STAT509 Homework 5 Solution

1. (a) population : All water wells in Texas

(b) See graph.

(c) $\hat{\mu} = \bar{x} = 9.735294$

$\hat{\sigma} = s = 11.52782$

(d) Based on the shape of the histogram and boxplot, exponential distribution might be a reasonable model, because the histogram and boxplot both show skewness to the right and the shape of the histogram resembles the density function of an exponential distribution.

(e) Easy way : $\hat{P}(X>20) = \dfrac{\text{number of wells in sample where arsenic levels exceed 20 ppb}}{\text{sample size } (n=102)}$

$= 6/102 = 0.0588$

Interesting way : Fit an exponential model, where $\hat{\lambda}^{-1} = \hat{\mu} = 9.735294$, then $\hat{\lambda} = 0.102719$.

$$\hat{P}(X>20) = 1 - \hat{P}(X \le 20)$$
$$= 1 - (1 - e^{-0.102719 * 20})$$
$$= e^{-0.102719 \times 20}$$
$$= 0.1281723$$

2. (a) See graph.

$\mu \pm \sigma \Rightarrow [125 - 15, 125 + 15] \Rightarrow [110, 140]$

68% of the population have SBP between 110 and 140.

$\mu \pm 2\sigma \Rightarrow [125 - 15 \times 2, 125 + 15 \times 2] \Rightarrow [95, 155]$

95% of the population have SBP between 95 and 155.

$\mu \pm 3\sigma \Rightarrow [125 - 15 \times 3, 125 + 15 \times 3] \Rightarrow [80, 170]$

99.7% of the population have SBP between 80 and 170.

(b) $P(X > 140) = 0.1586553$ See R-code attached

$X \sim N(125, 15^2)$

(c) $P(\bar{X} > 140) = 2.866516 \times 10^{-7}$ See R-code attached.

$X \sim N(125, 15^2) \Rightarrow \bar{X} \sim N(125, 15^2/n) = N(125, 9)$

Because the standard deviation of $\bar{X}$ is different from the standard deviation of X.

3. (a) See graph.

Normal distribution might be a reasonable model, because the histogram and boxplot show symmetric pattern and the shape of the histogram resembles normal distribution.

Population : All shock absorbers from the specific brand.

(b) $\bar{y} = 17006.84$ See R-code

$S = 6494.654$

(c)
$$t = \frac{\bar{y} - \mu}{s/\sqrt{n}}$$
$$= \frac{17006.84 - 18000}{6494.654/\sqrt{38}}$$
$$= -0.9426598$$

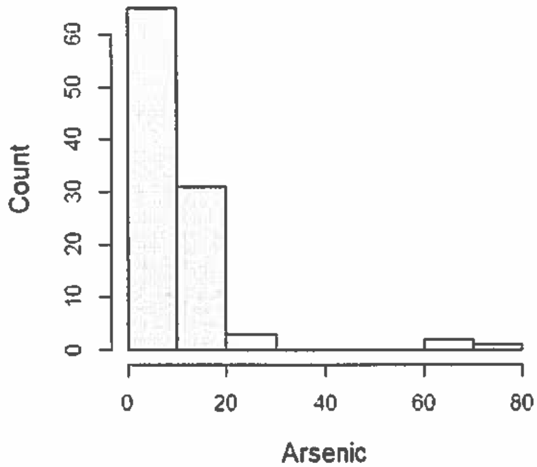
See graph.

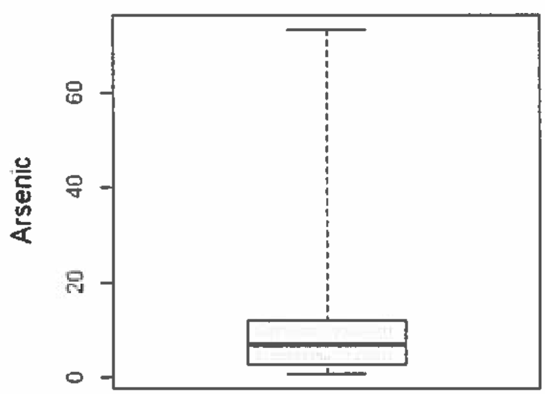
No, the t statistic is not an unusual value. It suggests that we shouldn't suspect quality assessment team's claim that mean distance to failure is 18,000 km.

