

# Devoir 1

Aloui Mohamed, Jafrou Douba  
26 novembre 2018

## Compte rendu du devoir 1

### Les données

On considère un tableau de données sur des personnes anorexiques qui ont été suivies pendant quelque temps. Certains patients ont suivi une thérapie (c'est une thérapie Cognitive Behavioural Treatment "CBT"), d'autres n'ont suivi aucun traitement (c'est le groupe de contrôle "Cont"). Le tableau contient pour chaque individu le type de la thérapie et son poids au début et à la fin de l'étude (en livres).

- Téléchargez le fichier `DonneesAnorexie.txt` et importez les données sous forme d'un tableau (dataframe). Familiarisez-vous avec les données.
- Changez l'unité de mesure du livre au kilogramme (1 livre = 0.453 kg).

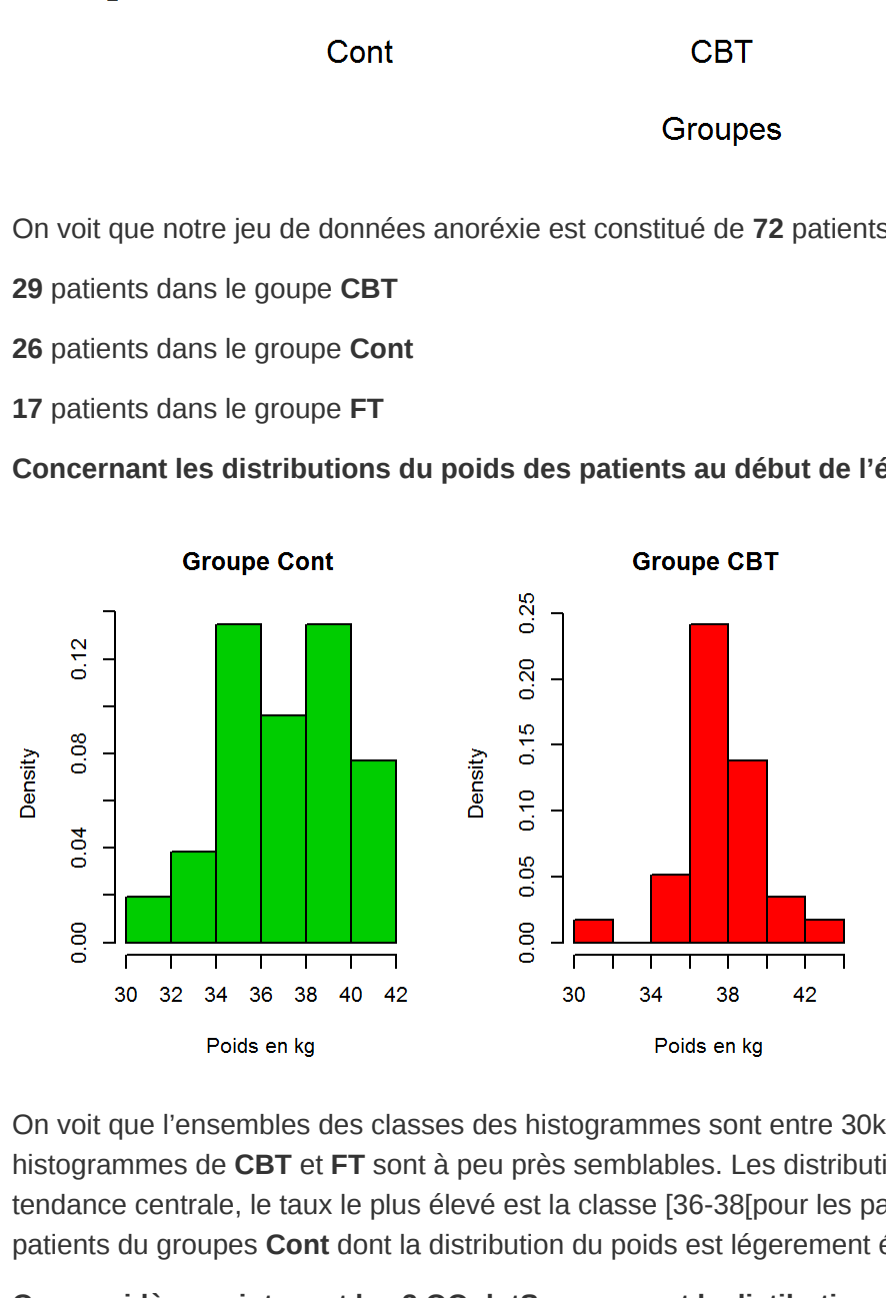
### Analyse des données

Ce tableau de données sur l'anorexie a été établi lors d'une étude clinique. L'objectif de l'étude est de comparer les trois types de traitements et leurs effets sur les patients. Quelle thérapie est la plus efficace ? Est-ce que certains traitements n'ont pas de résultat positif (voire ont des résultats négatifs) sur la santé des patients ?

