STAT 3115 – Homework 1 Solutions

1. (a) The population of interest is all S&T employees.

   (b) Variables include:
   - Lot letter: Categorical
   - Distance: Quantitative
   - Satisfaction: Categorical

   (c) Conducting a **census** would require information from every employee. A **sample** would only require a subset of the employees to respond to the survey. We would hope that our sample is representative of the population, however, so that we can make inference on the population.

   (d)
   - *Simple Random Sample.* We could collect a list of all employees and randomly select ones to complete the survey. This would be an acceptable sample since every employee would have equal chance of being chosen for our survey, however it may be interesting to compare responses across gender, age, etc. In that case, a stratified random sample would be more effective.
   - *Stratified Random Sample.* We could obtain a list of all employees and then divide them into groups. Groups we might consider would be employee type, age, and gender. From these groups we will randomly pick employees to answer the survey. This is a good choice if comparison among employees is desired and would ensure that each group is represented equally.
   - *Convenience Sample.* For this type, we could only conduct the survey among people who work in the sample building as the person conducting the study. While it may be easier, it will most likely be biased. For example, the building the survey is conducted in might be very close to one of the parking lots and so the people in the building would be very satisfied with their parking situation since they do not have to walk very far.

2. (a) First, the ordered data set:

Men:	3.0	3.3	3.3	3.4	3.6	3.7	3.8	3.8	3.8	3.9	4.0	4.0

Mean: $\dfrac{3.0 + 3.3 + 3.3 + 3.4 + \cdots + 4.0}{12} = \dfrac{43.6}{12} = 3.63\text{cm}.$

Median: Since n is even, we must average the $0.5n$ and $0.5(n+1)$ observations:

$$\frac{3.7 + 3.8}{2} = 3.75\text{cm}$$

IQR: First, we need to find Q1 and Q3.

Q1: Does not include median since n is even. Our "lower data set" is {3.0, 3.3, 3.3, 3.4, 3.6, 3.7} which still has n even so that the median is

$$\frac{3.3 + 3.4}{2} = 3.35.$$

Q3: Does not include median since n is even. Our "upper data set" is {3.8, 3.8, 3.8, 3.9, 4.0, 4.0} which still has n even so that the median is

$$\frac{3.8 + 3.9}{2} = 3.85.$$

Thus IQR $= 3.85 - 3.35 = 0.5$.

Variance:

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]$$
$$= \frac{1}{11}\left[159.52 - \frac{1}{12}(43.6)^2\right]$$
$$= 0.1006$$

Standard Deviation:

$$s = \sqrt{s^2} = \sqrt{0.1006} = 0.3172.$$

Now for the women:

Women:	2.5	3.0	3.0	3.1	3.1	3.1	3.2	3.5	3.7	4.2

Mean: $\frac{2.5 + 3.0 + 3.0 + \cdots + 4.2}{10} = \frac{32.4}{10} = 3.24\text{cm}.$

Median: Since n is even, we must average the $0.5n$ and $0.5(n+1)$ observations:

$$\frac{3.1 + 3.1}{2} = 3.1\text{cm}$$

IQR: First, we need to find Q1 and Q3.

Q1: Does not include median since n is even. Our "lower data set" is $\{2.5, 3.0, 3.0\}$ which has n odd so that the median is 3.0.

Q3: Does not include median since n is even. Our "upper data set" is $\{3.2, 3.5, 3.7, 4.2\}$ which still has n even so that the median is

$$\frac{3.5 + 3.7}{2} = 3.6.$$

Thus IQR $= 3.6 - 3.0 = 0.6$.

Variance:

$$s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right]$$
$$= \frac{1}{9}\left[106.9 - \frac{1}{10}(32.4)^2\right]$$
$$= 0.2138$$

Standard Deviation:

$$s = \sqrt{s^2} = \sqrt{0.2138} = 0.4624.$$

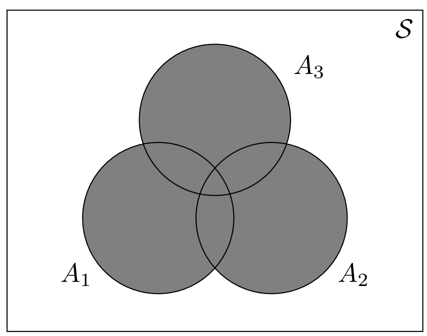
(b) Among the patients studied, both the mean and the median aortic root diameter is smaller in women than men, however women had a larger variance than men. The spread of the middle 50% of the data (IQR) is similar for both women and men though the women's was slightly higher (0.1cm).

(c) For outliers in the men's group, we require observations to be less than Q1−1.5IQR= $3.35 - 0.75 = 2.6$ or greater than Q3+1.5IQR= $3.85 + 0.75 = 4.6$ but none of our observations meet this criteria so **there are no outliers in the men's observations.**
For outliers in the women's group, we require observations to be less than Q1−1.5IQR= $3.0 - 0.9 = 2.1$ or greater than Q3+1.5IQR= $3.6 + 0.9 = 4.5$ but none of our observations meet this criteria so **there are no outliers in the women's observations.**

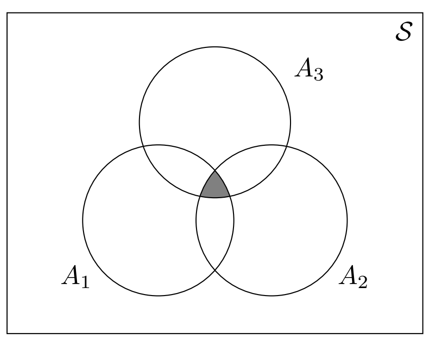
(d) If the 2.5 observation were instead 1.5, the median and thus IQR would be unaffected by this change. However, the mean, variance, and standard deviation would be pulled down towards this value.

