Lab 08: Proportions and Inference

Tests of changes in sex ratios based on a single sample

There is a long literature studying changes in sex-ratios of births due to stressful events, such as 9/11. In today's lab, we consider a relatively small study that recorded biomarkers of stress on pregnancy. In the group of subjects that had the highest markers of stress (based on cortisol), there were 14 births to males out of a total of 38.

In this lab, we will compare the four methods we learned to calculate CIs for proportions. Recall that two of these methods involved hand calculations (though we can treat R as if it were a calculator) and two of the methods used built-in R functions.

1. Use the Normal approximation to construct a 95% confidence interval in this high stress group. We also called this specific method of constructing the CI the "large sample method"

```
# your code here
```

Write the CI here.

2. Create the 95% CI again, this time using the R function that implements the Wilson Score method with a continuity correction.

```
# your code here
```

Write the CI here.

3. Create the 95% CI again, this time using the Plus 4 method.

```
# your code here
```

Write the CI here.

4. Create the 95% CI again, this time using the R function that implements the Clopper Pearson (Exact) method.

```
# your code here
```

Write the CI here.

5. Summarize the four methods' estimates in the following table. Do they include the null hypothesized value for the sex ratio?

Method	95% Confidence Interval
Large sample	AA.A% to AA.A%
Wilson Score*	AA.A% to AA.A%
Plus four	AA.A% to AA.A%
Exact	AA.A% to AA.A%

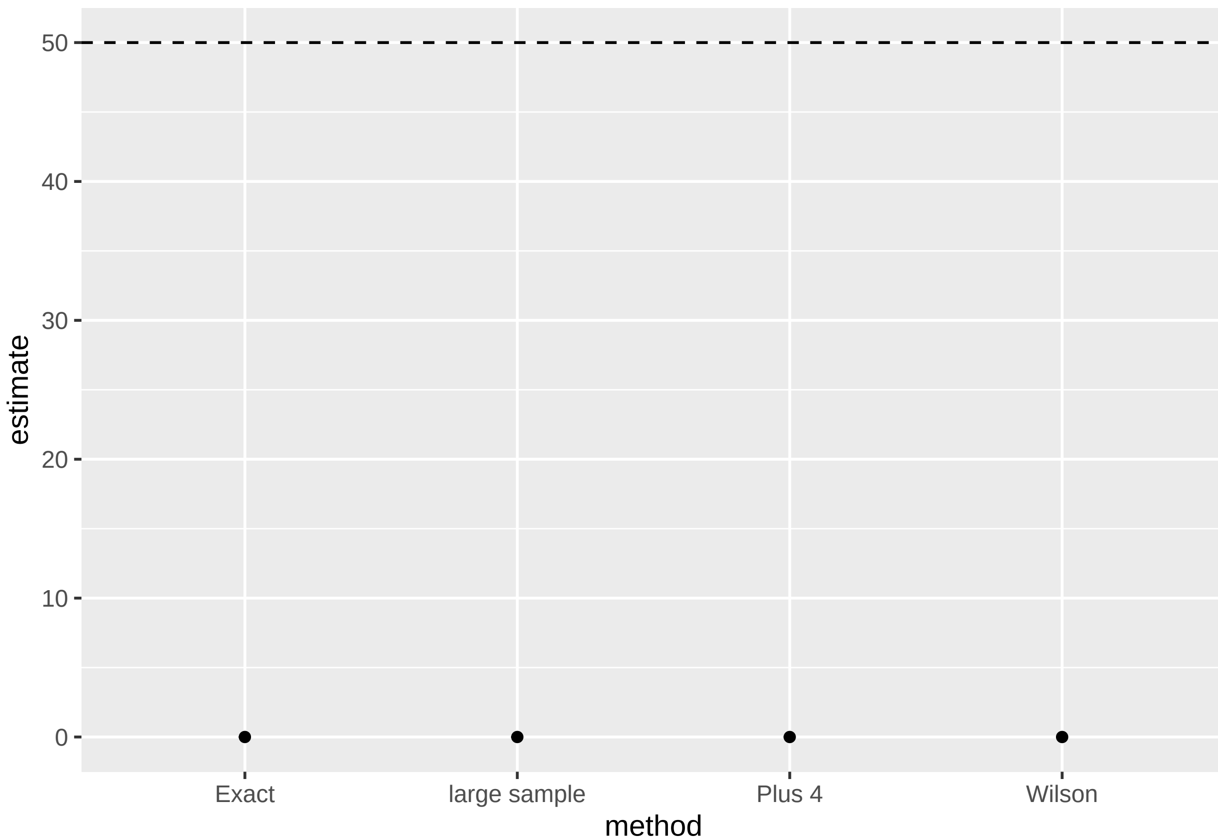
- with continuity correction.

6. Here is a code template to help you to graphically present these estimates. Graphical presentations of estimates and their CIs is very useful for assessing whether the CIs overlap the null hypothesized value and tends to be better than presenting tables of estimates to readers of your research.

