

# Age Analysis of Registered Voters in Tyrell County, North Carolina, United States of America

## Introduction

This project analyzed the distribution of age for registered voters in Tyrell County, North Carolina. The data used in this analysis was sourced from the North Carolina State Board of Elections. A small subset of this data was used to for this analysis. Please note that this is not a representative sample of registered voters in Tyrell County, North Carolina.

## Descriptive Statistics

```
In [9]: # import packages
import pandas as pd
import matplotlib.pyplot as plt
import matplotlib.ticker as ticker
```

```
In [10]: # Define Function to Read NC Voter Registration Data
def read_csv_ncvoterdata(voterdata):
    return pd.read_csv(
        voterdata, sep="\t", header=0, encoding="unicode_escape", low_memory
    )

# Load Data
df = read_csv_ncvoterdata("ncvoter89.txt")
display(df)
```

	county_id	county_desc	voter_reg_num	ncid	last_name	first_name	mid
0	89	TYRRELL	2941	EE4964	ABBOTT	DOUGLAS	
1	89	TYRRELL	2289	EE4315	ABBOTT	LARA	
2	89	TYRRELL	5002	EE6577	ABDUL-WAHAB	ISAIAH	
3	89	TYRRELL	705	EE2736	ABEL	JUDY	
4	89	TYRRELL	5822	BG42384	ACREMAN	CHELSEA	
...	...	...	...	...	...	...	...
595	89	TYRRELL	5037	EE5070	COOPER	BRENDA	
596	89	TYRRELL	5490	EE6874	COOPER	CHARLIE	
597	89	TYRRELL	4289	EE6132	COOPER	CHRISTOPHER	
598	89	TYRRELL	927	EE2957	COOPER	DONALD	
599	89	TYRRELL	925	EE2955	COOPER	DURWOOD	

600 rows x 67 columns

```
In [11]: # Define Functions to Calculate Mean, Median, and Standard Deviation of Age
def mean_age(df):
    # calculate mean of column with "age" in it
    age_column = [col for col in df.columns if "age" in col]
    if age_column:
        # Assuming there's only one age column in NC voter file data
        column_name = age_column[0]
        # Calculate the mean of the identified column
        result = df[column_name].mean()
        return result
    else:
        result = print("No column containing 'age' found.")
        return result

def median_age(df):
    # calculate median of column with "age" in it
    age_column = [col for col in df.columns if "age" in col]
    if age_column:
        # Assuming there's only one age column in NC voter file data
        column_name = age_column[0]
```

```

        # Calculate the mean of the identified column
        result = df[column_name].median()
        return result
    else:
        result = print("No column containing 'age' found.")
        return result

def std_age(df):
    # calculate standard deviation of column with "age" in it
    age_column = [col for col in df.columns if "age" in col]
    if age_column:
        # Assuming there's only one age column in NC voter file data
        column_name = age_column[0]
        # Calculate the mean of the identified column
        result = df[column_name].std()
        return result
    else:
        result = print("No column containing 'age' found.")
        return result

# Calculate Mean, Median, and Standard Deviation
summary = {
    "Statistic": [
        "Mean Age",
        "Median Age",
        "Standard Deviation of Age",
        "Count of Sample of Registered Voters",
    ],
    "Value (Rounded)": [
        round(mean_age(df), 2),
        round(median_age(df), 2),
        round(std_age(df), 2),
        round(len(df), 2),
    ],
}

# Create DataFrame
summarydf = pd.DataFrame(summary)
print(summarydf)

```

	Statistic	Value (Rounded)
0	Mean Age	57.48
1	Median Age	60.00
2	Standard Deviation of Age	19.74
3	Count of Sample of Registered Voters	600.00

## Data Visualization

To visualize the distribution of age among this non-representative sample of 600 registered voters in Tyrell County, North Carolina, we use a histogram created with the matplotlib Python package.

```
In [15]: def generate_histogram_age(df):
age_column = [col for col in df.columns if "age" in col]
plt.figure(figsize=(10, 6))
bins = 6
plt.hist(df[age_column], color="orange", bins=bins, edgecolor="black")
plt.title("Age Distribution for Registered Voters in Tyrell County, NC")
plt.xlabel("Age")
plt.ylabel("Frequency")
plt.gca().yaxis.set_major_formatter(
    ticker.FuncFormatter(lambda x, _: f"{int(x):,}")
)

plt.savefig("output.png")

plt.show()
```

## Conclusion

For this non-representative sample of 600 Tyrell County, NC registered voters, the mean age is 57.48, the median age is 60, and the standard deviation of age is 19.74. Although registered voters tend to be older than the general population, the mean age appears to be higher than expected. It is likely that the true mean age of all registered voters in Tyrell County, NC is lower than 57.48.

Further analysis of voter registration data from the North Carolina State Board of Elections (NCSBE) could be supplemented by building in a function to scrape the zipped files present on the website and download the stored .txt files for analysis. The NCSBE updates the voter registration records weekly on Saturday mornings. Another automation could be built to download, analyze, and visualize these records weekly after they are updated.