

Exam 1: US President Inauguration Speech Length and Party Affiliation

Kyle Ke

10/27/2021

Abstract

The analysis investigates whether there is a statistically significant difference in the inauguration speech word count between US presidents who are Republican vs Democrats. It is performed using a two-sample t-test, Mann–Whitney U test, approximate permutation test, and bootstrap test. The results conclude that there is no statistically significant difference in inauguration speech length between Republican and Democratic presidents.

Contents

Introduction	2
Methdology	2
Data Description	3
Results	5
Conclusions	6
Recommedations for Future Studies	6
References	7
Appendix	8
Code	8

Introduction

The majority (35 out of 47) of all US presidents, are affiliated with either the Republican or the Democratic Party. Since the inauguration speech word count of all US presidents are available, it would be interesting to see if there is a difference in inauguration speech length between Republican presidents and Democratic presidents.

Methdology

The inauguration speech word count of all 24 inauguration speeches by Republican presidents and 23 inauguration speeches by Democratic presidents were collected from the website by (Summers (n.d.)). We have 24 observations, denoted x_1, x_2, \dots, x_{24} in the Democratic group, and 23 observations, denoted y_1, y_2, \dots, y_{23} in the Republican group. A two-sample t-test, Mann-Whitney U test, approximate permutation test, and bootstrap test were performed during the analysis.

The approximate permutation test (Helwig (2017)) was choosen because the number of combinations for $\binom{47}{24}$ is 1.612×10^{13} . The number of all combinations is too big for us to compute all of the test statistics (difference in means) for each combination due to limited computing resources. Instead, a random sample of 24 word counts was taken 10000 times. Each sample of 24 was taken without replacement. However, the same combination could theoretically appear in more than one sample because we could not enumerate all combinations. The mean of each sample was taken and subtracted from the mean of its complement (i.e. the other 23 word counts) to obtain our test statistic: difference in mean. All 10000 test statistics were then pooled as a single distribution. The p-value was obtained by counting all of the absolute value of the test statistics that are greater than the absolute value of the original test statistic and dividing that count by 10000.

A bootstrap (randomization) test was performed taking a random sample of 47 word counts 10000 times. However, each sample of 47 word counts was taken with replacement, meaning the same word count could appear more than once within a sample. Within each sample, the mean of the word counts indexed [1:24] was subtracted from the mean of the word counts indexed [25:47] to obtain our test-statistic. All 10000 test statistics were then pooled as a single distribution. The p-value was obtained by counting all of the absolute value of the test statistics that are greater than the the absolute value of the original test statistic and dividing that count by 10000.

The null and alternative hypothesis are as follows:

Two-sample t-test, Approximate Permutation Test, and Bootstrap Test: (Helwig (2017))

$H_o : \mu_x = \mu_y$ The mean inauguration speech word count is the same between Democratic and Republican presidents

$H_A : \mu_x \neq \mu_y$ The mean inauguration speech word count is different between Democratic and Republican presidents

Significance Level: .05

Mann-Whitney U test:(Shier (n.d.))

$H_o : P(x_i > y_j) = .5$ The median inauguration speech word count is the same between all Democratic and Republican presidents

$H_A : P(x_i > y_j) \neq .5$ The median inauguration speech word count is not the same between all Democratic and Republican Presidents

Significance Level: .05

Data Description

Summary Statistics

Table 1: Summary Statistics for Inauguration Speech Word Counts

	Republicans	Democrats
Minimum	700.00	559.00
Maximum	5434.00	4809.00
1. Quartile	1641.50	1436.50
3. Quartile	3406.00	2334.00
Mean	2532.17	2036.09
Median	2373.50	1808.00
Stdev	1192.25	955.61

