Analysis of Presidential Approval Factors

Written by:

Shirley Chew

STA 104

University of California, Davis

Questions of Interest

The questions of interested in my study are:

1. Is there a significant difference in GDP (per billion of dollars) between democratic and republican presidents?
2. Which two quarters have the greatest difference in approval ratings?
3. Is an increase in GDP positively associated with higher presidential approval ratings?

Statistical Methods & Assumptions

Question 1:

In question one, I used a two sample test for independence to compare the GDP of the democratic party (first sample) against the republican party (second sample). GDP is adjusted for inflation, and therefore, variables are independent as well as observations, which mean it is suitable for this test. The two sample permutation test is suited for this situation to find the p-value without distributional assumptions since we do not know anything about the distribution. Since the observations for the democratic and republican party are large, I used a random sampling of R permuted data sets to find an approximate permutation p-value. I took a random sample of R = 10000 permutations, calculated the mean as the statistic of interest for each permutation sample, and then approximated the p-value. I am using the permutation test to test whether the mean GDP of democratic is significantly larger than the mean GDP of the Republican.

Question 2:

In question two, to compare all quarters to the approval ratings, I used a four sample method to compare the four quarters. I chose to use the permutation F-test with a random sampling of R = 1000 from the permutation distribution of the approval ratings. The sample sizes of the democratic approval ratings and the republican approval ratings are too large to permute all the permutations of the Permutation F-test. I am also not willing to make any normality assumptions because I do not know anything about the distribution of the data, so it is satisfied by the permutation F-test. Using the permutation F-test, I can see if there is a difference between all the means of approval ratings for each quarter. Then, by using the Tukey HSD permutation test to determine which quarters differ from others and to identify which quarter differs from the others.

Question 3:

In question three, to see if an increase in GDP is positively associated with higher presidential approval ratings, I first used Pearson's correlation coefficient to see if there was a positive linear association between GDP and approval ratings. After, I attempted to check for normality of the residuals using normal QQ plot to see if t-distribution test for slope is appropriate. Then I went on to use Kendall's tau since it is not based on normality assumption of the response variable. Kendall's tau correlation tells if there was a positive association so that even if Pearson's found the data to be nonlinear, there could still be a positive relationship. We can then measure the extent to which approval ratings increases with GDP. Both Pearson and Kendall's tau do not have to make assumptions about a monotonic relationship or a normal distribution, which makes them suitable for this situation.

Statistical Analysis and Results

Question 1:

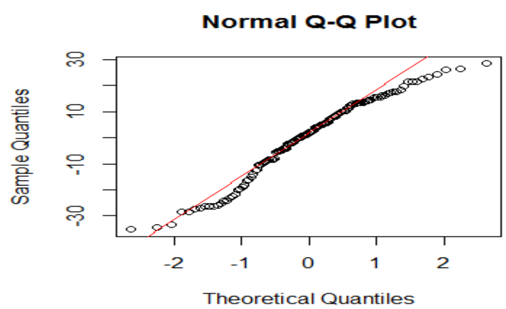
In the two sample permutation test with alpha = 0.05, I attempted to find the difference in GDP between democratic (first sample) and republican (second sample) presidents. The parameters of interest is $\Delta = \mu D - \mu R$. I tested [H0: mean of democratic GDP is equal to mean of republican GDP ($\Delta = 0$)] vs. [Ha: mean of democratic GDP is greater than mean of republican GDP ($\Delta > 0$)] The test statistic is 3007.752 while the p-value is 2.526x10^-5. Since the p-value is 2.526x10^-5 < 0.05, we reject H0 and conclude that there is enough evidence to say that when the president from the democratic party, the GDP is significantly larger than when the president is republican.

Question 2:

To answer which quarter has the strongest association with approval ratings, I used the permutation F-test with alpha = 0.05 to find if there was a difference in means of approval ratings within the quarters. I tested [H0 : $\mu 1 = \mu 1 = \mu 3 = \mu 4$] vs. [Ha: μi are not all equal i = 1,2,3,4]. The test statistic is F = 0.6893 and the p-value is 0.5616. Since the p-value is 0.5616 > 0.05 = alpha, we fail to reject H0 and conclude that there is not enough evidence to conclude that not all approval ratings in each quarter are equal. To check if there are differences between quarters, I applied Tukey HSD permutation test with alpha = 0.05 and the p-values were all above 0.4, which shows that there is no significant difference of approval ratings between the four quarters. All quarters have approximately the same approval ratings.

Question 3:

To see if higher GDP is positively correlated to higher presidential approval ratings, I assigned GDP as the predictor variable and approval rating as the response. I first found the Pearson's correlation coefficient, which is -0.1377827. This means that there is a slight negative linear association. As GDP decreases, presidential approval decreases too. Then when I attempted to check for normality of the residuals using normal QQ plot, it was found that the distribution of residuals are right-skewed and not normal, and thus t-distribution test for slope is not appropriate.



The Kendall's tau permutation test was used because it was not based on normality assumption of approval ratings. The result of the Kendall's tau is $\tau = -0.0426$, which implies almost no association, so pairs are equally likely to be discordant or concordant. Therefore, GDP has no association with presidential approval ratings.

Conclusion

In conclusion, I do trust my results for the most part. When a president from the democratic party serves, the GDP is higher than if the president was republican. All quarters were found to have the same approval rating, as there is no significant difference between any two quarters. It was also found that there is no association between GDP and approval ratings. Although I trust my results, there is still much room for improvement in the ways of testing to answer my questions. There may be slightly better alternatives to performing the tests such a bootstrapping, or separating the time by years and looking at the variables such as GDP and approval ratings. It may lead to a more in-depth analysis that would be great for future studies.

