HW3- Solution

8.1 (a) Naturally, the first digit of each number goes to the stem, and the second digit goes to the leaf. The leaf on the left represents the blocked intrusion attempts during the first 14 days, and the leaf on the right represents 20 days after the change.

	2	1
7 8	3	2 2 3 5 6 6 7 8 9 9
3 3 7 9	4	3 4 5 5 6 8 9
0 0 4 6 6 8 9	5	3 3
0	6	

(b) Before the change of firewall settings,

$$\min(X) = 37,\ \hat{Q}_1 = 43,\ \hat{M} = 50,\ \hat{Q}_3 = 56,\ \max(X) = 60.$$

The sample interquartile range is $56 - 43 = 13$, all the data are within $1.5(\widehat{IQR})$ from $\hat{Q}_1$ or $\hat{Q}_3$, and thus, we do not suspect any outliers.

After the change,

$$\min(X) = 21,\ \hat{Q}_1 = 35,\ \hat{M} = 39,\ \hat{Q}_3 = 46,\ \max(X) = 53.$$

In fact, according to Definition 8.7, any number between 35 and 36 is the first quartile, and any number between 45 and 46 is the third quartile. Next, $\widehat{IQR} = 10$, and again, we see no outliers in this sample.

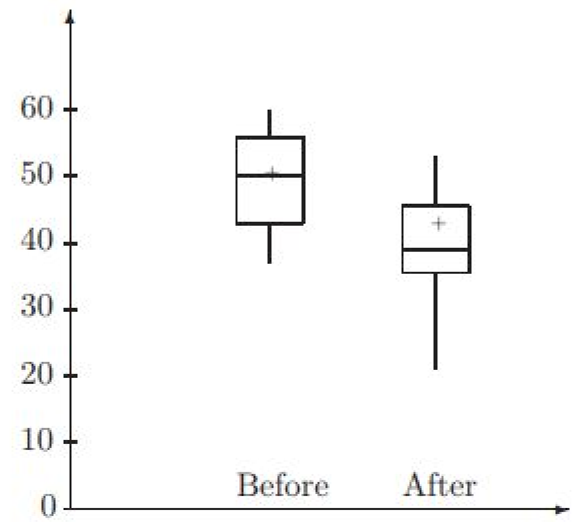
For the boxplots, compute sample means: $\bar{X} = 50$ before the change, and $\bar{X} = 40.2$ after the change. Parallel boxplots are in Figure 4.

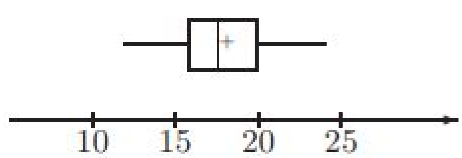
(c) The stem-and-leaf plots and boxplots show that the daily number of intrusion attempts has decreased in general. Every element of the five-number summary has reduced following the change in firewall settings, suggesting that any number of intrusion attempts is now exceeded with a lower probability.

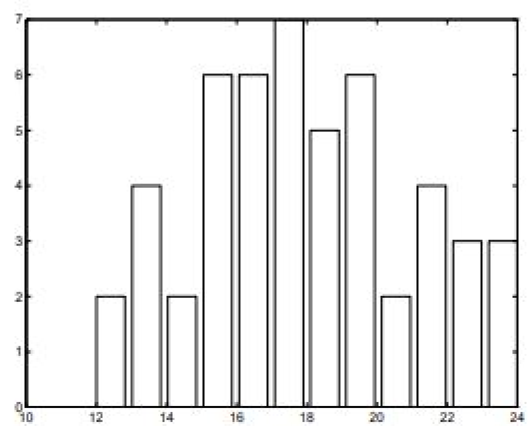
8.2 (a) By Definitions 8.3 and 8.8,

$$\bar{X} = 17.9540, \quad s^2 = 9.9682, \quad s = 3.1573.$$

(b) $s(\bar{X}) = s/\sqrt{n} = 3.1573/\sqrt{50} = \boxed{0.4465}$







(c) Order the data from smallest to largest. The minimum is 11.9, the maximum is 24.1. For the sample of size 50, the median $\hat{M}$ is any number between the 25th and the 26th smallest, i.e., between 17.5 and 17.6. The first quartile $\hat{Q}_1$ is the 13th smallest, i.e., 15.8 (exceeds $22\% \leq 25\%$ of the sample; is exceeded by $74\% \leq 75\%$ of the sample), and the third quartile $\hat{Q}_3$ is the 37th smallest, i.e., 19.9. The five-point summary is