Nous formulons ici quelques questions précises, auxquelles vous devez réfléchir et chercher des réponses en utilisant des outils de la statistique descriptive, notamment des représentations graphiques et des indicateurs statistiques. C'est à vous de choisir les bons outils pour répondre aux cinq questions suivantes.

### Question 1

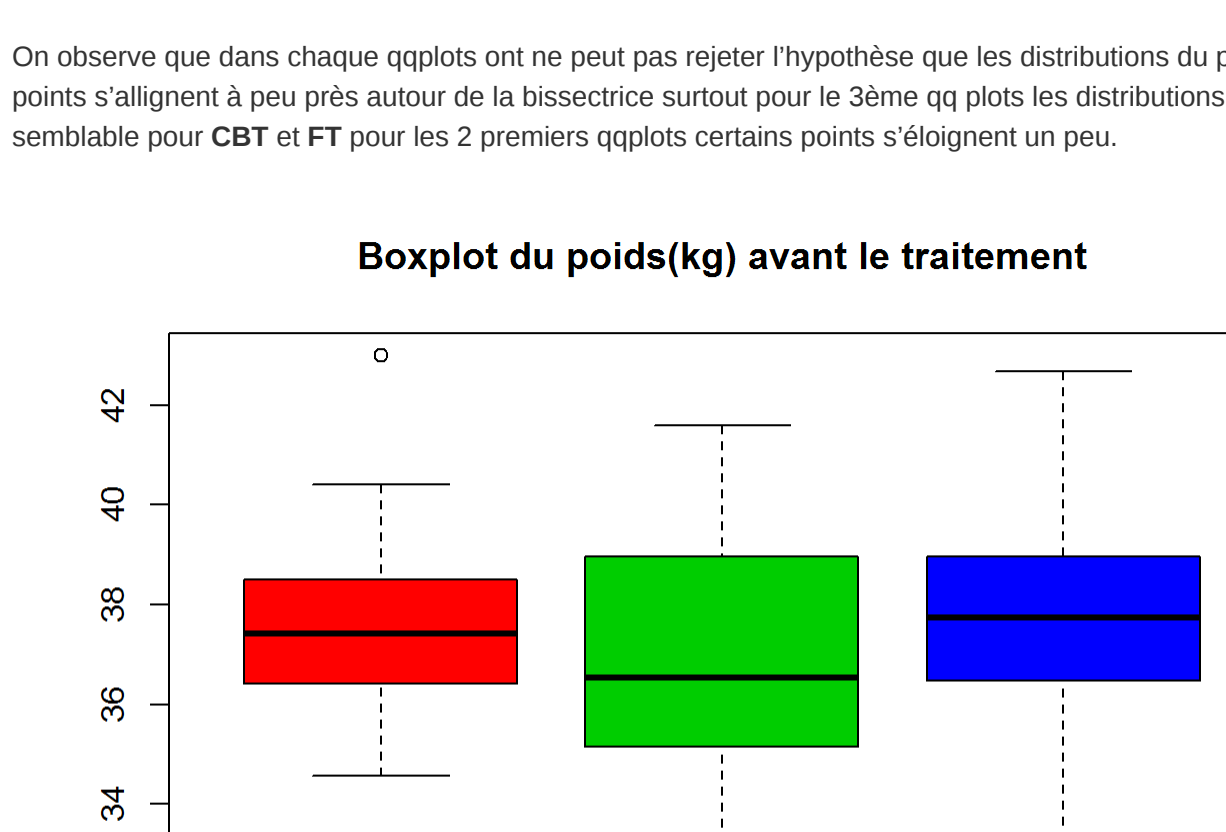
Tout d'abord, on veut savoir si l'étude a été correctement menée. En particulier, on veut vérifier si la composition des trois groupes de patients qui suivent des traitements différents a été bien faite. Plus précisément, être en mesure de comparer les résultats, il faut qu'au début de l'étude les trois groupes soient similaires. Comparez les nombres de patients par groupe, et comparez les distributions du poids des patients au début de l'étude.



