⊙ **Guided Practice 3.7**   A manager is about to oversee the mass production of a new tire model in her factory, and she would like to estimate what proportion of these tires will be rejected through quality control. The quality control team has monitored the last three tire models produced by the factory, failing 1.7% of tires in the first model, 6.2% of the second model, and 1.3% of the third model. The manager would like to examine enough tires to estimate the failure rate of the new tire model to within about 2% with a 90% confidence level.[4]

(a) There are three different failure rates to choose from. Perform the sample size computation for each separately, and identify three sample sizes to consider.

(b) The sample sizes vary widely. Which of the three would you suggest using? What would influence your choice?

⊙ **Guided Practice 3.8**   A recent estimate of Congress' approval rating was 17%.[5] If we were to conduct a new poll and wanted an estimate with a margin of error smaller than about 0.04 with 95% confidence, how big of a sample should we use?[6]

3.2   Difference of two proportions

We would like to make conclusions about the difference in two population proportions $(p_1 - p_2)$ using the normal model. In this section we consider three such examples. In the first, we compare the approval of the 2010 healthcare law under two different question phrasings. In the second application, a company weighs whether they should switch to a higher quality parts manufacturer. In the last example, we examine the cancer risk to dogs from the use of yard herbicides.

In our investigations, we first identify a reasonable point estimate of $p_1 - p_2$ based on the sample. You may have already guessed its form: $\hat{p}_1 - \hat{p}_2$. Next, in each example we verify that the point estimate follows the normal model by checking certain conditions; as before, these conditions relate to independence of observations and checking for sufficiently large sample size. Finally, we compute the estimate's standard error and apply our inferential framework.

---

[4](a) For the 1.7% estimate of p, we estimate the appropriate sample size as follows:

$$1.65 \times \sqrt{\frac{p(1-p)}{n}} \approx 1.65 \times \sqrt{\frac{0.017(1 - 0.017)}{n}} \leq 0.02 \qquad \rightarrow \qquad n \geq 113.7$$

Using the estimate from the first model, we would suggest examining 114 tires (round up!). A similar computation can be accomplished using 0.062 and 0.013 for p: 396 and 88.
  (b) We could examine which of the old models is most like the new model, then choose the corresponding sample size. Or if two of the previous estimates are based on small samples while the other is based on a larger sample, we should consider the value corresponding to the larger sample. (Answers will vary.)
    [5]www.gallup.com/poll/155144/Congress-Approval-June.aspx
    [6]We complete the same computations as before, except now we use 0.17 instead of 0.5 for p:

$$1.96 \times \sqrt{\frac{p(1-p)}{n}} \approx 1.96 \times \sqrt{\frac{0.17(1 - 0.17)}{n}} \leq 0.04 \qquad \rightarrow \qquad n \geq 338.8$$

A sample size of 339 or more would be reasonable.

3.2.1 Sample distribution of the difference of two proportions

We must check two conditions before applying the normal model to $\hat{p}_1 - \hat{p}_2$. First, the sampling distribution for each sample proportion must be nearly normal, and secondly, the samples must be independent. Under these two conditions, the sampling distribution of $\hat{p}_1 - \hat{p}_2$ may be well approximated using the normal model.

Conditions for the sampling distribution of $\hat{p}_1 - \hat{p}_2$ to be normal
The difference $\hat{p}_1 - \hat{p}_2$ tends to follow a normal model when

- each proportion separately follows a normal model, and
- the two samples are independent of each other.

The standard error of the difference in sample proportions is

$$SE_{\hat{p}_1 - \hat{p}_2} = \sqrt{SE_{\hat{p}_1}^2 + SE_{\hat{p}_2}^2} = \sqrt{\frac{p_1(1-p_1)}{n_1} + \frac{p_2(1-p_2)}{n_2}} \tag{3.9}$$

where p_1 and p_2 represent the population proportions, and n_1 and n_2 represent the sample sizes.

3.2.2 Intervals and tests for $p_1 - p_2$

In the setting of confidence intervals, the sample proportions are used to verify the success-failure condition and also compute standard error, just as was the case with a single proportion.

● **Example 3.10** The way a question is phrased can influence a person's response. For example, Pew Research Center conducted a survey with the following question:[7]

> As you may know, by 2014 nearly all Americans will be required to have health insurance. [People who do not buy insurance will pay a penalty] while [People who cannot afford it will receive financial help from the government]. Do you approve or disapprove of this policy?

For each randomly sampled respondent, the statements in brackets were randomized: either they were kept in the order given above, or the two statements were reversed. Table 3.2 shows the results of this experiment. Create and interpret a 90% confidence interval of the difference in approval.

First the conditions must be verified. Because each group is a simple random sample from less than 10% of the population, the observations are independent, both within the samples and between the samples. The success-failure condition also holds for each sample. Because all conditions are met, the normal model can be used for the point estimate of the difference in support, where p_1 corresponds to the original ordering and p_2 to the reversed ordering:

$$\hat{p}_1 - \hat{p}_2 = 0.47 - 0.34 = 0.13$$

---

[7]www.people-press.org/2012/03/26/public-remains-split-on-health-care-bill-opposed-to-mandate/.
Sample sizes for each polling group are approximate.

