

ANALYSE DE L'INITIATION DE LA MARCHÉ CHEZ DES ENFANTS AGES DE 6 A 12 ANS, NEUROTYPQUES ET ATTEINTS DE TROUBLE DU SPECTRE DE L'AUTISME

Effet du TSA

Auteurs

DAOUDA DIOP

BRUN BAHOUN HOUTOUKPE

UBS
ETUDE DE CAS

Introduction

Dans le cadre de cette étude, nous nous sommes intéressés à l'analyse de l'initiation de la marche chez des enfants âgés de 6 à 12 ans, à travers la comparaison entre un groupe d'enfants neurotypiques (NT) et un groupe d'enfants présentant un trouble du spectre de l'autisme (TSA). Le trouble du spectre de l'autisme se caractérise par une grande hétérogénéité des symptômes et une variété d'intensités, notamment en matière de communication, de comportements sociaux et, dans certains cas, de capacités motrices.

L'objectif principal de cette seconde partie était de **comparer les deux groupes** sur plusieurs aspects (données générales, paramètres biomécaniques, scores cliniques), d'**explorer les relations entre variables motrices et scores TSA**, et d'**identifier d'éventuels profils d'enfants similaires** à l'aide d'une approche de clustering non supervisé.

Toutes les analyses ont été réalisées avec le logiciel R. Le code correspondant ainsi que les fichiers de données ont été fournis en annexe. Ce rapport présente, pour chaque question posée, les traitements réalisés, les justifications méthodologiques des choix effectués (tests statistiques, visualisations, regroupements) ainsi que les résultats obtenus sous forme de tableaux et de graphiques.

Q2.1 Appariement des enfants NT et TSA

1. Comment a été effectué l'appariement ?

Dans cette étude, il a été nécessaire de sélectionner **30 enfants neurotypiques (NT)** parmi **40 initialement disponibles**, afin de les appairer avec **30 enfants atteints de trouble du spectre de l'autisme (TSA)** en fonction de trois critères :

- **L'âge** (Age),
- **La taille** (Height),
- **Le poids** (Weight).

Plusieurs approches peuvent être envisagées pour réaliser un appariement entre deux groupes :

1. **Appariement exact :**

Cette méthode consiste à sélectionner les enfants NT ayant exactement les mêmes valeurs d'âge, de taille et de poids que les enfants TSA. Toutefois, cette approche est souvent **trop restrictive**, car elle limite fortement les possibilités d'appariement et risque d'exclure de nombreux sujets.

2. **Appariement par le plus proche voisin ("Nearest Neighbor Matching") :**

Cette technique consiste à comparer chaque enfant TSA aux enfants NT disponibles et à sélectionner celui dont les caractéristiques sont les plus proches, en minimisant la **distance** entre les valeurs des trois critères d'appariement.

- Cette distance est généralement calculée à l'aide de la **distance euclidienne** ou du **score de propension**.
- Une seule correspondance est attribuée à chaque enfant TSA.

3. **Appariement avec une marge de tolérance ("Caliper Matching") :**

Cette méthode impose une **marge de tolérance** (appelée "caliper") pour limiter l'écart

entre les valeurs des variables d'appariement, garantissant ainsi une correspondance plus stricte. Cependant, si les données ne sont pas suffisamment nombreuses, cette méthode peut **réduire le nombre d'individus appariés**.

2. Justification du choix de l'appariement par le plus proche voisin

L'appariement a été réalisé en utilisant la méthode du **plus proche voisin** ("Nearest Neighbor Matching") pour plusieurs raisons :

- **Souplesse** : Contrairement à l'appariement exact, cette approche permet d'obtenir un appariement optimal sans exclure trop d'individus.
- **Précision** : Elle minimise l'écart entre les enfants NT et TSA en optimisant la similarité sur les critères définis (âge, taille et poids).
- **Fiabilité statistique** : Elle permet d'obtenir deux groupes équilibrés, réduisant ainsi les biais dans l'analyse des résultats.

Cette méthode garantit un échantillon **comparatif équilibré**, essentiel pour analyser les effets du TSA sans introduire de biais liés aux caractéristiques démographiques et anthropométriques.

3. Comment vérifier que l'appariement a été correctement effectué ?

Pour s'assurer que l'appariement est bien réalisé, plusieurs analyses ont été effectuées :

a) Comparaison des statistiques descriptives entre NT et TSA

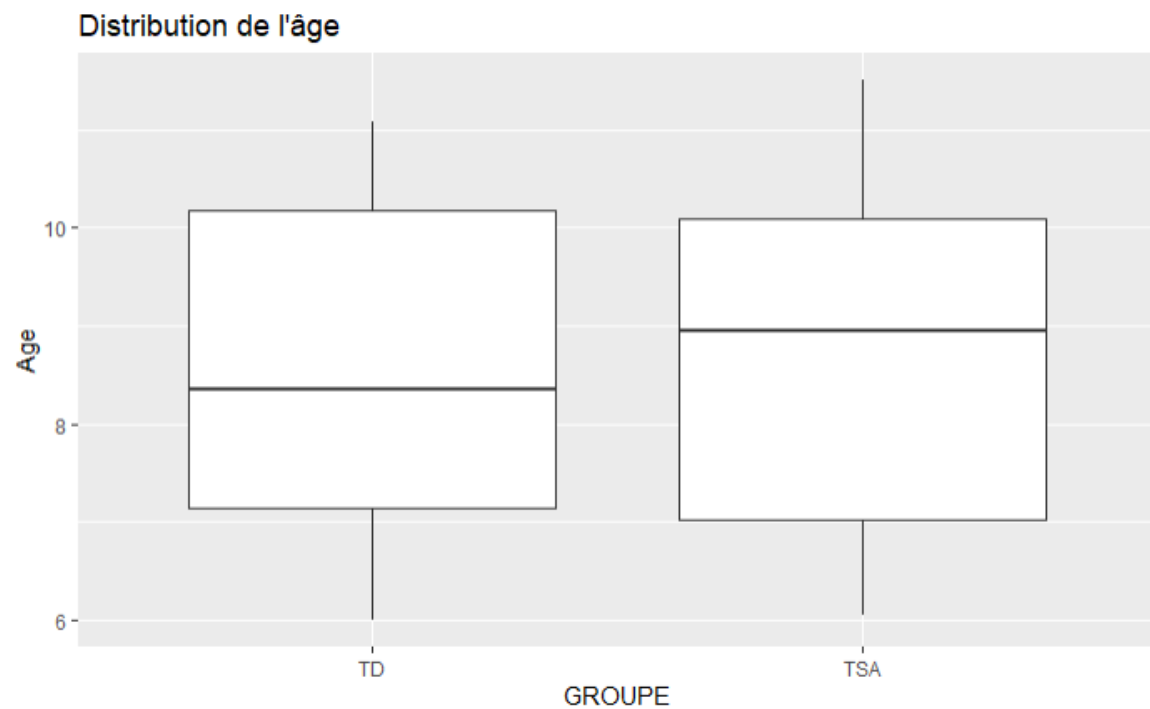
Avec R, nous avons calculé les **moyennes et écarts-types** des variables **âge, taille et poids** pour chaque groupe :

GROUPE	Age_Moyenne	Age_EcartType	Taille_Moyenne	Taille_EcartType	Poids_Moyen	Poids_EcartType
TD	8.609333	1.714623	134.7333	10.57871	28.76500	6.02319
TSA	8.755000	1.778569	133.4833	14.23599	29.80309	11.96507

