a4

Mushi Wang

15/06/2020

q1

```r
options(contrasts = c('contr.sum', 'contr.poly'))

time1 = c(9, 12, 10, 8, 15)
time2 = c(20, 21, 23, 17, 30)
time3 = c(6, 5, 8, 16, 7)
y = c(time1, time2, time3)
type = as.factor(c(rep(1, 5), rep(2, 5), rep(3, 5)))

model = lm(y ~ type)
summary(model)
```

```
##
## Call:
## lm(formula = y ~ type)
##
## Residuals:
##    Min     1Q Median     3Q    Max
##   -5.2   -2.3   -1.2    1.0    7.8
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   13.800      1.061  13.001 1.97e-08 ***
## type1         -3.000      1.501  -1.999 0.068833 .
## type2          8.400      1.501   5.596 0.000117 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.111 on 12 degrees of freedom
## Multiple R-squared:  0.7283, Adjusted R-squared:  0.683
## F-statistic: 16.08 on 2 and 12 DF,  p-value: 0.0004023
```

(a) $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$ vs H_a : at least one of them is not 0.

From summary, F-statistic is 16.08 on 2 and 12 DF. The p-value is 0.0004023. So there is tons of evidence reject H_0. Hence, the three circuit types have different response time.

(b) $\theta = \tau_2 - \frac{\tau_1+\tau_3}{2}$ $\tilde{\theta} = \tilde{\tau}_2 - \frac{\tilde{\tau}_1+\tilde{\tau}_3}{2}$

(c) Form summary, $\hat{\tau}_1 = -3$, $\hat{\tau}_2 = 8.4$, and $\hat{\tau}_3 = -(-3 + 8.4) = -5.4$. Since, $\bar{y}_{i+} = \bar{y}_{++} + \tau_i$ and $Var(\bar{Y}_{1+}) = Var(\bar{Y}_{2+}) = Var(\bar{Y}_{3+})$. Therefore the widths of confidence intervals of $\bar{Y}_{1+}, \bar{Y}_{2+}, \bar{Y}_{3+}$ has the same width.So we choose circuit type 3.

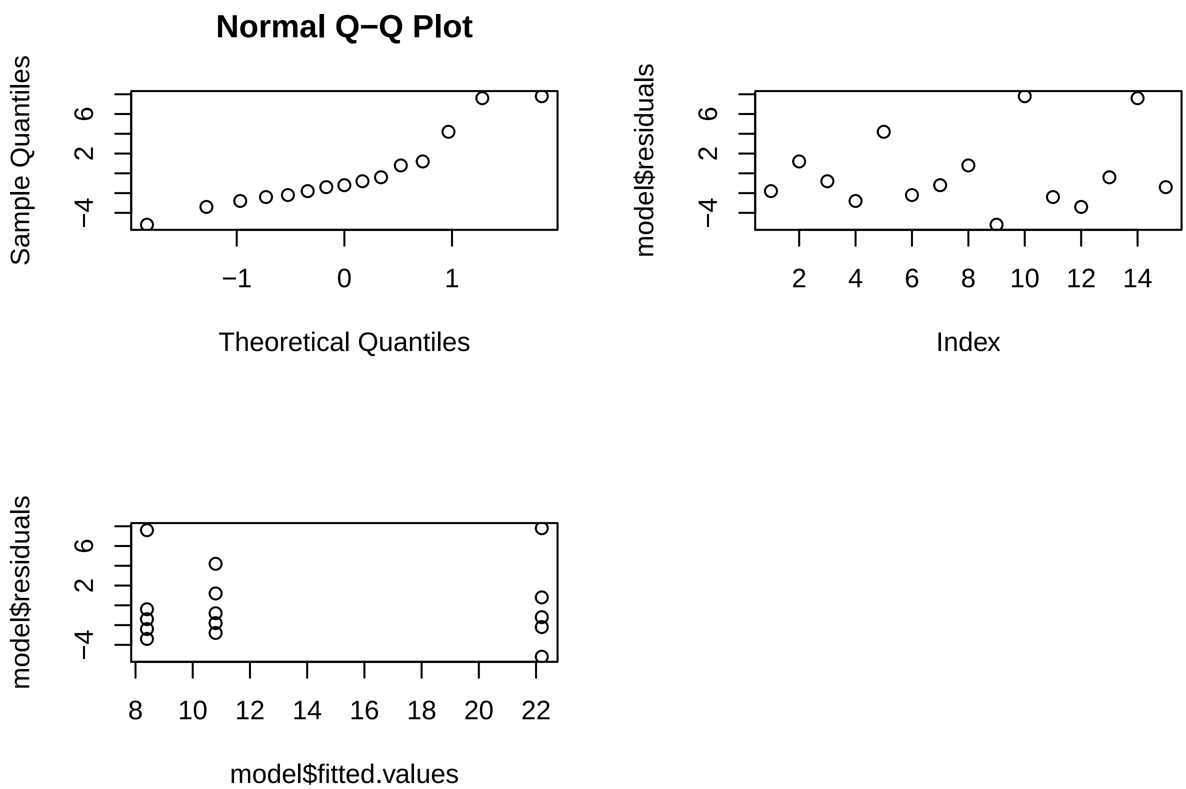
(d)

```r
anova(model)
```

```
## Analysis of Variance Table
##
## Response: y
```

```
##           Df Sum Sq Mean Sq F value    Pr(>F)
## type       2  543.6   271.8  16.083 0.0004023 ***
## Residuals 12  202.8    16.9
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```r
par(mfrow=c(2,2))
qqnorm(model$residuals)
plot(model$residuals)
plot(model$fitted.values, model$residuals)
```



