



جامعة سيدي محمد بن عبد الله
كلية العلوم ظهر المهرار

جامعة سيدي محمد بن عبد الله
كلية العلوم ظهر المهرار

Université Sidi Mohamed Ben Abdellah
Faculté des Sciences Dhar Mahraz



Master Web Intelligence & Sciences des Données

WISD



Modélisation & Simulation

Thème :

**La génétique de l'obésité : mythes et réalités
(Two-Way ANOVA)**



Réaliser par :

- EL AYADY ANASS
- AIT TOUCHENT ABDELLATIF

Sous la direction de :

- Pr. ALJ ABDEL KAMEL

Mai, 2023

Table des matières

Introduction générale.....	3
I - Problématique.....	5
II - Etude de l'existant :.....	6
1 - Jeu de données globale :.....	6
2 - Preprocessing :.....	7
i - Pandas :.....	7
ii - Jeu de données final :.....	8
iii - Hypothèse dans le test Two-Way ANOVA :.....	8
3 - Régression Linéaire Simple :.....	9
i - Définition :.....	9
ii - Objectif :.....	9
iii - Modèle de RLS :.....	9
iv - Interprétation de la sortie :.....	9
4 - Homogénéité des variances :.....	10
i - Test de Levene :.....	10
ii - Test de Bartlett :.....	10
5 - Test de normalité :.....	11
6 - Ajustement du modèle ANOVA :.....	12
7 - Synthèse :.....	12
III - Analyse of variance (ANOVA) :.....	13
1 - Le modèle d'ANOVA :.....	13
i - Hypothèse nulle et alternative :.....	14
ii - Les moyennes entre les groupes : (Voir l'annexe).....	14
iii - Boxplot des moyennes de poids par variables qualitatives :.....	15
2 - Le test d'ANOVA :.....	16
i - ANOVA en R :.....	16
ii - La violence de l'homogénéité des variances :.....	17
iii - Welch ANOVA :.....	17
3 - Tests de comparaison :.....	18
i - TukeyHSD test :.....	18
ii - TukeyHSD en R : (Voir l'annexe).....	18
iii - Estimated marginal means (emmeans) : (Voir annexe).....	19
Conclusion :.....	21

Table des Figures

Figure 1 : Les causes de l'obésité	5
Figure 2: obésité dans le monde	6
Figure 3: global Dataset.....	6
Figure 4: Data Preprocessing.....	7
Figure 5: logo de pandas	7
Figure 6 : Tableaux de données	8
Figure 7: RLS Model en R.....	9
Figure 8: Levene Test.....	10
Figure 9: Bartlett Test.....	11
Figure 10: les variances des groupes.....	11
Figure 11: test de normalité.....	11
Figure 12: Renvoie les groupes en R	14
Figure 13: les 4 groupes	14
Figure 14: les moyennes entre les groupes	14
Figure 15: les moyennes de poids.....	15
Figure 16: ANOVA en R.....	16
Figure 17: table d'ANOVA	16
Figure 18: Le modèle de Welch ANOVA.....	17
Figure 19: La sortie en R	17
Figure 20: TukeyHSD en R	18
Figure 21: Différences dans les niveaux moyens de traitement.....	19
Figure 22: Emmeans en R.....	19
Figure 23: les antécédents familiaux en fonction de sexe.....	20
Figure 24: le sexe en fonction des antécédents familiaux	20

Introduction générale

Des millions de personnes dans le monde souffrent d'obésité, un grave problème de santé publique qui augmente la probabilité de maladies chroniques telles que certains cancers, les accidents vasculaires cérébraux, l'hypertension artérielle, le diabète et les maladies cardiovasculaires. Bien que la génétique contribue de manière significative à l'apparition de l'obésité, il existe une pléthore de mythes et d'idées fausses entourant cette question.

Des facteurs génétiques, environnementaux et comportementaux contribuent tous à la condition complexe de l'obésité. Contrairement à la croyance populaire, elle n'est pas uniquement causée par un manque de volonté, d'indiscipline ou de motivation pour manger sainement et faire de l'exercice.

Génétiquement parlant, comment se transmet l'obésité et pourquoi son expression héréditaire est-elle influencée par des facteurs environnementaux ?



I - Problématique

La mesure dans laquelle la génétique influence l'obésité est une question urgente, qui peut être abordée en analysant les variances à l'aide de l'ANOVA. L'accent devrait être mis sur l'étude des variations génétiques au sein de différentes populations présentant divers degrés d'obésité et sur la détermination de l'incidence de ces variations sur les différences observées. La contribution de divers groupes génétiques à l'ensemble de l'obésité peut être analysée et quantifiée à l'aide d'une méthode statistique telle que l'ANOVA. De plus, cette étude évaluera comment les gènes et l'environnement interagissent dans le développement de l'obésité, ce qui pourrait avoir une signification importante pour sa gestion et son évitement.



