hw5

Sam Mottahedi

July 29, 2017

1)

a)

```
data = read.table('marketshare.txt',header=T, colClasses = c('numeric','numeric', 'numeric', 'factor',

share = data[,1]; A = as.factor(data[,4]); B = as.factor(data[,5])
tapply(share,A:B,length)
```

```
## 0:0 0:1 1:0 1:1
##   8   7   8  13
```

$y_{ijk} = \mu_{..} + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \epsilon_{ijk}$

$i : 1 \ldots 2$

$j : 1 \ldots 2$

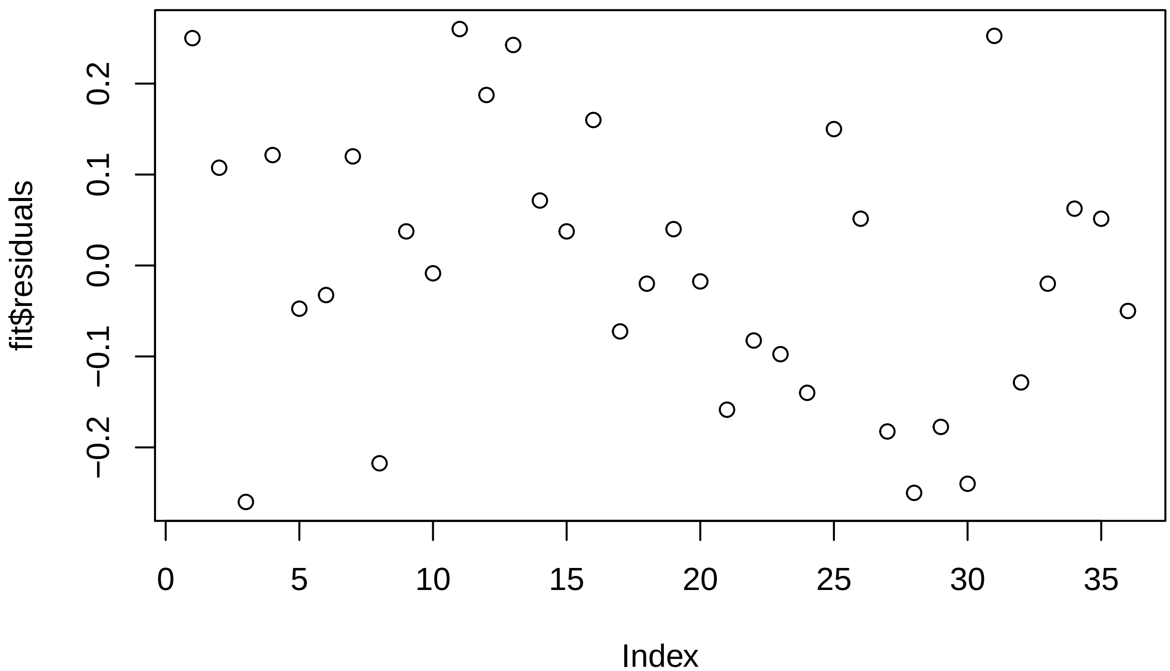
$k : 1 \ldots 8$

b)

