

# Statistique Descriptive

Kossi Tonyi Wobubey ABOTSI

default

```
## -- Attaching core tidyverse packages ----- tidyverse 2.0.0 --
## v dplyr      1.1.4      v readr      2.1.5
## v forcats    1.0.0      v stringr   1.5.1
## v ggplot2    3.5.1      v tibble    3.2.1
## v lubridate  1.9.3      v tidyr     1.3.1
## v purrr      1.0.2
## -- Conflicts ----- tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()     masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
## Le chargement a nécessité le package : emmeans
##
## Welcome to emmeans.
## Caution: You lose important information if you filter this package's results.
## See '? untidy'
##
## The 'lsmeans' package is now basically a front end for 'emmeans'.
## Users are encouraged to switch the rest of the way.
## See help('transition') for more information, including how to
## convert old 'lsmeans' objects and scripts to work with 'emmeans'.
##
## Le chargement a nécessité le package : carData
##
##
## Attachement du package : 'car'
##
##
## L'objet suivant est masqué depuis 'package:dplyr':
##
##   recode
##
##
## L'objet suivant est masqué depuis 'package:purrr':
##
##   some
##
##
## Attachement du package : 'MASS'
##
##
## L'objet suivant est masqué depuis 'package:dplyr':
##
```

```
##      select
##
##
## Le chargement a nécessité le package : mvtnorm
##
## Le chargement a nécessité le package : survival
##
## Le chargement a nécessité le package : TH.data
##
##
## Attachement du package : 'TH.data'
##
##
## L'objet suivant est masqué depuis 'package:MASS':
##
##      geyser
```

## Traitement des données

```
# Read data from the Excel file
data <- read_excel("data_send.xlsx")
#Selection des colonnes utiles
data = data %>% dplyr::select(college:classe, `taille_cm`: gender, sb:pmvpa, time, CA:`ecart_MVPA`)

#Renommage des colonnes
colnames(data)[23:24] = c("CSP_père", "CSP_mère")

#Caste
data$weight_kg = as.double(data$weight_kg)
data$taille_cm = as.double(data$taille_cm)
data$age = as.integer(data$age)
data$time = as.double(data$time)

#Ajout de colonne des IMC
data$IMC_kg_m2 <- data$weight_kg / (data$taille_cm * 10^-2)^2

# Ajout d'une nouvelle colonne "IPS_categorie"
data$IPS_categorie <- ifelse(data$IPS < 89, "Faible",
                             ifelse(data$IPS >= 90 & data$IPS <= 114, "Moyenne", "Élevée"))

#Résolution du problème de facteur double
data$gender = ifelse(data$gender != 'M' & data$gender != 'F', 'M', data$gender)

head(data)
```

```
## # A tibble: 6 x 28
##   college classe taille_cm weight_kg  age gender  sb   lpa   mpa   vpa psb
##   <chr>   <chr>    <dbl>    <dbl> <int> <chr>  <dbl> <dbl> <dbl> <dbl> <chr>
## 1 aigle   3.P         157      55    15 F    26.3  4.83  21.2  2.33 47,88~
## 2 aigle   3.P         178      61    14 M     14   8.33  28    4    25,45~
## 3 aigle   3.P         170      75    15 M    20.3  7.33  21.8  5    36,97~
## 4 aigle   3.P         153      68    15 F    26.2  7.33  18.7  2.83 47,58~
## 5 aigle   3.P         181      95    15 M    12.2 12.3  22.3  6.17 22,12~
## 6 aigle   3.P         164      51    15 F    20.5  6.5   20.3  4.83 37,27~
```

```
## # i 17 more variables: pla <chr>, pmpa <chr>, pvpa <chr>, mvpa <dbl>,
## #   pmvpa <chr>, time <dbl>, CA <dbl>, activites <chr>, IPS <dbl>,
## #   cat_IPS <chr>, Geographie <chr>, CSP_père <chr>, CSP_mère <chr>,
## #   moyenne_MVPA_classe <dbl>, ecart_MVPA <dbl>, IMC_kg_m2 <dbl>,
## #   IPS_categorie <chr>
```

### 1) Les écarts de MVPA entre filles et garçons selon les CA (CA1, CA2, CA3, CA4)

```
#Selection des colonnes importantes
data_1 = data %>%
  dplyr::select(gender,ecart_MVPA,CA)

#Assurons nous que le types des variables qualitatives sont des facteurs
data_1$gender = as.factor(data_1$gender)
data_1$CA = as.factor(data_1$CA)

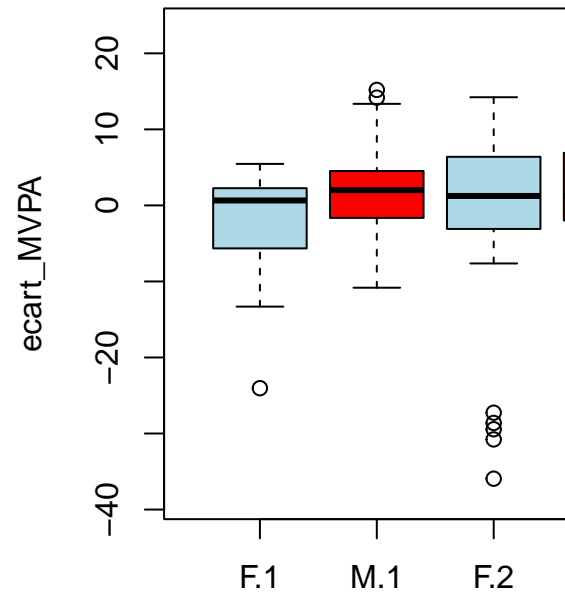
data_1
```

#### Selection des colonnes importantes

```
## # A tibble: 462 x 3
##   gender ecart_MVPA CA
##   <fct>      <dbl> <fct>
## 1 F      -5.28  3
## 2 M       3.55  3
## 3 M      -1.78  3
## 4 F      -7.62  3
## 5 M       1.38  3
## 6 F      -1.12  3
## 7 M      -0.615 3
## 8 M       3.38  3
## 9 M      20.9   3
## 10 F     -0.782 3
## # i 452 more rows
```

```
#write_xlsx(data_1, "data_gpt.xlsx")
```

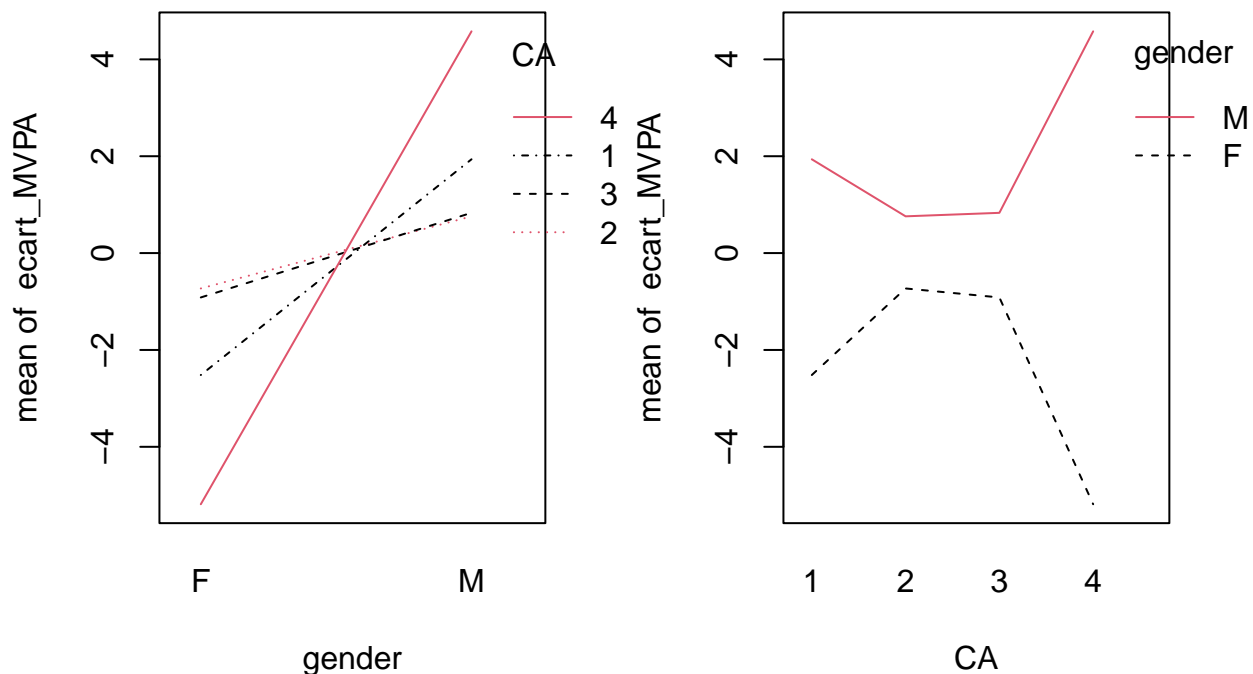
```
boxplot(ecart_MVPA~gender*CA,data=data_1,col=c(rep('lightblue',1),rep('red',1)))
```



