HYPOTHESIS TESTING

A PROCEDURE OF STATISTICAL TESTING

A.1 LOGIC OF STATISTICAL TESTING

Statistical testing functions much like a court trial.

- **The Assumption (Innocence):** We start by assuming there is no effect or no difference until proven otherwise.
- **The Evidence (Data):** We collect data from a sample.
- **The Verdict:** If the evidence is strong enough (beyond reasonable doubt), we reject the initial assumption.

Important: In hypothesis testing, we never "prove" H_0 is true. We either *reject* H_0 or *fail to reject* it because the evidence is not strong enough.

Definition **Null and Alternative Hypotheses**

- The **Null Hypothesis** (H_0) represents the status quo, "no difference," or "no effect." It always includes an equality ($\mu = k$, $\mu \leqslant k$, or $\mu \geqslant k$).
- The **Alternative Hypothesis** (H_1) is the claim we are trying to find evidence for. It uses strict inequalities ($\mu \neq k$, $\mu > k$, or $\mu < k$).

Ex: A company claims that the average battery life of their new phone is 24 hours. A consumer group suspects the battery life is actually shorter. Write the null and alternative hypotheses.

Answer: Let μ be the true mean battery life.

- $H_0 : \mu = 24$ (The company's claim is true; there is no difference).
- $H_1 : \mu < 24$ (The consumer group's suspicion; the mean is less than claimed).

A.2 P-VALUE AND SIGNIFICANCE LEVEL

Definition **p-value and α**

- The **significance level** (α) is the threshold for evidence (usually 0.05, 0.01, or 0.10). It is the probability of rejecting H_0 when it is actually true (a Type I error).
- The **p-value** is the probability of obtaining sample results at least as extreme as the ones observed, assuming H_0 is true.

A *small* p-value means the data would be very unlikely if H_0 were true, so it is evidence against H_0.**Decision Rule:**

- If p-value $\leqslant \alpha \implies$ **Reject** H_0. (The result is statistically significant).
- If p-value $> \alpha \implies$ **Fail to reject** H_0. (Not enough evidence against H_0).

Ex: A researcher performs a hypothesis test to check if a new fertilizer increases plant growth.
The calculated p**-value is** 0.042.
Make a conclusion for the following significance levels:

1. $\alpha = 0.05$ (5% significance level).
2. $\alpha = 0.01$ (1% significance level).

Answer:

1. **For $\alpha = 0.05$:**
$$p\text{-value} = 0.042 < 0.05$$
Since the p-value is smaller than α, the result is statistically significant at the 5% level. We **reject the null hypothesis** (H_0). There is sufficient evidence at the 5% significance level to support the claim.
2. **For $\alpha = 0.01$:**
$$p\text{-value} = 0.042 \geqslant 0.01$$
Since the p-value is greater than α, the result is not statistically significant at this level. We **fail to reject the null hypothesis** (H_0). There is not enough evidence at the 1% significance level to support the claim.

A.3 5-STEP PROCEDURE

Method Performing a Hypothesis Test

1. **State the Hypotheses:** Define the parameter (e.g., μ) and write H_0 and H_1.
2. **State the Test and Level:** Identify the test (e.g., t-test) and the significance level α.
3. **Calculate Statistics:** Use the GDC to find the test statistic and the **p-value**.
4. **Compare:** Explicitly compare the p-value to α (usually check whether $p \leqslant \alpha$ or $p > \alpha$).
5. **Conclude:** Write a conclusion in the context of the problem, referring to the original claim.

Ex: A coffee machine is supposed to dispense 250 ml per cup. A manager suspects it is dispensing **less**. He measures a sample of 10 cups and finds a mean of $\bar{x} = 248$ ml with a standard deviation of $s_{n-1} = 3$ ml.
Test the manager's suspicion at the 5% significance level.

Answer: Let μ be the population mean volume of coffee.

1. $H_0 : \mu = 250$ and $H_1 : \mu < 250$.
2. One-sample t-test at $\alpha = 0.05$.
3. Using GDC (T-Test with $\mu_0 = 250, \bar{x} = 248, s = 3, n = 10, < \mu_0$):
   $t \approx -2.108$ and p-value ≈ 0.032.
4. Since $0.032 < 0.05$, we reject H_0.
5. There is sufficient evidence to suggest the machine is dispensing less than 250 ml.

A.4 TYPE I AND TYPE II ERRORS

Definition Error Types

When making a decision based on a statistical test, there is always a risk of error.

- **Type I Error (α):** Rejecting H_0 when H_0 is actually true (False Positive). The probability of this error is the significance level α.
- **Type II Error (β):** Failing to reject H_0 when H_0 is actually false (False Negative).

