Introduction to Statistics for Engineers
Homework 3

Name: Rhea Mae Edwards

OSU-ID: 932-389-303

Instructions

- The homework is due on Monday May. 2nd and must be submitted on Canvas before midnight. (Please read the policies on late homeworks on the syllabus)
- Homeworks must be submitted to Canvas as a Word or PDF document. Any other format (including JPEG) will not be accepted and your homework will be considered late. (Please read the policies on late homeworks on the syllabus)
- You must show your work and provide complete answers in order to receive full credit. Solutions restricted only the final numerical values that do not reflect your statistical reasoning will not receive full credit. The homework is worth 25 points.
- If not using the space assigned for each question, you must clearly indicate the problem that you are working.
- You must include your name and OSU-ID number in your homework document.
- Failing to follow any of these instructions may result in a delay in the grading or a penalization in your final score.

1. The following data are lifetimes (in hundreds of hours) of twenty-six 40-watt 110-volt internally frosted incandescent lamps taken from forced life tests:

4.1	8.6	9.1	9.2	9.2	9.3	9.4	9.5	9.5	9.6	9.7	9.7	9.8
9.9	10.4	10.5	10.6	10.7	10.9	10.9	11.0	11.3	11.6	11.7	12.4	13.4

(a) Find the first quartile Q_1, the median Q_2, the third quartile Q_3 and the interquartile region IQR.

$Q_2 = (9.8 + 9.9)/2 = 9.85$

$Q_1 = ((9.85 - 4.1)/2) + 4.1 = 6.975$

$Q_3 = ((13.4 - 9.85)/2) + 9.85 = 11.625$

$IQR = Q_3 - Q_1 = 11.625 - 6.975 = 4.65$

$Q_1 = 6.975$
Q_2 (median) $= 9.85$
$Q_3 = 11.625$
$IQR = 4.65$

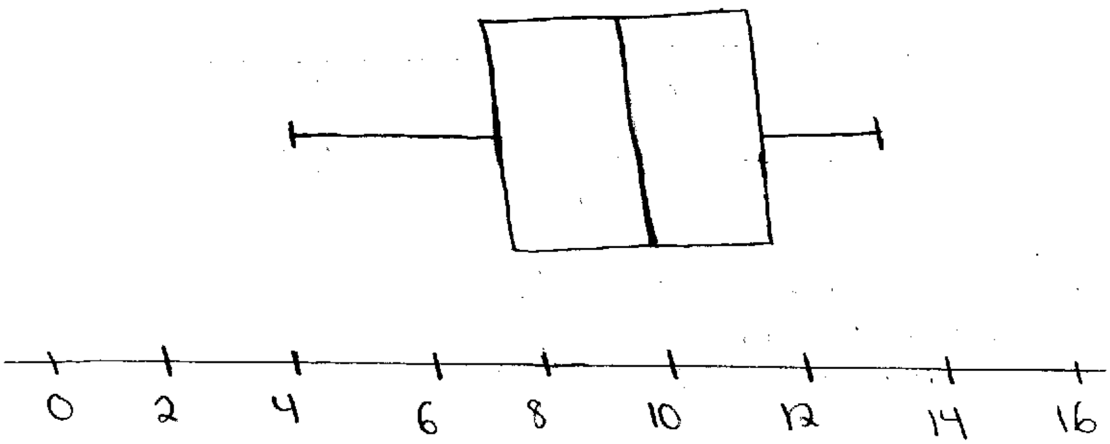
(b) Using you answers from part (a) determine the boundaries for the lower and upper inner fences.

