Solutions STA 100 HW 1

Dr. Erin K. Melcon

1. (a) Since this has a natural gap, it is a discrete random variable.
   (b) Since this has a natural minimum gap of 1 cent, it is a discrete random variable.
   (c) Since these are labels and do not have a natural rank, it is a nominal variable.
   (d) Since width can take on (in theory) any number within an interval, it is continuous.

---

2. (a) The mean of the data set is : $\bar{X} = \frac{1}{n}\sum X_i = \frac{1}{10}(299) = 29.9$
   (b) Since $0.50(n+1) = 0.50(10+1) = 5.5$, we average the $(5)^{th}$ and $(6)^{th}$. Thus, the median is equal to: 27.5.
   (c) The variance of this dataset is: $s^2 = \frac{1}{n-1}(\sum X_i^2 - n\bar{X}^2) = \frac{1}{10-1}(10111 - 10(29.9)^2) = \frac{1}{9}(1170.9) = 130.1$
   (d) The standard deviation is: $s = \sqrt{s^2} = \sqrt{130.1} = 11.4061$.
   (e) Thus, the typical deviation of number of leaves falling from the mean is 11.4061

---

3. Note, the ordered data is:

$$21, 22, 22, 24, 27, 28, 30, 32, 33, 60$$

   (a) Since $0.25(n+1) = 0.25(10+1) = 2.75$, we average the $(2)^{th}$ and $(3)^{th}$. Thus, Q_1 is equal to: 22.
   (b) Since $0.75(n+1) = 0.50(10+1) = 8.25$, we average the $(8)^{th}$ and $(9)^{th}$. Thus, Q_3 is equal to: 32.5.
   (c) The cutoff value is:
   lower cutoff: $Q_1 - 1.5(Q_3 - Q_1) = 22 - 1.5 * (32.5 - 22) = 6.25$
   (d) The cutoff value is:
   upper cutoff: $Q_3 + 1.5(Q_3 - Q_1) = 32.5 + 1.5 * (32.5 - 22) = 48.25$
   (e) Thus, the value which is an outlier is: 60
   (f) Since $0.30(n+1) = 0.30(10+1) = 3.3$, we average the $(3)^{th}$ and $(4)^{th}$. Thus, $y^{(30)}$ is equal to: 23.

---

4. (a) The mean of this dataset can be calculated as:
   $\bar{X} = \frac{1}{n}\sum X_i = \frac{1}{100}(0*20 + 1*40 + 2*24 + 3*14 + 10*2) = \frac{1}{100}(150) = 1.5$
   (b) Since $0.50(n+1) = 0.50(100+1) = 50.5$, we average the $(50)^{th}$ and $(51)^{th}$. Thus, the median is equal to: 1.
   (c) The variance of this dataset can be calculated as:
   $s^2 = \frac{1}{100-1}([0^2 * 20 + 1^2 * 40 + 2^2 * 24 + 3^2 * 14 + 10^2 * 2] - 100(1.5)^2) = 2.3939$
   (d) The standard deviation of this dataset can be found as:
   $s = \sqrt{s^2} = \sqrt{2.3939} = 1.5472$

---

5. (a) Since $0.25(n+1) = 0.25(100+1) = 25.25$, we average the $(25)^{th}$ and $(26)^{th}$. Thus, Q_1 is equal to: 1.
   (b) Since $0.75(n+1) = 0.50(100+1) = 75.75$, we average the $(75)^{th}$ and $(76)^{th}$. Thus, Q_3 is equal to: 2.
   (c) The cutoff value is:
   lower cutoff: $Q_1 - 1.5(Q_3 - Q_1) = 1 - 1.5 * (2 - 1) = -0.5$
   (d) The cutoff value is:
   upper cutoff: $Q_3 + 1.5(Q_3 - Q_1) = 2 + 1.5 * (2 - 1) = 3.5$
   (e) Thus, the values which are an outliers are: 10, 10

---

6. (a) FALSE. The standard deviation gives information about how far away values are from the mean, but no information on the mean itself. You could have a very large mean, but very small standard deviation, for example.

(b) FALSE. Since outliers will necessarily be one of the largest or smallest observations, they will have a large influence on the range.

(c) TRUE. The 90^{th} percentile has 90% below it, so 10% must be above it.

(d) TRUE. Since the mean equally weights all ovservations in a dataset, outliers tend to change the mean much more than the median.