In Q-Q plot, all the points are lie in a straight line with some exceptions on the right. So there could be a violation against assumption.

In indices vs residuals, all points are lie in a band between 0 and there is no obvious pattern.

In fitted values vs residuals, all points are lie in a band between 0 and there is no obvious pattern. But the could be a violation against assumption since we only have three distinct fitted values.

q2

```r
effectiveness = c(c(13, 22, 18, 39), c(16, 24, 17, 44), c(5, 4, 1, 22))
type = as.factor(c(rep(1, 4), rep(2, 4), rep(3, 4)))
block = as.factor(c(seq(from = 1, to = 4), seq(from = 1, to = 4), seq(from = 1, to = 4)))

effect = lm(effectiveness ~ type + block)
```

(a) $H_0 : \tau_1 = \tau_2 = \tau_3 = 0$ vs H_a : at least one of them is not 0

```r
anova(effect)
```

```
## Analysis of Variance Table
##
## Response: effectiveness
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## type       2  703.50  351.75  40.717 0.0003232 ***
## block      3 1106.92  368.97  42.711 0.0001925 ***
## Residuals  6   51.83    8.64
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Form anova table, $f = 40.717$, and p-value $= 0.0003232$. So there is tons of evidence reject H_0. Hence solutions have different effectiveness.

(b) $H_0 : \beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ vs H_a : at least one of them is not 0

```r
anova(effect)
```

```
## Analysis of Variance Table
##
## Response: effectiveness
##           Df  Sum Sq Mean Sq F value    Pr(>F)
## type       2  703.50  351.75  40.717 0.0003232 ***
## block      3 1106.92  368.97  42.711 0.0001925 ***
## Residuals  6   51.83    8.64
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Form anova table, $f = 42.711$, and p-value $= 0.00019252$. So there is tons of evidence reject H_0. Hence solutions have different effectiveness.

(c) $H_0 : \bar{y}_{2+} > 30$ vs $H_a : \bar{y}_{2+} \leq 30$

Let $\theta = \bar{Y}_{2+}$

```r
summary(effect)
```

```
##
## Call:
## lm(formula = effectiveness ~ type + block)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.583 -1.854 -0.250  1.250  4.417
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)  18.7500     0.8485  22.099 5.61e-07 ***
## type1         4.2500     1.1999   3.542  0.01219 *
```

```
## type2          6.5000     1.1999   5.417  0.00164 **
## block1        -7.4167     1.4696  -5.047  0.00234 **
## block2        -2.0833     1.4696  -1.418  0.20608
## block3        -6.7500     1.4696  -4.593  0.00372 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.939 on 6 degrees of freedom
## Multiple R-squared:  0.9722, Adjusted R-squared:  0.949
## F-statistic: 41.91 on 5 and 6 DF,  p-value: 0.0001371
```

