R Assignment 4 SOLUTIONS

A confidence interval for one proportion

Schwab

2025-01-09

Intro/Goal:

In this assignment you will do make a 95% confidence interval for a single proportion. You will be guided to do an exploratory data analysis (EDA), check conditions for the central limit theorem, build the confidence interval, and interpret the confidence interval and finally write a conclusion. I will make a 90% confidence interval.

Diseases are contagious. Some more so than others. I would like to know if someone is exposed to smallpox, what is the chance they will die from the disease. We will explore this together twice. Once for people who are vaccinated against smallpox and once for people are are not. I assume the people who are unvaccinated have a lower rate of survival. Really we are trying to get a handle on how effective an early vaccine against the disease is. (This data is very old). We are trying to figure out the true proportion of individuals who die from smallpox after being exposed to it.

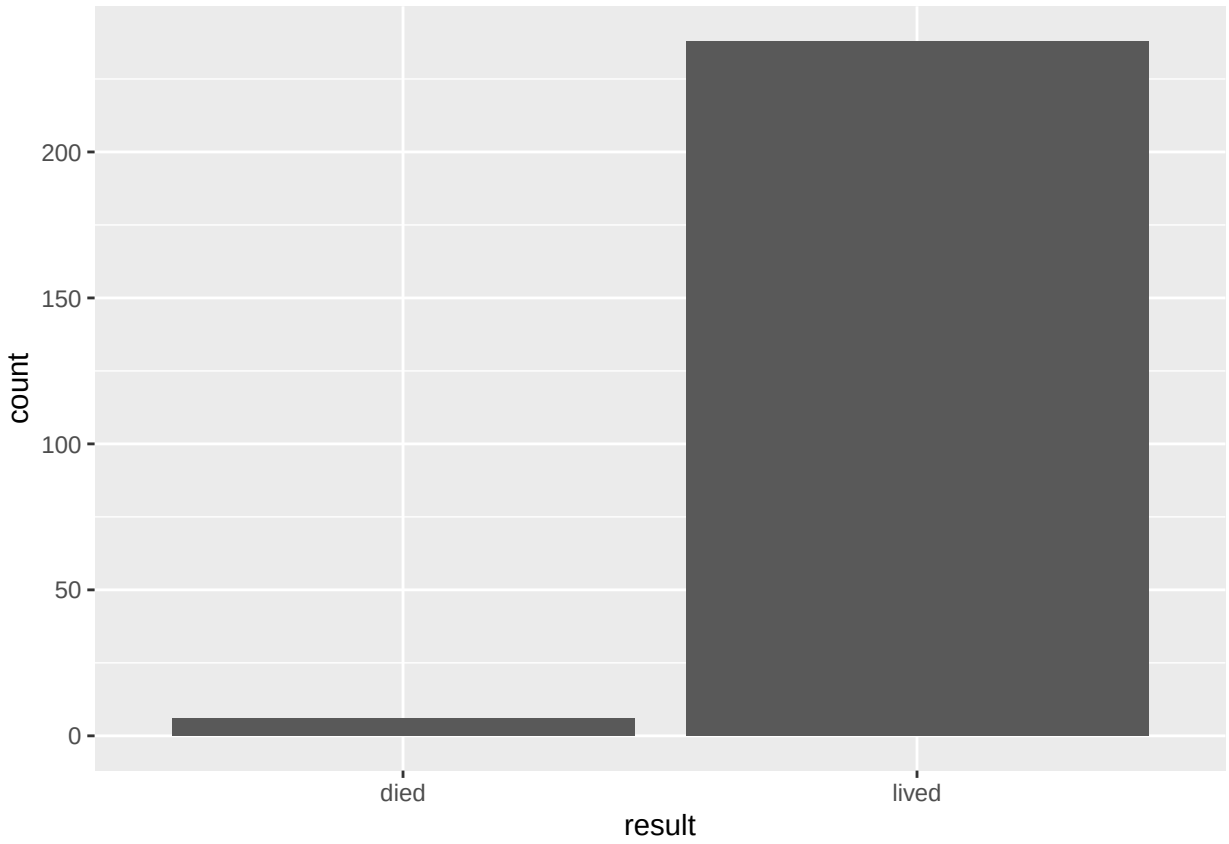
The data come from a dataset called smallpox. As part of your EDA you should look at the documentation for smallpox. Make a chunk and do that below.

