Abalone Short Analysis

August 2, 2020

[48]:
```
import pandas as pd
import numpy as np
import seaborn as sns
pd.set_option('display.max_columns', 500)
pd.set_option('display.max_columns', 500)
pd.set_option('display.width', 1000)

df = pd.read_csv('http://archive.ics.uci.edu/ml/machine-learning-databases/
 ↪abalone/abalone.data')
df.columns = ["Sex", "Length", "Diameter", "Height", "Whole_Weight",␣
 ↪"Shucked_Weight", "Viscera_Weight", "Shell_Weight", "Rings"]
print(df.head())
```

```
  Sex  Length  Diameter  Height  Whole_Weight  Shucked_Weight  Viscera_Weight
Shell_Weight  Rings
0   M   0.350     0.265   0.090        0.2255          0.0995          0.0485
0.070      7
1   F   0.530     0.420   0.135        0.6770          0.2565          0.1415
0.210      9
2   M   0.440     0.365   0.125        0.5160          0.2155          0.1140
0.155     10
3   I   0.330     0.255   0.080        0.2050          0.0895          0.0395
0.055      7
4   I   0.425     0.300   0.095        0.3515          0.1410          0.0775
0.120      8
```

0.1 The following ___ functions calculate _____ from a numpy array(s) or other values provided:

- **np.mean**: the average value
- **np.std** : the standard deviation of values
- **np.var**: the variance of values
- **np.median**: the median value
- **np.argmin**: the index of the minimum value
- **np.argmax**: the index of the maximum value
- **np.percentile**: the nth percentile value of all rows

[49]:
```
print("Data regarding the Length value of the abalone mollusc dataset:")
print("{}: {}".format("mean".rjust(20," "),round(np.mean(df['Length']),4)))
print("{}: {}".format("median".rjust(20," "),round(np.median(df['Length']),4)))
print("{}: {}".format("standard deviation".rjust(20," "),round(np.
 ↪std(df['Length']),4)))
print("{}: {}".format("variance".rjust(20," "),round(np.var(df['Length']),4)))
```

```
Data regarding the Length value of the abalone mollusc dataset:
                mean: 0.524
              median: 0.545
  standard deviation: 0.1201
            variance: 0.0144
```

0.2 What are the average statistics by sex of abalone?

0.3 and which sex has the biggest shell on average?

[50]:
```
gbs = df.groupby("Sex").mean().reset_index().sort_values(by=['Shell_Weight'])
print(gbs)
```

```
  Sex    Length  Diameter    Height  Whole_Weight  Shucked_Weight
Viscera_Weight  Shell_Weight      Rings
1   I  0.427746  0.326494  0.107996      0.431363        0.191035
0.092010      0.128182   7.890462
2   M  0.561460  0.439335  0.151418      0.991772        0.433083
0.215620      0.282056  10.702685
0   F  0.579093  0.454732  0.158011      1.046532        0.446188
0.230689      0.302010  11.129304
```

According to the documentation, "I" stands for infant, so it is no surprise that category has the smallest shell weight. Female abalone were on average the largest on every scale, apparently.

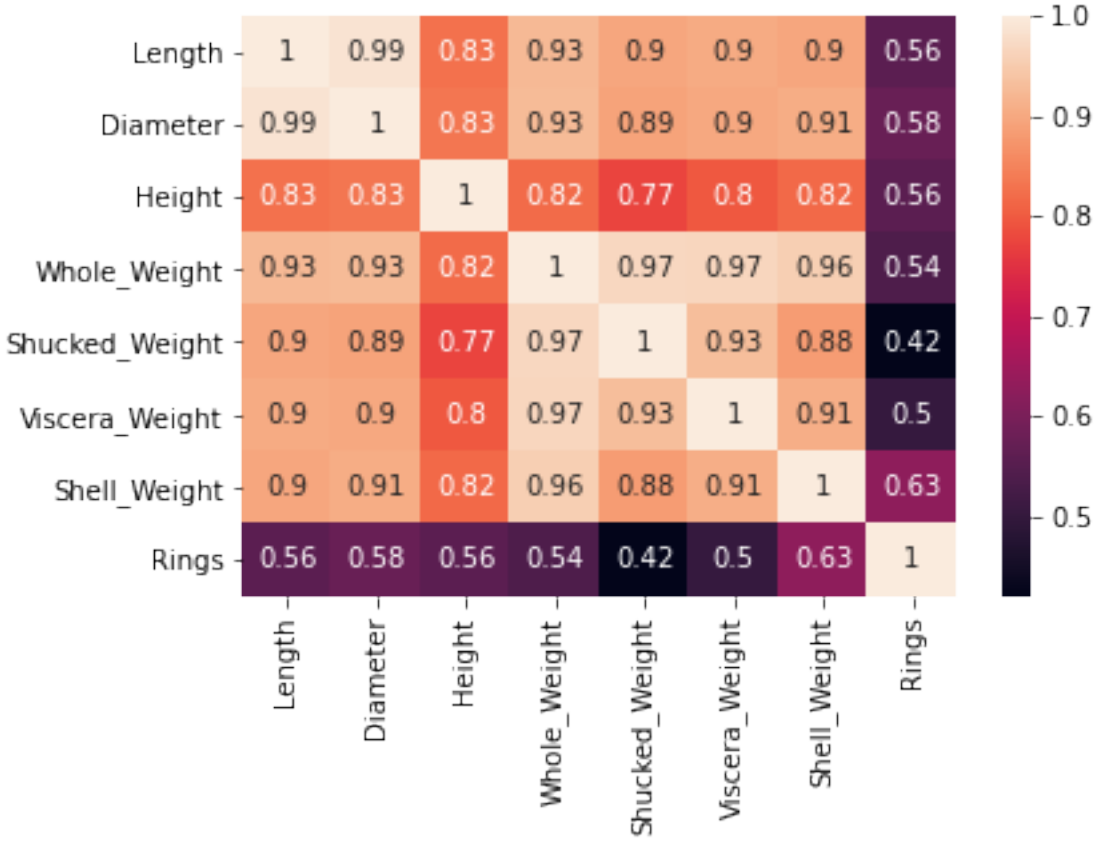
1 How do the different variables relate to each other?