Figure 1 : Les causes de l'obésité

II - Etude de l'existant :

Les taux d'obésité dans le monde continuent de grimper.

Plus de 640 millions de personnes sont maintenant grosses, selon les données de BMI.

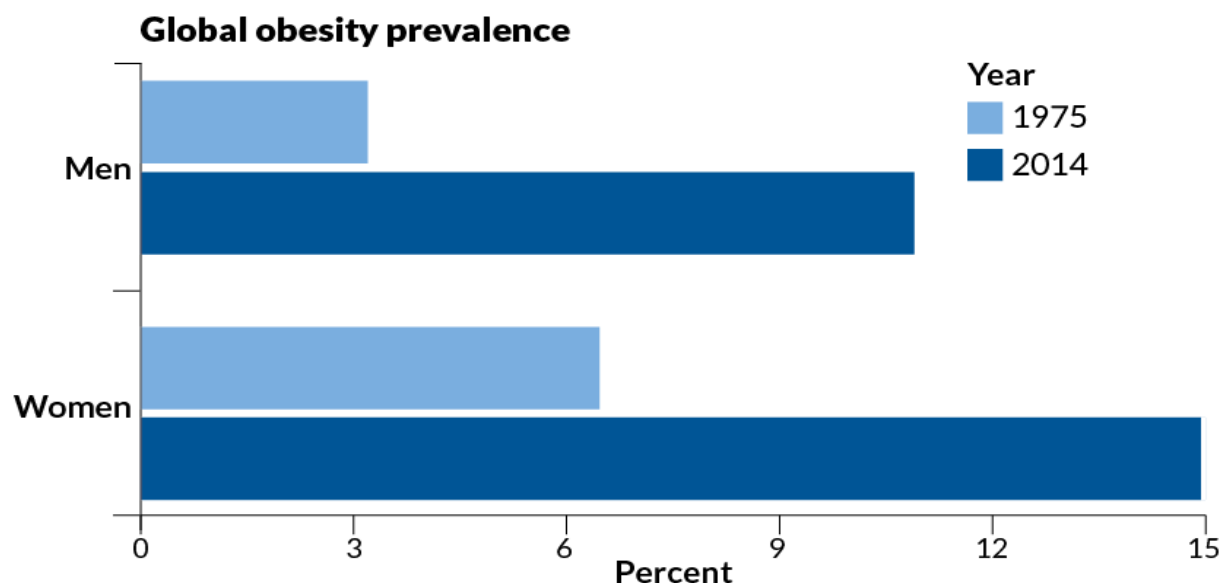


Figure 2: obésité dans le monde

1 - Jeu de données globale :

	Gender	Age	Height	Weight	family_history_with_overweight	FAVC	FCVC	NCP	CAEC	SMOKE
0	Female	21.000000	1.620000	64.000000	yes	no	2.0	3.0	Sometimes	no
1	Female	21.000000	1.520000	56.000000	yes	no	3.0	3.0	Sometimes	yes
2	Male	23.000000	1.800000	77.000000	yes	no	2.0	3.0	Sometimes	no
3	Male	27.000000	1.800000	87.000000	no	no	3.0	3.0	Sometimes	no
4	Male	22.000000	1.780000	89.800000	no	no	2.0	1.0	Sometimes	no
...
2106	Female	20.976842	1.710730	131.408528	yes	yes	3.0	3.0	Sometimes	no
2107	Female	21.982942	1.748584	133.742943	yes	yes	3.0	3.0	Sometimes	no
2108	Female	22.524036	1.752206	133.689352	yes	yes	3.0	3.0	Sometimes	no
2109	Female	24.361936	1.739450	133.346641	yes	yes	3.0	3.0	Sometimes	no
2110	Female	23.664709	1.738836	133.472641	yes	yes	3.0	3.0	Sometimes	no

2111 rows × 11 columns

Figure 3: global Dataset

2 - Preprocessing :

Pour obtenir l'ensemble de données final, nous avons effectués un nettoyage mineur et vérifier s'il y a des valeurs manquantes sur le jeu de données globale en python avec pandas.

```
import pandas as pd

data=pd.read_csv('ObesityDataSet.csv',usecols=['Gender','Weight','family_history_with_overweight'])

data['Gender'].replace({'Female': 'F', 'Male': 'M'}, inplace=True)

data['Weight']=data['Weight'].astype(int)

miss_val=data.isnull()

miss_val.sum()

Gender                0
Weight                0
family_history_with_overweight  0
dtype: int64
```

