

Python seminar Homework for Chap. 5.4

1. Please plot a box-and-whisker diagram with testing significant differences between the data. It is acceptable to use your own research data or any data you are interested in.

(1) Make a box-and-whisker diagram of the data of several groups (preferably three or more groups).

(2) After creating the box-and-whisker diagram, determine whether there is a significant difference in the means between the groups using ANOVA or other methods. If there is a significant difference between the groups, please do multiple comparisons if you are able to do.

When selecting a test method, it is necessary to check the distribution for normality.

For the case, you do not have any good data, an example data is listed in below.

Body length of stag beetles captured at 3 stations (Kurume, Kimitsu, Otaki).

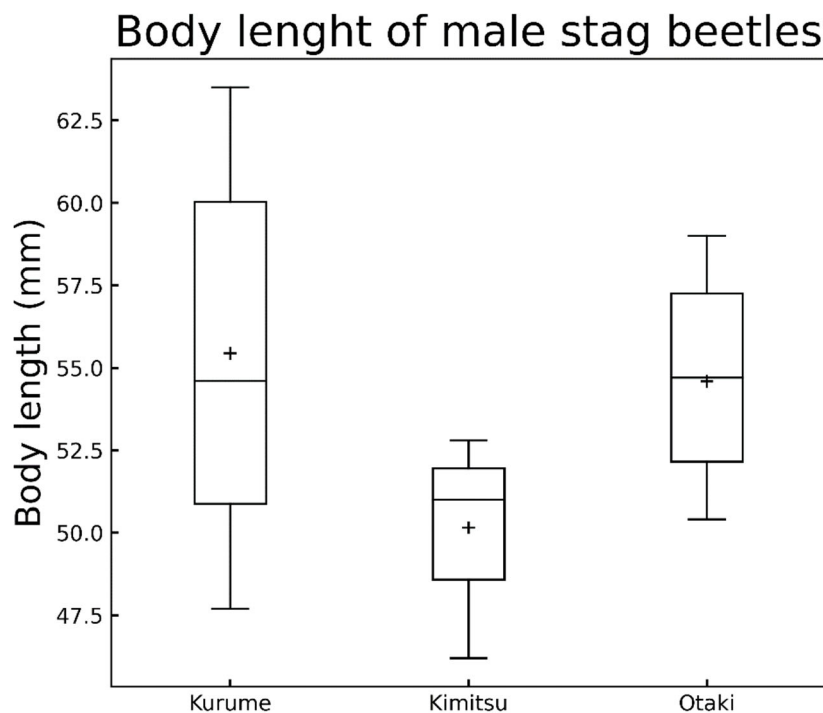
```
kurume = np.array([52.3,60.4,47.7,48.7,58.9,54.2,63.5,55.0,63.3,50.4])
```

```
kimitsu = np.array([49.7,51.3,46.2,52.8,52.1,51.8,46.7,52.0,48.2,50.7])
```

```
otaki = np.array([57.9,50.4,53.2,51.8,55.3,59.0,57.7,50.6,55.9,54.1])
```

```
BLdata = pd.DataFrame({'kurume':kurume, 'kimitsu':kimitsu, 'otaki':otaki})
```

```
print(BLdata)
```



normal distribution test for kurume, $p = 0.394291$

The null hypothesis cannot be rejected: normal distribution (D'Agostino and Pearson's test)

normal distribution test for kimitsu, $p = 0.423392$

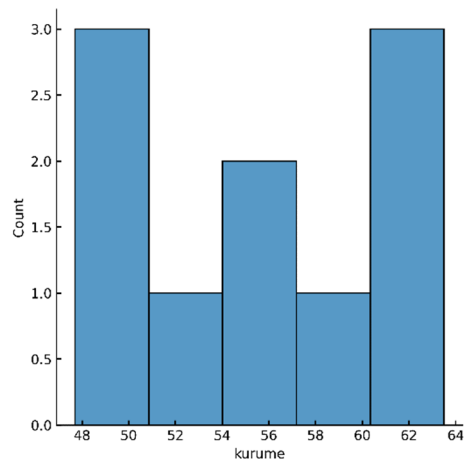
The null hypothesis cannot be rejected: normal distribution (D'Agostino and Pearson's test)

