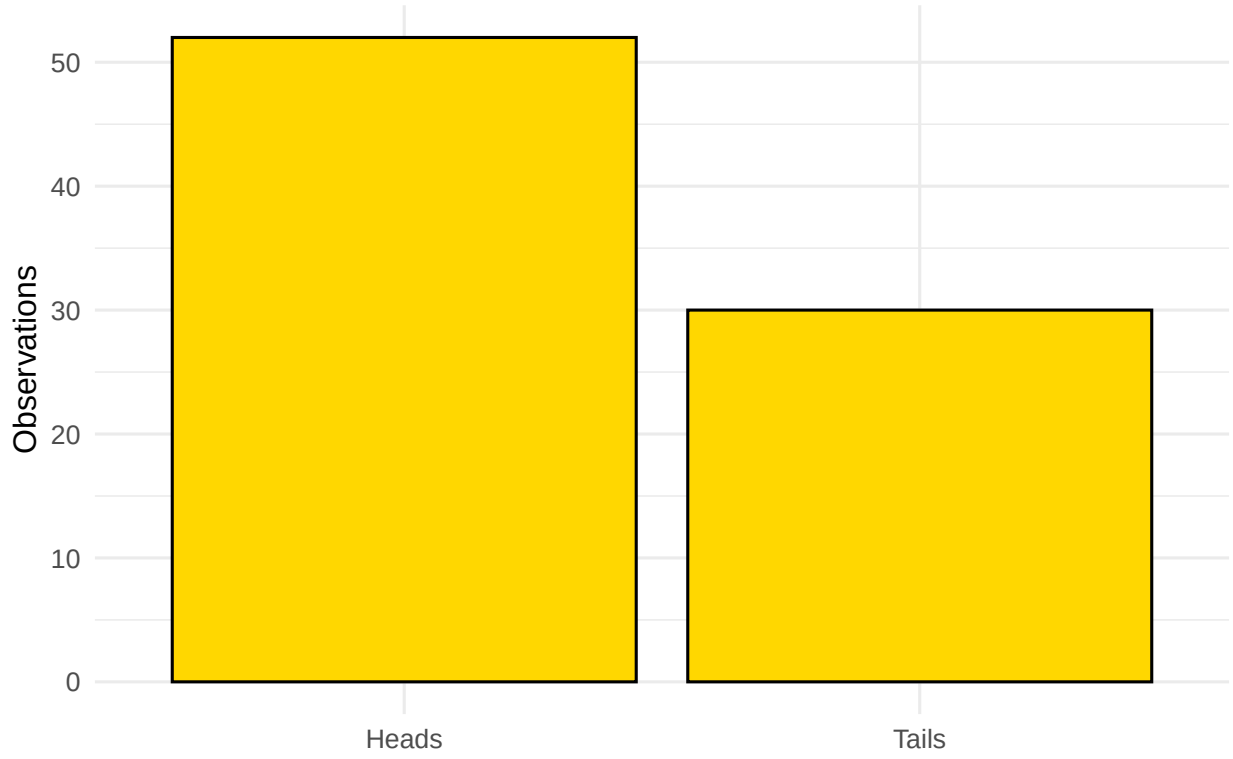
Chi-Squared Test

What is the Chi-Squared Test?

The Chi-Squared Test is a statistical test used to determine whether there is a significant association between categorical variables. It is very useful in answering the question - "Is the frequency of these observations unusual or due to chance?".



Is my coin biased?

When to use a Chi-Squared test?

Use a Chi-Squared test when your data consist of **categorical** variables and you want to determine whether there is a significant relationship between them or whether the observed frequencies differ from what you would expect by chance.

Examples could include; testing whether a dice is fair, comparing patient satisfaction across clinics and frequency of dental conditions across age groups.

Assumptions of the Chi-Squared test?

The assumptions for using the Chi-Squared test are as follows:

1) Categorical data
2) Expected cell counts are sufficiently large (At least ten coin flips)
3) Independence of observations
4) Each observation must belong to only one category and must not influence others.
5) Random sampling

Different types of Chi-Squared test

Type of test: Chi-Squared goodness-of-fit test

Example: Testing whether a coin is fair by comparing observed counts of heads and tails to the expected 50/50 distribution.

Type of test: Chi-Squared test of homogeneity

Example: Comparing smoking status (non-smoker, light, heavy) across three different regions.

Type of test: Fisher's exact test

Example: Same as the Chi-Squared goodness-of-fit test but with lower sample sizes.

How to implement in RStudio

```r
# Chi-Squared Test of Independence
chisq.test(table(df$categorical_var1, df$categorical_var2))
```

Interpreting results

Interpreting Chi-Squared results involves examining whether the observed frequencies differ from the expected frequencies under the assumption of no association.

The p-value indicates how likely it is to observe your data if the two variables are actually independent. A low p-value (commonly < 0.05) suggests that the observed pattern of counts would be very unlikely if the variables were unrelated.

The Chi-Squared Statistic (X^2) tells you how far your observed counts are from what would be expected by chance. Larger values means a bigger deviation from expectation and more evidence of a relationship.