🏠 Course / Course Wrap-up and Comprehensive Assess... / Comprehensive Assessmen...

< Previous   ✎✓   ✎✓   ✎   ▤   Next >

Brexit poll analysis - Part 3

🔖 Bookmark this page

Final Assessment due Jun 15, 2021 14:57 +03

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

```
# suggested libraries
library(tidyverse)

# 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 9: Chi-squared p-value

2/2 points (graded)

Define `brexit_hit`, with the following code, which computes the confidence intervals for all Brexit polls in 2016 and then calculates whether the confidence interval covers the actual value of the spread $d = -0.038$:

```
    brexit_hit <- brexit_polls %>%
  mutate(p_hat = (spread + 1)/2,
       se_spread = 2*sqrt(p_hat*(1-p_hat)/samplesize),
       spread_lower = spread - qnorm(.975)*se_spread,
       spread_upper = spread + qnorm(.975)*se_spread,
       hit = spread_lower < -0.038 & spread_upper > -0.038) %>%
  select(poll_type, hit)
```

Use `brexit_hit` to make a two-by-two table of poll type and hit status. Then use the `chisq.test()` function to perform a chi-squared test to determine whether the difference in hit rate is significant.

What is the p-value of the chi-squared test comparing the hit rate of online and telephone polls?

0.00101   ✔ **Answer:** 0.00101

0.00101

Answer code

```
    brexit_chisq <- table(brexit_hit$poll_type, brexit_hit$hit)
```

```
chisq.test(brexit_chisq)$p.value
```

Determine which poll type has a higher probability of producing a confidence interval that covers the correct value of the spread. Also determine whether this difference is statistically significant at a p-value cutoff of 0.05. Which of the following is true?

- ● Online polls are more likely to cover the correct value of the spread and this difference is statistically significant.

- ○ Online polls are more likely to cover the correct value of the spread, but this difference is not statistically significant.

- ○ Telephone polls are more likely to cover the correct value of the spread and this difference is statistically significant.

- ○ Telephone polls are more likely to cover the correct value of the spread, but this difference is not statistically significant.

✔

Answer code

```
          # online > telephone
hit_rate <- brexit_hit %>%
    group_by(poll_type) %>%
    summarize(avg = mean(hit))
hit_rate$avg[hit_rate$poll_type == "Online"] > hit_rate$avg[hit_rate$poll_type == "Telephone"]

# statistically significant
chisq.test(brexit_chisq)$p.value < 0.05
```

Submit   You have used 1 of 10 attempts

ℹ Answers are displayed within the problem

Question 10: Odds ratio of online and telephone poll hit rate

3/3 points (graded)
Use the two-by-two table constructed in the previous exercise to calculate the odds ratio between the hit rate of online and telephone polls to determine the magnitude of the difference in performance between the poll types.

Calculate the odds that an online poll generates a confidence interval that covers the actual value of the spread.

1.30   ✔ **Answer:** 1.30

1.30

Answer code

```
          # from previous question
brexit_chisq <- table(brexit_hit$poll_type, brexit_hit$hit)

# convert to data frame
chisq_df <- as.data.frame(brexit_chisq)

online_true <- chisq_df$Freq[chisq_df$Var1 == "Online" & chisq_df$Var2 == "TRUE"]
```

```
online_false <- chisq_df$Freq[chisq_df$Var1 == "Online" & chisq_df$Var2 == "FALSE"]

online_odds <- online_true/online_false
online_odds
```

Calculate the odds that a telephone poll generates a confidence interval that covers the actual value of the spread.

| 0.312 | ✔ **Answer:** 0.312 |

0.312

Answer code

```
        phone_true <- chisq_df$Freq[chisq_df$Var1 == "Telephone" & chisq_df$Var2 == "TRUE"]
  phone_false <- chisq_df$Freq[chisq_df$Var1 == "Telephone" & chisq_df$Var2 == "FALSE"]

  phone_odds <- phone_true/phone_false
  phone_odds
```

Calculate the odds ratio to determine how many times larger the odds are for online polls to hit versus telephone polls.

| 4.15 | ✔ **Answer:** 4.15 |

4.15

Answer code