However you will not use the smallpox dataframe as we want to work with vaccinated and unvaccinated people separately (we can work with them together, but that for the next R assignment. For this assignment you will use the not_vaccinated data frame. I will demonstrate a 90% confidence interval with the vaccinated data. Both dataframes have one column called result and it shows whether the person lived or died.

An exploratory data analysis.

You should have a look at the data, perhaps make a bar graph or table to see the current distribution of the data.

```
ggplot(vaccinated) +
  geom_bar(aes(x=result))
```



Check the conditions.*

We are using the Central Limit Theorem, thus we need to check conditions for a single proportion. We will do this by examining the data. This is a one proportion test so you should consult chapter 16 of IMS. I will produce code that counts the levels (lived or died) of the variable result. I do this with `count()` in the video. Be sure to type out whether the conditions are met or not and justify. Continue with the confidence interval even if conditions are not met.

```
count(vaccinated, result   )
```

```
## # A tibble: 2 x 2
##   result     n
##   <fct>  <int>
## 1 died       6
## 2 lived    238
```

My success-failure condition is questionable, as I do not have 10 successes and 10 failures. However I think its okay to relax this condition as often 5 successes and 5 failures is considered fine.

I have to assume independence of the data, and our sample size is large.

Make the confidence interval

Make a 90% confidence interval for the proportion of people who survived small pox exposure.

To find the confidence interval I'll have to do some math. First I'll find p-hat, then I'll calculate the standard error and margin of error. Finally I'll find the lower and upper bounds. See chapter 16 for the formulas.

```
p_hat = 238/244 # I got these numbers by looking at the count() above.
```

```r
SE = sqrt((1-p_hat)*p_hat/244)   # This is from the standard error formula
z_star = qnorm(p = 0.95)  # Yes that is the correct percent value for a 90% CI.
ME = z_star*SE   # The Margin of Error is the z_star times the Standard error.

print("Lower Bound")
```

```
## [1] "Lower Bound"
```

```r
p_hat - ME  # This calculate the lower bound for the confidence interval
```

```
## [1] 0.9591016
```

```r
print("Upper Bound")
```

```
## [1] "Upper Bound"
```

```r
p_hat + ME # This is the upper bound.
```

```
## [1] 0.9917181
```

Interprete the CI

With 90% confidence we can say that the true survival rate for people who were exposed to small pox who survived the disease is between 95.9% and 99.2%. This is a high survival rate. However we may want to consider bias in the data. People who were vaccinated may have been overall heather than those who were not.

