

Report: A Problem with Presidents

Author: Harish Gureddygar

Date: 11/12/2021

Summary

The analysis in this report deals with data containing the birth and death information of all presidents who have served the United States to see if we can discover any new information from the data provided. We perform bivariate analysis on top 10 presidents ranked by longest lived and shortest lived to see if we can find any trend. We found that a straight line comfortably fits through the data hence a linear relationship exists between 'year_of_birth' and 'days_lived' variables.

We also look at the central tendency of the data set as a whole to know more about the distribution of data by performing univariate analysis on the 'lived_days' column. The result was that data strongly follows Gaussian or normal distribution with the mean of 26370 days and a median of 26227 days. The weighted mean came out to be 27121 days. The mean is close to the median which indicates that data is symmetrically distributed. The data has a maximum of 35470 days and a minimum of 26227 days. The standard deviation of 4566 days suggest that data is clustered well around the mean.

Prerequisites & Assumptions

The analysis in this report is done using Statistical R programming language on MacOS environment. The R code and the required libraries can be found in the .R file.

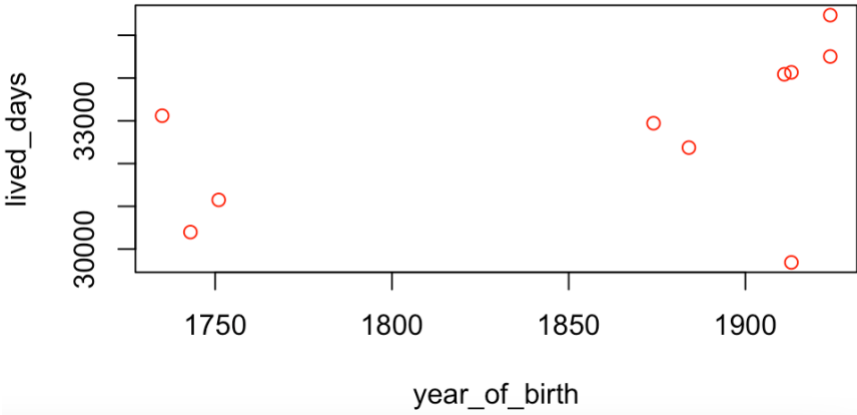
Since there was no information available about weights to calculate weighted mean, we assumed different weights for presidents who were alive and those who no longer were alive. The weights were chosen by solving for the equation $39x + 6y = 1$ since there were 39 presidents who were not alive anymore and 6 presidents who are still alive. Our assumption was that sum of all weights should equal to 1.

Bivariate analysis

Top 10 Presidents from longest lived to shortest lived

PRESIDENT	BIRTH DATE	BIRTH PLACE	DEATH DATE	LOCATION OF DEATH	year_of_birth	lived_years	lived_months	lived_days
Jimmy Carter	Oct 1, 1924	Plains, Georgia			1924	97.11*	1165.33*	35470*
George Bush	June 12, 1924	Milton, Mass.	Nov 30, 2018	Houston, Texas	1924	94.47	1133.6	34504
Gerald Ford	July 14, 1913	Omaha, Nebraska	Dec 26, 2006	Rancho Mirage, Cal.	1913	93.45	1121.39	34133
Ronald Reagan	Feb 6, 1911	Tampico, Illinois	June 5, 2004	Los Angeles, Cal.	1911	93.33	1119.97	34088
John Adams	Oct 30, 1735	Quincy, Mass.	July 4, 1826	Quincy, Mass.	1735	90.68	1088.13	33119
Herbert Hoover	Aug 10, 1874	West Branch, Iowa	Oct 20, 1964	New York, New York	1874	90.19	1082.32	32943
Harry S. Truman	May 8, 1884	Lamar, Missouri	Dec 26, 1972	Kansas City, Missouri	1884	88.63	1063.58	32373
James Madison	Mar 16, 1751	Port Conway, Va.	June 28, 1836	Orange Co., Va.	1751	85.28	1023.4	31150
Thomas Jefferson	Apr 13, 1743	Albemarle Co., Va.	July 4, 1826	Albemarle Co., Va.	1743	83.22	998.7	30397
Richard Nixon	Jan 9, 1913	Yorba Linda, Cal.	Apr 22, 1994	New York, New York	1913	81.28	975.43	29688