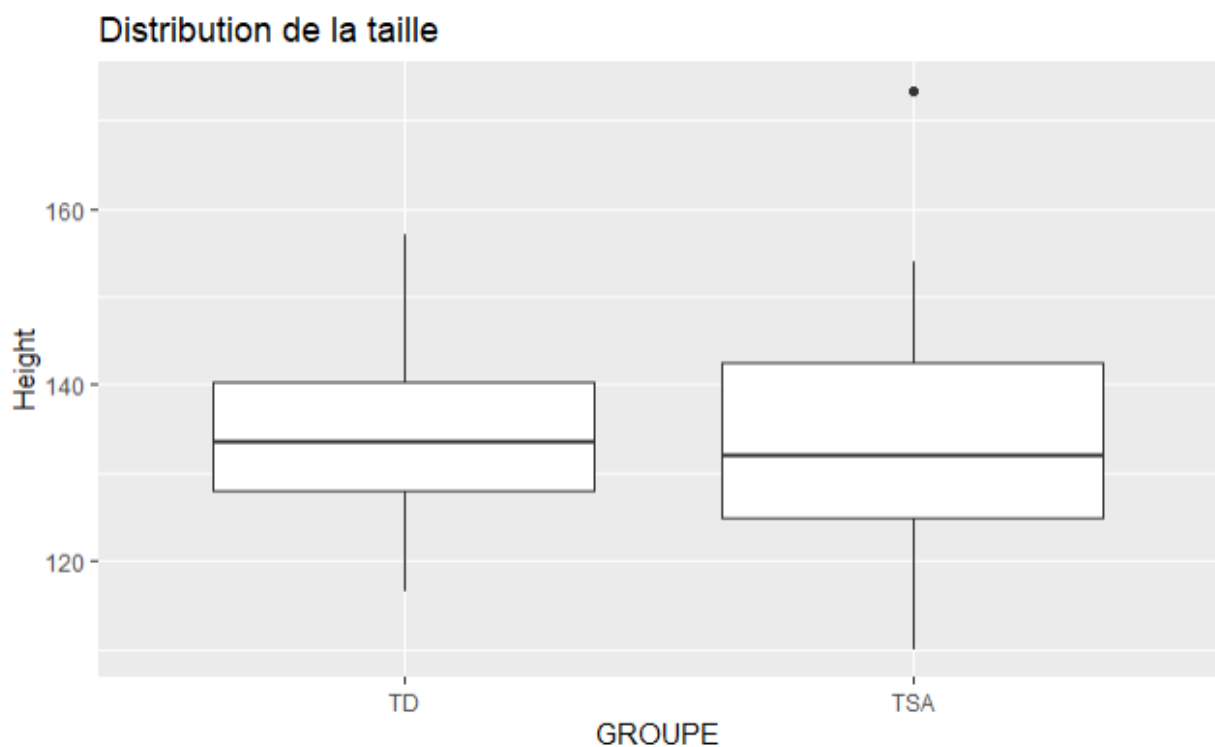
Les moyennes et les écarts-types sont proches entre NT et TSA, indiquant un bon équilibre entre les groupes.

b) Visualisation des distributions

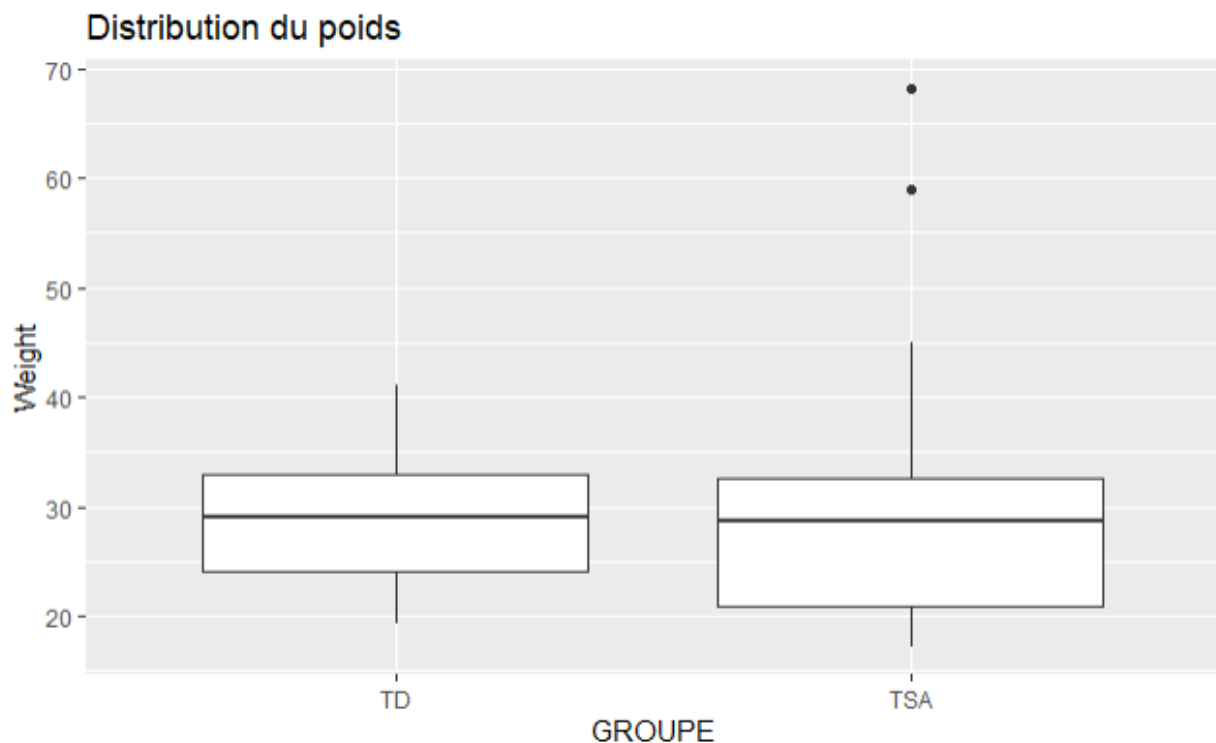
Nous avons également réalisé des boxplots pour visualiser la répartition des variables.



Les distributions sont similaires entre les deux groupes.



Au niveau de la taille, il y a quelques variations mais pas de différences majeures entre les groupes.



Pour le poids, il y a quelques valeurs extrêmes détectées, mais la distribution générale est homogène.

Les boxplots confirment que les distributions des variables sont comparables entre les groupes NT et TSA.

c) Tests statistiques pour valider l'appariement

Nous avons réalisé deux types tests de comparaison des groupes :

- **Test de Student (t-test)** pour comparer les moyennes
- **Test de Mann-Whitney (Wilcoxon Rank Sum Test)** pour vérifier si les distributions sont similaires

Résultats des test de Student avec R

```
Welch Two Sample t-test

data: Age by GROUPE
t = -0.32295, df = 57.922, p-value = 0.7479
alternative hypothesis: true difference in means between group TD and group TSA is not equal to 0
95 percent confidence interval:
-1.0485564  0.7572231
sample estimates:
mean in group TD mean in group TSA
      8.609333      8.755000
```

```
Welch Two Sample t-test

data: Height by GROUPE
t = 0.38602, df = 53.543, p-value = 0.701
alternative hypothesis: true difference in means between group TD and group TSA is not equal to 0
95 percent confidence interval:
-5.243412  7.743412
sample estimates:
mean in group TD mean in group TSA
      134.7333      133.4833
```

```
Welch Two Sample t-test

data: Weight by GROUPE
t = -0.42446, df = 42.811, p-value = 0.6734
alternative hypothesis: true difference in means between group TD and group TSA is not equal to 0
95 percent confidence interval:
-5.970918  3.894742
sample estimates:
mean in group TD mean in group TSA
      28.76500      29.80309
```

Le **test de Student** compare les moyennes des groupes NT et TSA pour chaque variable :

- **Âge** : $p=0.7479$
- **Taille** : $p=0.701$
- **Poids** : $p=0.6734$

Les p-values sont toutes supérieures à 0.05, indiquant qu'il n'y a pas de différence statistiquement significative entre NT et TSA pour ces trois variables. Cela signifie que les groupes sont bien équilibrés et comparables.

Résultats des tests de Mann-Whitney

```

Wilcoxon rank sum test with continuity correction
data: Age by GROUPE
W = 438, p-value = 0.8649
alternative hypothesis: true location shift is not equal to 0

Wilcoxon rank sum test with continuity correction
data: Height by GROUPE
W = 475, p-value = 0.717
alternative hypothesis: true location shift is not equal to 0

Wilcoxon rank sum test with continuity correction
data: Weight by GROUPE
W = 491.5, p-value = 0.5443
alternative hypothesis: true location shift is not equal to 0

```

Ce test non paramétrique vérifie si les distributions des groupes sont similaires, sans supposer qu'elles suivent une loi normale :

- **Âge** : $p=0.8649$
- **Taille** : $p=0.717$
- **Poids** : $p=0.5443$

Encore une fois, les p-values sont toutes supérieures à 0.05, confirmant qu'il n'existe pas de différence significative entre NT et TSA.

L'ensemble des résultats indique que :

- Les groupes **ne présentent pas de différences significatives** sur les variables d'appariement.
- L'appariement a été **réalisé correctement** et **ne crée pas de biais dans l'analyse des résultats**.