Box plots

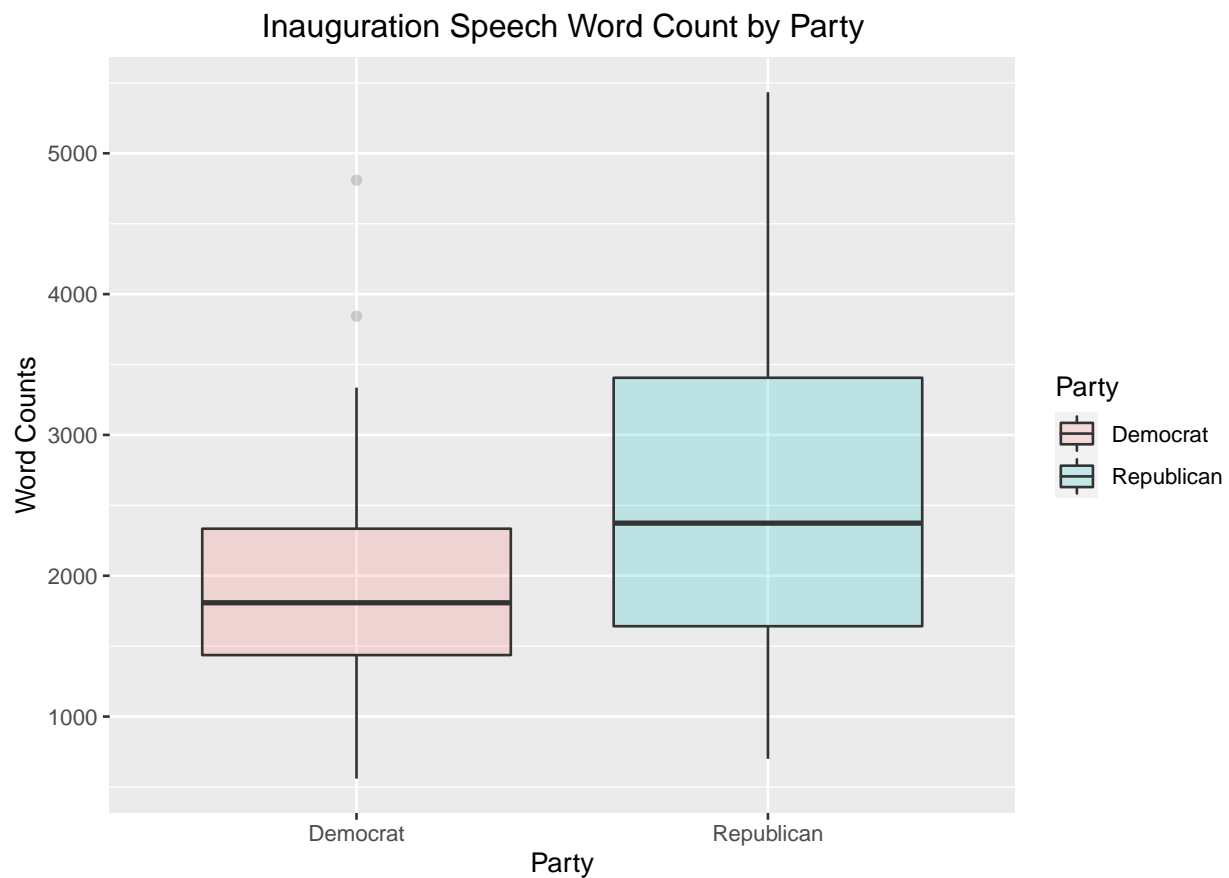


Figure 1: Inauguration Speech Word Count Boxplots by Party

We observe from the box plots and summary statistics that the median word count for Republican presidents is higher than the median word count for Democratic presidents. Moreover, the variance in word count is higher for Republican president compared to Democratic presidents.

