In statistics, a **hypothesis** is a claim or statement about a property of the population.

A **hypothesis test** is a statistical method for testing a hypothesis.

**Examples:**

Medical researchers claim that the mean body temperature of healthy adults is 98.6°F.

The average lifespan of Macbook Pro laptops is 10 years.

The percentage of people with green eyes is less than half.

**Rare Event Rule for Inferential Statistics**

If, under a given assumption, the probability of a particular observed event is exceptionally small, we conclude that the assumption is probably not correct.
Example 1: The cookie company claims the average cookie weight is 11g. If we observe a bag of 32 cookies has an average cookie weight of 10.9g, is our sample mean just random variation from a population mean, or did we get ripped off? Suppose that the sd is known to be 0.5g.

6-Step Method for hypothesis tests

1. State the hypothesis (always in terms of population parameters; steps 1-3 in book)
2. Determine the level of significance (0.05 unless otherwise specified)
3. Determine the test statistic (something that can be looked up in a table)
4. Compute the test statistic and either the critical region or the p-value
5. Reject or fail to reject
6. State conclusions in the context of the original problem
1) Claim: either $\mu = 11$ or $\mu < 11$

★ We use the one with the equality as the **null hypothesis**, and the other as the **alternative hypothesis**.
★ The null is the default (what we assume to be true).
★ We only conclude the alternative if there is enough evidence.
  i.e. Like in a trial the defendant is innocent until proven guilty.
★ Failure to reject the null is a lack of evidence; it does **NOT** mean the null is necessarily true.

◇ If we are trying to **prove that something is true**, we must make it our **alternative** hypothesis.
◇ If we are just trying to **show that an assumption is reasonable**, we can make it the **null** hypothesis.
★ If we show that the null is NOT reasonable, we can conclude the alternative is **HIGHLY LIKELY** to be true.
Write the hypothesis in terms of population parameters, and define them!!!

So, 

\( H_0 : \mu = 11 \) (claim)

\( H_1 : \mu < 11 \)

where \( \mu \) is the population mean cookie weight

2) Level of significance is \( \alpha = 0.05 \) unless otherwise specified

3) Test statistic: here we are testing a mean with \( \sigma \) known, so

\[
z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}
\]

(normal)

If \( \sigma \) is unknown, use

\[
t = \frac{\bar{x} - \mu}{s / \sqrt{n}}
\]

(t with n-1 degrees of freedom)

For a proportion, use

\[
z = \frac{\hat{p} - p}{\sqrt{p(1 - p)/n}}
\]

(normal)
4) The **critical region** is the set of values of the test statistic that would cause us to reject the null hypothesis - those values that would be highly unusual if the null were true.

- Unusual is defined by the **level of significance**, the probability we are willing to be wrong when the null is true.
- If the null is true, then \( z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} \) follows the standard normal distribution.
- For the cookie example, if our sample mean is too small, we will reject the null. How small is too small?

\[
\rightarrow P(z < -1.645) = 0.05 \text{ (from z table)}
\]

So, we will reject the null hypothesis if \( z < -1.645 \). The actual test statistic is \( z = \frac{10.9 - 11}{0.5/\sqrt{32}} = -1.13 \).
5) Fail to reject, since $-1.13 \not< -1.645$

6) Fail to reject the claim that the population mean cookie weight is 11g and: **Conclude that there is not sufficient evidence to warrant rejection of the claim that the population mean weight is 11g.** (see p.327 in book)
Type I and Type II errors:

- A **Type I** error is rejecting the null when it is actually true (convicting an innocent guy)
- A **Type II** error is failing to reject when the null is actually false (not convicting a guilty guy)

- We generally consider Type I error to be worse, so we limit those to the fixed significance level ($\alpha$) and try to minimize the probability of a Type II error. ($\beta$)

- The **power** of a test $\beta$ is the probability of rejecting when the alternative is true ($1 - \beta$).
Example 2: A manufacturer is concerned that their soda-filling machine may not be properly calibrated. A sample of 18 20oz bottles is found to have an average content of 19.96oz with a sd of 0.04oz. Test the claim that the machine is properly calibrated?

