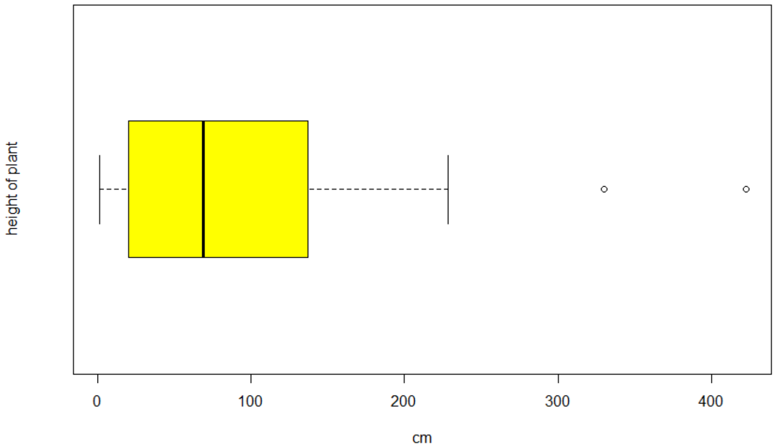
Jiayu Wang 1039580

Q1

(a)

```
> A <- c(173.1, 61.5, 123.3, 100.4, 20.4, 20.9, 228.4, 1.0, 6.8,
+                  11.4, 7.7, 40.7, 15.8, 422.4, 58.2, 19.9, 38.8, 121.0,
+                  118.6, 174.9, 87.2, 14.0, 204.7, 81.9,
+                  57.3, 177.0, 14.1, 137.0, 76.4, 330.2)
> boxplot(A, horizontal = TRUE, xlab = c("cm"), ylab = c("height of plant"),
+         col = c("yellow"))
```




```
> summary(A)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1.00   20.02   68.95   98.17  133.57  422.40
```

So for center, since the graph is screwed, I'd better use median (68.95) to describe center;

For shape, it is asymmetric, right-skewed;

For spread, its range is 422.40 – 1.00 = 421.40, IQR is 133.57 – 20.02 = 113.55

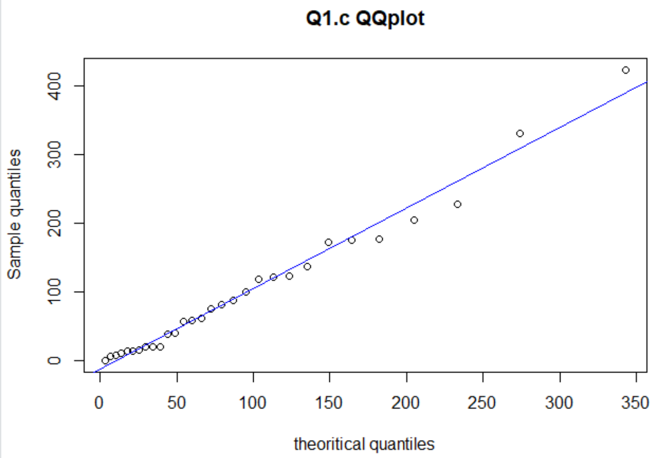
(b)

```
> library(MASS)
> exponential.fit <- fitdistr(A, "exponential")
>
> theta.hat <- 1/exponential.fit$estimate
> theta.hat
    rate
98.16667
> mean(A)
[1] 98.16667
```

Here we can know maximum likelihood estimate for theta is 98.16667 (which has the same value of the mean).

(c)

