Winter TEE

Ryan Cheng

2024-02-24

```
nfl.fanbases = read.csv("final_NFL_fanbases.csv")
head(nfl.fanbases)
```

```
##   X           TeamFan      Response  Race Political
## 1 1 Arizona Cardinals DemocratAsian Asian  Democrat
## 2 2 Arizona Cardinals DemocratAsian Asian  Democrat
## 3 3 Arizona Cardinals DemocratBlack Black  Democrat
## 4 4 Arizona Cardinals DemocratBlack Black  Democrat
## 5 5 Arizona Cardinals DemocratBlack Black  Democrat
## 6 6 Arizona Cardinals DemocratBlack Black  Democrat
```

What percentage of NFL fans do you think are Democrats?

Assumptions and conditions: We're assuming that this sample of NFL fans is either random or representative of the entire NFL fan base since the description of the data set does not say so.

```
nrow(nfl.fanbases)
```

```
## [1] 6364
```

n = 6364 n > 30: Yes p or q would need to be less than 0.00157133878 for np or nq to be less than 10, so we'll proceed with significant confidence. n < 10% of all NFL fans: Yes

Hypotheses: p = true proportion of NFL fans that are Democrats H0: p = 0.4 Ha: p != 0.4 alpha = 0.05

```
x = sum(nfl.fanbases$Political == "Democrat")
n = nrow(nfl.fanbases)
p = 0.4

prop.test(x, n, p, alternative = "two.sided")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  x out of n, null probability p
## X-squared = 150.28, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is not equal to 0.4
## 95 percent confidence interval:
##  0.3131655 0.3363247
## sample estimates:
##         p
## 0.3246386
```

p-value < 2.2e-16

Because p-value < 2.2e-16 is less than alpha = 0.05, we reject the null hypothesis and accept the alternative hypothesis, which is that the true proportion of NFL fans that are Democrats is not 0.4.

Now let's create a 95 percent confidence interval for the true proportion of NFL fans that are Democrats.

```
prop.test(x, n, conf.level = 0.95, correct = FALSE)
```

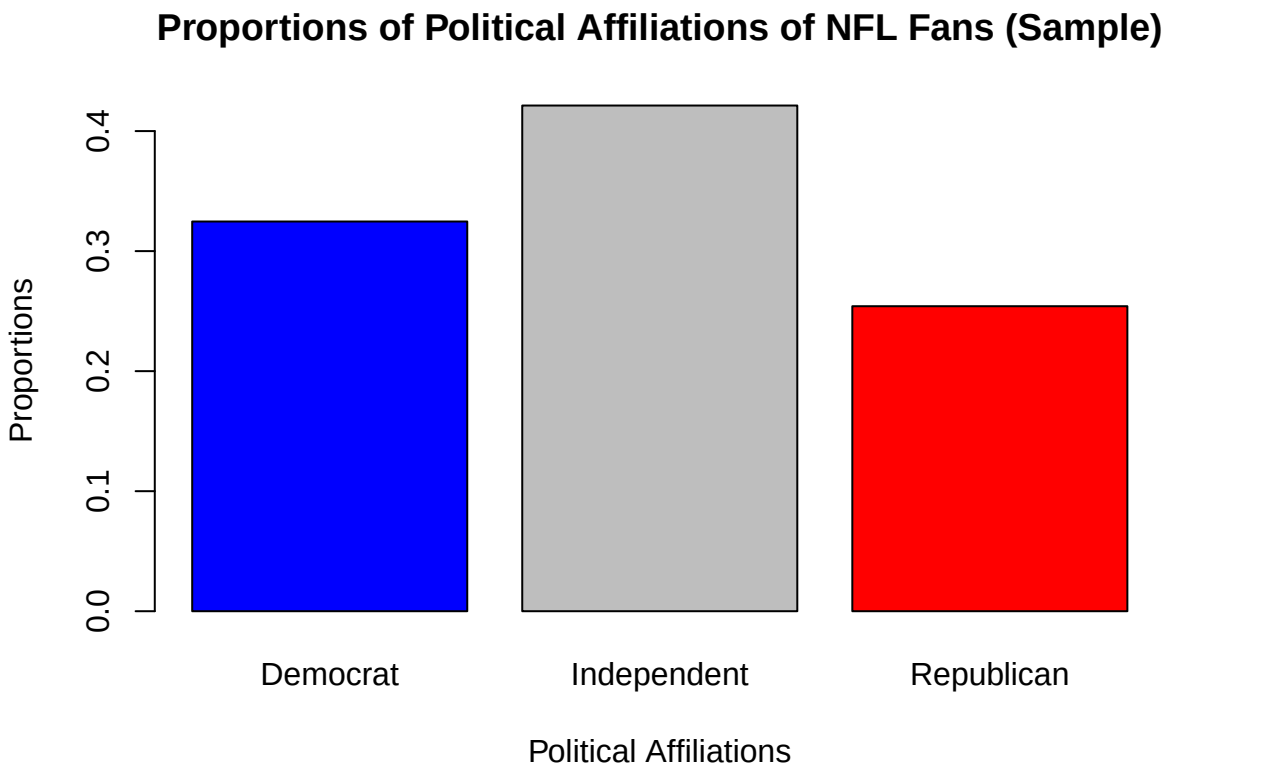
```
##
##  1-sample proportions test without continuity correction
##
## data:  x out of n, null probability 0.5
## X-squared = 782.81, df = 1, p-value < 2.2e-16
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.3132433 0.3362455
## sample estimates:
##         p
## 0.3246386
```