Upper Inner Fence $= Q_3 + 1.5 \times IQR$
$= 11.625 + (1.5 \times 4.65) = 18.6$

Lower Inner Fence $= Q_1 - 1.5 \times IQR$
$= 6.975 - (1.5 \times 4.65) = 0$

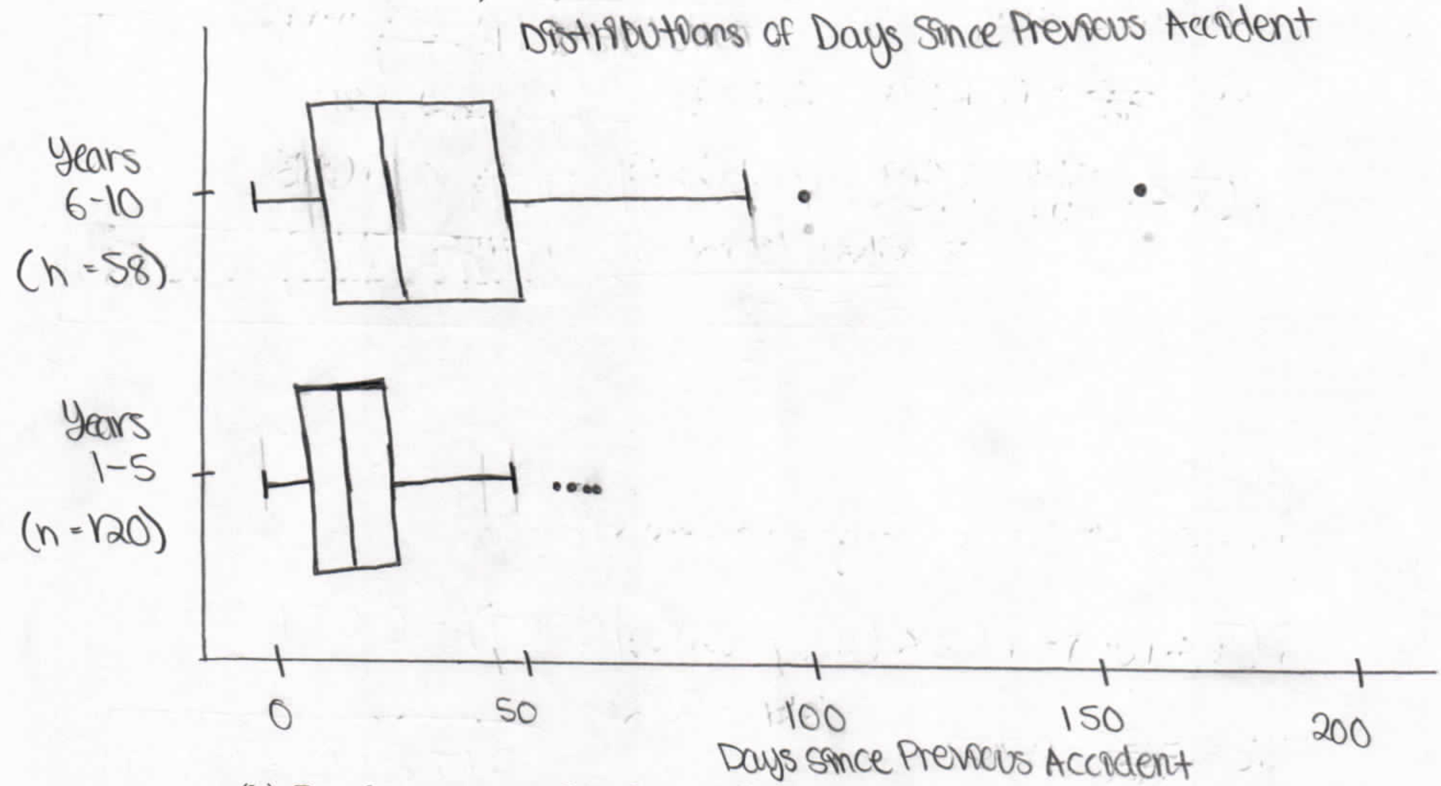
Lower Inner Fence = 0
Upper Inner Fence = 18.6

(c) Construct (by hand) the boxplot for these data.



2. A certain study followed for over a period of 10 years the times between accidents at a major chemical facility in the US. After 5 years (from the beginning of the study), some important changes in the company's safety policy were implemented. The file "time_accidents.csv" contains the data for the first and last 5 years of the study, where each entry corresponds to the number of days since the previous accident.