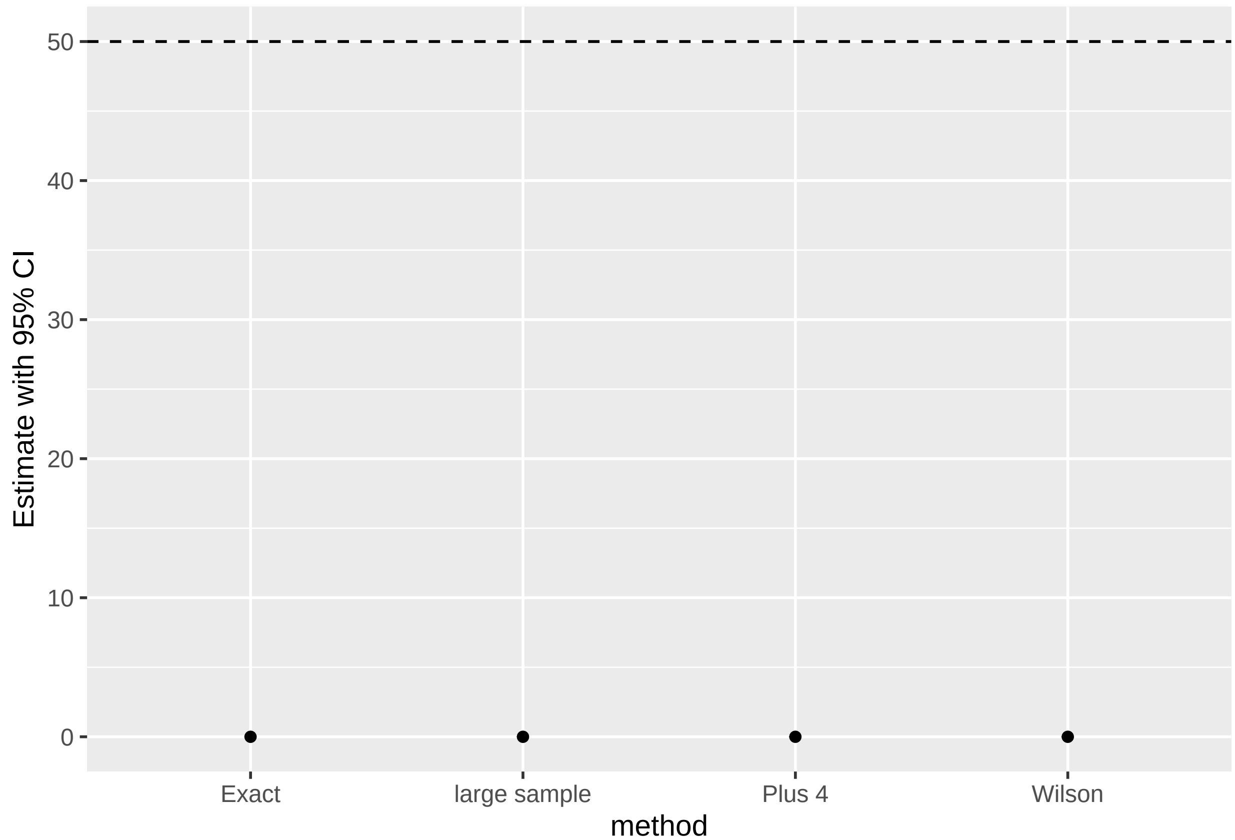
```
# First make a tibble (an easy way to make a data frame) with the data about
# each confidence interval. To do this, replace each instance of 0.00 with the
# estimate from your calculations above.
```

```r
library(ggplot2)
library(tibble)
sex_CIs <- tibble(method   = c("large sample", "Exact", "Wilson", "Plus 4"),
                  lower_CI = c(0.0          , 0.0    ,  0.0    ,  0.0),
                  upper_CI = c(0.0          , 0.0    ,  0.0    ,  0.0),
                  estimate = c(0.0          , 0.0    ,  0.0    ,  0.0)
                  )

# Build the ggplot incrementally, to understand how it works.
# Step 1: (qu: why do we put a horizontal line at 50?)
ggplot(data = sex_CIs, aes(x = method, y = estimate)) +
  geom_point() +
  geom_hline(aes(yintercept = 50), lty = 2)
```




```r
# Step 2:
ggplot(data = sex_CIs, aes(x = method, y = estimate)) +
  geom_point() +
  geom_hline(aes(yintercept = 50), lty = 2) +
  geom_segment(aes(x = method, xend = method, y = lower_CI, yend = upper_CI)) +
  labs(y = "Estimate with 95% CI")
```



What does `geom_segment()` do? In particular, what do `x`, `xend`, `y` and `yend` specify in this case?

Write your answer here.

7. Based on this plot, what can you say about the confidence intervals for the sex ratio in the high stress group?

Write your answer here.

---

8. If you have time, repeat the above analysis for the group with low stress. There were 25 births to this group, of which 17 of them were to males.

```
# your code here
```

7. If you recreated the graph for the low stress group, what can you say about the confidence intervals for the sex ratio in this group?

Write your answer here.

Submission

Please submit this lab file *directly* to Gradescope under the "Lab 8" Assignment. You can do this by knitting your file and downloading the PDF to your computer. Then navigate to Gradescope.com to submit your assignment. Here is a tutorial if you need help: https://www.gradescope.com/get_started. Scroll down on that page to "For students:submitting homework".