

Stat 477 - HW 3

Jay Maxwell

2/8/2022

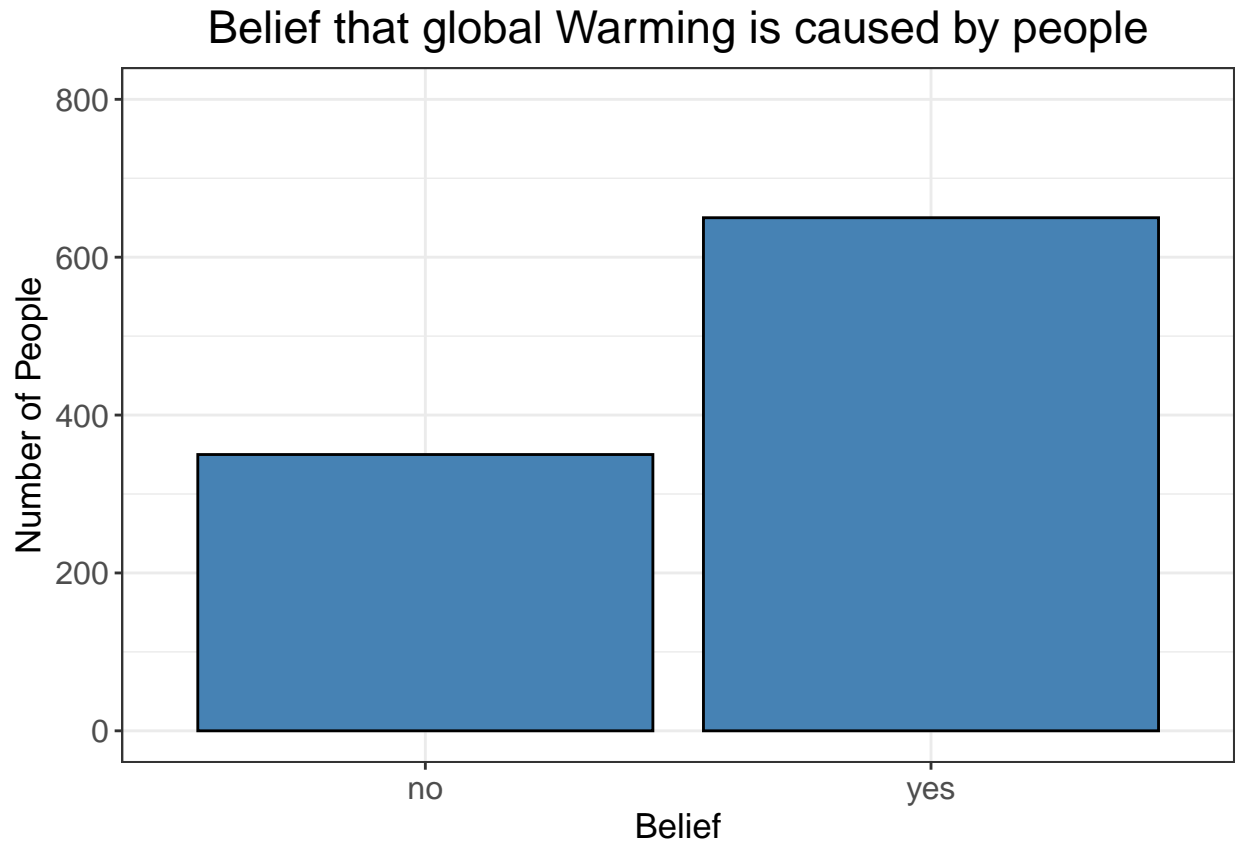
1. According to a Gallup poll, of 1,000 randomly selected adults aged 18 or older in the United States, 65% believe that global warming is more a result of human actions than natural causes.