On voit que notre jeu de données anorexie est constitué de 72 patients anorexiques dont:

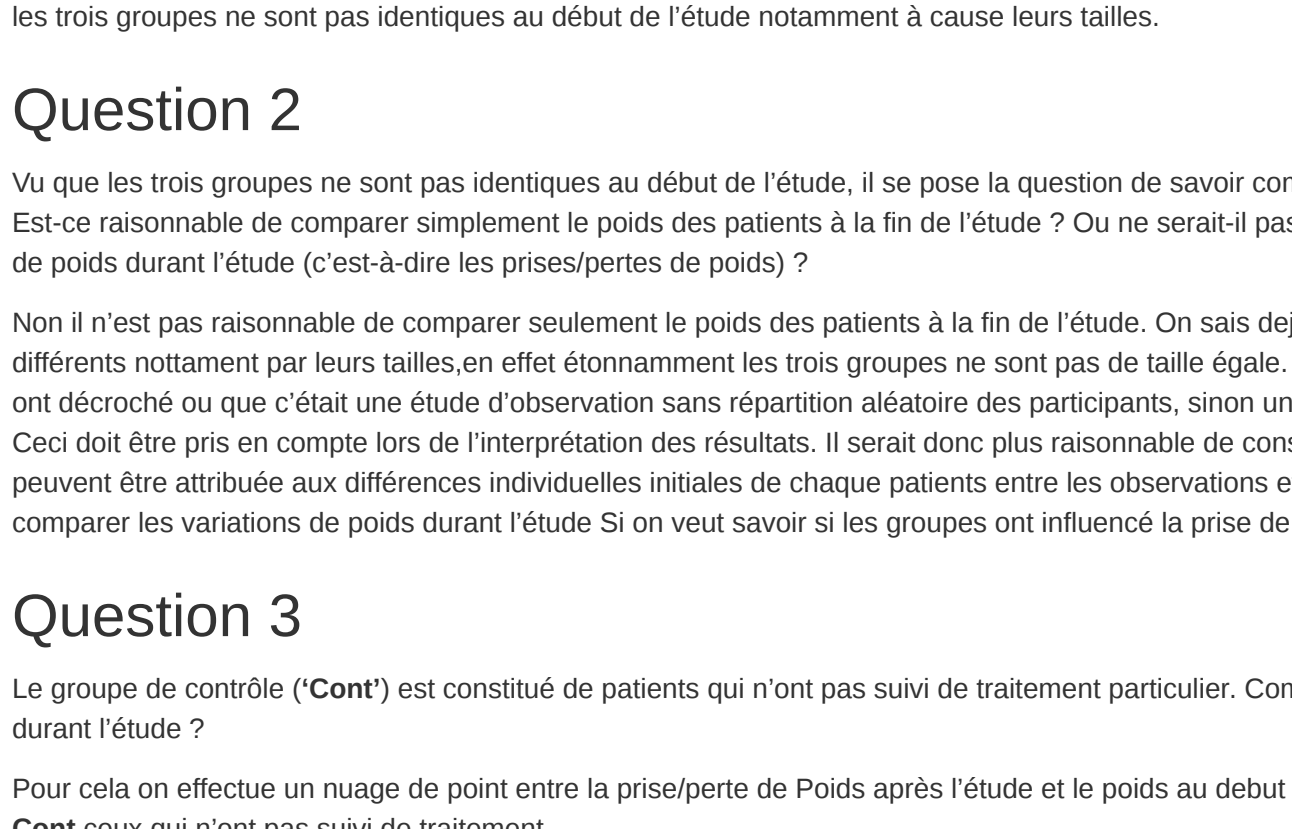
- 29 patients dans le groupe CBT
- 26 patients dans le groupe Cont
- 17 patients dans le groupe FT

Concernant les distributions du poids des patients au début de l'étude On observe l'histogrammes suivant:



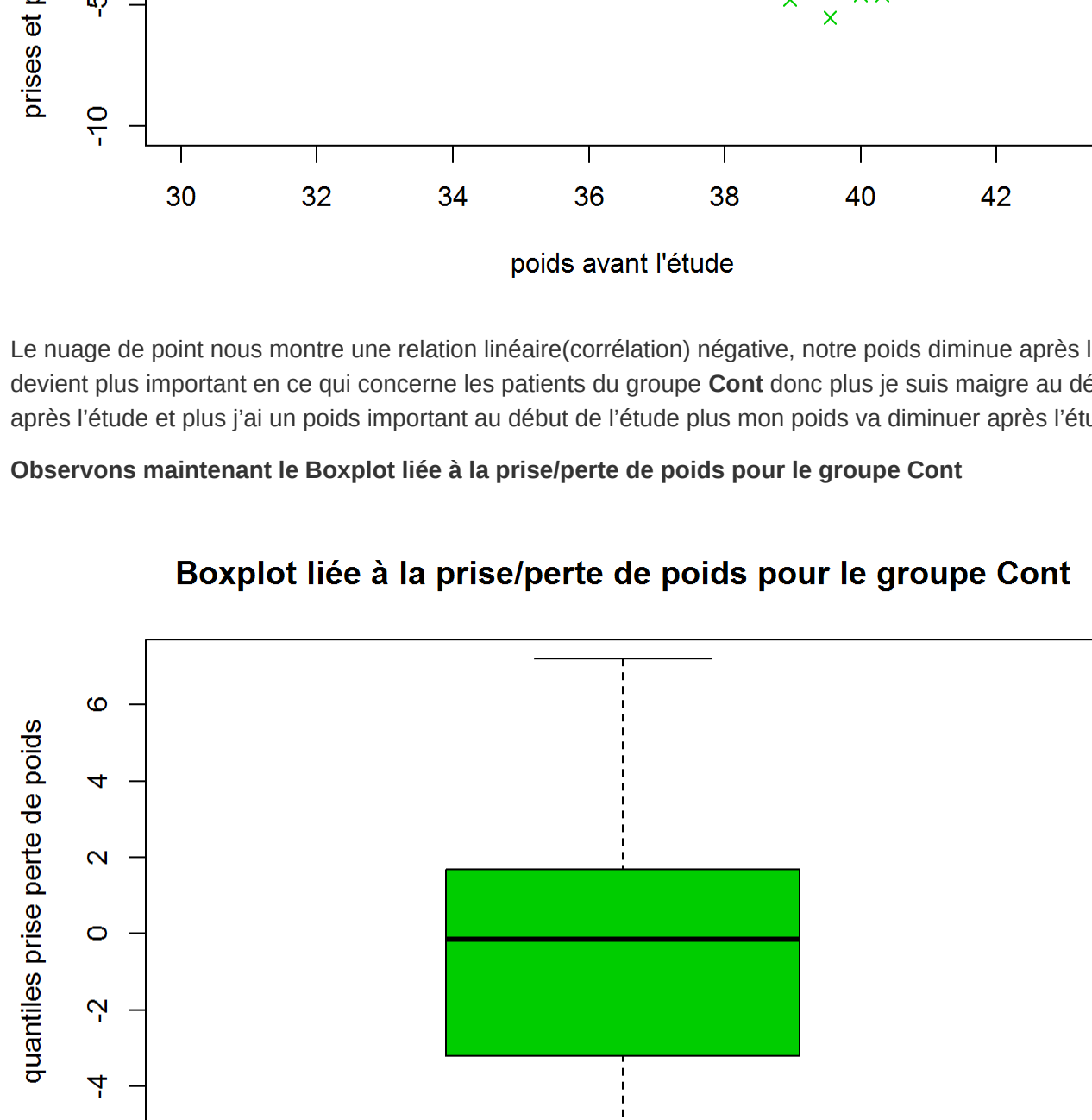
On voit que l'ensemble des classes des histogrammes sont entre 30kg et 44kg les patients sont belles et bien anorexiques, On remarque que les histogrammes de CBT et FT sont à peu près semblables. Les distributions du poids sont assez homogène pour les personnes des 3 groupes. En revanche, contrairement, le taux le plus élevé est la classe [35-38] pour les patients des groupes CBT et FT et les classes [34-36] [38-40] pour les patients du groupe Cont dont la distribution du poids est légèrement décalé vers la droite au delà de 34kg.

On considère maintenant les 3 QQplots comparant la distribution du poids avant l'étude dans chaque groupe



On observe que dans chaque qqplot on ne peut pas rejeter l'hypothèse que les distributions du poids avant l'étude semblent équilibrées les points s'alignent à peu près autour de la bissectrice surtout pour le 3ème qqplot les distributions du poids avant le traitement semble être semblable pour CBT et FT pour les 2 premiers qqplots certains points s'éloignent un peu.

### Boxplot du poids(kg) avant le traitement



Les boxplots de CBT et FT sont symétriques leurs distribution pourraient être semblables de plus ils ont à peu près la même médiane néanmoins, on voit que les poids avant l'étude dans chaque groupe ont des ordres de grandeurs différentes concernant le poids avant la thérapie celui de CBT est particulièrement plus petit que les autres.

les trois groupes ne sont pas identiques au début de l'étude notamment à cause leurs tailles.

### Question 2

Vu que les trois groupes ne sont pas identiques au début de l'étude, il se pose la question de savoir comment évaluer le succès d'un traitement. Est-ce raisonnable de comparer simplement le poids des patients à la fin de l'étude ? Ou ne serait-il pas plus pertinent de comparer les variations de poids durant l'étude (c'est-à-dire les prises/pertes de poids) ?

