Problem Set 1

lalonde <- read.csv("lalonde.csv", header = TRUE)  
attach(lalonde)

#### 1.

t.test(age[treat == 1], age[treat == 0])  
t.test(education[treat == 1], education[treat == 0])  
t.test(black[treat == 1], black[treat == 0])  
t.test(hispanic[treat == 1], hispanic[treat == 0])  
t.test(married[treat == 1], married[treat == 0])  
t.test(nodegree[treat == 1], nodegree[treat == 0])  
t.test(re74[treat == 1], re74[treat == 0])  
t.test(re75[treat == 1], re75[treat == 0])  
t.test(u74[treat == 1], u74[treat == 0])  
t.test(u75[treat == 1], u75[treat == 0])

a = 0.05

Age: P-value = 0.2659, Fail to reject null hypothesis, cannot conclude that a significant difference exists.

Education: P-value = 0.1502, Fail to reject null hypothesis, cannot conclude that a significant difference exists.

Black: P-value = P-value = 0.6474, Fail to reject null hypothesis, cannot conclude that a significant difference exists.

Hispanic: P-value = 0.06404, Fail to reject null hypothesis, cannot conclude that a significant difference exists.

Married: P-value = 0.3342, Fail to reject null hypothesis, cannot conclude that a significant difference exists.

No Degree: P-value = 0.002037, Reject null hypothesis, can conclude that a significant difference exists.

RE74: P-value = 0.982, Fail to reject null hypothesis, cannot conclude that a significant difference exists.

RE75: P-value = 0.3857, Fail to reject null hypothesis, cannot conclude that a significant difference exists.

U74: P-value = 0.3303, Fail to reject null hypothesis, cannot conclude that a significant difference exists.

U75: P-value = 0.06803, Fail to reject null hypothesis, cannot conclude that a significant difference exists.

#### 2.

t.test(re78[treat == 1], re78[treat == 0])  
mean(re78[treat == 1]) - mean(re78[treat == 0])

ATE = 1793.609

95% Confidence interval: (473.2551, 3113.9627)

#### 3.

t.test(re78[treat == 1 & re74 == 0 & re75 == 0], re78[treat == 0 & re74 == 0 & re75 == 0])  
mean(re78[treat == 1 & re74 == 0 & re75 == 0]) - mean(re78[treat == 0 & re74 == 0 & re75 == 0])  
mod\_data <- lalonde[which(re74 > 0 | re75 > 0),]  
t.test(mod\_data$re78[mod\_data$treat == 1], mod\_data$re78[mod\_data$treat == 0])  
mean(mod\_data$re78[mod\_data$treat == 1]) - mean(mod\_data$re78[mod\_data$treat == 0])

Zero earnings in 1975 and 1974:

ATE = 1842.076

95% Confidence interval: (366.4982, 3317.6536)

Positive earnings in 1975 or 1974:

ATE = 1508

95% Confidence interval: (-1000.282, 4016.283)

#### 4.

lalonde$pos78 <- ifelse(lalonde$re78 > 0, 1, 0)  
mean(lalonde$pos78[treat == 1]) - mean(lalonde$pos78[treat == 0])

ATE = 0.1106029

#### 5.

**i)**

Expected value = 6\*(6/21) + 7\*(7/21) + 8\*(8/21) = 7.095238

Variance = [6^2\*(6/21) + 7^2\*(7/21) + 8^2\*(8/21)] - 7.095238^2 = 0.6575977

Standard deviation = sqrt(0.6575977) = 0.810924

**ii)**

Expected value = 6^n\*(6/21) + 7^n\*(7/21) + 8^n\*(8/21)

**iii)**

6^3 + 2\*6 = 228

7^3 + 2\*7 = 357

8^3 + 2\*8 = 528

Expected value = 228\*(6/21) + 357\*(7/21) + 528\*(8/21) = 385.2857

Variance = [228^2\*(6/21) + 357^2\*(7/21) + 528^2\*(8/21)] - 385.2857^2 = 15093.93

Standard deviation = sqrt(15093.93) = 122.8574

**iv)**

E[Y] = E[X^3 + 2X] = E[X^3] + 2\*E[X]

Expected value = (6^3)\*(6/21) + (7^3)\*(7/21) + (8^3)\*(8/21) + 2\*7.095238 = 385.2857

Variance = [228^2\*(6/21) + 357^2\*(7/21) + 528^2\*(8/21)] - 385.2857^2 = 15093.93

Standard deviation = sqrt(15093.93) = 122.8574

#### 6.

**i)**

80 - 74 = 6

Z-score = 6/20 = 0.3

61.79% chance that a house will be on the market 80 days or less.

**ii)**

50 - 74 = -24

Z-score = -24/20 = -1.2

88.49% chance that a house will be on the market more than 50 days.

**iii)**

as.Date("October 1 2016", "%B %d %Y") - as.Date("August 1 2016", "%B %d %Y")

## Time difference of 61 days

as.Date("October 31 2016", "%B %d %Y") - as.Date("August 1 2016", "%B %d %Y")

## Time difference of 91 days

61 - 74 = -13

Z-score = -13/20 = -0.65

91 - 74 = 17

Z-score = 17/20 = 0.85

54.45% chance that the house will sell at any time during October of 2016.

**iv)**

Z-score = -0.44

-0.44\*20 = -8.8

-8.8 + 74 = 65.2

The fastest selling 33% of homes are on the market for 65.2 days.

**v)**

Z-score = 0.77

0.77\*20 = 15.4

15.4 + 74 = 89.4

The slowest selling 22% of homes are on the market for at least 89.4 days.

#### 7.

**i)**

ATE = 0.79 - 0.44 = 0.35

**ii)**

0.35 + sqrt[(1.7^2/4405) + (1.3^2/1040)] = 0.3977606

0.35 - sqrt[(1.7^2/4405) + (1.3^2/1040)] = 0.3022394

95% Confidence interval: (0.3022394, 0.3977606)

**iii)**

I would reject the null hypothesis that there was no treatment effects because 0 is not inside the 95% confidence interval.