

## Complete Examples

1. Is a soda-dispensing machine performing according to specification? Pepsi's dispensing machine is designed to fill bottles with exactly 2 liters of their product. To test if the machine is performing according to specification, we collect a sample of 100 "2-liter" bottles. The average quantity contained in the sample bottles is  $\bar{x} = 1.985$  liters. The sample standard deviation of the fill is  $s = 0.05$ . Test whether the machine is in control, at the 5% level of significance.

(a) What are the population and the sample?

**Solution:** Population: all bottles filled by the machine. Sample: the  $n = 100$  measured bottles.

(b) What are the null and alternative hypotheses?

**Solution:** The population we are interested in is all bottles filled by the machine (past and future); the mean fill level of the population is  $\mu$ . Our null and alternative hypotheses are as follows:

$$H_0 : \mu = 2$$

$$H_a : \mu \neq 2.$$

The nominal (null) value of the population mean is  $\mu_0 = 2$ .

(c) What is the test statistic?

**Solution:** Our test statistic is based on the sample of  $n = 100$  bottles. Let  $\bar{x}$  denote the mean fill level from this sample. Our test statistic is

$$\begin{aligned} t &= \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \\ &= \frac{1.985 - 2}{(.05)/\sqrt{100}} \\ &= -3. \end{aligned}$$

(d) Approximately what is the  $p$ -value?

**Solution:**

$$p \approx P(|Z| > 3) = 0.002700.$$

(e) What assumptions are you making?

**Solution:** We are assuming that the sample bottles were drawn independently without bias from the population.

(f) What is  $\alpha$ ? What is the result of the test?

**Solution:** The significance level is  $\alpha = 0.05$ . Since  $p < \alpha$ , we reject  $H_0$ .

2. Before Facebook's recent redesign, the mean number of ad clicks per day was 100K. In the 49 days after the redesign, the mean number of ad clicks per day was 105K and the standard deviation was 35K. Is there significant evidence that the redesign affected the expected number of ad clicks? Perform a test at the 5% level.

(a) What is the sample? What is the population?

**Solution:** The sample is the number of ad clicks on the measured  $n = 49$  days. The population is the number of ad clicks on all days after the redesign.

(b) What are the null and alternative hypotheses?

**Solution:** Let  $\mu$  be the expected clicks per day after the redesign. We use "thousands of clicks" as the units for all relevant quantities.

The null hypothesis is that the redesign had no effect on expected ad clicks. The alternative hypothesis is that  $\mu$  changed after the redesign:

$$H_0 : \mu = 100$$

$$H_a : \mu \neq 100$$

(c) What is the test statistic?

**Solution:** The test statistic is based on the sample, the observed clicks in the  $n = 49$  days after the redesign. Let  $\bar{x}$  denote the mean clicks per day in the sample; let  $s$  denote the sample standard deviation. Our test statistic is

$$\begin{aligned} t &= \frac{\bar{x} - \mu_0}{s/\sqrt{n}} \\ &= \frac{105 - 100}{35/\sqrt{49}} \\ &= 1. \end{aligned}$$

(d) Approximately what is the  $p$ -value?

**Solution:**

$$\begin{aligned} p &\approx P(|Z| > 1) \\ &= 0.3137. \end{aligned}$$

(e) What assumptions are you making?

**Solution:** We need to assume that the observed sample is a simple random sample from the population.

Admittedly, the assumption does *not* hold, since there is strong selection bias in the sample: we are sampling days right after the website redesign with higher probability than days far into the future. For example, we have no chance of sampling a day three years into the future.

Since the assumption does not hold, we are in a bit of an awkward position. The hypothesis test may not be valid. One way in which it may not be valid is the following: it usually takes people a few weeks to adjust to a website redesign, so the ad click behavior in our sample of 49 days may not be representative of all future ad click behavior.

(f) What is  $\alpha$ ? What is the result of the test?

**Solution:** Despite the caveats mentioned in part (d), we will proceed with the test. For a level-5% test,  $\alpha = 0.05$ . Since  $p \geq 0.05$ , we do not reject  $H_0$ . If there were no difference in expected ad clicks before and after the redesign, there would be a 31.37% chance of seeing data like we observed. There is no evidence of a difference.

## Types of Errors

3. In a hypothesis test, our decision will either be “reject  $H_0$ ” or “do not reject  $H_0$ ”. Under what situations will each of these decisions be in error?

**Solution:** Type I error:  $H_0$  is true, but we reject it.  
Type II error:  $H_0$  is false, and we fail to reject it.

4. We reject  $H_0$  when the  $p$ -value is below  $\alpha$ .  
(a) If  $H_0$  is true, what is the probability of making a Type I error?

**Solution:** We reject  $H_0$  when the  $p$ -value is less than  $\alpha$ . This happens when  $|T| > t_{\alpha/2, n-1}$ . So, if the null hypothesis is true, then the probability of making a Type I error is

$$P(|T| > t_{\alpha/2, n-1}) = \alpha.$$

- (b) If  $H_0$  is false, what is the probability of *not* making a Type II error?

**Solution:** We cannot give a direct answer to this question, because it depends on the true value of  $\mu$ . In general, the probability of not making a Type II error is called the “power” of the test; it is given the symbol  $\beta$  or  $\beta(\mu)$ . If  $\mu$  is close to  $\mu_0$ , then  $\beta$  will be small (close to  $\alpha$ ); if  $\mu$  is far from  $\mu_0$ , then  $\beta$  will be large (close to 1).

## More $p$ -values

5. Suppose we perform a hypothesis test and we observe a  $p$ -value of  $p = .02$ . True or false: There is a 2% chance that the null hypothesis is true.

**Solution:** False. The  $p$ -value is the probability of getting a test statistic at least as extreme as what was observed. Heuristically, we can think of this as

$$P(\text{Data} \mid H_0 \text{ is true}) = 2\%.$$

The statement in the problem is

$$P(H_0 \text{ is true} \mid \text{Data}) = 2\%.$$

Clearly, this is not the same.

6. Suppose we perform a hypothesis test and we observe a  $p$ -value of  $p = .02$ . True or false: If we reject the null hypothesis, then there is a 2% chance of making a type I error.

**Solution:** False. We can only make a type I error when the null hypothesis is true. Thus, the statement in question 6 is *exactly the same* as the statement in question 5.

7. Suppose we perform a hypothesis test and we observe a  $T$  test statistic  $t = -2.02$ , corresponding to a  $p$ -value of  $p = .02$ . True or false: If we were to repeat the experiment and the null hypothesis were actually true, then there would be a 2% chance of observing a test statistic at least as extreme as  $t = -2.02$ .

**Solution:** True. The  $p$ -value is the probability of getting a test statistic as least as extreme as the observed value if the null hypothesis were true. Note: for a one-sided less-than alternative, extreme means “less than or equal to.”