	H_0 is True	H_0 is False
Reject H_0	Type I Error	Correct Decision
Fail to Reject H_0	Correct Decision	Type II Error

Ex: A factory produces parachutes. The null hypothesis states that a batch of parachutes is safe.

- H_0: The parachutes are safe.
- H_1: The parachutes are defective.

Describe the Type I and Type II errors. Which error is more dangerous in this context?

Answer:

- **Type I Error:** The inspector concludes the parachutes are defective (rejects H_0) when they are actually safe (H_0 is true). Consequence: Financial loss from stopping production or destroying good products.
- **Type II Error:** The inspector concludes the parachutes are safe (fails to reject H_0) when they are actually defective (H_0 is false). Consequence: Potentially fatal accidents.

In this context, a **Type II error is much more dangerous** because human lives are at risk.

B t-TEST

The **t-test** is one of the most commonly used statistical tests. It is used to determine if there is a significant difference between means, especially when the population variance is unknown and the sample size is small ($n < 30$).

B.1 ONE-SAMPLE T-TEST

Definition One-Sample t-test Formula

The **one-sample t-test** compares the mean of a single sample ($\bar{x}$) to a known or hypothesized population mean (μ_0). It is used when:

- The data are quantitative (continuous).
- The population follows a normal distribution (or the sample size is large, $n \geqslant 30$).
- The population standard deviation σ is **unknown** (we use the sample standard deviation s).

The test statistic t is calculated as:

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

Method Steps for a One-Sample t-test using Calculator

1. **Step 1: Write the Hypotheses**
   - $H_0 : \mu = k$ (The population mean equals a specific value k).
   - $H_1 : \mu \neq k$ (or $\mu < k$, $\mu > k$ depending on the question).
2. **Step 2: Enter data into GDC**
   - Enter the data into List 1 (or use summary statistics $\bar{x}, s, n$).
   - Select **T-Test** (or 1-Sample t-test).
   - Enter the value of μ_0 (from H_0).
3. **Step 3: Decision Rule**
   Compare the p-value with the significance level α.
   - If p-value $< \alpha$: **Reject** H_0.
   - If p-value $\geqslant \alpha$: **Fail to reject** H_0.
4. **Step 4: Conclusion**
   State whether there is sufficient evidence to support the alternative hypothesis, in context.

Ex: A factory produces screws with a target length of 50 mm. A quality control manager takes a random sample of 15 screws and finds a mean length of 49.8 mm with a standard deviation of 0.5 mm.
Conduct a t-test at the 5% significance level to see if the mean length is different from 50 mm.

Answer:

1. **Step 1: Hypotheses**
   - $H_0 : \mu = 50$ (The mean is 50 mm).
   - $H_1 : \mu \neq 50$ (The mean is different from 50 mm).
2. **Step 2: Calculator**
   Using GDC with inputs: $\mu_0 = 50, \bar{x} = 49.8, s_x = 0.5, n = 15$.

$$t \approx -1.549$$

$$p\text{-value} \approx 0.143$$

3. **Step 3: Decision**
   $0.143 > 0.05$. Since $p > \alpha$, we **fail to reject** H_0.
4. **Step 4: Conclusion**
   There is insufficient evidence at the 5% level to claim that the mean length of the screws is different from 50 mm.

B.2 TWO-SAMPLE T-TEST (INDEPENDENT)

Definition Two-Sample t-test Formula

The **two-sample t-test** compares the means of two **independent** groups to see if they are significantly different.The test statistic is given by:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

Where:

- $\bar{x}_1, \bar{x}_2$ are the sample means.
- s_1^2, s_2^2 are the sample variances.
- n_1, n_2 are the sample sizes.

Method Steps for a Two-Sample t-test using Calculator

1. **Step 1: Write the Hypotheses**
   - $H_0 : \mu_1 = \mu_2$ (The population means are equal).
   - $H_1 : \mu_1 \neq \mu_2$ (or $\mu_1 < \mu_2$, $\mu_1 > \mu_2$).
2. **Step 2: Enter data into GDC**
   - Enter the data into List 1 and List 2 (or use summary statistics $\overline{x_1}, s_1, n_1, \overline{x_2}, s_2, n_2$).
   - Select **2-Sample t-test**.
   - **Pooled Setting:**
     - Choose **No** (Default): This assumes variances are *not* necessarily equal. It uses the formula involving $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$. Use this unless told otherwise.
     - Choose **Yes**: Only use this if the question explicitly states to "assume the population variances are equal".
3. **Step 3: Decision Rule**
   Compare the p-value with the significance level α.
   - If p-value $< \alpha$: **Reject** H_0.
   - If p-value $\geqslant \alpha$: **Fail to reject** H_0.
4. **Step 4: Conclusion**
   State whether there is sufficient evidence to support the alternative hypothesis.