*Still alive

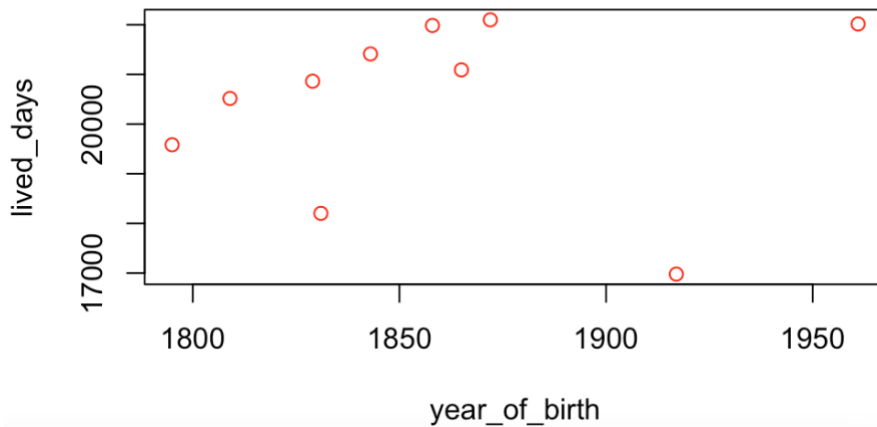


As seen in the scatter plot, the relationship between ‘lived_days’ and ‘year_of_birth’ is linear. We can also say that as ‘year_of_birth’ increases, ‘lived_days’ also increases proportionately. We need to notice that there are couple of outliers of which one data point is of a living president.

Top 10 Presidents from shortest lived to longest lived

PRESIDENT	BIRTH DATE	BIRTH PLACE	DEATH DATE	LOCATION OF DEATH	year_of_birth	lived_years	lived_months	lived_days
John F. Kennedy	May 29, 1917	Brookline, Mass.	Nov 22, 1963	Dallas, Texas	1917	46.48	557.8	16978
James A. Garfield	Nov 19, 1831	Cuyahoga Co., Ohio	Sep 19, 1881	Elberon, New Jersey	1831	49.83	598	18202
James K. Polk	Nov 2, 1795	Mecklenburg Co., N.C.	June 15, 1849	Nashville, Tennessee	1795	53.62	643.43	19583
Abraham Lincoln	Feb 12, 1809	LaRue Co., Kentucky	Apr 15, 1865	Washington, D.C.	1809	56.17	674.1	20516
Chester Arthur	Oct 5, 1829	Fairfield, Vermont	Nov 18, 1886	New York, New York	1829	57.12	685.43	20863
Warren G. Harding	Nov 2, 1865	Morrow County, Ohio	Aug 2, 1923	San Francisco, Cal.	1865	57.75	693	21091
William McKinley	Jan 29, 1843	Niles, Ohio	Sep 14, 1901	Buffalo, New York	1843	58.62	703.53	21412
Theodore Roosevelt	Oct 27, 1858	New York, New York	Jan 6, 1919	Oyster Bay, New York	1858	60.19	722.32	21985
Barack Obama	Aug 4, 1961	Honolulu, Hawaii			1961	60.27*	723.23*	22014*
Calvin Coolidge	July 4, 1872	Plymouth, Vermont	Jan 5, 1933	Northampton, Mass.	1872	60.51	726.03	22099

