Chapter 2 Solutions

LRP

Question 1

(a) We can use the transformation $y = \frac{x-35}{10} \cdot 15 + 100 = 1.5x + 47.5$

(b) There is another linear transformation $y = \frac{x-35}{10} \cdot (-15) + 100 = -1.5x + 152.5$ which also has mean 100 and standard deviation 15. We do not use this because it has negative correlations between the original scores and the transformed scores.

Question 2

(a)

```
girlbirth <- c(.4777, .4875, .4859, .4754, .4874, .4864, .4813, .4787,
               .4895, .4797, .4876, .4859, .4857, .4907, .5010, .4903,
               .4860, .4911, .4871, .4725, .4822, .4870, .4823, .4973)
girlsd_obs <- sd(girlbirth)
sprintf("Standard deviation of the proportions is %f", girlsd_obs)
```

```
## [1] "Standard deviation of the proportions is 0.006410"
```

```
constprob<- mean(girlbirth)
girlsd_pro<-sqrt(constprob*(1-constprob)/3900)
sprintf("Standard deviation assuming constant birth probability is %f", girlsd_pro)
```

```
## [1] "Standard deviation assuming constant birth probability is 0.008003"
```

(b)

First we need to prove that the distritbution of variable $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$, assuming $s^2 = \frac{\sum_1^n (X_i - \bar{X})^2}{\sigma^2}$ is the sample variance and $X_i \sim N(\mu, \sigma^2)$.

To prove this, we know that $\frac{\sum_1^n (X_i - \mu)^2}{\sigma^2} \sim \chi^2(n)$ and we have

$$\begin{aligned}
\frac{\sum_1^n (X_i - \mu)^2}{\sigma^2} &= \frac{\sum_1^n (X_i - \bar{X} + \bar{X} - \mu)^2}{\sigma^2} \\
&= \frac{\sum_1^n (X_i - \bar{X})^2}{\sigma^2} + \frac{\sum_1^n (\bar{X} - \mu)^2}{\sigma^2} + \frac{2\sum_1^n (X_i - \bar{X})(\bar{X} - \mu)}{\sigma^2} \\
&= \frac{\sum_1^n (X_i - \bar{X})^2}{\sigma^2} + \frac{\sum_1^n (\bar{X} - \mu)^2}{\sigma^2}
\end{aligned}$$

For the equation above, assuming LHS = Z, RHS = X+Y, we have $E(e^{tZ}) = E(e^{tX})E(e^{tY})$. Since the MGF of Z,Y are $(1-2t)^{-n/2}$ and $(1-2t)^{-1/2}$, we have $E(e^{tX}) = (1-2t)^{-(n-1)/2}$. Thus, $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2(n-1)$.

For a confidence interval $1-\alpha$, we have $\chi^2_{1-\alpha/2} \le \frac{(n-1)s^2}{\sigma^2} \le \chi^2_{\alpha/2}$, therefore $\frac{(n-1)s^2}{\chi^2_{\alpha/2}} \le \sigma^2 \le \frac{(n-1)s^2}{\chi^2_{1-\alpha/2}}$.

```
sd_upper = (23*girlsd_obs^2/qchisq(0.025,23))^0.5
sd_lower = (23*girlsd_obs^2/qchisq(0.975,23))^0.5
sprintf("Confidence interval for theoretical standard deviation (%f, %f)", sd_lower,sd_upper)
```

```
## [1] "Confidence interval for theoretical standard deviation (0.004982, 0.008991)"
```

Therefore, the theoretical SD 0.0064 is in the range and the difference between actual and theoretical SD is not significant.

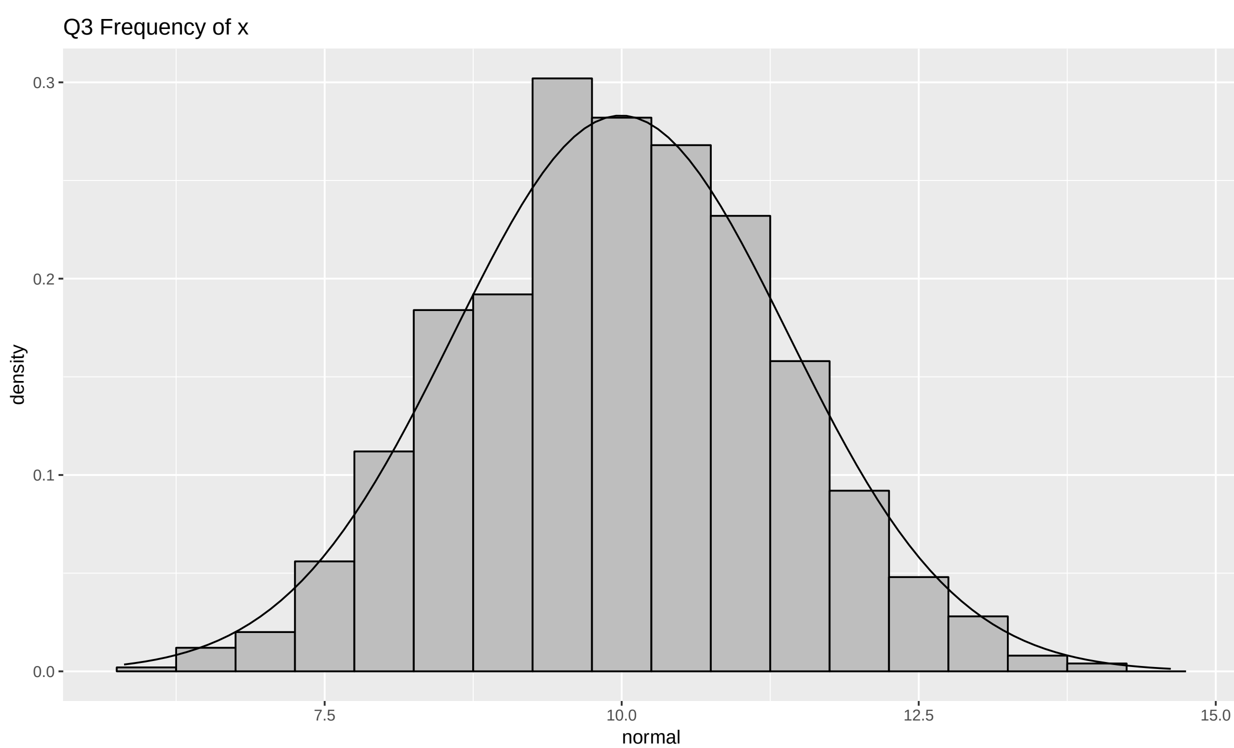
Question 3

```
library(ggplot2)

sim = 1000
x=rep(0,sim)
for (i in 1:sim){
  x[i]=sum(runif(20, min=0, max=1))
}

df<-data.frame(normal=rnorm(sim,mean=mean(x),sd=sd(x)),sumu=x)

ggplot(df,aes(x=normal)) +
  geom_histogram(aes(x = sumu, y = ..density..),
                 binwidth = 0.5, fill = "grey", color = "black") +
  stat_function(fun = dnorm, args = list(mean = mean(df$normal), sd = sd(df$normal)))+
  ggtitle("Q3 Frequency of x")
```