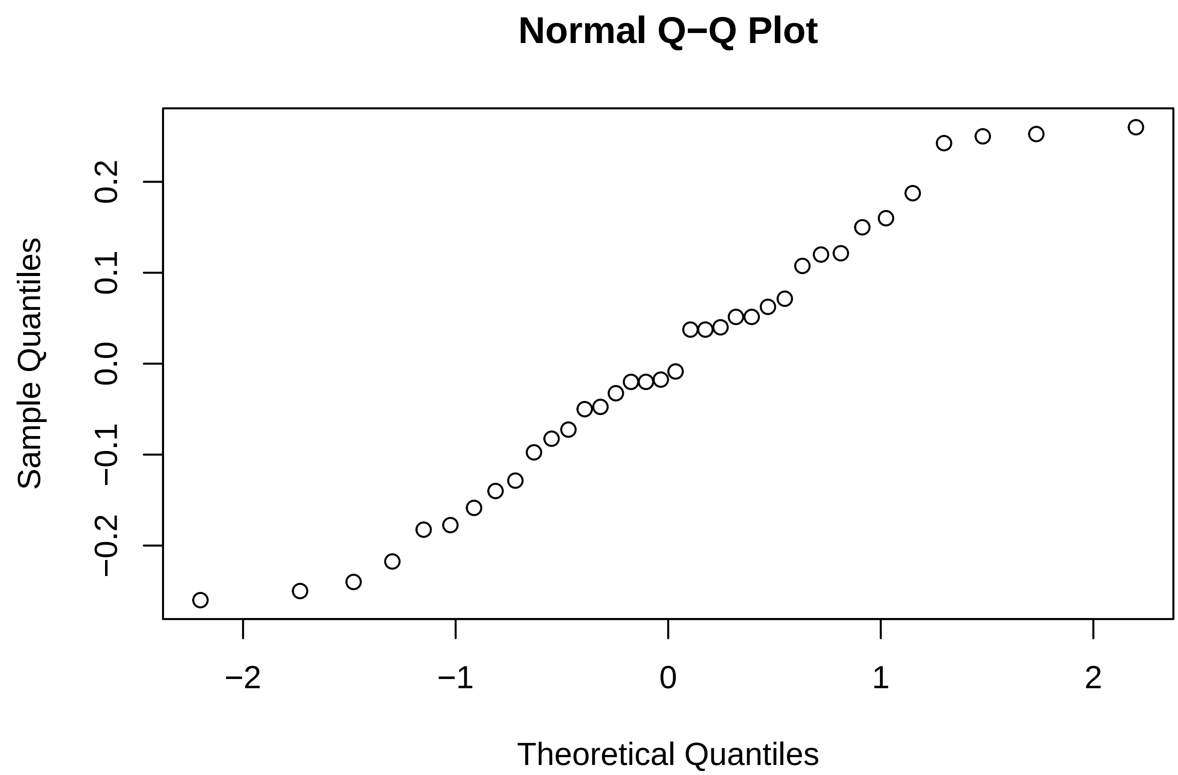
```
fit <- lm(share ~ A*B)
anova(fit)
```

```
## Analysis of Variance Table
##
## Response: share
##           Df  Sum Sq Mean Sq F value  Pr(>F)
## A          1 1.52953 1.52953 62.3515 5.2e-09 ***
## B          1 0.08610 0.08610  3.5097 0.07017 .
## A:B        1 0.04564 0.04564  1.8606 0.18208
## Residuals 32 0.78499 0.02453
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
plot(fit$residuals)
```




```r
qqnorm(fit$residuals)
```



The result does not show major deviation from normal dist residual and constant variance assumption.

c)

$y_{ijk} = \mu_{..} + \alpha_i + \beta_j + \epsilon_{ijk}$

```
fit2 <- lm(share ~ A + B)
anova(fit2)
```

```
## Analysis of Variance Table
##
## Response: share
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## A          1 1.52953 1.52953 60.7668 5.543e-09 ***
## B          1 0.08610 0.08610  3.4205   0.07336 .
## Residuals 33 0.83063 0.02517
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(fit2,fit)
```

```
## Analysis of Variance Table
##
## Model 1: share ~ A + B
## Model 2: share ~ A * B
##   Res.Df     RSS Df Sum of Sq      F Pr(>F)
## 1     33 0.83063
## 2     32 0.78499  1  0.045641 1.8606 0.1821
```

$F^* = 1.8605078$

$F = 4.1490974$

$F^* < F$

concluding H_0 the interaction term is not significant.

d)