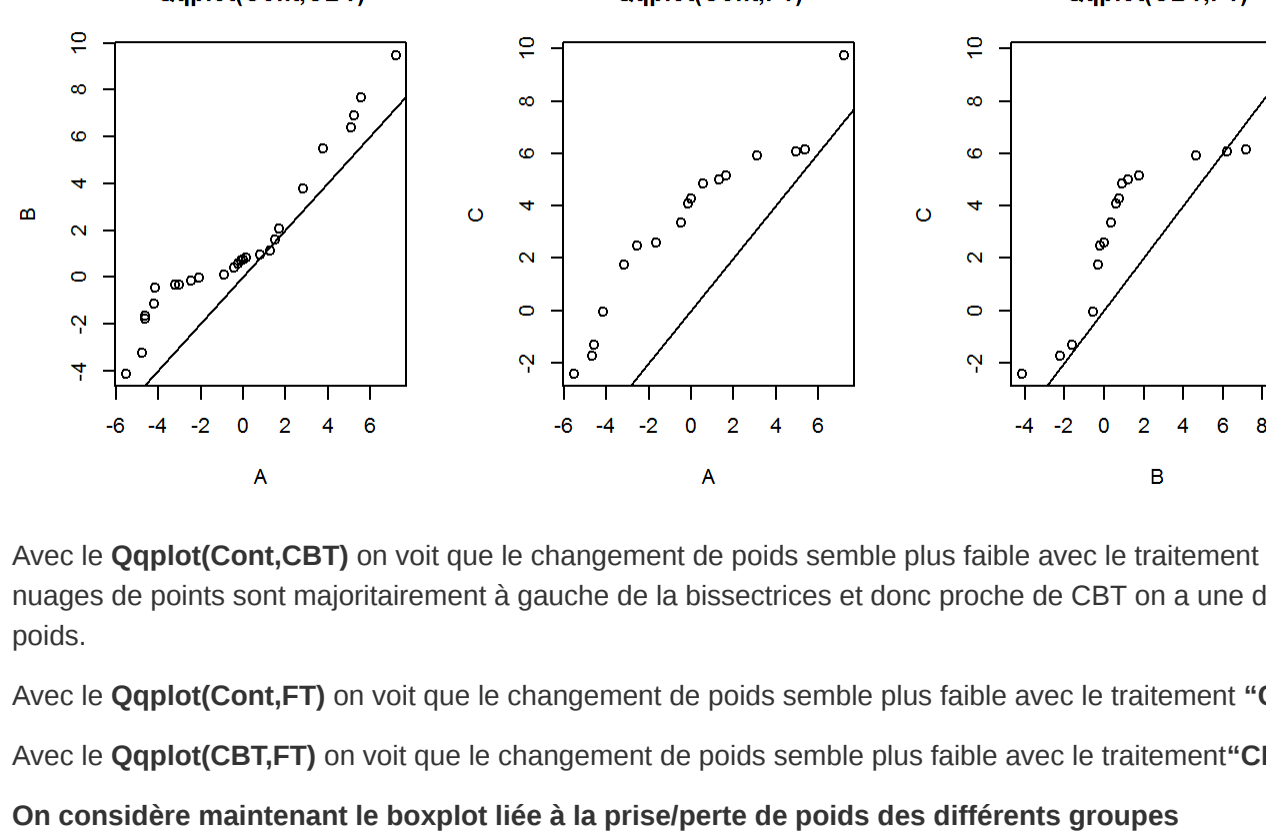
Non il n'est pas raisonnable de comparer seulement le poids des patients à la fin de l'étude. On sait déjà qu'au début de l'étude les 3 groupes sont différents notamment par leurs tailles en effet, disons que les patients du groupe Cont ont plus de poids que ceux des groupes CBT et FT. Cela peut signifier que certains patients ont décroché ou que c'était une étude d'observation sans répartition aléatoire des participants, sinon une autre explication pourrait être donnée. Ceci doit être pris en compte lors de l'interprétation des résultats. Il serait donc plus raisonnable de considérer que les résultats des trois groupes peuvent être attribués aux différences individuelles initiales de chaque patient entre les observations et ainsi il serait donc plus pertinent de comparer les variations de poids durant l'étude. Si on veut savoir si les groupes ont influencé la prise de poids de nos patients.

### Question 3

Le groupe de contrôle ("Cont") est constitué de patients qui n'ont pas suivi de traitement particulier. Comment a évolué le poids de ses patients durant l'étude ?