4. (a) $\mathcal{S}$ = {FFFF, FFFV, FFVF, FVFF, VFFF, FFVV, FVFV, VFFV, FVVF, VFVF, VVFF, FVVV, VFVV, VVFV, VVVF, VVVV}.
   (b) B = {FFFV, FFVF, FVFF, VFFF}.
   (c) C = {FFFF, VVVV}.
   (d) D = {FFFF, FFFV, FFVF, FVFF, VFFF}.
   (e) $C \cup D$ = {FFFF, VVVV, FFFV, FFVF, FVFF, VFFF}.
   $C \cap D$ = {FFFF}.
   (f) $B \cup C$ = {FFFV, FFVF, FVFF, VFFF, FFFF, VVVV}.
   $B \cap C = \emptyset$.

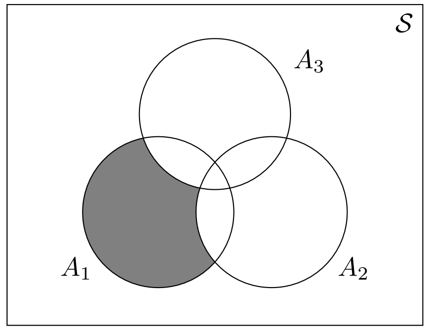
8. (a) $A_1 \cup A_2 \cup A_3$



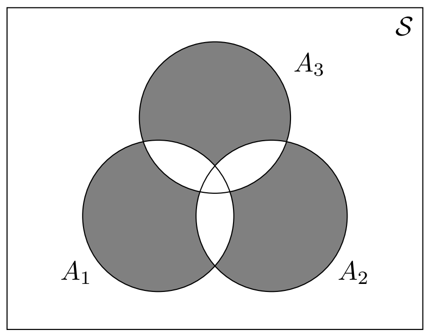
   (b) $A_1 \cap A_2 \cap A_3$



(c) $A_1 \cap A_2' \cap A_3'$



(d) $(A_1 \cap A_2' \cap A_3') \cup (A_1' \cap A_2 \cap A_3') \cup (A_1' \cap A_2' \cap A_3)$



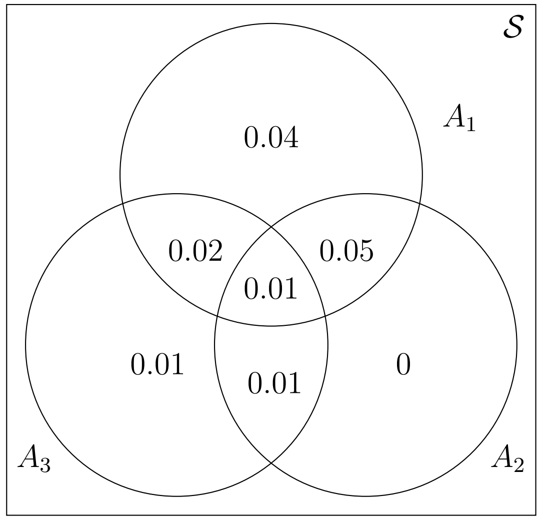
Blood Types. $\mathcal{S} = \{A^+, A^-, B^+, B^-, AB^+, AB^-, O^+, O^-\}$.

Section 2.2

12. (a) $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.5 + 0.4 - 0.25 = 0.65$
    (b) $P(A' \cap B') = P[(A \cup B)'] = 1 - P(A \cup B) = 1 - 0.65 = 0.35$
    (c) $P(A \cap B') = P(A) - P(A \cap B) = 0.5 - 0.25 = 0.25$

20. (a) Let D be the day shift, S be the swing shift, N be the night shift, UnS be unsafe working conditions, and UnR be unrelated to working conditions. Then $\mathcal{S} = \{D/UnS,\ D/UnR,\ S/UnS,\ S/UnR,\ S/UnS,\ S/UnR\}$
    (b) $P(UnS) = 0.1 + 0.08 + 0.05 = 0.23$
    (c) $P(D') = 1 - P(D) = 1 - (0.1 + 0.35) = 0.55$

26. I suggest filling out a *complete* Venn Diagram for this. (We will cover this in class.)



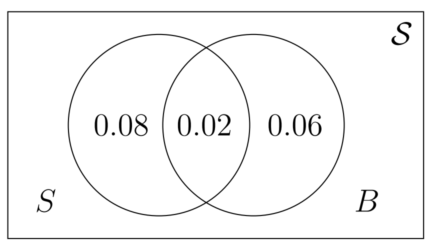
(a) $P(A_1') = 1 - P(A_1) = 1 - 0.12 = 0.88$

(b) $P(A_1 \cap A_2) = 0.05 + 0.01 = 0.06$

(c) $P(A_1 \cap A_2 \cap A_3') = 0.05$

(d) $P[(A_1 \cap A_2 \cap A_3)'] = 1 - 0.01 = 0.99$

Aircraft Defects. Again, with this problem it can be helpful to draw a completed Venn diagram.



(a) $P(\text{at least one}) = 0.08 + 0.06 + 0.02 = 0.16$

(b) $P(\text{B only or S only}) = 0.06 + 0.08 = 0.14$

(c) $P(\text{neither}) = 1 - P(\text{at least one}) = 1 - 0.14 = 0.86$