### Représentation des données et Analyse graphique de l'interaction

On remarque que généralement le mvpa des est supérieur dans les CA 1 et 2 que dans les CA 3 et 4.

```
par(mfrow=c(1,2))
with(data_1, interaction.plot(gender,CA,ecart_MVPA,col=1:2))
with(data_1, interaction.plot(CA,gender,ecart_MVPA,col=1:4))
```



Les profils ne sont pas parallèles. Nous constatons que les modalités **2-F** et **4-F** sont éloigné de la position qu'elle auraient dû occuper si les profils étaient exactement parallèles. Les filles dans le CA 2 sont propice à un mvpa élevé. L'effet de ma variable genre sur le mvpa depend du CA. Donc ce graphique suggère une interaction entre CA et Sexe.

Maintenant est ce que cette interaction observée est significative ?

```
table(data_1$gender,data_1$CA)
```

Modèle :

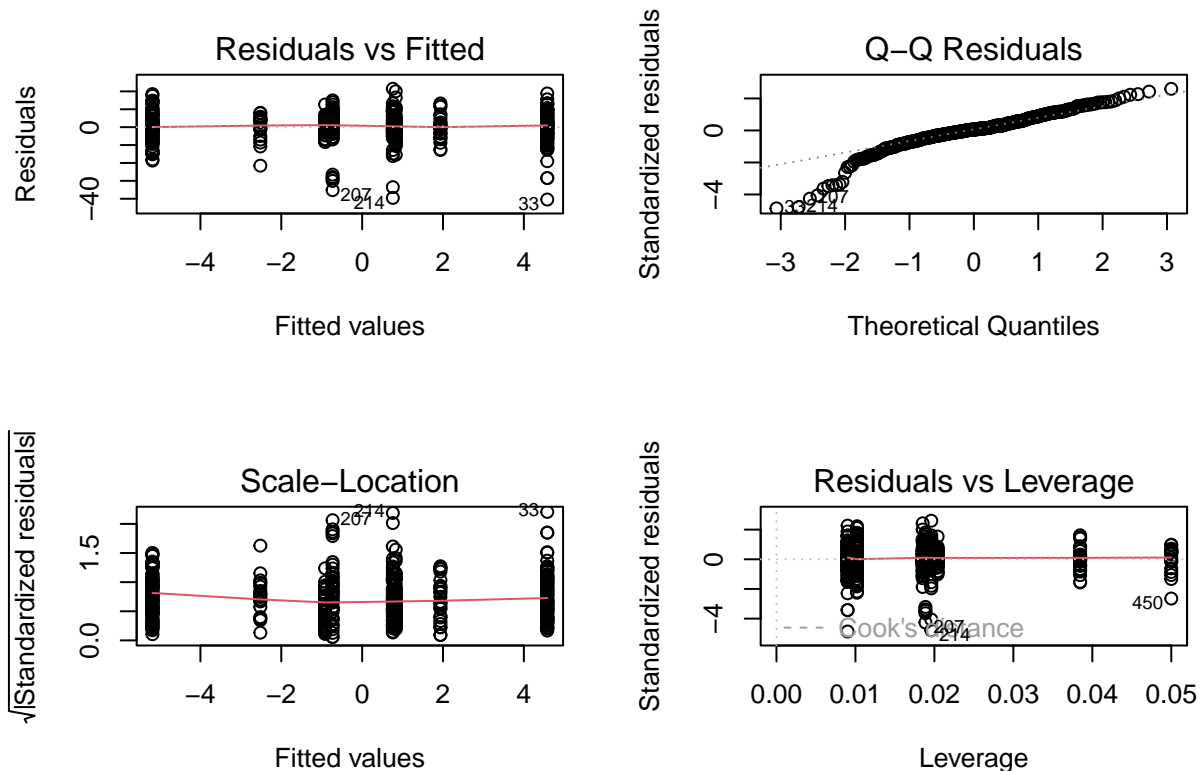
```
##
##      1  2  3  4
##  F  20 53 49 98
##  M  26 51 54 111
```

Dans notre cas , le nombre d'observation diffère selon le croisement des modalités de **CA** et **gender**. Donc nos données sont **déséquilibrées**. Lors de la décomposition de la variabilité, ce déséquilibre a un impact sur le calcul des différentes sommes de carrés, et par suite sur les tests qui sont construits. Il est donc nécessaire de choisir de quel type de somme de carré seront utilisés (type I, typeII, ou type III). Pour construire notre modèle j'utilise la somme des carré de type III avec la contrainte de somme des coefficients nulle.

```
modele_complet_sortie = AovSum(ecart_MVPA~CA+gender+CA:gender,data = data_1)
```

## Analyse des résidus

```
# Ajustement du modèle ANOVA
modele_complet <- aov(ecart_MVPA ~ gender + CA + gender:CA, data = data_1)
par(mfrow = c(2,2))
plot(modele_complet)
```



```
# Diagnostic des résidus
# Test de normalité des résidus
#residus <- residuals(modele_complet)
#shapiro_test <- shapiro.test(residus)
#print(shapiro_test)

# Test d'homogénéité des variances
#levene_test <- leveneTest(mvpa ~ gender*CA, data = data_1)
#print(levene_test)
```

Le graphique en haut à gauche, qui représente en abscisse les valeurs prédites et en ordonnée les résidus permet de vérifier que la variance des résidus est similaire dans tous les groupes. Le graphique du QQ-plot, permet de vérifier que l'hypothèse de normalité des résidus est raisonnable.

Avec ces hypothèses validées nous pouvons regarder les tests.

