Solution of HW 1

Problem 1: (1.1)

(a) The population of interest in this case would be U.S. citizens between 22 and 35, and the inferential objective would be finding out what percentage of the population is interested in starting their own business. A sample can be collected by conducting a quantitative survey in randomly selecting a few individuals (U.S. citizens between the ages of 22-35)

(b) The population is single-family homes in the city. the inferential objective would be finding out the approximate average amount of water being consumed per week in the homes. A sample can be collected by randomly selecting households in the city, and reporting the consumed water amout.

(c) The population is transistors of specific type. The inferential objective is figuring out if the average life of the transistor is more than 500 hours. A sample could be ~~coll~~ collected by randomly selecting transistors and testing them.

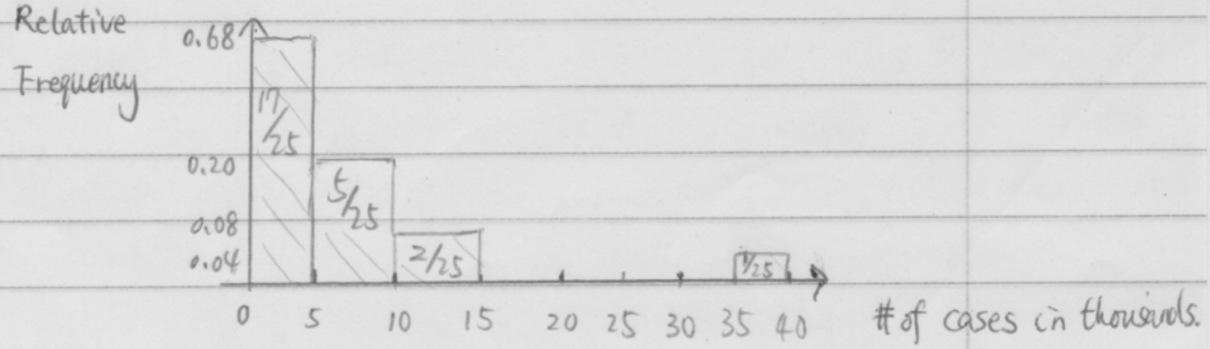
(5)
Problem 2:

A population, by definition, is a large body of data that is the target of interest While a sample is a subset selected from this large body of data. This means that a population is much larger than a sample and consists of every unit in a group that has similar characteristics. Since it is very costly and time consuming to test/survey an entire population, or every unit in a group, a smaller set of randomly selected individuals from this population, or a sample, would be tested instead to make inferences about the ~~general~~ overall results of the entire group. This is the logic used to make an inference about a population from a sample. Parts of a sample come from the population, so the outcomes collected from the sample is assumed to reflect the population.

Problem 3:

A measure of goodness is necessary when making inferences about a population based on a sample because it allows others to understand the Likelihood, or chance, that the inference made from the sample is true about the population. If an inference from a sample is made with no measure of goodness, then one wouldn't know how likely true the inference is about the population

(10)
Problem 4:



(b) X = # of cases of AIDS reported

$$P(X > 10{,}000) = \frac{3}{25}$$

(c)

$$P(X < 3000) = 11/25$$

because 11 cities have the numbers of AIDS cases less than 3000,

(10)
Problem 5.

For complete data:

Sample mean:

$$\bar{y} = \sum_{i=1}^{25} y_i / 25 = 5.68$$

Standar Deviation:

$$S = sd = \sqrt{S^2} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(y_i-\bar{y})^2}$$

$$= \sqrt{\frac{1}{24}\sum_{i=1}^{25}(y_i-5.68)^2}$$

$$= 7.48$$

$$(\bar{y}-2s,\ \bar{y}+2s) = (+5.68-2\times 7.48,\ 5.68+2\times 7.48)$$

$$= (-9.28,\ 20.6)$$

24 out of the 25 cities fall under this interval

which is 96% of the cities. The empirical rule states that 2 standard deviations from the mean there should be 95% of the observations. So it looks like the data falls within the bounds of the empirical rule.