Figure 4: Data Preprocessing

i - Pandas :



Figure 5: logo de pandas

Les programmeurs Python ne jurent que par Pandas, une bibliothèque open source qui facilite la manipulation et l'analyse des données. Grâce à ses structures de données dynamiques, Pandas facilite la gestion des données tabulaires et des données de séries chronologiques à partir de fichiers CSV ou Excel. Ce n'est pas tout, Pandas propose des outils de nettoyage et de transformation des données, associés à un arsenal d'analyses statistiques et de fonctionnalités de visualisation des données. Il n'est pas étonnant que cette bibliothèque continue d'être le premier choix des scientifiques des données, des analystes et des experts en apprentissage automatique.

ii - Jeu de données final :

L'obésité est une maladie complexe qui est influencée par de nombreux facteurs, notamment le sexe et les antécédents familiaux de surpoids.

	Gender	Weight	family_history_with_overweight
1	F	64	yes
2	F	56	yes
3	M	77	yes
4	M	87	no
5	M	89	no
6	M	53	no
7	F	55	yes
8	M	53	no
9	M	64	yes
10	M	68	yes
11	M	105	yes
12	F	80	yes
13	M	56	no
14	M	99	no
15	M	60	yes
16	F	66	yes
17	M	102	yes
18	F	78	no
19	F	82	yes
20	F	70	yes
21	M	80	yes
22	F	87	yes
23	F	60	yes
24	F	82	yes
25	M	68	yes

Figure 6 : Tableaux de données

iii - Hypothèse dans le test Two-Way ANOVA:

- H0 : Les moyennes sont égales pour les deux variables (c.-à-d. Variable factorielle)
- Ha : Les moyennes sont différentes pour les deux variables

3 - Régression Linéaire Simple :

i - Définition :

La régression linéaire est une technique d'analyse de données qui prédit la valeur de données inconnues en utilisant une autre valeur de données apparentée et connue.

ii - Objectif :

Nous avons effectué une régression linéaire simple pour estimer l'effet de chaque variable qualitative sur la variable dépendante.

Cela nous permettra de déterminer s'il y a une relation significative entre les variables qualitatives et la variable quantitative.

iii - Modèle de RLS :

Nous avons créé un modèle de régression linéaire simple en utilisant la fonction `lm()` de R. Et utilisé la variable de poids comme variable dépendante et les variables binaires encodées comme variables indépendantes.

```
# Importer les données
data <- read.csv("C:\\Users\\lenovo\\Desktop\\R project\\obesity.csv")

# Encodage des variables qualitatives en variables binaires
gender <- ifelse(data$gender == "F", 0, 1)
data$family_history_with_overweight <- ifelse(data$family_history_with_overweight == "no", 0, 1)

# Créer un modèle de régression linéaire simple
modele <- lm(data$weight ~ data$gender + data$family_history_with_overweight, data = data)
```

Figure 7: RLS Model en R

iv - Interprétation de la sortie :

Voir l'annexe. (RLS annexe)

- Les résultats de la régression linéaire montrent que les variables qualitatives ont un effet significatif sur le poids des individus.
- Le coefficient de détermination (R^2) est de 0.2587, ce qui indique que les variables qualitatives expliquent environ 26% de la variation du poids des individus.
- Le test F montre que le modèle dans son ensemble est significatif ($p\text{-value} < 2.2e-16$), ce qui indique que le modèle permet de prédire le poids des individus de manière significative.

4 - Homogénéité des variances :

L'hypothèse d'homogénéité des variances, également connue sous le nom d'hypothèse d'égalité des variances, est l'une des hypothèses de base de l'analyse de la variance (ANOVA). Il indique que les variances des différents groupes sont égales. Si l'hypothèse d'homogénéité des variances n'est pas respectée, cela peut affecter la validité des tests statistiques basés sur l'ANOVA.

i - Test de Levene :

Le test de Levene est un test statistique utilisé pour vérifier l'hypothèse d'homogénéité des variances dans un ensemble de données. Il teste si les variances des différents groupes sont égales. Ce test est souvent utilisé en conjonction avec l'ANOVA pour s'assurer que les hypothèses de l'ANOVA sont respectées.

```
from scipy.stats import levene

grouped_data = data.groupby(['Gender', 'family_history_with_overweight'])
# Appliquer le test de Levene pour tester l'homogénéité de variance
test_result = levene(*[group['Weight'] for name, group in grouped_data])

if test_result.pvalue > 0.05:
    print("Les variances sont homogènes pour toutes les combinaisons de variables.")
else:
    print("Il y a une différence significative dans les variances pour au moins une combinaison de variables.")

Il y a une différence significative dans les variances pour au moins une combinaison de variables.
```

