Anova. KidsIQ

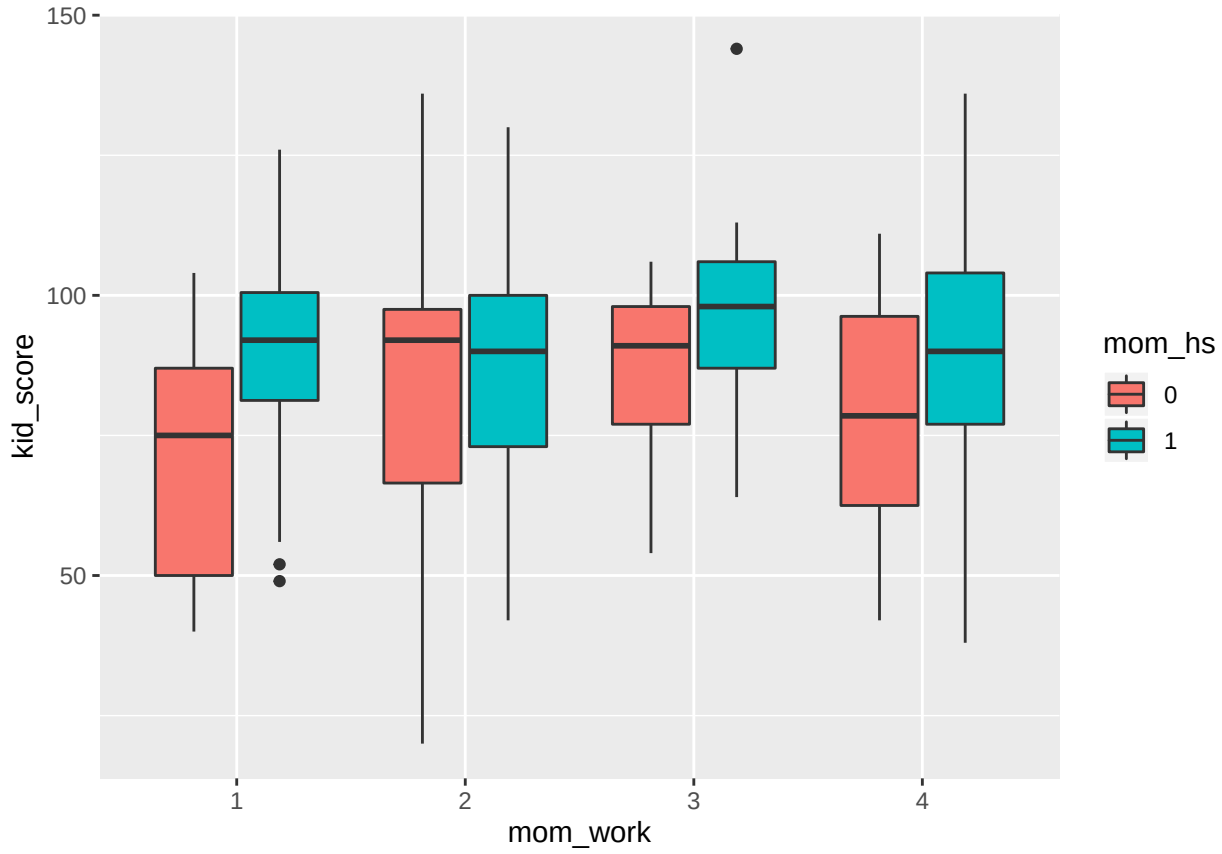
Sophie Donnet

Mars 2020

```
data_kids <- as.data.frame(read.table("kidiq.txt",head=TRUE))
data_kids$mom_hs <- as.factor(data_kids$mom_hs)
data_kids$mom_work <- as.factor(data_kids$mom_work)
library(ggplot2)
```

```
## Warning: package 'ggplot2' was built under R version 3.5.3
```

```
ggplot(data = data_kids,aes(x=mom_work,y=kid_score,fill=mom_hs))+geom_boxplot()
```




```
res_anova_complet <- lm(kid_score ~ mom_hs*mom_work,data = data_kids)
summary(res_anova_complet)
```

```
##
## Call:
## lm(formula = kid_score ~ mom_hs * mom_work, data = data_kids)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -62.652 -13.281   2.868  13.567  53.348
##
```

```
## Coefficients:
##                   Estimate Std. Error t value Pr(>|t|)
## (Intercept)         70.818      3.428  20.661  < 2e-16 ***
## mom_hs1             19.568      4.534   4.316 1.98e-05 ***
## mom_work2           11.834      5.348   2.213   0.0275 *
## mom_work3           14.909      6.855   2.175   0.0302 *
## mom_work4            7.297      5.163   1.413   0.1583
## mom_hs1:mom_work2  -15.357      6.537  -2.349   0.0193 *
## mom_hs1:mom_work3   -9.895      8.026  -1.233   0.2183
## mom_hs1:mom_work4   -9.153      6.135  -1.492   0.1365
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.69 on 426 degrees of freedom
## Multiple R-squared:  0.08442,    Adjusted R-squared:  0.06938
## F-statistic: 5.611 on 7 and 426 DF,  p-value: 3.319e-06
```

```
res_anova_add <- lm(kid_score ~ mom_hs + mom_work,data = data_kids)
anova(res_anova_complet,res_anova_add)
```

```
## Analysis of Variance Table
##
## Model 1: kid_score ~ mom_hs * mom_work
## Model 2: kid_score ~ mom_hs + mom_work
##   Res.Df    RSS Df Sum of Sq      F Pr(>F)
## 1    426 165157
## 2    429 167367 -3   -2209.2 1.8994  0.129
```

On ne rejette pas H_0. On peut garder le modèle sans interactions.

```
res_anova_1<- lm(kid_score ~ mom_hs ,data = data_kids)
anova(res_anova_1,res_anova_add)
```

```
