Statistical inference draws a conclusion about a population parameter based on a sample statistic. The two major procedure types are confidence intervals and hypothesis tests.

- A confidence interval estimates the value of a parameter with a range of values.
- A hypothesis test assesses the plausibility of a particular claim about the parameter.

The elements of a statistical hypothesis test are:
- The null hypothesis, denoted by $H_0$, is the claim being tested. The purpose of the test is to assess the evidence against the null hypothesis, which is typically a statement of “no effect” or “no difference.” The null hypothesis is a statement about a population parameter.
- The alternative hypothesis, denoted by $H_a$, is what the researchers suspect to be true, as opposed to the null hypothesis, about the population parameter.
- The p-value is the probability, if the null hypothesis were true, that the sample data would be as or more extreme as the one actually observed.
  - The smaller the p-value, the stronger the evidence against the null hypothesis.
  - A p-value can be calculated through simulation or from a sampling distribution.

Example 13-1: Baseball Big Bang
A reader wrote in to the “Ask Marilyn” column in Parade magazine to say that his grandfather told him that in $\frac{3}{4}$ of all baseball games, the winning team scores more runs in one inning than the losing team scores in the entire game. (This phenomenon is known as a “big bang.”) Marilyn responded that this $\frac{3}{4}$ value seemed to be too high to be believable. Let $\pi$ denote the proportion of all Major League Baseball games in which a “big bang” occurs.

a) Restate the grandfather’s assertion as the null hypothesis, in symbols and in words.

b) Report Marilyn’s response as an alternative hypothesis, in symbols and in words.

To investigate this claim, I selected a random sample of 95 games played in the 2006 Major League Baseball season. I examined the box scores of those games to determine which had a big bang and which did not.

c) Specify and sketch the sampling distribution of the sample proportion $\hat{p}$ of games with a big bang, assuming that the null hypothesis is true. Also check whether the conditions required for this result hold here.
Of the 95 games in my random sample, 47 contained a big bang.

d) Determine the sample proportion of games with a big bang. Use the appropriate symbol to denote it.

e) Calculate and interpret the \( z \)-score corresponding to this sample proportion, assuming that \( \frac{3}{4} \) of all games have a big bang.

f) Use the sampling distribution to determine the (approximate) probability that the sample proportion \( \hat{p} \) of games with a big bang would be (answer to d) or smaller in a random sample of 95 games if in fact \( \frac{3}{4} \) of all games have a big bang.

g) By what term is this probability in e) known?

h) Is this p-value small enough to convince you that the grandfather is wrong that \( \frac{3}{4} \) of all games have a big bang, and Marilyn is right that the true proportion is less than \( \frac{3}{4} \)? Explain the reasoning process underlying your answer.

More terminology related to hypothesis testing:
- A **test statistic** measures how far the observed sample result falls from the hypothesized value of the parameter.
  - Here the test statistic is the \( z \)-score associated with the sample proportion \( \hat{p} \), assuming the null hypothesis to be true.
- The p-value is always calculated in the direction of the alternative hypothesis.
  - Here the p-value is the area to the left of the test statistic \( z \) under the standard normal curve.
  - If the alternative had been in the > direction, then the p-value would have been the area to the right of the test statistic \( z \) under the standard normal curve.
- A **significance level** \( \alpha \) is a cut-off for how small the p-value must be in order for the sample data to be considered decisive. The most common value is .05, followed by .01 and .10.
  - If the p-value is less than or equal to \( \alpha \), then we **reject** the null hypothesis.
    - If the p-value is greater than \( \alpha \), then we **fail to reject** the null hypothesis.
  - If the p-value is less than or equal to \( \alpha \), then the sample result is said to be **statistically significant** at the \( \alpha \) level.
i) At the .01 significance level, is the sample proportion of games with a big bang statistically significantly less than \( \frac{3}{4} \)? Explain.

j) At the .001 significance level, would you reject the null hypothesis that \( \frac{3}{4} \) of all baseball games have a big bang? Explain.

