

26/02/2025

ÉTUDE DE CAS

Effet de l'âge et d'autres
paramètres sur l'anticipation

Auteurs

Daouda DIOP & Brun Bahoun Houtoukpe

Introduction

L'initiation de la marche est une phase clé du mouvement humain, caractérisée par un ensemble de mécanismes d'anticipation et de gestion du déséquilibre avant le déplacement effectif. Cette étude vise à analyser l'effet de l'âge et d'autres paramètres sur cette phase, en se concentrant sur les déplacements du centre de pression (CoP) et les durées des différentes étapes de l'initiation de la marche chez des enfants âgés de 6 à 12 ans.

Dans cette étude, nous avons utilisé un ensemble de données, permettant de mesurer les paramètres spatio-temporels de l'initiation de la marche. Afin d'évaluer l'impact de l'âge sur ces paramètres, nous avons appliqué différentes méthodes statistiques et exploratoires, en nous appuyant sur les langages Python et R. Ces outils ont permis de réaliser des analyses approfondies à travers des tests statistiques, des visualisations graphiques et des modèles de régression.

Données descriptives

L'essai le plus représentatif

L'objectif de cette question est de déterminer quelle valeur statistique est la plus représentative pour caractériser la performance d'un enfant à partir de ses 5 essais.

Lors de l'initiation de la marche, chaque enfant a réalisé **cinq essais** afin d'évaluer ses performances. L'objectif de cette première analyse est de sélectionner une mesure représentative parmi ces cinq essais.

Dans une étude de données, la gestion des valeurs manquantes est une étape essentielle pour garantir la fiabilité des analyses. Le choix entre la moyenne et la médiane présente des avantages et des inconvénients selon la nature des données et la présence des valeurs extrêmes.

La moyenne a l'avantage de prendre en compte l'ensemble des valeurs, ce qui permet d'avoir une vision globale des données. Elle est particulièrement utile lorsque la distribution est équilibrée et qu'il n'y a pas de valeurs aberrantes. Cependant, elle est sensible aux valeurs extrêmes, ce qui peut fausser les résultats si certaines mesures sont anormalement basses.

La médiane, quant à elle, est moins influencée par les valeurs aberrantes. Elle est donc plus robuste dans le cas d'une distribution asymétrique ou contenant des valeurs extrêmes. Toutefois, elle ne tient pas compte de toutes les valeurs de l'échantillon, ce qui peut limiter l'analyse de la variabilité des données.

Dans le cadre de cette étude, la médiane est retenue comme mesure représentative des performances de chaque enfant. Cette décision est justifiée par sa robustesse face aux valeurs extrêmes, ce qui permet d'assurer une meilleure fiabilité des résultats et d'éviter l'influence de potentielles erreurs de mesure.

Détection des valeurs manquantes

Dans le cadre de l'étude de l'initiation à la marche, il est essentiel de s'assurer de la qualité des données recueillies. Une mauvaise qualité des données peut fausser les conclusions et affecter la fiabilité des résultats. Ainsi, nous avons procédé à une analyse rigoureuse des données pour identifier et corriger les valeurs manquantes et les valeurs aberrantes.

L'analyse exploratoire des données avec R et Python a révélé la présence de **valeurs manquantes** dans plusieurs variables, notamment celles liées aux caractéristiques comportementales et aux indices d'évaluation cognitive. Le tableau ci-dessous résume les variables concernées :

Variable	Nombre de valeurs manquantes
Fctelle	1
MANUALITE	1
QTAC	1
TAC?	1
Inattention	1
Hyperactivité	1
Pb apprentissage	1
Pb fct exécutif	1
Agressivité	1
Pb sociaux	1
AWR	27
COG	27
COM	27
MOT	27
RRB	27
SCI	27
Module ADOS	30

Stratégie de traitement :

Pour les **variables avec peu de valeurs manquantes**, nous avons procédé à une **imputation par la médiane**, une approche qui préserve la distribution initiale des données.

Pour les **variables fortement incomplètes**, nous avons choisi de les exclure de l'analyse afin d'éviter tout biais potentiel dans l'interprétation des résultats.

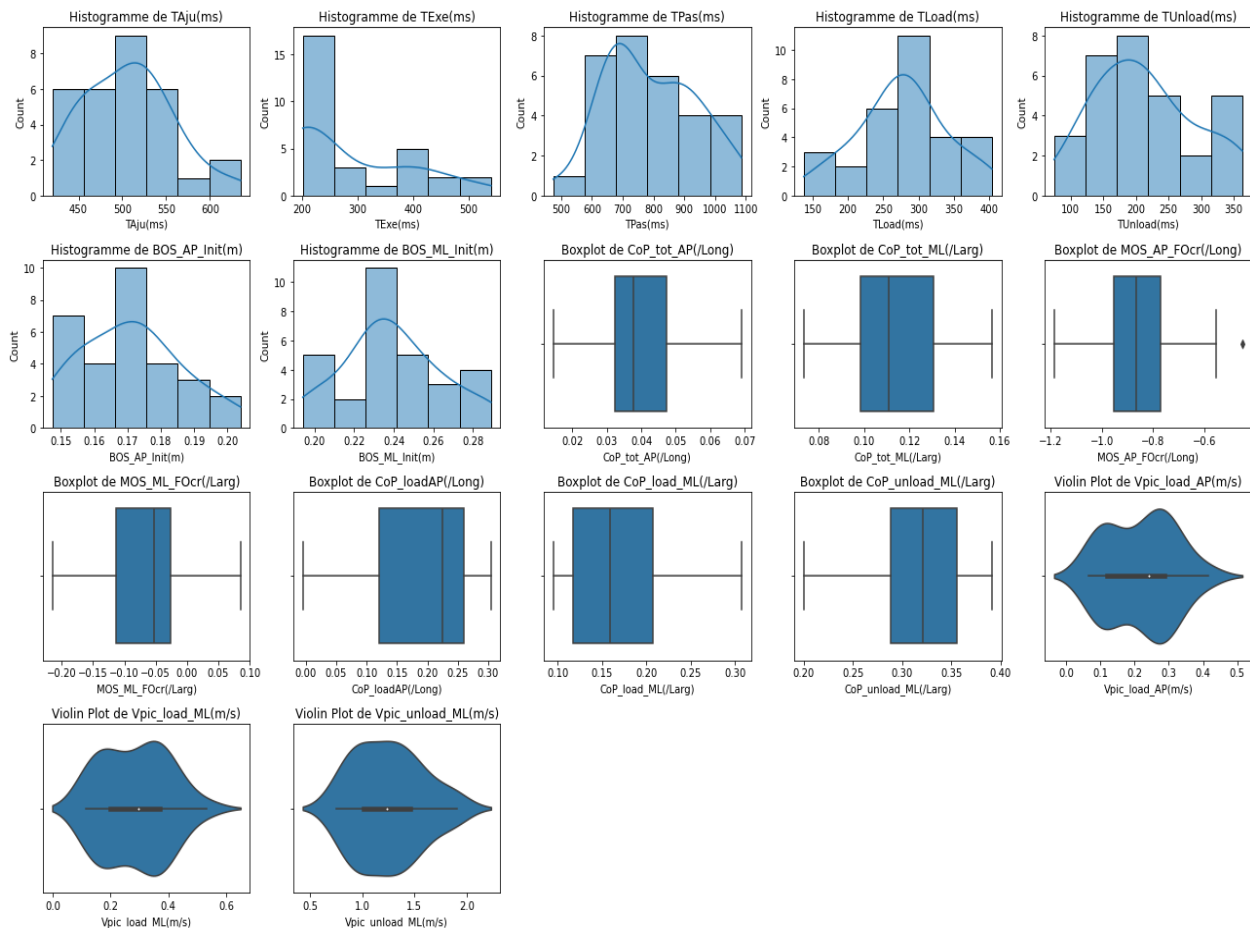
Identification des valeurs aberrantes

Les valeurs aberrantes ont été identifiées à l'aide de la méthode des **interquartiles (IQR)**, où toute valeur située en dehors de l'intervalle **$[Q1 - 1.5 \cdot IQR, Q3 + 1.5 \cdot IQR]$** est considérée comme aberrante. Nous avons donc décidé de remplacer les valeurs aberrantes par la médiane des autres valeurs car nous trouvons ce choix judicieux et aussi ce choix aura juste un impact minimal sur notre dataset. Le tableau suivant présente quelques valeurs considérées comme aberrantes :

Variable	Valeurs Aberrantes
TAju(ms)	273, 827, 301
TLoad(ms)	475, 464
TUnload(ms)	-168
CoP_tot_AP(Long)	0.0731
CoP_unload_ML(Larg)	0.1616
MOS_AP_FOcr(Long)	-0.315, -1.353, -0.329
MOS_ML_FOcr(Larg)	0.1537
Weight	49.8
Main	31

Visualisation des valeurs aberrantes avec des graphiques

Nous avons également visualisé les différentes variables à l'aide de boxplots pour identifier les valeurs aberrantes. Ces graphiques permettent de repérer rapidement les observations qui s'écartent significativement des autres valeurs, en apparaissant sous forme de points isolés au-delà des moustaches des boîtes.