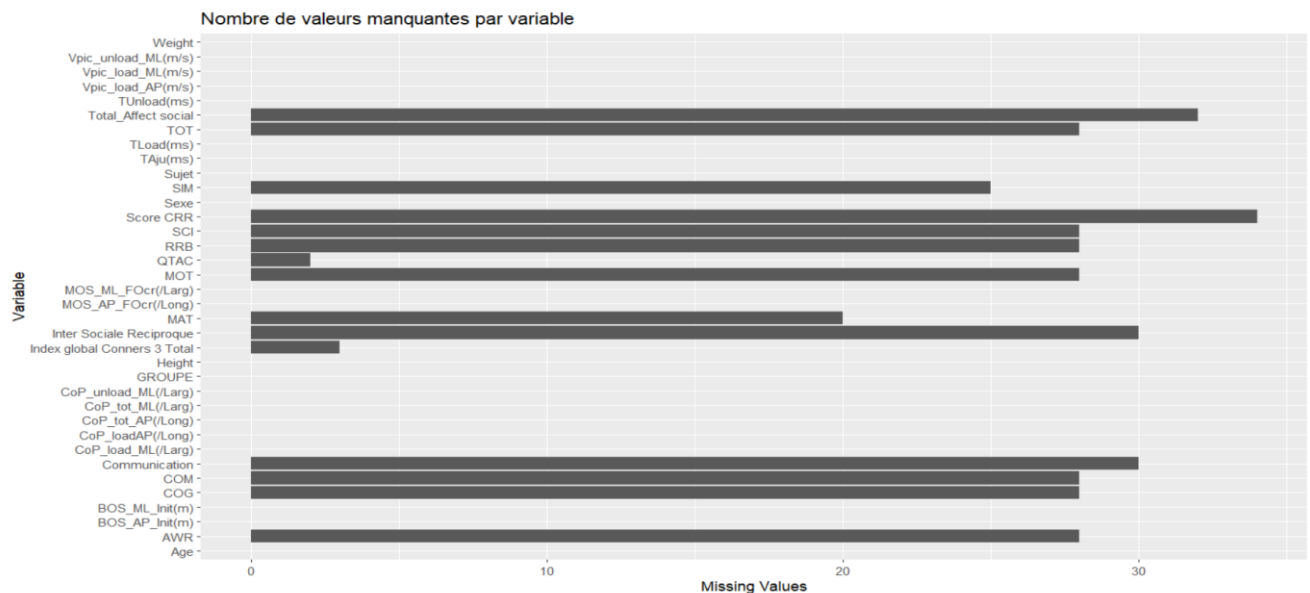
Nous pouvons donc conclure que l'appariement entre NT et TSA est valide et statistiquement justifié.

Q2.2 Détection d'anomalies

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.

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 jeu de données.

L'analyse exploratoire des données avec R a révélé la présence de valeurs manquantes dans plusieurs variables, notamment celles liées aux caractéristiques comportementales et aux indices d'évaluation cognitive. Le diagramme en barre ci-dessous nous montre bien quelles variables sont concernées.



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.

Q2.3 : Synthèse des données descriptives

Afin de mieux comprendre les caractéristiques des participants des deux groupes (TSA et TD), nous avons procédé à une analyse descriptive des principales variables de l'étude. Cette étape exploratoire permet de dégager une première vue d'ensemble des distributions et de détecter d'éventuelles valeurs atypiques. Pour ce faire, plusieurs représentations ont été mobilisées : tableaux statistiques, boxplots et histogrammes.

1. Tableaux statistiques descriptifs

Afin de décrire les caractéristiques des enfants des groupes NT et TSA, nous avons réalisé une synthèse statistique sous forme de trois tableaux descriptifs. Les variables ont été regroupées selon trois catégories principales : les données générales, les données motrices et les scores relatifs aux troubles du spectre de l'autisme (TSA).

Pour chaque variable numérique sélectionnée, nous avons calculé les moyennes, écarts-types et médianes, ce qui permet de disposer à la fois d'indicateurs de tendance centrale et de dispersion. Ces mesures sont classiques en statistique descriptive et permettent de résumer efficacement les distributions observées.

Tableau des Données Générales

GROUPE	Age_Moyenne	Age_EcartType	Age_Mediane	Height_Moyenne	Height_EcartType	Height_Mediane	Weight_Moyenne	Weight_EcartType	Weight_Mediane
TD	8.609333	1.714623	8.35	134.7333	10.57871	133.5	28.76500	6.02319	29.00000
TSA	8.755000	1.778569	8.94	133.4833	14.23599	132.0	29.80309	11.96507	28.62124

Ce tableau regroupe les données générales : âge, taille (Height) et poids (Weight).

Tableau des Données Motrices

GROUPE	TAju(ms)_Moyenne	TAju(ms)_EcartType	TAju(ms)_Mediane	TLoad(ms)_Moyenne	TLoad(ms)_EcartType	TLoad(ms)_Mediane
TD	502.8667	97.40273	507.5	289.2000	80.54658	281.0
TSA	557.9000	128.76184	571.0	291.6667	112.50665	259.5

Ce **deuxième tableau** concerne les données motrices, telles que les temps de latence (*TAju*, *TLoad*, *TUnload*) et les mesures de stabilité posturale (*CoP*).

Tableau des Scores TSA

GROUPE	AWR_Moyenne	AWR_EcartType	AWR_Mediane	COG_Moyenne	COG_EcartType	COG_Mediane	COM_Moyenne	COM_EcartType	COM_Mediane	MOT_Moyenne	MOT_EcartType	MOT_Mediane
TD	40.33333	7.637626	42	39.66667	0.5773503	40	40.66667	1.527525	41	42.33333	5.859465	40
TSA	69.37931	10.651673	70	75.93103	7.4972902	75	74.06897	8.971905	73	69.17241	9.308479	71

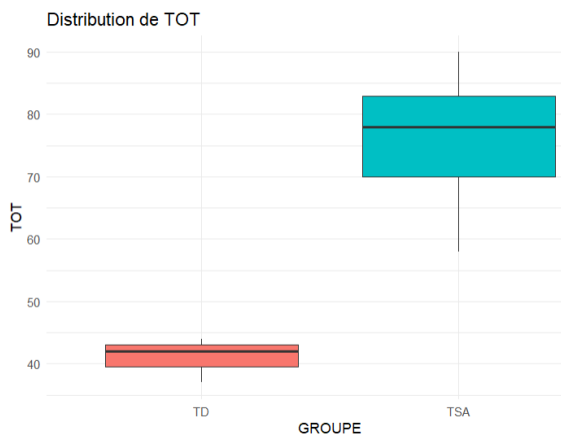
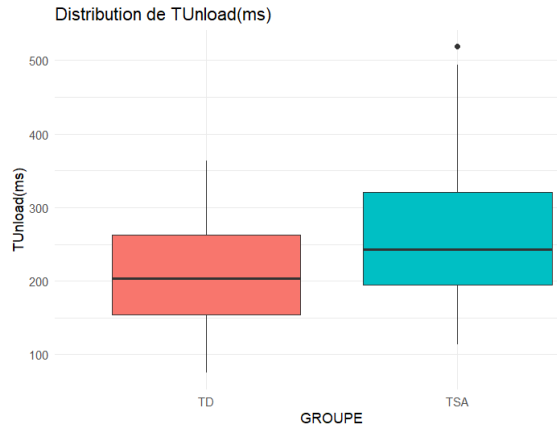
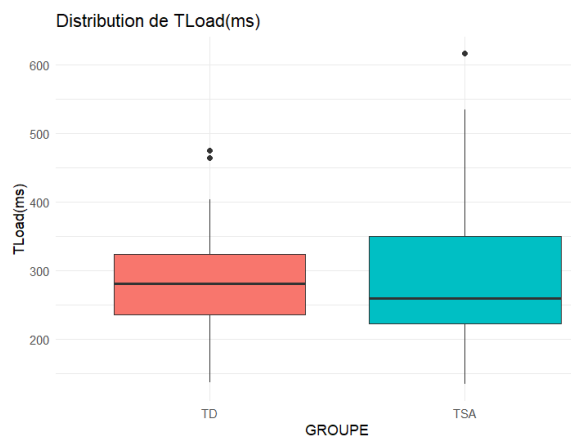
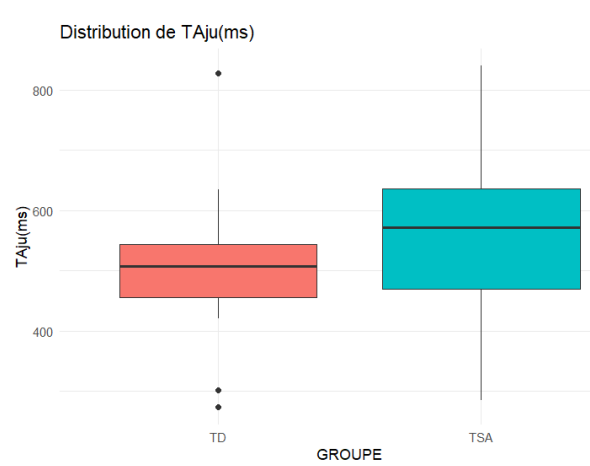
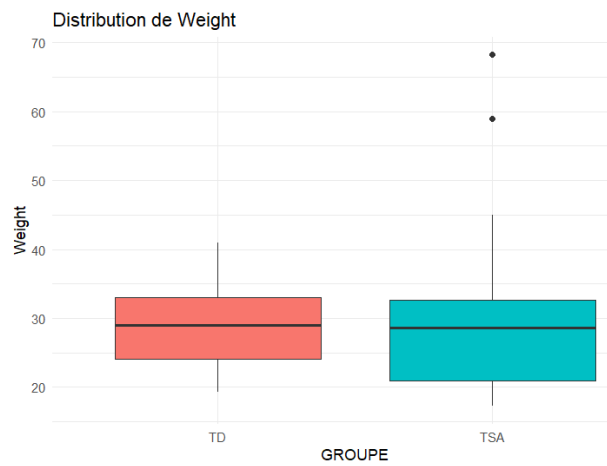
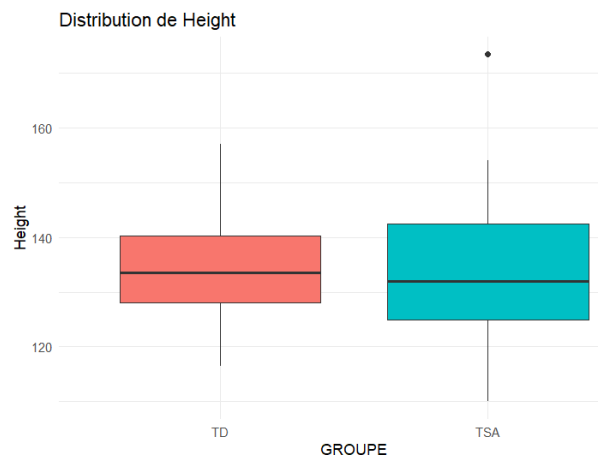
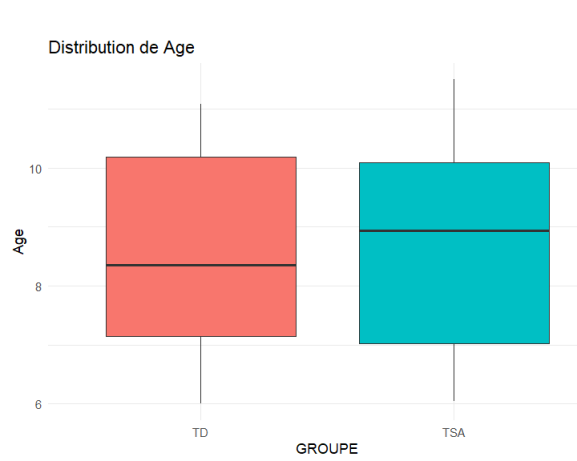
Finalement, le troisième tableau rassemble les scores issus de différentes dimensions évaluées chez les enfants TSA : communication (*COM*), motricité (*MOT*), comportement répétitif (*RRB*), etc.