$$
\begin{aligned}
E(\tilde{\theta}) &= E(\tilde{\tau_2} + \bar{Y}) \\
&= E(\tilde{\tau_2}) + E(\bar{Y}) \\
&= \tau_2 + \mu \quad \text{since unbiased}
\end{aligned}
$$

and,

$$
\begin{aligned}
Var(\tilde{\theta}) &= Var(\tilde{\tau_2} + \bar{Y}) \\
&= Var(\bar{Y_2} - \bar{Y} + \bar{Y}) \\
&= Var(\bar{Y_2}) \\
&= \frac{\sigma^2}{4}
\end{aligned}
$$

So, $\hat{\theta} = \hat{\mu} + \hat{\tau_2} = 18.75 + 6.5 = 25.25$

$$
\begin{aligned}
d &= \frac{\hat{\theta} - 30}{se(\tilde{\theta})} \\
&= \frac{25.25 - 30}{2.939/\sqrt{4}} \\
&= -3.232392
\end{aligned}
$$

Since $D \sim t_{12-7+2-1} = t_6$

```r
pt(-3.232392, 6)
```

```
## [1] 0.008928362
```

p-value $= 0.008928362$, so there is tons of evidence reject H_0. Hence the mean of solution 2 is less than or equal 30.

(d) $H_0 : \tau_1 - 2\tau_2 + \tau_3 = 0$ vs $H_a : \tau_1 - 2\tau_2 + \tau_3 \neq 0$

```r
summary(effect)
```

```
##
## Call:
## lm(formula = effectiveness ~ type + block)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -2.583 -1.854 -0.250  1.250  4.417
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  18.7500     0.8485  22.099 5.61e-07 ***
## type1         4.2500     1.1999   3.542  0.01219 *
## type2         6.5000     1.1999   5.417  0.00164 **
## block1       -7.4167     1.4696  -5.047  0.00234 **
## block2       -2.0833     1.4696  -1.418  0.20608
## block3       -6.7500     1.4696  -4.593  0.00372 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 2.939 on 6 degrees of freedom
## Multiple R-squared:  0.9722, Adjusted R-squared:  0.949
## F-statistic: 41.91 on 5 and 6 DF,  p-value: 0.0001371
```

Let $\theta = \tau_1 - 2\tau_2 + \tau_3$

$$\begin{aligned} E(\tilde{\theta}) &= E(\tilde{\tau_1}) - 2E(\tilde{\tau_2}) + E(\tilde{\tau_3}) \\ &= \tau_1 - 2\tau_2 + \tau_3 \quad \text{since unbiased} \end{aligned}$$

and,

$$\begin{aligned} Var(\tilde{\theta}) &= Var(\tilde{\tau_1} - 2\tilde{\tau_2} + \tilde{\tau_3}) \\ &= Var(\tilde{\tau_1}) + 4Var(\tilde{\tau_2}) + Var(\tilde{\tau_3}) \\ &= \frac{\sigma^2}{4} + \sigma^2 + \frac{\sigma^2}{4} \\ &= \frac{3}{2}\sigma^2 \end{aligned}$$

So,

$$\begin{aligned} d &= \frac{\hat{\tau_1} - 2\hat{\tau_2} + \hat{\tau_3} - 0}{\sqrt{\frac{3}{2}\hat{\sigma}^2}} \\ &= \sqrt{2}\frac{4.25 - 2 \times 6.5 - (4.25 + 6.5)}{\sqrt{3} \times 2.939} \\ &= -5.417381 \end{aligned}$$

Since $D \sim t_{12-7+2-1} = t_6$

```r
2 * (1 - pt(5.417381, 6))
```

```
## [1] 0.001636828
```

p-value $= 2p(D > |d|) = 0.001636828$

So we have tons of evidence to reject H_0. Hence $\tau_1 - 2\tau_2 + \tau_3 \neq 0$