## Analysis of Variance Table
##
## Model 1: kid_score ~ mom_hs
## Model 2: kid_score ~ mom_hs + mom_work
##   Res.Df    RSS Df Sum of Sq      F  Pr(>F)
## 1    432 170261
## 2    429 167367  3    2894.6 2.4732 0.06114 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Il n'y a pas d'effet significatif de la variable mom_work au dela de momo_hs

```
res_anova_1_bis <- lm(kid_score ~ mom_work ,data = data_kids)
anova(res_anova_1_bis,res_anova_add)
```

```
## Analysis of Variance Table
##
## Model 1: kid_score ~ mom_work
## Model 2: kid_score ~ mom_hs + mom_work
##   Res.Df    RSS Df Sum of Sq      F   Pr(>F)
## 1    430 175978
## 2    429 167367  1    8611.4 22.073 3.54e-06 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Si on a WORK il faut ajouter hs (p value tout petit). Si on a HS, pas besoin d'ajouter WORK (p value grande)

```
summary(res_anova_1)
```

```
##
## Call:
## lm(formula = kid_score ~ mom_hs, data = data_kids)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -57.55 -13.32   2.68  14.68  58.45
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   77.548      2.059  37.670  < 2e-16 ***
## mom_hs1       11.771      2.322   5.069 5.96e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 19.85 on 432 degrees of freedom
## Multiple R-squared:  0.05613,    Adjusted R-squared:  0.05394
## F-statistic: 25.69 on 1 and 432 DF,  p-value: 5.957e-07
```

```
summary(res_anova_1_bis)
```

```
##
## Call:
## lm(formula = kid_score ~ mom_work, data = data_kids)
##
## Residuals:
##    Min     1Q Median     3Q    Max
## -65.85 -12.85   2.79  14.15  50.50
##
## Coefficients:
##             Estimate Std. Error t value Pr(>|t|)
## (Intercept)   82.000      2.305  35.568   <2e-16 ***
## mom_work2      3.854      3.095   1.245   0.2137
## mom_work3     11.500      3.553   3.237   0.0013 **
## mom_work4      5.210      2.704   1.927   0.0547 .
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 20.23 on 430 degrees of freedom
## Multiple R-squared:  0.02444,    Adjusted R-squared:  0.01763
## F-statistic:  3.59 on 3 and 430 DF,  p-value: 0.01377
```