[54]:
```
corr = df.corr()
print(corr)
sns.heatmap(corr, xticklabels=corr.columns, yticklabels=corr.columns,␣
 ↪annot=True)
```

```
                  Length  Diameter    Height  Whole_Weight  Shucked_Weight
Viscera_Weight  Shell_Weight     Rings
Length          1.000000  0.986813  0.827552      0.925255        0.897905
0.903010      0.897697  0.557123
Diameter        0.986813  1.000000  0.833705      0.925452        0.893159
0.899726      0.905328  0.575005
Height          0.827552  0.833705  1.000000      0.819209        0.774957
0.798293      0.817326  0.558109
Whole_Weight    0.925255  0.925452  0.819209      1.000000        0.969403
```

```
0.966372       0.955351   0.540818
Shucked_Weight   0.897905   0.893159   0.774957       0.969403         1.000000
0.931956       0.882606   0.421256
Viscera_Weight   0.903010   0.899726   0.798293       0.966372         0.931956
1.000000       0.907647   0.504274
Shell_Weight     0.897697   0.905328   0.817326       0.955351         0.882606
0.907647       1.000000   0.628031
Rings            0.557123   0.575005   0.558109       0.540818         0.421256
0.504274       0.628031   1.000000
```

[54]: `<matplotlib.axes._subplots.AxesSubplot at 0x1c795a05dc0>`



It's interesting that everything is highly correlated with everything else, except rings, which in turn means age. Apparently abalone stay the same size/weight more or less after they are no longer infants. The only exception is that they seem to lose soft body weight as they age, and in proportion they get slightly bigger shells. Unless there's more to it…

1.1 What are the average values when grouping by rings/age?

[56]:
```
df['age_group'] = pd.cut(df['Rings'], 3, labels=['Immature','Adult',␣
↪'Advanced'])
df.groupby('age_group')['Rings'].mean()
```

[56]:
```
age_group
Immature     8.141758
Adult       12.945087
Advanced    21.532258
Name: Rings, dtype: float64
```

Wow there are pretty wide age ranges! OK let's find those average values.

[58]:
```
gbs_age = df.groupby(["Sex", "age_group"]).mean().reset_index().
↪sort_values(by=['Shell_Weight'])
print(gbs_age)
```

```
  Sex age_group    Length  Diameter    Height  Whole_Weight  Shucked_Weight
Viscera_Weight  Shell_Weight      Rings
3   I  Immature  0.414372  0.315333  0.103231      0.387357        0.174700
0.082670      0.114271   7.225779
6   M  Immature  0.534437  0.415798  0.140930      0.855129        0.389885
0.188452      0.233082   8.757750
4   I     Adult  0.529836  0.411678  0.144046      0.764618        0.315283
0.163099      0.234184  12.835526
0   F  Immature  0.559650  0.436607  0.149888      0.937981        0.420732
0.208247      0.259687   8.961310
5   I  Advanced  0.546667  0.426667  0.166667      0.958000        0.359000
0.185833      0.261667  20.333333
1   F     Adult  0.599104  0.473483  0.165978      1.153973        0.474038
0.254287      0.341458  12.981758
7   M     Adult  0.596614  0.469658  0.164817      1.168459        0.491892
0.251701      0.343052  12.936407
8   M  Advanced  0.614259  0.492222  0.177593      1.283611        0.456574
0.251463      0.440926  21.407407
2   F  Advanced  0.610313  0.482031  0.178437      1.301500        0.455953
0.257281      0.447437  21.750000
```

OK this gives a clearer overall picture.

Size and weight both change dramatically over an abalone's lifetime. Withr more time and effort, we could understand more about what differentiates a more advanced abalone than just "bigger".