# Biostatistics Homework 1

1.

Two drugs, amantadine (A) and rimantadine (R), are being studied for use in combatting the influenza virus. A single 100-milligram dose is administered orally to healthy adults. The variable studied is Tmax, the time in minutes required to reach maximum plasma concentration. These data are obtained (based on information found in "Drug Therapy", Gordon Douglas, Jr., New England Journal of Medicine, vol. 322, February 1990, pp. 443–449):

$T_{max}(A)$			$T_{max}(R)$		
105	123	12.4	221	227	280
126	108	134	261	264	238
120	112	130	250	236	240
119	132	130	230	246	283
133	136	142	253	273	516
145	156	170	256	271	
200					

(a) Construct a side-by-side box plot and identify outliers. (10%)

(b) Calculate $\bar{x}$ and s^2 for the data of set A. (5%)

(c) Assume that the outlier of set A is the result of a misplaced decimal point. Correct the error by deleting the decimal and see what changes this makes in your box plot. Recompute $\bar{x}$ and s^2, using the correct data point, and compare your results to those of part b. (10%)

(d) Is there an outlier in set R? If so, is there an obvious legitimate reason to delete it from the data set. (5%)
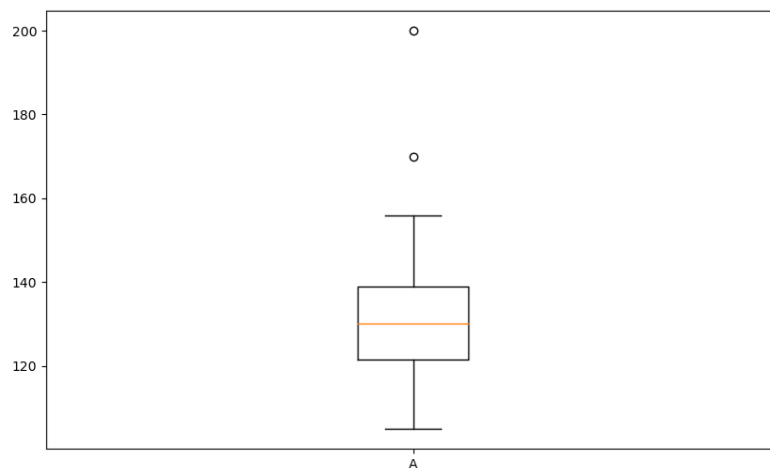
(a) A 的 outlier12.4、170、200

R 的 outlier516
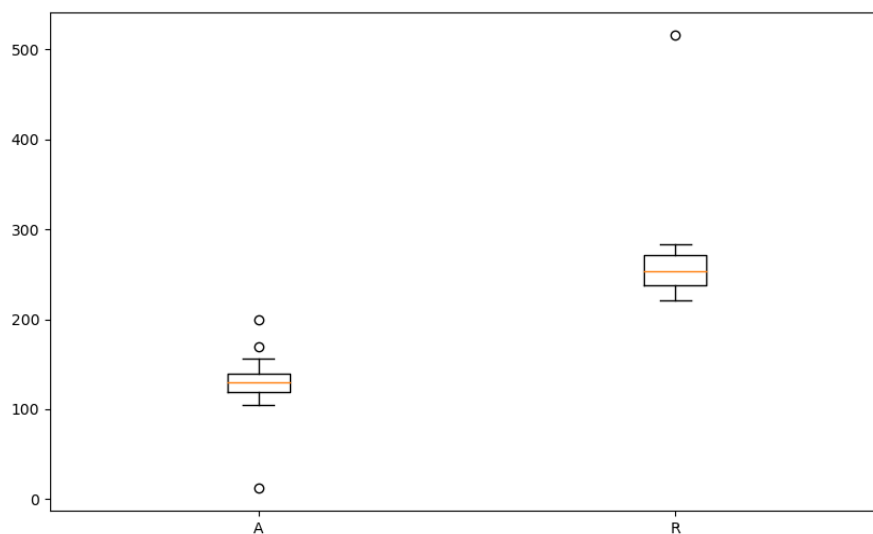
(b)

$\bar{x}$=(105+123+12.4+126+108+134+120+112+130+119+132+130+133+136+142+145+156+170+200)/19=128.07

s^2=$((105 - 128.07)^2 + (123 - 128.07)^2 + (12.4 - 128.07)^2 + (126 - 128.07)^2 + (108 - 128.07)^2 + (134 - 128.07)^2 + (120 - 128.07)^2 + (112 - 128.07)^2 + (130 - 128.07)^2 + (119 - 128.07)^2 + (132 - 128.07)^2 + (130 - 128.07)^2 + (133 - 128.07)^2 + (136 - 128.07)^2 + (142 - 128.07)^2 + (145 - 128.07)^2 + (156 - 128.07)^2 + (170 - 128.07)^2 + (200 - 128.07)^2)/(19 - 1) = 1282._{68}$

