

## 7.4: Claims on Population Proportions

### Learning Objectives

- Conduct hypothesis testing on claims regarding population proportions using the sampling distribution of sample proportions
- Conduct hypothesis testing on claims regarding population proportions using test statistics

### Review and Preview

Recall that proportions measure the percentage of observations that admit a certain quality. We might be interested in the percentage of the population (who are registered to vote) that will actually vote in an upcoming election. Each registered voter either will vote or will not vote; the registered voter either has the quality or does not have the quality. Remember that we denote the proportion of the population with a quality of interest using  $p$ . Since there are only two states regarding the quality, everybody else does not have the quality; we denote this proportion with  $q$ . The entire population is covered between the observations with the quality and those without the quality; we thus know that  $p + q = 1$ .

We may be even more interested in the proportion of registered voters who will vote for a particular candidate or a particular item on the ballot. In certain cases, particular proportions of affirmative votes are required for an item to pass. Can we test that there is enough support for a particular candidate or item to pass before the election occurs or before all of the ballots are counted? These questions center on claims about population proportions and is the topic of this section.

Testing claims on population proportions intimately involves the sampling distribution of sample proportions. In order to compute  $p$ -values, we need to know the approximate shape of the sampling distribution. We have seen and utilized the fact that the sampling distribution of sample proportions is approximately normal with  $\mu_{\hat{p}} = p$  and  $\sigma_{\hat{p}} = \sqrt{\frac{pq}{n}}$  when our sample size  $n$  is large enough that we expect more than 5 observations with the quality and more than 5 observations without the quality to be in our sample. To check this condition, we checked that the following two inequalities were satisfied:  $np > 5$  and  $nq > 5$ . With such a preview and having several sections of hypothesis testing under our belts, let us begin testing claims on population proportions.

### Testing Claims on Population Proportions

When conducting hypothesis testing, we do not know the value of the population proportion  $p$ , but this is okay because we compute the  $p$ -value under the assumption that the null hypothesis is true. We will thus be operating under the assumption that the population proportion is equal to some particular value which we denote as  $p_0$ . This notation leads to  $q_0$  which is the hypothesized proportion of the population without the quality. So, in order to conduct hypothesis testing on claims about population proportions, we need our samples to be randomly chosen and of such a size that  $np_0 > 5$  and  $nq_0 > 5$ . When these conditions are met, we can conduct the probability assessment using a normal distribution with  $\mu_{\hat{p}} = p_0$  and  $\sigma_{\hat{p}} = \sqrt{\frac{p_0q_0}{n}}$ .

#### ? Text Exercise 7.4.1

The success of a manufacturing plant that produces tens of thousands of motion-detecting sensors each week requires a high degree of quality assurance and quality control. As such, the plant sets the standard that at most 2.5% of the sensors produced at the plant will be defective. To test that the plant is meeting its production standards, random samples of 500 sensors are taken each week and tested. The company tests at the 0.03 level of significance. Last week, the sample contained 20 defective sensors.

1. Conduct the hypothesis test using the sampling distribution of sampling proportions and interpret the conclusions within the context of the problem.

#### Answer

We are considering a claim on population proportions because we are considering the percentage of sensors that have the quality that they are defective. The company set the standards that  $p < 0.025$ . This forms one of our hypotheses. The competing hypothesis would thus be that  $p \geq 0.025$ . The company does not want to have the default position that the machinery is not working; otherwise, they will frequently be conducting unnecessary maintenance.

$$H_0 : p \leq 0.025$$

$$H_1 : p > 0.025$$

We note that under the assumption that the null hypothesis is true, the largest  $p$ -value will be computed when the value is assumed to be 0.025. We thus set  $p_0 = 0.05$  to ensure the conditions for the test are met. Noting that  $q_0 = 1 - p_0 = 0.95$  and  $n = 500$ , we have  $np_0 = 12.5$  and  $nq_0 = 487.5$ . With the two inequalities met and the sample being randomly chosen, we can conduct the hypothesis test.

