Descriptive Statistics For pandas Dataframe

chrisalbon.com/python/data_wrangling/pandas_dataframe_descriptive_stats

December 20, 2017

Import modules

```
import pandas as pd
```

Create dataframe

```
data = {'name': ['Jason', 'Molly', 'Tina', 'Jake', 'Amy'],
        'age': [42, 52, 36, 24, 73],
        'preTestScore': [4, 24, 31, 2, 3],
        'postTestScore': [25, 94, 57, 62, 70]}
df = pd.DataFrame(data, columns = ['name', 'age', 'preTestScore', 'postTestScore'])
df
```

0	Jason	42	4	25
1	Molly	52	24	94
2	Tina	36	31	57
3	Jake	24	2	62
4	Amy	73	3	70

5 rows × 4 columns

The sum of all the ages

```
df['age'].sum()
```

```
227
```

Mean preTestScore

```
df['preTestScore'].mean()
```

```
12.800000000000001
```

Cumulative sum of preTestScores, moving from the rows from the top

```
df['preTestScore'].cumsum()
```

```
0     4
1    28
2    59
3    61
4    64
Name: preTestScore, dtype: int64
```

Summary statistics on preTestScore

```
df['preTestScore'].describe()
```

```
count     5.000000
mean     12.800000
std      13.663821
min       2.000000
25%       3.000000
50%       4.000000
75%      24.000000
max      31.000000
Name: preTestScore, dtype: float64
```

Count the number of non-NA values

```
df['preTestScore'].count()
```

```
5
```

Minimum value of preTestScore

```
df['preTestScore'].min()
```

```
2
```

Maximum value of preTestScore

```
df['preTestScore'].max()
```

```
31
```

Median value of preTestScore

```
df['preTestScore'].median()
```

df['preTestScore'].median()

4.0

Sample variance of preTestScore values

df['preTestScore'].var()

186.69999999999999

Sample standard deviation of preTestScore values

df['preTestScore'].std()

13.663820841916802

Skewness of preTestScore values

df['preTestScore'].skew()

0.74334524573267591

Kurtosis of preTestScore values

df['preTestScore'].kurt()

-2.4673543738411525

Correlation Matrix Of Values

df.corr()

age	1.000000	-0.105651	0.328852
preTestScore	-0.105651	1.000000	0.378039
postTestScore	0.328852	0.378039	1.000000

3 rows × 3 columns

Covariance Matrix Of Values

df.cov()

age	340.80	-26.65	151.20

preTestScore	-26.65	186.70	128.65
postTestScore	151.20	128.65	620.30

3 rows × 3 columns

Find an error or bug?

Everything on this site is available on GitHub. Head to and submit a suggested change. You can also message me directly on .