Nous avons utilisé plusieurs types de graphiques pour mieux comprendre la répartition des variables :

- **Histogrammes** : sont parfaits pour visualiser la **répartition** ou **distribution** des données, c'est-à-dire pour voir où se trouvent la majorité des valeurs, si elles sont symétriques, biaisées, ou s'il y a des valeurs extrêmes.
- **Boxplots** : Permettent d'identifier visuellement les valeurs aberrantes et d'analyser la dispersion des données.
- **Violin plots** : Fournissent une alternative aux boxplots en visualisant la densité des valeurs.

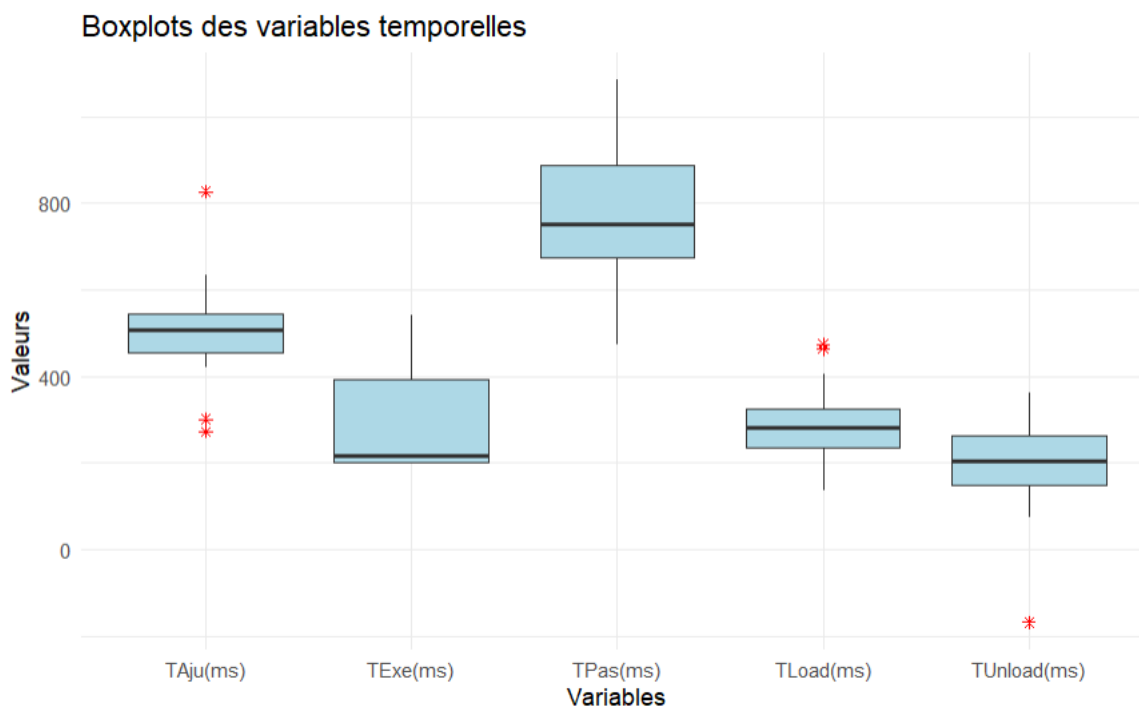
Pour les variables exprimées en ms (millisecondes) , L'histogramme est idéal pour observer comment les ajustements posturaux anticipés sont répartis (répartition des durées). Un histogramme avec une **courbe de densité** permettra de mieux voir la forme de la distribution.

Pour ceux exprimés en m (mètres), L'histogramme est aussi adapté pour voir comment les déplacements sont répartis et si certaines valeurs sont plus fréquentes que d'autres. L'ajout de la **courbe de densité** permettra d'analyser cette répartition de manière plus fluide.

Les variables exprimées en /long sont des **variables continue**, mais souvent les bases de support peuvent avoir une **répartition concentrée** ou des valeurs limites bien définies. Le boxplot met en évidence les **quartiles**, la **médiane** et les éventuelles **valeurs aberrantes**, ce qui est utile pour observer la distribution de ces variables sans qu'elles soient influencées par une échelle comme un histogramme.

Les variables exprimées en /larg (comme l'écart entre le CoP et le CoM) sont des **marges de stabilité** qui peuvent être sensible aux petites variations. Un **boxplot** permet de bien visualiser comment cette variable est distribuée et s'il y a des **valeurs aberrantes** dans la stabilité posturale (puisque la **marge de stabilité** peut fluctuer pour des raisons externes).

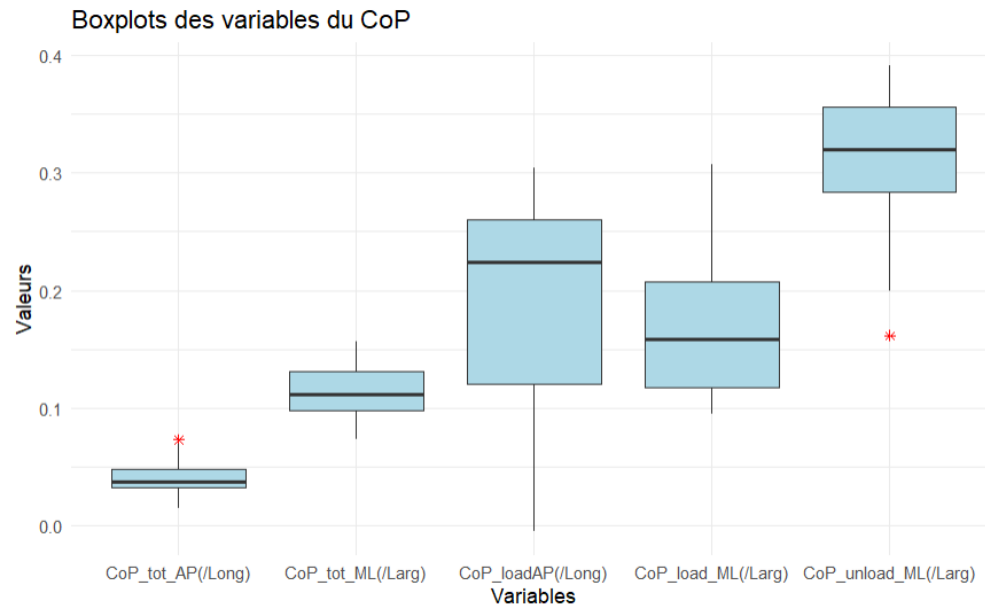
Comme les pics de vitesse sont souvent représentés par des distributions **asymétriques** ou avec des densités plus concentrées autour de certaines valeurs, le **violin plot** aide à comprendre où se situent ces concentrations et donne une meilleure vue d'ensemble de la forme de la distribution par rapport à un simple boxplot. Ce type de graphique permet aussi d'analyser la symétrie des données.



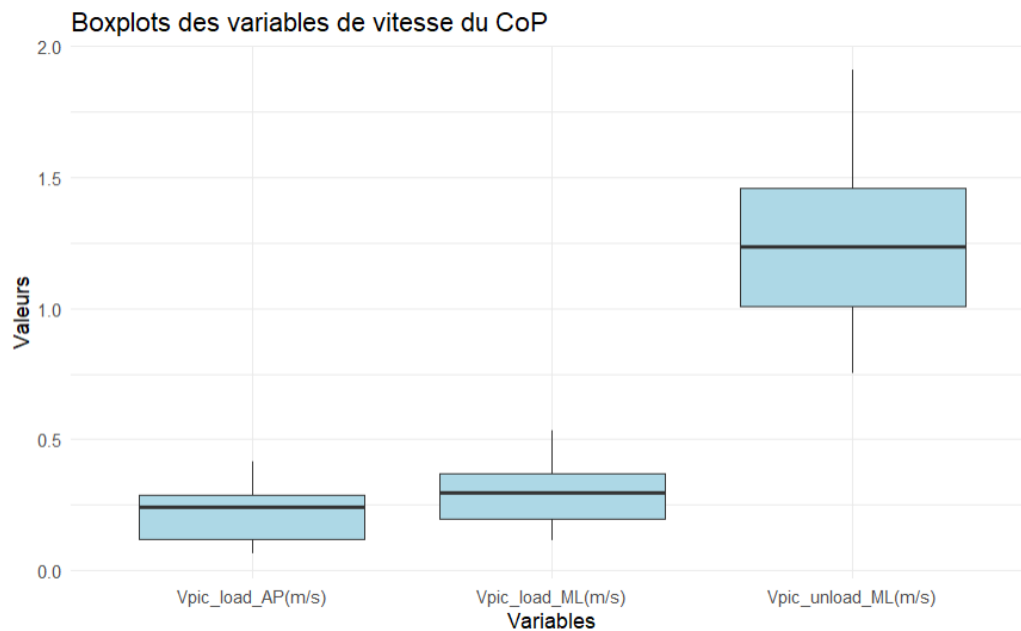
Nous observons plusieurs valeurs aberrantes, notamment pour TAju(ms) et TLoad(ms), où des points isolés se situent en dehors des limites habituelles des données.

