

stats_primer

April 11, 2024

0.1 The dataframe

```
[ ]: import pandas as pd

dict_measures = { 'age' : [15,18,25,25,40,55,58,60,80],
                  'height' : [160,162,165,168,172,174,174,174,176]}

df_measures = pd.DataFrame.from_dict(dict_measures)

df_measures
```

```
[ ]:   age  height
0   15     160
1   18     162
2   25     165
3   25     168
4   40     172
5   55     174
6   58     174
7   60     174
8   80     176
```

1 Summary measures

1.1 Position measures

1.1.1 Mean

```
[ ]: df_measures['age'].mean()
```

```
[ ]: 41.77777777777778
```

1.1.2 Median

```
[ ]: df_measures['age'].median()
```

```
[ ]: 40.0
```

1.1.3 Mode

```
[ ]: print('test')
```

```
test
```

1.2 Dispersion measures

1.2.1 Variance

```
[ ]: df_measures['age'].var()
```

```
[ ]: 509.94444444444446
```

1.2.2 Standard Deviation

```
[ ]: df_measures['age'].std()
```

```
[ ]: 22.58194952709895
```

1.2.3 Coefficient of Variation

```
[ ]: df_measures['age'].std() / df_measures['age'].mean() * 100
```

```
[ ]: 54.05253876167302
```

```
[ ]: df_measures['height'].std() / df_measures['height'].mean() * 100
```

```
[ ]: 3.4803552812368785
```

1.3 Shape Measures

1.3.1 Skewness

```
[ ]: df_measures.age.skew()
```

```
[ ]: 0.368108517895537
```

1.3.2 Kurtosis

```
[ ]: df_measures.age.kurtosis()
```

```
[ ]: -1.1344461075421046
```

1.4 Summary of the most useful stats measures

```
[ ]: df_measures.describe()
```

```
[ ]:
      count      age      height
count  9.000000  9.000000
```

mean	41.777778	169.444444
std	22.581950	5.897269
min	15.000000	160.000000
25%	25.000000	165.000000
50%	40.000000	172.000000
75%	58.000000	174.000000
max	80.000000	176.000000