Under the assumption that the null hypothesis is true and in the situation that produces the largest  $p$ -value, we have  $\mu_{\hat{p}} = 0.025$  and  $\sigma_{\hat{p}} = \sqrt{\frac{0.025 \cdot 0.975}{500}} \approx 0.007$ . The random sample of sensors from last week had 20 defective sensors. This is not the sample proportion. Proportions fall inclusively between 0 and 1. The proportion of defective sensors in the sample is the percent of defective sensors in the sample; thus,  $\hat{p} = \frac{20}{500} = 0.04$ . Given the hypotheses, we have a right-tailed test and thus visualize the test in the following figure.

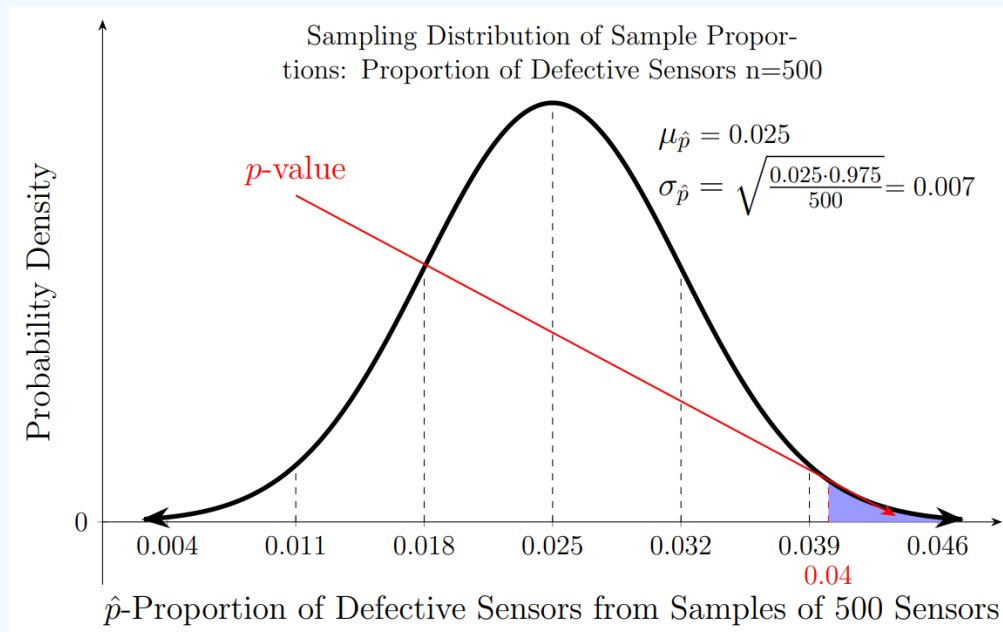


Figure 7.4.1 Sampling distribution of sample proportions

$$p\text{-value} \approx 1 - \text{NORM.DIST}(0.04, 0.025, 0.007, 1) \approx 0.0158$$

Since the  $p$ -value is smaller than the  $\alpha$  value, we have sufficient evidence to reject the null hypothesis that proportion of defective sensors is within the limit given the quality control standards set by the company. As such further investigation should happen regarding the sensors produced last week and the machinery should be checked before continuing production.

- Determine a transformation that takes the sampling distribution of sample proportions to a common distribution and thus determine the formula for the test statistic within the context of hypothesis testing with claims on population proportions. Verify that your solution is correct by applying it in the context of this text exercise and obtaining the same  $p$ -value.

### Answer

Since the sampling distribution of sample means is approximately normal when we are able to conduct hypothesis testing on claims about population parameters and we know the mean and the standard deviation, we can use the  $z$ -score transformation to map the sampling distribution to the standard normal distribution. We can thus define our test statistic as follows

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$$

We now apply this formula to the context of the text exercise to obtain a little validation. Using the values computed in the previous part, we have  $n = 500$ ,  $\mu_{\hat{p}} = 0.025$ ,  $\sigma_{\hat{p}} = \sqrt{\frac{0.025 \cdot 0.975}{500}} \approx 0.007$ , and  $\hat{p} = \frac{20}{500} = 0.04$ .