Pour TUnload(ms), une valeur extrêmement basse semble être un outlier évident.

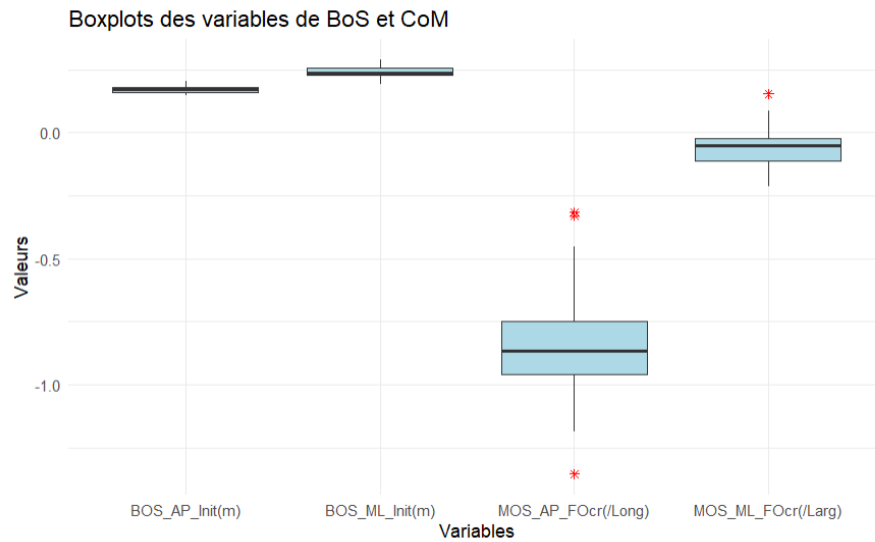
À l'inverse, TExe(ms) et TPas(ms) ne montrent pas d'anomalies marquées.



CoP_tot_AP(/Long) et CoP_unload_ML(/Larg) présentent quelques valeurs extrêmes, tandis que les autres variables sont globalement contenues dans les bornes normales.

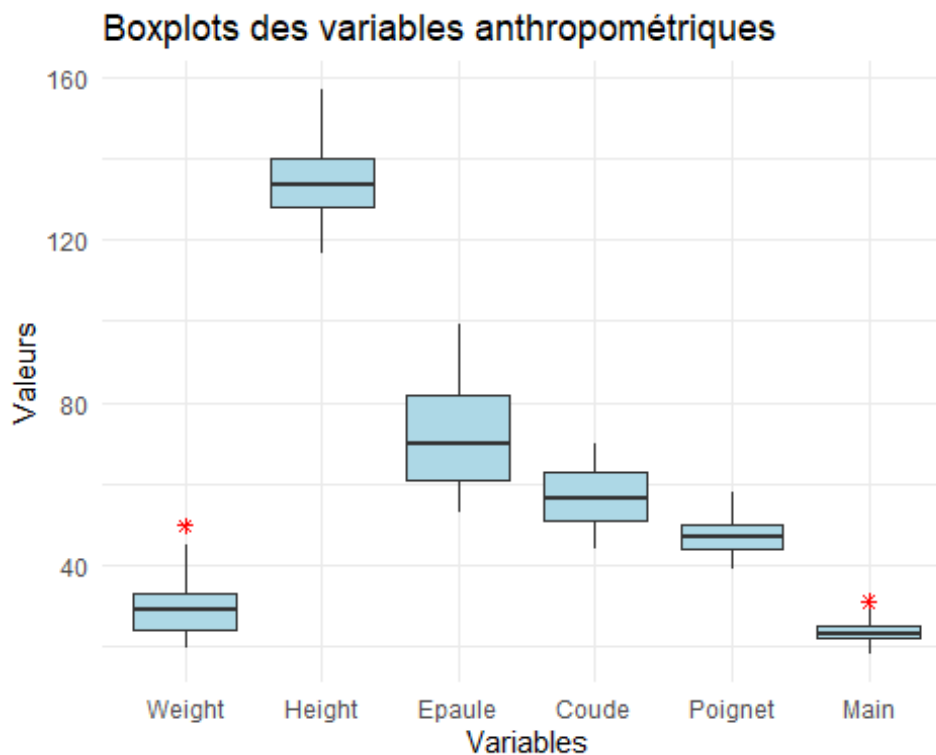


Aucune valeur aberrante significative n'a été identifiée sur ces variables, ce qui suggère une distribution relativement homogène.



Une anomalie est visible pour MOS_AP_FOcr(/Long), qui présente plusieurs valeurs éloignées des autres observations.

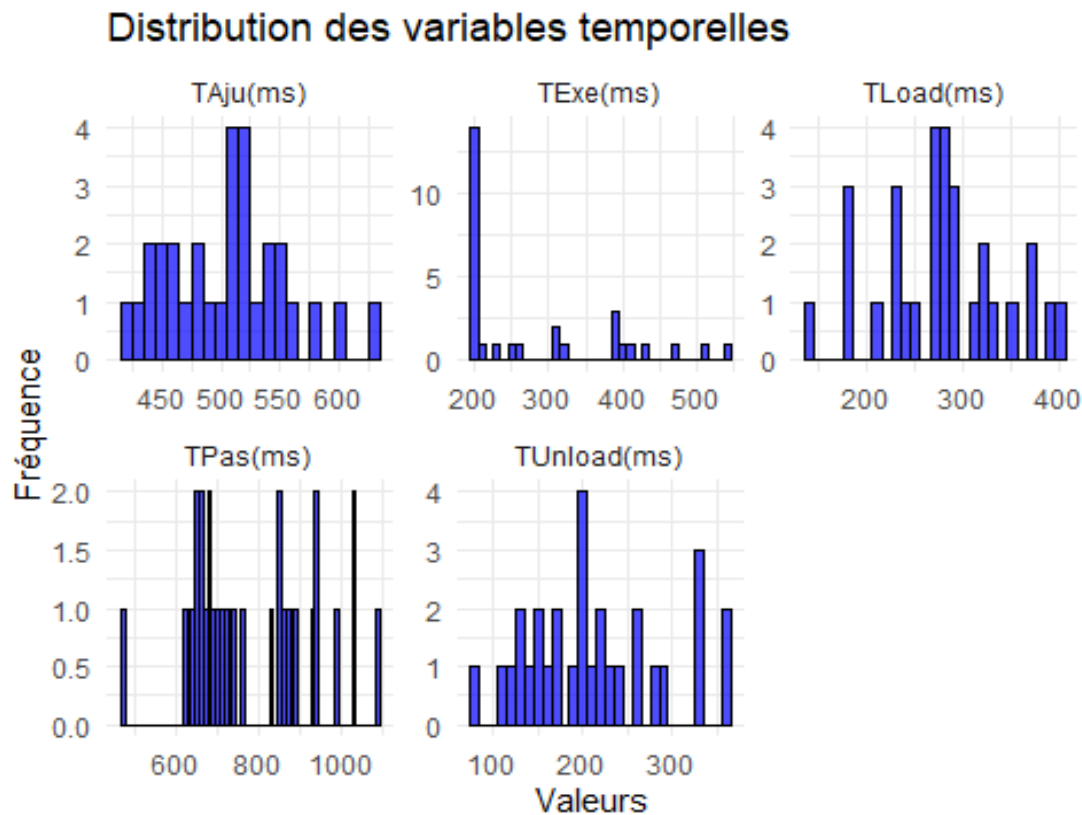
MOS_ML_FOcr(/Larg) montre également une dispersion importante avec une valeur qui pourrait être considérée comme un outlier.



