

## Portfolio assignment 8

15 min:

- Calculate the 90%, 95%, 99% and 99.99% confidence interval for at least 2 columns with numerical data in the dataset that you chose in portfolio assignment 4. Do you see the impact the confidence has on the interval?
- Now calculate the 95% confidence interval again but use only the first 10% of your rows. Compare this interval to the previous 95% confidence interval you calculated. Do you see the impact of having less data?

In [1]:

```
import pandas as pd
import scipy.stats as st
pokemons = pd.read_csv('../Pokemon.csv')
```

In [2]:

```
confidence = 0.90
st.t.interval(confidence, len(pokemons)-1, loc=pokemons['Attack'].mean(), scale=st.sem(poke
```

Out[2]:

```
(77.11152171030228, 80.89097828969771)
```

In [3]:

```
confidence = 0.95
st.t.interval(confidence, len(pokemons)-1, loc=pokemons['Attack'].mean(), scale=st.sem(poke
```

Out[3]:

```
(76.74869843886454, 81.25380156113546)
```

In [4]:

```
confidence = 0.99
st.t.interval(confidence, len(pokemons)-1, loc=pokemons['Attack'].mean(), scale=st.sem(poke
```

Out[4]:

```
(76.03830265303164, 81.96419734696836)
```

In [5]:

```
confidence = 0.90
st.t.interval(confidence, len(pokemons)-1, loc=pokemons['Defense'].mean(), scale=st.sem(pok
```

Out[5]:

```
(72.02693852111383, 75.65806147888617)
```

In [6]:

```
confidence = 0.95  
st.t.interval(confidence, len(pokemons)-1, loc=pokemons['Defense'].mean(), scale=st.sem(pok
```

Out[6]:

```
(71.67835509837681, 76.00664490162319)
```

The confidence interval gets lower when the level get higher

In [7]:

```
confidence = 0.95  
st.t.interval(confidence, len(pokemons)/10, loc=pokemons['Attack'].mean(), scale=st.sem(pok
```

Out[7]:

```
(76.71757028293392, 81.28492971706608)
```

In [8]:

```
st.t.interval(confidence, len(pokemons)/10, loc=pokemons['Defense'].mean(), scale=st.sem(po
```

Out[8]:

```
(71.64844863980046, 76.03655136019954)
```

The confidence interval get a lot closer when the amount of data is smalled down