Chapter 8

Section 8.1

We might be interested in

- the proportion of US males who have health insurance,
- the proportion of imported cars in the US,
- the proportion of Americans who have Type II diabetes,

Notations:

p: Population Proportion (the proportion of the entire population that has the specified attribute.)

$\hat{P} = \frac{x}{n}$: Sample Proportion (the proportion of a sample from the population that has the specified attribute.)

Example 1: A sample of 200 married couples selected from throughout the United States showed that for 84 of the couples, both the husband and wife held full-time jobs. Use the sample results to estimate the proportion of all married couples in the United States for which both the husband and wife hold full-time jobs. Solution. 0.42

The Sampling Distribution of the proportion

In selecting random samples of size n from a particular population with a proportion of p, the sampling distribution of the sample proportion $\hat{P}$ approaches a normal distribution with mean of **p** and the standard deviation of $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ as the sample size becomes large; $\hat{P} \sim N\left(p, \sqrt{\frac{p(1-p)}{n}}\right)$. When is the sample size considered large? If np ≥ 10 and n(1 – p) ≥ 10

(1) If p is unknown, then $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$ is estimated by $\sigma_{\hat{p}} = \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

(2) A $(1-\alpha)\%$ C.I. for estimating p is $\hat{p} \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$

Example 2: A particular county in West Virginia has a 9% unemployment rate. To monitor the unemployment rate in that county, a monthly survey of 800 individuals is conducted by a state agency.

(a) What is the sampling distribution of $\hat{P}$ when a sample size of 800 is used?

Solution (a) $\hat{P}$ **~ N(0.09, 0.0101)**

Example 3. In a Roper Organizational poll of 2,000 adults, 1,280 have money in regular saving accounts. Find a 95% confidence interval for the true proportion of adults who have money in regular saving accounts.

Solution. $\hat{p} = \frac{x}{n} = \frac{1280}{2000} = 0.64$, n=2000

$$\hat{p} \pm Z_{\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1 - \hat{p})}{n}} \Rightarrow 0.64 \pm (1.96)\sqrt{\frac{0.64(1 - 0.64)}{2000}} \Rightarrow 0.64 \pm (1.96)(0.011)$$
$$\Rightarrow 0.64 \pm 0.022 \Rightarrow (0.618, 0.662)$$

We are 95% confident that the percentage of adults having money in the saving account is **61.8% to 66.2%.**

Determination of the Sample Size: **Let m =** $Z_{\frac{\alpha}{2}}\sqrt{\frac{p(1 - p)}{n}}$

Solving for n gives us, n = $\left(\frac{Z_{\frac{\alpha}{2}}}{m}\right)^2 p(1-p)$ Note: If p is unknown take it to be $p = 0.5$. $p = 0.5$ it gives the maximum sample size.

Example 4. We want to estimate with a maximum error of 3%, the true proportion of VSU students who would like Friday classes during the summer semester and we want 96% confidence in our results. How many VSU students must we survey?

Solution. n = $\left(\frac{Z_{\frac{\alpha}{2}}}{m}\right)^2 \hat{p}(1-\hat{p}) = \left(\frac{2.05}{0.03}\right)^2 (0.5)(0.5) = 1167.36$. So, we need to survey 1168 students.

Example 5. A National survey of registered voters is being conducted to determine the proportion of voters who favor a particular candidate. Assume that the desired confidence level is 95% and that the desired maximum sampling error is 0.02 percent.

(a) How large a sample is needed if it is believed that approximately 35% of the population currently supports the candidate?

Solution. n = $\left(\frac{Z_{\frac{\alpha}{2}}}{m}\right)^2 \hat{p}(1-\hat{p}) = \left(\frac{1.96}{0.02}\right)^2 (0.35)(0.65) = 2184.91$.

So, we need to survey 2185 voters.

(b) How large a sample is needed if no information is available on the proportion of voters currently supporting the candidate?

Solution. $n = \left(\frac{Z_{\frac{\alpha}{2}}}{m}\right)^2 \hat{p}(1-\hat{p}) = \left(\frac{1.96}{0.02}\right)^2 (0.5)(0.5) = 2401$.

So, we need to survey 2401 voters.