95 percent confidence interval: 31.3% to 33.6%

Our CI supports the results of our hypothesis test because the 40% from our null hypothesis is not in it.

Let's create a barplot to help us better understand the political affiliations of NFL fans.

```
barplot(prop.table(table(nfl.fanbases$Political)),
        main = "Proportions of Political Affiliations of NFL Fans (Sample)",
        xlab = "Political Affiliations",
        ylab = "Proportions",
        col = c("blue", "grey", "red"))
```



It's quite surprising to see how high the proportion of NFL fans that are Independent is. I bet this high proportion can be attributed to the fact that this survey/poll was conducted in 2017, right after the 2016 Presidential election, the nature of which caused many voters not to outright affiliate themselves with a political party.

```r
sanfran.fanbase = nfl.fanbases[nfl.fanbases$TeamFan == "San Francisco 49ers",]
```

Is the 49ers' fan base more liberal than the overall NFL fan base? Is it more liberal than the city itself?

We'll be conservative and say that the true proportion of NFL fans that are Democrats is the higher end of its 95 percent confidence interval, 33.6%.

Hypotheses: p = true proportion of 49ers fans that are Democrats H0: p = 0.336 Ha: p > 0.336 alpha = 0.05

```r
x = sum(sanfran.fanbase$Political == "Democrat")
n = nrow(sanfran.fanbase)
p = 0.336

prop.test(x, n, p, alternative = "greater")
```

```
##
##  1-sample proportions test with continuity correction
##
## data:  x out of n, null probability p
## X-squared = 4.9728, df = 1, p-value = 0.01287
## alternative hypothesis: true p is greater than 0.336
## 95 percent confidence interval:
##  0.3527427 1.0000000
## sample estimates:
##         p
## 0.4039216
```

p-value = 0.01287

p-value of 0.01287 is less than alpha of 0.05, so we reject the null hypothesis and accept the alternative hypothesis. The true proportion of 49ers fans that are Democrats is greater than the true proportion of NFL fans that are Democrats, which is 33.6%.

Now let's create a 95 percent confidence interval for the true proportion of 49ers fans that are Democrats.

```r
prop.test(x, n, conf.level = 0.95, correct = FALSE)
```

```
##
##  1-sample proportions test without continuity correction
##
## data:  x out of n, null probability 0.5
## X-squared = 9.4157, df = 1, p-value = 0.002151
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.3455538 0.4651411
## sample estimates:
##         p
## 0.4039216
```

95 percent confidence interval: 34.6% to 46.5%

Our CI supports the results of our hypothesis test because the 33.6% proposed in our null hypothesis is not present in it and is lower than the smaller end.

In 2017, 57.72% of registered voters in the San Francisco area registered as Democratic (source: https://elections.cdn.sos.ca.gov/ror/ror-pages/ror-odd-year-2017/county.pdf). Based on the confidence interval we just created, we can confidently say that the 49ers' fan base is less liberal than the city the team hails from.