	Sample size (n_i)	Approve law (%)	Disapprove law (%)	Other
"people who cannot afford it will receive financial help from the government" is given second	771	47	49	3
"people who do not buy it will pay a penalty" is given second	732	34	63	3

Table 3.2: Results for a Pew Research Center poll where the ordering of two statements in a question regarding healthcare were randomized.

The standard error may be computed from Equation (3.9) using the sample proportions:

$$SE \approx \sqrt{\frac{0.47(1-0.47)}{771} + \frac{0.34(1-0.34)}{732}} = 0.025$$

For a 90% confidence interval, we use $z^\star = 1.65$:

$$\text{point estimate} \pm z^\star SE \quad \rightarrow \quad 0.13 \pm 1.65 \times 0.025 \quad \rightarrow \quad (0.09, 0.17)$$

We are 90% confident that the approval rating for the 2010 healthcare law changes between 9% and 17% due to the ordering of the two statements in the survey question. The Pew Research Center reported that this modestly large difference suggests that the opinions of much of the public are still fluid on the health insurance mandate.

⊙ **Guided Practice 3.11** A remote control car company is considering a new manufacturer for wheel gears. The new manufacturer would be more expensive but their higher quality gears are more reliable, resulting in happier customers and fewer warranty claims. However, management must be convinced that the more expensive gears are worth the conversion before they approve the switch. If there is strong evidence of a more than 3% improvement in the percent of gears that pass inspection, management says they will switch suppliers, otherwise they will maintain the current supplier. Set up appropriate hypotheses for the test.[8]

● **Example 3.12** The quality control engineer from Guided Practice 3.11 collects a sample of gears, examining 1000 gears from each company and finds that 899 gears pass inspection from the current supplier and 958 pass inspection from the prospective supplier. Using these data, evaluate the hypothesis setup of Guided Practice 3.11 using a significance level of 5%.

First, we check the conditions. The sample is not necessarily random, so to proceed we must assume the gears are all independent; for this sample we will suppose this assumption is reasonable, but the engineer would be more knowledgeable as to whether this assumption is appropriate. The success-failure condition also holds for each sample. Thus, the difference in sample proportions, $0.958 - 0.899 = 0.059$, can be said to come from a nearly normal distribution.

---

[8]H_0: The higher quality gears will pass inspection no more than 3% more frequently than the standard quality gears. $p_{highQ} - p_{standard} = 0.03$. H_A: The higher quality gears will pass inspection more than 3% more often than the standard quality gears. $p_{highQ} - p_{standard} > 0.03$.
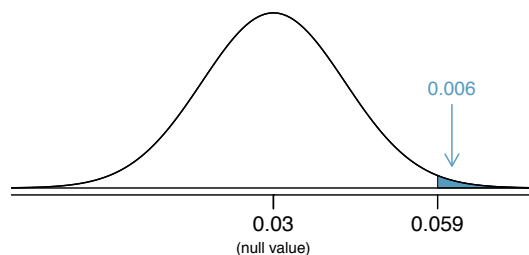
Figure 3.3: Distribution of the test statistic if the null hypothesis was true. The p-value is represented by the shaded area.

The standard error can be found using Equation (3.9):

$$SE = \sqrt{\frac{0.958(1 - 0.958)}{1000} + \frac{0.899(1 - 0.899)}{1000}} = 0.0114$$

In this hypothesis test, the sample proportions were used. We will discuss this choice more in Section 3.2.3.

Next, we compute the test statistic and use it to find the p-value, which is depicted in Figure 3.3.

$$Z = \frac{\text{point estimate} - \text{null value}}{SE} = \frac{0.059 - 0.03}{0.0114} = 2.54$$

Using the normal model for this test statistic, we identify the right tail area as 0.006. Since this is a one-sided test, this single tail area is also the p-value, and we reject the null hypothesis because 0.006 is less than 0.05. That is, we have statistically significant evidence that the higher quality gears actually do pass inspection more than 3% as often as the currently used gears. Based on these results, management will approve the switch to the new supplier.

3.2.3 Hypothesis testing when $H_0 : p_1 = p_2$

Here we use a new example to examine a special estimate of standard error when $H_0 : p_1 = p_2$. We investigate whether there is an increased risk of cancer in dogs that are exposed to the herbicide 2,4-dichlorophenoxyacetic acid (2,4-D). A study in 1994 examined 491 dogs that had developed cancer and 945 dogs as a control group.[9] Of these two groups, researchers identified which dogs had been exposed to 2,4-D in their owner's yard. The results are shown in Table 3.4.

---

[9]Hayes HM, Tarone RE, Cantor KP, Jessen CR, McCurnin DM, and Richardson RC. 1991. Case-Control Study of Canine Malignant Lymphoma: Positive Association With Dog Owner's Use of 2, 4-Dichlorophenoxyacetic Acid Herbicides. Journal of the National Cancer Institute 83(17):1226-1231.

	cancer	no cancer
2,4-D	191	304
no 2,4-D	300	641

Table 3.4: Summary results for cancer in dogs and the use of 2,4-D by the dog's owner.