(d) Yes. See graph.

qq plot suggests that normal distribution seems a reasonable model, because it shows linear pattern. It doesn't affect my analysis in part (c). t-distribution is robust to normality assumption, so the underlying assumption of normality is not an absolute requirement.

1. (b)





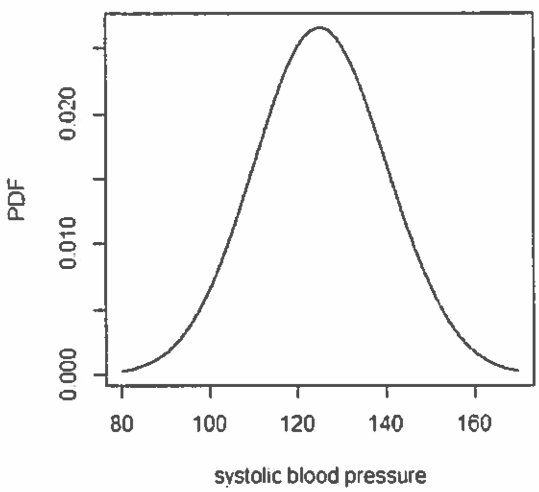

```
arsenic = c(17.6,10.4,13.5,4,19.9,16,12,12.2,11.4,12.7,3,10.3,21.4,19.4,9,6.
5,10.1,8.7,9.7,6.4,
            9.7,63,15.5,10.7,18.2,7.5,6.1,6.7,6.9,0.8,73.5,12,28,12.6,9.4,6.
2,15.3,7.3,10.7,15.9,
            5.8,1,8.6,1.3,13.7,2.8,2.4,1.4,2.9,13.1,15.3,9.2,11.7,4.5,1,1.2,
0.8,1,2.4,4.4,2.2,2.9,
            3.6,2.5,1.8,5.9,2.8,1.7,4.6,5.4,3,3.1,1.3,2.6,1.4,2.3,1,5.4,1.8,
2.6,3.4,1.4,10.7,18.2,
            7.7,6.5,12.2,10.1,6.4,10.7,6.1,0.8,12,28.1,9.4,6.2,7.3,9.7,62.1,1
5.5,6.4,9.5)
> # Create histogram
> hist(arsenic,xlab="Arsenic",xlim=c(0,80),ylab="Count",main="",col="lightblu
e",cex.lab=1.25)
> # Create boxplot
> # Use range=1.5 to detect outliers
```

```
> boxplot(arsenic,range=0,xlab="",ylab="Lifetime (in hours)",col="lightblue",
cex.lab=1.25)
> quantile(arsenic)
  0%  25%  50%  75% 100%
 0.8  2.9  7.1 12.0 73.5
```

(c)

```
> mean(arsenic) ## sample mean
[1] 9.735294
> sd(arsenic) ## sample standard deviation
[1] 11.52782
```

2. (a)




```
> y = seq(80,170,0.01)
> # Plot population distribution
> plot(y,dnorm(y,125,15),type="l",lty=1,xlab="systolic blood pressure",ylab="PDF")
```

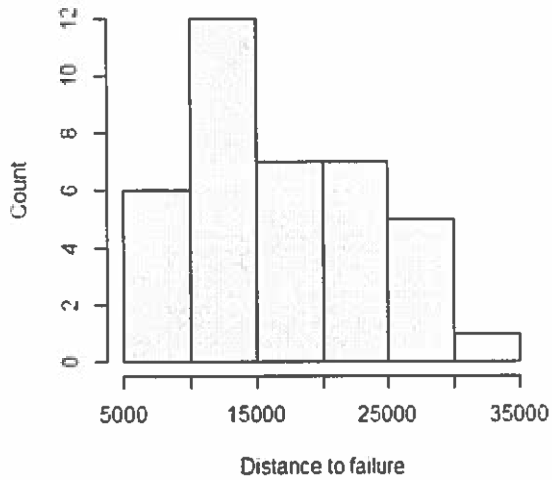
(b)