Une valeur extrême est détectée pour Weight, suggérant un individu dont le poids est nettement plus élevé que le reste de l'échantillon

Q2.3

Histogrammes pour observer la distribution des variables temporelles

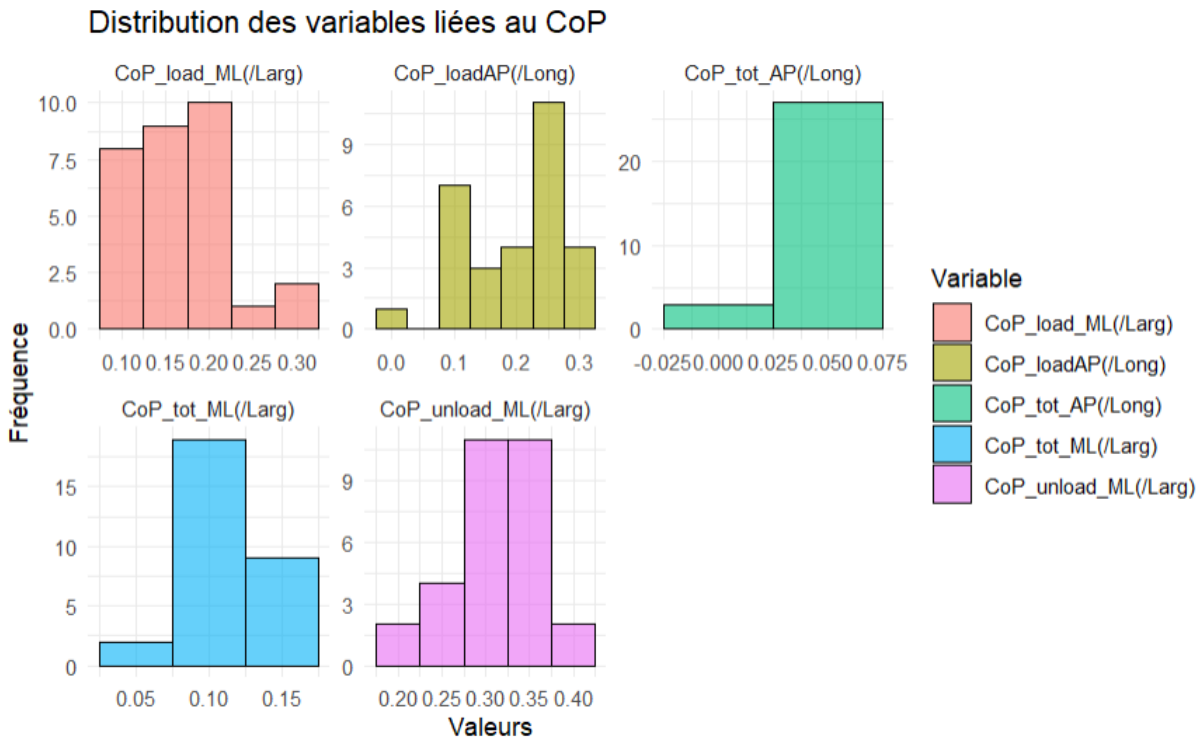


Les histogrammes présentés montrent la distribution des différentes variables temporelles liées à l'initiation de la marche, à savoir **TAju(ms)**, **TExe(ms)**, **TPas(ms)**, **TLoad(ms)**, et **TUnload(ms)**.

Explication et justification du choix des histogrammes

Les histogrammes sont un excellent outil pour visualiser la **répartition** des valeurs et détecter d'éventuelles **asymétries** ou **présences de valeurs extrêmes**. Ils permettent aussi d'observer si la distribution suit une **loi normale**, ce qui est essentiel pour choisir des tests statistiques appropriés par la suite.

Analyse des graphiques de distribution des variables liées au CoP



Ces histogrammes représentent la distribution des différentes variables associées au **Centre de Pression (CoP)**, qui est un indicateur clé du contrôle postural et de l'équilibre. L'objectif est de comprendre la répartition des déplacements du CoP et d'identifier d'éventuelles anomalies ou asymétries.

Effet de l'âge

Analyse des caractéristiques anthropométriques par groupe d'âge

Effectifs et caractéristiques anthropométriques par groupe d'âge

Groupe_Age	Effectif	Poids_Moyen	Taille_Moyenne	Poids_Médiane	Taille_Médiane
6-8 ans	13	24.67692	126.9615	24	128
9-11 ans	17	32.18529	140.6765	33	138

Ce tableau présente les statistiques descriptives du poids et de la taille des individus selon leur groupe d'âge. Il permet de mieux comprendre les différences morphologiques entre les enfants de **6-8 ans** et ceux de **9-11 ans**.

Poids :

- Le poids moyen des enfants **de 6-8 ans** est d'environ **24.68 kg**, tandis que celui des **9-11 ans** est **32.19 kg**.
- L'écart de **7.5 kg en moyenne** entre les deux groupes est significatif et cohérent avec le développement physiologique normal.
- La **médiane** suit la même tendance : **24 kg** pour les plus jeunes contre **33 kg** pour les plus âgés, ce qui confirme une répartition équilibrée sans valeurs extrêmes.

Taille :

- La taille moyenne des **6-8 ans** est **126.96 cm**, contre **140.68 cm** pour les **9-11 ans**.
- L'augmentation de la taille est également significative (**+13.7 cm en moyenne**).
- La **médiane** suit la même tendance (128 cm vs 138 cm), confirmant que la répartition des tailles est homogène

Analyse des performances selon l'âge

Groupe_Age	TAju(ms)	TExe(ms)	TPas(ms)	TLoad(ms)	TUnload(ms)
6-8 ans	504.2308	271.5385	775.7692	287.3846	201.9231
9-11 ans	508.9706	300.9412	802.7647	268.4118	224.7059

Ce tableau présente une comparaison des performances temporelles entre les groupes d'âge **6-8 ans** et **9-11 ans**, en fonction des différentes phases de mouvement mesurées en millisecondes (**ms**).

Il **synthétise** efficacement l'évolution des temps de réponse et d'exécution en fonction de l'âge. Ce tableau met en évidence **les différences de performances** entre les deux groupes d'âge sur plusieurs paramètres clés. Il permet également de formuler des **hypothèses** sur l'évolution des capacités motrices et posturales des enfants.

Vérification des conditions pour le test statistique

L'objectif est de comparer deux groupes d'enfants âgés de **6-8 ans** et **9-11 ans** sur différentes variables. Pour effectuer cette comparaison, il est important de choisir un test statistique approprié.

Le choix du test dépend de deux conditions fondamentales :

1. **La normalité des données**, évaluée à l'aide du test de **Shapiro-Wilk**.
2. **L'homogénéité des variances**, vérifiée avec le test de **Levene**.

- Si les données suivent une distribution normale et que les variances sont homogènes, nous utilisons le **test t de Student** (pour variances égales ou inégales selon les résultats du test de Levene).
- Si les données ne suivent pas une distribution normale, nous optons pour le **test de Mann-Whitney** (test non paramétrique).

Nous avons appliqué le **test de Shapiro-Wilk** à chaque variable pour vérifier la normalité des données au sein de chaque groupe d'âge.

- Une p-value **supérieure à 0,05** indique que la variable suit une distribution normale.
- Une p-value **inférieure à 0,05** signifie que la normalité est rejetée, nécessitant un test non paramétrique.

Ensuite, nous avons appliqué le **test de Levene** pour vérifier l'homogénéité des variances.

- Une p-value **supérieure à 0,05** indique que les variances sont homogènes, ce qui permet d'utiliser le test t de Student avec variances égales.
- Une p-value **inférieure à 0,05** implique des variances hétérogènes, nécessitant une correction dans le test t.

Application des tests

Les résultats des tests de normalité et d'homogénéité des variances ont guidé notre choix de tests :

- Pour la majorité des variables, la normalité a été respectée, et nous avons appliqué le **test t de Student**.
- Certaines variables ne respectaient pas la normalité, et nous avons utilisé le **test de Mann-Whitney**.

Résultats des tests

Les résultats des tests appliqués aux variables d'anticipation sont résumés dans le tableau ci-dessous :

Résultats des tests statistiques et moyennes par groupe

Variable	Test_utilisé	p_value	Significatif	Groupe_Age	Moyenne
TAju(ms)	Test t de Student (var égal)	0.8066251	Non	6-8 ans	504.2307692
TAju(ms)	Test t de Student (var égal)	0.8066251	Non	9-11 ans	508.9705882
TExe(ms)	Test de Mann-Whitney	0.7489421	Non	6-8 ans	271.5384615
TExe(ms)	Test de Mann-Whitney	0.7489421	Non	9-11 ans	300.9411765
TPas(ms)	Test t de Student (var égal)	0.6294428	Non	6-8 ans	775.7692308
TPas(ms)	Test t de Student (var égal)	0.6294428	Non	9-11 ans	802.7647059
TLoad(ms)	Test t de Student (var égal)	0.4300492	Non	6-8 ans	287.3846154
TLoad(ms)	Test t de Student (var égal)	0.4300492	Non	9-11 ans	268.4117647
TUnload(ms)	Test t de Student (var égal)	0.4368950	Non	6-8 ans	201.9230769
TUnload(ms)	Test t de Student (var égal)	0.4368950	Non	9-11 ans	224.7058824
CoP_tot_AP(/Long)	Test t de Student (var égal)	0.7124539	Non	6-8 ans	0.0389123
CoP_tot_AP(/Long)	Test t de Student (var égal)	0.7124539	Non	9-11 ans	0.0406420
CoP_tot_ML(/Larg)	Test t de Student (var égal)	0.4070435	Non	6-8 ans	0.1096158
CoP_tot_ML(/Larg)	Test t de Student (var égal)	0.4070435	Non	9-11 ans	0.1164093
CoP_load_ML(/Larg)	Test t de Student (var égal)	0.1913699	Non	6-8 ans	0.1507763
CoP_load_ML(/Larg)	Test t de Student (var égal)	0.1913699	Non	9-11 ans	0.1779306
CoP_unload_ML(/Larg)	Test t de Student (var égal)	0.8429170	Non	6-8 ans	0.3121257
CoP_unload_ML(/Larg)	Test t de Student (var égal)	0.8429170	Non	9-11 ans	0.3157475
Vpic_load_AP(m/s)	Test de Mann-Whitney	0.0249143	Oui	6-8 ans	0.1740515
Vpic_load_AP(m/s)	Test de Mann-Whitney	0.0249143	Oui	9-11 ans	0.2512397
Vpic_load_ML(m/s)	Test t de Student (var égal)	0.1893836	Non	6-8 ans	0.2519010
Vpic_load_ML(m/s)	Test t de Student (var égal)	0.1893836	Non	9-11 ans	0.3081270

Analyse des résultats

La **grande majorité des tests** montrent des p-values **supérieures à 0,05**, indiquant qu'il n'y a **pas de différence significative** entre les groupes d'âges pour la plupart des variables.