*Still alive

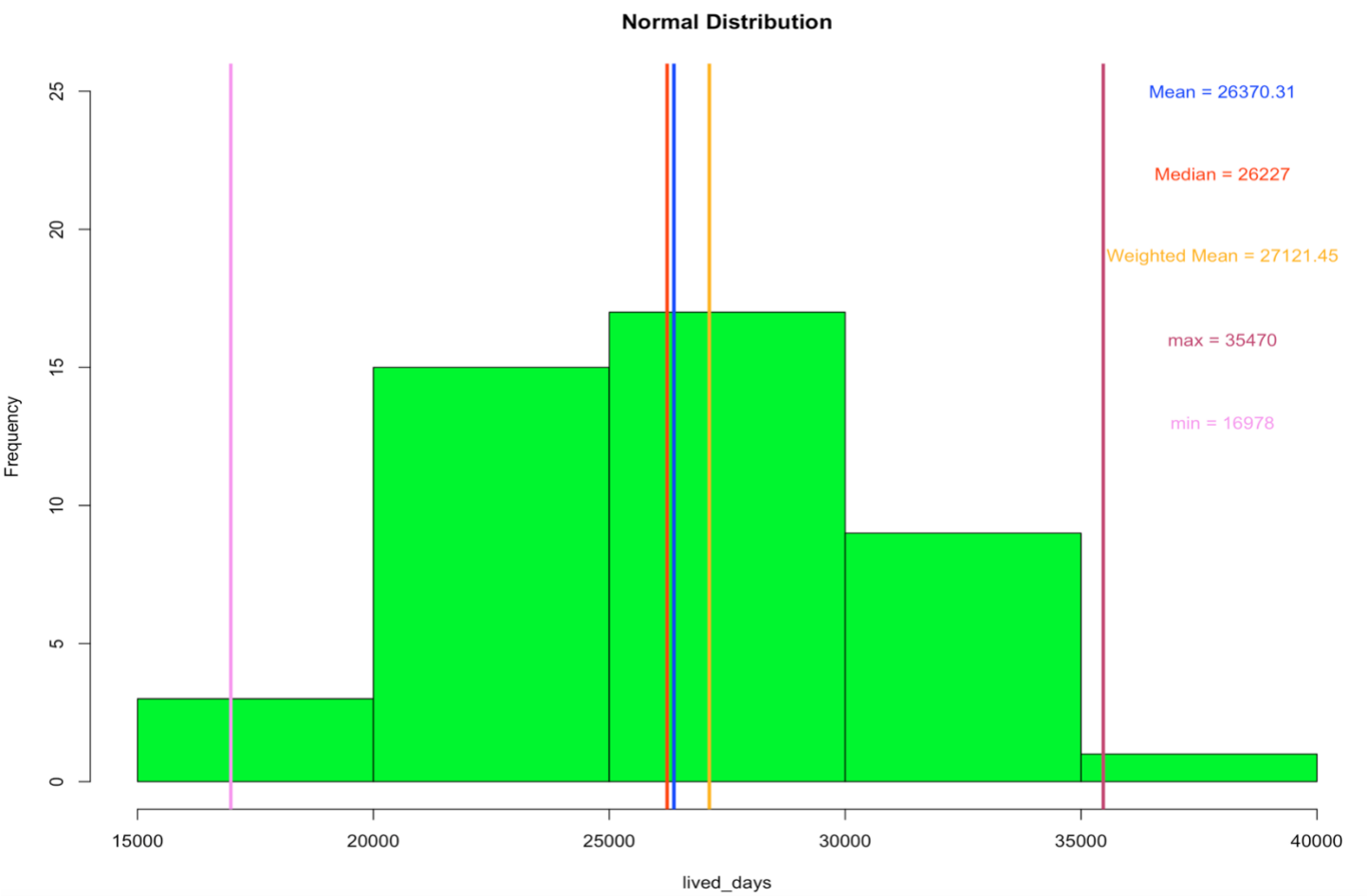


The scatter plot for shortest lived presidents also follows linear pattern (straight line). The three outliers we see in the above scatter plot are unfortunately victims of assassinations which skewed the plot. But otherwise, the trend we notice is that as 'year_of_birth' increases, 'lived_days' increases in a proportional way.

Univariate analysis

Basic statistics- mean, median, mode, max, min and standard deviation

	lived_days
Mean	26370.31
Weighted Mean	27121.45
Median	26227
Mode	16978,18202,19583,20516,20863,21091,21412,21985,22014,22099,23082,23098,23524,23695,23967,24320,24507,24676,24767,24890,25673,26030,26227,26471,26729,27088,27478,27522,27544,28163,28574,28654,28846,29085,29446,29688,30397,31150,32373,32943,33119,34088,34133,34504,35470
Max	35470
Min	16978
Standard deviation	4566.33



Conclusion

After completing the exploratory data analysis on the data set, we found out that 'year_of_birth' and 'lived_days' have a linear relationship and we observed that as 'year_of_birth' increases, the 'lived_days' increased as well. There were few outliers as well but those were mainly a result of presidents either still living or being assassinated which skewed the data. We can therefore conclude that modern presidents tend to live a longer life which can probably be attributed to better healthcare advancement as well as tougher security around them.