Apr. 15  Statistic for the day:  
There were 2.1 million prisoners in the U.S. at the end of 2003.  
(including 0.47% of all white males, 1.23% of all hispanic males, and 3.41% of all black males)

Assignment:  
Reread Chapter 22  
Exercises pp. 428-429:  3, 5, 6, 11, 13

The steps in a hypothesis test  
From the book:  p. 414  
1. Determine the null hypothesis and the alternative hypothesis.  
2. Collect data and summarize them with a single number known as a test statistic.  
3. Determine how unlikely the test statistic is if the null hypothesis is true.  
4. Make a decision.

Recall Step3:  
Determine how unlikely the test statistic is if the null hypothesis is true.  
The probability that we compute to see how unlikely the test statistic is is called the  
p-value.  
\[ p-value = \text{probability of observing the statistic you observed when the null hypothesis is true.} \]

Definition of p-value (p. 415)  
Assume that the null hypothesis is true.  The probability of observing a test statistic at least as extreme as the one actually observed, under this assumption, is called the  
p-value.

Byzantine coins  
Historically, fluctuations in the amount of rare metals found in coins are common.  We consider a certain twelfth century Byzantine coin minted a two different times during the reign of Manuel I (1143 – 1180).  Discovered in Cyprus.

Statistics can help decide if the king’s minter was cheating by reducing the amount of silver in the coins from one mintage to the next.
Byzantine coins: Research question

To analyze the silver content of a large find of coins, coins must be destroyed (dissolved in nitric acid).

The measurement on a coin consists in the % silver.

9 coins were identified as being from the first coinage
7 coins were identified as being from the fourth coinage.

Question of interest: Did the true mean silver content decrease between the first coinage and the fourth coinage?

Summary of data: % silver content in coins

<table>
<thead>
<tr>
<th></th>
<th>First</th>
<th>Fourth</th>
</tr>
</thead>
<tbody>
<tr>
<td>mean</td>
<td>6.7</td>
<td>5.6</td>
</tr>
<tr>
<td>SD</td>
<td>.56</td>
<td>.36</td>
</tr>
<tr>
<td>sample size</td>
<td>9</td>
<td>7</td>
</tr>
</tbody>
</table>

Let’s set up appropriate statistical hypotheses (remember, a hypothesis is a theory)

Statistical hypotheses are ALWAYS written about the TRUE population values (never the sample estimates).

Null (nothing) Hypothesis:

The null hypothesis is the theory of no difference, or nothing happening, or... however “NOTHING” should be interpreted for a given problem!

For the coins:

The null hypothesis states that there is no difference between the two mintings in terms of true mean silver content.

Alternative Hypothesis

The alternative hypothesis is a theory that is different from the null hypothesis. Three possibilities in this case:

- Two-sided
  - The true mean silver content (TMSC) of the first minting is different from the TMSC of the fourth minting.
- One-sided
  - TMSC 1 is smaller than TMSC 4.
  - TMSC 1 is larger than TMSC 4.

Which of the 3 seems most appropriate? (You’re not allowed to look at the data to decide! Base the decision on the research question only.)

Test statistic

Clearly, this problem is concerned with the difference in mean silver content.

Thus, a logical choice of statistic will be based on the difference in sample means!

The test statistic will be a STANDARDIZED SCORE based on the difference in sample means.

Anatomy of a test statistic

The test statistic in this class is a STANDARDIZED SCORE:

\[
\text{Estimate} - \text{Null value} / \text{Standard error of estimate}
\]

Sounds easy! The challenge is figuring out what those three numbers are.
We rejected the null hypothesis.

Could we be making a type 1 error (rejecting when we shouldn’t)?

YES!

Could we be making a type 2 error (failing to reject when we should)?

NO! (That can only happen when we fail to reject the null.)

P-value (see top of p. 376)

The p-value is the probability, under the assumption that the null hypothesis is true, of observing a test statistic at least as extreme (supportive of the alternative hypothesis) as the one observed.

Null: No difference in TMSC (which means that the standardized test statistic should be roughly standard normal).

Observed test statistic: 4.6

Alternative: TMSC 1 bigger than TMSC 4 (which should lead to a POSITIVE test statistic)

Put it all together: p-value is probability of seeing standard normal bigger than 4.6

Decision time!

SMALL p-values indicate EVIDENCE AGAINST the null.

LARGE p-values indicate LACK OF EVIDENCE AGAINST the null.

Traditionally, .05 is the dividing line between LARGE and SMALL.

In our case, the p-value is about 0. We have strong evidence against the null.

Conclusion: We have evidence of that TMSC 1 is larger than TMSC 4.

Summary:

State the alternative and null hypotheses.

(Research Advocate vs the Skeptic)

Collect and summarize data.

Calculate a test statistic.

Ascertain the p-value.

Decide whether the data support the alternative hypothesis.