Seule **Vpic_load_AP(m/s)** présente une p-value inférieure à 0,05 (**0.0249**), ce qui indique une différence significative entre les groupes pour cette variable. Cela signifie que les enfants plus âgés ont une vitesse de charge du CoP plus élevée.

Pour les autres variables, il n'y a pas de preuve statistique permettant d'affirmer que l'âge influence significativement les différences observées.

Q3.2

Création de trois groupes d'âge

10-11 ans	6-7 ans	8-9 ans
11	13	6

Sélection du test statistique

Pour comparer ces trois groupes, nous avons choisi un **test ANOVA** pour les variables suivant une **distribution normale** et avec **homogénéité des variances**. En revanche, pour les variables qui ne respectaient pas ces conditions, le **test de Kruskal-Wallis** a été utilisé.

Test de normalité (Shapiro-Wilk)

Groupe_Age3 <chr>	p_value <dbl>	Variable <chr>
10-11 ans	8.984362e-01	TAju(ms)
6-7 ans	3.485266e-01	TAju(ms)
8-9 ans	8.884202e-01	TAju(ms)
10-11 ans	3.598525e-05	TExe(ms)
6-7 ans	2.006489e-03	TExe(ms)
8-9 ans	1.301150e-01	TExe(ms)
10-11 ans	4.588542e-01	TPas(ms)
6-7 ans	1.563582e-01	TPas(ms)
8-9 ans	2.082514e-01	TPas(ms)
10-11 ans	9.251084e-01	TLoad(ms)
6-7 ans	5.960574e-01	TLoad(ms)
8-9 ans	5.607355e-01	TLoad(ms)
10-11 ans	3.477375e-01	TUnload(ms)
6-7 ans	4.075086e-01	TUnload(ms)
8-9 ans	4.390303e-01	TUnload(ms)
10-11 ans	6.784177e-01	CoP_tot_AP(/Long)
6-7 ans	7.845126e-01	CoP_tot_AP(/Long)
8-9 ans	8.072918e-01	CoP_tot_AP(/Long)
10-11 ans	7.405572e-01	CoP_tot_ML(/Larg)
6-7 ans	7.296329e-01	CoP_tot_ML(/Larg)
8-9 ans	9.894529e-01	CoP_tot_ML(/Larg)
10-11 ans	7.314157e-01	CoP_load_ML(/Larg)
6-7 ans	1.276788e-01	CoP_load_ML(/Larg)
8-9 ans	7.528587e-01	CoP_load_ML(/Larg)
10-11 ans	3.752550e-01	CoP_unload_ML(/Larg)
6-7 ans	2.154465e-01	CoP_unload_ML(/Larg)
8-9 ans	7.453629e-01	CoP_unload_ML(/Larg)
10-11 ans	4.027844e-01	Vpic_load_AP(m/s)
6-7 ans	3.877639e-02	Vpic_load_AP(m/s)
8-9 ans	9.845856e-02	Vpic_load_AP(m/s)
10-11 ans	8.054970e-01	Vpic_load_ML(m/s)
6-7 ans	9.544958e-02	Vpic_load_ML(m/s)
8-9 ans	4.005584e-01	Vpic_load_ML(m/s)

Test d'homogénéité des variances (Levene)

Variable <chr>	p_value <dbl>
TAju(ms)	0.1557481
TExe(ms)	0.9405211
TPas(ms)	0.5874998
TLoad(ms)	0.9914516
TUnload(ms)	0.3572654
CoP_tot_AP(/Long)	0.3798946
CoP_tot_ML(/Larg)	0.5484490
CoP_load_ML(/Larg)	0.3574218
CoP_unload_ML(/Larg)	0.6549964
Vpic_load_AP(m/s)	0.9988317
Vpic_load_ML(m/s)	0.5866309
Vpic_unload_ML(m/s)	0.2912578

Vérification des conditions du test

Avant d'appliquer l'ANOVA, les hypothèses suivantes ont été vérifiées :

- **Normalité des données** : Test de Shapiro-Wilk.
- **Homogénéité des variances** : Test de Levene.

Les résultats ont montré que :

- Pour les variables normalement distribuées et homogènes en variances, l'**ANOVA** a été appliqué.
- Pour les variables non normales ou avec des variances hétérogènes, le **test de Kruskal-Wallis** a été utilisé.

Application des tests et résultats

Résultats des tests statistiques pour comparer les trois groupes d'âge

Variable	Test_utilisé	p_value	Significatif
TAju(ms)	ANOVA	0.9334501	Non
TExe(ms)	Kruskal-Wallis	0.0621794	Non
TPas(ms)	ANOVA	0.7609210	Non
TLoad(ms)	ANOVA	0.6842165	Non
TUnload(ms)	ANOVA	0.4526874	Non
CoP_tot_AP(/Long)	ANOVA	0.8702278	Non
CoP_tot_ML(/Larg)	ANOVA	0.2785121	Non
CoP_load_ML(/Larg)	ANOVA	0.1944756	Non
CoP_unload_ML(/Larg)	ANOVA	0.7820469	Non
Vpic_load_AP(m/s)	Kruskal-Wallis	0.0813666	Non
Vpic_load_ML(m/s)	ANOVA	0.2526505	Non
Vpic_unload_ML(m/s)	ANOVA	0.8542104	Non

Tests ANOVA

L'ANOVA a été appliqué aux variables suivantes :

- **TAju(ms), TPas(ms), TLoad(ms), TUnload(ms), CoP_tot_AP(Long), CoP_tot_ML(Larg), CoP_load_ML(Larg), CoP_unload_ML(Larg), Vpic_load_ML(m/s), Vpic_unload_ML(m/s).**

Les p-values obtenues étant toutes **supérieures à 0.05**, aucune différence significative n'a été trouvée entre les trois groupes pour ces variables.

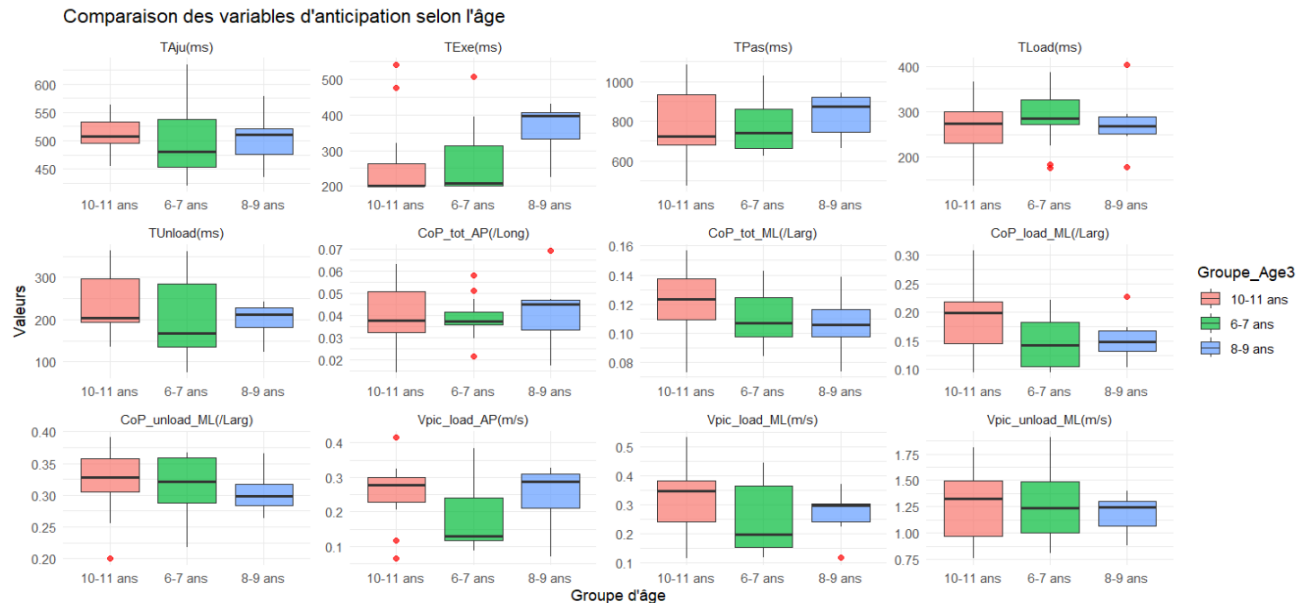
Tests de Kruskal-Wallis

Ce test a été appliqué aux variables ne respectant pas l'hypothèse de normalité :

- **TExe(ms) (p = 0.062)**
- **Vpic_load_AP(m/s) (p = 0.081)**

Ces valeurs étant **légèrement supérieures à 0.05**, on ne peut pas conclure à une différence statistiquement significative.

