CS 498 - Homework 1

Dylan Fleck
dpleck2

2.1 $\text{mean}(\{kx\}) = \frac{1}{N}\sum_{i=1}^{i=N}(kx_i)$

$= k\frac{1}{N}\sum_{i=1}^{i=N} x_i = k\,\text{mean}(\{x\})$

2.2 $\text{mean}(\{kx+c\}) = \frac{1}{N}\sum_{i=1}^{i=N}(x_i + c)$

$= \frac{1}{N}\sum_{i=1}^{i=N} x_i + \frac{N \cdot c}{N} = \text{mean}(\{x\}) + c$

2.3 $\sum_{i=1}^{N}(x_i - \text{mean}(\{x\})) = \sum_{i=1}^{N} x_i - N \cdot \frac{1}{N} \cdot \sum_{i=1}^{N} x_i$

$= \sum_{i=1}^{N} x_i - \sum_{i=1}^{N} x_i = 0$

(switched)

2.5 $\text{std}(kx) = \sqrt{\text{mean}(\{(kx_i - \text{mean}(\{kx\}))^2\})}$

$= \sqrt{\text{mean}(\{k^2(x_i - \text{mean}(\{x\}))^2\})}$

$= \sqrt{k^2\,\text{mean}(\{(x_i - \text{mean}(\{x\}))^2\})}$

$= k\,\text{std}(x)$

(switched)

2.4 $\text{std}(x+c) = \sqrt{\text{mean}(\{(x_i + c - \text{mean}(\{x+c\}))^2\})}$

$= \sqrt{\text{mean}(\{(x_i - \text{mean}(\{x\}) + c - c)^2\})}$

$= \text{std}(x)$

2.6 median($\{x+c\}$), every point is uniformly shifted, including the median point, so: $\Rightarrow$ median($\{x\}$) $+c$

2.7 median($\{kx\}$), every point is uniformly scaled by a common factor, including the median point, so: $\Rightarrow k$ median($\{x\}$)

2.8 $igr\{x+c\}$ = percentile($\{x+c\}$, 75) − percentile($\{x+c\}$, 25)

= percentile($\{x\}$, 75) $+c$ − percentile($\{x\}$, 25) $-c$

= $igr\{x\}$

2.9 $igr\{kx\}$ = percentile($\{kx\}$, 75) − percentile($\{kx\}$, 25)

= k percentile($\{x\}$, 75) − k percentile($\{x\}$, 25)

= $k\, igr\{x\}$

2.10 No, the data varies from 30 to 21,922, indicating a large variance over a relatively small set of 88 data points. This means that a large amount of the data is very far away from the mean.

Mean: 8582.7
Variance: 70141830

2.11 a. No, by cost most powerplants can't be considered an outlier. Efficiency might be a different story, though.

b. Mean = 461.6 , Std = 170.1

c. Mean = 0.570 , Std = 0.187

d. It appears to be skewed a little to the left, as higher CPMs are more favorable for efficiency / business.

2.12 The histograms show that poultry has significantly less calories than beef or meat, while meat had the most variance in calories. Beef has slightly less sodium than meat, while poultry had the most variance of sodium content.

2.13 a. Magazines in group 1 have a larger variance in how many large words are used, while group 2 has less of a variance and group 3 with almost none. Groups 1 & 2 use significaty more larger words then group 3.

b. The sentences show less of a difference between the groups, but group 3 shows much less variance in how many sentences are used.

2.14 a. After dividing the top third of the private colleges, I noticed that the student debt drops considerable at colleges in the 20th percentile range. They probably recieve more money from funding for scholarships, or have richer students.

b. It's apparent that the more highly ranked private schools have a better faculty-to-student ratio, especially those in the top 10th percentile.

c. Colleges which are ranked higher have the tendency to cost more.

d. Ranking doesn't play a significant factor in the cost of the college.