Just for fun, let's see how much more liberal the fan base of the 49ers is than that of the Tampa Bay Buccaneers. Let's create a confidence interval for the true proportion of Buccaneers fans that are Democrats.

```r
buccaneers.fanbase = nfl.fanbases[nfl.fanbases$TeamFan == "Tampa Bay Buccaneers",]
```

```r
x = sum(buccaneers.fanbase$Political == "Democrat")
n = nrow(buccaneers.fanbase)

prop.test(x, n, conf.level = 0.95, correct = FALSE)
```

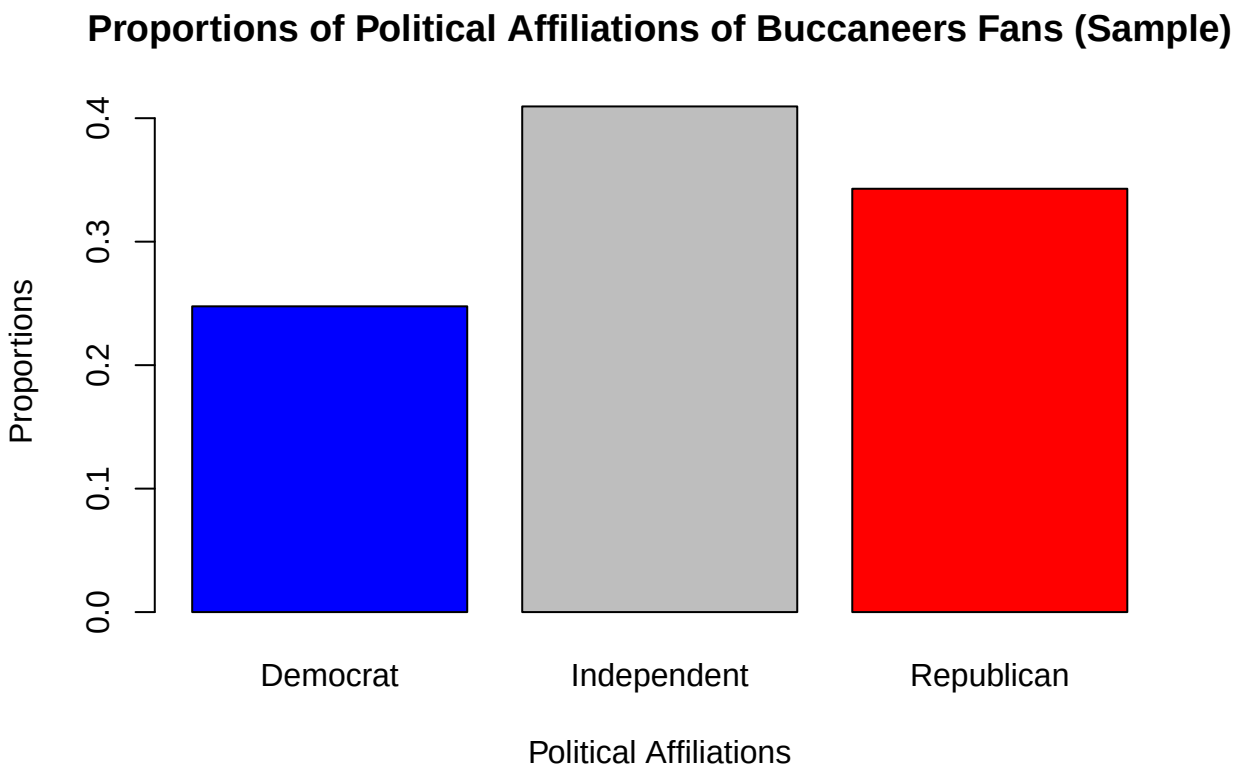
```
##
##  1-sample proportions test without continuity correction
##
## data:  x out of n, null probability 0.5
## X-squared = 26.752, df = 1, p-value = 2.313e-07
## alternative hypothesis: true p is not equal to 0.5
## 95 percent confidence interval:
##  0.1749498 0.3381034
## sample estimates:
##        p
## 0.247619
```

95 percent confidence interval: 17.5% to 33.8%

Both the CI for the 49ers and the Buccaneers are quite big so it's hard to say exactly how much more liberal the 49ers fan base is. However, we can say quite confidently that the 49ers fan base is indeed more liberal than the Buccaneers'.

