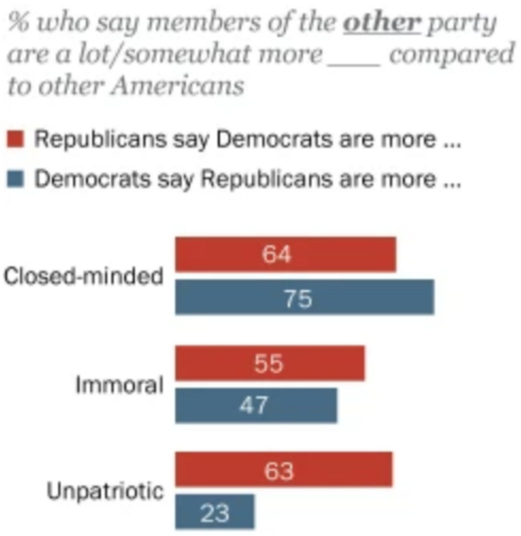
Confidence Intervals for p

U.S. POLITICS | OCTOBER 10, 2019

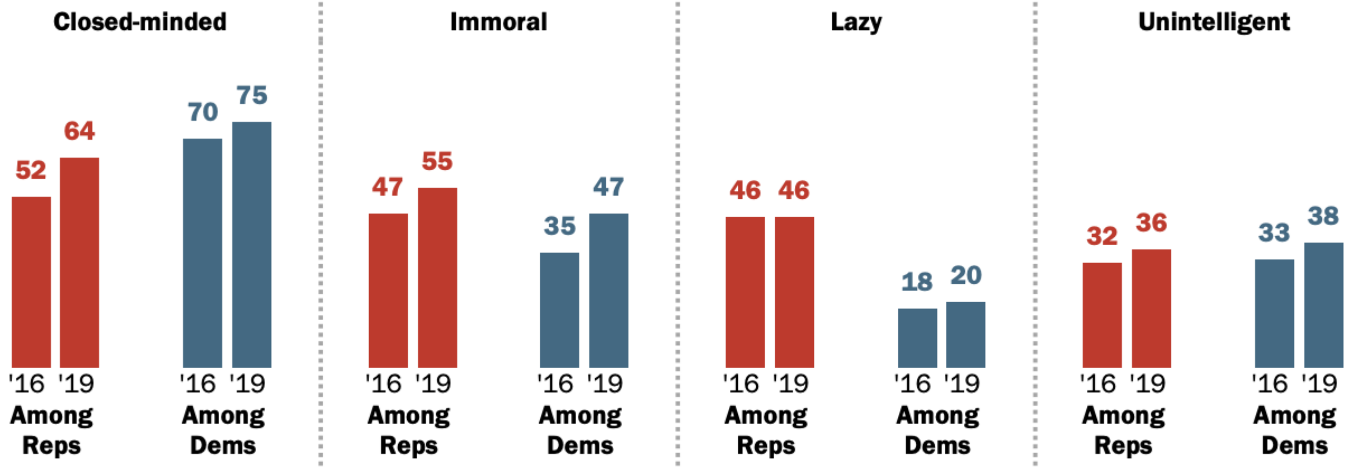
Partisan Antipathy: More Intense, More Personal

The share of Republicans who give Democrats a "cold" rating on a 0-100 thermometer has risen 14 percentage points since 2016. Similarly, 57% of Democrats give Republicans a very cold rating, up from 2016.



Increasing shares of partisans see members of the other party as 'closed-minded' and 'immoral'

*% who say members of the **other** party are a lot/somewhat more ___ compared to other Americans*



Note: Partisans do not include leaners.
Source: Survey of U.S. adults conducted Sept. 3-15, 2019.

PEW RESEARCH CENTER

Data in this report are drawn from the panel wave conducted September 3 to September 15, 2019. A total of 9,895 panelists responded out of 10,478 who were sampled, for a response rate of 94%. This does not include four panelists who were removed from the data due to extremely high rates of refusal or straightlining. The cumulative response rate accounting for nonresponse to the recruitment surveys and attrition is 5.9%. The break-off rate among panelists who logged onto the survey and completed at least one item is 0.8%. The margin of sampling error for the full sample of 9,895 respondents is plus or minus 1.5 percentage points.

American Trends Panel recruitment surveys

Recruitment Dates	Mode	Invited	Joined	Active panelists remaining
Jan. 23 to March 16, 2014	Landline/ cell RDD	9,809	5,338	2,272
Aug. 27 to Oct. 4, 2015	Landline/ cell RDD	6,004	2,976	1,298
April 25 to June 4, 2017	Landline/ cell RDD	3,905	1,628	652
Aug. 8 to Oct. 31, 2018	ABS/web	9,396	8,778	6,256
	Total	**29,114**	**18,720**	**10,478**

Note: Approximately once per year, panelists who have not participated in multiple consecutive waves or who did not complete an annual profiling survey are removed from the panel. Panelists also become inactive if they ask to be removed from the panel.