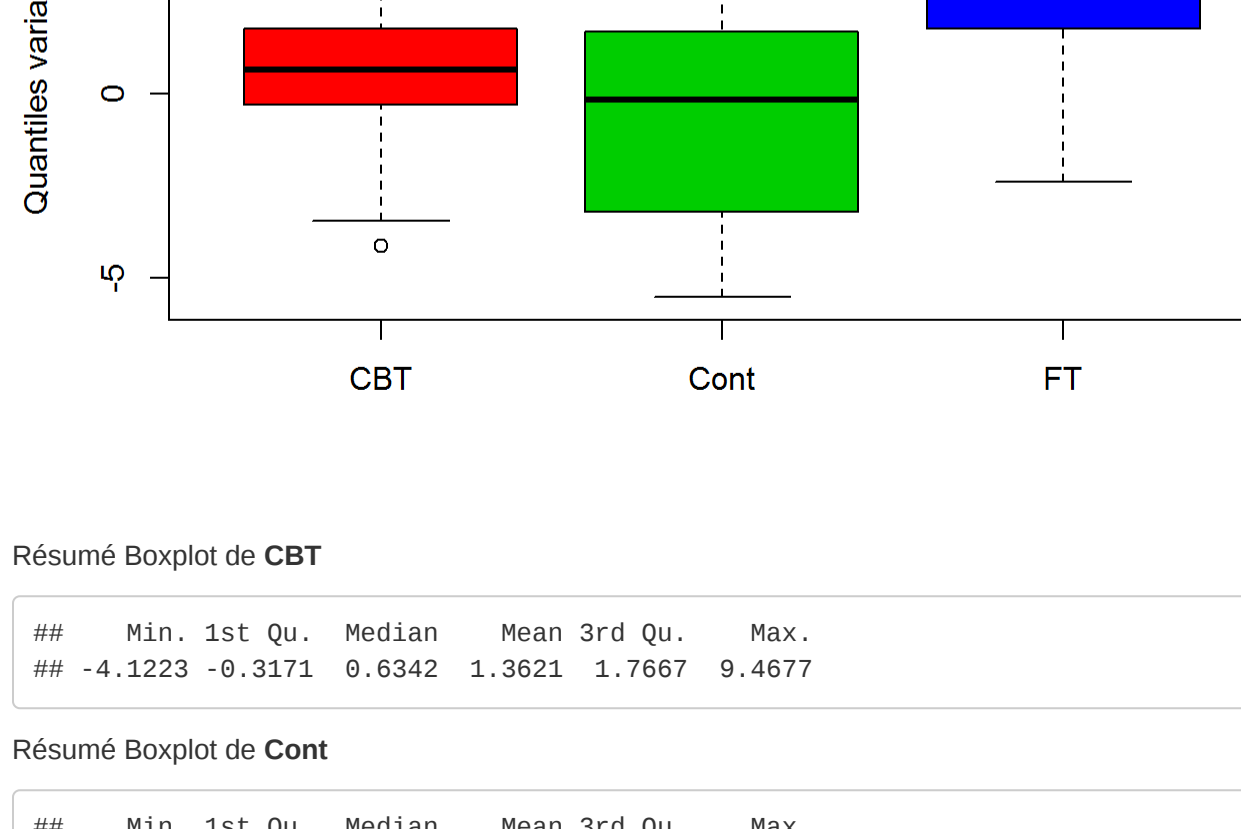
Pour cela on effectue un nuage de point entre la prise/perte de Poids après l'étude et le poids au début de l'étude pour les patients du groupe Cont ceux qui n'ont pas suivi de traitement

### Evolution du poids des patients du groupe Cont



Le nuage de point nous montre une relation linéaire (corrélation) négative, notre poids diminue après l'étude lorsque notre poids avant l'étude devient plus important ce qui concerne les patients du groupe Cont donc plus le poids au début de l'étude plus mon poids va augmenter après l'étude et plus le poids important au début de l'étude plus mon poids va diminuer après l'étude

Observons maintenant le Boxplot liée à la prise/perde de poids pour le groupe Cont



Résumé Boxplot de Cont

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	-5.5266	-3.3718	-0.1585	-0.2039	1.6306	7.2827

On observe que la moyenne est de -0.2039 donc durant l'étude il n'y a pas d'augmentation significative du poids pour les patients du groupe Cont, au contraire, nous avons donc en moyenne une perte de poids qui n'est pas importante pour les patients du groupe Cont durant l'étude.

### Question 4

Étudier l'impact des différents traitements sur le poids des patients. Comparer les trois méthodes. Quel traitement semble le meilleur ?

