

# Statistik un Data Science für die Informatik

Humboldt-Universität zu Berlin, SS 22, Prof. Dr. Alan Akbik

## Bearbeitung von Übungsblatt 2

### Namen und Matrikelnummer der Abgebenden:

Gruppe AG-24	Name	CMS-Benutzername	Matrikelnr
	Kristina Pianykh	pianykhk	617331
	Miguel Nuno Carqueijeiro Athouguia	carqueim	618203
	Winston Winston	winstonw	602307

### Aufgabe 3a)

2)

The output of the exact permutation test performed on the datasets variante\_a.txt and variante\_b.txt:

Differenz zwischen den Mittelwerten: -2.6

P-Wert (p=0.381)

The difference of 2.6 (with the mean in variante\_b.txt greater than that in variante\_a.txt) is not statistically significant as the p-value (p=0.381) is above the set alpha level of 0.05. Hence, the null hypothesis of no difference cannot be rejected.

3)

The output of the approximative permutation test for independant samples with the assumption of equal variance in the samples carried out using scipy.stats.ttest\_ind is as follows:

```
Ttest_indResult(statistic=0.9630818692995163,  
pvalue=0.3637101975769613)
```

The results indicate that the difference in the means of the samples is not significant based on the p-value of roughly 0.364. Therefore, the null hypothesis of no difference cannot be rejected.

### Aufgabe 3b)

2)

The output of the program permutationtest.py for an approximative t-test using the data from gmbh.txt and konkurrent.txt with 20 000 unqiue permutations is as follows:

Differenz zwischen den Mittelwerten: -1.3763

P-Wert (p=0.0122)

The p-value is below the significance level of 0.05 and thus we can reject the null hypothesis of no difference between the means.

3)

We now carry out the approximative permutation test from 2) 10 times. The results are below:

1.  $p=0.0109$
2.  $p=0.0115$
3.  $p=0.0112$
4.  $p=0.0114$
5.  $p=0.011$
6.  $p=0.0111$
7.  $p=0.0109$
8.  $p=0.0109$
9.  $p=0.0112$
10.  $p=0.0097$

Each of the tests delivers the p-value below the significance level, meaning the difference between the means of the given datasets is statistically significant.

### 5)

The results of the approximative permutation test from 2) for the following number of permutations:

```
n = 10: p = 0.0
n = 1000: p = 0.007
n = 10000: p = 0.011
n = 100000: p = 0.0115
```

In the first case with  $n=10$ , the number of permutations is too small to yield any mean differences whose absolute value is greater than the one observed in the actual datasets. Thus the extremely small p-value.

### 6)

The output of the approximative permutation test for independant samples with 20000 permutations and the assumption of equal variance in the samples carried out using `scipy.stats.ttest_ind` is as follows:

```
Ttest_indResult(statistic=2.559682021137714, pvalue=0.011)
```

The p-value yielded by the test is 0.011 and below the significance level of 0.05. With that, the difference in the means of both products is significant, and the null hypothesis can be rejected.