Exercise 3 - Question 2 Solution

Question 2

2.1

```
d <- read.csv("process.csv")
# Equal-variance t-test
t.test(d$output[d$temp==50],d$output[d$temp==100],var.equal=T)
```

```
##
##  Two Sample t-test
##
## data:  d$output[d$temp == 50] and d$output[d$temp == 100]
## t = -3.3407, df = 58, p-value = 0.001466
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -214.79367  -53.83427
## sample estimates:
## mean of x mean of y
##  899.8219 1034.1358
```

```
#Welch t-test
t.test(d$output[d$temp==50],d$output[d$temp==100],var.equal=F)
```

```
##
##  Welch Two Sample t-test
##
## data:  d$output[d$temp == 50] and d$output[d$temp == 100]
## t = -3.3407, df = 56.917, p-value = 0.00148
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -214.82630  -53.80164
## sample estimates:
## mean of x mean of y
##  899.8219 1034.1358
```

The equal-variance and Welch t-tests give the same result because the sample sizes in the two groups are equal. The large-sample Z-test gives a slightly different p-value.

2.2

```
d2 <- d[1:45,]
#Equal-variance t-test
with(d2,t.test(output[temp==50],output[temp==100],var.equal=T))
```

```
##
##  Two Sample t-test
##
## data:  output[temp == 50] and output[temp == 100]
```

```
## t = -2.5736, df = 43, p-value = 0.0136
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -225.78744  -27.39309
## sample estimates:
## mean of x mean of y
##  899.8219 1026.4121
```

```
#Welch t-test
result<-with(d2,t.test(output[temp==50],output[temp==100],var.equal=F))
result
```

```
##
##  Welch Two Sample t-test
##
## data:  output[temp == 50] and output[temp == 100]
## t = -2.7867, df = 34.728, p-value = 0.008571
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -218.83546  -34.34507
## sample estimates:
## mean of x mean of y
##  899.8219 1026.4121
```

```
#Large-sample Z-test
Z<-result$statistic
p=2*(1-pnorm(abs(Z)))
p
```

```
##           t
## 0.005324024
```

2.3

```
N<-5000
set.seed(102)
n1<-30
n2<-15
n<-n1+n2
sim_data<-data.frame(temp=c(rep(50,n1),rep(100,n2)),output=rep(NA,n))
sim_result<- data.frame(p.equal.var=rep(NA,N),p.Welch=rep(NA,N),p.Z=rep(NA,N))
for(i in 1:N){
    sim_data$output<-rnorm(n,mean=1000,sd=c(rep(80,n1),rep(20,n2)))
    result<-with(sim_data,t.test(output[temp==50],output[temp==100],var.equal=T))
    sim_result[i,1]<-result$p.value
    result<-with(sim_data,t.test(output[temp==50],output[temp==100],var.equal=F))
    sim_result[i,2]<-result$p.value
    Z<-result$statistic
    sim_result[i,3]<-2*(1-pnorm(Z))
}
apply(sim_result<0.05,2,mean)
```

```
## p.equal.var     p.Welch         p.Z
##      0.0102      0.0482      0.0282
```

2.4

```
N<-5000
set.seed(102)
n1<-30
n2<-15
n<-n1+n2
sim_data<-data.frame(temp=c(rep(50,n1),rep(100,n2)),output=rep(NA,n))
sim_result<- data.frame(p.equal.var=rep(NA,N),p.Welch=rep(NA,N),p.Z=rep(NA,N))
for(i in 1:N){
    sim_data$output<-rnorm(n,mean=1000,sd=c(rep(20,n1),rep(80,n2)))
    result<-with(sim_data,t.test(output[temp==50],output[temp==100],var.equal=T))
    sim_result[i,1]<-result$p.value
    result<-with(sim_data,t.test(output[temp==50],output[temp==100],var.equal=F))
    sim_result[i,2]<-result$p.value
    Z<-result$statistic
    sim_result[i,3]<-2*(1-pnorm(Z))
}
apply(sim_result<0.05,2,mean)
```

```
## p.equal.var     p.Welch         p.Z
##      0.1518      0.0428      0.0278
```