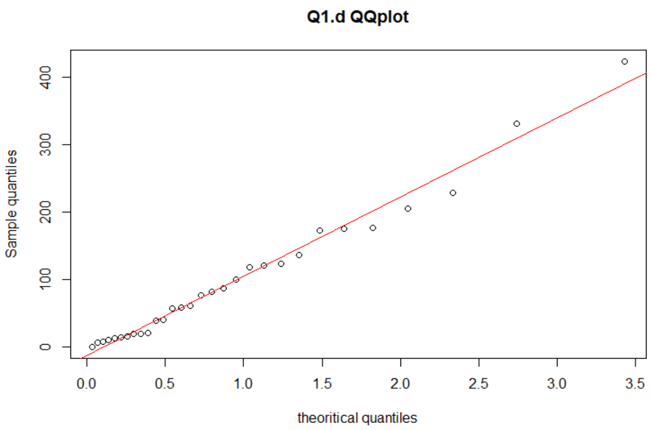
```
> p <- (1:30)/31
> qsam <- quantile(A, p, type = 6)
> qtheo <- qexp(p, 1/100)
> plot(qtheo, qsam, ylab = "Sample quantiles", xlab = "theoritical quantiles", main
 = "Q1.c QQplot")
> fit <- lm(qsam ~ qtheo)
> abline(fit, col = "blue")
```



The model is relatively fit to the data, since all the points can be considered as falling on a line.

(d)

```
> qtheo2 <- qexp(p,1)
> plot(qtheo2, qsam, ylab = "Sample quantiles", xlab = "theoritical quantiles",
 main = "Q1.d QQplot")
> fit2 <- lm(qsam ~ qtheo2)
> abline(fit2, col = "red" )
```



The approach will work. The slope will be approximately 100 and intercrept will be approximately 0.

Q2

(a) ① $f(x|\mu,\lambda)=\frac{1}{x\sqrt{2\pi\lambda}}\exp\left\{-\frac{(\ln x-\mu)^2}{2\lambda}\right\},\ x>0$

$$L(\mu)=\prod_{i=1}^{n}\frac{1}{x_i\sqrt{2\pi\lambda}}\exp\left\{-\frac{(\ln x_i-\mu)^2}{2\lambda}\right\}$$

$$\ln L(\mu)=-\frac{1}{2}n\ln(2\pi\lambda)-\frac{1}{2\lambda}\sum_{i=1}^{n}(\ln x_i-\mu)^2$$

$$\frac{d\ln L(\mu)}{d\mu}=\frac{1}{\lambda}\sum_{i=1}^{n}(\ln X_i-\mu)$$

setting $\frac{d\ln L(\mu)}{d\mu}=\frac{1}{\lambda}\sum_{i=1}^{n}(\ln X_i-\mu)=0$ and solving gives:

$$\sum_{i=1}^{n}\ln X_i-n\mu=0\Rightarrow\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}\ln X_i$$

② $f(x|\mu,\lambda)=\frac{1}{x\sqrt{2\pi\lambda}}\exp\left\{-\frac{(\ln x-\mu)^2}{2\lambda}\right\},\ x>0$

$$L(\lambda)=\prod_{i=1}^{n}\frac{1}{x_i\sqrt{2\pi\lambda}}\exp\left\{-\frac{(\ln x_i-\mu)^2}{2\lambda}\right\}$$

$$\ln L(\lambda)=-\frac{1}{2}n\cdot\ln(2\pi\lambda)-\frac{1}{2\lambda}\sum_{i=1}^{n}(\ln X_i-\mu)^2$$

$$\frac{d\ln L(\lambda)}{d\lambda}=-\frac{1}{2}n\frac{1}{\lambda}+\frac{1}{2\lambda^2}\sum_{i=1}^{n}(\ln X_i-\mu)^2$$

setting $\frac{d\ln L(\lambda)}{d\lambda}=-\frac{1}{2}n\frac{1}{\lambda}+\frac{1}{2\lambda^2}\sum_{i=1}^{n}(\ln X_i-\mu)^2=0$, solving gives:

$$\hat{\lambda}=\frac{1}{n}\sum_{i=1}^{n}(\ln X_i-\hat{\mu})^2$$

(b) i. let $Y_i=\ln X_i$, then $\hat{\mu}=\frac{1}{n}\sum_{i=1}^{n}\ln X_i=\frac{1}{n}\sum_{i=1}^{n}Y_i=\bar{Y}$

hense $\hat{\lambda}=\frac{1}{n}\sum_{i=1}^{n}(\ln X_i-\hat{\mu})^2=\frac{1}{n}\sum_{i=1}^{n}(Y_i-\bar{Y})^2$, and becanse $s^2=\frac{1}{n-1}\sum_{i=1}^{n}(Y_i-\bar{Y})^2$

