Assignment 1 Question 1

Sheen Thusoo

h. [4 marks] Suppose $\gamma(\mathcal{P})$ is a measure of dispersion/variation with the same units as the original data (standard deviation and inter-quartile range are some examples). Consider location, scale, and location-scale invariance/equivariance properties. Explain why each of these properties are desirable or undesirable for a measure of dispersion.

Explanation: Assuming that $\gamma(\mathcal{P})$ is a measure of dispersion/variation, it would be desirable for the attribute to be location invariant, scale equivariant, and neither location-scale equivariant nor invariant.

First, the attribute should ideally be location invariant because the spread of the data's distribution should not depend on its center. In a normal distribution, $N(\mu, 1)$, the standard deviation is 1 regardless of its mean. This is because the mean does not influence its spread.

Second, the attribute should ideally be scale equivariant because scaling influences the distribution's spread. As this attribute is a measure of variation, which measures the spread of a distribution, its value would be affected by scaling. For instance, the value of the standard deviation of 4 points could change if each point is scaled by a factor of 2, this can be seen in the code chunk below. This example shows that the original standard deviation is about 1.29 whereas the standard deviation of the points multiplied by 2 is 2.58. This is equivariant since $1.29 \times 2 = 2.58$.

```
points <- c(1,2,3,4)
sd(points)
```

```
## [1] 1.290994
```

```
scaled_points <- points * 2
sd(scaled_points)
```

```
## [1] 2.581989
```

Lastly, the attribute should ideally not be location-scale equivariant, nor invariant. This is because of what has been mentioned above, the attribute should be location invariant but also scale equivarant hence it cannot be location-scale invariant nor equivariant.