Chapter 5 Scenarios

1918 Spanish Flu

The 1918 flu pandemic was the first of two H1N1 pandemics to sweep the globe (the other was the 2009 "swine flu"). During the first year of the pandemic, life expectancy in the United States dropped by 12 years. While it is unclear where the virus first mutated into the pandemic strain, the disease was called the "Spanish flu" because Spain was uninvolved in WWI and thus did not censor reporting on the pandemic; thus, it seemed that Spain was hit harder than the rest of the world because articles were actually published about it. Globally, there were 3 waves of the H1N1 flu pandemic; by far the most deadly was the wave which hit in the fall of 1918.

One city which was particularly hard hit during the fall of 1918 was Philadelphia; about one week after the flu was introduced to the city via the Navy Yard, a wartime parade was held as scheduled. Shortly thereafter, influenza cases exploded across the city. It is estimated that 200,000 people attended the parade, and within 3 days, there were 635 new civilian cases. One of the biggest factors in Philadelphia's flu mortality rate during the pandemic was that medical facilities were already strained due to the war; as a result, there was no capacity to treat the sudden increase in severe influenza cases. It is estimated that in Philadelphia, a city of 1.8 million people, there were 12,000 deaths during the fall 1918 wave of the pandemic.

The city of Houston, by contrast, instituted a strict quarantine during the same period, cancelling public events, closing schools, and encouraging people to avoid public transportation. During the same period, Houston recorded 111 deaths due to flu, in a population of 130,000.

1. What is the observational unit?

   An individual living in Houston or Philadelphia

2. Fill in the contingency table below with the correct numbers.

	Died from flu	Survived	Total
Philadelphia	*12,000*	*1,788,000*	*1,800,000*
Houston	*111*	*129,889*	*130,000*
Total	*12,111*	*1,917,889*	*1,930,000*

3. For each city, calculate the marginal *crude fatality rate*, that is, the total number of deaths divided by the total population. The crude fatality rate is lower than the *case fatality rate*, which uses the number of confirmed cases as the denominator.

   Philadelpha: $\hat{p}_P = \frac{12000}{1800000} \approx 0.00667$

   Houston: $\hat{p}_H = \frac{111}{130000} \approx 0.000854$

4. Calculate the risk of dying of the 1918 flu pandemic in Philadelphia relative to the risk of dying of the flu in Houston during the same period. Interpret the number you get in the context of the problem.

   $$RR_{P,H} = \frac{\hat{p}_P}{\hat{p}_H} = \frac{0.00667}{0.000854} \approx 7.808$$

   Residents of Philadelphia were 7.8 times more likely to die during the fall 1918 wave of the Spanish flu pandemic than residents of Houston during the same period.

5. What type of study is this?

   This is an observational study; we did not assign individuals to live in Philadelphia or Houston in 1918, nor did we determine the public health measures taken in each respective city during that pandemic.

6. Public health authorities would like to know if the interventions in Houston ("social distancing" methods, in modern terms) are associated with a reduced death rate from the flu. Write out the appropriate null and alternative hypotheses in both words and symbols.

   $H_0 : \pi_P - \pi_H = 0$*, that is, there is no difference in the crude death rate from the flu in Philadelphia compared to the crude death rate from the flu Houston.*

   $H_A : \pi_P - \pi_H > 0$*, that is, the death rate in Houston is lower than the death rate in Philadelphia.*

   Note that you could also write $H_0 : \pi_P = \pi_H$ *and* $H_A : \pi_P > \pi_H$*; the interpretation remains the same.*

7. Do you think a simulation study is appropriate for this case? Why or why not? Note that the validity conditions might not be the only consideration: think about how you would physically implement a simulation study for this case.

   A simulation study is a valid option for this case, but because of the large populations involved, it will be extremely slow. Theory-based inference is a better choice for this data.

