## Test Your Understanding 1

Scientists wish to test whether a laboratory scale is accurate. To do so, they weigh a 1 gram test weight 12 times. Formulate the scientific hypothesis, the statistical model and the statistical hypotheses. Be sure to tell what the parameters in the model are.

## Solution:

- 1. The Scientific Hypothesis The scale is not accurate.
- 2. The Statistical Model C+E:

 $Y = \mu + \epsilon,$ 

where the random error,  $\epsilon$ , follows a  $N(0, \sigma^2)$  distribution model.

# 3. The Statistical Hypotheses

$$\begin{array}{rrrr} H_0: & \mu & = & 1 \\ H_a: & \mu & \neq & 1 \end{array}$$

#### Test Your Understanding 2

In the experiment described in TYU 1, the twelve weighings of the 1 g test weight have a mean of 0.974g and a standard deviation of 0.027g. For the hypotheses

$$\begin{array}{rcl} H_0: & \mu & = & 1 \\ H_a: & \mu & \neq & 1, \end{array}$$

a. Indicate how to obtain the *p*-value.

Solution: The standardized test statistic is

$$t = \frac{\overline{Y} - 1}{S/\sqrt{12}}$$

which has a  $t_{11}$  distribution under the null hypothesis. The observed value of the standardized test statistic is

$$t^* = \frac{0.974 - 1}{0.027/\sqrt{12}} = -3.3358.$$

Since this is a two-sided test, the p-value equals  $p \pm = 2min(p_-, p^+)$ , where  $p_- = P(t_{11} \le -3.3358)$ , and  $p^+ = P(t_{11} \ge 3.3358)$ .

b. Use the appropriate table to test the hypotheses at the 0.05 level of significance.

**Solution:** Since  $t_{11,0.995} = 3.1058$ , and  $t_{11,0.999} = 4.0250$ ,  $0.001 \le p_{-} \le 0.005$ , and  $0.995 \le p^{+} \le 0.999$ , so  $0.002 \le p \pm \le 0.010$ . Therefore  $H_0$  is rejected in favor of  $H_a$  at the 0.05 level of significance.

#### Test Your Understanding 3

As part of its "Quality 2000" quality improvement program, a manufacturer of microwave ovens has set a target of 0.1% defectives (i.e. ovens returned for replacement or repair) in the first year. To check if the target is met, a random sample of return and service records of 10,500 ovens sold the past year was taken, and 17 of the ovens were found to be defective. Is this good evidence that the target has not been achieved? To decide, conduct an hypothesis test at the 0.01 significance level. Be sure to give:

a. The scientific hypothesis.

**Solution:** That the target is not being met.

b. The statistical model.

**Solution:** The binomial model, b(10500, p), where p is the population proportion of defective ovens.

c. The statistical hypotheses being tested.

Solution:

$$\begin{array}{rcl} H_0: & p & = & 0.001 \\ H_a: & p & > & 0.001, \end{array}$$

d. The test statistic being used.

Solution: The test statistic is Y, the number of defective ovens in the sample.

e. The p-value and the decision rule being used.

Solution: The p-value may be computed exactly or using the normal approximation.

**Exact p-value:**  $p^+ = P(b(10500, 0.001) \ge 17) = 0.0395.$ 

Normal approximation:  $p^+ \approx P(N(0,1) \ge z_u^*)$ , where  $z_u^* = (17 - (10500)(0.001) - 0.5)/\sqrt{(10500)(0.001)(0.999)} = 1.853$ . Thus  $p^+ \approx P(N(0,1) \ge 1.853) = 0.0319$ .

f. The assumptions made, and why they are, or are not, justified.

**Solution:** For the normal approximation, we assume a large enough sample for a good approximation. The guidelines of  $Y \ge 10$  and  $n - Y \ge 10$  are satisfied.

g. Your conclusions.

**Solution:** We would reject  $H_0$  in favor of  $H_a$  at the 0.05 level of significance, but not at the 0.025 level. According to the interpretations found in Table 6.1, p. 292, the evidence against  $H_0$  and in favor of  $H_a$  is between "reasonably strong" and "strong".

### Test Your Understanding 4

In a test of the efficacy of SAT preparation courses, a group of 200 randomly-selected high school seniors were randomly divided into two groups: a treatment group of 100 took the SAT in the fall, were then given an SAT preparation course, and took the SAT again in the spring; a control group of 100 took the same SAT in the fall and again in the spring, without taking the preparation course in between. The experimenters suspect that the SAT preparation course is effective. The treatment group posted a mean increase in SAT scores of 52.6 with a standard deviation of 15.5, while the mean increase in SAT scores for the control group was 48.9 with a standard deviation of 22.7. Plots revealed no evidence of outliers or nonnormality in the data.