$$(11.9,\ 15.8,\ 17.55,\ 19.9,\ 24.1).$$

The boxplot is on Figure 5.

(d) $\widehat{IQR} = \hat{Q}_3 - \hat{Q}_1 = 4.1$. Compute

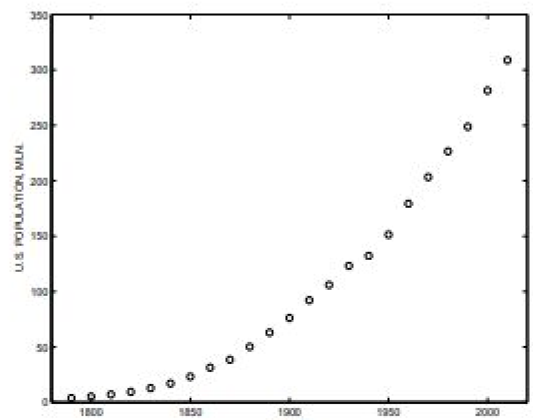
$$\hat{Q}_1 - 1.5(\widehat{IQR}) = 9.65 \text{ and } \hat{Q}_3 + 1.5(\widehat{IQR}) = 26.05.$$

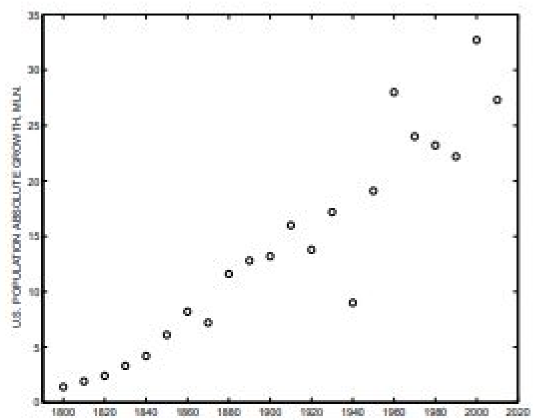
Both $\min(X) = 11.9$ and $\max(X) = 24.1$ are within the range $[9.65, 26.05]$, hence the sample has no obvious outliers.

(e) The histogram in Figure 5 does not have a bell shape and does not support the assumption of a Normal distribution.

8.4 Standard Normal quartiles equal about ± 0.675 (Table A4), and the Standard Normal IQR is $2(0.675) = 1.35$.

For a Normal(μ, σ) random variable, the quartiles equal $\mu \pm 0.675\sigma$, and the interquartile range equals 1.35σ. The probability to take a value within 1.5 interquartile ranges





from its quartiles equals

$$
\begin{aligned}
& \boldsymbol{P}\left\{Q_1-1.5(IQR) \leq X \leq Q_3+1.5(IQR)\right\} \\
& = \boldsymbol{P}\left\{-0.675-1.5(1.35) \leq \frac{X-\mu}{\sigma} \leq 0.675+1.5(1.35)\right\} \\
& = \boldsymbol{P}\{-2.70 \leq Z \leq 2.70\} \\
& = \Phi(2.70)-\Phi(-2.70)=0.9965-0.0035=\boxed{0.9930}
\end{aligned}
$$

8.7 (a) The sample mean is 0.22, the sample median is 0.21, and the variance is 0.01. That is, the U.S. population has increased by 22%, on the average, every 10 years.