Ex: A teacher wants to compare the effectiveness of two teaching methods. She randomly assigns 10 students to Method A and 12 students to Method B. The results of the final test are shown below:

- **Method A:** $75, 82, 90, 65, 88, 92, 78, 85, 70, 80$
- **Method B:** $60, 72, 68, 75, 62, 80, 70, 65, 78, 66, 74, 69$

Test at the 5% significance level whether there is a difference in the mean scores of the two methods.

Answer:

1. **Step 1: Hypotheses**
   - $H_0 : \mu_A = \mu_B$ (The mean scores are equal).
   - $H_1 : \mu_A \neq \mu_B$ (The mean scores are different).
2. **Step 2: Calculator**
   Enter the data into List 1 and List 2. Select **2-Sample t-test**.
   Since the question **does not** state that variances are equal, we choose **Pooled: No**.

$$\bar{x}_A = 80.5, \quad \bar{x}_B \approx 69.92$$

$$t \approx 3.19$$

$$p\text{-value} \approx 0.0057$$

3. **Step 3: Decision**
   $0.0057 < 0.05$. Since $p < \alpha$, we **reject** H_0.

4. **Step 4: Conclusion**
   There is sufficient evidence at the 5% level to suggest that the mean scores of the two methods are different.

B.3 PAIRED T-TEST

Definition Paired t-test

The **paired t-test** compares the means of two **dependent** groups (e.g., "Before and After" measurements on the *same* subject or matched pairs).
It is used when:

- The data consist of matched pairs (x_1, x_2).
- The differences $d = x_2 - x_1$ (or $x_1 - x_2$) are calculated.
- The differences follow a normal distribution.

The test is essentially a **one-sample t-test** performed on the differences d, testing if the mean difference μ_d is zero.The test statistic is:

$$t = \frac{\bar{d} - 0}{s_d/\sqrt{n}}$$

Method Steps for a Paired t-test

1. **Step 1: Calculate Differences**
   Calculate the difference for each pair: $d_i = x_{2i} - x_{1i}$ (be consistent with the chosen order).

2. **Step 2: Write the Hypotheses**
   Let μ_d be the population mean of the differences.
   - $H_0 : \mu_d = 0$ (No difference on average).
   - $H_1 : \mu_d \neq 0$ (or $\mu_d < 0$, $\mu_d > 0$).

3. **Step 3: Enter data into GDC**
   - Enter the calculated differences into List 1 (or put raw data in L1, L2 and define L3 = L2 - L1).
   - Select **T-Test** (1-Sample t-test) on the differences list.
   - Set $\mu_0 = 0$.

4. **Step 4: Decision and Conclusion**
   Compare the p-value to α and conclude in context.

Ex: A weight-loss program claims to reduce weight after one month. The weights of 5 participants are recorded before and after the program.

Participant	1	2	3	4	5
Before (kg)	80	95	88	102	90
After (kg)	78	94	85	100	91

Test at the 5% level if the program effectively reduces weight.

Answer:

1. **Calculate Differences (d = After − Before):**

$$d = \{78 - 80, 94 - 95, 85 - 88, 100 - 102, 91 - 90\}$$

$$d = \{-2, -1, -3, -2, 1\}$$

2. **Hypotheses:**
   Let μ_d be the mean difference.
   - $H_0 : \mu_d = 0$ (The program has no effect).
   - $H_1 : \mu_d < 0$ (The program reduces weight on average).

3. **Calculator:**
Perform a **1-Sample T-Test** on the list of differences $\{-2, -1, -3, -2, 1\}$ with $\mu_0 = 0$ and test $< \mu_0$.

$$\bar{d} = -1.4, \quad s_d \approx 1.517$$

$$t \approx -2.06$$

$$p\text{-value} \approx 0.0549$$

4. **Conclusion:**
$0.0549 > 0.05$. Since $p > \alpha$, we **fail to reject** H_0.
There is insufficient evidence at the 5% level to conclude that the program effectively reduces weight (although the mean difference is negative, the sample is small and the result is not statistically significant).

C CHI-SQUARED TEST (χ^2)

C.1 CHI-SQUARED TEST FOR INDEPENDENCE

The **Chi-squared** (χ^2) **test for independence** determines if there is a significant association between two categorical variables. It compares the observed frequencies in a contingency table to the frequencies we would expect if the variables were completely independent.

Definition Observed and Expected Frequencies

- **Observed Frequencies (f_o):** The actual data collected and recorded in a contingency table.
- **Expected Frequencies (f_e):** The theoretical counts calculated assuming the variables are independent.

