : Suppose there is a real difference in the time spent with obese patients, but the study does not detect it
- This failure to reject H_0 when it is false is a Type II error
- The probability of a Type II error is called β (Beta)
- The probability of correctly rejecting a false null hypothesis is $1-\beta$, known as statistical power.

Date / /
Page No.

Summary: One-Tailed vs Two-Tailed Tests in Hypothesis Testing

1) Understanding one-tailed and two-tailed test

One-tailed test: Tests if an effect is in one specific direction (eg better than chance)

Two-tailed test: Tests for an effect in either direction (eg significantly different from chance, whether higher or lower)

2) Example: James Bond case study

- Bond's performance: correct 13 out of 16 times in distinguishing martinis.
- One-tailed probability $P = 0.0106$ (only considering extreme success)
- Two-tailed probability $p = 0.0212$ (considering both success and failure extremes)

Use a two-tailed test if:

- We want to test whether he performs significantly different from chance, in either direction. Use a one-tailed test if we only care about whether he is better than chance (not worse).

3) Null and Alternative Hypotheses

Test Type	Null Hypothesis (H_0)	Alternative Hypothesis (H_A)
Two-tailed	$\pi = 0.5$	$\pi \neq 0.5$
One-tailed	$\pi \leq 0.5$	$\pi > 0.5$

4) When to Use a One-Tailed vs Two-Tailed Test

Use a one-tailed test when

- We only care about one direction (eg a new drug must be better than placebo, not worse)
- An effect in the opposite direction would not matter (eg predicting only improvement in performance)