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \approx \frac{0.04 - 0.025}{\sqrt{\frac{0.025 \cdot 0.975}{500}}} \approx 2.1483$$

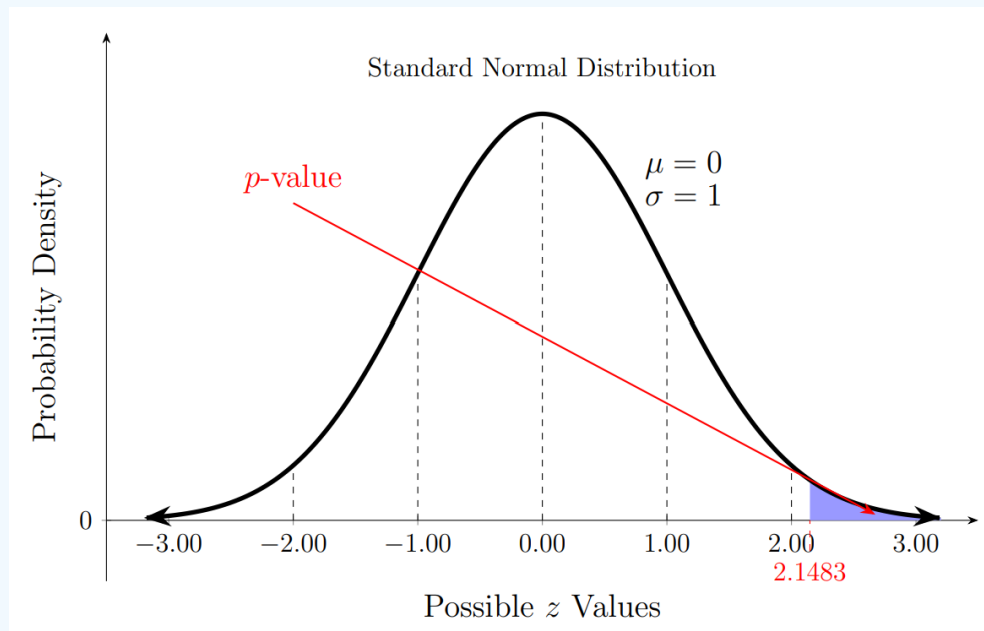


Figure 7.4.2 Standard normal distribution

$$p\text{-value} \approx 1 - \text{NORM.S.DIST}(2.1483, 1) \approx 0.0158$$

This produces the same  $p$ -value as computed from the sampling distribution of sample proportions in the previous part of the question. Indeed, we have settled on the proper formulation of test statistics in the realm of testing hypotheses on population proportions.

### ? Text Exercise 7.4.2

A large, public corporation with thousands of shareholders is considering purchasing another large corporation, but according to the bylaws by which the corporation was founded, to do so requires a two-thirds majority of shareholders to be in support of such a purchase. The chief operating officer is vehemently opposed to the acquisition and has been rallying the shareholders to vote against the purchase. The chief operating officer gets to set the agenda for the upcoming shareholder meeting and is trying to decide if the vote regarding the purchase should be held or postponed.

To facilitate this decision, the chief operating officer randomly selects 60 shareholders and has the human resources department contact them to assess their positions regarding the possible acquisition. After conducting these 60 conversations, the human resources department returns that 35 of the shareholders are planning to vote in favor of the acquisition. Conduct a hypothesis test from the perspective of the chief operating officer at the 0.05 significance level and decide, from his perspective, whether or not to schedule the vote during the upcoming meeting.

#### Answer

For the vote to pass, a two-thirds majority of shareholders need to vote in favor of the acquisition in order for it to pass. This fact determines our two hypotheses:  $p \geq \frac{2}{3}$  and  $p < \frac{2}{3}$ .

The chief operating officer does not want the acquisition to pass and, therefore, wants the second hypothesis to be true. He has the control over when the vote occurs. He does not want to assume his position is going to win out. We thus set the hypotheses as follows.

$$H_0 : p \geq \frac{2}{3}$$

$$H_1 : p < \frac{2}{3}$$

We note that under the assumption that the null hypothesis is true, the largest  $p$ -value will be computed when the value is assumed to be  $\frac{2}{3}$ . We thus set  $p_0 = \frac{2}{3}$  and note that  $q_0 = \frac{1}{3}$  and  $n = 500$ . Thus we have  $np_0 = 40$  and  $nq_0 = 20$ . With the two inequalities met and the sample being randomly chosen, we can conduct the hypothesis test.

We must determine the mean and standard deviation of the sampling distribution and the sample proportion in order to compute the test statistic and then compute the probability for this left-tailed test.  $\mu_{\hat{p}} = \frac{2}{3}$ ,  $\sigma_{\hat{p}} = \sqrt{\frac{\frac{2}{3} \cdot \frac{1}{3}}{60}} \approx 0.0609$ , and  $\hat{p} = \frac{35}{60} = \frac{7}{12} = 0.58\bar{3}$ .

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} = \frac{\frac{7}{12} - \frac{2}{3}}{\sqrt{\frac{\frac{2}{3} \cdot \frac{1}{3}}{60}}} \approx -1.3693$$

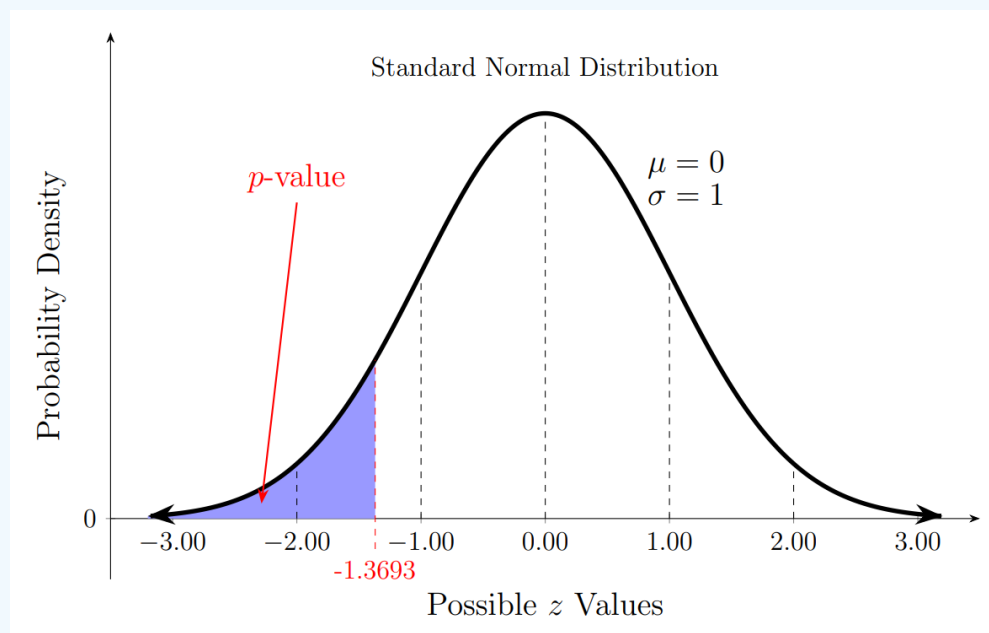


Figure 7.4.3 Standard normal distribution

$$p\text{-value} \approx \text{NORM.S.DIST}(-1.3693, 1) \approx 0.0855$$

The  $p$ -value is larger than the  $\alpha$  value. There is not sufficient evidence to reject the null hypothesis that there is enough support for the acquisition to pass. Since the chief operating officer cannot feel confident that the acquisition is going to fail, he should postpone the vote in order to have more time to convince shareholders of his point of view.

### ? Text Exercise 7.4.3

We often assume that the probability of having a female baby is 50%, but there is mounting evidence that indicates this assumption does not align with reality. The Centers for Disease Control (CDC) of the United States keeps track of birth records and makes the [data accessible to the public](#). In 2019, there were 3,747,540 births in the United States with 1,830,094 of those being female. This indicates that only 48.8345% of babies born in the United States in 2019 were female. In 2023, there were 3,519,017 births with 1,756,223 being females which again produces a proportion of 48.8380% of babies being female. What about on a global scale?

