Solutions 1

Jumping Rivers

Question 1

This question contains some maths equations, in particular two summations. If you're unsure about summations either ask your presenter or skip this question. We're not bothered about your maths theory skills, just your python skills!

```
total = 0
for i in range(1,6):
    total = total + i
total
```

The `for` loop above calculates

$$\sum_{i=1}^{5} i = 1 + 2 + 3 + 4 + 5$$

1. What is the final value of `total` in the above piece of code?

```
print(total)
```

```
## 15
```

2. Change the above loop to calculate the following summations:

$$\text{(i)} \sum_{i=1}^{20} (i + 1)$$

```
total = 0
for i in range(1,21):
    total = total + (i + 1)
total
```

```
## 230
```

$$\text{(ii)} \sum_{j=-5}^{15} j$$

```
total = 0
for j in range(-5,16):
    total = total + j
total
```

```
## 105
```

3. Rewrite the two loops using the `sum()` function from the **numpy** library and the `range()` function. For example, the for loop in the first example can be written as `np.sum(range(1,6))`

```python
import numpy as np

np.sum(range(2,22))

## 230

np.sum(range(-5,16))

## 105
```

Question 2

In the notes, we observed that it was straight forward to loop through a data set and select the maximum values. For instance, the maximum value of each column:

```python
import pandas as pd
d = {
"t1": [1,4,7,3,20],
"t2": [10,21,11,8,5],
"t3": [8,9,4,8,4]
}
df = pd.DataFrame(d)

max_cols = []
for i in [0,1,2]:
    max_cols.append(df.iloc[:, i].max())
print(max_cols)

## [20, 21, 9]
```

- Alter the above the code to calculate the **mean** instead of the maximum value
- Now, calculate the variance (via `var`) as well as the mean.

  You should only have a single loop!

```python
mean_cols = []
var_cols = []
for i in [0,1,2]:
    mean_cols.append(df.iloc[:, i].mean())
    var_cols.append(df.iloc[:, i].var())
```