Figure 8: Levene Test

ii - Test de Bartlett :

Le test de Bartlett est un test statistique qui évalue l'hypothèse d'homogénéité de la variance dans un ensemble de données, tout comme le test de Levene. Cependant, contrairement au test de Levene, le test de Bartlett est basé sur la distribution normale des données. Par conséquent, le test de Bartlett est plus adapté aux échantillons de données distribués normalement.

```

from scipy.stats import bartlett

groupes = data.groupby(["Gender", "family_history_with_overweight"])["Weight"].apply(list)
stat, p = bartlett(*groupes)

if p < 0.05:
    print("Les variances des groupes sont significativement différentes")
else:
    print("Les variances des groupes ne sont pas significativement différentes")

Les variances des groupes sont significativement différentes

```

Figure 9: Bartlett Test

Les deux tests indiquent que les variances des groupes sont significativement différentes.

```

variances = data.groupby(["Gender", "family_history_with_overweight"])["Weight"].var()
variances

```

Gender	family_history_with_overweight	
F	no	91.416555
	yes	748.835941
M	no	152.346319
	yes	419.690214

Figure 10: les variances des groupes

5 - Test de normalité :

L'hypothèse de normalité est bien validée par le test de Shapiro-Wilk.

```

> shapiro.test(resid(anova))

      shapiro-wilk normality test

data:  resid(anova)
W = 0.98839, p-value = 4.894e-12

```

Figure 11: test de normalité

Le test de normalité de Shapiro-Wilk a été utilisé pour évaluer si les résidus ANOVA étaient normalement distribués. Dans ce cas, le test montre que les résidus ne sont pas significativement différents d'une distribution normale car la valeur de p est inférieure au seuil de 0,05. Cela montre que l'ANOVA est robuste et répond à l'hypothèse de base de la normalité des résidus.

6 - Ajustement du modèle ANOVA :

L'ajustement du modèle ANOVA à l'aide d'une transformation des variables telles que le logarithme ou la racine carrée permettent d'améliorer l'homogénéité de la variance et la normalité des résidus du modèle. Cela permet également de réduire l'impact des valeurs aberrantes et des données extrêmes.

Remarque importante :

L'ANOVA est robuste aux violations de l'hypothèse d'homogénéité des variances, mais cela affecte la puissance statistique du test.

Il est recommandé d'utiliser une ANOVA ajustée, telle qu'une ANOVA à variance inégale ou une ANOVA de Welch.



7 - Synthèse :

Cette première partie dédiée pour le Data Preprocessing, nous avons effectué tous les tests nécessaires pour appliquer le modèle d'ANOVA. Ainsi de déterminer s'il y a une relation significative entre les variables qualitatives et la variable quantitative.

III -Analyse of variance (ANOVA) :

1 - Le modèle d'ANOVA :

Comme déjà mentionné ci-dessus, nous nous allons examiner la relation entre :

- ✚ Le sexe (Gender), qui est notre première variable explicative -> appelons-le "X".
- ✚ Les antécédents familiaux avec le surpoids (Family_history_with_overweight) appelons-le "Z".
- ✚ Le poids (Weight) est notre variable de réponse -> appelons-le "Y"
Mathématiquement, la relation peut être écrite comme ceci : $Y \sim X * Z$

ANOVA va comparer les moyennes du poids parmi les 4 groupes et vérifier si les différences sont statistiquement significatives.

Un modèle ANOVA pouvant être appliqué à notre ensemble de données peut être décrit comme suit :

$$Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij}$$

Telle que :

- ✚ Y_{ij} est le poids sexuel du i-ème individu dans le j-ème groupe (homme $j = 1$, femme $j = 2$)
- ✚ μ est la moyenne globale de tous les poids
- ✚ α_i est l'effet du i-ème antécédent familial de surpoids (par rapport à la moyenne globale)
- ✚ β_j est l'effet du j-ème groupe de sexe (par rapport à la moyenne de la population)
- ✚ ϵ_{ij} est le résidu associé à la i-ème observation chez le j-ème sexe et les antécédents familiaux.

i - Hypothèse nulle et alternative :

Hypothèse nulle : les quatre groupes signifient qu'ils sont égaux

✚ Il n'y a pas de relation entre le sexe, les antécédents familiaux et le poids, que nous pouvons écrire comme suit :

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4$$

Hypothèse alternative : les moyens des quatre groupes ne sont pas tous égaux

✚ Il existe une relation entre les groupes et le poids :

H1 : tous les μ ne sont pas égaux

Les 4 groupes sont :

```
grouped_data <- split(data, list(Gender,family_history_with_overweight))
for (i in 1:length(grouped_data)) {
  print(names(grouped_data)[i])
}
```