The United Nations maintains records and an organization called [Our World in Data](#) maintains an article addressing the gender ratio. It is the general trend throughout history, at least for the last century, that more males are born than females globally. A study cited by Our World in Data indicates that the proportion of females at the time of conception is indeed 50%, which implies that the difference is caused by events occurring during pregnancies. An interested reader is encouraged to examine the article linked above.

Suppose we took a random sample of newly born infants from across the world and 5460 of them were female while 5733 of them were male. Would this constitute significant evidence against the common assumption that 50% of babies born are female? Test the claim at a significance level of 0.01.

### Answer

The wording of the problem "evidence against the common assumption that 50% of babies born are female" indicates the null hypothesis. We thus have a two-tailed test with the following hypotheses.

$$H_0 : p = 0.50$$

$$H_1 : p \neq 0.50$$

In order to confirm the requirements for the test, we need to compute the sample size  $n = 5460 + 5733 = 11193$ . Since  $p_0 = q_0$ , we have only one inequality to check. Half of 11193 is much more than 5; so, we have the requirements met. Our sample is large enough and was randomly selected.

$$\mu_{\hat{p}} = 0.50, \sigma_{\hat{p}} = \sqrt{\frac{0.5 \cdot 0.5}{11193}} \approx 0.0047, \text{ and } \hat{p} = \frac{5460}{11193} \approx 0.4878.$$

$$z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0 q_0}{n}}} \approx \frac{0.4878 - 0.50}{0.0047} \approx -2.5804$$

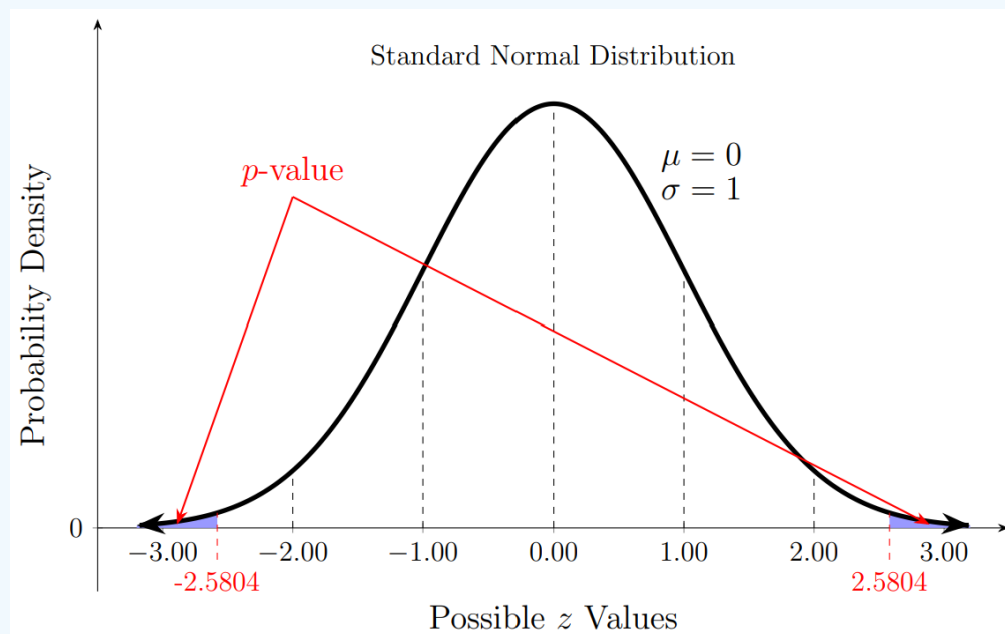


Figure 7.4.4 Standard normal distribution

$$p\text{-value} \approx 2 \cdot \text{NORM.S.DIST}(-2.5804, 1) \approx 0.0099$$

The  $p$ -value is just below the significance level of the test. We, therefore, have sufficient evidence to reject the null hypothesis in support of the notion that the proportion of females among newborn babies is not 50%.

- 3.3: Measures of Central Tendency by David Lane is licensed Public Domain. Original source: <https://onlinestatbook.com>.