Ces synthèses tabulaires permettent d'avoir une vue d'ensemble des données avant toute comparaison statistique ou modélisation.

2. Visualisation par boxplots

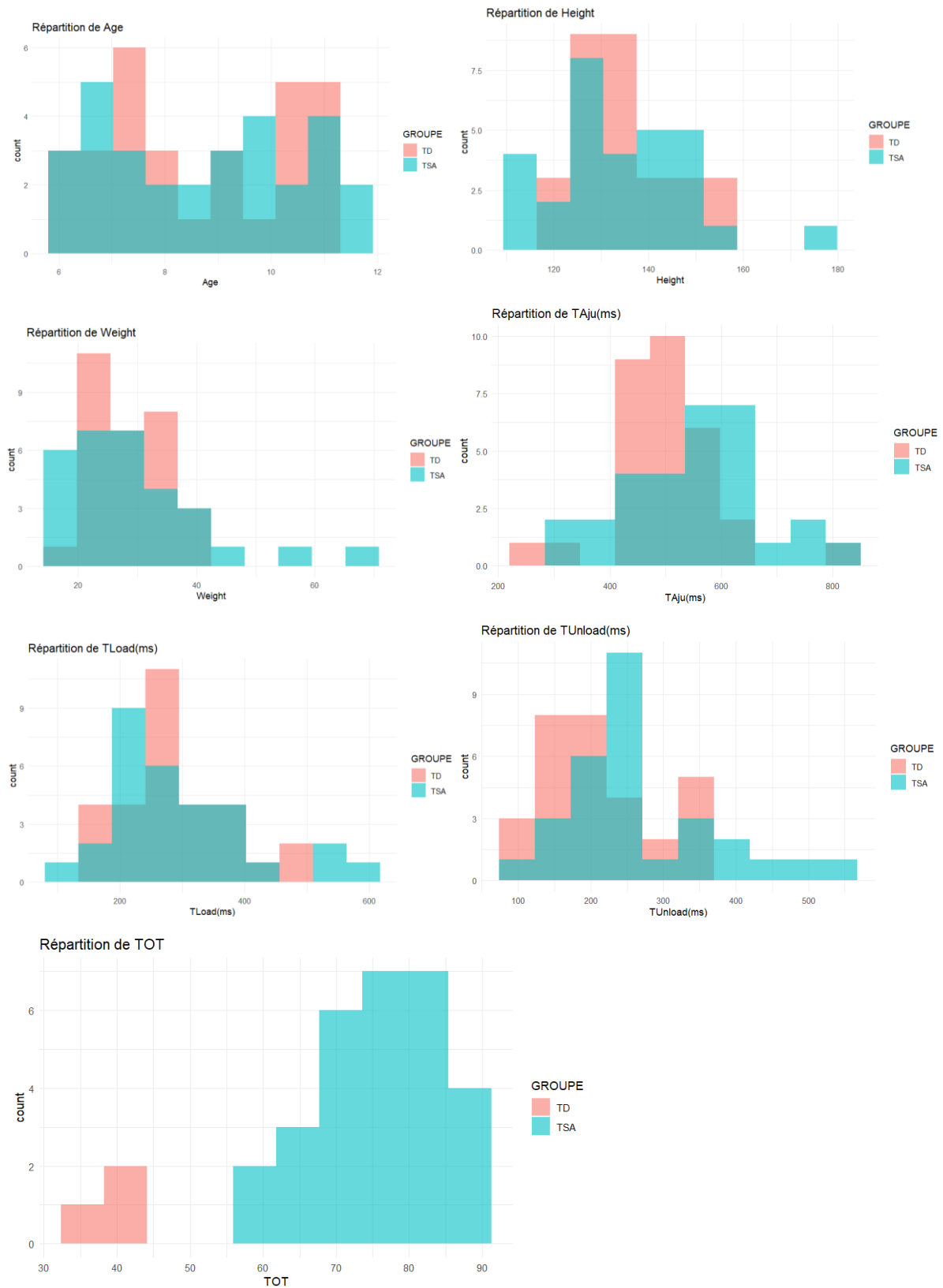
Pour compléter les tableaux, nous avons réalisé des **boxplots** afin de visualiser la distribution des principales variables par groupe. Ce type de graphique permet de :

- Visualiser la **médiane**, les **quartiles**, et les **valeurs extrêmes** (outliers)
- Comparer rapidement les **distributions** entre les groupes TSA et TD
- Repérer des **asymétries** ou des **écarts importants** de dispersion



4. Visualisation par histogrammes

Enfin, nous avons construit des **histogrammes** pour visualiser plus finement la distribution des valeurs au sein de chaque groupe. Contrairement aux boxplots, les histogrammes donnent une vision plus précise du **profil de distribution** (symétrie, concentration, étalement, etc.).



Q3.1 – Vérification des conditions d'application des tests statistiques

L'objectif de cette étape est de comparer les enfants des groupes TD et TSA sur un ensemble de variables quantitatives (âge, taille, poids, paramètres biomécaniques) à l'aide de tests statistiques appropriés. Afin de garantir la validité de ces tests, les conditions suivantes ont été vérifiées au préalable :

- La **normalité de distribution** des variables,
- L'homogénéité **des variances** entre les groupes.

Vérification de la normalité

Nous avons testé la normalité de chaque variable à l'aide du **test de Shapiro-Wilk** une méthode statistique adaptée aux petits échantillons. Ce test permet de savoir si la variable suit une distribution gaussienne, ce qui conditionne l'utilisation d'un test paramétrique (comme le test t de Student) ou non paramétrique (comme le test de Wilcoxon).

Les variables testées sont : Age, Height, Weight, TAju(ms), TLoad(ms) et TUnload(ms). Les résultats obtenus ont guidé le choix des tests dans la phase suivante.

Vérification de l'homogénéité des variances

L'égalité des variances entre les groupes a été évaluée avec le **test de Levene**. Ce test permet de savoir si les groupes présentent des dispersions similaires, ce qui est une condition nécessaire pour appliquer un test t de Student classique. Dans notre cas, les variances étaient jugées homogènes pour les variables concernées.

Choix des tests comparatifs

Le choix du test statistique a été réalisé selon les critères suivants :

- Lorsque la variable suivait une distribution normale et que les variances étaient homogènes, un **test t de Student** a été utilisé.
- Dans le cas contraire (distribution non normale ou variances inégales), un **test de Mann-Whitney-Wilcoxon**, qui est non paramétrique, a été préféré.