Visualisation des résultats avec des boxplots



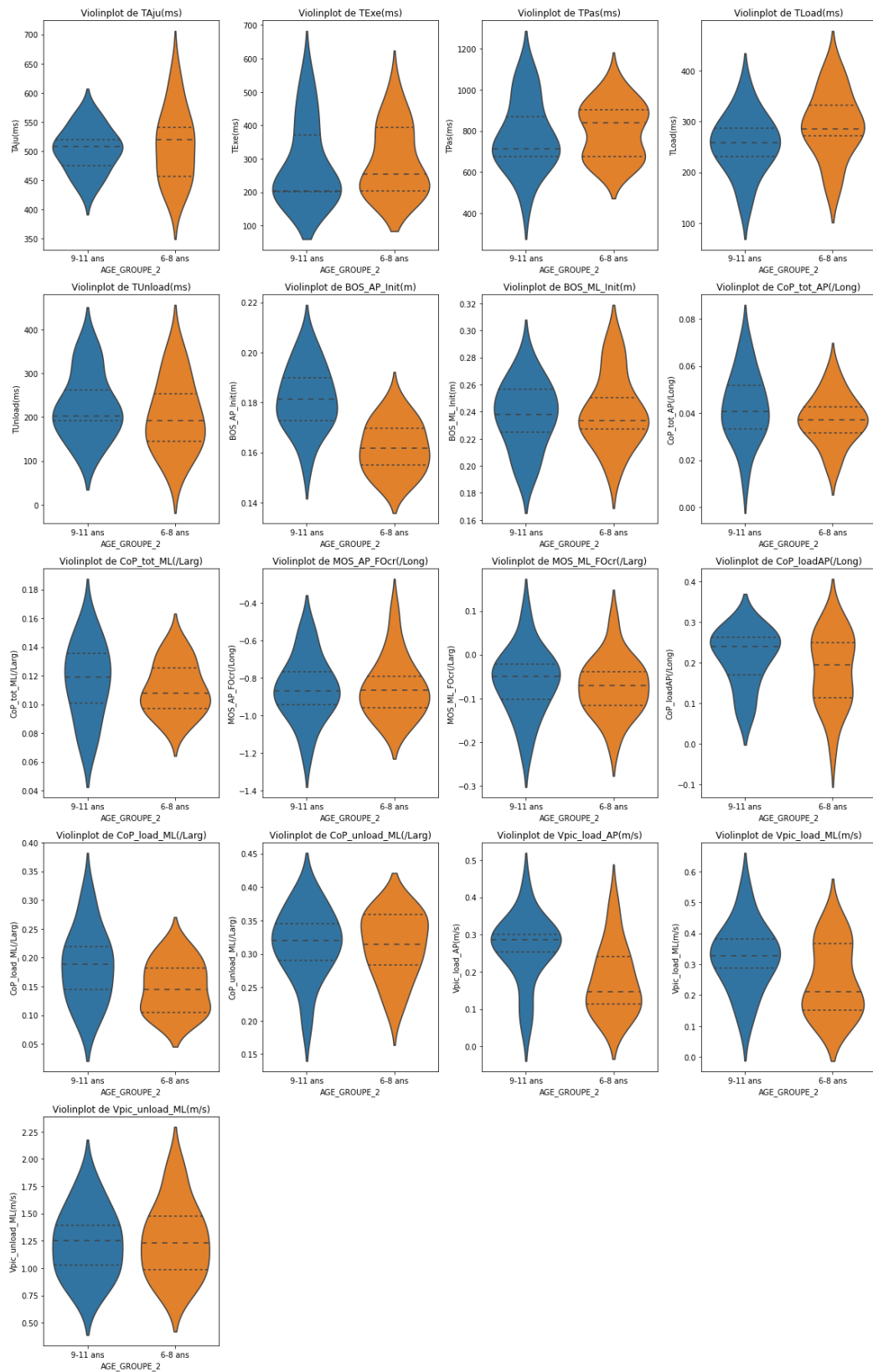
Le boxplot est un graphique qui résume la distribution d'une variable numérique à travers ses quartiles, sa médiane, et ses valeurs aberrantes (outliers). Voici pourquoi il est utile :

Comparaison rapide : Il permet de comparer la répartition des variables entre les groupes d'âge (6-7 ans, 8-9 ans et 10-11 ans) en un seul coup d'œil. On peut voir si l'un des groupes a des valeurs plus grandes ou plus petites que l'autre.

Identification des outliers : Les boxplots montrent clairement les valeurs aberrantes, ce qui peut être utile si on veut vérifier si certains points de données peuvent fausser l'interprétation des résultats.

Simplicité : C'est un graphique très simple à comprendre et qui met bien en évidence les différences entre les groupes, sans trop de détails superflus.

Visualisation avec des violinplot



Le violinplot est similaire au boxplot, mais il ajoute la densité de probabilité de la variable à chaque côté du boxplot, ce qui permet de mieux comprendre la forme de la distribution (symétrie, multimodalité, etc.). Voici pourquoi il est également pertinent :

Représentation complète de la distribution : Contrairement au boxplot, qui montre uniquement des résumés statistiques (quartiles, médiane, outliers), le violinplot montre l'ensemble de la distribution de la variable. Nous pouvons voir s'il y a des modes multiples (par exemple, une distribution bimodale, avec deux pics) ou des asymétries dans les données.

Meilleure visualisation des différences : En plus des quartiles et de la médiane, il montre où les données sont les plus concentrées. Cela peut être intéressant si nous voulons visualiser non seulement la médiane mais aussi la variation des données dans chaque groupe d'âge.

Esthétique : Les violinplots sont souvent plus visuellement attrayants et donnent un aperçu plus riche de la structure de la distribution des données.

Méthode alternative: Régression Linéaire

L'objectif de cette analyse est de tester l'effet de l'âge sur les variables d'anticipation sans avoir à regrouper les individus en deux ou trois catégories d'âge. Pour cela, nous avons utilisé une régression linéaire, qui permet de modéliser directement la relation entre l'âge et chaque variable d'anticipation.

Justification du choix de la régression linéaire

L'approche par régression linéaire présente plusieurs avantages :