Conduct an hypothesis test for this problem at the 0.05 level of significance. Be sure to give:

a. The scientific hypothesis.

**Solution:** The SAT preparation course is effective.

b. The statistical model.

**Solution:** The two-population C+E model.

c. The statistical hypotheses being tested.

Solution:

$$\begin{array}{rcl} H_0: & \mu_1 - \mu_2 & = & 0 \\ H_{a^+}: & \mu_1 - \mu_2 & > & 0, \end{array}$$

where  $\mu_1$  is the population mean increase for the SAT prep group and  $\mu_2$  is the population mean increase for the control group.

d. The test statistic being used.

Solution: Since the sample sizes are large, we can use the large sample test (Case 2, p. 349 of text). The standardized test statistic is

$$Z = \frac{\overline{Y}_1 - \overline{Y}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

e. The p-value and the decision rule being used. Solution: The observed value of the test statistic is

$$z^* = \frac{52.6 - 48.9}{\sqrt{\frac{15.5^2}{100} + \frac{22.7^2}{100}}} = 1.346.$$

The p-value is then  $p^+ = P(N(0, 1) \ge 1.346) = 0.0892$ .

f. The assumptions made, and why they are, or are not, justified.

**Solution:** That the test statistic is normally distributed or nearly so. This is true because of the lack of outliers or nonnormality in the data and the Central Limit Theorem.

g. Your conclusions.

**Solution:** The p-value is between 0.05 (reasonably strong evidence against  $H_0$  and in favor of  $H_a$ ) and 0.10 (borderline evidence against  $H_0$  and in favor of  $H_a$ ). We would not reject  $H_0$  in favor of  $H_a$  at the usual 0.05 level of significance.

### Test Your Understanding 5

It has been reported (The Boston Globe, July 31, 1997) that as much as 25% of all worker's compensation claims and 40% of all payments are attributable to lower back injuries. In an effort to reduce such injuries, employers have sent workers "back to school" to learn to lift heavy objects safely. Suppose that, in order to study the effectiveness of such a program, for one year after the program's completion, you monitor the incidence of back injury for a random sample of 200 workers who have taken the program, and a random sample of 200 workers who have not. Of the 200 workers who completed the course, 18 report at least one lower back injury during this period and of the other group, 21 report at least one lower back injury. Is this convincing evidence for the efficacy of the program? Formulate the scientific and statistical hypotheses, the statistical model, and the standardized test statistic. Obtain the *p*-value and use it to state your conclusion. What assumptions have you made and are they justified?

# Solution:

a. The scientific hypothesis. The program is effective.

## b. The statistical model.

The two-population binomial model. Let  $Y_1$  be the number in the course-graduate group and  $Y_2$  the number in the control group who reported at least one lower back injury. Then  $Y_1 \sim b(n_1, p_1)$  and  $Y_2 \sim b(n_2, p_2)$ , where  $n_1 = n_2 = 200$  are the sample sizes, and  $p_1$  and  $p_2$  are the corresponding population proportions reporting lower back pain.

c. The statistical hypotheses being tested.

d. The test statistic being used.

Since the sample sizes are large, and the null hypothesis states the equality of the population proportions, we use the large sample test given in Case 1, p. 350 of the text. The standardized test statistic is

$$Z = \frac{\hat{p}_1 - \hat{p}_2}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

where  $\hat{p} = (Y_1 + Y_2)/(n_1 + n_2)$ .

e. The p-value.

For these data,  $\hat{p} = (18 + 21)/(200 + 200) = 0.0975$ ,  $\hat{p}_1 = 18/200 = 0.09$ , and  $\hat{p}_1 = 21/200 = 0.105$ . Then the observed value of the test statistic is

$$z^* = \frac{0.09 - 0.105}{\sqrt{0.0975(1 - 0.0975)\left(\frac{1}{200} + \frac{1}{200}\right)}} = -0.5057.$$

The approximate p-value is then  $p_{-} = P(N(0, 1) \le -0.5057) = 0.3065$ .

f. Conclusions.

The p-value is much larger than 0.10 (borderline significant), so we do not reject  $H_0$ . There is not enough evidence to conclude that among workers who take the course, the population proportion reporting lower back pain is lower than among those who do not take the course.

g. The assumptions made, and why they are, or are not, justified.

That the test statistic is normally distributed or nearly so. This is true because both  $Y_1$  and  $Y_2$  exceed 10.