

# Red and Black

coop711

2015년 4월 7일

## Data Preparation

자료 읽혀들이기.

```
red.black<-read.table("red_black.txt", header=TRUE, sep="")
str(red.black)
```

```
## 'data.frame':    58 obs. of  23 variables:
##  $ Color: chr   "Curitiba" "Curitiba" "Curitiba" "Curitiba" ...
##  $ Q1    : int   3 2 3 4 4 3 3 3 2 1 ...
##  $ Q2_1   : int   3 5 4 3 4 3 4 5 5 4 ...
##  $ Q2_2   : int   4 5 4 4 5 4 4 5 4 5 ...
##  $ Q2_3   : int   3 5 3 2 3 4 3 5 3 4 ...
##  $ Q2_4   : int   4 5 4 4 3 4 5 5 5 5 ...
##  $ Q2_5   : int   3 5 4 4 5 4 4 5 4 4 ...
##  $ Q2_6   : int   3 5 5 4 5 3 5 5 5 4 ...
##  $ Q3     : int   3 3 NA 3 5 4 3 4 3 2 ...
##  $ Q4_1   : int   4 5 3 5 4 3 4 5 4 3 ...
##  $ Q4_2   : int   4 3 4 4 3 3 3 4 2 2 ...
##  $ Q4_3   : int   2 4 4 2 3 2 2 5 3 5 ...
##  $ Q4_4   : int   2 4 4 3 3 3 2 5 3 4 ...
##  $ Q4_5   : int   2 5 3 5 4 4 2 5 3 4 ...
##  $ Q4_6   : int   2 2 2 2 3 2 2 5 3 3 ...
##  $ Q5_1   : int   1 4 3 5 2 3 2 5 3 2 ...
##  $ Q5_2   : int   2 5 4 5 3 3 2 5 3 4 ...
##  $ Q5_3   : int   2 4 4 4 4 3 2 5 3 3 ...
##  $ Q5_4   : int   3 4 4 5 4 4 3 5 3 5 ...
##  $ Q5_5   : int   4 5 4 4 4 4 3 5 3 3 ...
##  $ Q6_1   : int   2 2 1 1 2 2 1 2 1 1 ...
##  $ Q6_2   : int   1 2 2 1 1 1 1 1 1 2 ...
##  $ Q6_4   : int   2 1 3 2 1 3 2 1 2 2 ...
```

```
head(red.black)
```

```
##      Color Q1 Q2_1 Q2_2 Q2_3 Q2_4 Q2_5 Q2_6 Q3 Q4_1 Q4_2 Q4_3 Q4_4 Q4_5
## 1 Curitiba 3    3    4    3    4    3    3    3    4    4    2    2    2
## 2 Curitiba 2    5    5    5    5    5    5    3    5    3    4    4    5
## 3 Curitiba 3    4    4    3    4    4    5 NA    3    4    4    4    3
## 4 Curitiba 4    3    4    2    4    4    4    3    5    4    2    3    5
## 5 Curitiba 4    4    5    3    3    5    5    5    4    3    3    3    4
## 6 Curitiba 3    3    4    4    4    4    3    4    3    3    2    3    4
##      Q4_6 Q5_1 Q5_2 Q5_3 Q5_4 Q5_5 Q6_1 Q6_2 Q6_4
## 1      2      1      2      2      3      4      2      1      2
## 2      2      4      5      4      4      5      2      2      1
## 3      2      3      4      4      4      4      1      2      3
## 4      2      5      5      4      5      4      1      1      2
## 5      3      2      3      4      4      4      2      1      1
## 6      2      3      3      3      4      4      2      1      3
```

### Curitiba 와 Veja 응답 평균값 비교

```
options(digits=2)
aggregate(red.black[, -c(1, 21:23)], by=list(red.black[, 1]), mean, na.rm=TRUE)
```

```
##      Group.1  Q1 Q2_1 Q2_2 Q2_3 Q2_4 Q2_5 Q2_6  Q3 Q4_1 Q4_2 Q4_3 Q4_4 Q4_5
## 1 Curitiba 3.1  4.0  4.2  3.8  4.3  4.3  4.2 3.1  4.2  3.2  3.1  3.0  3.3
## 2      Veja 3.4  4.1  3.9  3.7  4.3  4.2  4.1 3.5  3.6  3.0  3.6  3.7  3.5
##      Q4_6 Q5_1 Q5_2 Q5_3 Q5_4 Q5_5
## 1  2.8  3.0  3.5  3.4  4.1  4.0
## 2  3.5  3.5  3.8  3.6  4.0  3.7
```

### 21-23번의 응답 테이블

```
table(red.black[, 21])
```

```
##
##  1  2
## 33 25
```

```
table(red.black[, c(1, 21)])
```

```
##      Q6_1
## Color      1  2
## Curitiba 16 14
##  Veja     17 11
```

```
table(red.black[, c(1, 22)])
```

```
##      Q6_2
## Color      1  2
## Curitiba 16 14
##  Veja     16 12
```

```
table(red.black[,c(1,23)])
```

```
##           Q6_4
## Color      1  2  3  4
## Curitiba 17 10  2  1
## Veja      9 12  6  1
```

평균 점수에 차이가 있어보이는 Q3, Q4에 대하여 t-test 수행. default로 Welch's Approxiation 수행

```
t.test(Q3~Color, data=red.black)
```

```
##
## Welch Two Sample t-test
##
## data:  Q3 by Color
## t = -1.6, df = 54, p-value = 0.1205
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.780  0.093
## sample estimates:
## mean in group Curitiba      mean in group Veja
##                3.1                3.5
```

```
t.test(Q4_1~Color, data=red.black)
```

```
##
## Welch Two Sample t-test
##
## data:  Q4_1 by Color
## t = 2.6, df = 55, p-value = 0.01198
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.13 0.99
## sample estimates:
## mean in group Curitiba      mean in group Veja
##                4.2                3.6
```