Ces tests ont permis de comparer les groupes TD et TSA pour chaque variable de manière indépendante.

Présentation des résultats

Les résultats des tests sont synthétisés dans le tableau ci-dessous. Celui-ci indique, pour chaque variable, le test utilisé, la valeur p obtenue, ainsi que si la différence entre les groupes est jugée statistiquement significative (seuil de 5 %).

Tableau – Résultats des tests statistiques entre TD et TSA

Résultats des tests statistiques entre NT et TSA

Variable	Test	P_Value	Significatif
Age	Wilcoxon rank sum test with continuity correction	0.8649286	Non
Height	Two Sample t-test	0.7008940	Non
Weight	Wilcoxon rank sum test with continuity correction	0.5443125	Non
TAju(ms)	Two Sample t-test	0.0669600	Non
TLoad(ms)	Wilcoxon rank sum test with continuity correction	0.5201040	Non
TUnload(ms)	Wilcoxon rank sum test with continuity correction	0.0344855	Oui

Les tests ont été appliqués de manière systématique selon les conditions précédemment établies.

Comparaison graphique des distributions

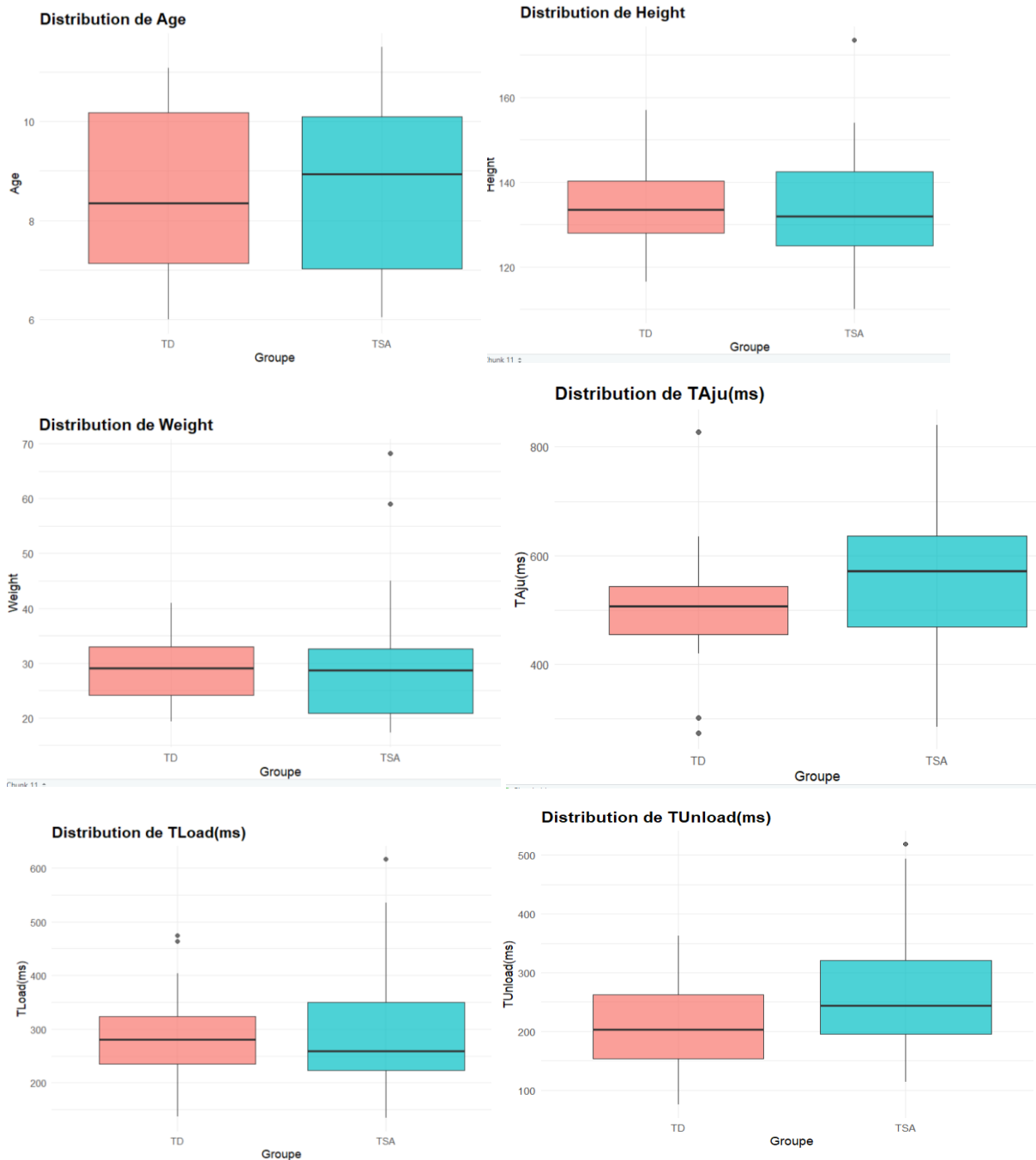
Afin de compléter les analyses statistiques réalisées sur les variables quantitatives, des **boxplots** ont été construits pour chacune des variables comparées entre les groupes TSA et TD.

Ce type de représentation graphique a été choisi car il permet de :

- visualiser la **médiane** de chaque groupe pour chaque variable,
- observer l'**étendue interquartile** (dispersion des 50 % centraux),
- repérer d'éventuelles **valeurs extrêmes** (outliers),
- **comparer visuellement** les distributions des groupes sur un même axe.

Les boxplots ont été réalisés pour les variables suivantes :

- Variables générales : Age, Height, Weight
- Variables motrices : TAju(ms), TLoad(ms), TUnload(ms)



Chaque graphique représente les données des deux groupes côte à côte, avec une couleur distincte par groupe pour améliorer la lisibilité. Ces visualisations viennent **illustrer** les résultats des tests précédents. Cette approche graphique renforce l'analyse exploratoire et permet de vérifier visuellement la pertinence des tests choisis. Elle constitue une étape complémentaire dans l'évaluation de la distribution des données et des éventuelles différences entre les groupes.

Q3.2 — Vérification d’une corrélation entre troubles de l’anticipation motrice et scores d’évaluation du TSA (SRS et ADOS)

Méthodologie

Pour évaluer les liens potentiels entre les variables biomécaniques et les scores d’évaluation du TSA, nous avons réalisé des tests de corrélation. Deux types de corrélation peuvent être appliqués selon la distribution des variables :

- La **corrélation de Pearson** est utilisée lorsque les deux variables suivent une distribution normale. Elle permet d'évaluer une corrélation linéaire.
- La **corrélation de Spearman** est préférable si au moins une des variables ne suit pas une distribution normale. Elle évalue une association monotone, sans supposer de linéarité.

Vérification de la normalité

Un test de normalité (Shapiro-Wilk) a été appliqué sur l’ensemble des variables biomécaniques (TA_{ju}, TLoad, TUnload, CoP_tot_AP, CoP_tot_ML) ainsi que sur les scores TSA (AWR, COG, COM, MOT, RRB, SCI, TOT). Les résultats (tableau ci-dessous) montrent que plusieurs variables ne suivent pas une distribution normale. Par conséquent, dans la majorité des cas, la **corrélation de Spearman** a été utilisée.

	Variable <chr>	P_Value <dbl>	Normalite <chr>
TA _{ju} (ms)	TA _{ju} (ms)	0.1749192282	Oui
TLoad(ms)	TLoad(ms)	0.0021411566	Non
TUnload(ms)	TUnload(ms)	0.0107135779	Non
CoP_tot_AP(/Long)	CoP_tot_AP(/Long)	0.6752287791	Oui
CoP_tot_ML(/Larg)	CoP_tot_ML(/Larg)	0.0038967914	Non
AWR	AWR	0.5176262704	Oui
COG	COG	0.0003866114	Non
COM	COM	0.0144515605	Non
MOT	MOT	0.2813268271	Oui
RRB	RRB	0.0032402757	Non

Résultats des corrélations

Ce tableau récapitulatif ci-dessous présente les tests réalisés pour chaque combinaison variable biomécanique / score TSA. Le test appliqué, le coefficient de corrélation, la p-value et la significativité sont indiqués.