## Les tests du modèle complet et des différents effets

```
modele_complet_sortie
```

```
## Ftest
##              SS   df      MS F value    Pr(>F)
## CA              9.6    3      3.19  0.0459    0.9869
## gender        1646.5    1 1646.46 23.6602 1.588e-06 ***
## CA:gender      1738.0    3  579.33  8.3252 2.122e-05 ***
## Residuals    31592.8  454    69.59
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Ttest
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   -0.155573    0.448593  -0.3468  0.72890
## CA - 1        -0.134955    0.985249  -0.1370  0.89111
## CA - 2         0.169898    0.732063   0.2321  0.81658
## CA - 3         0.113170    0.734737   0.1540  0.87766
## CA - 4        -0.148114    0.606927  -0.2440  0.80731
## gender - F     -2.182033    0.448593  -4.8642 < 2e-16 ***
## gender - M      2.182033    0.448593   4.8642 < 2e-16 ***
## CA - 1 : gender - F -0.045344    0.985249  -0.0460  0.96331
## CA - 2 : gender - F  1.437123    0.732063   1.9631  0.05024 .
## CA - 3 : gender - F  1.308541    0.734737   1.7810  0.07559 .
## CA - 4 : gender - F -2.700320    0.606927  -4.4492  1e-05 ***
## CA - 1 : gender - M  0.045344    0.985249   0.0460  0.96331
## CA - 2 : gender - M -1.437123    0.732063  -1.9631  0.05024 .
## CA - 3 : gender - M -1.308541    0.734737  -1.7810  0.07559 .
## CA - 4 : gender - M  2.700320    0.606927   4.4492  1e-05 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Pour le test de comparaison du modèle avec interaction (modèle complet) et du modèle sans interaction, on a les hypothèses :

- $H_0$  : Il n'y a pas d'interaction.
- $H_1$  : il y a des interactions.

On observe une p-valeur inférieur au seuil de 5% donc les interactions entre le genre et CA sont significativement différent de zéro.

## Test post hoc

```
lsmeans_results <- lsmeans(modele_complet, pairwise ~ gender:CA, adjust='Tukey')
summary(lsmeans_results, infer = c(TRUE, TRUE))
```

```
## $lsmeans
##   gender CA lsmean    SE   df lower.CL upper.CL t.ratio p.value
## F      1  -2.518 1.865 454    -6.18     1.15  -1.350  0.1777
## M      1   1.937 1.636 454    -1.28     5.15   1.184  0.2371
## F      2  -0.731 1.146 454    -2.98     1.52  -0.638  0.5241
## M      2   0.759 1.168 454    -1.54     3.05   0.650  0.5160
## F      3  -0.916 1.192 454    -3.26     1.43  -0.769  0.4426
## M      3   0.831 1.135 454    -1.40     3.06   0.732  0.4645
```

```
## F      4 -5.186 0.843 454 -6.84 -3.53 -6.154 <.0001
## M      4  4.579 0.792 454  3.02  6.13  5.783 <.0001
##
## Confidence level used: 0.95
##
## $contrasts
## contrast      estimate    SE df lower.CL upper.CL t.ratio p.value
## F CA1 - M CA1 -4.4548 2.48 454 -12.010    3.100 -1.795 0.6235
## F CA1 - F CA2 -1.7873 2.19 454 -8.454    4.879 -0.816 0.9922
## F CA1 - M CA2 -3.2771 2.20 454 -9.979    3.425 -1.489 0.8132
## F CA1 - F CA3 -1.6020 2.21 454 -8.342    5.138 -0.724 0.9963
## F CA1 - M CA3 -3.3490 2.18 454 -9.998    3.300 -1.534 0.7888
## F CA1 - F CA4  2.6681 2.05 454 -3.565    8.901  1.304 0.8973
## F CA1 - M CA4 -7.0966 2.03 454 -13.267   -0.926 -3.502 0.0118
## M CA1 - F CA2  2.6674 2.00 454 -3.415    8.750  1.335 0.8849
## M CA1 - M CA2  1.1776 2.01 454 -4.944    7.299  0.586 0.9990
## M CA1 - F CA3  2.8527 2.02 454 -3.311    9.016  1.409 0.8527
## M CA1 - M CA3  1.1058 1.99 454 -4.958    7.169  0.555 0.9993
## M CA1 - F CA4  7.1229 1.84 454  1.519   12.727  3.871 0.0031
## M CA1 - M CA4 -2.6418 1.82 454 -8.176    2.893 -1.454 0.8314
## F CA2 - M CA2 -1.4898 1.64 454 -6.473    3.493 -0.910 0.9850
## F CA2 - F CA3  0.1853 1.65 454 -4.849    5.220  0.112 1.0000
## F CA2 - M CA3 -1.5617 1.61 454 -6.473    3.350 -0.968 0.9786
## F CA2 - F CA4  4.4555 1.42 454  0.124    8.787  3.132 0.0387
## F CA2 - M CA4 -5.3092 1.39 454 -9.550   -1.068 -3.812 0.0039
## M CA2 - F CA3  1.6751 1.67 454 -3.406    6.757  1.004 0.9738
## M CA2 - M CA3 -0.0719 1.63 454 -5.032    4.888 -0.044 1.0000
## M CA2 - F CA4  5.9453 1.44 454  1.559   10.331  4.128 0.0011
## M CA2 - M CA4 -3.8194 1.41 454 -8.117    0.478 -2.707 0.1232
## F CA3 - M CA3 -1.7470 1.65 454 -6.759    3.265 -1.061 0.9643
## F CA3 - F CA4  4.2701 1.46 454 -0.174    8.715  2.926 0.0699
## F CA3 - M CA4 -5.4946 1.43 454 -9.851   -1.138 -3.840 0.0035
## M CA3 - F CA4  6.0171 1.41 454  1.712   10.322  4.256 0.0007
## M CA3 - M CA4 -3.7476 1.38 454 -7.962    0.467 -2.708 0.1229
## F CA4 - M CA4 -9.7647 1.16 454 -13.286   -6.244 -8.445 <.0001
##
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 8 estimates
## P value adjustment: tukey method for comparing a family of 8 estimates
```

### Diagramme en barre de la différence des écart\_mvpa des Filles et Garçons

```
# Calcul des moyennes marginales estimées et des comparaisons post hoc avec correction de Tukey
summary_data <- summary(lsmeans_results$contrasts)

# Filtrer les contrastes pour ne garder que ceux entre F et M au sein de chaque catégorie d'âge (CA)
plot_data <- summary_data[summary_data$contrast %in% c("F CA1 - M CA1", "F CA2 - M CA2", "F CA3 - M CA3"), ]

# Préparation des données pour le graphique
plot_data <- as.data.frame(plot_data[c("contrast", "estimate", "p.value")])
names(plot_data) <- c("Contrast", "Estimate", "P_Value")