한꺼번에 수행하려면 t.test 의 구조를 이용하여 함수 작성 후 apply() 적용.

```
t<-function(x) {t.test(x~Color, data=red.black, na.rm=TRUE)}
apply(red.black[,c(1,21:23)],2, t)
```

```
## $Q1
##
## Welch Two Sample t-test
##
## data:  x by Color
## t = -1.4, df = 56, p-value = 0.1743
## alternative hypothesis: true difference in means is not equal to 0
```

```

## 95 percent confidence interval:
## -0.71 0.13
## sample estimates:
## mean in group Curitiba      mean in group Veja
##                3.1                3.4
##
##
## $Q2_1
##
## Welch Two Sample t-test
##
## data:  x by Color
## t = -0.59, df = 56, p-value = 0.5604
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.48 0.27
## sample estimates:
## mean in group Curitiba      mean in group Veja
##                4.0                4.1
##
##
## $Q2_2
##
## Welch Two Sample t-test
##
## data:  x by Color
## t = 2.3, df = 55, p-value = 0.02525
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## 0.044 0.637
## sample estimates:
## mean in group Curitiba      mean in group Veja
##                4.2                3.9
##
##
## $Q2_3
##
## Welch Two Sample t-test
##
## data:  x by Color
## t = 0.47, df = 54, p-value = 0.6426
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.29 0.47
## sample estimates:
## mean in group Curitiba      mean in group Veja
##                3.8                3.7
##
##
## $Q2_4
##
## Welch Two Sample t-test
##
## data:  x by Color

```

```

## t = 0.054, df = 50, p-value = 0.9572
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.43  0.46
## sample estimates:
## mean in group Curitiba      mean in group Veja
##                4.3                4.3
##
##
## $Q2_5
##
## Welch Two Sample t-test
##
## data:  x by Color
## t = 0.091, df = 53, p-value = 0.9281
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.35  0.39
## sample estimates:
## mean in group Curitiba      mean in group Veja
##                4.3                4.2
##
##
## $Q2_6
##
## Welch Two Sample t-test
##
## data:  x by Color
## t = 0.12, df = 56, p-value = 0.9071
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.38  0.43
## sample estimates:
## mean in group Curitiba      mean in group Veja
##                4.2                4.1
##
##
## $Q3
##
## Welch Two Sample t-test
##
## data:  x by Color
## t = -1.6, df = 54, p-value = 0.1205
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.780  0.093
## sample estimates:
## mean in group Curitiba      mean in group Veja
##                3.1                3.5
##
##
## $Q4_1
##
## Welch Two Sample t-test

```

```

##
## data:  x by Color
## t = 2.6, df = 55, p-value = 0.01198
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  0.13 0.99
## sample estimates:
## mean in group Curitiba      mean in group Veja
##                4.2                3.6
##
##
## $Q4_2
##
## Welch Two Sample t-test
##
## data:  x by Color
## t = 0.55, df = 51, p-value = 0.5842
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.35  0.61
## sample estimates:
## mean in group Curitiba      mean in group Veja
##                3.2                3.0
##
##
## $Q4_3
##
## Welch Two Sample t-test
##
## data:  x by Color
## t = -2.1, df = 56, p-value = 0.04423
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -1.067 -0.014
## sample estimates:
## mean in group Curitiba      mean in group Veja
##                3.1                3.6
##
##
## $Q4_4
##
## Welch Two Sample t-test
##
## data:  x by Color
## t = -2.7, df = 55, p-value = 0.008063
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -1.11 -0.17
## sample estimates:
## mean in group Curitiba      mean in group Veja
##                3.0                3.7
##
##
## $Q4_5

```

```

##
##  Welch Two Sample t-test
##
## data:  x by Color
## t = -0.77, df = 51, p-value = 0.4452
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.73  0.33
## sample estimates:
## mean in group Curitiba      mean in group Veja
##                3.3                3.5
##
##
## $Q4_6
##
##  Welch Two Sample t-test
##
## data:  x by Color
## t = -2.5, df = 55, p-value = 0.01516
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.20 -0.13
## sample estimates:
## mean in group Curitiba      mean in group Veja
##                2.8                3.5
##
##
## $Q5_1
##
##  Welch Two Sample t-test
##
## data:  x by Color
## t = -1.8, df = 56, p-value = 0.07144
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -1.120  0.048
## sample estimates:
## mean in group Curitiba      mean in group Veja
##                3.0                3.5
##
##
## $Q5_2
##
##  Welch Two Sample t-test
##
## data:  x by Color
## t = -0.85, df = 54, p-value = 0.4016
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
##  -0.73  0.30
## sample estimates:
## mean in group Curitiba      mean in group Veja
##                3.5                3.8
##

```

```
##
## $Q5_3
##
## Welch Two Sample t-test
##
## data: x by Color
## t = -0.78, df = 56, p-value = 0.4362
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.74 0.33
## sample estimates:
## mean in group Curitiba      mean in group Veja
##                3.4                3.6
##
##
## $Q5_4
##
## Welch Two Sample t-test
##
## data: x by Color
## t = 0.3, df = 55, p-value = 0.7651
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.38 0.51
## sample estimates:
## mean in group Curitiba      mean in group Veja
##                4.1                4.0
##
##
## $Q5_5
##
## Welch Two Sample t-test
##
## data: x by Color
## t = 1.2, df = 53, p-value = 0.2353
## alternative hypothesis: true difference in means is not equal to 0
## 95 percent confidence interval:
## -0.21 0.85
## sample estimates:
## mean in group Curitiba      mean in group Veja
##                4.0                3.7
```