Résultats des tests de corrélation entre variables biomécaniques et scores TSA

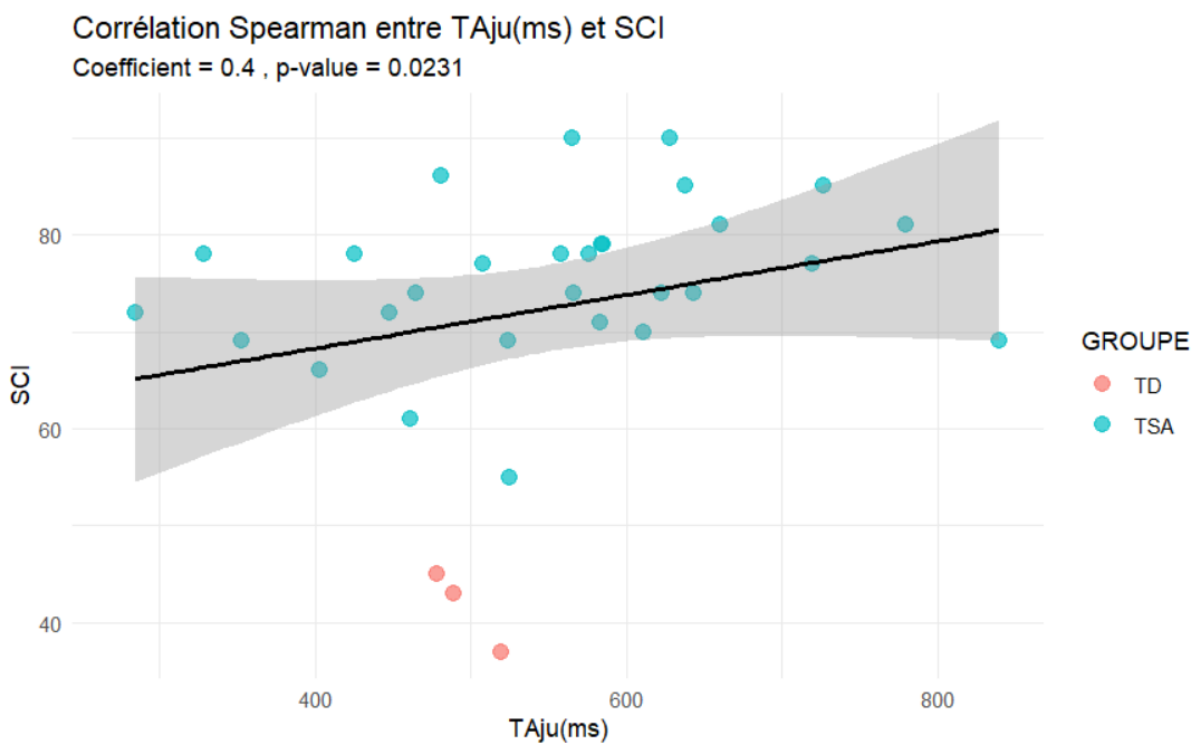
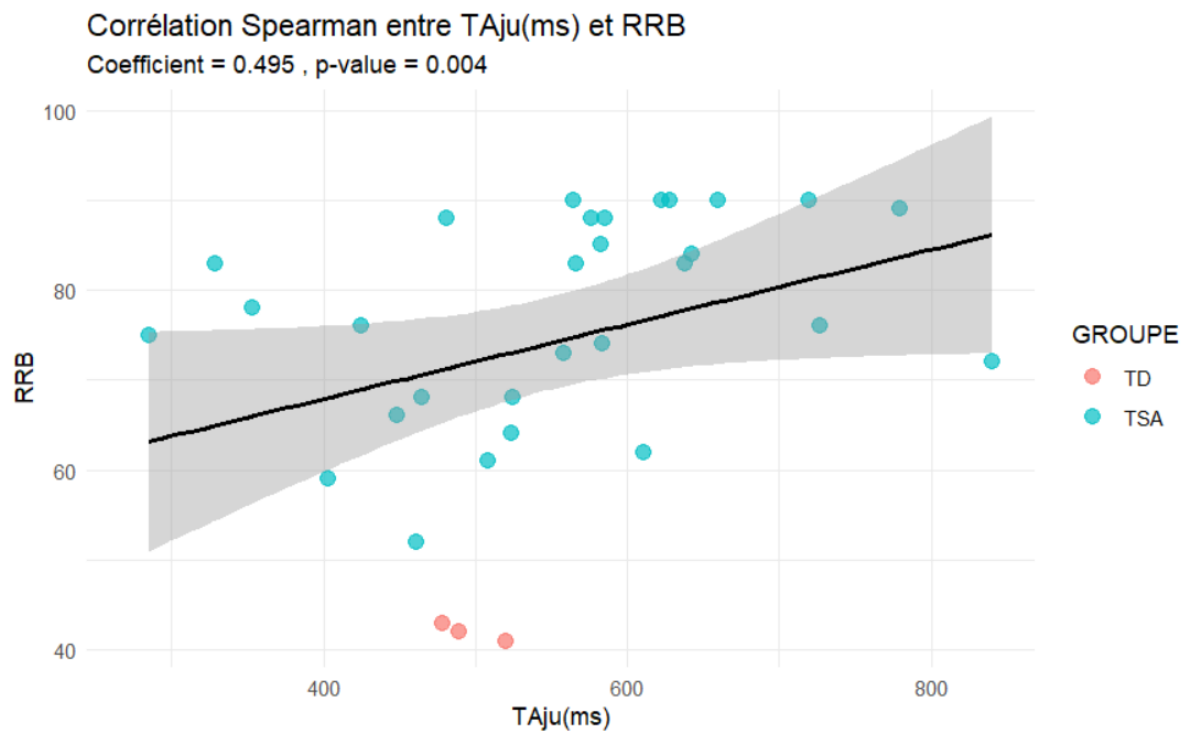
	Variable_Biomecanique	Variable_Score	Methode	Coefficient	P_Value	Significatif
cor	TAju(ms)	AWR	Pearson	0.319	0.0749	Non
rho	TAju(ms)	COG	Spearman	0.297	0.0989	Non
rho1	TAju(ms)	COM	Spearman	0.292	0.1044	Non
cor1	TAju(ms)	MOT	Pearson	0.294	0.1019	Non
rho2	TAju(ms)	RRB	Spearman	0.495	0.0040	Oui
rho3	TAju(ms)	SCI	Spearman	0.400	0.0231	Oui
rho4	TAju(ms)	TOT	Spearman	0.445	0.0106	Oui
rho5	TLoad(ms)	AWR	Spearman	0.094	0.6105	Non
rho6	TLoad(ms)	COG	Spearman	0.049	0.7898	Non
rho7	TLoad(ms)	COM	Spearman	0.048	0.7957	Non
rho8	TLoad(ms)	MOT	Spearman	0.128	0.4852	Non
rho9	TLoad(ms)	RRB	Spearman	-0.047	0.7974	Non
rho10	TLoad(ms)	SCI	Spearman	0.114	0.5355	Non
rho11	TLoad(ms)	TOT	Spearman	0.062	0.7345	Non
rho12	TUnload(ms)	AWR	Spearman	0.144	0.4305	Non
rho13	TUnload(ms)	COG	Spearman	0.124	0.4996	Non

Corrélations significatives observées

Cinq corrélations se sont révélées statistiquement significatives ($p < 0.05$) :

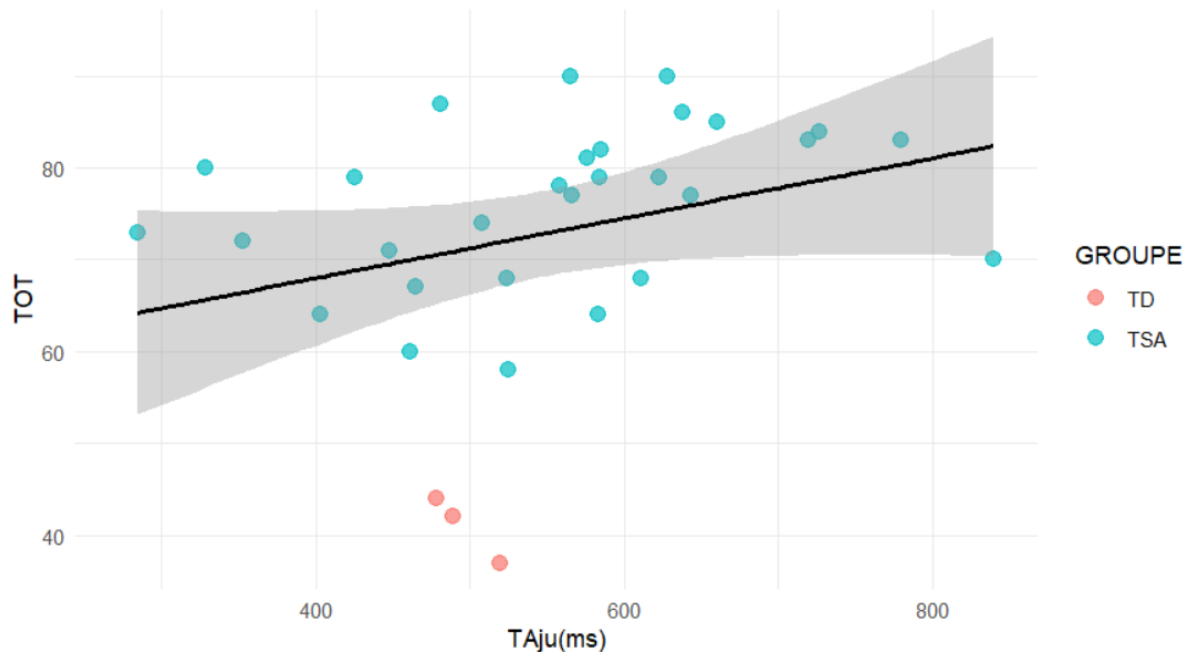
- **TAju(ms) – RRB** (Spearman, $\rho = 0.495$, $p = 0.004$)
- **TAju(ms) – SCI** (Spearman, $\rho = 0.400$, $p = 0.0231$)
- **TAju(ms) – TOT** (Spearman, $\rho = 0.445$, $p = 0.0106$)
- **TUnload(ms) – RRB** (Spearman, $\rho = 0.414$, $p = 0.0186$)
- **CoP_tot_AP(/Long) – AWR** (Pearson, $r = -0.453$, $p = 0.0092$)