   (a) Construct parallel boxplots for these data (you can do it by hand or use a statistical software).



   (b) Based on your results, do you think the change in the safety policies had any effect? Justify your answer.

Yes, the change in the safety policies had a positive effect with increasing the number of days since previous accidents at the major chemical facility in the U.S.
The spread of the data collected increased after the changes were implemented; the amount of data points decreased, length of the number of days increased, having a positive effect with the report of less accidents and longer time frames within the same length of time frame of five years of the two different distributions. Even though quite a bit of the data is concentrated towards the shorter day lengths of the Years 6-10 distribution, the comparison was improved compared to the greater condensed Years 1-5 distribution.

3. Bags of concrete mix labeled as containing 100 lb have a population mean weight of 100 lb and a population standard deviation of 0.5 lb.

$\mu = 100\text{lb}$
$\sigma = 0.5\text{lb}$

(a) What is the probability that the sample mean of weights of a random sample of 50 bags is less than 99.9 lb?

$n = 50$ bags → big enough sample to assume normally distributed

Standard Error (SE) $= \frac{\sigma}{\sqrt{n}}$

Central Limit Theorem

$\bar{X} \sim \text{Normal}\left(\mu, \frac{\sigma^2}{n}\right)$ where $\bar{X} = \frac{X_1 + X_2 + \ldots + X_n}{n}$

$SE = \frac{0.5}{\sqrt{50}} = 0.707$

$P(\bar{X} < 99.9) = P(Z < \underbrace{(99.9 - 100)/0.707}_{-1.414}) = 0.0793$

Z-score Application

$Z = \frac{\bar{X} - \mu}{SE}$

$\boxed{P(\bar{X} < 99.9) = 0.0793}$

(b) If the population mean weight is increased to 100.15 lb, what is the probability that the sample mean of weights of a sample of size 50 will be less than 100 lb?

$\mu = 100.15$ lb

$n = 50$ → Large enough to assume normal distribution

$P(\bar{X} < 100) = P(Z < \underbrace{(100 - 100.15)/0.707}_{-2.121}) = 0.017$

$SE = \frac{0.5}{\sqrt{50}} = 0.707$

$\boxed{P(\bar{X} < 100) = 0.017}$

(c) What is the standard error of the sampling distribution of the sample mean in this case?

Standard Error

$SE = \frac{\sigma}{\sqrt{n}}$

$SE = \frac{0.5}{\sqrt{50}} = 0.707$

$\boxed{SE = 0.707}$

It was the same for part (a) and (b). Even though the population mean weight was different for each part.

4. A major bottler of soft drinks historically averages ~~0.5%~~ 5% under-fills. Consider a random sample of 2000 bottles.

   (a) Check whether the sample size is large enough to use the normal approximation to the binomial distribution in this case.

   $n = 2000$ "Typically, we consider n to be large enough if both, np and $n(1-p)$ are at least 15"

   $p = 0.05$

   $np = (2000)(0.05) = 100 > 15$

   $n(1-p) = 2000(1-0.05) = 1900 > 15$

   Since both are greater than 15, we can use the normal approximation ✓

   (b) Using the normal approximation, find the probability that 10 or fewer bottles are under-filled. $11/2000 = 0.0055$

   -9.13

   $P(\bar{X} \leq 10) = P(\bar{X} < 11) = P(Z < (0.0055 - 0.05)/0.0049) = 0$

   $p = 0.05$

   $\boxed{P(\bar{X} \leq 10) = 0.00}$

   $SE = \sqrt{p(1-p)/n}$

   $= \sqrt{0.05(1-0.05)/2000} = 0.0049$ (Standard Error)

   (c) Using the normal approximation, find the probability that 15 or more are under-filled. $14/2000 = 0.007$

   $P(\bar{X} \geqslant 15) = P(\bar{X} > 14) = P(Z > (0.007 - 0.05)/0.0049) = 1$

   -8.823

   $\boxed{P(\bar{X} \geqslant 15) = 1.0}$