normal distribution test for otaki, $p = 0.456277$

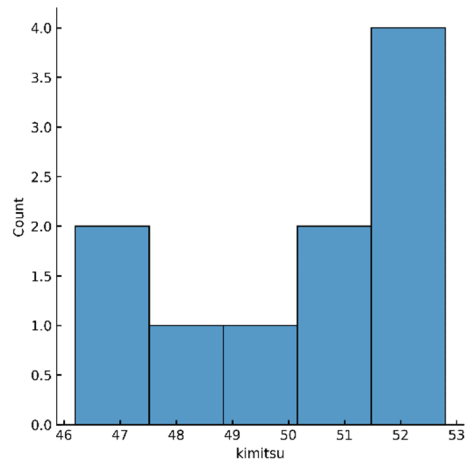
The null hypothesis cannot be rejected: normal distribution (D'Agostino and Pearson's test)

Normal distribution test cannot reject the normal distribution. However, the data distribution does not seem like a normal distribution.

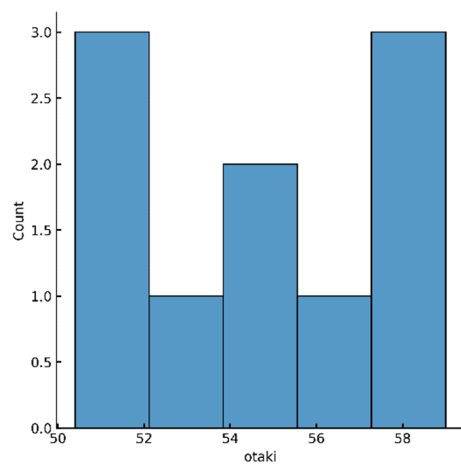
Kurume



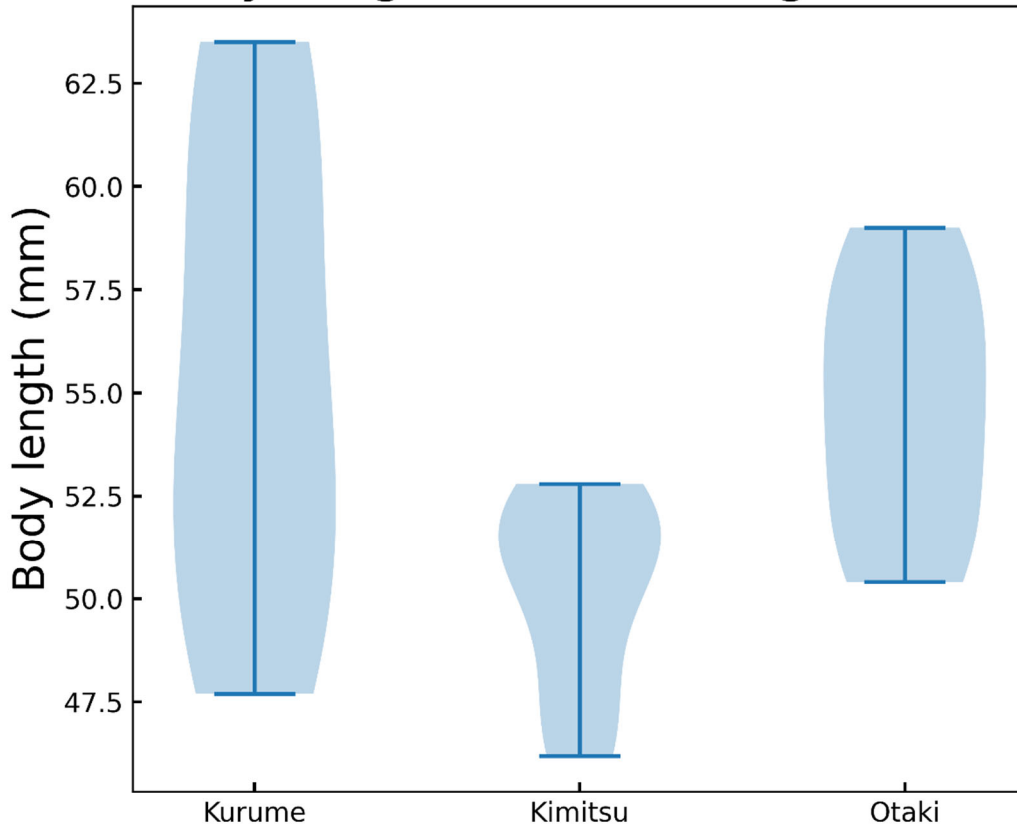
Kimitsu



Otaki



Body length of male stag beetles



In addition, Bartlett's test for homoscedasticity showed that kurume & kimitsu have different variance.

homoscedasticity test for kurume & kimitsu, $p = 0.0128705$

The null hypothesis can be rejected: variance is different (Bartlett's test)

homoscedasticity test for kurume & otaki, $p = 0.0720795$

The null hypothesis cannot be rejected: variation is not different (Bartlett's test)

homoscedasticity test for kimitsu & otaki, $p = 0.435696$

The null hypothesis cannot be rejected: variation is not different (Bartlett's test)