Quantile-Quantile Plots

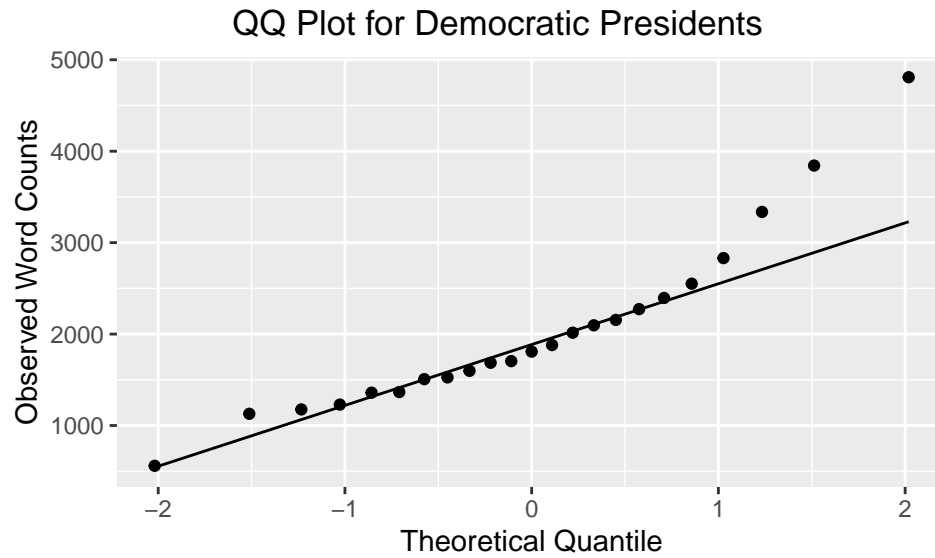


Figure 2: QQ Plot of Inauguration Speech Word Count for Democratic Presidents

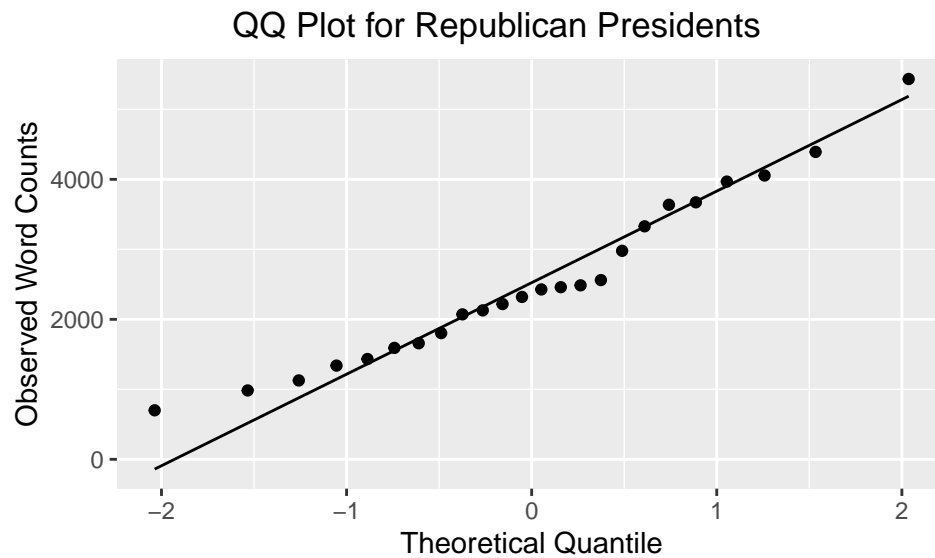


Figure 3: QQ Plot of Inauguration Speech Word Count for Republican Presidents

The quantile-quantile plots show that inauguration speech word counts for Democratic presidents is heavily right skewed. Thus, the data is exhibiting properties that might be more suitable for the non-parametric Mann-Whitney U test.

Results

Two sample t-test

The two sample t-test has a p-value of 0.123. Under a significance level of .05, we fail to reject the null hypothesis that the mean inauguration speech word count is the same between Democratic and Republican presidents.

Mann-Whitney U test

The (Mann-Whitney U test has a p-value of 0.134. Under a significance level of .05, we fail to reject the null hypothesis that the median inauguration speech word count is the same between Democratic and Republican presidents.

The results from the t-test and Man-Whitney U test are in congruent with each other. At the .05 significance level, both tests fail to reject the null hypothesis that there is a difference in the central tendency of inauguration speech length between Democratic and Republican presidents.