# Extraire la catégorie d'âge à partir du nom du contraste
plot_data$CA <- sub(".*CA([0-9]+).*", "\\1", plot_data$Contrast)
```



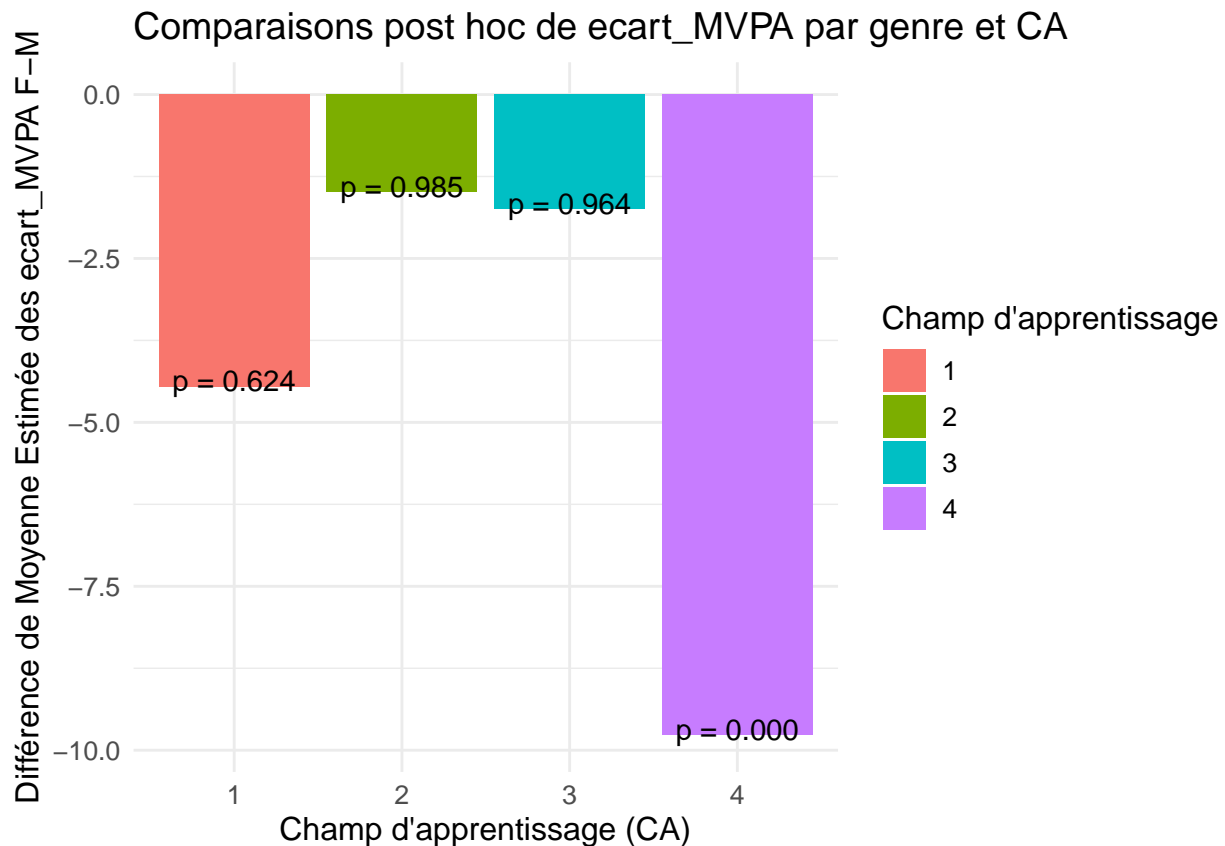
```

# Déterminer le décalage pour les étiquettes textuelles
max_estimate <- max(plot_data$Estimate) * 0.05

# Création du barplot avec légende pour les catégories d'âge
p <- ggplot(plot_data, aes(x = CA, y = Estimate, fill = CA)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  geom_text(aes(label = sprintf("p = %.3f", P_Value),
    y = Estimate + max_estimate), # Utilisation de la variable calculée hors de aes()
    position = position_dodge(width = 0.99), vjust = 0) +
  labs(title = "Comparaisons post hoc de ecart_MVPA par genre et CA",
    x = "Champ d'apprentissage (CA)",
    y = "Différence de Moyenne Estimée des ecart_MVPA F-M",
    fill = "Champ d'apprentissage") + # Ajout de la légende
  theme_minimal() +
  theme(text = element_text(size=12))

print(p)

```



On remarque un écart significatif entre les écarts de mvpa à la moyenne de chaque classe des filles et garçons dans le champs 4 avec une p-valeur inférieur au seuil de 5%. Donc la moyenne des écarts à la moyenne de chaque classe des garçons est plus élevé dans le champs 4 que celui des filles. Dans les autres champs d'apprentissage cet écart de la moyenne des écart n'est pas significatif.

2) Les écarts de MVPA entre filles et garçons selon la catégorie d'IPS du collège (élevé, moyen, faible)

```

#Selection des colonnes importantes
data_2 = data %>%
  dplyr::select(gender,ecart_MVPA,IPS_categorie)

#Assurons nous que le types des variables qualitatives sont des facteurs
data_2$gender = as.factor(data_2$gender)
data_2$IPS_categorie = as.factor(data_2$IPS_categorie)

data_2

```

### Selection des colonnes importantes

```

## # A tibble: 462 x 3
##   gender ecart_MVPA IPS_categorie
##   <fct>      <dbl> <fct>
## 1 F        -5.28 Moyenne
## 2 M         3.55 Moyenne
## 3 M        -1.78 Moyenne
## 4 F        -7.62 Moyenne
## 5 M         1.38 Moyenne
## 6 F        -1.12 Moyenne
## 7 M        -0.615 Moyenne
## 8 M         3.38 Moyenne
## 9 M        20.9  Moyenne
## 10 F       -0.782 Moyenne
## # i 452 more rows

table(data_2$gender,data_2$IPS_categorie)

```

```

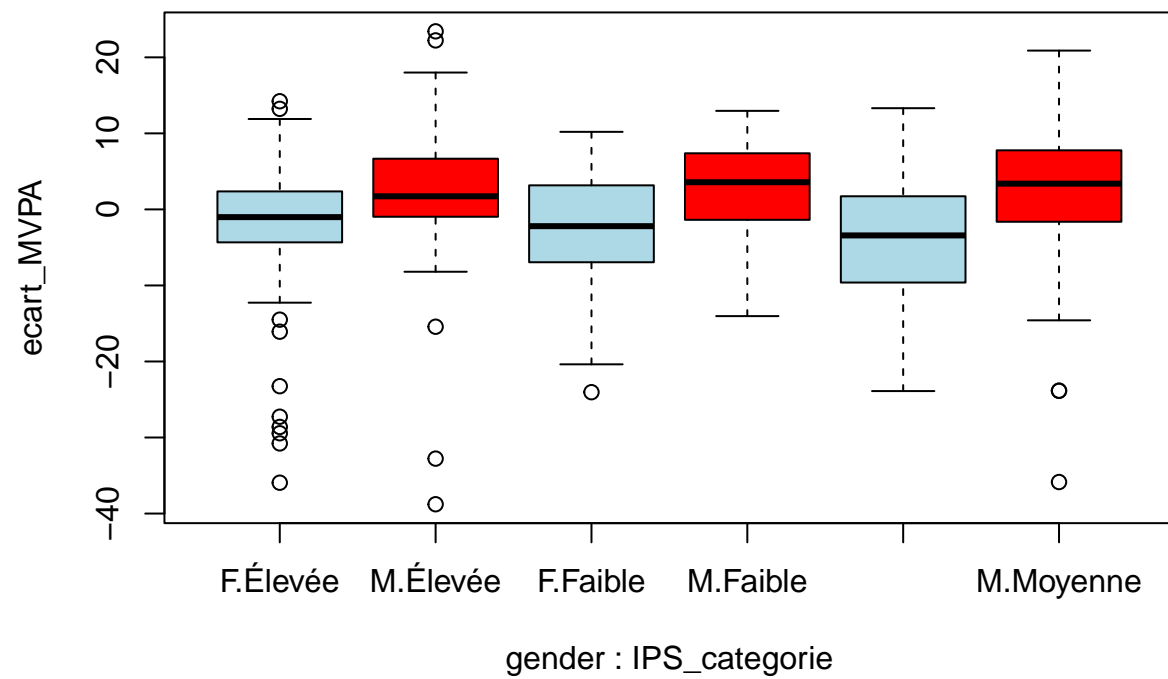
##
##   Élevée Faible Moyenne
##   F    93    48    79
##   M    89    44   109

```

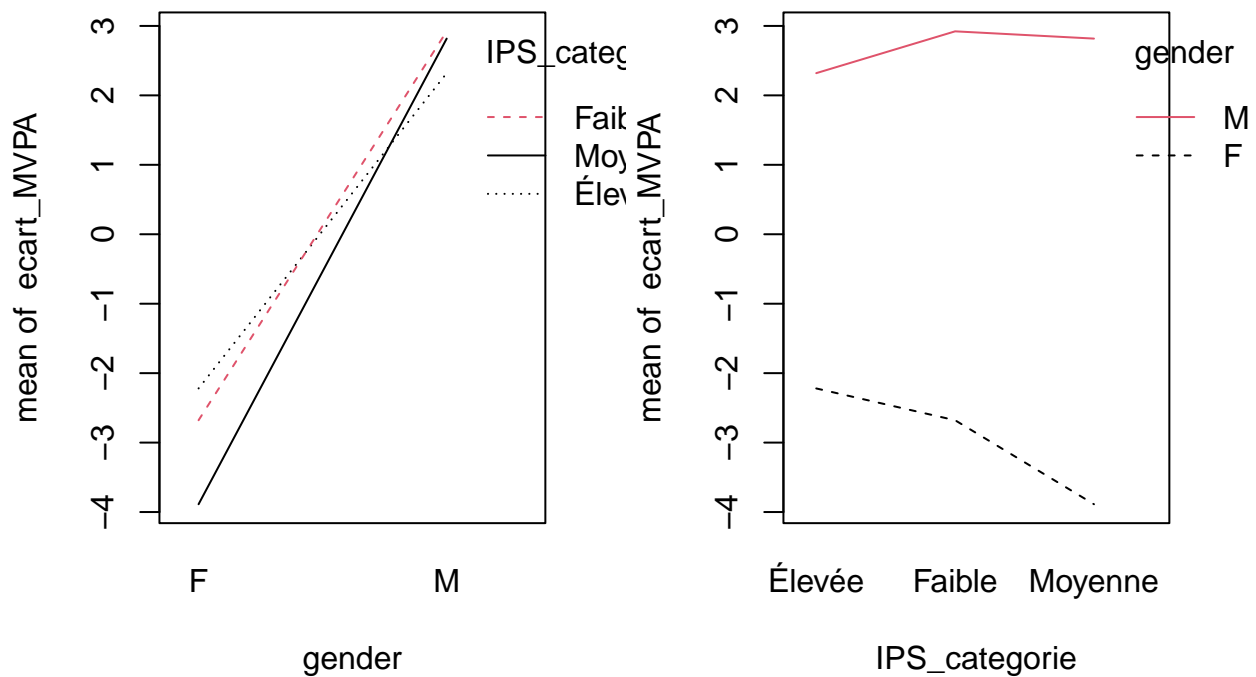
```