(a) Use R to obtain a summary table and bar graph of the sample data.

```
warming <- as.factor(c(rep("yes", 650), rep("no", 350)))
df <- data.frame(warming)
tab1(df$warming, graph = FALSE, cum.percent = FALSE)
```

```
## df$warming :
##           Frequency Percent
## no             350       35
## yes            650       65
## Total          1000      100
```

```
df %>%
  ggplot(aes(x = warming)) + geom_bar(fill = "steelblue", colour = "black") + ylim(0,
800) + labs(x = "Belief", y = "Number of People", title = "Belief that global Warming is caused by p
theme_bw() + theme(axis.title.y = element_text(size = rel(1.2)), axis.title.x = element_text(size =
axis.text.x = element_text(size = rel(1.4)), axis.text.y = element_text(size = rel(1.4)),
plot.title = element_text(hjust = 0.5, size = rel(1.5)))
```



(b) Describe the population proportion of interest p in words.

In this example, the population proportion (p) is the fraction of the entire US population that believes global warming is caused more by people than natural causes.

(c) Use R to calculate a 95% confidence interval for the population proportion p using the normal approximation method.

```
prop.ci(650, 1000, type = "normal", 0.95)
```

```
## 0.6204377 0.6795623
```

(d) Give the interpretation of the 95% confidence interval you calculated in part (c) in the context of the problem.

We are 95% confident that the true population proportion of people in the USE who believe global warming is caused more by people than nature is between 62.04% to approximately 67.95%.

(e) Use R to calculate a 95% confidence interval for the population proportion p using the Wilson's score method. Compare the center and width of this interval to the one you calculated in part (c).

```
prop.ci(650, 1000, type = "score", 0.95)
```

```
## 0.6199147 0.6789373
```

```
(0.6795623 + 0.6204377)/2
```

```
## [1] 0.65
```

```
(0.6789373 + 0.6199147)/2
```

```
## [1] 0.649426
```

The results of the score method have a slightly smaller range than the results of the normal approximation method. I would call the centers for each calculation approximately equal at .65.

(f) Gallup is planning to conduct another poll on global warming. They would like to have a 95% confidence interval with a margin of error of no more than 2.5%. What sample size do they need to obtain this margin of error? Make sure to specify how you are calculating the sample size

```
cat("Sample size needed:", nprop.ci(0.65, 0.025, 0.95))
```

```
## Sample size needed: 1399
```

To calculate the sample size, we solve for $n \geq (\frac{z_{1-\frac{\alpha}{2}}}{M})^2 \hat{p} * (1 - \hat{p})$. Because this is a follow up survey, the reasearches might use their prior findings to estimate $\hat{p} = .65$.

2. Unlike confidence intervals for other parameters... What do you notice about the coverage rates of the two methods? How do these results depend on the values of n and p?

p	Normal			Wilson's		
	n=25	n=250	n=1000	n=25	n=250	n=1000
0.5	95.651	95.145	94.656	95.766	95.008	94.729
0.75	89.468	93.999	94.66	93.901	94.365	94.758
0.9	91.851	93.104	95.17	96.33	95.54	94.893

When we compare the coverages calcuated above it appears that for most of our simulations the Wilson's score method results in coverage closer to the 95% goal more sore than the Normal approximation method. I have a hard time determining a pattern of relation between n and p for both methods. There is a lot of variation in coverage (between 89% and 95%) for the normal method for all the combinations of n and p. While the Wilson's score method seems to stay consistently close to our ideal coverage of 95% no matter what the combinations of n and p are changed to.

```
# coverage.ci(25,.5,'normal',0.95) coverage.ci(250,.5,'normal',0.95)
# coverage.ci(1000,.5,'normal',0.95) coverage.ci(25,.5,'score',0.95)
# coverage.ci(250,.5,'score',0.95) coverage.ci(1000,.5,'score',0.95)
# coverage.ci(25,.75,'normal',0.95) coverage.ci(250,.75,'normal',0.95)
# coverage.ci(1000,.75,'normal',0.95) coverage.ci(25,.75,'score',0.95)
# coverage.ci(250,.75,'score',0.95) coverage.ci(1000,.75,'score',0.95)
# coverage.ci(25,.9,'normal',0.95) coverage.ci(250,.9,'normal',0.95)
# coverage.ci(1000,.9,'normal',0.95) coverage.ci(25,.9,'score',0.95)
# coverage.ci(250,.9,'score',0.95) coverage.ci(1000,.9,'score',0.95)
```