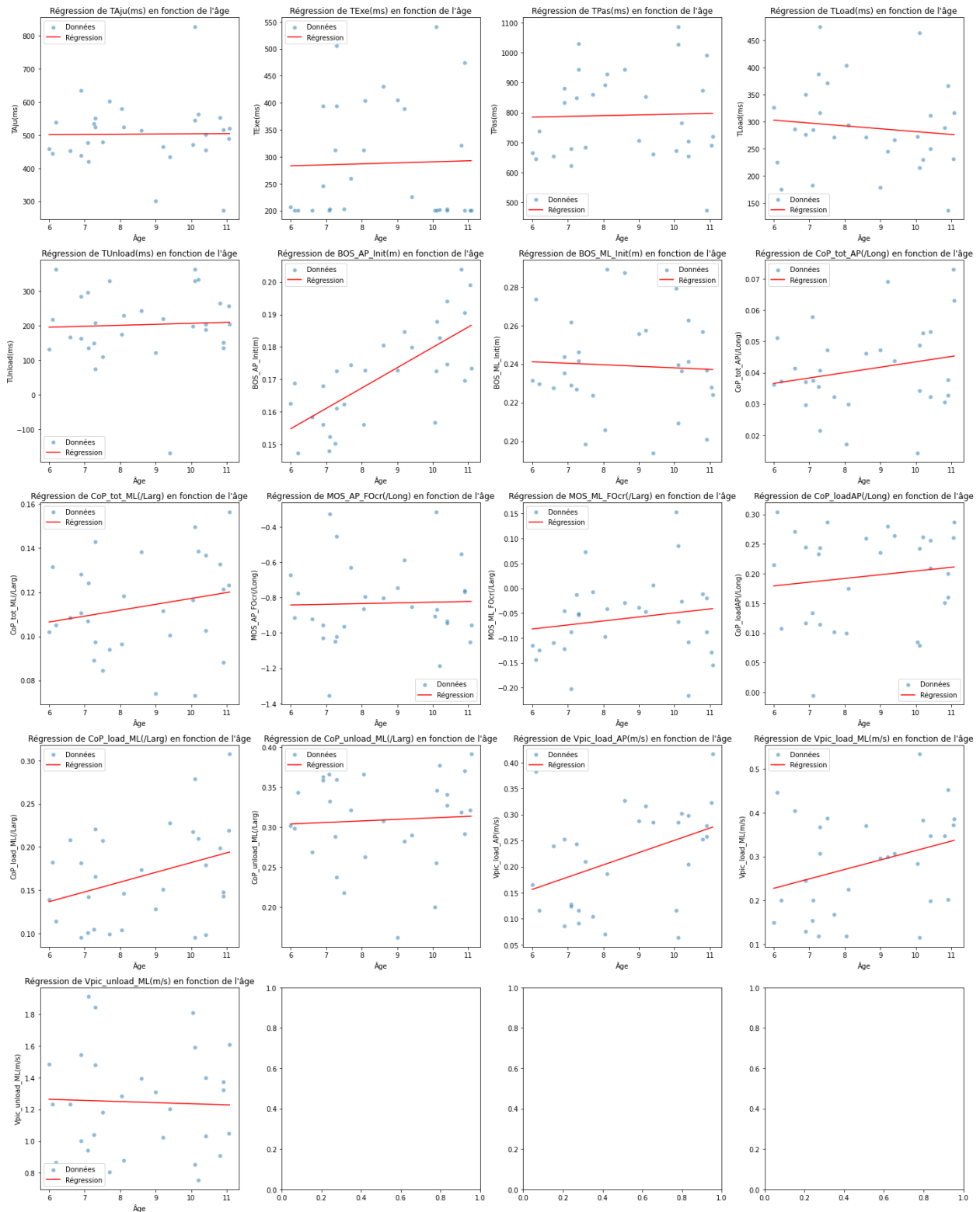
- **Précision accrue** : elle évite la perte d'information due à la catégorisation de l'âge.
- **Analyse plus fine** : elle permet de tester directement l'effet de l'âge sur chaque variable.
- **Visualisation des tendances** : elle met en évidence les relations linéaires entre l'âge et les variables mesurées.

Résultats de la régression

Le tableau et le graphique suivants résument les résultats des régressions linéaires pour chaque variable d'anticipation :

Résultats de la régression linéaire entre l'âge et les variables d'anticipation

Variable	Coefficient_Age	p_value
TAju(ms)	2.2503268	0.6923182
TExe(ms)	1.8717777	0.8776444
TPas(ms)	2.4991363	0.8794448
TLoad(ms)	-5.5281312	0.4341471
TUnload(ms)	6.2383207	0.4692639
CoP_tot_AP(/Long)	0.0007065	0.6081544
CoP_tot_ML(/Larg)	0.0026715	0.2649455
CoP_load_ML(/Larg)	0.0113120	0.0597847
CoP_unload_ML(/Larg)	0.0026358	0.6230687
Vpic_load_AP(m/s)	0.0235347	0.0238421
Vpic_load_ML(m/s)	0.0215570	0.0833588
Vpic_unload_ML(m/s)	-0.0069568	0.8446142



Analyse des résultats

- La plupart des p-values sont **supérieures à 0.05**, indiquant que l'effet de l'âge sur ces variables n'est pas statistiquement significatif.

- Seule la variable **Vpic_load_AP(m/s)** présente une **p-value inférieure à 0.05**, suggérant un effet significatif de l'âge sur cette variable.
- Certaines variables, comme **CoP_load_ML(/Larg)** ($p=0.059$), montrent des tendances proches de la significativité.

Les graphiques de dispersion ci-dessus montrent les tendances de chaque variable en fonction de l'âge avec une droite de régression superposée :

- La plupart des variables ne présentent pas de relation linéaire forte avec l'âge.
- Certaines variables (ex. Vpic_load_AP(m/s)) montrent une légère tendance croissante avec l'âge.
- Quelques variables, comme **TLoad(ms)**, semblent montrer une relation négative avec l'âge, mais sans significativité statistique.

Conclusion sur la regression linéaire simple

L'approche par régression linéaire a permis d'examiner directement l'effet de l'âge sur les variables d'anticipation. Cependant, la plupart des relations testées ne sont pas significatives, sauf pour **Vpic_load_AP(m/s)**, où l'âge semble avoir une influence significative.

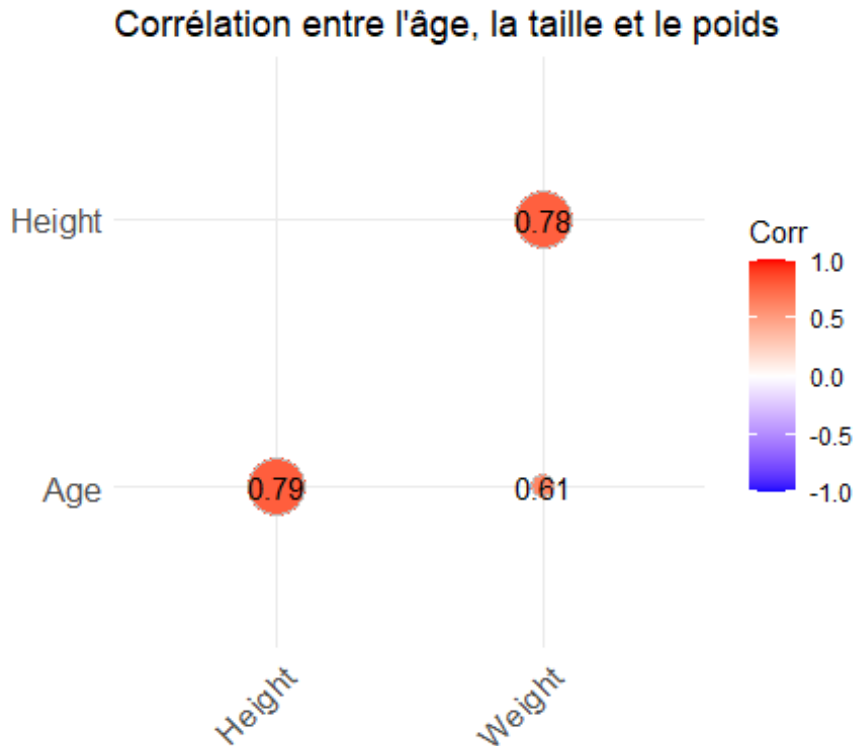
Cette méthode présente l'avantage de conserver toute l'information de l'âge sans perte due à la catégorisation, mais les résultats montrent qu'un modèle plus complexe (ex. régression non linéaire, interaction avec d'autres variables) pourrait être envisagé pour mieux comprendre les variations des performances en fonction de l'âge.

Effet de l'âge sur la taille et le poids

L'objectif de cette question est d'évaluer l'influence de l'âge sur la taille et le poids des enfants. Il est également nécessaire de s'assurer que les effets observés sont bien dus à l'âge et non à d'autres variables confondantes comme la taille et le poids eux-mêmes.

Analyse de la corrélation entre l'âge, la taille et le poids

Avant de réaliser une régression, nous avons exploré les relations entre ces variables en calculant des coefficients de corrélation.



L'âge est **fortement corrélé à la taille** (coefficient de corrélation : **0.79**).

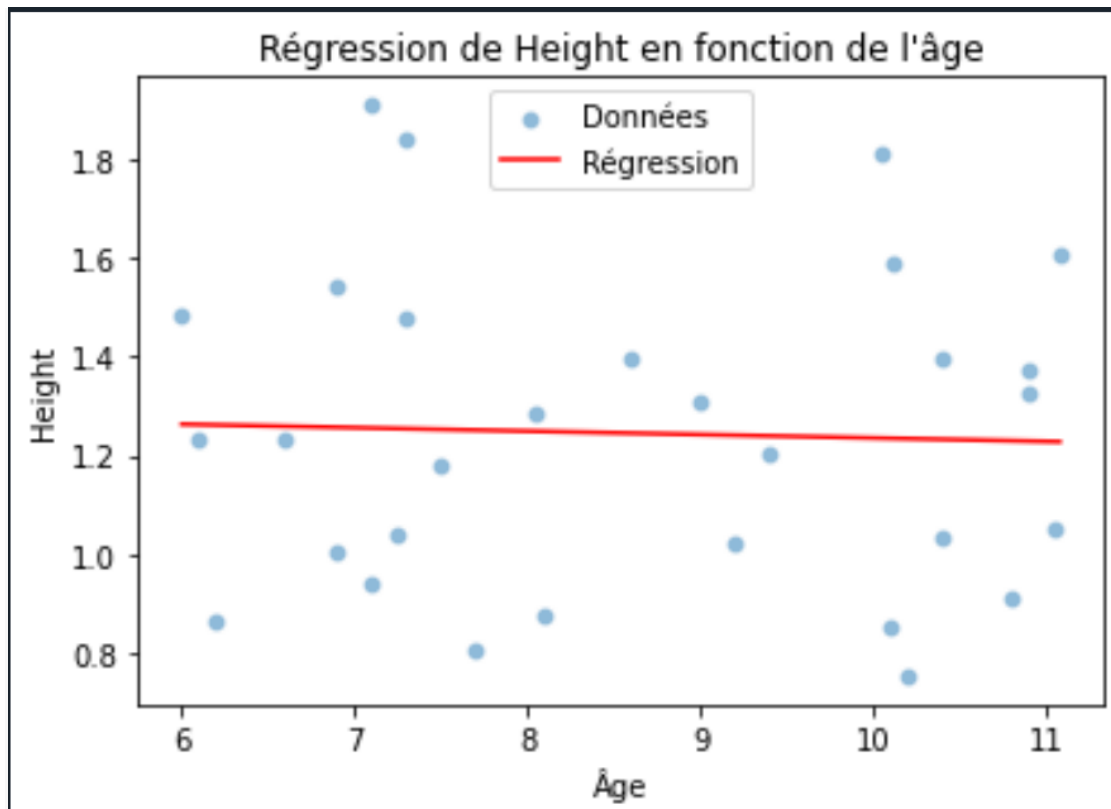
L'âge est **modérément corrélé au poids** (coefficient de corrélation : **0.61**).