Approximate Permutation Test

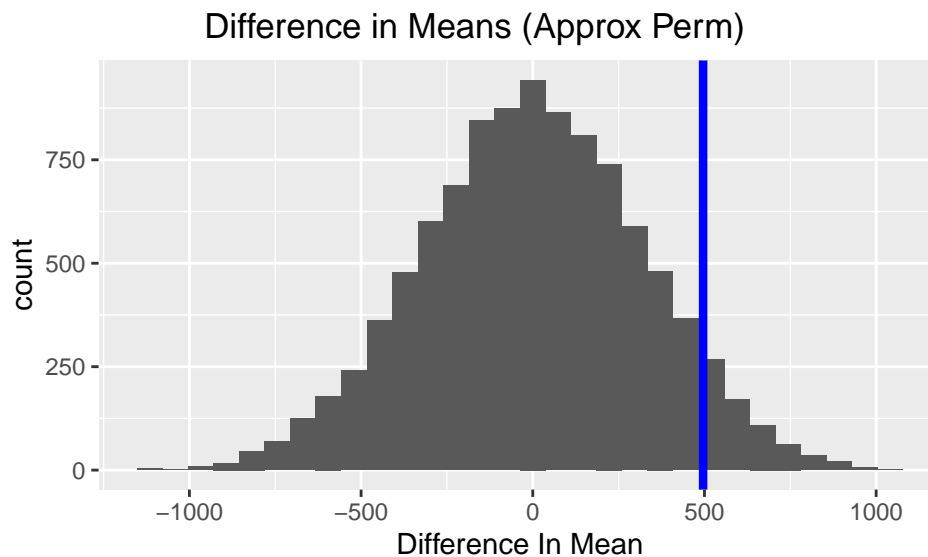


Figure 4: Histogram of Difference in Means from Approx Permutation Tests

The results of the approximate permutation test yielded a p-value (two-sided) of 0.126. Under the significance level of .05, we fail to reject the null hypothesis that the mean inauguration speech word count between Democratic and Republican presidents is the same.

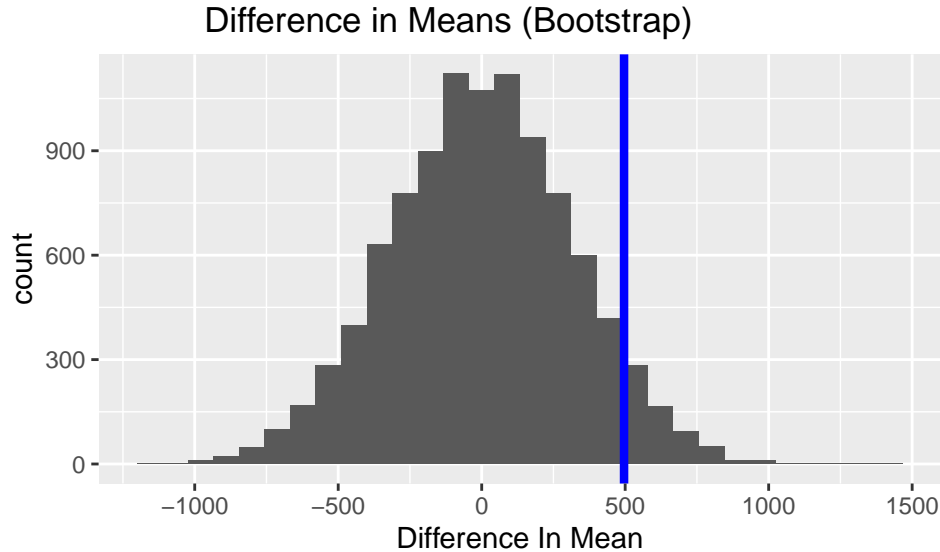


Figure 5: Histogram of Difference in Means from Bootstrap

The results of the bootstrap test yielded a p-value (two-sided) of 0.12. Under the significance level of .05, we fail to reject the null hypothesis that the mean inauguration speech word count between Democratic and Republican presidents is the same.

Conclusions

Although, through eye observations of the summary statistics and box plots, Republican presidents appear to have generally longer inauguration speeches, four different tests have failed to detect a statistically significant difference in inauguration speech word count between Democratic presidents and Republican presidents.

Recommendations for Future Studies

Future studies could investigate the two outliers in the Democratic president group as shown in Figure 1. It is possible the results of the analysis could change as the t-test is especially sensitive to outliers. Future studies could also consider conducting one-tailed tests to determine if the speech length of Republican presidents is greater than the speech length of Democratic presidents, instead of only investigating if there is any difference between the groups as done in this analysis. Moreover, future studies could investigate if there is difference across time (e.g., 1800s, 1900s, and 2000s).