boxplot(ecart_MVPA~gender*IPS_categorie,data=data_2,col=c(rep('lightblue',1),rep('red',1)))

```



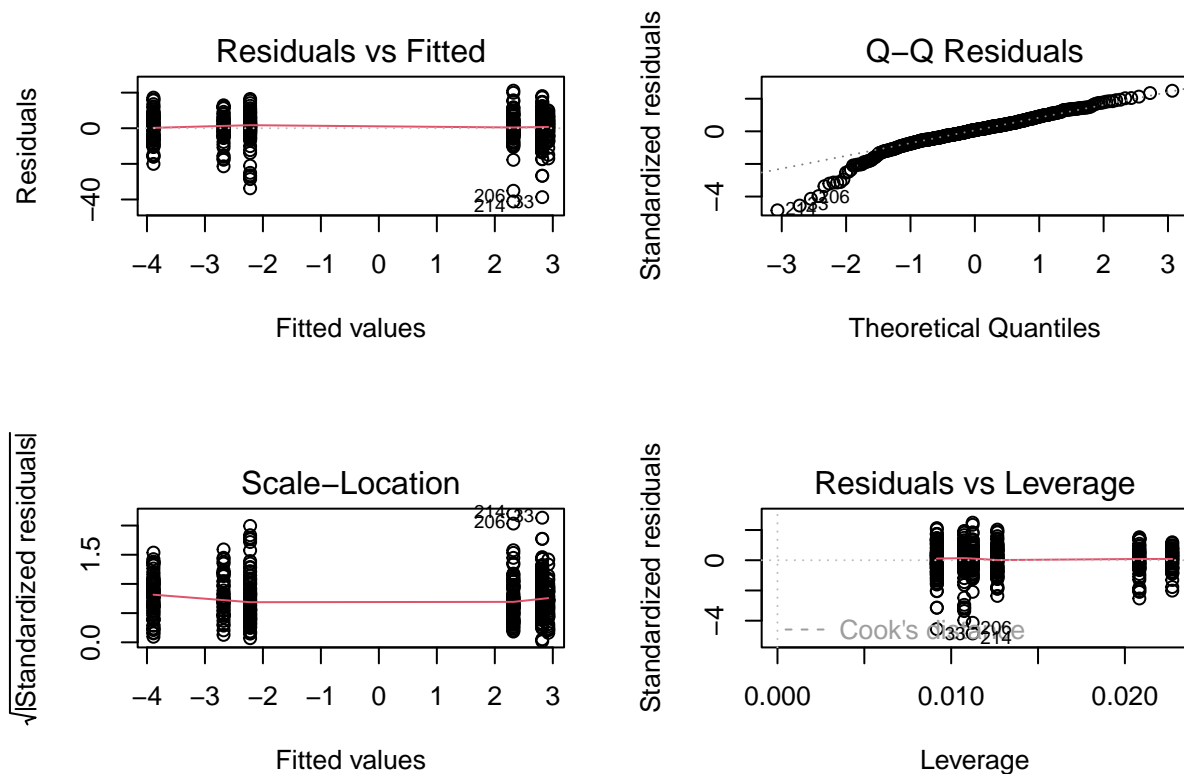
```
par(mfrow=c(1,2))
with(data_2, interaction.plot(gender, IPS_categorie, ecart_MVPA, col=1:2))
with(data_2, interaction.plot(IPS_categorie, gender, ecart_MVPA, col=1:3))
```



```
modele_complet_III = AovSum(ecart_MVPA~gender+IPS_categorie+gender:IPS_categorie,data = data_2)

# Ajustement du modèle ANOVA
modele_complet <- aov(ecart_MVPA ~ gender + IPS_categorie + gender:IPS_categorie, data = data_2)

# Résumé du modèle
par(mfrow = c(2,2))
plot(modele_complet)
```



```
# Diagnostic des résidus
# Test de normalité des résidus
#residus <- residuals(modele_complet)
#shapiro_test <- shapiro.test(residus)
#print(shapiro_test)