Without the large value 38.3 :

$$\tilde{y} = \sum_{i=1}^{24} y_i / 24 = 4.32$$

$$s = \sqrt{\frac{1}{24-1} \sum_{i=1}^{24} (y_i - \tilde{y})^2} = 3.19.$$

$$(\tilde{y} - 2s, \quad \tilde{y} + 2s) = (-2.37, 10.37)$$

22 out of 24 cities fall under this interval, which is 92% of the cities. Again, the relatively falls within the empirical rule.

When we take away outliers (large values here) from a set of data, the standard deviation usually gets smaller. But the empirical rule seems to hold well for either case.

Problem 6: (1.8)

(a) $\mu = 14$, $\sigma = 17$

$\therefore$ one standard deviation below the mean is $14 - 17 = -3$

(b). Let X = time the user spents online

$$P(X < -3) = P\left(\frac{X-\mu}{\sigma} < \frac{-3-\mu}{\sigma} \right)$$

$$= P\left(N(0,1) < \frac{-3-14}{17} \right)$$

$$= P(N(0,1) < -1)$$

$$= \Phi(-1)$$

$$= 16\%$$

(c)

No, it is not normally distributed since $X \geq 0$ is required in this question, but X could be any value if $X \sim$ Normal Distribution.

(5)
Problem 7.

Let X = weekly maintenance cost.

Then $X \sim N(420, 30^2)$

$$P(X > 450) = P\left(\frac{X-420}{30} > \frac{450-420}{30} \right)$$

$$= P(N(0,1) > 1)$$

$$= 1 - \Phi(1)$$

$$= 16\%$$

(10)
Problem 8.

$$A = \{ (1,2), (1,4), (1,6)$$
$$(2,2), (2,4), (2,6)$$
$$(3,2), (3,4), (3,6)$$
$$(4,2), (4,4), (4,6)$$
$$(5,2), (5,4), (5,6)$$
$$(6,2), (6,4), (6,6) \}$$

$$\bar{C} = \{(2,2), (2,4), (2,6)$$
$$(4,2), (4,4), (4,6)$$
$$(6,2), (6,4), (6,6)\}$$

$$A \cap B = \{(2,2), (2,4), (2,6)$$
$$(4,2), (4,4), (4,6)$$
$$(6,2), (6,4), (6,6)\}$$

$$A \cap \bar{B} = \{(1,2), (1,4), (1,6)$$
$$(3,2), (3,4), (3,6)$$
$$(5,2), (5,4), (5,6)\}$$

$$\bar{A} \cup B = \{(1,1), (1,3), (1,5), (2,2)\ (4,2), (6,2)$$
$$(2,1), (2,3), (2,5)\quad (2,4)\ (4,4), (6,4)$$
$$(3,1), (3,3), (3,5)\quad (2,6)\ (4,6), (6,6)$$
$$(4,1), (4,3), (4,5)$$
$$(5,1), (5,3), (5,5)$$
$$(6,1), (6,3), (6,5)\}$$

$$\bar{A} \cap C = \{(1,1), (1,3), (1,5)$$
$$(2,1), (2,3), (2,5)$$
$$(3,1), (3,3), (3,5)$$
$$(4,1), (4,3), (4,5)$$
$$(5,1), (5,3), (5,5)$$
$$(6,1), (6,3), (6,5)\}$$

Problem 9:

	On Campus	Off Campus
Undergraduate	n_{11}	n_{12}
Graduate	n_{21}	n_{22}

we have known

$$\begin{cases} n_{11} + n_{12} = 36 & \text{— undergraduate students} \\ n_{12} + n_{22} = 9 & \text{— living off campus} \\ n_{12} = 3 & \text{— undergradute \& off campus} \\ n_{11}+n_{12}+n_{21}+n_{22}=60 & \text{— total students} \end{cases}$$

$$\Rightarrow \quad \begin{aligned} n_{11} &= 33 \\ n_{12} &= 3 \\ n_{21} &= 18 \\ n_{22} &= 6 \end{aligned}$$

(a)

$$\underset{\substack{\uparrow \\ \text{UR}}}{n_{11}} + \underset{\substack{\uparrow \\ \text{UR\&Off}}}{n_{12}} + \underset{\substack{\uparrow \\ \text{G\&off}}}{n_{22}} = 42$$

(b) $n_{11} = 33$ → undergraduate & living on campus

(c) $n_{21} = 18$ → graduate & living on campus