$\Rightarrow\hat{\lambda}=\frac{n-1}{n}s^2$

since $\frac{n-1}{\sigma^2}\cdot s^2=\frac{n-1}{\lambda}s^2\sim\chi^2_{n-1}\Rightarrow Var\left[\frac{n-1}{\lambda}\cdot s^2\right]=2(n-1)$

$\Rightarrow Var\left(\frac{n\hat{\lambda}}{\lambda}\right)=\frac{n^2}{\lambda^2}Var(\hat{\lambda})=2(n-1)\Rightarrow Var(\hat{\lambda})=\frac{2\lambda^2(n-1)}{n^2}$

thus, $sd(\hat{\lambda})=\sqrt{Var(\hat{\lambda})}=\frac{\lambda}{n}\sqrt{2(n-1)}$

ii since $\frac{n-1}{\lambda}\cdot s^2=\frac{n\hat{\lambda}}{\lambda}\sim\chi^2_{n-1}$

$\Rightarrow Pr\left(F^{-1}(\frac{\alpha}{2})\le\frac{n\hat{\lambda}}{\lambda}\le F^{-1}(1-\frac{\alpha}{2})\right)=1-\alpha$, and then rearrange as usual,

hence $100(1-\alpha)\%$ CI for λ is $\left(\frac{n\hat{\lambda}}{F^{-1}(1-\frac{\alpha}{2})},\frac{n\hat{\lambda}}{F^{-1}(\frac{\alpha}{2})}\right)$, here F^{-1} is regarding χ^2_{n-1}

(c)

i.

```
> B <- c(12.9, 2.3, 2.4, 65.0, 6.7, 1.8, 1.5, 1.7, 248.7, 1.0, 2.0, 4.9, 3.6, 4.1, 6.8)
> mu.hat <- sum(log(B)) / 15
> lambda.hat <- sum((log(B) - mu.hat)^2) / 15
> # according to q2(b)i,we can know:
> se.lambda.hat <- lambda.hat * sqrt(2*(15-1))/15
> se.lambda.hat
[1] 0.7360108
```

ii.

```
> lambda.hat
[1] 2.086394
> upperbound <- 15 * lambda.hat / qchisq(p = 0.025, df = 15 -1 )
> lowerbound <- 15 * lambda.hat / qchisq(p = 0.975, df = 15 -1 )
> upperbound
[1] 5.560036
> lowerbound
[1] 1.198207
```

(for 95% CI)

Q3

(a) i. $E(X) = \frac{0+\theta}{2} = \frac{\theta}{2} \Rightarrow \theta = 2 \cdot E(X) \Rightarrow \tilde{\theta} = 2\bar{X}$

$E(\tilde{\theta}) = E(2\bar{X}) = 2E(\bar{X}) = 2E(X) = \theta$

$Var(\tilde{\theta}) = Var(2\bar{X}) = 2^2 Var(\bar{X}) = 4\frac{Var(X)}{n} = \frac{4}{n}\left(\frac{\theta^2}{12}\right) = \frac{\theta^2}{3n}$

since a single observation, which means $n=1$, then:

$\tilde{\theta} = 2X$

$E(\tilde{\theta}) = \theta$

$Var(\tilde{\theta}) = \frac{\theta^2}{3}$

ii $f(x|\theta) = \frac{1}{\theta - 0} = \frac{1}{\theta}$, $0 \le x \le \theta$

$L(\theta) = \prod_{i=1}^{n} f(x|\theta) = \left(\frac{1}{\theta}\right)^n$

$\ln L(\theta) = -n\ln\theta$

$\frac{d\ln L(\theta)}{d\theta} = -\frac{n}{\theta} < 0$

So $L(\theta)$ is a deceasing function for $\theta \ge x_{(n)}$, and $L(\theta)$ is maximized at $\theta = x_{(n)}$, hence the MLE for θ is given by

$\hat{\theta} = X_{(n)}$

$E(\hat{\theta}) = E(X_{(n)}) = \frac{n\theta}{n+1}$

$Var(\hat{\theta}) = Var(X_{(n)}) = E(X_{(n)}^2) - E^2(X_{(n)}) = \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2}$

since a single observation, which means $n=1$, then:

$\hat{\theta} = X$

$E(\hat{\theta}) = \frac{\theta}{2}$

$Var(\hat{\theta}) = \frac{\theta^2}{3} - \frac{\theta^2}{4} = \frac{\theta^2}{12}$

(b) i. According to Q3 (a) ii, $Var(\hat{\theta}) = \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2}$

$bias(\hat{\theta}) = E(\hat{\theta}) - \theta \Rightarrow bias(\hat{\theta})^2 = E^2(\hat{\theta}) - 2\theta \cdot E(\hat{\theta}) + \theta^2$

$MSE(\hat{\theta}) = E[(\hat{\theta} - \theta)^2] = E[\hat{\theta}^2 - 2\hat{\theta}\theta + \theta^2] = E[\hat{\theta}^2] - 2\theta E[\hat{\theta}] + E[\theta^2]$

$= E[\hat{\theta}^2] - 2\theta E[\hat{\theta}] + \theta^2$

$Var(\hat{\theta}) = E[\hat{\theta}^2] - E^2[\hat{\theta}]$

hence $Var(\hat{\theta}) + bias(\hat{\theta})^2 = E(\hat{\theta}^2) - E^2[\hat{\theta}] + E^2(\hat{\theta}) - 2\theta E(\hat{\theta}) + \theta^2$

$= E(\hat{\theta}^2) - 2\theta E(\hat{\theta}) + \theta^2$

$= MSE(\hat{\theta})$

ii. for MME: $MSE(\tilde{\theta}) = Var(\tilde{\theta}) + bias(\tilde{\theta})^2 = Var(\tilde{\theta}) + (E(\tilde{\theta}) - \theta)^2 = \frac{\theta^2}{3} + (\theta - \theta)^2 = \frac{\theta^2}{3}$

for MLE: $MSE(\hat{\theta}) = Var(\hat{\theta}) + bias(\hat{\theta})^2 = Var(\hat{\theta}) + (E(\hat{\theta}) - \theta)^2 = \frac{\theta^2}{12} + (\frac{\theta}{2} - \theta)^2 = \frac{\theta^2}{3}$

iii Accoring to above answers, $\tilde{\theta} = 2X$, $\hat{\theta} = X$

Setting $\hat{\theta}_2 = \frac{\hat{\theta} + \tilde{\theta}}{2} = \frac{3}{2}X$

hence $E(\hat{\theta}_2) = E(\frac{3}{2}X) = \frac{3}{2}E(X) = \frac{3}{2} \times \frac{\theta}{2} = \frac{3}{4}\theta$

$Var(\hat{\theta}_2) = Var(\frac{3}{2}X) = \frac{9}{4}Var(X) = \frac{9}{4} \times \frac{\theta^2}{12} = \frac{3}{16}\theta^2$

$MSE(\hat{\theta}_2) = Var(\hat{\theta}_2) + bias(\hat{\theta}_2)^2 = \frac{3}{16}\theta^2 + (\frac{3}{4}\theta - \theta)^2 = \frac{1}{4}\theta^2 < MSE(\tilde{\theta}) = MSE(\hat{\theta})$

(c)

i. for MME:

$E(X) = \frac{0+\theta}{2} = \frac{\theta}{2} \Rightarrow \theta = 2 \cdot E(X) \Rightarrow \tilde{\theta} = 2\bar{X}$

$E(\tilde{\theta}) = E(2\bar{X}) = 2E(\bar{X}) = 2E(X) = \theta$

$Var(\tilde{\theta}) = Var(2\bar{X}) = 2^2 Var(\bar{X}) = 4\frac{Var(X)}{n} = \frac{4}{n}(\frac{\theta^2}{12}) = \frac{\theta^2}{3n}$

$MSE(\tilde{\theta}) = Var(\tilde{\theta}) + bias(\tilde{\theta})^2 = \frac{\theta^2}{3n} + [\theta - \theta]^2 = \frac{\theta^2}{3n}$

ii. for MLE:

$f(x|\theta) = \frac{1}{\theta - 0} = \frac{1}{\theta}$, $0 \le x \le \theta$

$L(\theta) = \prod_{i=1}^{n} f(x|\theta) = (\frac{1}{\theta})^n$

$\ln L(\theta) = -n \ln\theta$

$\frac{d \ln L(\theta)}{d\theta} = -\frac{n}{\theta} < 0$

So $L(\theta)$ is a deceasing function for $\theta \ge X_{(n)}$, and $L(\theta)$ is maximized at $\theta = X_{(n)}$, hence the MLE for θ is given by $\hat{\theta} = X_{(n)}$

$E(\hat{\theta}) = E(X_{(n)}) = \frac{n\theta}{n+1}$

$Var(\hat{\theta}) = Var(X_{(n)}) = E(X_{(n)}^2) - E^2(X_{(n)}) = \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2} = \frac{n\theta^2}{(n+2)(n+1)^2}$

$MSE(\hat{\theta}) = Var(\hat{\theta}) + bias(\hat{\theta})^2 = \frac{n\theta^2}{n+2} - \frac{n^2\theta^2}{(n+1)^2} + (\frac{n\theta}{n+1} - \theta)^2 = \frac{2\theta^2}{(n+2)(n+1)}$

iii $E(a\hat{\theta}) = aE(\hat{\theta}) = \frac{an\theta}{n+1}$, $Var(a\hat{\theta}) = a^2 Var(\hat{\theta}) = \frac{a^2 n\theta^2}{(n+2)(n+1)^2}$

$MSE(a\hat{\theta}) = Var(a\hat{\theta}) + bias(a\hat{\theta})^2 = \frac{a^2 n\theta^2}{(n+2)(n+1)^2} + (\frac{an\theta}{n+1} - \theta)^2$

setting $\frac{d\,MSE(a\hat{\theta})}{da} = 2a\frac{n\theta^2}{(n+2)(n+1)^2} + 2a\frac{n^2\theta^2}{(n+1)^2} - \frac{2n\theta^2}{n+1} = 0$

solving and gives:

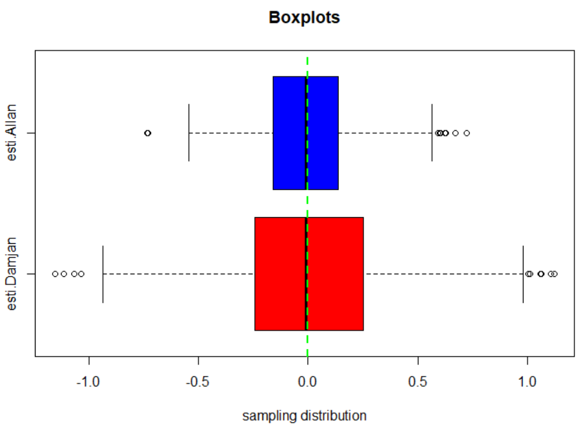
$a = \frac{n+2}{n+1}$

hence $a = \frac{n+2}{n+1}$ minimises MSE

Q4.

```
> num.simulation <- 1000
> samplesize <- 20
> esti.Damjan <- 1 : num.simulation
> esti.Allan <- 1 : num.simulation
> for (i in 1 : num.simulation) {
+   sample <- rnorm(samplesize)
+   esti.Damjan[i] <- (min(sample) + max(sample)) / 2
+   esti.Allan[i] <- mean(sample)
+ }
> bias.Damjan <- mean(esti.Damjan) - 0
> bias.Allan <- mean(esti.Allan) - 0
> var.Damjan<- var(esti.Damjan)
> var.Allan <- var(esti.Allan)
> bias.Damjan
[1] 0.008845386
> bias.Allan
[1] -0.003088627
> var.Damjan
[1] 0.1337073
> var.Allan
[1] 0.04974547
```

```
> boxplot(esti.Damjan,esti.Allan, col = c("red","blue"), xlab = "sampling distribution",
+     names = c("esti.Damjan","esti.Allan"), main = "Boxplots", horizontal = TRUE)
> abline(v = 0, col = "green", lty = 2, lwd = 2)
```