On essaye maintenant de tester si le changement de poids avant/après traitement diffère selon le traitement reçu. Plus précisément, on veut tester la différence entre les 3 traitements Cont, CBT et FT.

Pour étudier l'impact des différents traitements sur le poids des patients observons le nombre de patients par groupes qui ont pris et perdu du poids

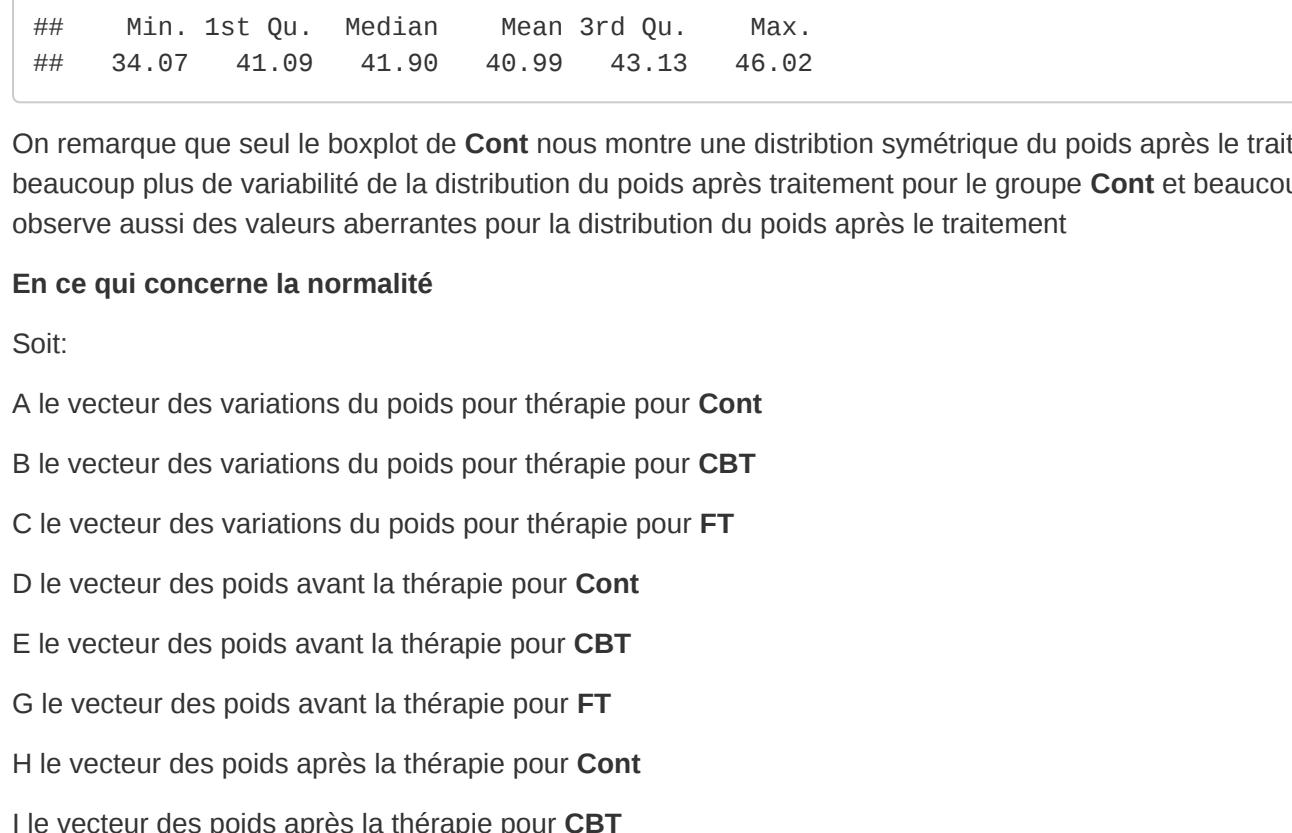


On voit qu'il y a:

- 18 personnes sur 29 qui ont pris du poids dans le groupe CBT soit 62% des personnes de ce groupe ont pris du poids et donc (11 personnes) 38% en ont perdu
- 11 personnes sur 26 qui ont pris du poids dans le groupe Cont soit 42% des personnes de ce groupe ont pris du poids et donc 54% (14 personnes) qui en ont perdu et 1 personnes sur 26 donc 4% qui a vu son poids inchangé
- 13 personnes sur 17 qui ont pris du poids dans le groupe FT soit 76% des personnes de ce groupe ont pris du poids et donc (4 personnes) 24% en ont perdu