Ces résultats sont illustrés par les nuages de points suivants :



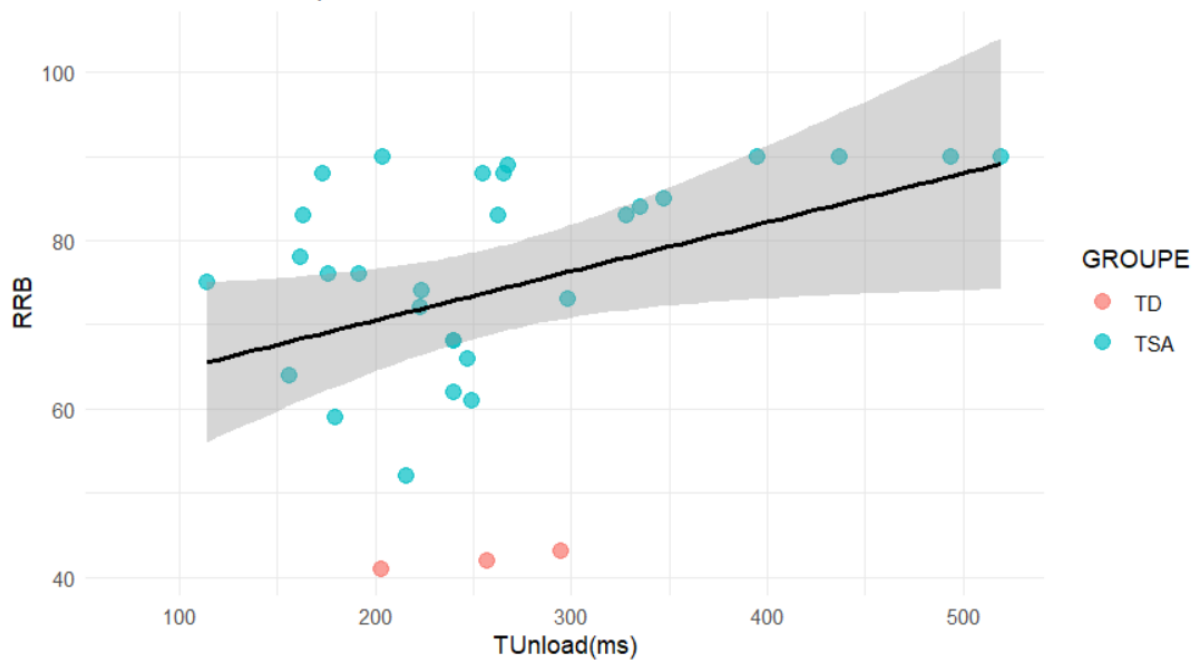
Corrélation Spearman entre TAju(ms) et TOT

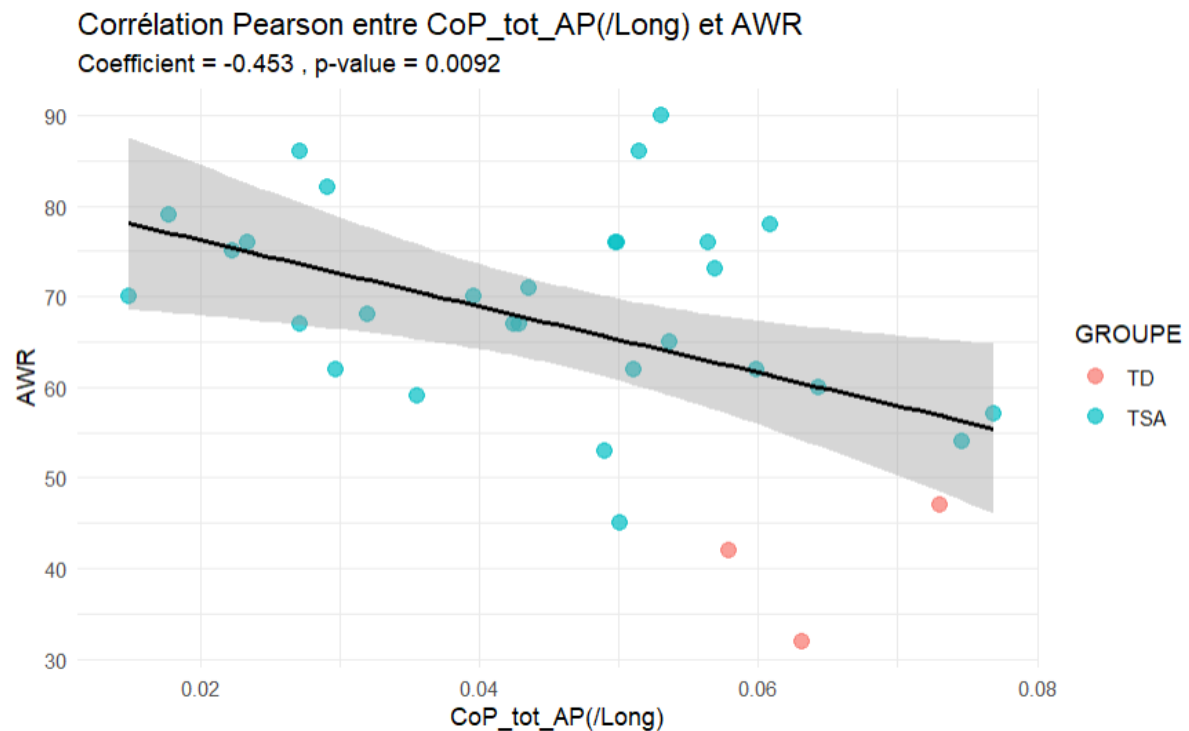
Coefficient = 0.445 , p-value = 0.0106



Corrélation Spearman entre TUnload(ms) et RRB

Coefficient = 0.414 , p-value = 0.0186





Ces associations, bien que significatives, ne permettent pas d'établir une relation causale. Elles ouvrent néanmoins la voie à des pistes d'interprétation et de réflexion sur le lien entre motricité et caractéristiques cliniques du TSA.

Q3.3 – Clustering des enfants selon leurs profils

L'objectif de cette analyse est de regrouper les enfants en fonction de la similarité de leurs profils, sans connaissance préalable des groupes (TSA ou TD). Pour cela, nous avons choisi une méthode de **clustering non supervisé**, en l'occurrence le **K-means**, appliquée aux **scores SRS** (AWR, COG, COM, MOT, RRB, SCI, TOT), jugés comme les plus représentatifs de l'intensité des symptômes TSA.