The formula for the expected frequency of a cell is:

$$f_e = \frac{\text{Row Total} \times \text{Column Total}}{\text{Grand Total}}$$

Definition Chi-squared Statistic

The test statistic χ^2_{calc} measures the total deviation between observed and expected values:

$$\chi^2_{calc} = \sum \frac{(f_o - f_e)^2}{f_e}$$

The **degrees of freedom** (df) for a table with r rows and c columns is:

$$df = (r-1)(c-1)$$

Method Steps for a Chi-squared Test

1. **Step 1: Write the Hypotheses**
   - H_0: The variables are **independent**.
   - H_1: The variables are **not independent** (associated).
2. **Step 2: Enter data into GDC**
   - Enter the observed contingency table into a **Matrix** (e.g., Matrix A).
   - Select χ^2**-Test** (usually under Stat → Tests).
3. **Step 3: Analyze Results**
   The calculator provides χ^2_{calc}, the p-value, and the degrees of freedom (df).
   Note: As a rule of thumb, all expected frequencies should be at least 5 for the test to be reliable.
4. **Step 4: Conclusion**
   Compare the p-value to the significance level α.
   - If p-value $< \alpha$: **Reject** H_0 (Variables are dependent/associated).
   - If p-value $\geqslant \alpha$: **Fail to reject** H_0 (No evidence of association).

Ex: A survey asked 200 people about their preferred type of movie and their age group. The results are shown below:

	Action	Comedy	Drama
Under 30	40	35	15
30 and over	20	45	45

Test at the 5% significance level whether age group and movie preference are independent.

Answer:

1. **Step 1: Hypotheses**
   H_0: Age group and movie preference are independent.
   H_1: Age group and movie preference are not independent (they are associated).
2. **Step 2: Calculator**
   Enter the 2×3 matrix into the calculator and run the χ^2-Test.
   - $\chi^2_{calc} \approx 21.1$
   - $df = (2-1)(3-1) = 2$
   - p-value ≈ 0.000026 (about 2.6×10^{-5})
3. **Step 3: Conclusion**
   $0.000026 < 0.05$. Since $p < \alpha$, we **reject** H_0.
   There is strong evidence to suggest that movie preference depends on age group.

C.2 χ^2 GOODNESS OF FIT TEST

The **Goodness of Fit (GOF)** test is used to determine whether a variable is likely to come from a specified distribution (such as Uniform, Binomial, Normal, or a specific ratio). It compares the observed data with what we would expect theoretically.

Definition Goodness of Fit Statistic

The test uses the same χ^2 statistic formula:

$$\chi^2_{calc} = \sum \frac{(f_o - f_e)^2}{f_e}$$

However, the **degrees of freedom (df)** calculation depends on the distribution:

$$df = k - 1 - m$$

Where:

- k is the number of categories (bins).
- m is the number of population parameters estimated from the sample data (e.g., if you calculate mean and standard deviation from the sample to fit a Normal distribution, $m = 2$).

Method Steps for a GOF Test

1. **Step 1: Hypotheses**
   - H_0: The data follow the specified distribution.
   - H_1: The data do not follow the specified distribution.
2. **Step 2: Expected Frequencies**
   Calculate the expected frequency for each category:
   $$f_e = n \times P(\text{category})$$
   Note: Usually done in List 2 of the GDC.
3. **Step 3: Calculator**
   - Enter observed values in List 1 (f_o).
   - Enter expected values in List 2 (f_e).
   - Select χ^2 **GOF Test**.
   - Enter the correct df.
4. **Step 4: Conclusion**
   Compare p-value to α and interpret in context.

Ex: A die is rolled 60 times. The results are:

Outcome	1	2	3	4	5	6
Frequency	8	12	15	9	10	6

Test at the 5% level if the die is fair (Uniform distribution).

Answer:

1. **Hypotheses:**
   H_0: The die is fair (Uniform distribution).
   H_1: The die is not fair.

2. **Expected Frequencies:**
   Total $n = 60$. If fair, $P(X = k) = 1/6$.

$$f_e = 60 \times \frac{1}{6} = 10 \quad \text{for all outcomes}$$

3. **Calculator / Working:**
   $L_1 : \{8, 12, 15, 9, 10, 6\}$
   $L_2 : \{10, 10, 10, 10, 10, 10\}$

$$\chi^2_{calc} = \sum \frac{(f_o - f_e)^2}{f_e} = \frac{(8-10)^2}{10} + \frac{(12-10)^2}{10} + \cdots + \frac{(6-10)^2}{10} = 5.0$$

   Degrees of freedom: $df = k - 1 = 6 - 1 = 5$ (no parameters estimated).
   p-value ≈ 0.416.

4. **Conclusion:**
   $0.416 > 0.05$. We fail to reject H_0. The die appears to be fair.