TUTORIAL 4 EXERCISE

PART 1

A biomaterials engineer has proposed two different topographies for a stent. Each of these two topographies have been investigated for their angiogenesis (blood vessel formation) potential. Angiogenesis activity was measured using a fluorescence-based assays (a.u.) where higher values suggest greater amounts of blood vessel formation. The engineer wants to decide if there is a differnce between the two different topographies.

What is the appropriate statistical test for this data and why?

Importing relevant libraries!

In [57]:
```
import pandas as pd #library to work with data frames
import numpy as np #library to work with data frames
import matplotlib.pyplot as plt #library to plot figure
import matplotlib.dates as mdates #library for visualization
import seaborn as sns #library to plot figures
import scipy
from scipy import stats
from scipy.stats import t
import statsmodels.api as sm
```

Reading the data and generating descriptive statistics.

In [60]:
```
#Creating dataset
topographyA = np.array([15.8,17.3,15.7,16.9,18.5,17.3,16.5,18.1])
topographyA_df = pd.DataFrame(topographyA)
topographyB = np.array([18.5,19.4,19.3,20.1,19.6,19,18.5,18])
topographyB_df = pd.DataFrame(topographyB)
df=pd.concat([topographyA_df,topographyB_df],axis="columns")
df.columns=["topographyA","topographyB"]
```

In [62]:
```
#Generating Descriptive Statistics for both Datasets
```

In [64]:
```
# Generating Plots for both Datasets
```

PART 2

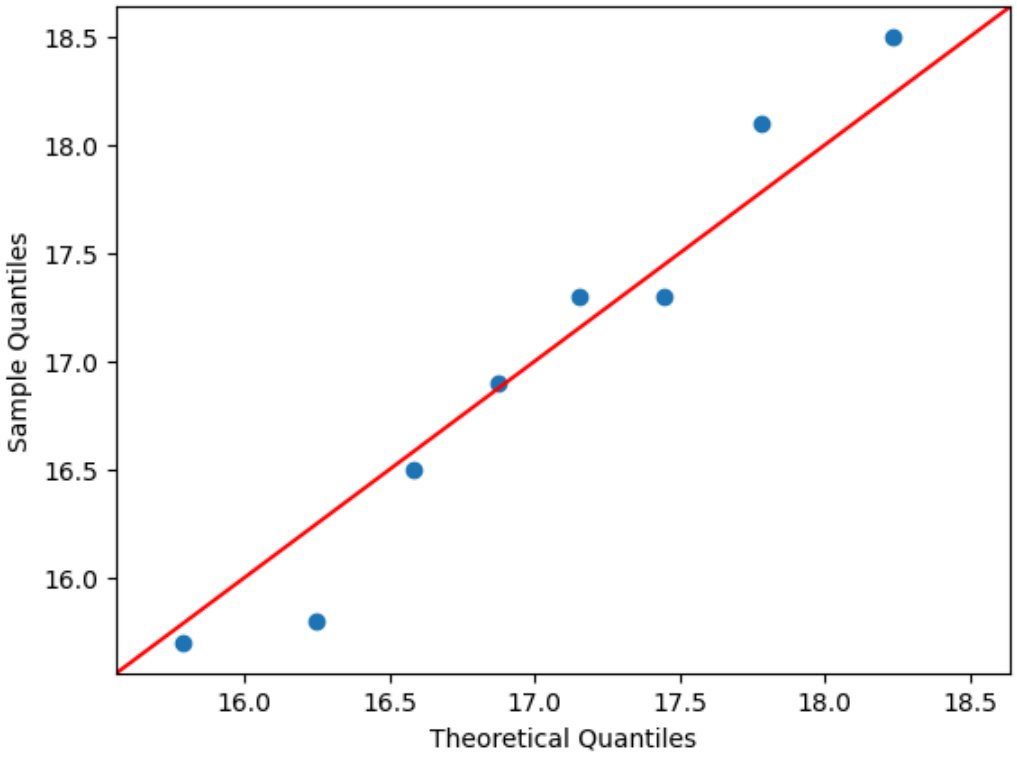
What assumptions are you making by choosing this test? Justify why they are acceptable.

Think thorugh how to demonstrate normally distributed. Use qq plot and shapiro Wilks test.

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.shapiro.html

In [68]:
```
print(sm.qqplot(topographyA,line='45',loc=np.mean(topographyA),scale=np.std(topographyA,ddof=1
```

```
Figure(640x480)
```