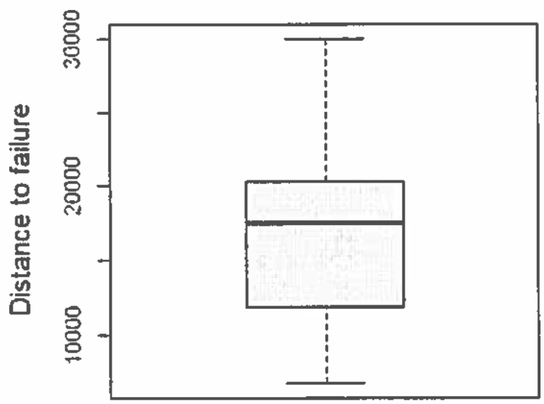
```
> 1-pnorm(140,125,15)
[1] 0.1586553
```

(c)

```
> 1-pnorm(140,125,3)
[1] 2.866516e-07
```

3. (a)





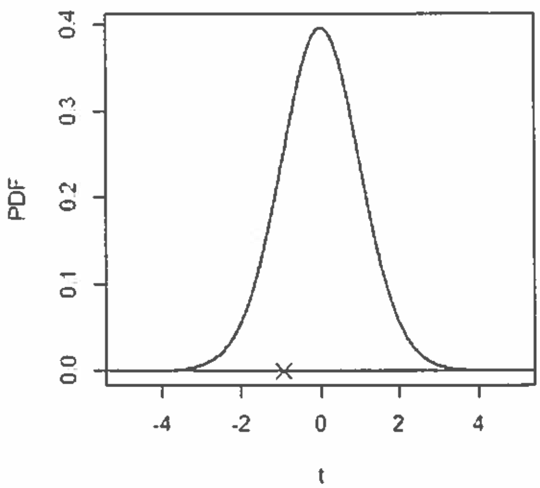

```
distance=c(6700, 6950, 7820, 9120, 9660, 9820, 11310, 11690, 11850, 11880,
12140, 12200, 12870, 13150, 13330 ,13470 ,14040, 14300, 17520, 17540,
17890 ,18450, 18960, 18980 ,19410 ,20100, 20100, 20150, 20320, 20900,
22700 ,23490 ,26510, 27410 ,27490, 27890, 28100, 30050)

> hist(distance,xlab="Distance to failure",xlim=c(5000,35000),ylab="Count",main="",co
l="lightblue")
> # Create boxplot
> # Use range=1.5 to detect outliers
> boxplot(distance,range=0,xlab="",ylab="Distance to failure",col="lightblue",cex.lab
=1.25)
```

(b)

```
> mean(distance)
[1] 17006.84
```

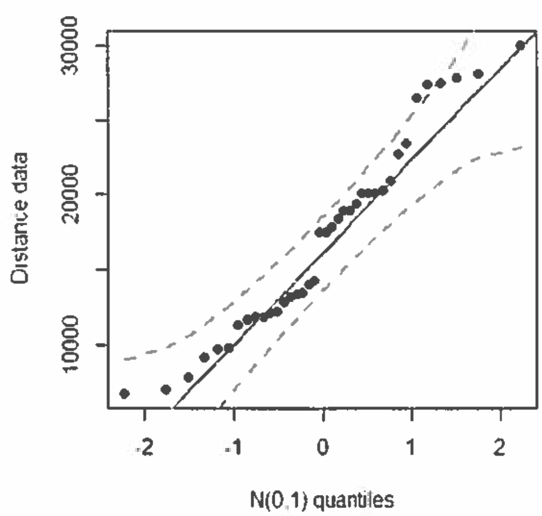
```
> sd(distance)
[1] 6494.654
```

(c)




```
> t = seq(-5,5,0.001)
> pdf = dt(t,37)
> plot(t,pdf,type="l",lty=1,xlab="t",ylab="PDF")
> abline(h=0)
> # Add points
> points(x=-0.9426598,y=0,pch=4,cex=1.5)
```

(d)




```
>qqPlot(distance,distribution="norm",xlab="N(0,1) quantiles",ylab="Distance d
ata",pch=16)
```