Hypothesis Testing – One Population Proportion

General form of Hypothesis Testing

Step 1	$Case 1.\quad H_0: p \geq p_0$ $H_a: p < p_0$	$Case 2.\quad H_0: p \leq p_0$ $H_a: p > p_0$	$Case 3.\quad H_0: p = p_0$ $H_a: p \neq p_0$
Step 2	$Z^* = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	$Z^* = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$	$Z^* = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$
Step 3	$\text{Reject } H_0 \text{ if } Z^* < -Z_\alpha$	$\text{Reject } H_0 \text{ if } Z^* > Z_\alpha$	$\text{Reject } H_0 \text{ if } \lvert Z^* \rvert > Z_{\frac{\alpha}{2}}$
Step 4	Conclusion: We do not reject H_0 and conclude that .... We reject H_0 and conclude that		

Example 6. Mr. Dixon, a Republican, claims that he has the support of 55% of all voters in the 23rd U.S. Congressional District. Can the Central Committee conclude that less than 55% of all voters support Mr. Dixon, if, out of a random sample of 500 registered voters, only 245 expressed their preference for Mr. Dixon? Use α = 0.01 as the level of significance.

1. $H_0: p = 0.55 \quad vs \quad H_a: p < 0.55 \quad (\textit{Lower Tail Test})$
2. Test Statistic: $Z^* = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{0.49-0.55}{\sqrt{\frac{0.55(1-0.55)}{500}}} = -2.7$
3. $\text{Reject } H_0 \text{ if } Z^* < -Z_{\alpha=0.01} = -2.33$
4. **Conclusion:** Since Z^* = -2.7 is less than –2.33, we reject H_0 and conclude that the Central Committee has enough evidence that less than 55% of all voters support Mr. Dixon.
Note: P-Value = P(Z < Z*) = P(Z < -2.7) = 0.0035   P-Value = 0.0035 < α = 0.01. Therefore, we reject H_0 and draw the same conclusion as in step 4 above.

Example 7. The manager of an Italian restaurant is considering opening a carryout food service. However, the manager is concerned that not all individuals placing order by phone actually pickup the order. If 90% or less of the phone orders will be picked up, the restaurant will not have a profitable operation. However, if it can be concluded that more than 90% of the phone orders will be picked up the carryout operation will be a worthwhile addition for the restaurant. During a 2-week period, 234 orders in a sample of 250 phone orders were picked up. Test the appropriate hypothesis at α = 0.05 level of significance.

1. $H_0 : p = 0.90 \quad vs \quad H_a : p > 0.90 \ \ (Upper\ Tail\ Test)$
2. Test Statistic: $Z^* = \dfrac{\hat{p} - p_0}{\sqrt{\dfrac{p_0(1-p_0)}{n}}} = \dfrac{0.936 - 0.90}{\sqrt{\dfrac{0.90(1-0.90)}{250}}} = 1.9$
3. $\text{Re}\,ject\ H_0\ if\ Z^* > Z_{\alpha=0.05} = 1.645$
4. **Conclusion:** Since Z^* = 1.9 is greater than 1.645, we reject H_0 and conclude that more than 90% of the phone orders will be picked up and the carryout operation should be implemented.
Note: P-Value = P(Z > Z*) = P(Z >1.9) = 1-P(Z <1.9) =1-0.9713 = 0.0287 P-Value = 0.0287 < α = 0.05. Therefore, we reject H_0 and draw the same conclusion as in step 4 above.

Homework: 8.30 (a), 8.31, 8.32, 8.33, 8.34, 8.35(a), 8.36, 8.37(a), 8.38(a and c), 8.39, 8.40, 8.41, 8.42 pages 503-504

Section 8.2 Comparing Two Population Proportions $\left(P_1 - P_2\right)$.

The estimator for the difference of two population proportions is $\left(\hat{P}_1 - \hat{P}_2\right)$.

$$\textit{Note}: \quad E\left(\hat{P}_1 - \hat{P}_2\right) = P_1 - P_2 \quad \textit{and} \quad \sigma_{\hat{P}_1 - \hat{P}_2} = \sqrt{\frac{P_1\left(1 - P_1\right)}{n_1} + \frac{P_2\left(1 - P_2\right)}{n_2}}$$

Based on the C.L.T.: If $n_1 p_1 \ge 10$, $n_1\left(1 - p_1\right) \ge 10$, $n_2 p_2 \ge 10$, $n_2\left(1 - p_2\right) \ge 10$, then

$$\left(\hat{P}_1 - \hat{P}_2\right) \sim N\left(P_1 - P_2 \, , \, \sqrt{\frac{P_1\left(1 - P_1\right)}{n_1} + \frac{P_2\left(1 - P_2\right)}{n_2}} \right).$$

$\left(1-\alpha\right)\%$ **Confidence Interval for** $\left(P_1 - P_2\right)$ **.**

$$\left(\hat{P}_1 - \hat{P}_2\right) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{P}_1\left(1 - \hat{P}_1\right)}{n_1} + \frac{\hat{P}_2\left(1 - \hat{P}_2\right)}{n_2}}$$

Example 1 : The following table shows the result on a recent survey among men and women in supporting the president's new tax law. Find a 95% confidence interval for the difference of the two population proportions that are currently supporting the president's new tax law.

Population	n	x	$\hat{P} = \frac{x}{n}$
1 (Men)	7180	1630	0.227
2 (Women)	9916	1684	0.170
Total	17096	3314	0.194

$$\left(\hat{P}_1 - \hat{P}_2\right) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{P}_1\left(1 - \hat{P}_1\right)}{n_1} + \frac{\hat{P}_2\left(1 - \hat{P}_2\right)}{n_2}} \Rightarrow \left(0.227 - 0.170\right) \pm \left(1.96\right)\sqrt{\frac{0.227\left(1 - 0.227\right)}{7180} + \frac{0.170\left(1 - 0.170\right)}{9916}}$$

$$\Rightarrow \left(0.057\right) \pm \left(1.96\right)\left(0.00622\right) \Rightarrow \left(0.057\right) \pm \left(0.012\right) \Rightarrow \left(0.045, \, 0.069\right)$$

We are 95% confident that the proportion of men that support the new tax law is 4.5% to 6.9% higher than the proportion of women.

Hypothesis Testing: $H_0 : P_1 - P_2 = 0$

We are doing hypothesis testing under the assumption that the Null Hypothesis, $H_0 : P_1 - P_2 = 0$, is true, meaning that the two population proportions are the same; i.e. $P_1 = P_2 = P$. Since the two population proportions are the same, the standard deviation of $\left(\hat{P}_1 - \hat{P}_2\right)$ is

$$\sigma_{\hat{P}_1 - \hat{P}_2} = \sqrt{\frac{P_1(1-P_1)}{n_1} + \frac{P_2(1-P_2)}{n_2}} = \sqrt{\frac{P(1-P)}{n_1} + \frac{P(1-P)}{n_2}} = \sqrt{P(1-P)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

. We estimate the common value of **P** , by the overall proportion of success in both samples. We are using $\hat{\mathbf{P}}$ to estimate **P** ; i.e. $\hat{\mathbf{P}} = \frac{X_1 + X_2}{n_1 + n_2}$. $\hat{\mathbf{P}}$ is called the pooled estimate because we are combining or pooling, the information from both samples for our estimate. Hence, the pooled standard deviation is denoted by $S_{\mathbf{P}}$ and is given by the following formula, $S_{\mathbf{P}} = \sqrt{\hat{P}\left(1-\hat{P}\right)\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ where

$$\hat{\mathbf{P}} = \frac{X_1 + X_2}{n_1 + n_2}.$$

Hypothesis Testing – Two Population Proportions

General form of Hypothesis Testing

Step 1	$Case 1.\quad H_0 : p_1 - p_2 = 0$ $H_a : p_1 - p_2 < 0$	$Case 2.\quad H_0 : p_1 - p_2 = 0$ $H_a : p_1 - p_2 > 0$	$Case 3.\quad H_0 : p_1 - p_2 = 0$ $H_a : p_1 - p_2 \neq 0$
Step 2	$Z^* = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{S_p}$ *where* $S_p = \sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$ *and* $\hat{\mathbf{P}} = \frac{X_1 + X_2}{n_1 + n_2}$		
Step 3	$Reject\ H_0\ if\ Z^* < -Z_\alpha$	$Reject\ H_0\ if\ Z^* > Z_\alpha$	$Reject\ H_0\ if\ \lvert Z^* \rvert > Z_{\frac{\alpha}{2}}$
Step 4	Conclusion: We do not reject H_0 and conclude that .... We reject H_0 and conclude that		

Example 2: In Example 1, do hypothesis testing. Do we have enough evidence that more men support the new tax law than women? Test it at $\alpha = 0.05$.