⊙ **Guided Practice 3.13**   Is this study an experiment or an observational study?[10]

⊙ **Guided Practice 3.14**   Set up hypotheses to test whether 2,4-D and the occurrence of cancer in dogs are related. Use a one-sided test and compare across the cancer and no cancer groups.[11]

● **Example 3.15**   Are the conditions met to use the normal model and make inference on the results?

(1) It is unclear whether this is a random sample. However, if we believe the dogs in both the cancer and no cancer groups are representative of each respective population and that the dogs in the study do not interact in any way, then we may find it reasonable to assume independence between observations. (2) The success-failure condition holds for each sample.

Under the assumption of independence, we can use the normal model and make statements regarding the canine population based on the data.

In your hypotheses for Guided Practice 3.14, the null is that the proportion of dogs with exposure to 2,4-D is the same in each group. The point estimate of the difference in sample proportions is $\hat{p}_c - \hat{p}_n = 0.067$. To identify the p-value for this test, we first check conditions (Example 3.15) and compute the standard error of the difference:

$$SE = \sqrt{\frac{p_c(1 - p_c)}{n_c} + \frac{p_n(1 - p_n)}{n_n}}$$

In a hypothesis test, the distribution of the test statistic is always examined as though the null hypothesis is true, i.e. in this case, $p_c = p_n$. The standard error formula should reflect this equality in the null hypothesis. We will use p to represent the common rate of dogs that are exposed to 2,4-D in the two groups:

$$SE = \sqrt{\frac{p(1 - p)}{n_c} + \frac{p(1 - p)}{n_n}}$$

---

[10]The owners were not instructed to apply or not apply the herbicide, so this is an observational study. This question was especially tricky because one group was called the *control group*, which is a term usually seen in experiments.

[11]Using the proportions within the cancer and no cancer groups may seem odd. We intuitively may desire to compare the fraction of dogs with cancer in the 2,4-D and no 2,4-D groups, since the herbicide is an explanatory variable. However, the cancer rates in each group do not necessarily reflect the cancer rates in reality due to the way the data were collected. For this reason, computing cancer rates may greatly alarm dog owners.

H_0: the proportion of dogs with exposure to 2,4-D is the same in "cancer" and "no cancer" dogs, $p_c - p_n = 0$.
H_A: dogs with cancer are more likely to have been exposed to 2,4-D than dogs without cancer, $p_c - p_n > 0$.

We don't know the exposure rate, p, but we can obtain a good estimate of it by *pooling* the results of both samples:

$$\hat{p} = \frac{\text{\# of "successes"}}{\text{\# of cases}} = \frac{191 + 304}{191 + 300 + 304 + 641} = 0.345$$

This is called the **pooled estimate** of the sample proportion, and we use it to compute the standard error when the null hypothesis is that $p_1 = p_2$ (e.g. $p_c = p_n$ or $p_c - p_n = 0$). We also typically use it to verify the success-failure condition.

Pooled estimate of a proportion
When the null hypothesis is $p_1 = p_2$, it is useful to find the pooled estimate of the shared proportion:

$$\hat{p} = \frac{\text{number of "successes"}}{\text{number of cases}} = \frac{\hat{p}_1 n_1 + \hat{p}_2 n_2}{n_1 + n_2}$$

Here $\hat{p}_1 n_1$ represents the number of successes in sample 1 since

$$\hat{p}_1 = \frac{\text{number of successes in sample 1}}{n_1}$$

Similarly, $\hat{p}_2 n_2$ represents the number of successes in sample 2.

TIP: Use the pooled proportion estimate when $\mathbf{H_0 : p_1 = p_2}$
When the null hypothesis suggests the proportions are equal, we use the pooled proportion estimate ($\hat{p}$) to verify the success-failure condition and also to estimate the standard error:

$$SE = \sqrt{\frac{\hat{p}(1 - \hat{p})}{n_1} + \frac{\hat{p}(1 - \hat{p})}{n_2}} \tag{3.16}$$

⊙ **Guided Practice 3.17** Using Equation (3.16), $\hat{p} = 0.345$, $n_1 = 491$, and $n_2 = 945$, verify the estimate for the standard error is $SE = 0.026$. Next, complete the hypothesis test using a significance level of 0.05. Be certain to draw a picture, compute the p-value, and state your conclusion in both statistical language and plain language.[12]

---

[12]Compute the test statistic:

$$Z = \frac{\text{point estimate} - \text{null value}}{SE} = \frac{0.067 - 0}{0.026} = 2.58$$

We leave the picture to you. Looking up $Z = 2.58$ in the normal probability table: 0.9951. However this is the lower tail, and the upper tail represents the p-value: $1 - 0.9951 = 0.0049$. We reject the null hypothesis and conclude that dogs getting cancer and owners using 2,4-D are associated.