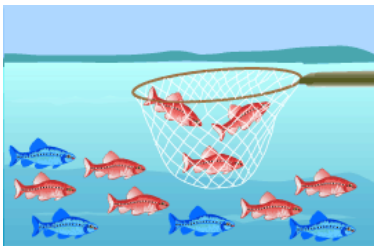
3.10 More Hypothesis Testing and the Risk of Being Wrong

Purpose of the Hypothesis Test

Verify the validity of a claim about a *population* based on a single *sample*.



Process of Determination

1. Develop your hypothesis statements H_0 and H_a.

These are statements about a population, so they are written in terms of a population parameter.

- The **null hypothesis**, H_0, is a statement of "no effect" or "no difference." It *cannot* be proven true but can be shown to be *untrue* with specific risks of error. These decisions are analogous to a courtroom finding a defendant not guilty or guilty.
- The **alternative hypothesis**, H_a, represents the result when the null hypothesis is rejected. Because H_a expresses the hypothesis we hope to find evidence for, begin with H_a and set up H_0 as the nonoccurrence of the "preferred" outcome.

2. Select a level of significance a.
   - Confidence level is $1 - a$.
3. Select a sample size n.
4. Select an appropriate test (one sample, two sided, etc.) for your hypothesis.
5. Calculate the standardized test statistic from the sample data (t, Z, etc.).
6. Use the test statistic to compute the area in the tail(s), or p-value.
7. Compare the p-value with a.
8. **Reject** or **fail to reject** the null hypothesis.
9. State your decision.

Decision Rule: If p Is Low, H_0 Must Go

1. If your p-value is $< \alpha$, **reject** H_0.
   - In this case, the data are statistically significant at the α level, and the observed difference is too large to be explained by chance alone.
2. If your p-value is $\geq \alpha$, **fail to reject** H_0.

Example: Packaging Vegetables

You are responsible for packaging vegetables and labeling them 227 grams. You sample four packs of vegetables and find the average weight to be 222 grams.

Variation exists, so we can't expect every veggie pack to weigh exactly 227 grams. Is this low average weight due to chance variation, or is it evidence that your sorting and packaging equipment require adjusting?

Solution: Packaging Vegetables

The packaging process has a known standard deviation, σ, of 5 grams. Assume a level of significance, α, of 0.10 ($1 - \alpha = 90\%$). So we are given:

1. $n = 4$
2. $\bar{x} = 222$
3. $\sigma = 5$

Since the equality condition is always in H_0, we have a one-sample, two-tailed test with known σ:

1. H_0: $\mu = 227$
2. H_a: $\mu \neq 227$
3. $Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{222 - 227}{\frac{5}{\sqrt{4}}} = -2$
   - From your book, use: Table C
   - In Excel, use:
     - `=STANDARDIZE(222,227,2.5)`, which gives -2
     - `=NORM.S.DIST(-2)`, which gives 0.0228

4. Since this is a two-tailed test, our p-value is 2×0.0228, or 0.0456.
5. Since $0.0456 < 0.10$, $p < a$, so reject H_0.
   - (If p is low, H_0 must go!)
6. Your packaging equipment probably needs adjusting.

Risk of a False Conclusion

Because we use a *sample* to draw a conclusion about an entire *population*, our conclusions might be false.

		The actual state of things (what actually happened)	
		H_0 is true	H_0 is false
The conclusion you draw (what you think happened)	Fail to reject H_0	Correct conclusion	Type II error, beta risk, or consumer's risk
	Reject H_0	Type I error, alpha risk, or producer's risk	Correct conclusion

In a **type I error**, you reject the null hypothesis (accept H_a), when you should have accepted the null hypothesis. You believe you discover something that is in fact false.

In a **type II error**, you fail to reject the null hypothesis when you should have done so. You fail to discover something that is true.

1. a is the probability of a type I error and the probability of incorrectly rejecting H_0 when H_0 is true.
2. β is the probability of a type II error and the probability of incorrectly failing to reject H_0 when H_0 is false.
3. The power of the test is $1 - \beta$, or the probability of correctly rejecting H_0 when H_0 is false.

Two Possible Conclusions

There are only two possible conclusions:

1. **Reject H_0.**
2. **Do not reject H_0.**

We do not have enough evidence based on a sample to *confirm* the null hypothesis.

A defendant is not found "innocent" by a jury, they are found "not guilty": the evidence is insufficient to prove either guilt or innocence.