APPENDIX 1

Q1 R code:

```
> pres = read.table("C:\\Users\\Shirley\\Desktop\\pres.txt", header = T)
> dem = pres[pres$party == 1,]
> rep = pres[pres$party == 2,]
> x = dem$gdp
> y = rep$gdp
> "pvalCalc"=function(teststat,teststat.obs){
+   mean(teststat<=teststat.obs)
+ }
>
> "permTestSum"=function(R,x,y){
+   m=length(x)
+   perms=replicate(R,sample(c(x,y),m))
+   ts=apply(perms,2,mean)
+   ts.o=mean(x)
+   pval=pvalCalc(ts,ts.o)
+   return(pval)
+ }
>
> perms=replicate(10000,sample(c(x,y),64))
> ts=apply(perms,2,mean)
> ts.o=mean(x)
>
> permTestSum(10000,x,y)
[1] 2.526e-05
```

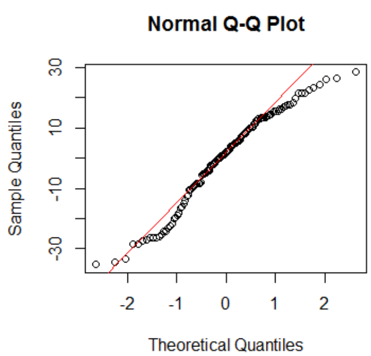
Q2 R code:

```
> q1 = pres[pres$quarter == 1,]
> q2 = pres[pres$quarter == 2,]
> q3 = pres[pres$quarter == 3,]
> q4 = pres[pres$quarter == 4,]
> q1u = q1$approval
> q2u = q2$approval
> q3u = q3$approval
> q4u = q4$approval
> x = c(q1u, q2u, q3u, q4u)
> grps = rep(1:4, each = 30)
> "perm.approx.F" <- function(x,grps,R)
+ {
+   N = length(x)
+   results = rep(NA,R)
+   for (i in 1:R){
+     smpl = sample(1:N,N) # generate random indices, length N
+     results[i] = getF(x[smpl], grps)
+   }
+   return(results)
+ }
>
> "getF" <- function(x,grps)
+ {
+   junk = table(grps)
+   N = length(x)
+   k = length(junk)
+   SST = sum(junk * (getmeans(x,grps))^2) - N*(mean(x))^2
+   F = (SST/(k-1))/((sum((x-mean(x))^2) - SST)/(N-k))
+   return(F)
+ }
>
> "getmeans" <- function(x,grps)
+ {
+   junk = table(grps) # evaluate n_i (counts)
+   k = length(junk)
+   meanvec = rep(NA,k) # create empty vector
+   for (i in 1:k){
+     meanvec[i] = mean(x[grps==names(junk)[i]])
+   }
+   return(meanvec)
+ }
```

```
> Fobs = getF(x,grps)
> Fobs
[1] 0.6893181
> perm.F = perm.approx.F(x, grps, R=10000)
> perm.pval = mean(perm.F >= Fobs)
> perm.pval
[1] 0.558


> Tukey.HSD(x, grps, k=4, 0.05, 1000)
$sig
      [,1]  [,2]  [,3] [,4]
[1,]    NA    NA    NA   NA
[2,] FALSE    NA    NA   NA
[3,] FALSE FALSE    NA   NA
[4,] FALSE FALSE FALSE   NA

$pvalsTij
      [,1]  [,2]  [,3] [,4]
[1,]    NA    NA    NA   NA
[2,] 0.891    NA    NA   NA
[3,] 0.866 1.000    NA   NA
[4,] 0.495 0.891 0.915   NA
```

Q3 R code:

```
> y = pres$approval
> x = pres$gdp
> fit = lm(y~x)
> summary(lm(y~x))

Call:
lm(formula = y ~ x)

Residuals:
    Min      1Q  Median      3Q     Max 
-35.240  -9.480   1.712  13.030  28.411 

Coefficients:
             Estimate Std. Error t value Pr(>|t|)    
(Intercept) 62.938635   4.586612  13.722   <2e-16 ***
x           -0.001959   0.001297  -1.511    0.133    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 15.59 on 118 degrees of freedom
Multiple R-squared:  0.01898,	Adjusted R-squared:  0.01067 
F-statistic: 2.283 on 1 and 118 DF,  p-value: 0.1334

> e = fit$residuals
> qqnorm(e)
> qqline(e, col = "red")
```



```
> cor(x,y , method = "kendall")
[1] -0.04250705
> cor(x,y , method = "pearson")
[1] -0.1377827
> cor.test(x,y, alternative = "less", method = "kendall", exact = NULL,
conf.level = 0.95, continuity = FALSE)

        Kendall's rank correlation tau

data:  x and y
z = -0.681, p-value = 0.2479
alternative hypothesis: true tau is less than 0
sample estimates:
        tau
-0.04250705

> cor.test(x,y, alternative = "less", method = "pearson", exact = NULL,
conf.level = 0.95, continuity = F)

        Pearson's product-moment correlation

data:  x and y
t = -1.5111, df = 118, p-value = 0.06672
alternative hypothesis: true correlation is less than 0
95 percent confidence interval:
 -1.00000000  0.01340134
sample estimates:
       cor
-0.1377827
```