La taille et le poids sont également corrélés entre eux (**0.78**), suggérant que la taille pourrait être une variable confondante dans l'analyse de l'effet de l'âge sur le poids.

Régression linéaire simple : Effet de l'âge sur la taille et le poids

Nous avons réalisé deux modèles de régression pour voir si l'âge influence significativement la taille et le poids.

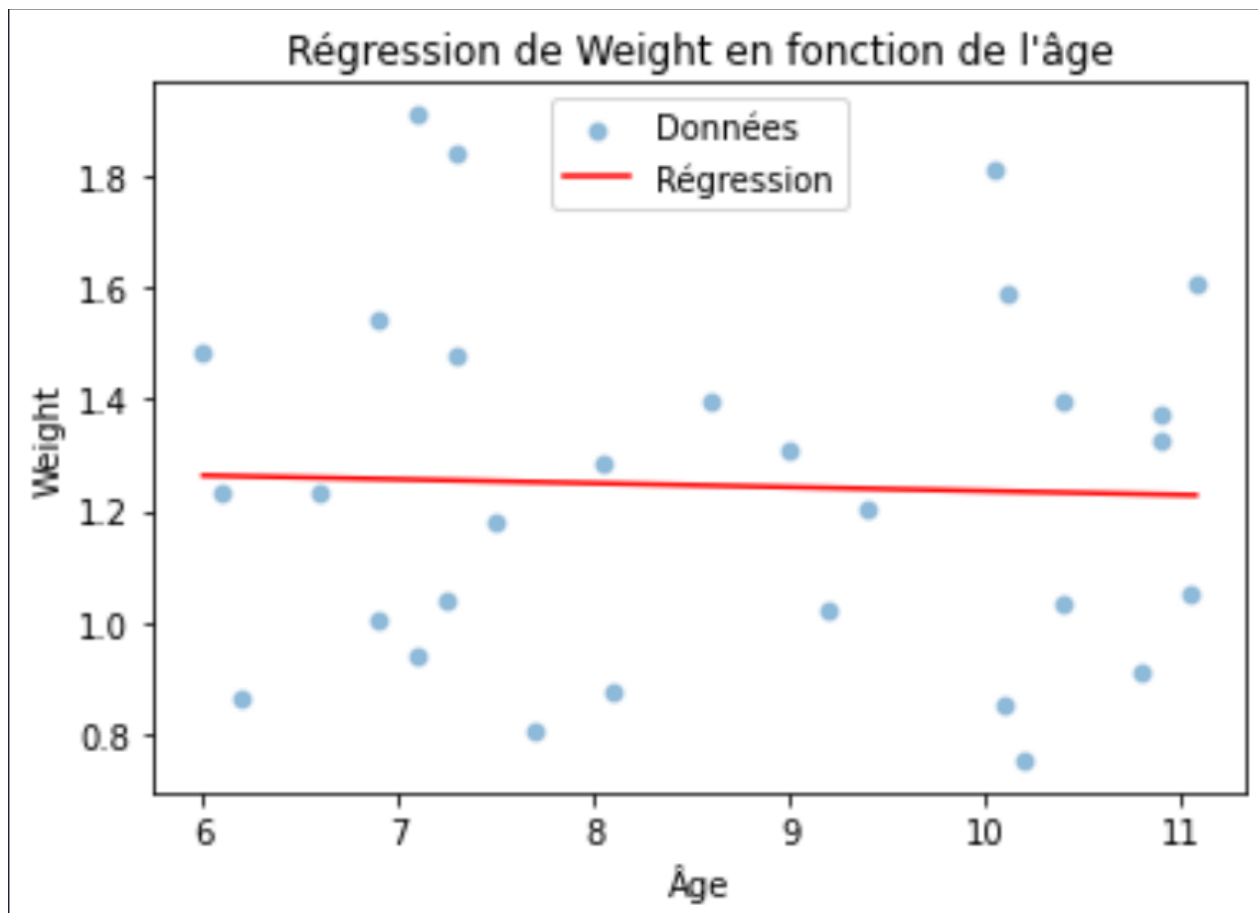
```
##
## Call:
## lm(formula = Height ~ Age, data = donnees)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -14.966  -3.935   1.168   3.439  11.524
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  92.5157     6.2075   14.90 7.67e-15 ***
## Age          4.9037     0.7076    6.93 1.56e-07 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```



```
## Residual standard error: 6.534 on 28 degrees of freedom
## Multiple R-squared:  0.6317, Adjusted R-squared:  0.6186
## F-statistic: 48.03 on 1 and 28 DF,  p-value: 1.559e-07
```

Régression du poids en fonction de l'âge

```
##
## Call:
## lm(formula = Weight ~ Age, data = donnees)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.9029 -3.2568 -0.5629  2.6440 12.0898
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   9.1199     4.9405   1.846 0.075501 .
## Age           2.3012     0.5632   4.086 0.000333 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.2 on 28 degrees of freedom
## Multiple R-squared:  0.3736, Adjusted R-squared:  0.3512
## F-statistic: 16.7 on 1 and 28 DF,  p-value: 0.0003334
```

L'effet de l'âge est **significatif** sur la taille (**p-value < 0.001**, coefficient positif).

L'effet de l'âge est également **significatif** sur le poids (**p-value = 0.0003**, coefficient positif).

L'âge explique **61.8 % de la variance de la taille** et **35.1 % de la variance du poids** (R^2 ajusté).

Régression linéaire multiple : Contrôler les effets confondants

Cette approche permet de vérifier si l'effet de l'âge demeure significatif après ajustement pour la taille et le poids.

Régression de la taille en fonction de l'âge et du poids

```
##  
## Call:  
## lm(formula = Height ~ Age + Weight, data = donnees)  
##
```

```
## Residuals:
##      Min       1Q   Median       3Q      Max
## -8.1571 -3.5995 -0.7351  3.4201 15.3275
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  85.5409      5.3118  16.104 2.28e-15 ***
## Age           3.1437      0.7223   4.352 0.000173 ***
## Weight        0.7648      0.1918   3.987 0.000459 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 5.279 on 27 degrees of freedom
## Multiple R-squared:  0.7682, Adjusted R-squared:  0.751
## F-statistic: 44.73 on 2 and 27 DF, p-value: 2.69e-09

# Régression du poids en fonction de l'âge et de la taille

##
## Call:
## lm(formula = Weight ~ Age + Height, data = donnees)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -10.5559 -2.3136 -0.5626  1.9861  9.0317
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) -35.70053    11.93058  -2.992 0.005854 **
## Age          -0.07446     0.74977  -0.099 0.921626
## Height        0.48446     0.12153   3.987 0.000459 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 4.201 on 27 degrees of freedom
## Multiple R-squared:  0.6057, Adjusted R-squared:  0.5765
## F-statistic: 20.73 on 2 and 27 DF, p-value: 3.501e-06
```

Après ajustement pour la taille, **l'effet de l'âge sur le poids devient non significatif ($p = 0.921$)**, indiquant que l'effet de l'âge sur le poids est principalement médié par la taille.

En revanche, l'effet de l'âge sur la taille reste **très significatif** même après ajustement pour le poids (**$p < 0.001$**), ce qui confirme que l'âge a un effet direct sur la croissance.

Documentations

[https://fr.wikipedia.org/wiki/M%C3%A9diane_\(statistiques\)](https://fr.wikipedia.org/wiki/M%C3%A9diane_(statistiques))

<https://sites.google.com/site/rgraphiques/4--stat/tests-statistiques-avec-r>

<https://www.sthda.com/french/wiki/tests-statistiques-avec-r>

<https://seaborn.pydata.org/generated/seaborn.violinplot.html>