1. $\mu = 20$ (claim) vs $\mu \neq 20$
   $H_0 : \mu = 20$ where $\mu =$ population mean bottle content
   $H_1 : \mu \neq 20$

2. $\alpha = 0.05$

3. test for a mean, with $\sigma$ UNKNOWN $\Rightarrow t = \frac{\bar{x} - \mu}{s/\sqrt{n}}$ $t$ with 17 degrees of freedom

4. Two-tailed test, reject if $t < -2.11$ OR $t > 2.11$ (critical region). Compute $t = \frac{19.96 - 20}{0.04/\sqrt{18}} = -4.24.$

5. Reject the null, since $-4.24 < -2.11$

6. Conclude that there is sufficient evidence to warrant rejection of the claim that the mean bottle content is 20oz.
**p-value:**

- An alternative to finding the critical region is to compute the **p-value**, a measure of how unusual our observed data would be if the null was true.

\[ p = \text{probability of observing a test statistic as or more extreme than the one we observed, assuming the null hypothesis is true.} \]

- We reject when \( p < \alpha \) !!

**Example 1:** \( H_0 : \mu = 11 \) vs \( H_1 : \mu < 11 \), reject if \( z < -1.645 \) so “extreme” is very negative

\[
z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}} = -1.13 \quad p = P(z < -1.13) = 0.1262 \quad p > 0.05
\]

\[ \Rightarrow \text{fail to reject} \]

If \( \mu = 11 \), the probability of seeing a test statistic at least as extreme as we got is 0.1292.
Example 2: $H_0 : \mu = 20$ vs $H_1 : \mu \neq 20$,
reject if $t < -2.11$ OR $t > 2.11$ so “extreme” is very negative OR very positive

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = -4.24$$

$$p = P(t < -4.24) + P(t > 4.24) = 2P(t < -4.24)$$
from t-table

$$p < 0.01$$ (p = 0.0055 from calculator)

$$p < 0.05 \Rightarrow \text{reject } H_0$$

If $\mu = 20$, the probability of seeing a test statistic at least as extreme as we got is $< 0.01$ (or 0.0055).
Example 3: In a recent poll of 1207 people, 53% disapproved of the way Bush is handling the economy. Test the claim that Bush’s approval rating is NOT DUE to random variation from an even split.

1. \( H_0 : p = 0.5 \) vs \( H_1 : p \neq 0.5 \) (claim)
2. \( \alpha = 0.05 \)
3. Recall: that if \( X \) is binomial with \( n \) and \( p \), then by CLT \( \hat{p} \) is approx. normal with mean \( p \) and sd \( \sqrt{p(1-p)/n} \). So, the test statistic is \( z = \frac{\hat{p} - p}{\sqrt{p(1-p)/n}} \) note: use of \( p \) rather than \( \hat{p} \) as in CI
4. \( z = \frac{0.53 - 0.5}{\sqrt{(0.5)(0.5)/1207}} = 2.08 \)
   (a) Critical Region: reject if \( z < -1.96 \) or \( z > 1.96 \) OR
   (b) p-value = \( P(|Z| > 2.08) = 2P(Z < -2.08) = 2(0.0188) = 0.0376 \)
5. Reject the null, (a) since 2.08 > 1.96 (b) since p-value < 0.05
6. Conclude that the sample data support the claim that Bush’s approval rating is NOT DUE to random variation from an even split.
**Note:** There are 3 equivalent ways to reach the same conclusion:

- Is the hypothesized population proportion, \( p \), outside of the 95% CI?
- Is the test statistic in the critical region?
- Is the p-value < 0.05?
Key Concepts!!!!!

- **Hypothesis Test**: for mean with $\sigma$ known and unknown AND for proportions
- **6 Step Method**
- **Type I and II errors**
- **p-values**