Figure 12: Renvoie les groupes en R

La sortie est :

```
('F', 'no')
('F', 'yes')
('M', 'no')
('M', 'yes')
```

Figure 13: les 4 groupes

ii - Les moyennes entre les groupes : (Voir l'annexe)

Il est important de noter que ces résultats sont basés sur une analyse descriptive des données et ne tiennent pas compte de la taille de l'échantillon ou de la variabilité des données.

```
moyennes<-aggregate(weight~Gender+family_history_with_overweight,data=data,FUN=mean)
moyennes
```

Figure 14: les moyennes entre les groupes

Il semble qu'il y ait une différence significative de poids entre les groupes de sexe et d'antécédents familiaux de surpoids.

Le calcul des moyennes intergroupes est une étape importante dans l'analyse exploratoire des données, car il vous permet de visualiser et de comparer les tendances générales des données entre différents groupes. La moyenne inter-groupes fournit des informations sur la distribution des données et aide à déterminer s'il existe des différences significatives entre les groupes.

Remarque :

Pour tirer des conclusions plus robustes, une analyse statistique plus formelle, telle qu'un test d'hypothèse, pourrait être nécessaire.

iii - Boxplot des moyennes de poids par variables qualitatives :

Les moyennes sont des points rouges.

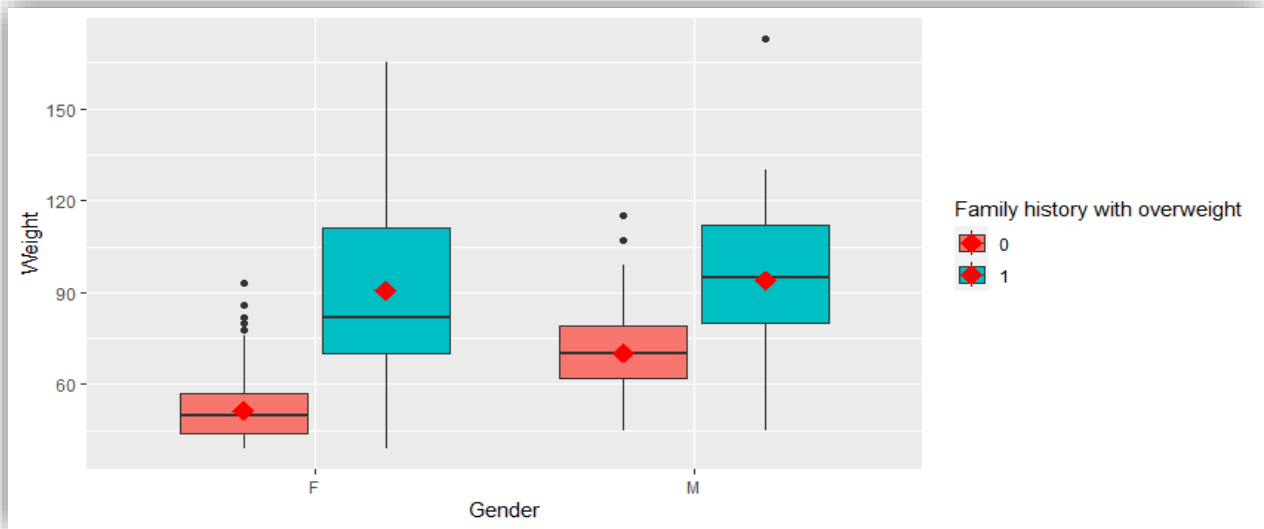


Figure 15: les moyennes de poids

La boîte à moustaches montre que les moyens sont différents (certains moins, d'autres plus). Mais cela montre également que chaque groupe présente une quantité différente de variation du poids, de sorte qu'il y a beaucoup de chevauchement des valeurs entre certains groupes. Par conséquent, les différences de moyennes auraient pu provenir du hasard et il ne faut pas rejeter le cas d'hypothèse nulle. C'est là qu'ANOVA vient nous aider.

Dans le premier cas, nous aurons des preuves solides contre l'hypothèse nulle (les moyennes sont toutes égales), tandis que dans le second cas, nous aurions peu de preuves contre l'hypothèse nulle.

Question ?

Les variations entre les moyennes des groupes sont-elles dues à de vraies différences sur les moyennes des populations ou simplement à la variabilité d'échantillonnage ?

