STA286 Problem Set 1 Solutions

Question 1.1

Part(a)

As the sample consists of 15 data records, the sample size is 15.

Part(b)

Computation for the sample mean:

$$\bar{x} = \frac{\sum_1^{15} x_i}{15} = 3.7867$$

Part (c)

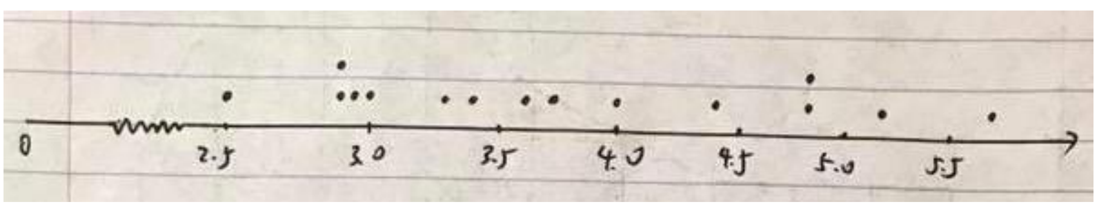
The sorted list in ascending order is as following:

2.5, 2.8, 2.8, 2.9, 3.0, 3.3. 3.4, 3.6, 3.7, 4.0, 4.4, 4.8, 4.8, 5.2, 5.6

With 15 elements within the sample, the median is the $\frac{15+1}{2} = 8$th element in the sorted list, and thus the medium is "3.6".

Part(d)

The dot plot:



Part(e)

With 15 elements, note that $15 \times 20\% = 3$, thus the three smallest and three largest elements would be trimmed.

2.5, 2.8, 2.8, 2.9, 3.0, 3.3. 3.4, 3.6, 3.7, 4.0, 4.4, 4.8, 4.8, 5.2, 5.6

The trimmed mean is then the mean of the remaining 9 elements:

$$\overline{x_{trimmed}} = \frac{\sum_1^9 x_i}{9} = \frac{2.9 + 3.0 + 3.3 + \cdots 4.4 + 4.8}{9} = 3.6778$$

Part(f)

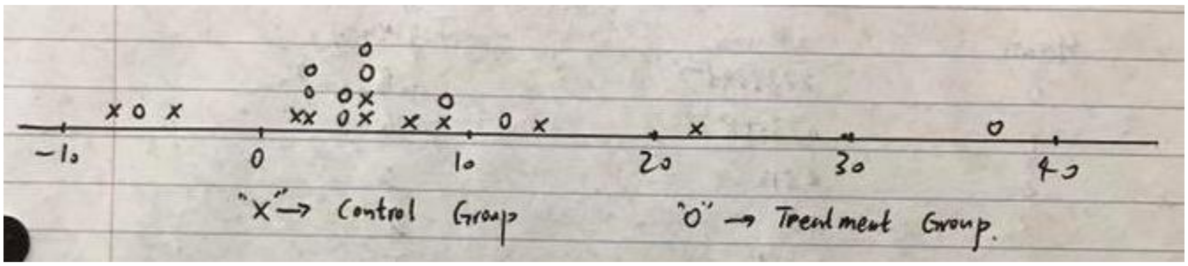
The sample mean is 3.7867 while the trimmed mean is 3.6778. The two values are really close to each other, due to the fact that there isn't outliers with extremely large or small values within the sample.

Thus, both are almost equally descriptive as a center of location.

Question 1.5

Part(a)

The dot plot:



Note in the figure above, "X" represents data points within control group; while "O" represents data points within the treatment group.

Part(b)

For control group:

Mean: $\bar{x} = \frac{\sum_1^{10} x_i}{10} = 5.6$

Median: Taking the average of 5th element "5" and 6th element "5", the medium is 5.