(c) A 的 outlier 只剩 170、200

$\bar{x} = (105 + 123 + 124 + 126 + 108 + 134 + 120 + 112 + 130 + 119 + 132 + 130 + 133 + 136 + 142 + 145 + 156 + 170 + 200)/19 = 133.95$

s^2=$((105 - 133.95)^2 + (123 - 133.95)^2 + (124 - 133.95)^2 + (126 - 133.95)^2 + (108 - 133.95)^2 + (134 - 133.95)^2 + (120 - 133.95)^2 + (112 - 133.95)^2 + (130 - 133.95)^2 + (119 - 133.95)^2 + (132 - 133.95)^2 + (130 - 133.95)^2 + (133 - 133.95)^2 + (136 - 133.95)^2+(145 - 133.95)^2 + (142 - 133.95)^2 + (156 - 133.95)^2 + (170 - 133.95)^2 + (200 - 133.95)^2)/(19 - 1) = 503.83$

(d) 516 是 outlier，因為 Q3:271、Q1:238，IQR:271-238=33，上限=

271+1.5*33=320.5

2. The data sets below are temperature readings from two different sensors (a Celsius sensor and a Fahrenheit sensor). Which sensor is better by picking the one with the least variance.(15%)

<p align="center">Celsius: [0, 10, 20, 30, 40]</p>

<p align="center">Fahrenheit: [32, 50, 68, 86, 104]</p>

Sol:

$$\bar{c} = \frac{0 + 10 + 20 + 30 + 40}{5} = 20$$

$$Vc = ((0 - 20)^2 + (10 - 20)^2 + (20 - 20)^2 + (30 - 20)^2 + (40 - 20)^2)/(5 - 1) = 250$$

$$\bar{f} = \frac{32 + 50 + 68 + 86 + 104}{5} = 68$$

$$Vf = ((32 - 68)^2 + (50 - 68)^2 + (68 - 68)^2 + (86 - 68)^2 + (104 - 68)^2)/(5 - 1) = 810$$

Ans:Vc,因為 Vc<Vf

3.**Coevolution.** Different varieties of the tropical flower Heliconia are fertilized by different species of hummingbirds. Over time, the lengths of the flowers and the form of the hummingbirds' beaks have evolved to match each other. Here are data on the lengths in millimeters of two variates of these flowers on the island of Dominica.

<table>
<thead>
<tr><th colspan="9">H. caribaea red</th></tr>
</thead>
<tbody>
<tr><td>41.90</td><td>42.01</td><td>41.93</td><td>43.09</td><td>41.47</td><td>41.69</td><td>39.78</td><td>40.57</td></tr>
<tr><td>39.63</td><td>42.18</td><td>40.66</td><td>37.87</td><td>39.16</td><td>37.40</td><td>38.20</td><td>38.07</td></tr>
<tr><td>38.10</td><td>37.97</td><td>38.79</td><td>38.23</td><td>38.87</td><td>37.78</td><td>38.01</td><td></td></tr>
</tbody>
</table>

<table>
<thead>
<tr><th colspan="8">H. caribaea yellow</th></tr>
</thead>
<tbody>
<tr><td>36.78</td><td>37.02</td><td>36.52</td><td>36.11</td><td>36.03</td><td>35.45</td><td>38.13</td><td>37.1</td></tr>
<tr><td>35.17</td><td>36.82</td><td>36.66</td><td>35.68</td><td>36.03</td><td>34.57</td><td>34.63</td><td></td></tr>
</tbody>
</table>

(a) Make a back-to-back stem-leaf plot to compare the two samples. That is, use one set of stems with two sets of leaves, one to the right and one to the left of the stems. (Draw a line on either side of the stems to separate stems and leaves.) Order both sets of leaves from smallest at the stem to largest away from the stem. (Please round the data to the first decimal place. Use non-fractional part as the stem, first decimal place as leaf. Ex: 41.93 → 41.9 → stem = 41, leaf = 9; 38.79 → 38.8, stem = 38, leaf = 8). (15%)

(b) What are the most important differences among the two varieties of flower? (5%)

(c) Find Q1, Median and Q3 for each group. (10%)

(a)

```
      yellow              |  |           red
                          | 43| 1
                          | 42| 0 2
                          | 41| 5 7 9 9
                          | 40| 6 7
                          | 39| 2 6 8
                          | 38| 0 0 1 1 2 2 8 9
              1  1  0| 37| 4 8 9
      8 8 7 5 1 0  0| 36|
              7  5  2| 35|
                 6  6| 34|
```

(b)

Red 品種長度 Q1、中位數、Q3 皆比 yellow 長

(c)

	Q1	Median	Q3
Red	38.1	39.2	41.6
yellow	35.6	36.1	36.8