Q5

(a). since X_1, X_2, X_3 are independent with each other

$$E[T_1] = E[\frac{1}{4}(X_1+X_2)+\frac{1}{2}X_3] = \frac{1}{4}E(X_1)+\frac{1}{4}E(X_2)+\frac{1}{2}E(X_3) = \frac{1}{4}u+\frac{1}{4}u+\frac{1}{2}u = u$$

hence T_1 is unbiased

$$E[T_2] = E[\frac{1}{3}(X_1+2X_2+3X_3)] = \frac{1}{3}E(X_1+2X_2+3X_3) = \frac{1}{3}E(X_1)+\frac{2}{3}E(X_2)+E(X_3)$$

$$= \frac{1}{3}u+\frac{2}{3}u+u = 2u$$

hence T_2 is biased

$$E[T_3] = E[\frac{1}{3}(X_1+X_2+X_3)] = \frac{1}{3}E(X_1+X_2+X_3) = \frac{1}{3}E(X_1)+\frac{1}{3}E(X_2)+\frac{1}{3}E(X_3) = \frac{1}{3}u\times 3 = u$$

hence T_3 is unbiased

$$E[T_4] = E[\frac{1}{2}(X_1+X_2)+\frac{1}{4}X_3^2] = \frac{1}{2}E(X_1)+\frac{1}{2}E(X_2)+\frac{1}{4}E(X_3^2) = \frac{1}{2}u+\frac{1}{2}u+\frac{1}{4}(u^2+\frac{\sigma^2}{3^2})$$

$$= u+\frac{1}{4}u^2+\frac{1}{36}\sigma^2 \neq u$$

hence T_4 is biased

Overall, T_1 and T_3 are unbiased

(b) $Var[T_1] = Var[\frac{1}{4}(X_1+X_2)+\frac{1}{2}X_3] = \frac{1}{16}Var[X_1]+\frac{1}{16}Var[X_2]+\frac{1}{4}Var[X_3] = \frac{1}{16}\sigma^2+\frac{1}{16}\frac{\sigma^2}{2^2}+\frac{1}{4}\frac{\sigma^2}{3^2} \approx 0.1060\sigma^2$

$Var[T_3] = Var[\frac{1}{3}(X_1+X_2+X_3)] = \frac{1}{9}Var[X_1]+\frac{1}{9}Var[X_2]+\frac{1}{9}Var[X_3] = \frac{1}{9}\times(\sigma^2+\frac{\sigma^2}{2^2}+\frac{\sigma^2}{3^2}) \approx 0.1510\sigma^2$

So $Var[T_1] < Var[T_3]$, T_1 has smallest variance

(c) Through comparison of T_1 and T_3 and their variances,

set $T_5 = \frac{1}{6}X_1+\frac{1}{3}X_2+\frac{1}{2}X_3$, (since $\frac{1}{6} < \frac{1}{3}$ and $\frac{1}{6}+\frac{1}{3} = \frac{1}{2}$)

hence $E[T_5] = E[\frac{1}{6}X_1+\frac{1}{3}X_2+\frac{1}{2}X_3] = \frac{1}{6}E[X_1]+\frac{1}{3}E(X_2)+\frac{1}{2}E[X_3] = \frac{1}{6}u+\frac{1}{3}u+\frac{1}{2}u = u$

, which proves T_5 is unbiased

$$Var[T_5] = Var[\frac{1}{6}X_1+\frac{1}{3}X_2+\frac{1}{2}X_3] = \frac{1}{36}Var[X_1]+\frac{1}{9}Var[X_2]+\frac{1}{4}Var[X_3]$$

$$= \frac{1}{36}\sigma^2+\frac{1}{9}\times\frac{\sigma^2}{2^2}+\frac{1}{4}\frac{\sigma^2}{3^2} = \frac{1}{12}\sigma^2 \approx 0.0833\sigma^2 < Var[T_1] < Var[T_3]$$

So estimator $T_5 = \frac{1}{6}X_1+\frac{1}{3}X_2+\frac{1}{2}X_3$ is what I suggest.