References

- Helwig, Nathaniel. 2017. “Permutation Tests.” University of Minnesota Twin Cities. https://statmoodle.byu.edu/pluginfile.php/168/mod_resource/content/2/perm-Notes.pdf.
- Shier, Rosie. n.d. *The Mann-Whitney U Test*. statstutoruk. <https://www.statstutor.ac.uk/resources/uploaded/mannwhitney.pdf>.
- Summers, Robert. n.d. “Inaugural Address Lengths.” *POTUS*. <https://www.potus.com/presidential-facts/inaugural-address-length/>.
- Wickham, Hadley. 2016. *Ggplot2: Elegant Graphics for Data Analysis*. Springer-Verlag New York. <https://ggplot2.tidyverse.org>.
- . 2021. *Rvest: Easily Harvest (Scrape) Web Pages*. <https://CRAN.R-project.org/package=rvest>.
- Wickham, Hadley, Winston Chang, Lionel Henry, Thomas Lin Pedersen, Kohske Takahashi, Claus Wilke, Kara Woo, Hiroaki Yutani, and Dewey Dunnington. 2021. *Ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics*. <https://CRAN.R-project.org/package=ggplot2>.
- Wickham, Hadley, Romain François, Lionel Henry, and Kirill Müller. 2021. *Dplyr: A Grammar of Data Manipulation*. <https://CRAN.R-project.org/package=dplyr>.
- Wuertz, Diethelm, Tobias Setz, and Yohan Chalabi. 2020. *fBasics: Rmetrics - Markets and Basic Statistics*. <https://www.rmetrics.org>.
- Xie, Yihui. 2014. “Knitr: A Comprehensive Tool for Reproducible Research in R.” In *Implementing Reproducible Computational Research*, edited by Victoria Stodden, Friedrich Leisch, and Roger D. Peng. Chapman; Hall/CRC. <http://www.crcpress.com/product/isbn/%0A%20%20%20%209781466561595>.
- . 2015. *Dynamic Documents with R and Knitr*. 2nd ed. Boca Raton, Florida: Chapman; Hall/CRC. <https://yihui.org/knitr/>.
- . 2021a. *formatR: Format r Code Automatically*. <https://github.com/yihui/formatR>.
- . 2021b. *Knitr: A General-Purpose Package for Dynamic Report Generation in r*. <https://yihui.org/knitr/>.

