CS3-mid-p1

In [12]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [13]:
```
csv_in = 'mid-p1.csv'
df = pd.read_csv(csv_in, sep=',', skiprows=4, header=0)  # (1)
```

In [14]:
```
print( df.shape )  # (2)
print( df.info() )  # (3)
display( df.head() )  # (4)
```

```
(108, 4)
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 108 entries, 0 to 107
Data columns (total 4 columns):
 #   Column  Non-Null Count  Dtype  
---  ------  --------------  -----  
 0   d1      108 non-null    float64
 1   d2      108 non-null    int64  
 2   d3      108 non-null    object 
 3   d4      108 non-null    object 
dtypes: float64(1), int64(1), object(2)
memory usage: 3.5+ KB
None
```

	d1	d2	d3	d4
0	1.75	22	S	j
1	-0.29	0	D	t
2	-0.48	19	S	p
3	-2.65	19	S	p
4	-0.01	20	D	e

(5) 108

In [15]:
```
print( df['d1'].head() )  # (6)
print( type(df['d1']) )
```

```
0    1.75
1   -0.29
2   -0.48
3   -2.65
4   -0.01
Name: d1, dtype: float64
<class 'pandas.core.series.Series'>
```

(7) Series

In [16]:
```
print( df['d1'].min() )  # (8)
```

```
-2.65
```

(9) -2.65

In [17]:
```
display( df.sort_values(by='d2', ascending=False).head() )  # (10)
```

	d1	d2	d3	d4
0	1.75	22	S	j
4	-0.01	20	D	e
2	-0.48	19	S	p
3	-2.65	19	S	p
106	-0.05	19	N	j

(11) 20

In [18]:
```
print( df['d3'].value_counts() )  # (12)
```

```
N    29
D    28
H    27
S    24
Name: d3, dtype: int64
```

(13) 27

In [19]:
```
df2 = df.drop(columns='d4')  # (14)
display( df2.groupby('d3').max() )  # (15)
```

	d1	d2
d3		
D	1.79	20
H	1.57	16
N	2.16	19
S	1.75	22

(16) 1.79

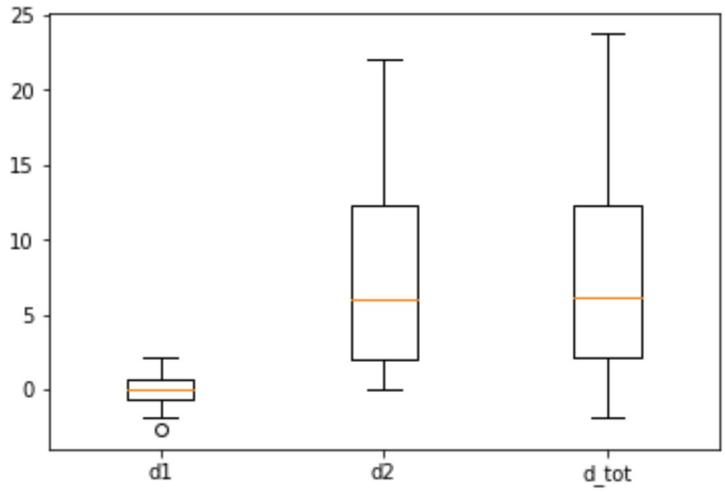
In [20]:
```
df['d_tot'] = df['d1'] + df['d2']  # (17)
display( df.head() )
```

	d1	d2	d3	d4	d_tot
0	1.75	22	S	j	23.75
1	-0.29	0	D	t	-0.29
2	-0.48	19	S	p	18.52
3	-2.65	19	S	p	16.35
4	-0.01	20	D	e	19.99

(18) 23.75

In [21]:

```
df3 = df[ ['d1', 'd2', 'd_tot'] ]  # (19)
plt.boxplot( df3, labels=df3.columns )  # (20)
plt.show()
```



In [ ]: