03_anova_test

March 31, 2022

1 ANOVA

Analysis of Variance

[ ]:
```
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
```

[ ]:
```
kashti = sns.load_dataset('titanic')
kashti.head()
```

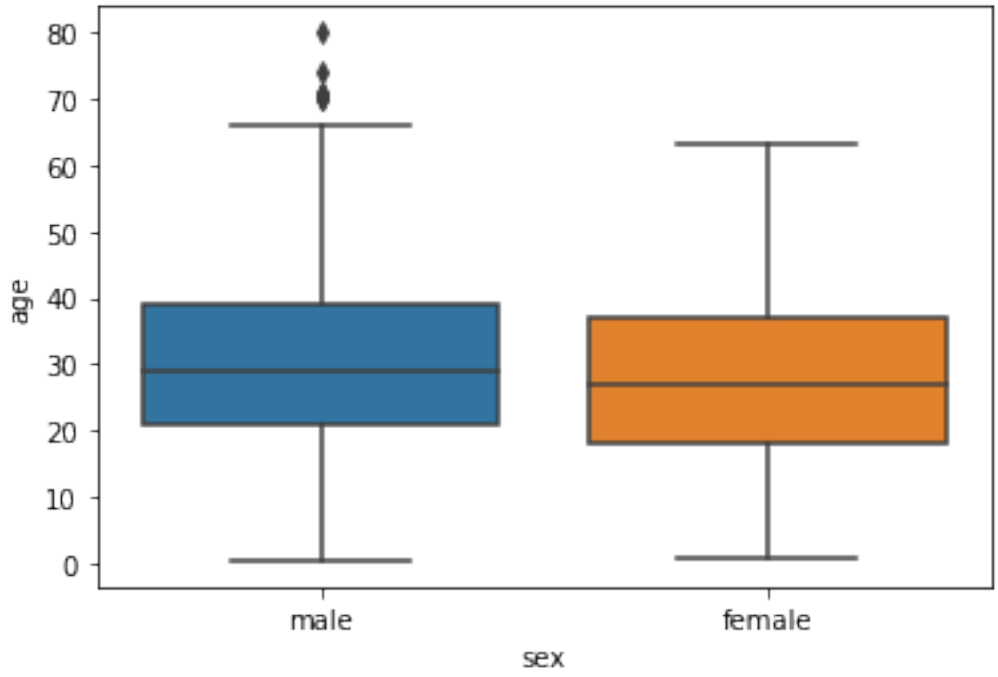
[ ]:

	survived	pclass	sex	age	sibsp	parch	fare	embarked	class	\
0	0	3	male	22.0	1	0	7.2500	S	Third	
1	1	1	female	38.0	1	0	71.2833	C	First	
2	1	3	female	26.0	0	0	7.9250	S	Third	
3	1	1	female	35.0	1	0	53.1000	S	First	
4	0	3	male	35.0	0	0	8.0500	S	Third	

	who	adult_male	deck	embark_town	alive	alone
0	man	True	NaN	Southampton	no	False
1	woman	False	C	Cherbourg	yes	False
2	woman	False	NaN	Southampton	yes	True
3	woman	False	C	Southampton	yes	False
4	man	True	NaN	Southampton	no	True

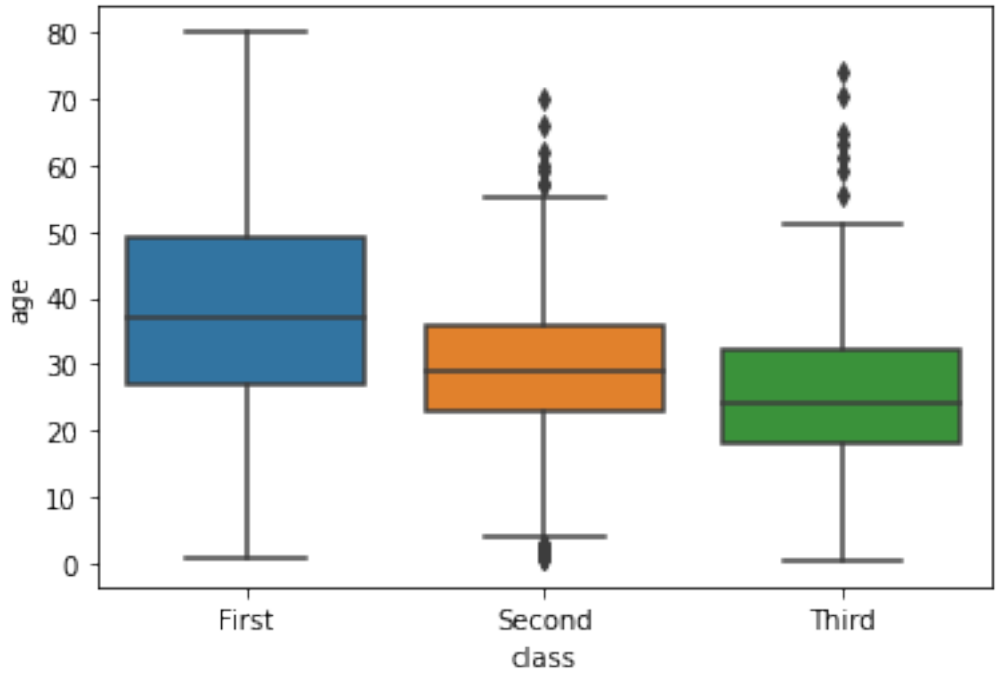
[ ]:
```
sns.boxplot(x='sex', y='age',data=kashti) #comparision betwwn two categorical␣
↪variable for one continous
```