8. Calculate the standard error you would use to create a 95% confidence interval for the difference betwen the crude death rates in the two cities. Note that because the numbers are small, you may need to use scientific notation. $\sqrt{\dfrac{0.00667(1-0.00667)}{1800000} + \dfrac{0.000854(1-0.000854)}{130000}} = \sqrt{3.68\text{e}{-9} + 6.56\text{e}{-9}} \approx 0.0001012$

9. Construct a 2SD confidence interval for the difference in the crude death rate between the two cities.

   $$\hat{p}_P - \hat{p}_H \pm 2 * (SE) = 0.00667 - 0.000854 \pm 2 * 0.0001012 = (0.005614, 0.006018)$$

10. Interpret the interval you got in 9. Note that when working with rates, it may be easier to interpret and think about the values as fatality rate per 1000 - e.g. a rate of 0.005 would be read as 5 per 1000. What can you say about your null hypothesis in 6?

    We can be 95% confident that per 1000 residents, between 5.6 and 6.0 more died in Philadelphia than in Houston. As this interval does not contain 0, we can conclude that the death rate in Philadelphia was significantly higher than in Houston; the difference in the death rates cannot be explained by chance alone.

11. Does this mean we can conclude that the public health interventions and social distancing measures used in Houston were effective at reducing the death rate? Can you think of any potential confounding variables?

    No, because this is an observational study. Philadelphia and Houston were vastly different sizes, with different population density and climates (this is especially true in the fall). It is possible that the difference in the crude mortality rate is due to one of these other differences.

COVID-19

Environment

The Diamond Princess cruise ship was quarantined outside of Yokohama, Japan after it was discovered that several passengers were infected with SARS-nCOV-2, the virus that causes the illness known as COVID-19. During its almost 4-week quarantine outside Japan, over 705 individuals tested positive for the virus, out of 3711 on board. Due to the quarantine, and subsequent testing of all individuals on board, the ship represents one estimate of the infection rate of the virus in a relatively isolated population. This unique situation allows us to get a point estimate based on what is essentially a census of a population. In this situation, we can say $\hat{p}_s = \frac{705}{3711} \approx 0.1900$.

It is hard to get estimates of true fatality rates and infection rates in a population during an epidemic, as those with mild cases may not seek medical care. After the epidemic, retrospective studies can identify portions of the population with antibodies to the virus, which is a much more accurate estimate of overall infection rates. Nevertheless, it is common to roughly estimate parameters such as infectivity during an epidemic, if only to provide rough guidance on how best to allocate medical resources. Of the other countries experiencing coronavirus outbreaks as of March 10, 2020, South Korea has done the most thorough job of testing its population for infection, with 189,236 individuals tested and 7,382 individuals with confirmed cases.

Assuming for the moment that South Koreans tested for coronavirus are representative samples of the country's population, epidemiologists would like to know whether the infection rate in South Korea is consistent with the infection rate in a closed environment such as a cruise ship.

1. What is an observational unit?

   An individual tested for COVID-19

2. What is the variable of interest?

   Test result (positive, negative) or infection status (infected, not infected) is the variable of interest.

3. What is the population parameter of interest, in words and in symbols?

   We are interested in the long-run average rate of coronavirus infection in South Korea. We might write this as π_i, the infection rate.

4. What should the null and alternate hypotheses be? Treat the infection rate on the cruise ship as a fixed number, not a variable or parameter.

   $H_0 : \pi_s = 0.1900$
   $H_A : \pi_s \neq 0.1900$

5. What is the value of the sample statistic? Show your work.

The sample statistic is the infection rate in South Korea, which is $\hat{p}_i = \frac{7382}{189236} \approx 0.03901$

6. Create a 95% confidence interval for the infection rate in South Korea.

$$\hat{p}_i \pm 2\sqrt{\frac{\hat{p}_i(1-\hat{p}_i)}{n}} = 0.03901 \pm 2\sqrt{\frac{0.03901(0.96099)}{189236}} = 0.03901 \pm 2(0.000445) = (0.0381, 0.0399)$$

7. Interpret the interval in the context of the problem. What can you say about the null hypothesis?
Epidemiologists can be 95% confident that the infection rate in South Korea is between 3.80% and 3.99%. This is significantly lower than the infection rate on the Diamond Princess, suggesting that under normal living conditions, the infection rate is not equal to the 19% observed on the cruise ship.