

[Course](#) > [Course Wrap-up an...](#) > [Comprehensive Ass...](#) > Brexit poll analysis ...

Brexit poll analysis - Part 2

This problem set is continued from the previous page. Make sure you have run the following code:

```
# suggested libraries and options
library(tidyverse)
options(digits = 3)
```

```
# load brexit_polls object and add x_hat column
library(dslabs)
data(brexit_polls)
brexit_polls <- brexit_polls %>%
  mutate(x_hat = (spread + 1)/2)
```

```
# final proportion voting "Remain"
p <- 0.481
```

Question 4: Confidence intervals for polls in June

4.0/4.0 points (graded)

Create the data frame `june_polls` containing only Brexit polls ending in June 2016 (`enddate` of "2016-06-01" and later). We will calculate confidence intervals for all polls and determine how many cover the true value of d .

First, use `mutate` to calculate a plug-in estimate `se_x_hat` for the standard error of the estimate $\hat{SE}[X]$ for each poll given its sample size and value of \hat{X} (`x_hat`). Second, use `mutate` to calculate an estimate for the standard error of the spread for each poll given the value of `se_x_hat`. Then, use `mutate` to calculate upper and lower bounds for 95% confidence intervals of the spread. Last, add a column `hit` that indicates whether the confidence interval for each poll covers the correct spread $d = -0.038$.

How many polls are in `june_polls`?

✓ Answer: 32

Create `june_polls` and count the number of rows with the following code:

```
june_polls <- brexit_polls %>%  
  filter(enddate >= "2016-06-01")  
  nrow(june_polls)
```

What proportion of polls have a confidence interval that covers the value 0?

✓ Answer: 0.625

Answer code

```
june_polls <- june_polls %>%  
  mutate(se_x_hat = sqrt(x_hat*(1-x_hat)/samplesize),  
         se_spread = 2*se_x_hat,  
         ci_lower_spread = spread - qnorm(0.975)*se_spread,  
         ci_upper_spread = spread + qnorm(0.975)*se_spread)  
  mean(june_polls$ci_lower_spread < 0 & june_polls$ci_upper_spread > 0)
```

What proportion of polls predict "Remain" (confidence interval entirely above 0)?

✓ Answer: 0.125

Answer code

```
mean(june_polls$ci_lower_spread > 0)
```

What proportion of polls have a confidence interval covering the true value of d ?

✓ Answer: 0.563

Answer code

```
june_polls <- june_polls %>%  
  mutate(hit = (2*p-1 > ci_lower_spread) & (2*p-1 < ci_upper_spread))  
  mean(june_polls$hit)
```

Submit

You have used 2 of 10 attempts

Question 5: Hit rate by pollster

2.0/2.5 points (graded)

Group and summarize the `june_polls` object by pollster to find the proportion of hits for each pollster and the number of polls per pollster. Use `arrange` to sort by hit rate.

Which of the following are TRUE?

Select ALL correct answers.

- ☒ Unbiased polls and pollsters will theoretically cover the correct value of the spread 50% of the time.
- ☐ Only one pollster had a 100% success rate in generating confidence intervals that covered the correct value of the spread.
- ☐ The pollster with the highest number of polls covered the correct value of the spread in their confidence interval over 60% of the time.
- ☐ All pollsters produced confidence intervals covering the correct spread in at least 1 of their polls.
- ☒ The results are consistent with a large general bias that affects all pollsters. ✓



Answer

Incorrect:

Revisit the definition of 95% confidence intervals.

Submit

You have used 2 of 2 attempts

i Answers are displayed within the problem

Question 6: Boxplot of Brexit polls by poll type

1.5/2.5 points (graded)

Make a boxplot of the spread in `june_polls` by poll type.

Which of the following are TRUE?

Select ALL correct answers.

- ☐ Online polls tend to show support for "Remain" (`spread > 0`).
- ☐ Telephone polls tend to show support "Remain" (`spread > 0`). ✓

☒ Telephone polls tend to show higher support for "Remain" than online polls (higher spread). ✓

☒ Online polls have a larger interquartile range (IQR) for the spread than telephone polls, indicating that they are more variable. ✓

☐ Poll type introduces a bias that affects poll results. ✓



Answer

Incorrect:

Did you make the boxplot correctly?

Did you make the boxplot correctly?

Submit

You have used 2 of 2 attempts

Answers are displayed within the problem

Question 7: Combined spread across poll type

2.0/2.0 points (graded)

Calculate the confidence intervals of the spread combined across all polls in `june_polls`, grouping by poll type. Recall that to determine the standard error of the spread, you will need to double the standard error of the estimate.

Use this code (which determines the total sample size per poll type, gives each spread estimate a weight based on the poll's sample size, and adds an estimate of p from the combined spread) to begin your analysis:

```
combined_by_type <- june_polls %>%  
  group_by(poll_type) %>%  
  summarize(N = sum(samplesize),  
            spread = sum(spread*samplesize)/N,  
            p_hat = (spread + 1)/2)
```

What is the lower bound of the 95% confidence interval for online voters?

-0.0165

✓ Answer: -0.0165

-0.0165

Create `june_polls` and count the number of rows with the following code:

```
combined_by_type <- june_polls %>%
  group_by(poll_type) %>%
  summarize(N = sum(samplesize),
            spread = sum(spread*samplesize)/N,
            p_hat = (spread + 1)/2,
            se_spread = 2*sqrt(p_hat*(1-p_hat)/N),
            spread_lower = spread - qnorm(.975)*se_spread,
            spread_upper = spread + qnorm(.975)*se_spread)
combined_by_type %>%
  filter(poll_type == "Online") %>%
  pull(spread_lower)
```

What is the upper bound of the 95% confidence interval for online voters?

0.00166

✓ Answer: 0.00165

0.00166

Answer code

```
combined_by_type %>%
  filter(poll_type == "Online") %>%
  pull(spread_upper)
```

Submit

You have used 1 of 10 attempts

i Answers are displayed within the problem

Question 8: Interpreting combined spread estimates across poll type

2.5/2.5 points (graded)

Interpret the confidence intervals for the combined spreads for each poll type calculated in the previous problem.

Which of the following are TRUE about the confidence intervals of the combined spreads for different poll types?

Select ALL correct answers.

☒ Neither set of combined polls makes a prediction about the outcome of the Brexit referendum (a prediction is possible if a confidence interval does not cover 0). ✓

☐ The confidence interval for online polls is larger than the confidence interval for telephone polls.

☒ The confidence interval for telephone polls is covers more positive values than the confidence interval for online polls. ✓

☐ The confidence intervals for different poll types do not overlap.

☒ Neither confidence interval covers the true value of $d = -0.038$. ✓



Submit

You have used 1 of 2 attempts

i Answers are displayed within the problem

Continue the comprehensive assessment on the next page.