10% trimmed mean: Trim the smallest single element "-7" and largest single element "22", computing the mean of the rest:

$$\overline{x_{trimmed}} = \frac{\sum_1^8 x_i}{8} = 5.125$$

For treatment group:

Mean: $\bar{x} = \frac{\sum_1^{10} x_i}{10} = 7.6$

Median: Taking the average of 5th element "4" and 6th element "5", the medium is 4.5.

10% trimmed mean: Trim the smallest single element "-6" and largest single element "37", computing the mean of the rest:

$$\overline{x_{trimmed}} = \frac{\sum_1^8 x_i}{8} = 5.625$$

Part(c)

Difference in mean is 2.0 in favor of treatment group, which appears to be evident for treatment's efficacy.

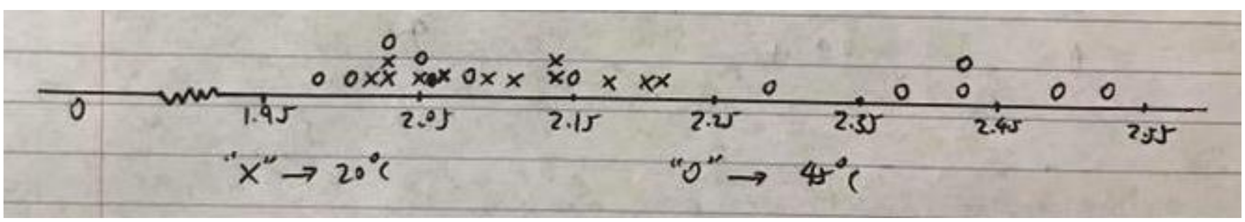
However when comparing medians and trimmed means, treatment group doesn't show apparent advantages (it even has a lower median than control group).

The reason is mostly due to the abnormally large outlier in the treatment group with the value "37", which is 15 more larger than the largest value in the control group. Abnormally large outlier would bring the sample mean up which might not provide the real representation of the sample data.

Question 1.6

Part(a)

The dot plot:



Note in the figure above, "X" represents data points with 20^oC temperature; while "O" represents data points with 45^oC temperature.

Part(b)

For 20^oC:

$$\bar{x} = \frac{\sum_1^{12} x_i}{12} = 2.1075$$

For 45^oC:

$$\bar{x} = \frac{\sum_1^{12} x_i}{12} = 2.2350$$

Part(c)

From the plot, it could be seen that data points labelled by "O", and thus under $45^o C$, are more often appeared in the higher scale region; whereas data points labelled by "X", and thus under $20^o C$, are more often clustered within lower scale region.

Thus it does appear that tensile strength tends to increase along with the temperature under which the experiment is carried on.

Part(d)

Furthermore, notice that data points labelled by "O" ($45^o C$) are more spread out, indicating that higher temperature leads to higher variation (or standard deviation) in tensile strength.

Question 1.7

Sample variance:

$$s^2 = \frac{\sum_1^n (x_n - \bar{x})^2}{n-1}$$

With n = 15 and $\bar{x} = 3.7867$, plug in the numbers and obtain:

$$s^2 = 0.94267$$

Sample standard deviation:

$$s = \sqrt{s^2} = 0.97091$$

Question 1.11

For control group:

Sample variance:

$$s^2 = \frac{\sum_1^n (x_n - \bar{x})^2}{n-1}$$

With n = 10 and $\bar{x} = 5.6$, plug in the numbers and obtain:

$$s^2 = 69.378$$

Sample standard deviation:

$$s = \sqrt{s^2} = 8.329$$

For treatment group:

Sample variance:

$$s^2 = \frac{\sum_1^n (x_n - \bar{x})^2}{n-1}$$

With n = 10 and $\bar{x} = 7.6$, plug in the numbers and obtain:

$$s^2 = 128.044$$

Sample standard deviation:

$$s = \sqrt{s^2} = 11.316$$

Question 1.12

For $20^o C$ group:

Sample variance:

$$s^2 = \frac{\sum_1^n (x_n - \bar{x})^2}{n-1}$$

With n = 12 and $\bar{x} = 2.1075$, plug in the numbers and obtain:

$$s^2 = 0.00502$$

Sample standard deviation:

$$s = \sqrt{s^2} = 0.07086$$

For $45^o C$ group:

Sample variance:

$$s^2 = \frac{\sum_1^n (x_n - \bar{x})^2}{n-1}$$

With n = 12 and $\bar{x} = 2.2350$, plug in the numbers and obtain:

$$s^2 = 0.04128$$

Sample standard deviation:

$$s = \sqrt{s^2} = 0.20318$$

As the variation (and trivially also the standard deviation) of 45^oC sample data is a lot larger than that of the 20^oC sample data, it is verified that increasing in temperature leads to increase of variability in tensile strength.

Question 1.16

First of all, note the definition of mean:

$$\bar{x} = \frac{\sum_1^n x_i}{n}$$

Simplify the expression and plug in the above definition for mean:

$$\sum_{i=1}^{n}(x_i - \bar{x})$$

$$= \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} \bar{x}$$

$$= \sum_{i=1}^{n} x_i - n\bar{x}$$

$$= \sum_{i=1}^{n} x_i - n\frac{\sum_1^n x_i}{n}$$

$$= \sum_{i=1}^{n} x_i - \sum_{i=1}^{n} x_i$$

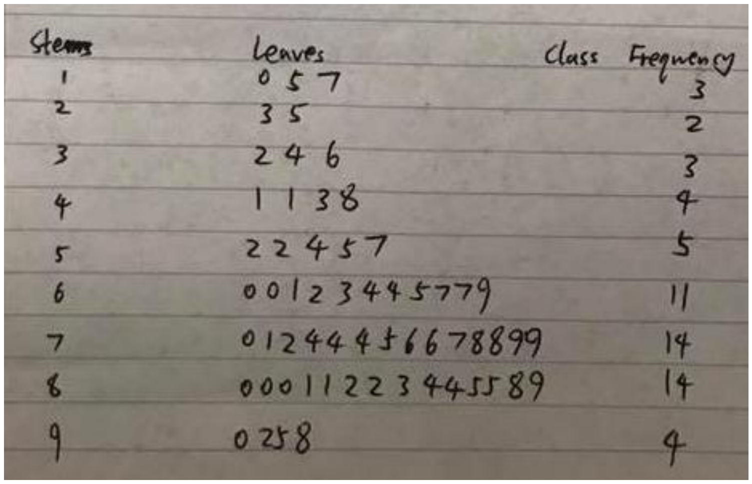
$$= 0$$

QED.

Question 1.18

Part (a)

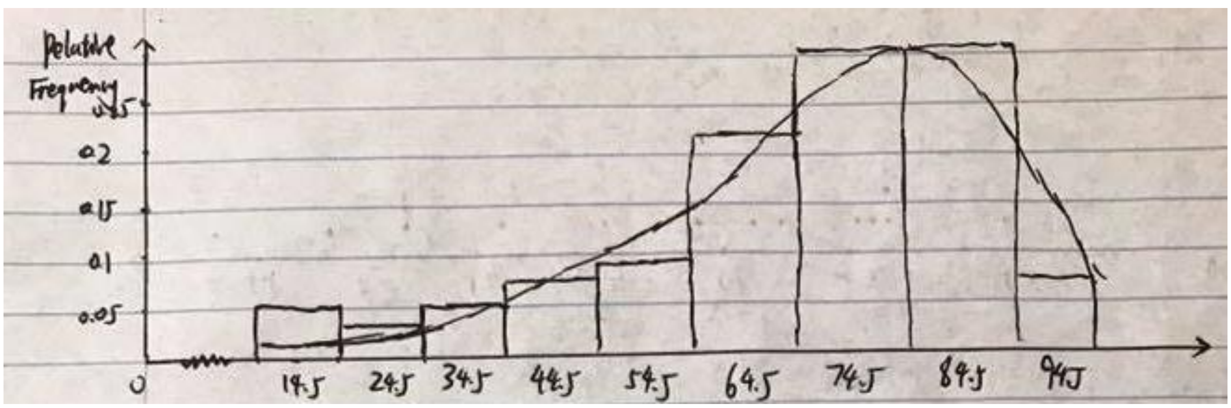
The stem-and-leaf plot:



Stems	Leaves	Class Frequency
1	0 5 7	3
2	3 5	2
3	2 4 6	3
4	1 1 3 8	4
5	2 2 4 5 7	5
6	0 0 1 2 3 4 4 5 7 7 9	11
7	0 1 2 4 4 4 5 6 6 7 8 8 9 9	14
8	0 0 0 1 1 2 2 3 4 4 5 5 8 9	14
9	0 2 5 8	4

Part (b)

The relative frequency histogram:



Note that each label on horizontal axis is taken as the midpoint of the corresponding interval (e.g. for first interval 10~19, the midpoint is 14.5).

The curve is the estimate of the graph of distribution.

As the graph has a long tail at the left, the distribution is skewed to the left.

Part (c)

Sample mean:

$$\bar{x} = \frac{\sum_{1}^{60} x_i}{60} = 65.483$$

Sample median:

With 60 samples, the median is the average between 30th smallest sample "71" and 31th smallest sample "72", and thus has the value **71.5**.

Sample standard deviation:

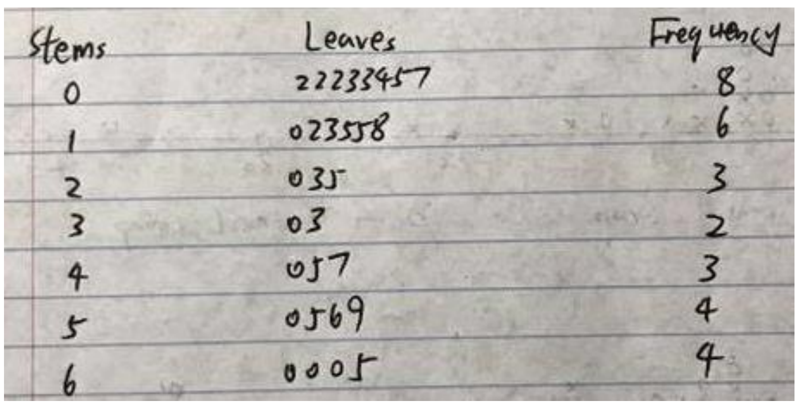
$$s^2 = \frac{\sum_1^n (x_n - \bar{x})^2}{n-1} = 446.6268$$

$$s = \sqrt{s^2} = 21.1335$$

Question 1.19

Part (a)

The stem-and-leaf plot:



Stems	Leaves	Frequency
0	22233457	8
1	023558	6
2	035	3
3	03	2
4	057	3
5	0569	4
6	0005	4

Part (b)

Relative frequency distribution could be derived through individual frequencies as shown in the table above.

Sample computation:

Total amount of data: 30

For stem 0:

It represents data range 0~0.9. As it has 8 occurrences out of total 30, it's relative frequency is 0.2667.

Following the same computation, we can obtain all relative frequencies.

The relative frequency is shown here:

Interval	0~0.9	1.0~1.9	2.0~2.9	3.0~3.9	4.0~4.9	5.0~5.9	6.0~6.9
Relative	0.2667	0.20	0.10	0.0667	0.10	0.1333	0.1333

Frequency							

Part (c)

Sample mean:

$$\bar{x} = \frac{\sum_1^{30} x_i}{30} = 2.7967$$

Sample range:

Max value within the sample: 6.5

Min value within the sample: 0.2

range: 6.5 - 0.2 = 6.3

Sample standard deviation:

Sample variance:

$$s^2 = \frac{\sum_1^n (x_n - \bar{x})^2}{n-1} = 4.9610$$

Sample standard deviation:

$$s = \sqrt{s^2} = 2.2273$$

Question 2.1

Part (a)

According to the numerical rule, the sample space would be:

S = {8, 16, 24, 32, 40, 48}

Part (b)

To obtain the set, we need to solve the quadratic equation. Note the following factorization of the quadratic terms:

$$x^2 + 4x - 5 = (x+5)(x-1) = 0$$

Thus the solutions are trivially:

$$x1 = -5$$

$$x2 = 1$$

Thus the sample space is:

S = {-5, 1}

Part (c)

According to the rule, just enumerate all possible cases:

S = {T, HT, HHT, HHH}

Part (d)

According to the 7-continent standard, listing all the continents:

S = {North America, South America, Antarctica, Asia, Africa, Australia (Oceania), Europe}

Part (e)

The elements within the sample space should satisfy both inequalities simultaneously.

Solve the inequalities one by one:

First inequality:

$$2x - 4 \geq 0$$

$$x \geq 2$$

Second inequality:

$$x < 1$$

It is apparent that there is no x that would satisfy both inequalities simultaneously. Thus, the corresponding sample space is an empty set:

S = ∅

Question 2.3

To answer the question purposed, first we need to list out the elements of each event:

Part (a)

The event is:

A = {1,3}

Part (b)

Enumerating all numbers on a die, the event is:

B = {1,2,3,4,5,6}

Part (c)

Solving the quadratic equation, note the factorization of the expression:

$$x^2 - 4x + 3 = (x - 1)(x - 3) = 0$$

From this factorization, we can trivially obtain the solutions, and thus the elements of the event:

C = {1,3}

Part (d)

With 6 coin tosses, the amount of heads could be from 0 heads (no heads at all) to 6 heads (all tosses are heads).

Thus, the event it describes include the following elements:

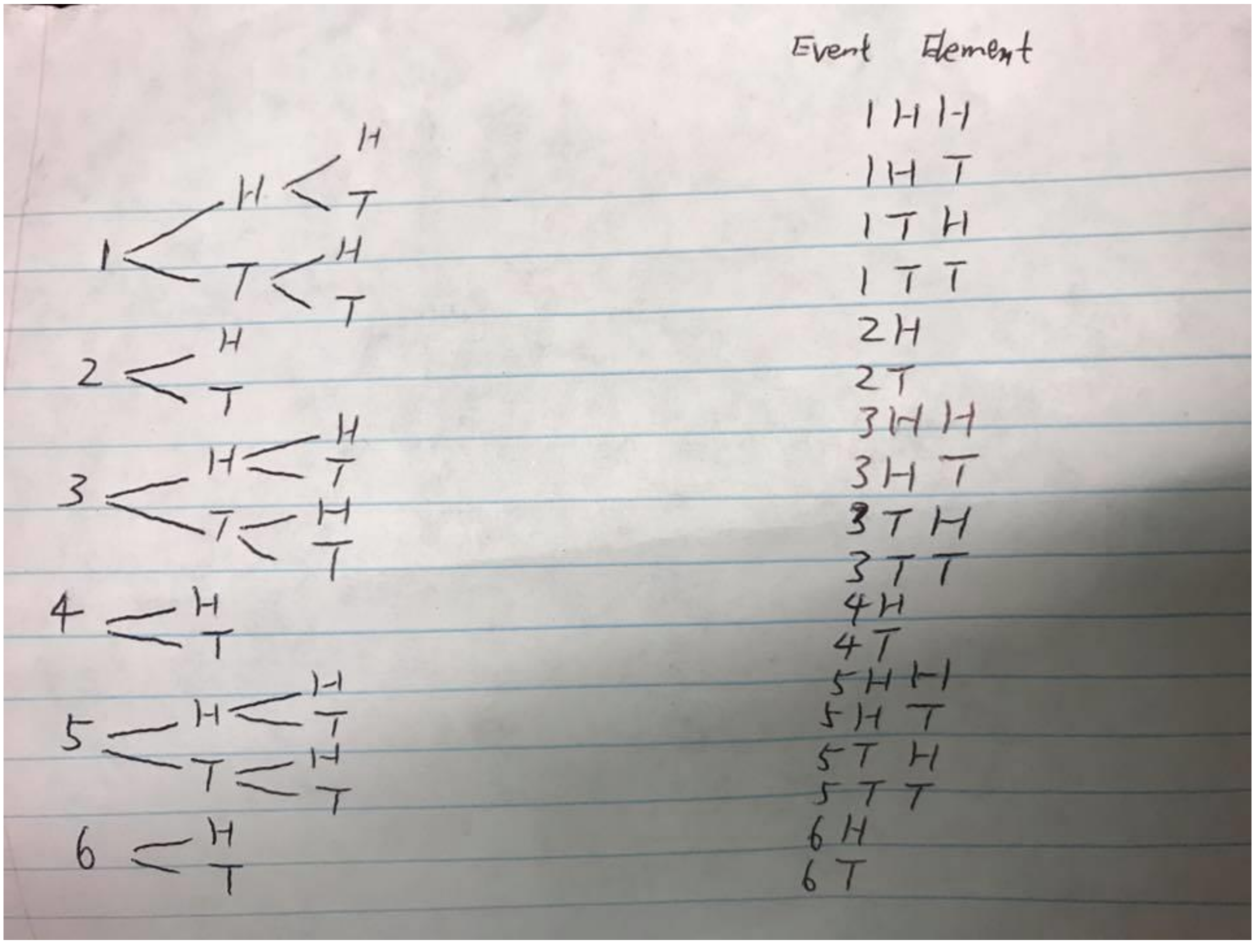
D = {0,1,2,3,4,5,6}

Comparing the events from four parts, it is obvious that the following pair of events are equal:

A = C

Question 2.5

The tree diagram is shown as following:



Question 2.9

Part (a)

The rule of event A is trivially conditioning on the first die toss result, listed as following:

A = {1HH, 1HT, 1TH, 1TT, 2H, 2T}

Part (b)

Trivially, the event B is as following:

B = {1TT, 3TT, 5TT}

Part (c)

A', as the complement of A, would consist all elements within the sample space that are not in A. Thus, A' is as following:

$$A' = \{3HH, 3HT, 3TH, 3TT, 4H, 4T, 5HH, 5HT, 5TH, 5TT, 6H, 6T\}$$

Part (d)

With A' and B listed above, their intersection is readily obtained:

$$A' \cap B = \{3TT, 5TT\}$$

Part (e)

The union is trivially taken given A and B already listed as above:

$$A \cup B = \{1HH, 1HT, 1TH, 1TT, 2H, 2T, 3TT, 5TT\}$$

Question 2.14

Part (a)

Taking the union is straight forward:

$$A \cup C = \{0, 2, 3, 4, 5, 6, 8\}$$

Part (b)

Note that A consists only even numbers; while B consists only odd numbers, the intersection thus would be empty set:

$$A \cap B = \emptyset$$

Part (c)

Taking the compliment gives:

$$C' = \{0, 1, 6, 7, 8, 9\}$$

Part (d)

Compute step by step.

First of all, taking the intersection given the C compliment above:

$$C' \cap D = \{1, 6, 7\}$$

Then take the union:

$$(C' \cap D) \cup B = \{1, 6, 7\} \cup \{1, 3, 5, 7, 9\}$$

$$= \{1, 3, 5, 6, 7, 9\}$$

Part (e)

Note that as C is a strict subset of S the sample space, intersection of S and C would give back C. Thus the event essentially is compliment of C, which is the same as in part (c):

$$(S \cap C)' = \{0, 1, 6, 7, 8, 9\}$$

Part (f)

Compute step by step:

First of all:

$$A \cap C = \{2, 4\}$$

Then take the union:

$$A \cap C \cap D' = \{2, 4\} \cap \{0, 2, 3, 4, 5, 8, 9\}$$

$$= \{2, 4\}$$

Question 2.16

Part (a)