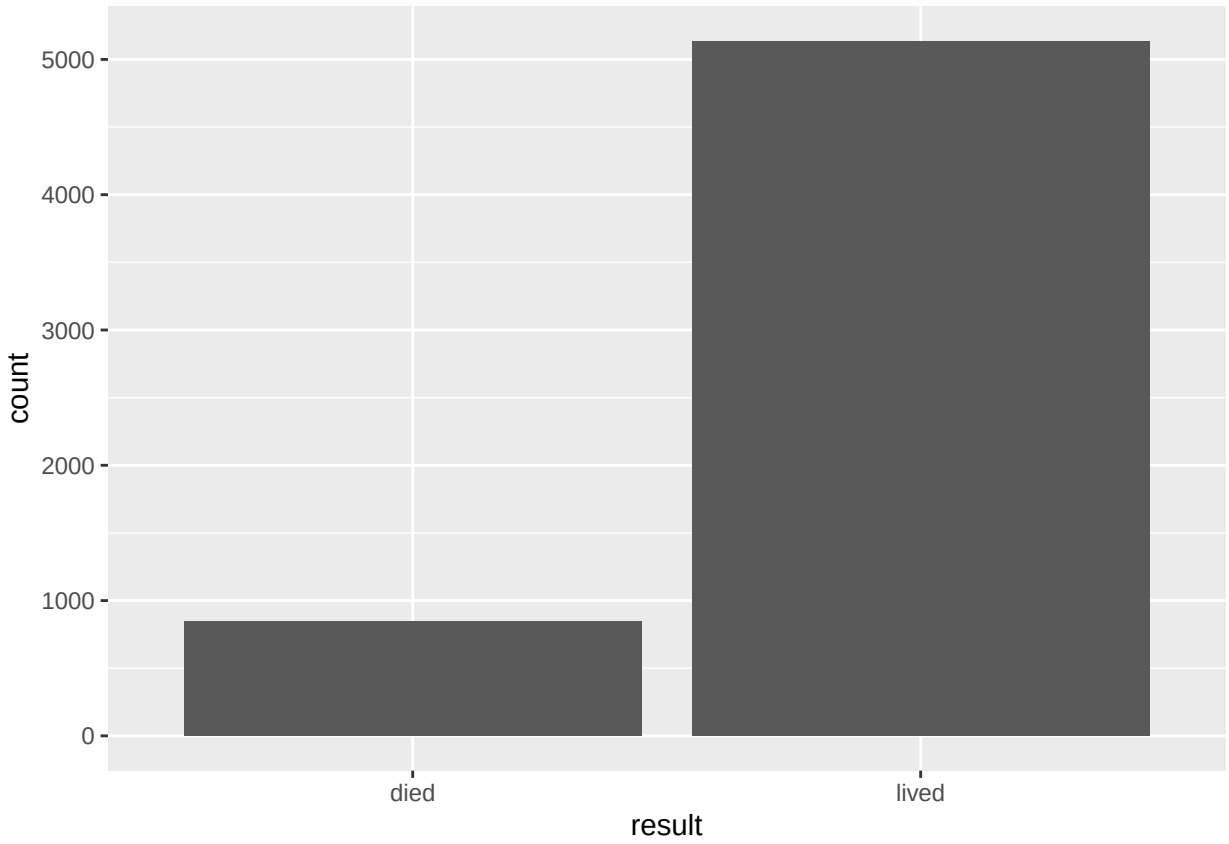
Exercises

You will make a 95% confidence interval for the proportion of people who survived exposure to smallpox and were not vaccinated. You may use and edit any code above, but you have to go through the steps. I've abbreviated the instructions slightly.

An exploratory data analysis.

Have a look at the not_vaccinated data, perhaps make a bar graph or table.

```r
ggplot(data = not_vaccinated)+
  geom_bar(aes(x=result))
```



Check the conditions.

You need to check that the conditions for this confidence interval are met by examining the data.

```
count(not_vaccinated, result   )
```

```
## # A tibble: 2 x 2
##   result     n
##   <fct>  <int>
## 1 died     844
## 2 lived   5136
```

Answer: Both the success failure condition and independence are met. We also have a large sample.

Make the confidence interval

Make a 95% confidence interval for the proportion of people who survived small pox exposure but were unvaccinated. Note: You may carefully copy and edit the code above. I made a 90% CI and you will make a 95% CI, also our p_hats will be different as you are working with the not_vaccinated data.

```
p_hat = 5136/5980 # I got these numbers by looking at the count() above.

SE = sqrt((1-p_hat)*p_hat/5980)    # This is from the standard error formula
z_star = qnorm(p = 0.975)  # Yes that is the correct percent value for a 95% CI.
ME = z_star*SE   # The Margin of Error is the z_star times the Standard error.

print("Lower Bound")
```

```
## [1] "Lower Bound"
```

```r
p_hat - ME  # This calculate the lower bound for the confidence interval
```

```
## [1] 0.8500386
```

```r
print("Upper Bound")
```

```
## [1] "Upper Bound"
```

```r
p_hat + ME # This is the upper bound.
```

```
## [1] 0.8676872
```

Interpret the CI

What does your confidence interval mean for those who survived smallpox?

Answer: The true probability of someone who is not vaccinated living after being exposed to small pox is between 85% and 87%. This is a good survival rate, but not nearly as good as those who were vaccinated. It seems that the vaccine works.

Compare the confidence intervals.

Does it seem like vaccination helps the chance of survival? What other factors might be confounding our test? (Note: It might help to this of the data's provenance, ie where the data came from)

Answer: From the basic documentation we don't know much about the 244 individuals that were given this experimental vaccine. They may have been in good health. Perhaps they had characteristics which would have made them more likely to survive small pox. Otherwise the vaccine seems like a great success.

Keep in mind these are people who lived after exposure. Just because they were exposed to smallpox does not mean they would have caught the disease. Small pox has a 70% survivability rate, if all those unvaccinated actually caught the disease after exposure we would expenct our confidence interval to be centered around 70%.