The "property" referred to in the preceding definitions is often the value of a population parameter, so here are some examples of typical hypotheses (or claims):

- $\mu < 98.6°F$ The mean body temperature of humans is less than 98.6°F.
- $p > 0.5$ The XSORT method of gender selection increases the probability that a baby will be born a girl, so the probability of a girl is greater than 0.5.
- $\sigma = 15$ The population of college students has IQ scores with a standard deviation equal to 15.

Example 1 **Testing the Claim That the XSORT Gender-Selection Method Is Effective**

Assume that 100 babies are born to 100 couples treated with the XSORT method of gender selection that is claimed to make girls more likely. If 58 of the 100 babies are girls, test the claim that "with the XSORT method, the proportion of girls is greater than the proportion of 0.5 that occurs without any treatment." Using p to denote the proportion of girls born with the XSORT method, the claim is that $p > 0.5$.

The big picture In Example 1 we see that getting 58 girls in 100 births is more than the 50 girls that we would expect with no treatment or an ineffective treatment. But is 58 girls high enough to justify the conclusion that the XSORT method is effective? The method of hypothesis testing allows us to answer that question. We will see that without an effective treatment, there is a 0.0548 probability of getting 58 *or more* girls. Because that probability is not small, such as 0.05 or less, we will conclude that it is easy to get 58 girls in 100 births by random chance, so 58 girls is not quite high enough to justify the conclusion that the XSORT method is effective. (The actual XSORT results are more extreme, so in reality it does appear that the XSORT method is effective.)

Using technology It is easy to obtain hypothesis-testing results using technology. The accompanying screen displays show results from four different technologies, so we can use computers or calculators to do all of the computational heavy lifting. Examining the four screen displays, we see some common elements. They all display a "test statistic" of $z = 1.60$, and they all include a "P-value" of 0.055 (rounded). These two results are important, but *understanding* the hypothesis-testing procedure is critically important. Focus on *understanding* how the hypothesis-testing procedure

Aspirin Not Helpful for Geminis and Libras

Physician Richard Peto submitted an article to *Lancet*, a British medical journal. The article showed that patients had a better chance of surviving a heart attack if they were treated with aspirin within a few hours of their heart attacks. *Lancet* editors asked Peto to break down his results into subgroups to see if recovery worked better or worse for different groups, such as males or females. Peto believed that he was being asked to use too many subgroups, but the editors insisted. Peto then agreed, but he supported his objections by showing that when his patients were categorized by signs of the zodiac, aspirin was useless for Gemini and Libra heart-attack patients, but aspirin is a lifesaver for those born under any other sign. This shows that when conducting multiple hypothesis tests with many different subgroups, there is a very large chance of getting some wrong results.

STATDISK

```
Claim:   p > p(hyp)

Sample proportion: 0.58
Test Statistic, z: 1.6000
Critical z:        1.6449
P-Value:           0.0548
```

MINITAB

```
Test of p = 0.5 vs p > 0.5

                              95% Lower
Sample  X    N  Sample p      Bound   Z-Value  P-Value
1      58  100  0.580000   0.498817     1.60    0.055
```

TI-83/84 PLUS

```
1-PropZTest
 prop>.5
 z=1.6
 p=.0547992894
 p̂=.58
 n=100
```

STATCRUNCH

Hypothesis test results:

p : proportion of successes for population

$H_0 : p = 0.5$

$H_A : p > 0.5$

Proportion	Count	Total	Sample Prop.	Std. Err.	Z-Stat	P-value
p	58	100	0.58	0.05	1.6	0.0548