Name: Abrar Fauzan Hamzah
ID: 28551494

Assignment 2

Q1.

a. The average mean is 213.1 and by using the function **calcCI()** studio 4 we get the 95% CI which is (172.5715, 412.8785).

#g1 = calcCI(alist, alpha=0.05)

This means that we are certain that the mean could represent the population.

b. The estimated difference in mean between Boston not having house (sample size = 8) or having house on Charles River (sample size = 8) is -79.625. We are 95% confident that the mean is between -181.91919 up to 22.66919. As it includes zero, it means that we cannot assume that there is no difference in mean of population between having house or not on Charles River.

```
alist <- c(270.5, 230.9, 180.2, 230.3, 210.6, 200.6, 160.8, 220.9)
n1 <- length(alist)

g1 = calcCI(alist, alpha=0.05)
g1

# Q1b
blist <- c(290.0, 500.0, 500.0, 210.7, 130.4, 330.1, 230.3, 150.3)
n2 <- length(blist)
g2 = calcCI(blist, alpha=0.05)
g2

diff = g1$mu.hat - g2$mu.hat

se.diff = sqrt(g1$sigma2.hat/n1 + g2$sigma2.hat/n2)
se.diff
CI.diff = diff + c(-1.96*se.diff, 1.96*se.diff)
```

Name: Abrar Fauzan Hamzah
ID: 28551494

c. Just by looking from **diff** and **se.diff** value, we could guest that it's likely to be a weak evidence against the null hypothesis because there is no significant difference between diff and the standard error differences. Also, **p** is not small enough to be rejected (0.1270974). Therefore, we can conclude that we can accept the null hypothesis which means that the difference of means is likely to be the same.

```
# H0: θx= θy
# Ha: θx≠ θy
z = diff/se.diff
p = 2*pnorm(-abs(z))
```

Q2.

a.

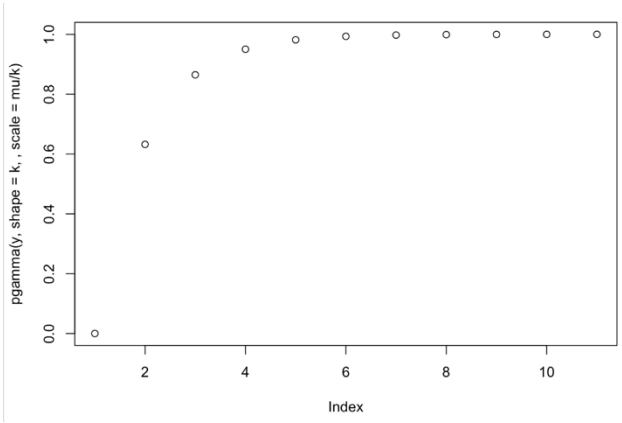

Fig1: mu=1 & k=1

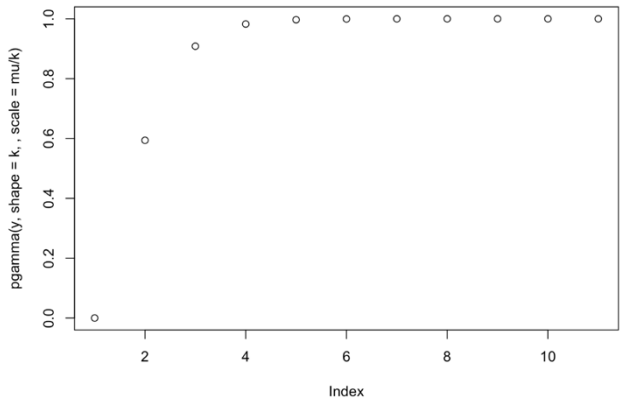

Fig2: mu=1 & k=2

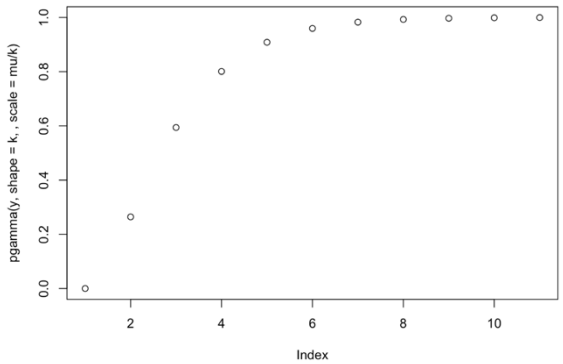

Fig2: mu=2 & k=2

Name: Abrar Fauzan Hamzah
ID: 28551494

b. $\prod_{i=1}^{n} P(Yi\,|(k,\mu)) = (\frac{k}{\mu})^k \cdot \left(\frac{1}{(k-1)!}\right) \cdot y_1 \cdot \exp\left(-\frac{k \times y_1}{\mu}\right) \times \left(\frac{k}{\mu}\right)^k \cdot \left(\frac{1}{(k-1)!}\right) \cdot y_2 \cdot \exp\left(-\frac{k \times y_2}{\mu}\right) \times \ldots \times \left(\frac{k}{\mu}\right)^k \cdot \left(\frac{1}{(k-1)!}\right) \cdot y_n \cdot \exp\left(-\frac{k \times y_n}{\mu}\right)$

⇨

$$\left((\frac{k}{\mu})^k \cdot \left(\frac{1}{(k-1)!}\right)\right)^n * m * \exp\left(\frac{-\ell}{\mu}\right), where\ \ell\ is\ \sum_{i=1}^{n} y_i\ \ and\ \ m\ is\ \prod_{i=1}^{n} y_i$$

c. L(y|(k,mu)) = -ln(P(y | (k,mu))