$$S_p = \sqrt{0.194(1-0.194)\left(\tfrac{1}{7180}+\tfrac{1}{9916}\right)} = 0.006126 \quad and \quad \hat{\mathbf{P}} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{1630+1684}{7180+9916} = \frac{3314}{17096} = 0.194$$

1. $H_0 : P_1 - P_2 = 0 \quad \text{vs} \quad H_a : P_1 - P_2 > 0 \quad (\textit{Upper Tail Test})$
2. $Z^* = \dfrac{(\hat{p}_1 - \hat{p}_2) - 0}{S_p} = \dfrac{(0.227 - 0.170) - 0}{0.006126} = 9.34 \quad where\ S_p = 0.006126\ and\ \hat{\mathbf{P}} = 0.194$
3. $\text{Re}\,ject\ H_0\ if\ Z^* > Z_{\alpha=0.05} = 1.645$
4. **<u>Conclusion:</u>** Since Z^*= 9.34 is greater than 1.645, we reject H_0 and conclude that more men will support the new tax law than women.
Note: P-Value = P(Z > Z^*) = P(Z >9.34) = 1-P(Z <9.34) =1-1 = 0 P-Value = 0 < α = 0.05. Therefore, we reject H_0 and draw the same conclusion as in step 4 above.

Example 3: The power take-off(PTO) driveline on farm tractors is a potentially serious hazard to farmers. A shield covers the driveline on new tractors, but for a variety of reasons, the shield is often missing on older tractors. Two types of shield are the bolt-on and the flip-up. A study initiated by the National Safety Council took a sample of older tractors to examine the proportions of shields removed. The study found that 35 shields had been removed from the 83 tractors having bolt-on shields and 15 had been removed from 136 tractors with flip-up shields.

(a) Test the hypothesis that there is no difference in the two proportions of the type of shields. (use α = 0.01)

$$S_p = \sqrt{0.2283(1-0.2283)\left(\tfrac{1}{83}+\tfrac{1}{136}\right)} = 0.0585 \quad and \quad \hat{\mathbf{P}} = \frac{X_1 + X_2}{n_1 + n_2} = \frac{35+15}{83+136} = \frac{50}{219} = 0.2283$$

1. $H_0: P_1 - P_2 = 0 \quad vs \quad H_a: P_1 - P_2 \neq 0 \quad (\textit{Two Tail Test})$
2. $Z^* = \frac{(\hat{p}_1 - \hat{p}_2) - 0}{S_p} = \frac{(0.4217 - 0.1103) - 0}{0.0585} = 5.33 \quad where\ S_p = 0.0585\ and\ \hat{\mathbf{P}} = 0.2283$
3. $\text{Re}\,ject\ H_0\ if\ \lvert Z^* \rvert > Z_{\frac{\alpha}{2}=0.005} = 2.575$
4. **<u>Conclusion:</u>** Since Z^* = 5.33 is greater than 2.575, we reject H_0 and conclude that there is a significant difference in the two shield types.
Note: P-Value = $2P\left(Z < -\lvert Z^* \rvert\right)$ = 2P(Z <-5.33) = 2.0=0 P-Value = 0 < α = 0.01. Therefore, we reject H_0 and draw the same conclusion as in step 4 above.

(b) Give a 90% confidence interval for the proportions of removed shields for the bolt-on and the flip-up types. Based on the data, what recommendation would you make about the type of shield to be used on new tractors?

$$\left(\hat{P}_1 - \hat{P}_2\right) \pm Z_{\frac{\alpha}{2}} \sqrt{\frac{\hat{P}_1\left(1 - \hat{P}_1\right)}{n_1} + \frac{\hat{P}_2\left(1 - \hat{P}_2\right)}{n_2}} \Rightarrow \left(0.4217 - 0.1103\right) \pm \left(1.645\right) \sqrt{\frac{0.4217\left(1 - 0.4217\right)}{83} + \frac{0.1103\left(1 - 0.1103\right)}{136}}$$

$$\Rightarrow (0.3114) \pm (1.645)(0.0605) \Rightarrow (0.3114) \pm (0.0995) \Rightarrow (0.2119,\ 0.4109)$$

We are 90% confident that the proportion of bolt-on shied removed is 21.19% to 41.09% higher than the proportion of flip-up shied removed. Hence, the flip-up shield shields are much more likely to remain on the tractor.

Homework: 8.63(a,b,c,d, and e - Only), 8.67 (a,b,c,d, and e - Only), 8.74 pages 520-522