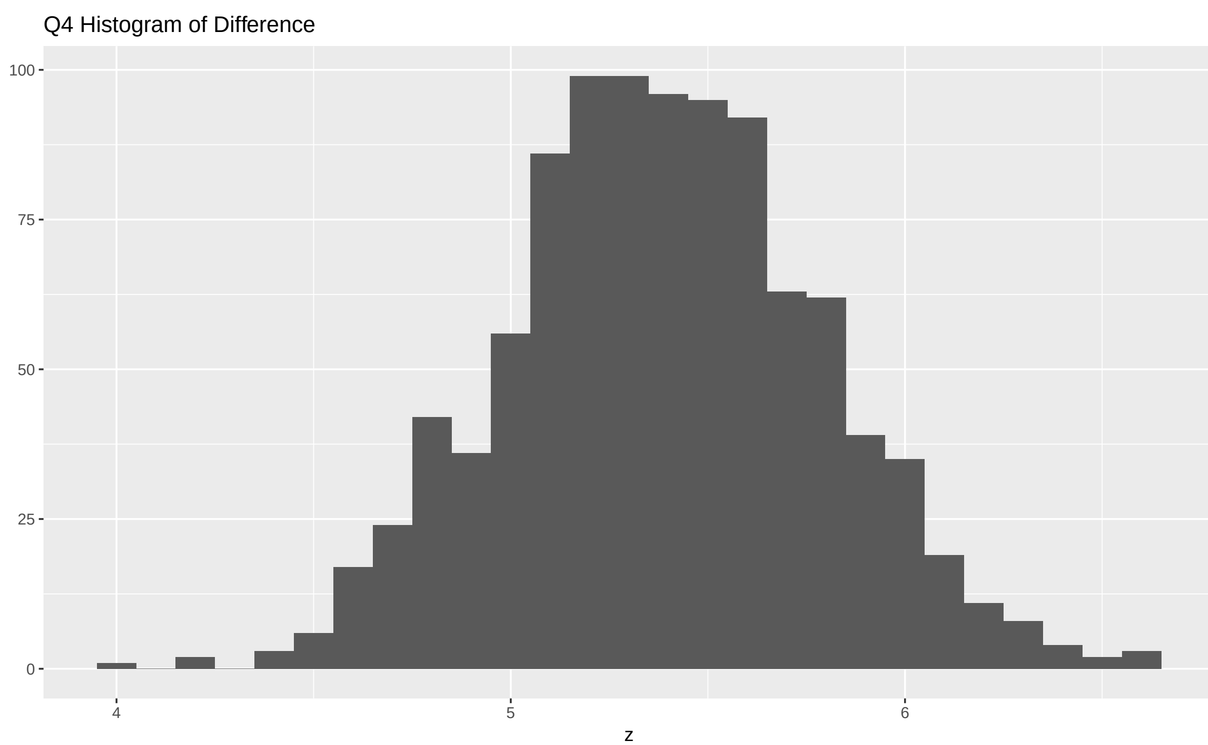


From the histogram and normal density curve, we can see that they are close.

Question 4

```r
sim = 1000
z=rep(0,sim)
for (i in 1:sim){
  x=mean(rnorm(100,mean = 69.1, sd=2.9))
  y=mean(rnorm(100,mean=63.7, sd=2.7))
  z[i]=x-y
}

qplot(z, geom="histogram", main="Q4 Histogram of Difference", binwidth = 0.1)
```




```r
actualmean=69.1-63.7
actualsd=(2.9^2+2.7^2)^0.5
sprintf("The simulated mean is %f and the actual mean is %f", mean(z), actualmean)
```

```
## [1] "The simulated mean is 5.386651 and the actual mean is 5.400000"
```

```r
sprintf("The simulated SD is %f and the actual SD is %f", sd(z), actualsd)
```

```
## [1] "The simulated SD is 0.396593 and the actual SD is 3.962323"
```

Therefore, the simulated mean and standard deviation are close to the exact values.

Question 5

The mean is $E(\frac{x+y}{2}) = (69.1 + 63.7)/2 = 66.4$.

The standard deviation is $sd(\frac{x+y}{2}) = \frac{(\sigma_x^2+\sigma_y^2+2\rho\sigma_x\sigma_y)^2}{2} = 2.2582$