# Test d'homogénéité des variances
#levene_test <- leveneTest(residus, data_2$IPS_categorie) # Assurez-vous de remplacer 'CA' par la vari
#print(levene_test)
```

Le graphique en haut à gauche, qui représente en abscisse les valeurs prédites et en ordonnée les résidus permet de vérifier que la variance des résidus est similaire dans tous les groupes. Le graphique du QQ-plot, permet de vérifier que l'hypothèse de normalité des résidus est raisonnable.

Avec ces hypothèses validées nous pouvons regarder les tests.

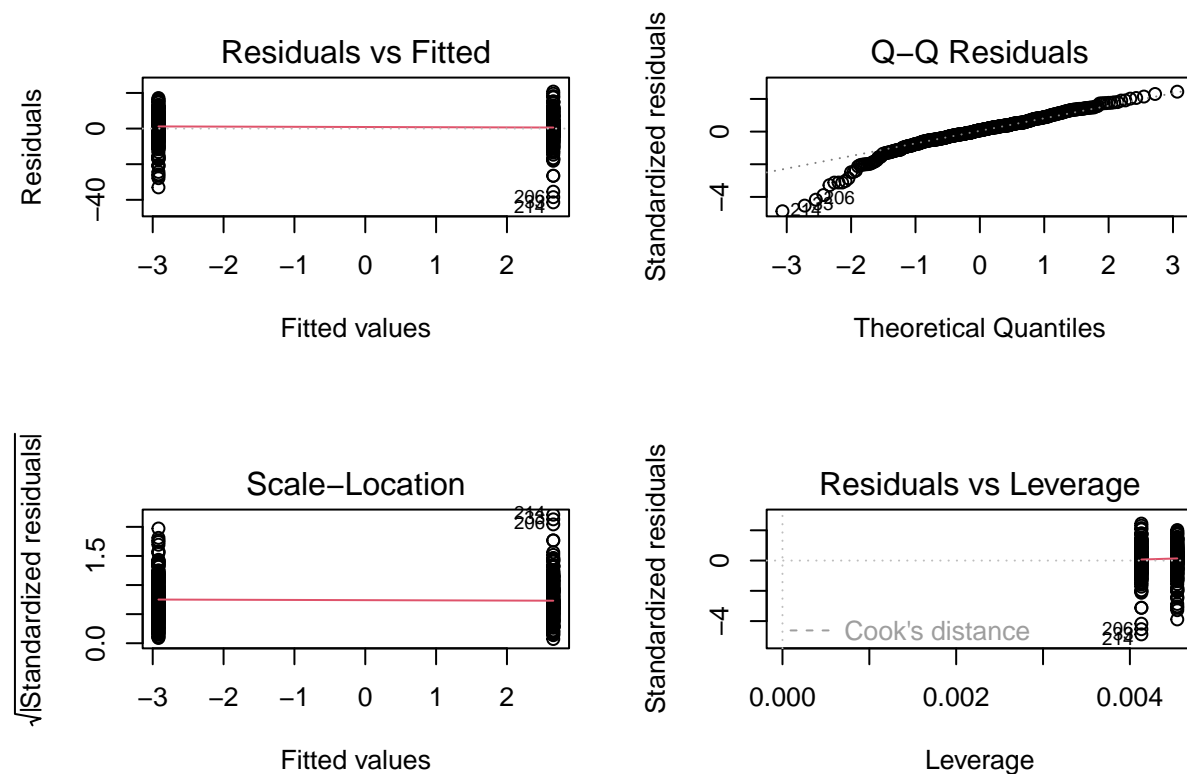
```
modele_complet_III
```

```
## Ftest
##               SS   df      MS F value    Pr(>F)
## gender       3248    1 3247.8 44.6101 7.016e-11 ***
## IPS_categorie    41    2   20.5  0.2817  0.7546
## gender:IPS_categorie 107    2   53.5  0.7349  0.4801
## Residuals     33199 456   72.8
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Ttest
```

```
##                                Estimate Std. Error t value Pr(>|t|)
## (Intercept)                   -0.1211264  0.4203765  -0.2881   0.7734
## gender - F                    -2.8077290  0.4203765  -6.6791  <2e-16 ***
## gender - M                     2.8077290  0.4203765   6.6791  <2e-16 ***
## IPS_categorie - Élevée         0.1710201  0.5568865   0.3071   0.7589
## IPS_categorie - Faible         0.2428752  0.6640788   0.3657   0.7147
## IPS_categorie - Moyenne       -0.4138953  0.5560355  -0.7444   0.4570
## gender - F : IPS_categorie - Élevée  0.5375673  0.5568865   0.9653   0.3349
## gender - M : IPS_categorie - Élevée -0.5375673  0.5568865  -0.9653   0.3349
## gender - F : IPS_categorie - Faible  0.0075062  0.6640788   0.0113   0.9910
## gender - M : IPS_categorie - Faible -0.0075062  0.6640788  -0.0113   0.9910
## gender - F : IPS_categorie - Moyenne -0.5450735  0.5560355  -0.9803   0.3275
## gender - M : IPS_categorie - Moyenne  0.5450735  0.5560355   0.9803   0.3275
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Les interactions entre IPS\_categorie et genre et le facteur IPS\_categorie ne sont pas significative, en effet leurs p-valeurs respectifs sont inférieure à 5% seuil du test. Donc le modèle optimal est celui avec le facteur genre uniquement.

```
modele = aov(ecart_MVPA ~ gender, data = data_2)
par(mfrow = c(2,2))
plot(modele)
```



Il n'y pas de problème d'égalité de variance ou de normalité sur ce modèle.

En ce qui concerne les tests :

```
modele_III = AovSum(ecart_MVPA~gender,data = data_2)
modele_III
```

```
## Ftest
##              SS  df      MS F value    Pr(>F)
## gender      3579   1 3578.7  49.381 7.642e-12 ***
## Residuals 33337 460   72.5
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Ttest
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -0.13268    0.39651 -0.3346  0.7381
## gender - F   -2.78636    0.39651 -7.0272 <2e-16 ***
## gender - M    2.78636    0.39651  7.0272 <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

### Test post hoc

```
lsmeans(modele, pairwise ~ gender, adjust='Tukey')
```

```
## $lsmeans
##   gender lsmean    SE  df lower.CL upper.CL
## F      -2.92 0.574 460    -4.05    -1.79
## M       2.65 0.547 460     1.58     3.73
##
## Confidence level used: 0.95
##
## $contrasts
##   contrast estimate    SE  df t.ratio p.value
## F - M      -5.57 0.793 460   -7.027 <.0001
```

Donc ici l'IPS\_categorie n'a aucune influence sur les écarts de mvpa à la moyenne de chaque classe des filles et garçons.

### 3) Les écarts de MVPA entre filles et garçons selon le milieu géographique (urbain, rural)

```
#Selection des colonnes importantes
data_3 = data %>%
  dplyr::select(gender,ecart_MVPA,Geographie)

#Assurons nous que le types des variables qualitatives sont des facteurs
data_3$gender = as.factor(data_3$gender)
data_3$Geographie = as.factor(data_3$Geographie)

data_3
```

#### Selection des colonnes importantes

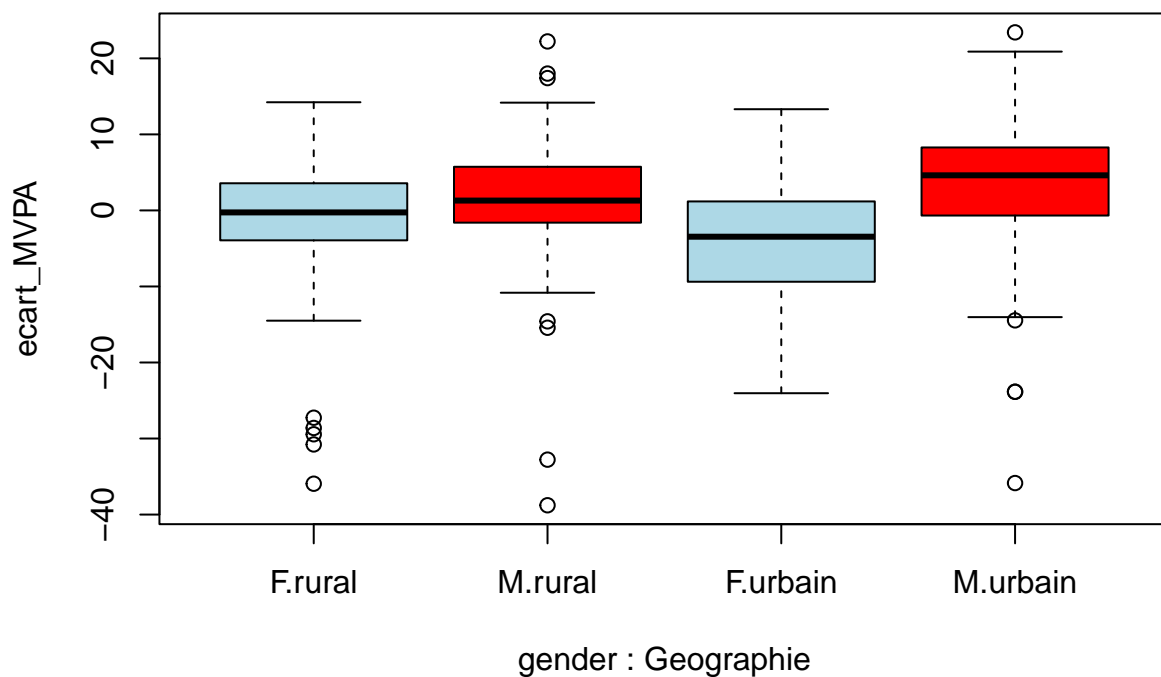
```
## # A tibble: 462 x 3
##   gender ecart_MVPA Geographie
##   <fct>      <dbl> <fct>
## 1 F      -5.28  urbain
```