Therefore, non-parametric Kruskal-Wallis test was performed.

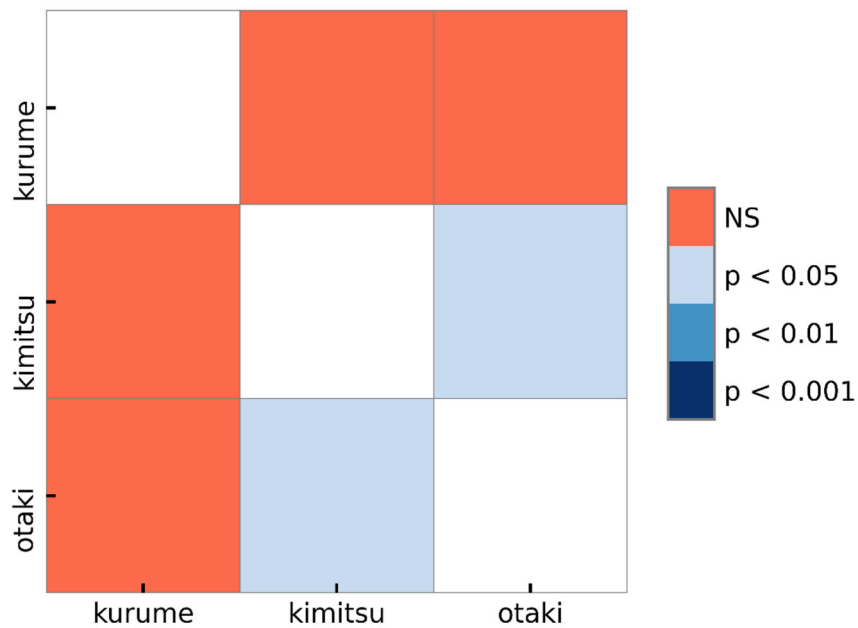
Kruskal-Wallis test, $p = 0.0227623$

Kruskal-Wallis test was true. Therefore, Steel-Dwass test was performed.

Steel-Dwass test

	kurume	kimitsu	otaki
kurume	1.000000	0.086558	0.900000

kimitsu	0.086558	1.000000	0.024645
otaki	0.900000	0.024645	1.000000



* Normal distribution test and Kruskal-Wallis test can be conducted using scipy. However, Steel-Dwass test is not prepared in scipy. There are several ways including scikit_posthocs and Rpy2. If you want to use scikit_posthocs, it can be installed by the command “pip install scikit_posthocs” from anaconda prompt.

Tips for multiple comparison

Frequently used methods for multiple comparison.

Case	A priori comparison	Post hoc comparison
Normal, homoscedasticity	ANOVA one way	Tukey-Kramer test
Normal, homoscedasticity	ANOVA one way	Steel-Dwass or Games-Howel test
Normal, heteroscedasticity	Welch t-test	Games-Howel test
Non-normal, heteroscedasticity	Kruskal-Wallis test	Steel-Dwass test