⇨ $-\ln\left(\left((\frac{k}{\mu})^k \cdot \left(\frac{1}{(k-1)!}\right)\right)^n * m * \exp\left(\frac{-\ell}{\mu}\right)\right)$

⇨ $-n \log\left(\left(\frac{k}{mu}\right)^k \left(\frac{1}{(k-1)!}\right)\right) - \log(m) + \log\left(\exp\left(\frac{-l}{mu}\right)\right)$

⇨ $-n(klog(k) - \log\big((k-1)!\big) - k\log(mu) - \log(m) + \frac{l}{mu}$

d. $\frac{d}{dmu}\left(L\big(y\big|(k,mu)\big)\right) = n\,(\frac{d}{dm}\big(k\ln(k) - \ln\big((k-1)!\big) - k\ln(mu)\big) + \frac{d}{dm}\frac{l}{mu}$

Since we assume k is a constant then,

$\frac{d}{dm} k\ln(k)\ and \frac{d}{dm}\ln\big((k-1)!\big) = 0$

Therefore,

$$n\left(0 - 0 - \frac{d}{dm} k \ln(mu)\right) + \frac{d}{dm}\frac{l}{mu}$$

⇨ $n\left(0 - 0 - k * \frac{1}{mu}\right) + -\frac{l}{mu^2}$

$= \frac{nk}{mu} - \frac{l}{mu^2}$

Name: Abrar Fauzan Hamzah
ID: 28551494

e. we solve the answer above and plug k = 0. Therefore the estimated variance of MLE is

$$\frac{1}{mu^2} \sum_{i=1}^{n} y_i$$

Q3.

a. The success rate of German's of converting penalty is 17/18.

b. H_0 : German rate <= Population's rate (0.71)
H_a : German rate > 0.71

The approximate p-value is 0.01418839 which is smaller than 0.05. Therefore, we can reject the null hypothesis and conclude that German's rate of converting penalty is greater than the average in World Cup.

c. The exact p-value by using binom.test() we get 0.01756 which is still smaller than 0.05 and we can also reject the null hypothesis

```
#Q1a
g = (5+4+4+4)/(6+4+4+4)
g
g1= 5+4+4+4
x0 = 0.71
n1 = 6+4+4+4
#Q2b

#H0: german's rate <= 71%
#HA: german's rate > than 71%

z = (g - x0) / sqrt(x0*(1-x0)/n1)
z
p = 1-pnorm(z)
p
e <- binom.test(g1, n = n1 , p = 0.71, alternative = "g")
e
```

Name: Abrar Fauzan Hamzah
ID: 28551494

d. H_0 : German conversion rate = Opponent conversion rate
   H_a : German conversion rate != Opponent conversion rate

   The p-value from calculating from the difference in mean between German's conversion rate and opponent's conversion rate is 0.007053638. This p-value suggest that the average German's conversion will likely not be the same as its Opponent's rate and therefore we can reject the null hypothesis.

e. The first possible problem is that the sample data is only collected from world cup, it did not record all penalty in friendly match or other matches. Second, different years were taken which means that each penalty shootout, both team players could be different which could affect the outcome of the penalties. Other possibility is that small factors like weather, condition of the players could also affect the outcome.

Q4.

a. Based on the p-value after summarising the data, the three most significant predictors are **Cement** (p= 7.15e-07), **Blast.Furnace.Slage** (p = 1.16e-05) and **Age** (p= 2.86e-14) and these p-value is smaller than 0.01 which indicates they are important variables for strength.

```
data <-read.csv("/Users/abrarfauzanhamzah/Desktop/Monash/FIT2086/concrete.csv")
model <- lm(Strength ~.,data=data)
summary(model)
```

Name: Abrar Fauzan Hamzah
ID: 28551494

b. By using the formula p.adjust() to assess the 3 significant predictors in alpha = 0.05, the p-value is not drastically changed and this is tested to each predictors by the following code:

```
p.adjust(7.15e-07, method = "bonferroni", n = 8) #Cement
p.adjust(1.16e-05, method = "bonferroni", n = 8) #Blast
p.adjust(2.86e-14, method = "bonferroni", n = 8) #Age
```

c. Adding more water to the concrete could lower the mean compressive strength while adding a day of age could increase the mean compressive strength

d. Equation: Strength ~ Cement + Blast.Furnace.Slag + Age + Water

```
#StepWISE
null = lm(Strength ~ 1,data=data)
full = lm(Strength ~ Cement + Blast.Furnace.Slag + Fly.Ash
+Water+Superplasticizer+Coarse.Aggregate+Fine.Aggregate +Age, data=data)
step(null, scope = list(lower=null,upper=full),
     direction="both", criterion="BIC", k=log(250))
```

e. To improve the strength, we need to increase the amount of **cement**, **Blast.Furnace.Slag** and the amount of **days** as well as decreasing the amount of **water**. Although the final equation ignores unimportant variables, these variable could possibly contribute to the strength of the cement.

f. By calculating the mean from Table 3 in the assignment sheet, we get **E[X] = 54.27** and by using the formula **calcCI()**, we are 95% confidence that the mean is between **51.98531 and 56.55469.** This answers the next question, the new proposed mix is better than the current mix. Based on the CI , the highest mean is 56.56. Although the lowest is 51.98531, the difference between the proposed mix and current mix is only by around 0.3.