$y_{ijk} = \mu_{..} + \beta_j + \epsilon_{ijk}$

The full model for this test is the reduced model from part c because the two factor are included in the full model and the reduced model does not include the factor A since we want to test factor A effect.

```
fit3 <- lm(share ~ B)
anova(fit3)
```

```
## Analysis of Variance Table
##
## Response: share
##           Df  Sum Sq  Mean Sq F value  Pr(>F)
## B          1 0.22756 0.227556  3.4871 0.07049 .
## Residuals 34 2.21870 0.065256
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
anova(fit2,fit3)
```

```
## Analysis of Variance Table
##
## Model 1: share ~ A + B
```

```
## Model 2: share ~ B
##   Res.Df     RSS Df Sum of Sq      F    Pr(>F)
## 1     33 0.83063
## 2     34 2.21870 -1   -1.3881 55.147 1.562e-08 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$F^* = 55.1464671$

$F = 4.1392525$

$F^* > F$

e)

$y_{ijk} = \mu_{..} + \beta_j + \epsilon_{ijk}$

```r
fit4 <- lm(share ~ A)
anova(fit4)
```

```
## Analysis of Variance Table
##
## Response: share
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## A          1 1.52953 1.52953  56.728 9.584e-09 ***
## Residuals 34 0.91672 0.02696
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
anova(fit2,fit4)
```

```
## Analysis of Variance Table
##
## Model 1: share ~ A + B
## Model 2: share ~ A
##   Res.Df     RSS Df Sum of Sq      F  Pr(>F)
## 1     33 0.83063
## 2     34 0.91672 -1 -0.086097 3.4205 0.07336 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$F^* = 3.4202593$

$F = 4.1392525$

$F^* < F$

concluding H_0 the factor B is not a significant effect in promotion package.

2)

a)

Fixed effect model is more appropriate since if we decide to repeat the experiment using the 4 types of retirement as treatments, the new sample exhaust the population. Also, the effect (retirement plan) is the main interest of this study.

b)

Random effect model is more appropriate since the sample size is small part of the population and repeating the experiment would result in a new sample each time.

3)

```
data = read.table('coils.dat', header=T, colClasses = c('numeric', 'factor', 'factor'))
y = data[,1]
machine = as.factor(data[,2])
coil = as.factor(data[,3])
```

a)

$Y_ij = \mu_i + \epsilon_{ij}$

$\mu_i \& \epsilon_{ij}$: independent random variables

$i = 1, \ldots, 4; j = 1, \ldots, 10$

b)

```
anova(lm(y ~ machine, data))
```

```
## Analysis of Variance Table
##
## Response: y
##           Df Sum Sq Mean Sq F value   Pr(>F)
## machine    3  602.5  200.83  28.089 1.54e-09 ***
## Residuals 36  257.4    7.15
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

$H_0 : \sigma^2_\mu = 0$

$H_a : \sigma^2_\mu > 0$

$F^* = 28.0881119 > F = 2.2426052$

we conclude H_a, and mean measured characteristic is different for different machines.

c)

$\bar{Y}_{..} = 205.05$

$s^2\{\bar{Y}_{..}\} = 5.02075$

$s\{\bar{Y}_{..}\} = 2.240703$

$t(0.95, 3) = 2.3533634$

$199.7768115 \leq \mu_. \leq 210.3231885$

d)

$\sigma_\mu^2 = 19.368$

e)

```
L = 1/10 * ((200.83/7.15) * (1/ (qf(.95, 3, 4*9))) -1)
U = 1/10 * ((200.83/7.15) * (1/ (qf(0.5, 3, 4*9))) -1)
```

$L = 0.879955$

$U = 3.3943718$

$L^* = 0.4680724$

$U^* = 0.7724362$

$0.4680724 \leq \frac{\sigma_\mu^2}{\sigma_\mu^2 + \sigma^2} \leq 0.7724362$

The result indicates that the variability in the measured characteristic in products from the four machines accounts for between 47 to 77 percent of the total variability in measure characteristics.

f)

$H_0 : \sigma_\mu^2 = \sigma^2$

$H_a : \sigma_\mu^2 \neq \sigma^2$

$F^* = 1$

$F = 2.2426052 > F^*$

concluding the H_0.

also, $0.879955 \leq \sigma_\mu^2 / \sigma^2 \leq 3.3943718$, we can see that the 90% interval includes 1.