- H_0: Null hypothesis. The hypothesis is that there is no association, no effect or no difference.
- H_A: Alternative hypothesis. The hypothesis is that there is an association, effect or difference.
- **P-value**: We measure the "consistency" of the observed data with the claim using a p-value. The p-value is the probability of observing the value of the test statistic, or a value more extreme, calculated under the assumption that H_0 is true. A small p-value indicates we would be unlikely to see the data we did if the null hypothesis were true. i.e., the smaller the p-value is, the easier to reject H_0. If the p-value is less than α we reject H_0. If the p-value is greater than or equal to α we do not reject H_0.
- **Test statistic (t - statistic)**: A test statistic is the standardised value of the sample value.
$$T = \frac{observed\ sample\ value - null\ value}{estimated\ standard\ error}$$
When studying hypothesis testing, we usually use π to denote the value of the overall parameter, instead of using p.
- **Z Statistic vs t-statistic**:
  - Z-statistic: When to use: Large samples (n > 30) or known population standard deviation (σ). Basis: Uses population standard deviation (σ).
  - t-statistic: When to use: Small samples (n < 30) or unknown population standard deviation (σ). Basis: Uses sample standard deviation (s).
- **How to conduct a hypothesis test**:
  1. Set H_0 and H_A.
  2. Calculate t-statistic(if it is sample) or Z-statistic(if it is population).
  3. Calculate p-value.
  4. Calculate 95% CI with (3).
  5. Draw a conclusion based on the p-value (reject H_0 or not).
- **The difference between p and p***:
  - p (Sample Proportion): Frequency or proportion of events in a sample. Example: If 50% support a policy in a sample, (p = 0.5).
  - p* (Estimate of Overall Proportion): Estimate of the overall proportion based on sample data. Example: If 50% support a policy in a sample, (p* = 0.5).
- **When to use chi square**: When trying to control groups in experiments - looking for differences between men and women in each group, etc. Looking for differences between categorical variables - maybe you want to know if there is a difference between men and women for their favourite type of ice cream.
- **How to conduct chi-square test**:

1. Define the Null-Hypothesis and Alternative Hypothesis. H_0 : The treatment and response are independent (i.e. no association). H_A : The treatment and response are dependent in some way (i.e. there is some association).
2. Calculating expected cell counts.
3. Calculating the χ^2 test statistic.
4. Get the degree of freedom.
5. Calculate the p-value with R.
6. Reject / not reject H_0.
7. Draw conclusion.

- **Expected cell counts**:

$$E_{(row\ i, col\ j)} = \frac{r_i \times c_j}{n}$$

  - r_i is the row total, for row i
  - c_j is the column total, for column j
  - n is the total number (of trials, patients, etc.)

  We worked out what we would have expected to see under the null hypothesis in each cell given the observed row and column totals.

- **Formula for chi-square:**

$$\chi^2 = \sum_{ij} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

- **Degree of freedom for chi square**: $\nu = (number of rows - 1) * (number of columns - 1)$
- **Range for p-value**: (0, 1)
- **Belief (interpretation/decision):**

	Fail to reject H_0	Reject H_0
Null is true	Correct interpretation (No error)	Error (Type I)
Null is false	Error (Type II)	Correct interpretation (No error)

- **Type I error (a false positive result)**: Concluding that there is an association between exposure and outcome, where there is not. Type I error is controlled when we set the significance level (usually 0.05).
- **Type II error (a false negative result)**: Concluding that there is not an association between exposure and outcome, where there is. Type II error is primarily controlled through the sample size. Ideally, power should be between 80 and 90%.