PEW RESEARCH CENTER

Boardwork

Pew Data

```r
pew <- data.frame(party = "Republican",
                  closed = rep(c("yes", "no"),
                               c(3166, 1781)
head(pew)
```

```
##        party closed
## 1 Republican    yes
## 2 Republican    yes
## 3 Republican    yes
## 4 Republican    yes
## 5 Republican    yes
## 6 Republican    yes
```

Bootstrapping p-hat I

```
pew %>%
  specify(response = closed, success = "yes") %>%
  generate(reps = 100, type = "bootstrap")
```

```
## Response: closed (factor)
## # A tibble: 494,700 x 2
## # Groups:   replicate [100]
##    replicate closed
##        <int> <fct>
##  1         1 no
##  2         1 yes
##  3         1 no
##  4         1 yes
##  5         1 yes
##  6         1 yes
##  7         1 no
##  8         1 no
##  9         1 no
## 10         1 no
```

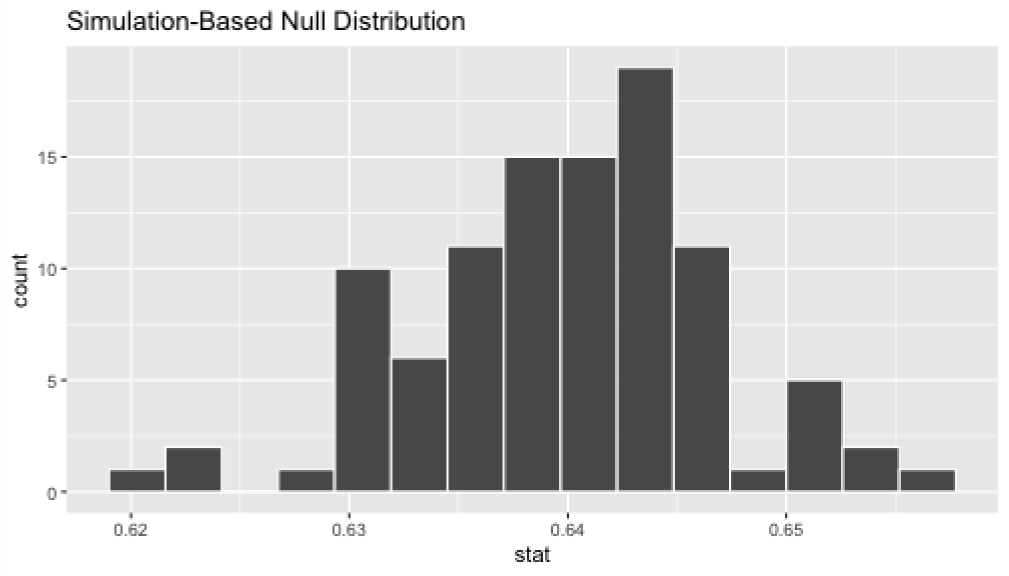
Bootstrapping p-hat II

```
pew %>%
  specify(response = closed, success = "yes") %>%
  generate(reps = 100, type = "bootstrap") %>%
  calculate(stat = "prop")
```

```
## # A tibble: 100 x 2
##    replicate  stat
##        <int> <dbl>
##  1         1 0.643
##  2         2 0.629
##  3         3 0.635
##  4         4 0.650
##  5         5 0.642
##  6         6 0.639
##  7         7 0.634
##  8         8 0.627
##  9         9 0.628
## 10        10 0.646
## # … with 90 more rows
```

Bootstrapping p-hat III

```
pew %>%
  specify(response = closed, success = "yes") %>%
  generate(reps = 100, type = "bootstrap") %>%
  calculate(stat = "prop") %>%
  visualize()
```



Bootstrapping p-hat III

```r
boot_se <- pew %>%
  specify(response = closed, success = "yes") %>%
  generate(reps = 100, type = "bootstrap") %>%
  calculate(stat = "prop") %>%
  summarize(se = sd(stat)) %>%
  pull()
```

```r
boot_se
```

```
## [1] 0.006723887
```

Constructing a CI

$$\hat{p} \pm 1.96 \times SE$$

```
p_hat <- pew %>%
  specify(response = closed, success = "yes") %>%
  calculate(stat = "prop") %>%
  pull()
```

```
p_hat
```

```
## [1] 0.6399838
```

```
c(p_hat - 1.96 * boot_se, p_hat + 1.96 * boot_se)
```

```
## [1] 0.6268050 0.6531626
```

Normal Approximation Method

The sampling distribution of the point estimate $\hat{p}$ can be approximated with:

$\hat{p} \sim N(\mu = p, \sigma = \sqrt{\frac{p(1-p)}{n}})$ if

1. Observations are independent
2. The number of observed successes and failures is more than 10.

$$n\hat{p} \geq 10$$
$$n(1-\hat{p}) \geq 10$$

Question

We are given that $n = 4947, \hat{p} = 0.64$. Which of the below is the correct calculation of the 95% confidence interval?

1. $0.64 \pm 1.96\sqrt{(.64 \times 0.36)/4947}$
2. $0.64 \pm 1.65\sqrt{(.64 \times 0.36)/4947}$
3. $0.64 \pm 1.96((.64 \times 0.36)/\sqrt{4947})$
4. $64 \pm 1.96\sqrt{(64 \times 36)/4947}$

Constructing a CI

```
nrow(pew) * p_hat
```

```
## [1] 3166
```

```
nrow(pew) * (1 - p_hat)
```

```
## [1] 1781
```

```
approx_SE <- sqrt((p_hat * (1 - p_hat)) /
                    nrow(pew))
```

```
approx_SE
```

```
## [1] 0.006824558
```

Constructing a CI, cont.

```
z_star <- qnorm(.025) # for a 95% CI
```

```
z_star
```

```
## [1] -1.959964
```

```
MoE <- z_star * approx_SE
```

```
c(p_hat - MoE, p_hat + MoE)
```

```
## [1] 0.6533597 0.6266079
```

Choosing between methods

Bootstrap SE

- Must have sufficiently large sample (n > 30)
- Can be calculated for any statistic

Normal approximation

- Must satisfy assumptions (independence, sample size)
- Most commonly used (still)