**2 - Le test d'ANOVA :****i - ANOVA en R :**

Après cet excursus théorique, il est temps d'effectuer une ANOVA sur nos données et d'essayer d'interpréter les résultats.

```
anova=aov(weight~Gender*family_history_with_overweight)
summary(anova)
```

Figure 16: ANOVA en R

Le résultat de test :

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
Gender	1	38445	38445	77.51	< 2e-16	***
family_history_with_overweight	1	332899	332899	671.17	< 2e-16	***
Gender:family_history_with_overweight	1	19258	19258	38.83	5.58e-10	***
Residuals	2107	1045067	496			

 Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Figure 17: table d'ANOVA

a - Interprétations :

La valeur p indique la probabilité que l'effet observé soit dû au hasard si l'hypothèse nulle (les moyennes sont égales entre les groupes) est vraie. Si la valeur p est inférieure au niveau de signification (par exemple, 0,05), l'hypothèse nulle est rejetée et on conclut que les moyennes sont significativement différentes entre les groupes.

Dans ce cas, les résultats montrent que le sexe, l'histoire familiale de surpoids et l'interaction entre les deux ont tous un effet significatif sur le poids moyen ($p < 0,001$ dans tous les cas). Cela suggère que les moyennes de poids diffèrent significativement entre les groupes définis par ces variables.

ii - La violence de l'homogénéité des variances :

La condition de l'égalité des variances est violée, cela peut affecter la fiabilité du test ANOVA standard. Ce test suppose que les écarts entre tous les groupes sont égaux. Dans de tels cas, il peut être préférable d'utiliser une version d'ANOVA qui est robuste comme Welch ANOVA.

iii - Welch ANOVA :

L'analyse de la variance (ANOVA) est une méthode statistique utilisée pour comparer les moyennes de plusieurs groupes. Dans certains cas, les groupes peuvent avoir des variances différentes, ce qui peut réduire la fiabilité des résultats de l'ANOVA. Dans ce cas, la correction de Welch peut être utilisée pour ajuster les résultats de l'ANOVA en tenant compte des différences de variance.

a - *Welch ANOVA en R :*

```
model <- aov(weight~Gender*family_history_with_overweight, data = mydata)
results <- Anova(model, type = "III", white.adjust = TRUE)
```

Figure 18: Le modèle de Welch ANOVA

Le résultat de test :

```
Analysis of Deviance Table (Type III tests)

Response: weight
              Df      F    Pr(>F)
(Intercept)    1 6645.981 < 2.2e-16 ***
Gender          1  259.908 < 2.2e-16 ***
family_history_with_overweight  1 1169.068 < 2.2e-16 ***
Gender:family_history_with_overweight  1  90.728 < 2.2e-16 ***
Residuals      2107
---
signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 19: La sortie en R

b - *Interprétation :*

Les résultats montrent que toutes les variables indépendantes (le sexe, les antécédents familiaux de surpoids et l'interaction entre les deux) sont significativement associées au poids, avec des valeurs de p pour chaque variable inférieure à 0,001. Un score F élevé pour chaque variable indique qu'elles ont une grande influence sur la variation de poids. Enfin, la ligne

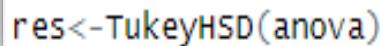
Résidus représente la somme résiduelle des carrés, représentant la variation non expliquée par le modèle.

3 - Tests de comparaison :

i - TukeyHSD test :

Le test de Tukey, également connu sous le nom de test HSD (Honest Significant Difference), est une méthode statistique utilisée pour comparer toutes les paires de moyennes dans un ensemble de données. Il est souvent utilisé après ANOVA pour déterminer quelles paires de moyennes sont significativement différentes les unes les autres.

ii - TukeyHSD en R : (Voir l'annexe)