```
## 2 M      3.55 urbain
## 3 M     -1.78 urbain
## 4 F     -7.62 urbain
## 5 M      1.38 urbain
## 6 F     -1.12 urbain
## 7 M     -0.615 urbain
## 8 M      3.38 urbain
## 9 M     20.9  urbain
## 10 F    -0.782 urbain
## # i 452 more rows
```

```
table(data_3$gender,data_3$Geographie)
```

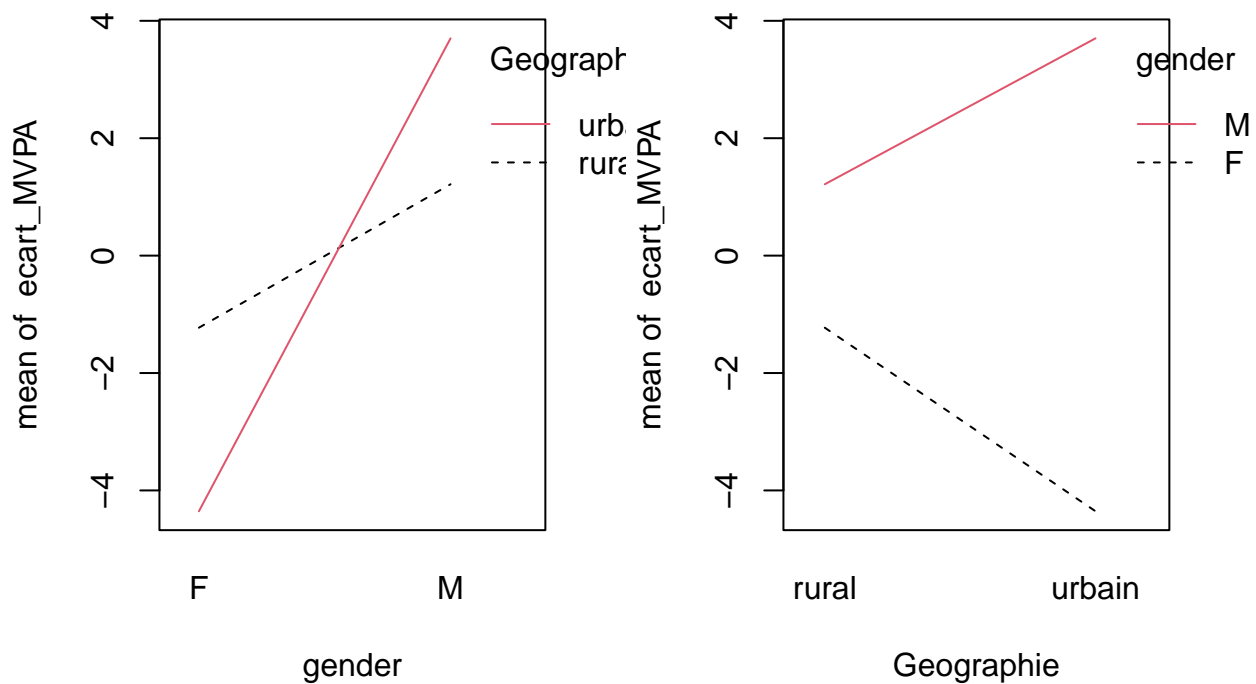
```
##
##      rural urbain
## F    101   119
## M    102   140
```

```
boxplot(ecart_MVPA~gender*Geographie,data=data_3,col=c(rep('lightblue',1),rep('red',1)))
```



```
par(mfrow=c(1,2))
with(data_3, interaction.plot(gender,Geographie,ecart_MVPA,col=1:2))
with(data_3, interaction.plot(Geographie,gender,ecart_MVPA,col=1:3))
```



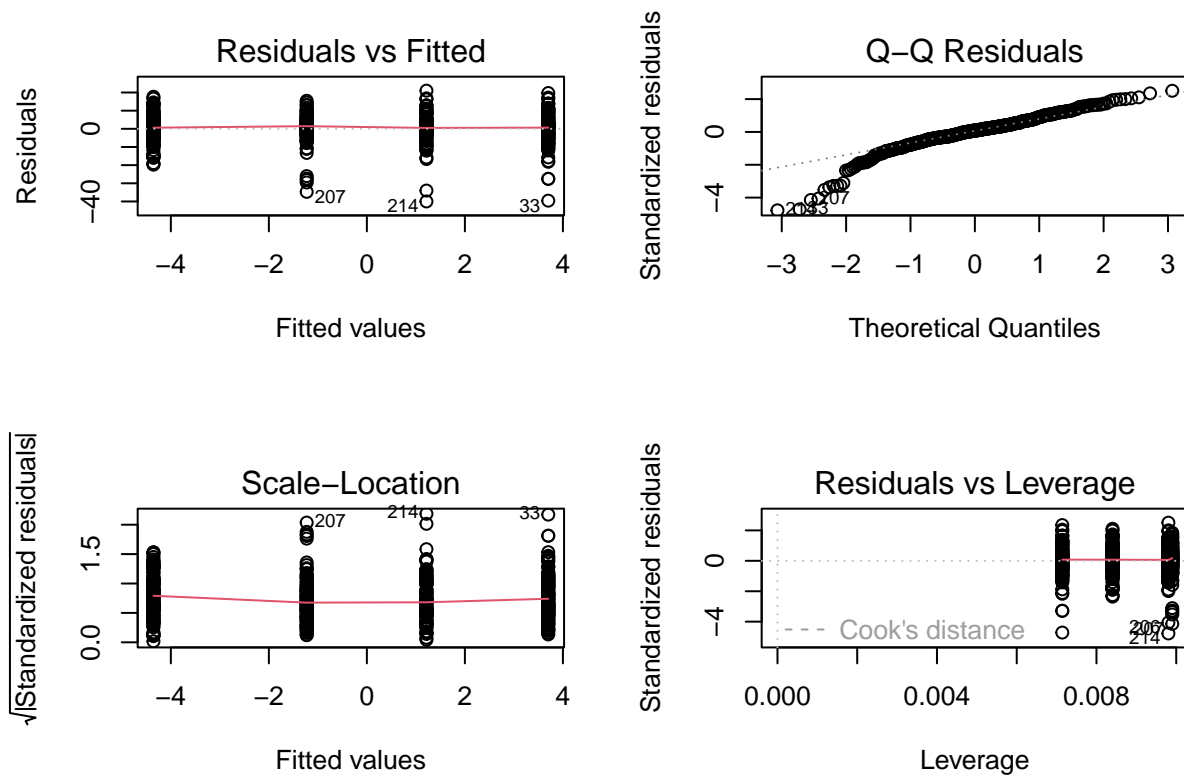


```
modele_complet_III = AovSum(ecart_MVPA~gender+Geographie+gender:Geographie,data = data_3)
modele_complet_III$Ftest
```

```
##              SS   df      MS F value    Pr(>F)
## gender          3127    1 3127.17 44.1525 8.645e-11 ***
## Geographie         12    1   11.66  0.1646 0.6851157
## gender:Geographie   893    1  893.44 12.6145 0.0004224 ***
## Residuals        32439  458   70.83
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
# Ajustement du modèle ANOVA
modele_complet <- aov(ecart_MVPA ~ gender + Geographie + gender:Geographie, data = data_3)

# Résumé du modèle
par(mfrow = c(2,2))
plot(modele_complet)
```



```
# Diagnostic des résidus
# Test de normalité des résidus
#residus <- residuals(modele_complet)
#shapiro_test <- shapiro.test(residus)
#print(shapiro_test)

# Test d'homogénéité des variances
#levene_test <- leveneTest(residus, data_2$IPS_categorie) # Assurez-vous de remplacer 'CA' par la vari
#print(levene_test)
```

Les hypothèses de normalité et d'égalité semble être validée.