```
online_odds/phone_odds
```

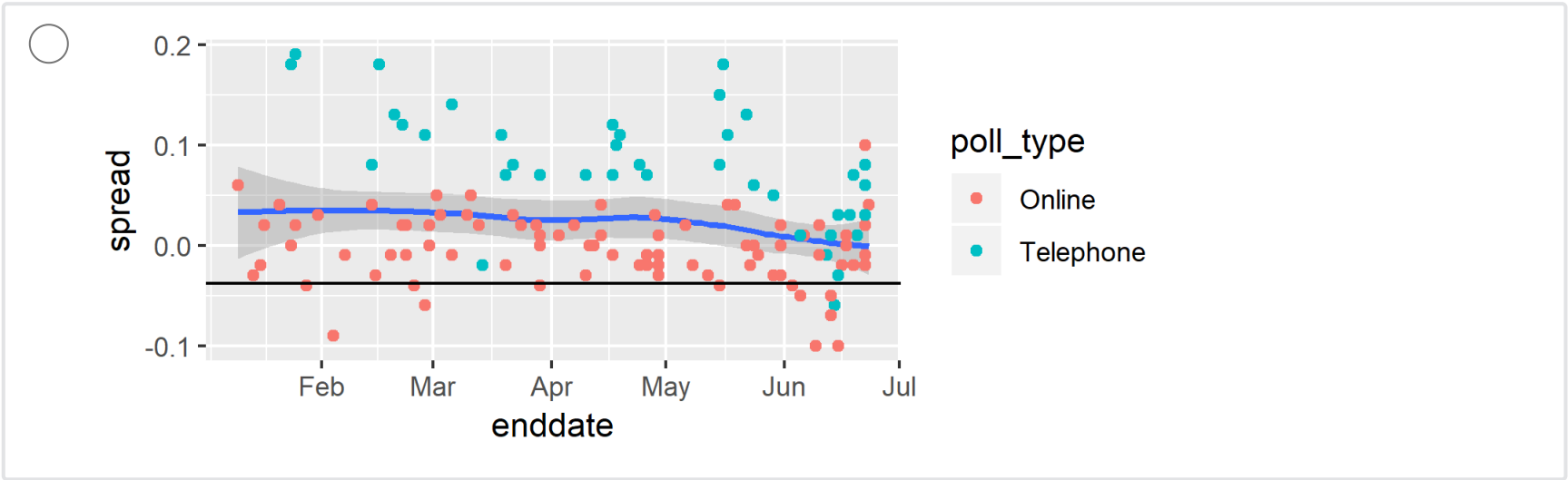**Submit** You have used 1 of 10 attempts

ⓘ Answers are displayed within the problem

Question 11: Plotting spread over time

1/1 point (graded)
Use `brexit_polls` to make a plot of the spread (`spread`) over time (`enddate`) colored by poll type (`poll_type`). Use `geom_smooth()` with `method = "loess"` to plot smooth curves with a span of 0.4. Include the individual data points colored by poll type. Add a horizontal line indicating the final value of $d = -.038$.

Which of the following plots is correct?

✔
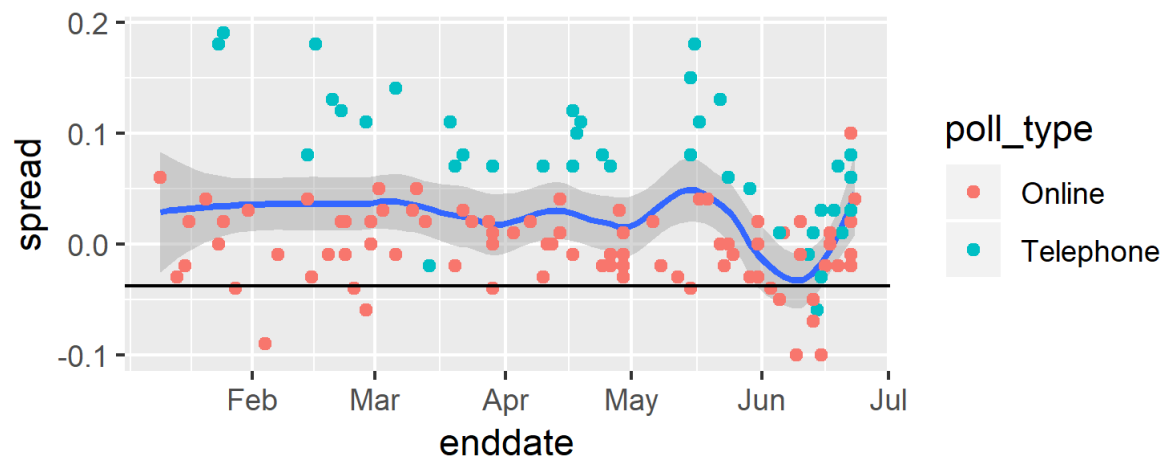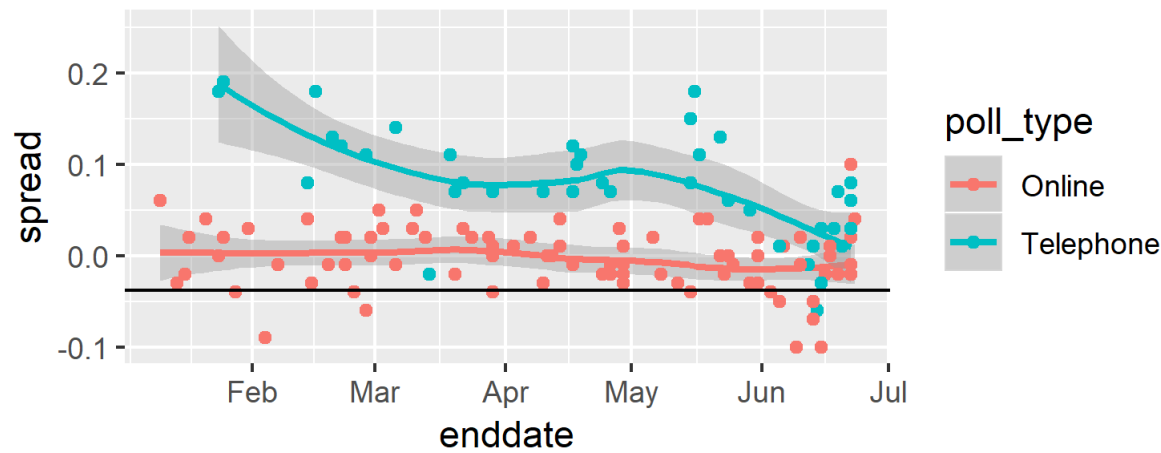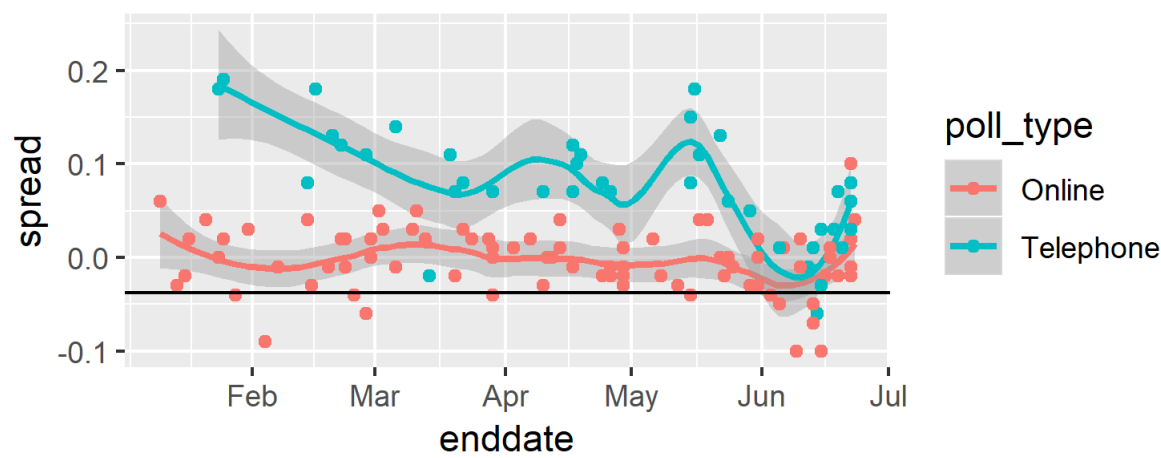
Explanation

The plot can be made using the following code:

```
brexit_polls %>%
    ggplot(aes(enddate, spread, color = poll_type)) +
    geom_smooth(method = "loess", span = 0.4) +
    geom_point() +
    geom_hline(aes(yintercept = -.038))
```

Submit   You have used 1 of 2 attempts

ℹ Answers are displayed within the problem

Question 12: Plotting raw percentages over time

2.5/2.5 points (graded)

Use the following code to create the object `brexit_long`, which has a column `vote` containing the three possible votes on a Brexit poll ("remain", "leave", "undecided") and a column `proportion` containing the raw proportion choosing that vote option on the given poll:

```
brexit_long <- brexit_polls %>%
    gather(vote, proportion, "remain":"undecided") %>%
    mutate(vote = factor(vote))
```

Make a graph of proportion over time colored by vote. Add a smooth trendline with `geom_smooth()` and `method = "loess"` with a span of 0.3.

Which of the following are TRUE?
Select ALL correct answers.

- ☑ The percentage of undecided voters declines over time but is still around 10% throughout June.

- ☑ Over most of the date range, the confidence bands for "Leave" and "Remain" overlap.

- ☑ Over most of the date range, the confidence bands for "Leave" and "Remain" are below 50%.

- ☑ In the first half of June, "Leave" was polling higher than "Remain", although this difference was within the confidence intervals.

- ☐ At the time of the election in late June, the percentage voting "Leave" is trending upwards.

✔

Submit You have used 1 of 2 attempts

ⓘ Answers are displayed within the problem

‹ Previous   Next ›

edX

edX

About
Affiliates
edX for Business
Open edX
Careers
News

Legal

Terms of Service & Honor Code
Privacy Policy
Accessibility Policy
Trademark Policy
Sitemap

Connect

Blog

Contact Us

Help Center

Media Kit

Donate