```
res<-TukeyHSD(anova)
```

Figure 20: TukeyHSD en R

Le test de Tukey est réalisé sur un modèle linéaire de la forme Poids ~ Sexe * historique_familial_avec_surpoids avec un niveau de confiance de 95% pour la famille testée.

Les résultats ont montré qu'il y avait une différence significative dans le poids corporel moyen des hommes et des femmes, avec une différence moyenne de 8,54 kg.

De plus, la différence de poids moyenne entre ceux qui ont des antécédents familiaux de surpoids et ceux sans antécédents familiaux de surpoids est également significative, avec une différence moyenne de 32,35 kg.

Enfin, de multiples comparaisons entre groupes ont révélé des différences moyennes de poids entre les groupes M:non-F:non, F:oui-F:non, M:oui-F:non, F:oui-M:non et M:oui- M :no était également significatif, avec des différences moyennes allant de 19,06 kg à 42,53 kg. De plus, la différence moyenne de poids corporel entre les groupes M:oui-F:oui était également significative, avec une différence moyenne de 3,13 kg.

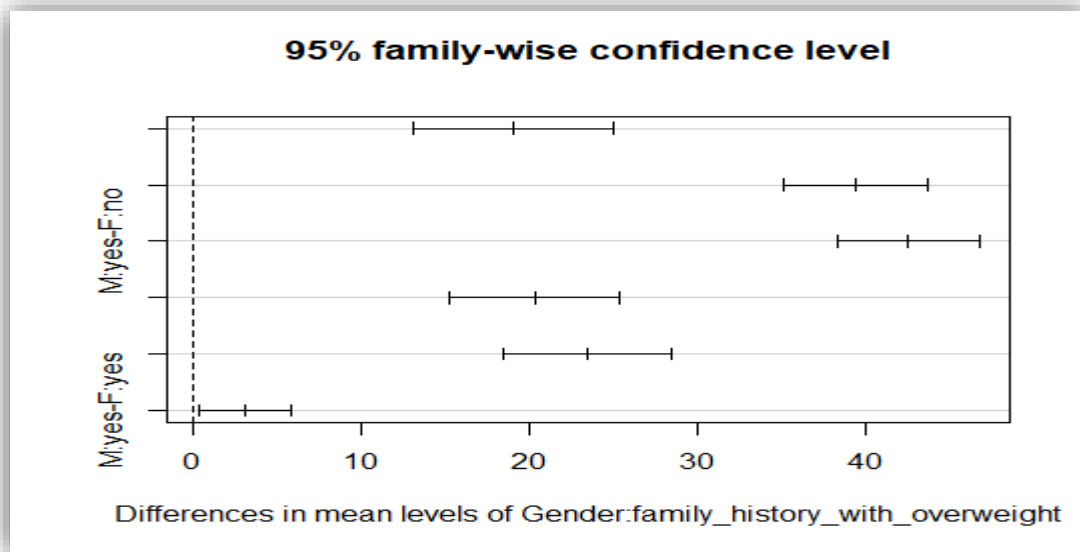


Figure 21: Différences dans les niveaux moyens de traitement

iii - Estimated marginal means (emmeans) : (Voir annexe)

a - Le test en R :