```
lsmeans_results <- lsmeans(modele_complet, pairwise ~ gender:Geographie, adjust='Tukey')
```

```
# Affichage des résultats avec les intervalles de confiance
summary(lsmeans_results, infer = c(TRUE, TRUE))
```

```
## $lsmeans
##   gender Geographie lsmean    SE  df lower.CL upper.CL t.ratio p.value
## F      rural      -1.23 0.837 458  -2.873    0.418  -1.466  0.1433
## M      rural       1.22 0.833 458   -0.422    2.853   1.459  0.1452
## F      urbain     -4.35 0.771 458  -5.871   -2.838  -5.644 <.0001
## M      urbain      3.70 0.711 458   2.304    5.099   5.204 <.0001
##
## Confidence level used: 0.95
##
## $contrasts
```

```
## contrast      estimate    SE  df lower.CL upper.CL t.ratio p.value
## F rural - M rural    -2.44 1.18 458   -5.490    0.602  -2.068  0.1650
## F rural - F urbain    3.13 1.14 458    0.191    6.062   2.746  0.0317
## F rural - M urbain   -4.93 1.10 458   -7.762   -2.096  -4.486  0.0001
## M rural - F urbain    5.57 1.14 458    2.642    8.498   4.905  <.0001
## M rural - M urbain   -2.49 1.10 458   -5.310    0.339  -2.269  0.1070
## F urbain - M urbain   -8.06 1.05 458  -10.761   -5.350  -7.677  <.0001
##
## Confidence level used: 0.95
## Conf-level adjustment: tukey method for comparing a family of 4 estimates
## P value adjustment: tukey method for comparing a family of 4 estimates

# Calcul des moyennes marginales estimées et des comparaisons post hoc avec correction de Tukey
lsmeans_results <- lsmeans(modele_complet, pairwise ~ gender:Geographie, adjust='Tukey')
summary_data <- summary(lsmeans_results$contrasts)

# Filtrer les contrastes pour ne garder que ceux entre F et M au sein de chaque niveau de géographie
plot_data <- summary_data[summary_data$contrast %in% c("F rural - M rural", "F urbain - M urbain"), ]

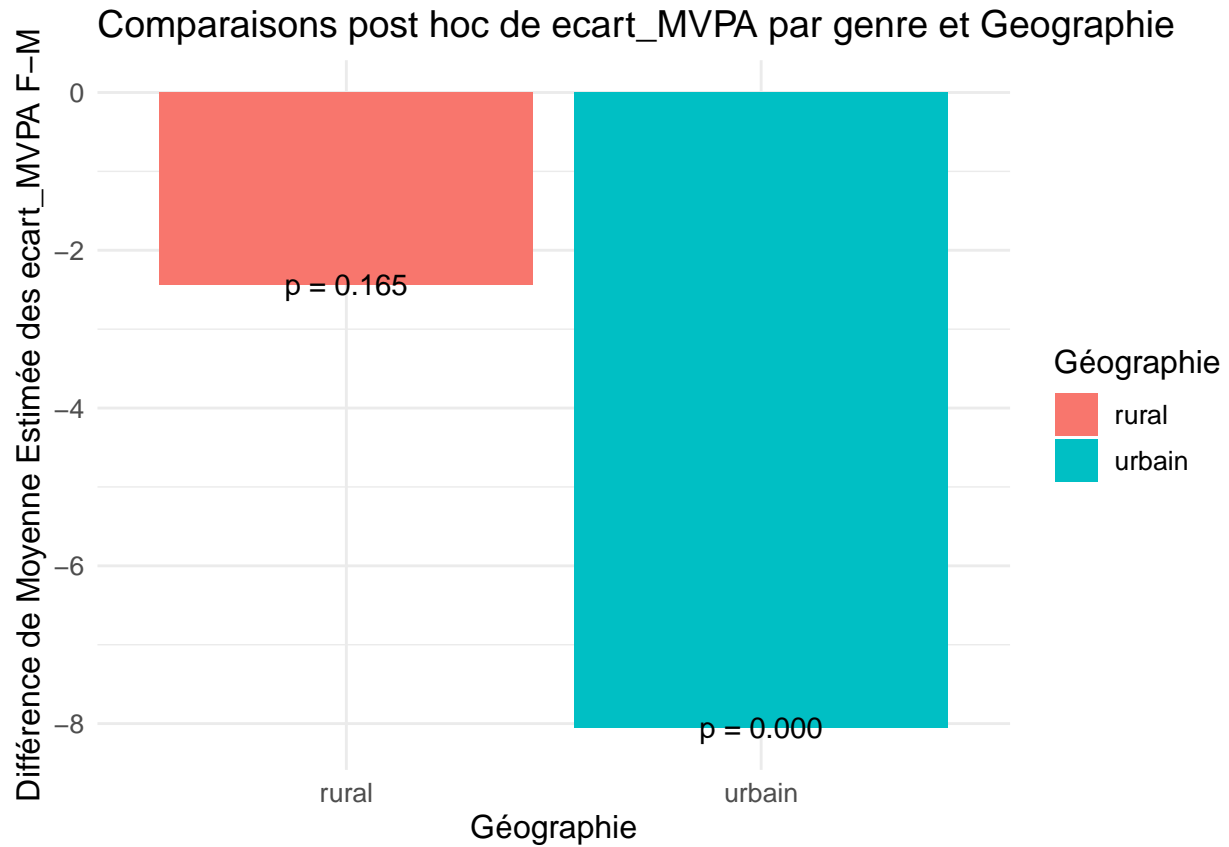
# Préparation des données pour le graphique
plot_data <- as.data.frame(plot_data[c("contrast", "estimate", "p.value")])
names(plot_data) <- c("Contrast", "Estimate", "P_Value")

# Extraire le niveau de géographie à partir du nom du contraste
plot_data$Geographie <- sub(".* (rural|urbain) -.*", "\\1", plot_data$Contrast)

# Déterminer le décalage pour les étiquettes textuelles
max_estimate <- max(plot_data$Estimate) * 0.05

# Création du barplot avec légende pour les niveaux de géographie
p <- ggplot(plot_data, aes(x = Geographie, y = Estimate, fill = Geographie)) +
  geom_bar(stat = "identity", position = position_dodge()) +
  geom_text(aes(label = sprintf("p = %.3f", P_Value),
    y = Estimate + max_estimate), # Utilisation de la variable calculée hors de aes()
    position = position_dodge(width = 0.99), vjust = 0) +
  labs(title = "Comparaisons post hoc de ecart_MVPA par genre et Geographie",
    x = "Géographie",
    y = "Différence de Moyenne Estimée des ecart_MVPA F-M",
    fill = "Géographie") + # Ajout de la légende
  theme_minimal() +
  theme(text = element_text(size=12))

print(p)
```



On remarque un écart significatif entre les écarts de mvpa à la moyenne de chaque classe des filles et garçons dans le milieu urbain avec une p-valeur inférieure au seuil de 5%. Donc dans le milieu urbain en moyenne les écarts à la moyenne de chaque classe des garçons est plus élevé que celui des filles par contre cette différence entre la moyenne des écarts n'est pas significatif dans le milieu rural.