Appendix

Code

```
knitr::opts_chunk$set(message = FALSE)
library(rvest)
library(dplyr)
library(ggplot2)
library(fBasics)
library(knitr)
library(formatR)
knitr::write_bib(c('rvest','dplyr','ggplot2','fBasics','knitr','formatR'),
                 file='packages.bib',width = 60)
length_of_inaugural_url = "https://www.potus.com/presidential-facts/inaugural-address-length/"
length_of_inaugural_list = rvest::read_html(length_of_inaugural_url)
length_of_inaugural_df = length_of_inaugural_list %>% rvest::html_table() %>% as.data.frame(.)

party_affiliation_url = "https://www.potus.com/presidential-facts/political-party/"
party_affiliation_list = rvest::read_html(party_affiliation_url)
party_affiliation_df = party_affiliation_list %>% rvest::html_table() %>% as.data.frame(.)
length_of_inaugural_df = length_of_inaugural_df[-c(60:64),]
party_affiliation_df = party_affiliation_df[-c(1,21,38:51),]
party_affiliation_df$Party = c(rep('Republican',19),rep('Democrat',16))

joint_df = merge(party_affiliation_df,length_of_inaugural_df,by="President")
joint_df$Length.in.Words = as.numeric(gsub(",", "", joint_df$Length.in.Words))
joint_df = joint_df[,-c(2,3,4,6,7,8)]
Republicans=joint_df[joint_df$Party=="Republican",]$Length.in.Words
Democrats = joint_df[joint_df$Party=="Democrat",]$Length.in.Words
options(scipen = 0,digits=6)
summary_stats = cbind(basicStats(Republicans),basicStats(Democrats))
summary_stats = summary_stats[c("Minimum","Maximum","1. Quartile","3. Quartile",
                                "Mean","Median","Stdev"),]
knitr::kable(summary_stats,caption = "Summary Statistics for Inauguration Speech Word Counts",
              col.names=colnames(summary_stats))

ggplot(joint_df, aes(x=as.factor(Party), y=Length.in.Words,fill=Party)) +
  geom_boxplot(alpha=0.2) +
  ggtitle('Inauguration Speech Word Count by Party ') +
  ylab('Word Counts') + xlab('Party')
ggplot(joint_df[joint_df$Party=="Democrat",],aes(sample = Length.in.Words )) +
  stat_qq() +
  stat_qq_line() + ggtitle('QQ Plot for Democratic Presidents ') +
  xlab('Theoretical Quantile') + ylab('Observed Word Counts')

ggplot(joint_df[joint_df$Party=="Republican",],aes(sample = Length.in.Words )) +
  stat_qq() +
  stat_qq_line() + ggtitle('QQ Plot for Republican Presidents ') +
  xlab('Theoretical Quantile') + ylab('Observed Word Counts')

t_Test_results <- t.test(Republicans,Democrats,"two.sided" ,paired = FALSE,var.equal = TRUE,fig.align='left')
t_Test_results
wilcox_test_results <- wilcox.test(Republicans,Democrats, exact = TRUE, correct = TRUE,
```



```

                                alternative = 'two.sided')
set.seed(5)
meandif <- function(x,y) mean(x) - mean(y)
perm2samp <- function(x,y,myfun=meandif,nsamp=10000,
                      alternative=c("two.sided","less","greater")){
  theta.hat = myfun(x,y)
  m = length(x)
  n = length(y)
  N = m + n
  z = c(x,y)
  gmat = replicate(nsamp,sample.int(N,m))
  theta.mc = apply(gmat,2,function(c,z){myfun(z[c],z[-c])},z=z)
  if(alternative[1]=="less"){
    aslperm = sum(theta.mc <= theta.hat) / nsamp
  } else if(alternative[1]=="greater"){
    aslperm = sum(theta.mc >= theta.hat) / nsamp
  } else{
    aslperm = sum(abs(theta.mc) >= abs(theta.hat)) / nsamp
  }
  list(theta.hat=theta.hat,theta.mc=theta.mc,asl=aslperm)
  return(list(theta.hat,theta.mc,aslperm))
}

perm_results = perm2samp(Republicans,Democrats,myfun=meandif,nsamp=10000,alternative="two.sided")
theta.sampled=perm_results[[2]]
ggplot() +aes(theta.sampled)+ geom_histogram() + ggtitle('          Difference in Means (Approx Perm)

set.seed(5)
meandif <- function(x,y) mean(x) - mean(y)
perm2samp <- function(x,y,myfun=meandif,nsamp=10000,
                      alternative=c("two.sided","less","greater")){
  theta.hat = myfun(x,y)
  m = length(x)
  n = length(y)
  N = m + n
  z = c(x,y)
  gmat = replicate(nsamp,sample.int(N,N,replace=TRUE))
  theta.mc = apply(gmat,2,function(c,z){myfun(z[c[1:23]],z[c[24:47]])},z=z)
  if(alternative[1]=="less"){
    aslperm = sum(theta.mc <= theta.hat) / nsamp
  } else if(alternative[1]=="greater"){
    aslperm = sum(theta.mc >= theta.hat) / nsamp
  } else{
    aslperm = sum(abs(theta.mc) >= abs(theta.hat)) / nsamp
  }
  list(theta.hat=theta.hat,theta.mc=theta.mc,asl=aslperm)
  return(list(theta.hat,theta.mc,aslperm))
}

perm_results = perm2samp(Republicans,Democrats,myfun=meandif,nsamp=10000,alternative="two.sided")
theta.sampled=perm_results[[2]]

```

```
ggplot() +aes(theta.sampled)+ geom_histogram() + ggtitle('          Difference in Means (Bootstrap) ')
  xlab('Difference In Mean') + geom_vline(xintercept = perm_results[[1]], color = "blue", size=1.5)

knitr::opts_chunk$set(tidy.opts=list(width.cutoff=100), tidy=TRUE)
```