The union would be all the regions in S that are covered by either M or N:

$$M \cup N = \{x \mid 0 < x < 9\}$$

Part (b)

The intersection would be taking all the regions covered by both M and N:

$$M \cap N = \{1 < x < 5\}$$

Part (c)

First of all, taking the compliments of both events:

$$M' = \{x \mid 0 < x \leq 1 \text{ or } 9 \leq x < 12\}$$

$$N' = \{x \mid 5 \leq x < 12\}$$

Then, taking the union of the two leads to:

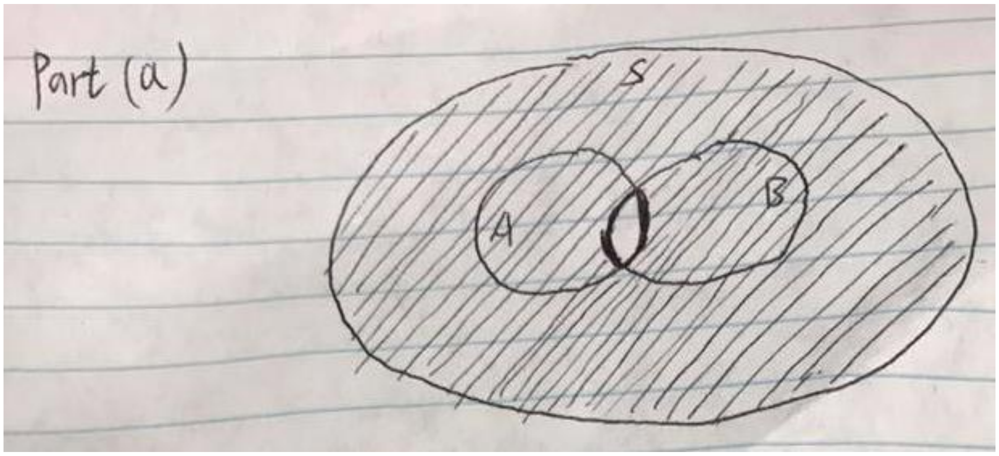
$$M' \cap N' = \{x \mid 9 \leq x < 12\}$$

Question 2.17

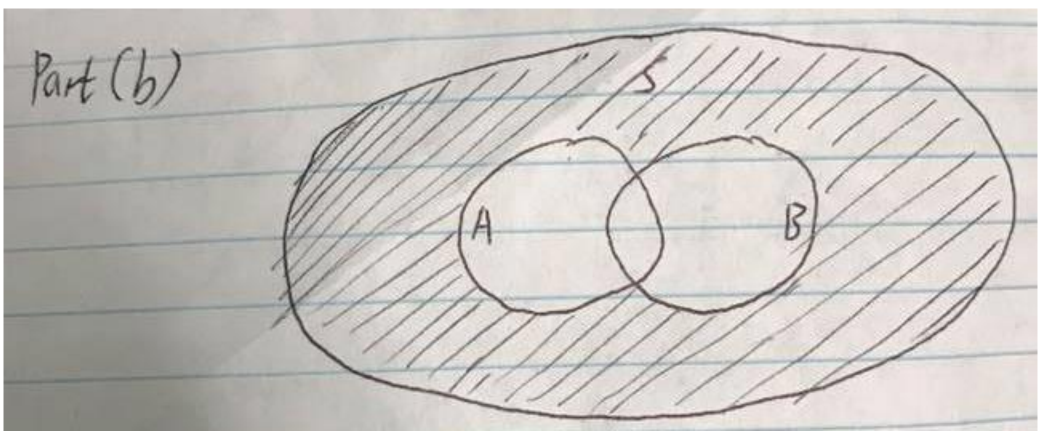
The Venn diagrams and corresponding shaded regions are shown in below:

Part (a).

Essentially every part except for the intersection region is included:



Part (b).



Part (c).

Compute step by step. First get the intersection region, then take the union using B and that intersection region, we then obtain:

Part (C)