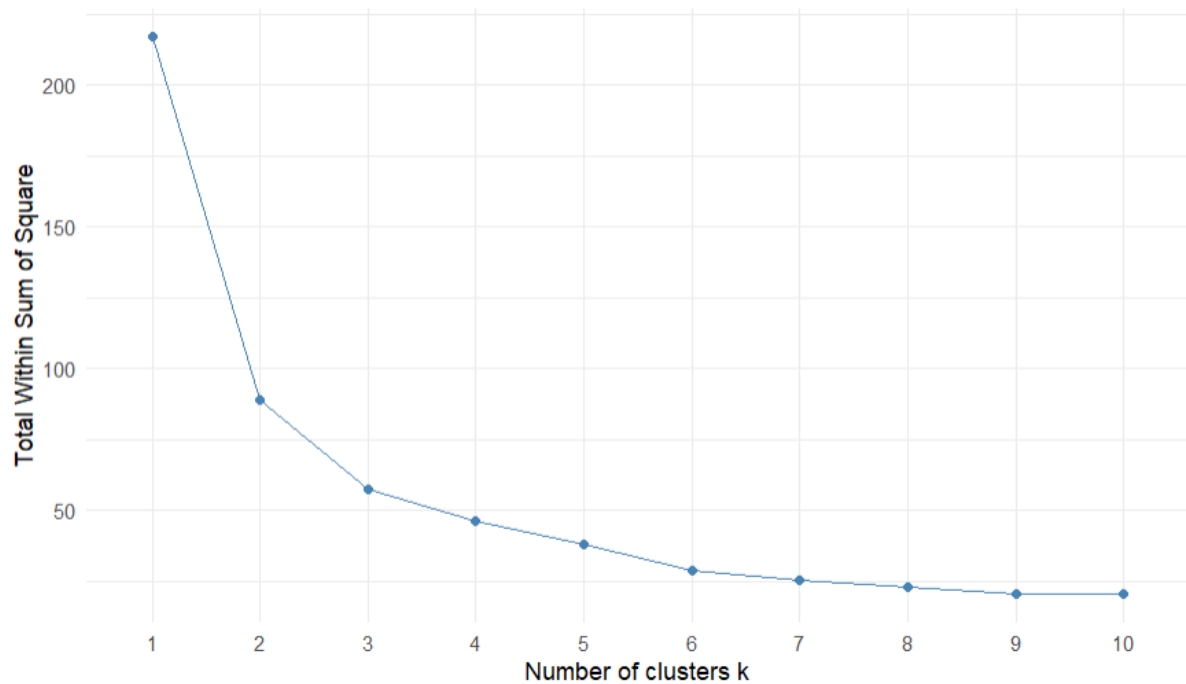
Préparation des données

Nous avons d'abord sélectionné les variables SRS d'intérêt et conservé uniquement les lignes sans valeurs manquantes. Les données ont ensuite été **centrées et réduites**, comme requis pour le K-means et l'ACP.

Choix du nombre de clusters

Le nombre optimal de clusters a été déterminé à l'aide de la **méthode du coude**, qui consiste à observer la courbe du total des carrés intra-cluster (Within Sum of Squares) en fonction du nombre de clusters. Le coude de la courbe se situe à **k = 3**, ce qui indique que trois clusters distincts permettent une bonne représentation des données (voir graphique ci-dessous).

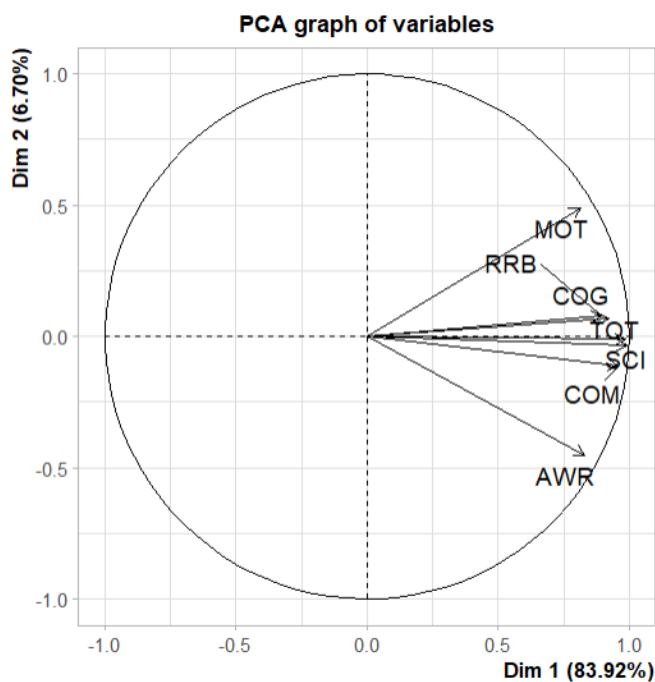
Méthode du coude : choix optimal du nombre de clusters

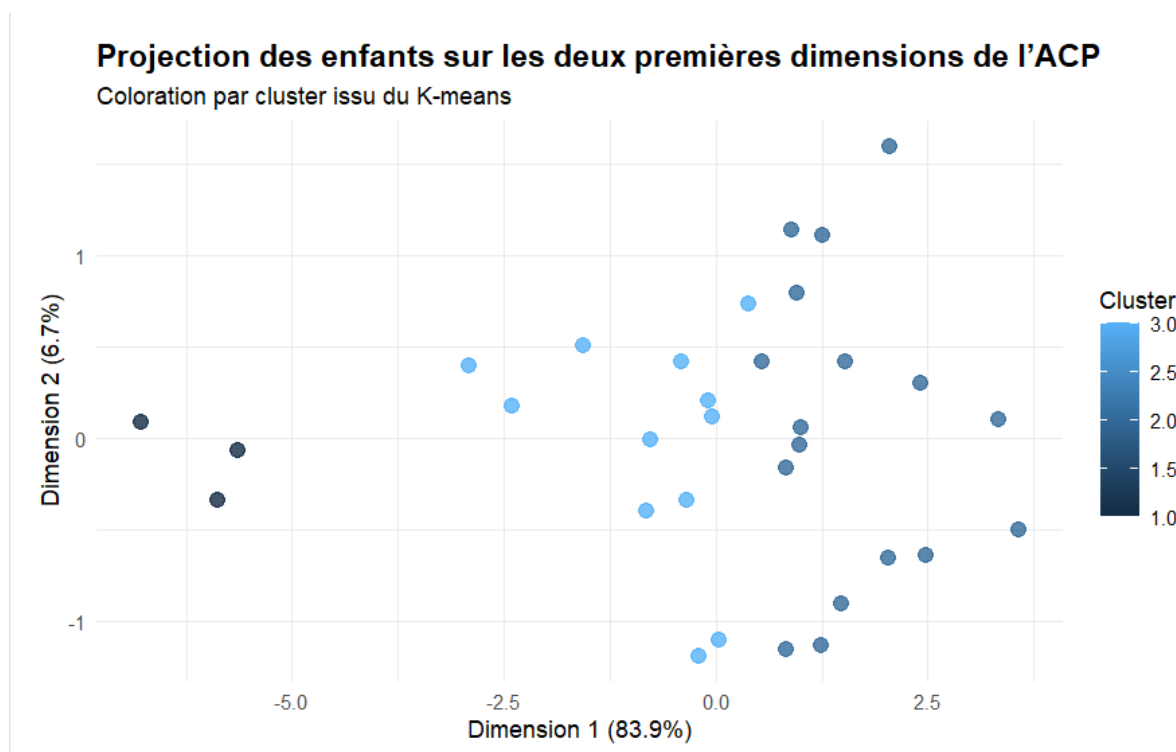


Graphique : Méthode du coude – Choix optimal de k

Visualisation des clusters à l'aide de l'ACP

Une **Analyse en Composantes Principales (ACP)** a été réalisée pour projeter les individus sur les deux premières dimensions principales. Cette projection permet de visualiser la répartition des clusters dans l'espace factoriel.





Les résultats de l'ACP montrent que la **première dimension explique à elle seule 83.9 %** de la variance totale, ce qui justifie son utilisation pour l'interprétation des clusters. Les variables les plus contributives à cette dimension sont notamment **TOT, COM, SCI, COG**, ce qui suggère que les clusters sont majoritairement différenciés sur l'intensité globale des symptômes TSA.

Interprétation des clusters

Un tableau récapitulatif des **moyennes et écarts-types des scores SRS** par cluster permet d'interpréter plus précisément les profils :

cluster	AWR_Moyenne	AWR_EcartType	COG_Moyenne	COG_EcartType	COM_Moyenne	COM_EcartType	MOT_
1	40.33	7.64	39.67	0.58	40.67	1.53	
2	72.53	9.96	79.88	6.58	79.35	6.75	
3	64.92	10.36	70.33	4.68	66.58	5.85	

Tableau : Moyennes et écarts-types des scores TSA par cluster

Conclusion

Ce travail a permis de mobiliser différentes méthodes statistiques pour répondre à une série de questions portant sur les différences et les similarités entre enfants neurotypiques et enfants atteints de TSA.

Après avoir vérifié la qualité des données et la validité de l'appariement entre les groupes, nous avons exploré les distributions des variables générales, motrices et comportementales. Des tests statistiques adaptés ont été appliqués pour comparer les groupes, et des corrélations ont été recherchées afin d'évaluer les liens entre certaines variables motrices et les scores d'évaluation du TSA. Enfin, une analyse de clustering a été conduite pour identifier des regroupements d'enfants en fonction de leurs profils comportementaux.

L'ensemble des analyses a été mené de manière rigoureuse, en tenant compte des hypothèses de chaque test et en choisissant les outils graphiques les plus adaptés à chaque étape.

Documentation

https://documentation-snds.health-data-hub.fr/snds/fiches/appariement_donnees_externes.html

<https://healthcaredatainstitute.com/2020/01/31/enrichissements-de-bases-de-donnees-relier-les-dossiers-cliniques-les-registres-ou-les-informations-declarees-par-les-patients-la-consommation-de-soins-de-sante/>

<https://www.qualtrics.com/fr/gestion-de-l-experience/etude-marche/analyse-cluster/>

<https://thinkr.fr/premiers-pas-en-machine-learning-avec-r-volume-3-k-means-clustering/>