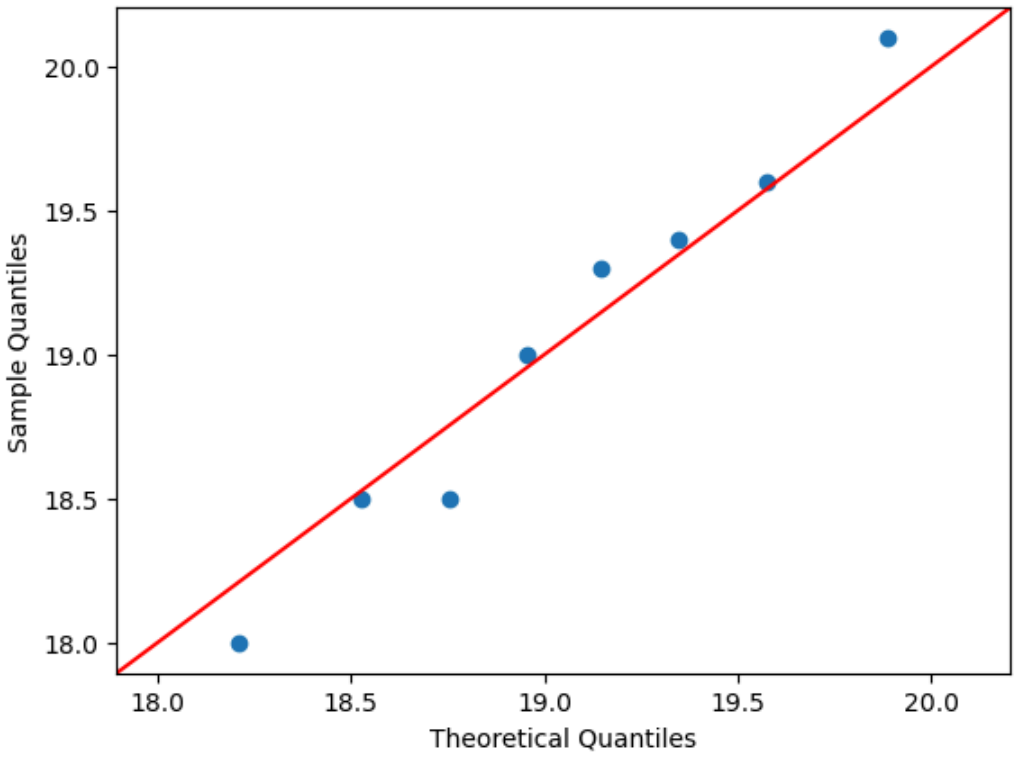


To use the student's t-test, you must assume that the variances are equal between the groups. Think thorugh how to demonstrate equal variances, levene's test

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.levene.html

In [82]:
```
print(sm.qqplot(topographyB,line='45',loc=np.mean(topographyB),scale=np.std(topographyB,ddof=1
```

```
Figure(640x480)
```



PART 3: Provide the 7 steps of the Procedures for Hypothesis Tests

1. Paramater of Interest

2. State the null hypothesis

3. State the null hypothesis

4. Determine appropriate test statistic

5. State the rejection criteria for null hypothesis

6. Computations

7. Draw Conclusions

In [112…
```
# Step 1: parameter of interest is difference in means
```

Step 2: state the null hypothesis

Step 3: state the alternative hypothesis

Step 4: test statistic is:

The t statistic to test whether the means are different can be calculated as follows:

$$t = \frac{\bar{X}_1 - \bar{X}_2}{s_p \cdot \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

where, with equal varaince,

$$s_p = \sqrt{\frac{(n_1 - 1)\, s_{X_1}^2 + (n_2 - 1)\, s_{X_2}^2}{n_1 + n_2 - 2}}$$

Step 5: Reject Ho if .....

The critical values at this alhpa value are t=.....

n − 1 is the number of degrees of freedom for each group, and the total sample size minus two (that is, n1 + n2 − 2) is the total number of degrees of freedom, which is used in significance testing.

14 dof, n1+n2-2 = ...

In [120…
```
# critical value of t at dof=14 and at alpha/2 = 0.025
stats.t.ppf(0.025,14)
```

Out[120… `-2.1447866879169277`

In [122…
```
# critical value of t at dof=14 and at alpha/2 = 0.975
stats.t.ppf(0.975,14)
```

Out[122… 2.1447866879169273

Step 6 Computations

We will calculate a pooled estimate of the combined standard deviations, which is a weighted average of the two standard deviations, used when the variances are equivalent to each other

$$s_p = \sqrt{\frac{(n_1 - 1)\, s_{X_1}^2 + (n_2 - 1)\, s_{X_2}^2}{n_1 + n_2 - 2}}$$

In [136…
```
# find the s_pooled value
```

In [138…
```
# find the t-statistic
```

Step 7: Conclusions

find the p-value of the t-statistic, make a conclusion, and put it into words using the context of the problem.

In [ ]:
```

```

PART 4

Use python to validate the hyptheisis test

https://docs.scipy.org/doc/scipy/reference/generated/scipy.stats.ttest_ind.html

e.g.: stats.ttest_ind(topographyA, topographyB, equal_var=True, alternative="two-sided))

PART 5

Compute the corresponding confidence interval for the difference of means.

Make a conclusion based on the CI, and put into words in the context of the problem.

In [145…
```
cm = sm.stats.CompareMeans(sm.stats.DescrStatsW(topographyA), sm.stats.DescrStatsW(topographyB
print("The 95% difference in mean CI is:", cm.tconfint_diff(usevar='pooled'))
```

```
The 95% difference in mean CI is: (-2.9586152548128006, -1.1163847451872022)
```

In [ ]:
```

```

In [ ]:
```

```