**Example 17-2: Halloween Treats**

Stemming from concern over the nation’s obesity epidemic, researchers investigated whether children might be as tempted by toys as by candy for Halloween treats. Test households in five Connecticut neighborhoods offered children two bowls: one with lollipops or fruit candy and one containing small, inexpensive Halloween toys, like plastic bugs that glow in the dark.

a) State the appropriate null and alternative hypotheses. Be sure to explain in words what the parameter represents here.

- Because we do not have a specific direction in mind for the parameter in this case. We therefore use a **two-sided alternative** instead of a **one-sided alternative**.

Of the 283 children with ages from 3 to 14 whose reactions were observed in this study, 148 chose candy and 135 chose toys.

b) What proportion of the children in the sample chose a toy? What symbol do we use for this value?

c) Check whether the CLT conditions are satisfied in this study, assuming the null hypothesis to be true.

d) Calculate the value of the test statistic.

- To determine the \( p \)-value with a two-sided alternative, we find the probability of being as extreme as this sample result in either direction. In other words, we add the areas in the two tails of the normal distribution.
e) Calculate the two-sided p-value.

f) Write a sentence interpreting what this p-value says (probability of what, assuming what?).

g) At the .05 significance level, would you reject the null hypothesis? Explain the reasoning behind this decision.

h) Summarize the conclusions that you would draw from this study.

- Failing to reject a null hypothesis is not equivalent to accepting the null hypothesis.

i) Use Minitab (Stat> Basic Statistics> 1 Proportion…) or the Theory-Based Inference applet to verify your calculations.

**Summarizing the components of a hypothesis test about a population proportion:**
- The null hypothesis is $H_0: \pi = \text{hypothesized value}$
- The alternative hypothesis is $H_a: \pi > \text{hypothesized value}$ or $H_a: \pi < \text{hypothesized value}$ or $H_a: \pi \neq \text{hypothesized value}$.
- The test statistic is $z = \frac{\hat{p} - \text{hypothesized value}}{\sqrt{\frac{\text{hypothesized value}(1 - \text{hypothesized value})}{n}}}$.
- The p-value is the area under the standard normal curve more extreme than the test statistic, in the direction indicated by the alternative hypothesis.
- The test decision is to reject $H_0$ whenever p-value $< \alpha$ (significance level).
- The technical conditions required for this procedure to be valid are:
  - that the data can be regarded as a random sample from the population
  - $n \times \text{hypothesized value} \geq 10$ and $n \times (1 - \text{hypothesized value}) \geq 10$

Here are some criteria for evaluating strength of evidence based on a p-value:
- A p-value of .10 or less provides *some* evidence against the null hypothesis.
- A p-value of .05 or less provides *fairly strong* evidence against the null hypothesis.
- A p-value of .01 or less provides *very strong* evidence against the null hypothesis.

**Example 17-3: Cat Households**
A sample survey of 80,000 households in 2001 found that 31.6% of American households own a pet cat.
a) What are the observational units and variable here? What kind of variable? Is this number (31.6\%) a parameter or a statistic? Explain, and indicate the symbol used to represent it.

b) Conduct a test of whether the sample data provide evidence that the population proportion of all American households that own a pet cat differs from one-third. (Feel free to use Minitab or an applet.) State the hypotheses, and report the test statistic and p-value. Also indicate the test decision at the .01 significance level, and summarize your conclusion in the context of this study.

c) Produce a 99.9\% CI for the population proportion of all American households that own a pet cat. Interpret this interval.

d) Are the test decision and confidence interval consistent with each other? Explain.

e) Do the sample data provide very strong evidence that the population proportion who own a pet cat is not one-third? Explain whether the p-value or the CI helps you to decide.

f) Do the sample data provide strong evidence that the population proportion who own a pet cat is very different from one-third? Explain whether the p-value or the CI helps you to decide.

- This example reveals the distinction between statistical significance and practical importance.
  - A hypothesis test reveals whether an observed difference is “statistically significant,” meaning that it is unlikely to have occurred by chance.
    - A statistically significant result may or may not be practically important.
    - Especially with large sample sizes, even a very small difference can be statistically significant.
  - A confidence interval estimates the magnitude of the difference, helps to judge practical importance.