Ivana Zdravevska

Project 6

November 24, 2019

Confidence Interval Function

In [7]:
```
import numpy as np
import scipy.stats

def confidence_interval(data, confidence=0.95):
    a = 1.0 * np.array(data)
    n = len(a)
    m, se = np.mean(a), scipy.stats.sem(a)
    h = se * scipy.stats.t.ppf((1 + confidence) / 2., n-1)
    return m, m-h, m+h
```

Red Wine Data Set

In [8]:
```
import pandas as pd
import numpy as np

df_red = pd.read_csv('/Users/ivanazdravevska/Desktop/winequality-red.csv', sep=';')
df_red.head(10)
```

Out[8]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol
0	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4
1	7.8	0.88	0.00	2.6	0.098	25.0	67.0	0.9968	3.20	0.68	9.8
2	7.8	0.76	0.04	2.3	0.092	15.0	54.0	0.9970	3.26	0.65	9.8
3	11.2	0.28	0.56	1.9	0.075	17.0	60.0	0.9980	3.16	0.58	9.8
4	7.4	0.70	0.00	1.9	0.076	11.0	34.0	0.9978	3.51	0.56	9.4
5	7.4	0.66	0.00	1.8	0.075	13.0	40.0	0.9978	3.51	0.56	9.4
6	7.9	0.60	0.06	1.6	0.069	15.0	59.0	0.9964	3.30	0.46	9.4
7	7.3	0.65	0.00	1.2	0.065	15.0	21.0	0.9946	3.39	0.47	10.0
8	7.8	0.58	0.02	2.0	0.073	9.0	18.0	0.9968	3.36	0.57	9.5
9	7.5	0.50	0.36	6.1	0.071	17.0	102.0	0.9978	3.35	0.80	10.5

In [9]:
```
df_red.describe()
```

Out[9]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulf dioxi
count	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.000000	1599.0000
mean	8.319637	0.527821	0.270976	2.538806	0.087467	15.874922	46.4677
std	1.741096	0.179060	0.194801	1.409928	0.047065	10.460157	32.8953
min	4.600000	0.120000	0.000000	0.900000	0.012000	1.000000	6.0000
25%	7.100000	0.390000	0.090000	1.900000	0.070000	7.000000	22.0000
50%	7.900000	0.520000	0.260000	2.200000	0.079000	14.000000	38.0000
75%	9.200000	0.640000	0.420000	2.600000	0.090000	21.000000	62.0000
max	15.900000	1.580000	1.000000	15.500000	0.611000	72.000000	289.0000

Confidence Interval for Red Wine Fixed Acidity

In [10]:
```
red_fixed_acidity = df_red[["fixed acidity"]]

#Calling the function to calculate the confidence interval for the fixed acidity in red
confidence_interval(data=red_fixed_acidity,confidence=0.95)
```

Out[10]:
```
(8.31963727329581, array([8.23423376]), array([8.40504078]))
```

White Wine Data Set

In [11]:
```
df_white = pd.read_csv('/Users/ivanazdravevska/Desktop/winequality-white.csv', sep=';')
df_white.head(10)
```

Out[11]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulfur dioxide	density	pH	sulphates	alcohol
0	7.0	0.27	0.36	20.7	0.045	45.0	170.0	1.0010	3.00	0.45	8.8
1	6.3	0.30	0.34	1.6	0.049	14.0	132.0	0.9940	3.30	0.49	9.5
2	8.1	0.28	0.40	6.9	0.050	30.0	97.0	0.9951	3.26	0.44	10.1
3	7.2	0.23	0.32	8.5	0.058	47.0	186.0	0.9956	3.19	0.40	9.9
4	7.2	0.23	0.32	8.5	0.058	47.0	186.0	0.9956	3.19	0.40	9.9
5	8.1	0.28	0.40	6.9	0.050	30.0	97.0	0.9951	3.26	0.44	10.1
6	6.2	0.32	0.16	7.0	0.045	30.0	136.0	0.9949	3.18	0.47	9.6
7	7.0	0.27	0.36	20.7	0.045	45.0	170.0	1.0010	3.00	0.45	8.8
8	6.3	0.30	0.34	1.6	0.049	14.0	132.0	0.9940	3.30	0.49	9.5
9	8.1	0.22	0.43	1.5	0.044	28.0	129.0	0.9938	3.22	0.45	11.0

In [12]:
```
df_white.describe()
```

Out[12]:

	fixed acidity	volatile acidity	citric acid	residual sugar	chlorides	free sulfur dioxide	total sulf dioxi
count	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.000000	4898.0000
mean	6.854788	0.278241	0.334192	6.391415	0.045772	35.308085	138.3606
std	0.843868	0.100795	0.121020	5.072058	0.021848	17.007137	42.4980
min	3.800000	0.080000	0.000000	0.600000	0.009000	2.000000	9.0000
25%	6.300000	0.210000	0.270000	1.700000	0.036000	23.000000	108.0000
50%	6.800000	0.260000	0.320000	5.200000	0.043000	34.000000	134.0000
75%	7.300000	0.320000	0.390000	9.900000	0.050000	46.000000	167.0000
max	14.200000	1.100000	1.660000	65.800000	0.346000	289.000000	440.0000

Confidence Interval for White Wine Fixed Acidity

In [13]:
```
white_fixed_acidity = df_white[["fixed acidity"]]

#Calling the function to calculate the confidence interval for the fixed acidity in white
confidence_interval(data=white_fixed_acidity,confidence=0.95)
```

Out[13]:
```
(6.854787668436097, array([6.83114913]), array([6.87842621]))
```

The mean quality of the fixed acidity is higher in the red wine. The fixed acidity in the red wine has a mean of 8.3196 while the mean for the white is 6.8548. This is also evident in the confidence interval, because the 95% CI for the red wine is between 8.234 and 8.405 while the 95% CI for the white wine is between 6.831 and 6.878. Also the .describe() tells us the count, mean, standard deviation, minimum, 25th, 50th, and 75th percentile, and the max. The min, percentiles, and max values are all higher in the red wine fixed acidity quality.