[ ]:
```
<AxesSubplot:xlabel='sex', ylabel='age'>
```



[ ]:
```
sns.boxplot(x='class', y='age',data=kashti) #ANOVA
```

[ ]:
```
<AxesSubplot:xlabel='class', ylabel='age'>
```



[ ]:
```
#example
phool = sns.load_dataset('iris')
phool.head()
```

[ ]:

	sepal_length	sepal_width	petal_length	petal_width	species
0	5.1	3.5	1.4	0.2	setosa
1	4.9	3.0	1.4	0.2	setosa
2	4.7	3.2	1.3	0.2	setosa
3	4.6	3.1	1.5	0.2	setosa
4	5.0	3.6	1.4	0.2	setosa

[ ]:
```
phool.sample(30)
phool.columns
```

[ ]:
```
Index(['sepal_length', 'sepal_width', 'petal_length', 'petal_width',
       'species'],
      dtype='object')
```

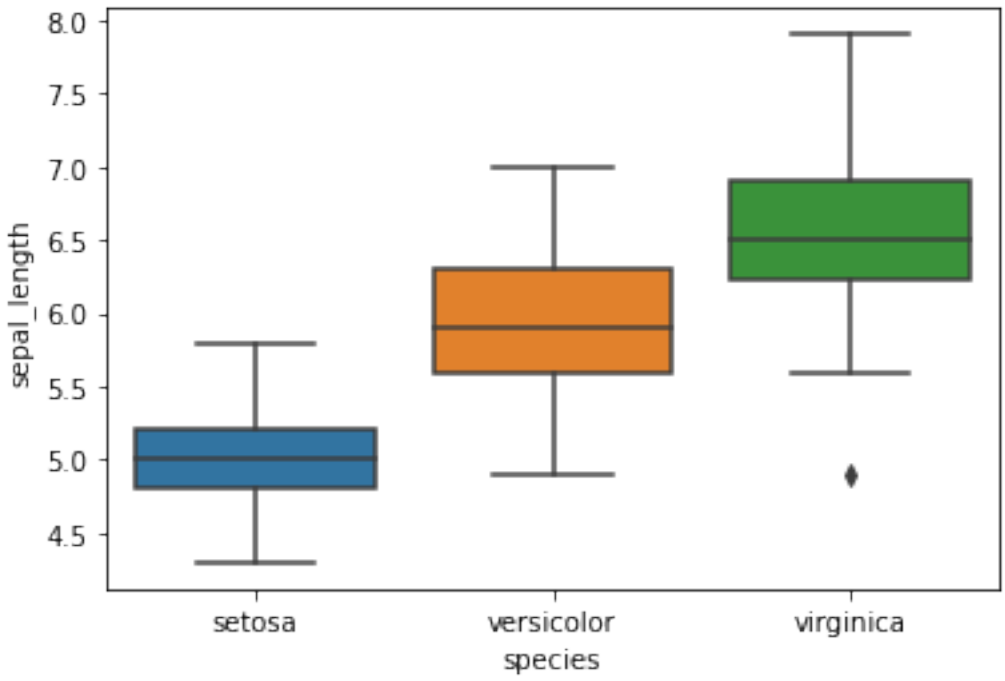
[ ]:
```
phool.describe()
```

[ ]:

	sepal_length	sepal_width	petal_length	petal_width
count	150.000000	150.000000	150.000000	150.000000
mean	5.843333	3.057333	3.758000	1.199333
std	0.828066	0.435866	1.765298	0.762238
min	4.300000	2.000000	1.000000	0.100000
25%	5.100000	2.800000	1.600000	0.300000
50%	5.800000	3.000000	4.350000	1.300000
75%	6.400000	3.300000	5.100000	1.800000
max	7.900000	4.400000	6.900000	2.500000

[ ]:
```
sns.boxplot(x ='species', y='sepal_length',data =phool)
```

[ ]:
```
<AxesSubplot:xlabel='species', ylabel='sepal_length'>
```



[ ]:
```
#stat
import statsmodels.api as sm
from statsmodels.formula.api import ols
```

[ ]:
```
# one way ANOVA

mod = ols('sepal_length ~ species', data=phool).fit()

aov_table = sm.stats.anova_lm(mod, type=2) #assignment why type = 2
print(aov_table)
```

```
                df      sum_sq    mean_sq           F        PR(>F)
species        2.0   63.212133  31.606067  119.264502  1.669669e-31
Residual     147.0   38.956200   0.265008         NaN           NaN
```

[ ]:
```
#pairwise comparision          #Assignment ANOVA tabel read kse krte hn
pair_t = mod.t_test_pairwise('species', method ='bonferroni') #sidak
pair_t.result_frame

pair_t = mod.t_test_pairwise('species', method ='sidak') #sidak
```

[ ]:
```
# tukey test hsd test
import pingouin as pg
aov = pg.anova(data= phool, dv = 'sepal_length', between ='species',␣
 ↪detailed=True)
```

```
print(aov)
```

```
    Source          SS   DF         MS           F         p-unc       np2
0  species   63.212133    2  31.606067  119.264502  1.669669e-31  0.618706
1   Within   38.956200  147   0.265008         NaN           NaN       NaN
```

[ ]:
```
# tukey HSD
pt = (data= phool, dv= 'sepal_length', between='species')
print(pt)
```

```
            A           B  mean(A)  mean(B)   diff        se          T  \
0      setosa  versicolor    5.006    5.936 -0.930  0.102958  -9.032819
1      setosa   virginica    5.006    6.588 -1.582  0.102958 -15.365506
2  versicolor   virginica    5.936    6.588 -0.652  0.102958  -6.332686

        p-tukey    hedges
0  2.420286e-14 -1.792703
1  2.153833e-14 -3.049522
2  8.287554e-09 -1.256820
```