Stats 10 Lab 5 Submission
Name: Enmin Zhou
UID: 104756697

Exercise 1

a)

Ho: p = 10%, Ha: p > 10%, one-sided

b)

```
> flint <-read.csv("~/Desktop/stats10/lab5/flint_2015.csv", header = TRUE)
> n <- nrow(flint)
> dangerous_lead_indicator <- (flint$Pb >= 15)
> p_hat <- mean(dangerous_lead_indicator)
> sd_sample<-sqrt(p_hat*(1-p_hat)/n)
> p_hat
[1] 0.1238447
> sd_sample
[1] 0.0141622
```

c)

```
> p_null <- 0.10
> se_null <- sqrt(p_null*(1-p_null)/n)
> z_stat <- (p_hat-p_null)/se_null
> se_null
[1] 0.01289801
> z_stat
[1] 1.848714
```

d)

```
> p_value <- 1-pnorm(z_stat, sd=1, mean=0)
> p_value
[1] 0.03224953
```

e)

p_value < 0.05, Reject Ho

f)

Reject Ho, p > 10%, we should take actions.

g)

```
> prop.test(x=sum(dangerous_lead_indicator), n=n, p=0.1,alternative = "greater")

	1-sample proportions test with continuity correction

data:  sum(dangerous_lead_indicator) out of n, null probability 0.1
X-squared = 3.1579, df = 1, p-value = 0.03778
alternative hypothesis: true p is greater than 0.1
95 percent confidence interval:
 0.101559 1.000000
sample estimates:
        p
0.1238447
```

h)

```
> prop.test(x=sum(dangerous_lead_indicator), n=n,p=0.1,alternative = "greater", conf.level = 0.99)

	1-sample proportions test with continuity correction
```

```
data:  sum(dangerous_lead_indicator) out of n, null probability 0.1
X-squared = 3.1579, df = 1, p-value = 0.03778
alternative hypothesis: true p is greater than 0.1
99 percent confidence interval:
 0.09376523 1.00000000
sample estimates:
        p
0.1238447
```

Exercise 2
a)
Ho: p1=p2, Ha: p1 != p2, two-sided
b)

```
> flint_north <- flint[flint$Region=="North",]
> n_north <- nrow(flint_north)
> flint_south <- flint[flint$Region == "South",]
> n_south <- nrow(flint_south)
> p_hat_north <- mean(flint_north$Pb>=15)
> p_hat_south <-mean(flint_south$Pb>=15)
> p_hat_pooled <- mean(flint$Pb >=15)
> SE <- sqrt(p_hat_pooled*(1-p_hat_pooled)*(1/n_north + 1/n_south))
> z_stat <- (p_hat_north-p_hat_south-0)/SE
> z_stat
[1] 3.572283
```

c)

```
> p_value <- 2*(1-pnorm(z_stat, sd=1, mean=0))
> p_value
[1] 0.0003538831
```

d)
p_value < 0.05, reject the null hypothesis. The proportion of the lead level greater than 15 in the north is different from that in the south.

e)

```
> x_north <- sum(flint_north$Pb >= 15)
> x_south <- sum(flint_south$Pb >= 15)
> prop.test(x=c(x_north, x_south), n = c(n_north, n_south), alternative = "two.sided")

	2-sample test for equality of proportions with continuity correction

data:  c(x_north, x_south) out of c(n_north, n_south)
X-squared = 11.845, df = 1, p-value = 0.0005781
alternative hypothesis: two.sided
95 percent confidence interval:
 0.04196839 0.16052203
sample estimates:
   prop 1    prop 2
0.1762452 0.0750000
```

The p_value changes a little, but it is still smaller than 0.05. The result does not change and we still reject the null hypthesis.

Exercise 3
a)

Ho: M=40
Ha: M != 40
Two-sided

b)
```
> xbar <- mean(flint$Cu)
> s <- sd(flint$Cu)
> xbar
[1] 54.58102
> s
[1] 133.3042
```

c)
```
> n <- nrow(flint)
> SE <- s/sqrt(n)
> SE
[1] 5.731197
```

d)
```
> t_stat <- (xbar-40)/SE
> t_stat
[1] 2.54415
> p_value <- (1-pt(t_stat, df=n-1))*2
> p_value
[1] 0.01123183
```

e)
0.011 > 0.01, Fail to reject the Ho. We don't have strong evidence that the mean copper is different from 40 ppm.

f)
```
> t.test(flint$Cu, mu=40, alternative = "two.sided")

	One Sample t-test

data:  flint$Cu
t = 2.5441, df = 540, p-value = 0.01123
alternative hypothesis: true mean is not equal to 40
95 percent confidence interval:
 43.32285 65.83920
sample estimates:
mean of x 
 54.58102 
```

0.011 < 0.05, reject Ho => the mean is different from 40 ppm.

Extra Credit
a)
b

b)
Ho: b = 0

Ha: b != 0

c)

```
> soil<-read.table("http://www.stat.ucla.edu/~nchristo/statistics_c173_c273/soil_complete.txt",
header=TRUE)
> linear_model <- lm(soil$lead ~ soil$zinc)
> summary(linear_model)

Call:
lm(formula = soil$lead ~ soil$zinc)

Residuals:
     Min       1Q   Median       3Q      Max 
-79.853 -12.945  -1.646  15.339 104.200 

Coefficients:
                Estimate Std. Error t value Pr(>|t|)    
(Intercept) 17.367688   4.344268   3.998 9.92e-05 ***
soil$zinc    0.289523   0.007296  39.681  < 2e-16 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 33.24 on 153 degrees of freedom
Multiple R-squared:  0.9114,	Adjusted R-squared:  0.9109 
F-statistic:  1575 on 1 and 153 DF,  p-value: < 2.2e-16
```

p_value is 2.2e-16

d)

2.2e-16 < 0.05. Reject Ho => b != 0 and there is a relationship between lead and zinc values.