ISyE 6739 – Group Activity 9
solutions

1. (a)
```
n = length(data1$Score_Diff)
t.a<-t.test(data1$Score_Diff, alternative = "two.sided", mu = 0, conf.level = 0.95)
t.a$conf.int
```
$\Rightarrow$ 0.95% confidence interval for the mean of score difference is $\big(-5.611, 7.968\big)$. Therefore, we can conclude that on average the score difference is 0 (because $-5.611 < 0 < 7.968$).

   (b) The sample size n_{new} for which the estimation error is 0.1 is

$$n_{new} = \left(\frac{Z_{\alpha/2}s}{E}\right)^2 = \left(\frac{1.96 \cdots 17.5}{0.1}\right)^2 \approx 117778.$$

here s is a standard deviation.

   (c)
```
t.b <- t.test(data1$Attend, alternative = "greater", mu = 0, sigma.x = NULL,
              sigma.y = NULL, conf.level = 0.95)
t.b$conf.int
```
$\Rightarrow$ 0.95% confidence interval for the attendance is $\big(6322.4, +\infty\big)$. It is more appropriate to use a lower bound because we are interested in increasing the attendance. Also, attendance is already bounded from above by the number of seats.

2.
```
data2 <- c(.9765, .9961, 1.0, .9922, .9961, 1.0, .9922, .9843, .9804, 1.0)
df <- length(data2) - 1
L = var(data2) * df / qchisq(0.05/2, df, lower.tail = FALSE)
U = var(data2) * df / qchisq(1 - 0.05/2, df, lower.tail = FALSE)
# CI for the variance
c(lower = L,  upper = U)
```
$\Rightarrow$ 0.95% confidence interval for the variance of the pixel values is $\big(3.47 \cdot 10^{-5}, 2.442 \cdot 10^{-4}\big)$.

```
# CI for the std deviation
c(lower = L^.5,  upper = U^.5)
```

$\Rightarrow$ 0.95% confidence interval for the standard deviation of the pixel values is $\big(0.00589, 0.01562\big)$.

3. Fun size packets of M&M's in general contains 10 candies. There are 1 red and 2 yellow candies in the first packet, 1 red and 1 yellow candies in the second one.

   (a) Point estimates for the proportions of red and yellow candies:

$$\hat{p}_{red} = \frac{1+1}{10+10} = 0.1, \quad \hat{p}_{yellow} = \frac{2+1}{10+10} = 0.15.$$

   (b) The confidence interval for population proportion is

$$\hat{p} - Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}} \le p \le \hat{p} + Z_{\alpha/2}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

Then 95% confidence interval for the proportions of red and yellow candies are respectively

$$-0.0315 \le p_{red} \le 0.2315,$$
$$-0.00649 \le p_{yellow} \le 0.30649.$$

Based on two fun size packets, we can conclude that on average proportion of yellow candies is greater then proportion of red ones.

(c)

$$n_{new} = \left(\frac{Z_{\alpha/2}}{E}\right)^2 \hat{p}_{red}(1-\hat{p}_{red}) \approx 3458$$

$$n_{new} = \left(\frac{Z_{\alpha/2}}{E}\right)^2 \cdot 0.25 \approx 9604.$$

4. $X \sim Exp(\lambda)$, $\mathrm{E}[X] = \frac{1}{\lambda}$, $\mathrm{Var}[X] = \frac{1}{\lambda^2}$. We know that the MLE for parameter λ is $\hat{\lambda} = \frac{1}{\bar{X}}$. Let $\theta = \frac{1}{\lambda}$. Then by the invariance property of MLE

$$\hat{\theta} = \frac{1}{\hat{\lambda}} = \bar{X},$$

$$\mathrm{Var}[\hat{\theta}] = \mathrm{Var}[\bar{X}] = \frac{1}{n\lambda^2}$$

We can approximate $\mathrm{Var}[\hat{\theta}]$ by setting $\lambda \approx \hat{\lambda}$:

$$\mathrm{Var}[\hat{\theta}] = \frac{(\bar{X})^2}{n}.$$

Then the confidence interval for parameter θ is

$$\hat{\theta} - Z_{\alpha/2}\sqrt{\mathrm{Var}[\hat{\theta}]} \leq \theta \leq \hat{\theta} + Z_{\alpha/2}\sqrt{\mathrm{Var}[\hat{\theta}]}$$

$$\bar{X} - Z_{\alpha/2}\frac{\bar{X}}{\sqrt{n}} \leq \frac{1}{\lambda} \leq \bar{X} + Z_{\alpha/2}\frac{\bar{X}}{\sqrt{n}}$$

where s is a sample standard deviation, n is a sample size. $\Rightarrow$ If $\bar{X} - Z_{\alpha/2}\frac{\bar{X}}{\sqrt{n}} > 0$ then the confidence interval for parameter λ is

$$\frac{1}{\bar{X} + Z_{\alpha/2}\frac{\bar{X}}{\sqrt{n}}} \leq \frac{1}{\lambda} \leq \frac{1}{\bar{X} - Z_{\alpha/2}\frac{\bar{X}}{\sqrt{n}}}$$

otherwise

$$\frac{1}{\lambda} \geq \frac{1}{\bar{X} + Z_{\alpha/2}\frac{\bar{X}}{\sqrt{n}}}$$

Now we have that $n = 100$, $\bar{X} = 10$. Then 95% confidence interval for parameter λ is

$$0.0836 \leq \lambda \leq 0.1244.$$

5. $X_i \sim Beta(\alpha, 1)$

$$\Rightarrow \mathrm{E}[X_i] = \frac{\alpha}{\alpha+1}, \quad \mathrm{Var}(X_i) = \frac{\alpha}{(\alpha+1)^2(\alpha+2)}.$$

By MOM:

$$\bar{X} = \frac{\alpha}{\alpha+1}$$

$$\Rightarrow \hat{\alpha} = \frac{\bar{X}}{1-\bar{X}}, \tag{1}$$

$$\widehat{\mathrm{Var(X_i)}} = \frac{\hat{\alpha}}{(\hat{\alpha}+1)^2(\hat{\alpha}+2)} = \frac{\bar{X}(1-\bar{X})^2}{2-\bar{X}}.$$

Then CI for the mean is

$$\hat{X} - Z_{\alpha/2}\sqrt{\frac{\widehat{\mathrm{Var}(X)}}{n}} \leq \frac{\hat{\alpha}}{1+\hat{\alpha}} \leq \hat{X} + Z_{\alpha/2}\sqrt{\frac{\widehat{\mathrm{Var}(X)}}{n}}$$

$$\Rightarrow 1 - \hat{X} + Z_{\alpha/2}\sqrt{\frac{\widehat{\mathrm{Var}(X)}}{n}} \geq \frac{1}{1+\hat{\alpha}} \geq 1 - \hat{X} - Z_{\alpha/2}\sqrt{\frac{\widehat{\mathrm{Var}(X)}}{n}}, \tag{2}$$

If $1-\hat{X}-Z_{\alpha/2}\sqrt{\frac{\widehat{\mathrm{Var}(X)}}{n}}>0$ then CI for parameter α is

$$-1+\frac{1}{1-\hat{X}+Z_{\alpha/2}\sqrt{\frac{\widehat{\mathrm{Var}(X)}}{n}}}\leq\hat{\alpha}\leq-1+\frac{1}{1-\hat{X}-Z_{\alpha/2}\sqrt{\frac{\widehat{\mathrm{Var}(X)}}{n}}}$$

otherwise

$$\hat{\alpha}\geq-1+\frac{1}{1-\hat{X}+Z_{\alpha/2}\sqrt{\frac{\widehat{\mathrm{Var}(X)}}{n}}}$$