02_ttestsAssignemnt

April 1, 2022

[ ]:
```
# one sample t-test
# import libraries

import seaborn as sns
import pandas as pd
from scipy.stats import ttest_1samp

# load dataset

df = sns.load_dataset('penguins')
df
```

[ ]:
```
     species     island  bill_length_mm  bill_depth_mm  flipper_length_mm  \
0     Adelie  Torgersen            39.1           18.7              181.0
1     Adelie  Torgersen            39.5           17.4              186.0
2     Adelie  Torgersen            40.3           18.0              195.0
3     Adelie  Torgersen             NaN            NaN                NaN
4     Adelie  Torgersen            36.7           19.3              193.0
..       ...        ...             ...            ...                ...
339   Gentoo     Biscoe             NaN            NaN                NaN
340   Gentoo     Biscoe            46.8           14.3              215.0
341   Gentoo     Biscoe            50.4           15.7              222.0
342   Gentoo     Biscoe            45.2           14.8              212.0
343   Gentoo     Biscoe            49.9           16.1              213.0

     body_mass_g     sex
0         3750.0    Male
1         3800.0  Female
2         3250.0  Female
3            NaN     NaN
4         3450.0  Female
..           ...     ...
339          NaN     NaN
340       4850.0  Female
341       5750.0    Male
342       5200.0  Female
343       5400.0    Male
```

```
[344 rows x 7 columns]
```

[ ]:
```
df1 = df[['species', 'bill_length_mm','bill_depth_mm']]
df1.head()
```

[ ]:
```
  species  bill_length_mm  bill_depth_mm
0  Adelie            39.1           18.7
1  Adelie            39.5           17.4
2  Adelie            40.3           18.0
3  Adelie             NaN            NaN
4  Adelie            36.7           19.3
```

[ ]:
```
df1.describe()
```

[ ]:
```
       bill_length_mm  bill_depth_mm
count      342.000000     342.000000
mean        43.921930      17.151170
std          5.459584       1.974793
min         32.100000      13.100000
25%         39.225000      15.600000
50%         44.450000      17.300000
75%         48.500000      18.700000
max         59.600000      21.500000
```

[ ]:
```
# check the age and compare witht a known value of 45 years

ttest_1samp(df1['bill_length_mm'], 50)

stat, p = ttest_1samp(df1['bill_length_mm'], 50)

print('stat=%.3f, p=%.3f' % (stat, p))

# make a conditional arguement for ease
if p > 0.05:
    print('Probably the same distribution')
else:
    print('Probably different Distribution')
```

```
stat=nan, p=nan
Probably different Distribution
```

[ ]:
```
# we will compare

#splitting dataset
df_G= df1.loc[df1['species']== 'Gentoo']
df_A= df1.loc[df1['species']== 'Adeli']
```

```
# library
from scipy.stats import ttest_ind
stat, p = ttest_ind(df_G['bill_length_mm'], df_A['bill_length_mm'])

print('stat=%.3f, p=%.3f' % (stat, p))

# make a conditional arguement for ease
if p > 0.05:
    print('Probably the same distribution')
else:
    print('Probably different Distribution')
```

```
stat=nan, p=nan
Probably different Distribution
```

[ ]:
```
df_A.describe
```

[ ]:
```
<bound method NDFrame.describe of Empty DataFrame
Columns: [species, bill_length_mm, bill_depth_mm]
Index: []>
```

[ ]:
```
df_G.info
```

[ ]:
```
<bound method DataFrame.info of     species  bill_length_mm  bill_depth_mm
220  Gentoo            46.1           13.2
221  Gentoo            50.0           16.3
222  Gentoo            48.7           14.1
223  Gentoo            50.0           15.2
224  Gentoo            47.6           14.5
..      …               …             …
339  Gentoo             NaN            NaN
340  Gentoo            46.8           14.3
341  Gentoo            50.4           15.7
342  Gentoo            45.2           14.8
343  Gentoo            49.9           16.1

[124 rows x 3 columns]>
```

[ ]:
```
df_G= df1.loc[df1['species']== 'Gentoo']
df_G.head()
```

[ ]:
```
    species  bill_length_mm  bill_depth_mm
220  Gentoo            46.1           13.2
221  Gentoo            50.0           16.3
222  Gentoo            48.7           14.1
223  Gentoo            50.0           15.2
```

```
224  Gentoo            47.6          14.5
```