Code for "Proportions of Political Affiliations of Buccaneers Fans (Sample)" graph:

```r
barplot(prop.table(table(buccaneers.fanbase$Political)),
        main = "Proportions of Political Affiliations of Buccaneers Fans (Sample)",
        xlab = "Political Affiliations",
        ylab = "Proportions",
        col = c("blue", "grey", "red"))
```



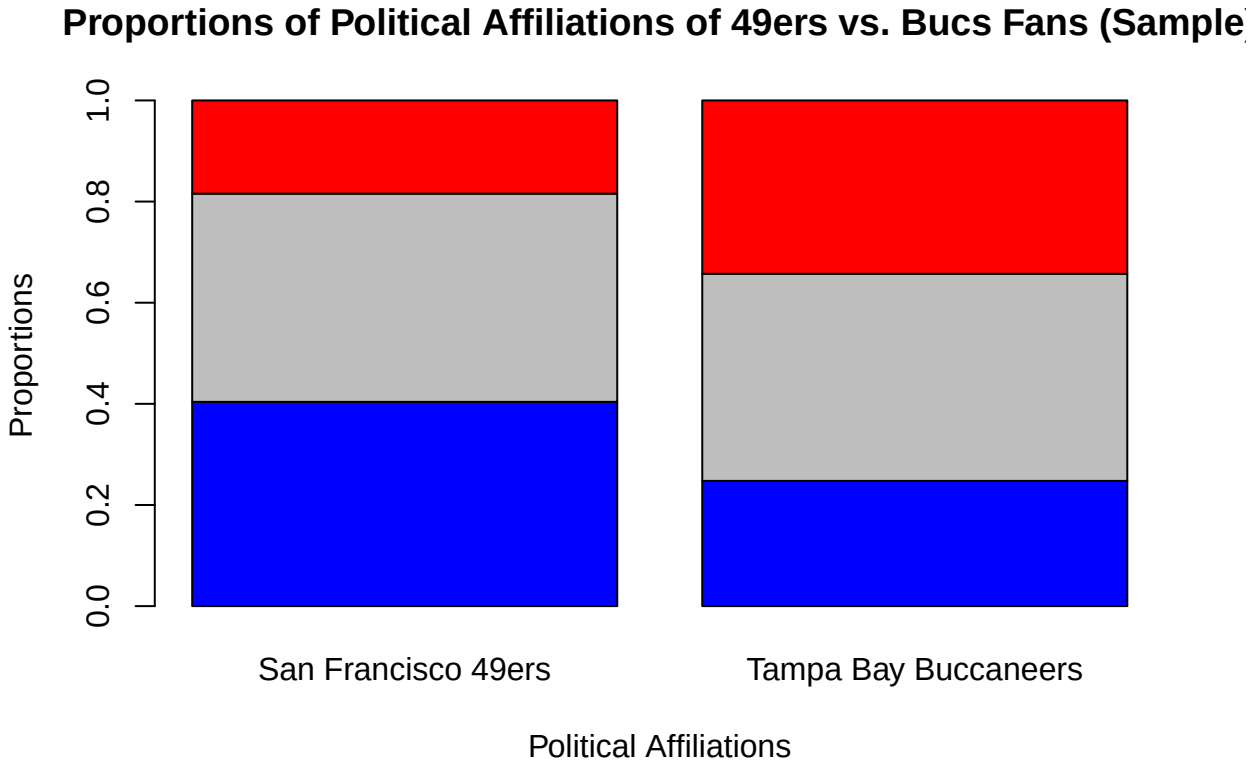
Code for subsetting the entire NFL fanbase into the 49ers and Buccaneers fanbases:

```
sanfran.and.buccaneers.fanbases = nfl.fanbases[nfl.fanbases$TeamFan == "San Francisco 49ers" | nfl.fanba
head(sanfran.and.buccaneers.fanbases)
```

```
##         X             TeamFan      Response  Race Political
## 5431 5431 San Francisco 49ers DemocratAsian Asian  Democrat
## 5432 5432 San Francisco 49ers DemocratAsian Asian  Democrat
## 5433 5433 San Francisco 49ers DemocratAsian Asian  Democrat
## 5434 5434 San Francisco 49ers DemocratAsian Asian  Democrat
## 5435 5435 San Francisco 49ers DemocratAsian Asian  Democrat
## 5436 5436 San Francisco 49ers DemocratAsian Asian  Democrat
```

Code for "Proportions of Political Affiliations of 49ers vs. Bucs Fans (Sample)" graph:

```
barplot(prop.table(table(sanfran.and.buccaneers.fanbases$Political, sanfran.and.buccaneers.fanbases$Tea
        main = "Proportions of Political Affiliations of 49ers vs. Bucs Fans (Sample)",
        xlab = "Political Affiliations",
        ylab = "Proportions",
        col = c("blue", "grey", "red"))
```


Proportions of Political Affiliations of 49ers vs. Bucs Fans (Sample
Proportions
0.0
0.2
0.4
0.6
0.8
1.0
San Francisco 49ers
Tampa Bay Buccaneers
Political Affiliations