```
library(emmeans)
res.g<-emmeans(anova,~Gender | family_history_with_overweight)
res.f<-emmeans(anova,~family_history_with_overweight | Gender)
```

Figure 22: Emmeans en R

Les résultats de l'analyse des moyennes marginales indiquent que parmi les personnes ayant des antécédents familiaux de surpoids, les femmes ont un poids corporel estimé moyen significativement plus élevé que les hommes, avec une moyenne de 90,7 kg et 93,8 kg, respectivement. Les femmes sans antécédents familiaux d'obésité avaient des poids moyens estimés significativement inférieurs à ceux des hommes, 51,3 kg et 70,3 kg, respectivement. L'analyse suggère donc que des antécédents familiaux d'obésité ont un impact sur le poids corporel moyen et que le facteur fonctionne différemment selon sur le genre.

b - Plots :

Les graphiques (plots) peuvent être très utiles pour visualiser les résultats des tests EMMeans. Ils permettent de représenter graphiquement les différences entre les moyennes estimées pour chaque niveau des variables d'intérêt, ainsi que les intervalles de confiance associés.

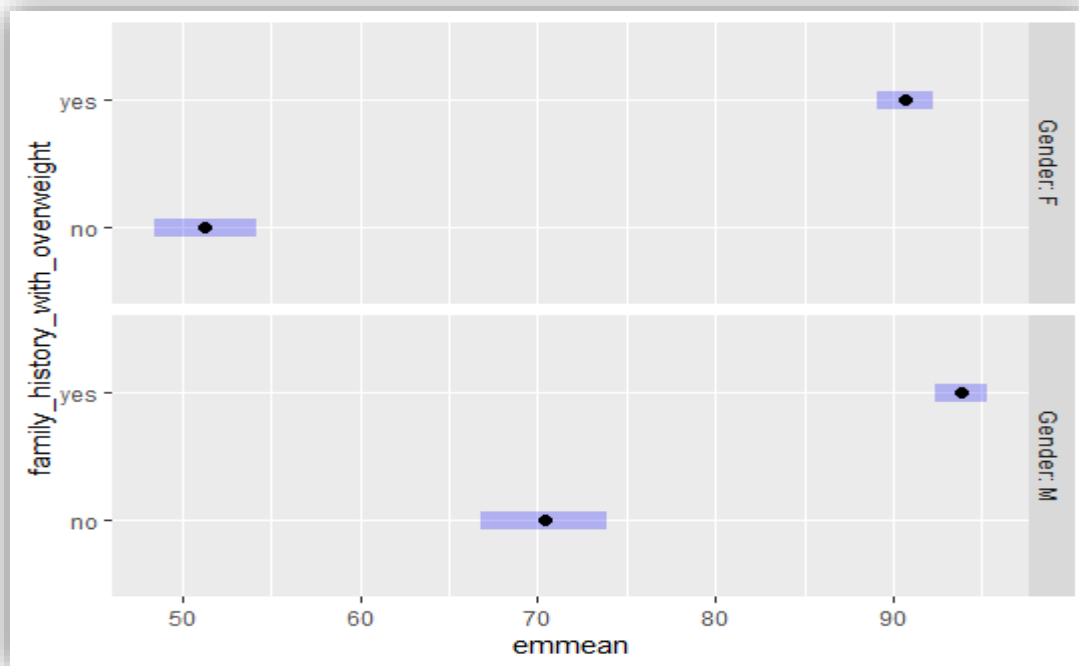


Figure 23: les antécédents familiaux en fonction de sexe

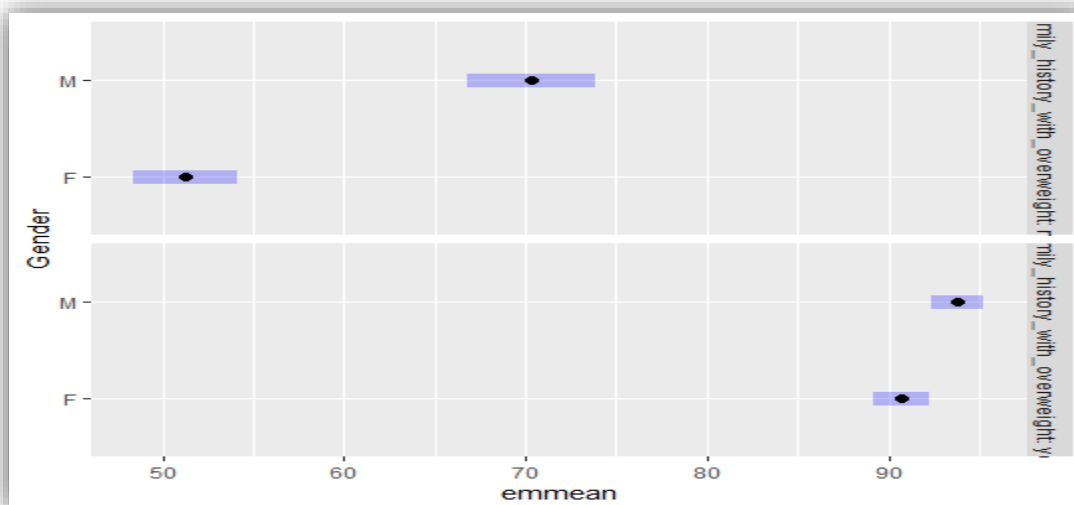


Figure 24: le sexe en fonction des antécédents familiaux

En interprétant les résultats, on peut conclure sur les différences significatives ou non entre les moyennes marginales pour chaque niveau de chaque variable, ainsi que sur les effets des interactions entre les variables. Les résultats peuvent donc aider à mieux comprendre les relations entre les variables dans un modèle.

Conclusion :

L'ANOVA (Analyse de la Variance) est une méthode statistique utilisée pour comparer les moyennes de plusieurs groupes et déterminer si les différences observées sont statistiquement significatives. L'ANOVA est souvent utilisé pour comparer les moyennes de groupes qui ont une variable indépendante en commun, appelée variable de groupe.

Dans ce travail dédié, l'ANOVA a été utilisé pour comparer les moyennes de poids de différents groupes définis par le sexe et l'historique familial de surpoids. Les résultats ont montré que le sexe, l'historique familial de surpoids et l'interaction entre les deux avaient tous un effet significatif sur le poids moyen. Les valeurs de p pour chaque variable étaient inférieures à 0,001, ce qui indique que les moyennes de poids diffèrent significativement entre les groupes définis par ces variables.

En guise de conclusion, Il est important de prendre en compte ces facteurs lorsqu'on étudie l'obésité, car cela peut aider à mieux comprendre les différences de poids corporel entre les individus et à élaborer des stratégies pour prévenir ou traiter l'obésité.

Références :

-  <https://www.kaggle.com/datasets/ankurbajaj9/obesity-levels>
-  <https://www.wikipedia.org/>
-  <https://www.optimyself.com/obesite-genetique/>
-  <https://delladata.fr/anova-a-2-facteurs-principe/>
-  <https://towardsdatascience.com/two-way-anova-test-with-python-a112e2396d78>
-  <https://pandas.pydata.org/docs/>
-  <https://rdr.io/r/>
-  <https://www.sciencenews.org/article/global-obesity-rates-continue-climb>