(b) The time plot is on Figure 7. We see that in general, the proportional population change decreases, and the trend is almost linear. In 1800–1860, the population increased by about 35% each decade whereas its relative growth never exceeded 20% since 1920.

(c) An increasing trend of the absolute increments in Exercise 8.6 corresponds to a decreasing trend of the relative increments in Exercise 8.7. We should expect a rather strong *negative correlation*: large absolute increments correspond to small relative increments, and vice versa.

Indeed, the correlation coefficient equals −0.79.

A steady reduction of relative increments is not surprising. Had the population

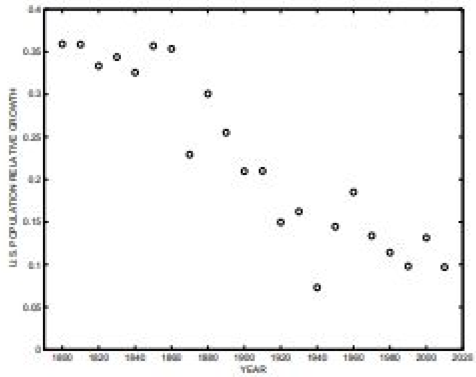


FIGURE 7: Time plot for Exercise 8.7

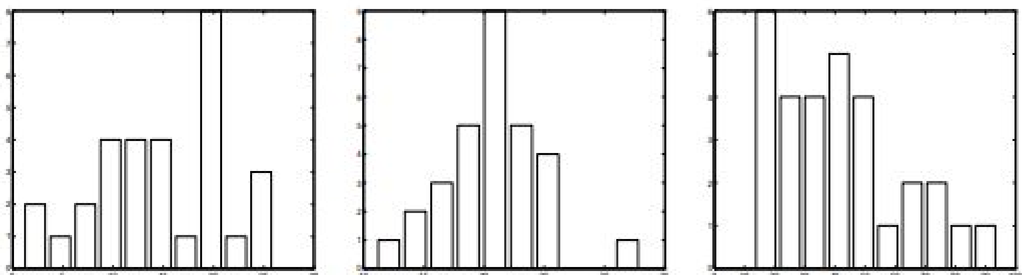


FIGURE 8: Histograms of three data sets in Exercise 8.8

increased by a constant percent every decade, it would have grown exponentially fast. However, the human population does not grow at an exponential rate.

On the other hand, had the population increased by a constant absolute number every decade, it would have grown linearly. The increments increase, therefore the population grows faster than a linear function.

Thus, increasing absolute increments and decreasing relative increments show that the U.S. population growth is faster than linear but slower than exponential. As we'll see in Section 11.3, a *quadratic* model is rather accurate for the U.S. population.

8.9 (a) $\bar{X} = 49.6$, $\hat{M} = 47.5$, $\hat{Q}_1 = 43$, $\hat{Q}_3 = 52$, and $s = 23.5$. For a sample of size 10, any number between the 5th and the 6th smallest observations is a median, the 3rd smallest observation is $\hat{Q}_1$, and the 3rd largest observation is $\hat{Q}_3$.

(b) Compute $\widehat{IQR} = \hat{Q}_3 - \hat{Q}_1 = 9$, $\hat{Q}_1 - 1.5(\widehat{IQR}) = 29.5$, and $\hat{Q}_3 + 1.5(\widehat{IQR}) = 65.5$. One observation, 105, is far outside of the 1.5(IQR) range. So many registered new accounts during the sixth day is likely an outlier.

(c) Without 105, we have $\bar{X} = 43.4$, $\hat{M} = 45$, $\hat{Q}_1 = 43$, $\hat{Q}_3 = 51$, and $s = 13.9$.

Now, for a sample of size 9, the median is the 5th smallest observation, the 3rd smallest is $\hat{Q}_1$, and the 3rd largest observation is $\hat{Q}_3$.

(d) When we deleted the outlier, the mean and the standard deviation decreased significantly. Especially the standard deviation because variation of the data is much smaller without $x_6 = 105$. Quartiles did not change that much. They are *robust* measures, not very sensitive to outliers.