FT a tendance à être le traitement le plus efficace

Pour comparer les 3 méthodes utilisons la fonction qqplot deux à deux pour A,B et C les changements de poids avant/après traitement pour les individus sous traitement Cont, CBT et FT respectivement

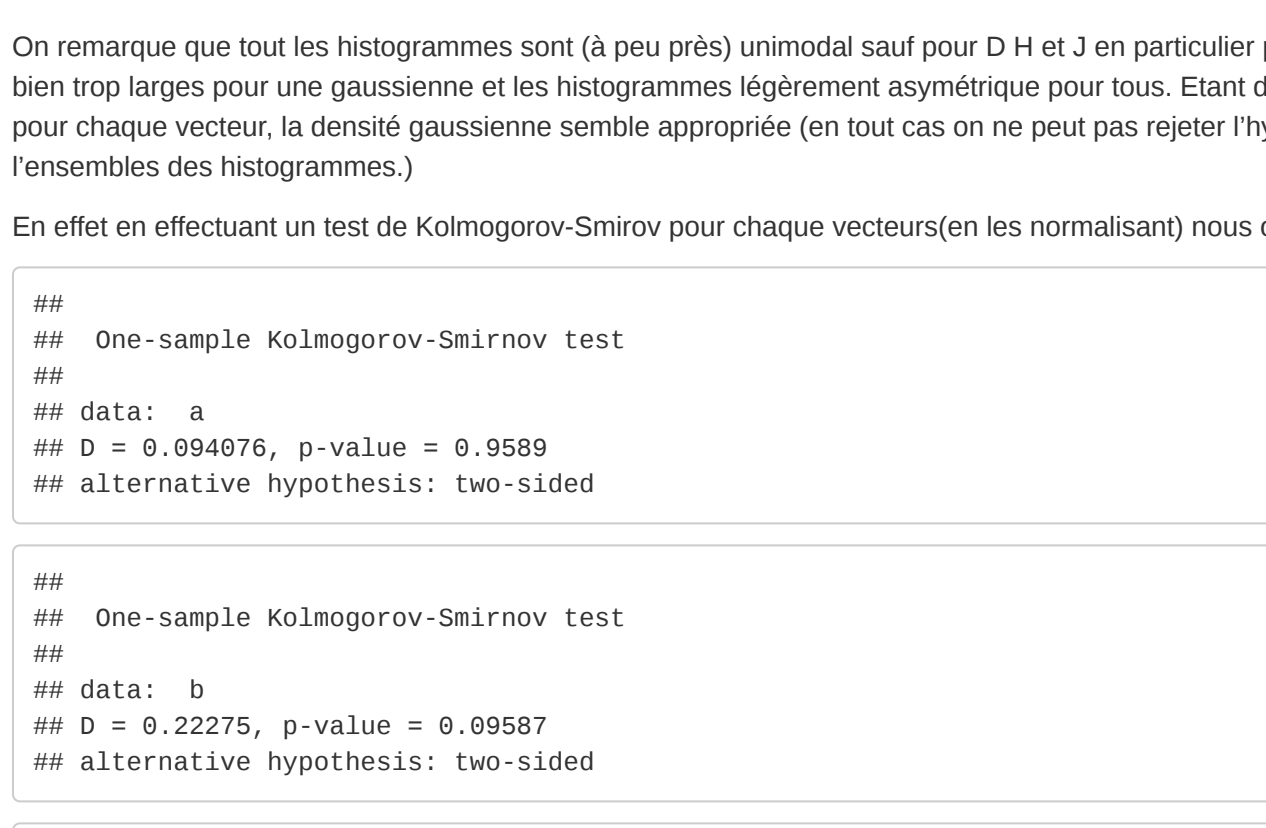


Avec le qqplot(Cont,CBT) on voit que le changement de poids semble plus faible avec le traitement "Cont" qu'avec le traitement "CBT" car les nuages de points sont majoritairement à gauche de la bissectrice et donc proche de CBT on a une distribution plus forte du changement de poids.

Avec le qqplot(Cont,FT) on voit que le changement de poids semble plus faible avec le traitement "Cont" qu'avec le traitement "FT".

Avec le qqplot(CBT,FT) on voit que le changement de poids semble plus faible avec le traitement "CBT" qu'avec le traitement "FT".

On considère maintenant le boxplot liée à la prise/perde de poids des différents groupes



Résumé Boxplot de CBT

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	-4.1223	-0.3371	0.6342	1.3023	1.7607	9.4877

Résumé Boxplot de Cont

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	-5.5266	-3.3718	-0.1585	-0.2039	1.6306	7.2827

Résumé Boxplot de FT

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	-2.401	1.767	4.077	3.291	5.164	9.739

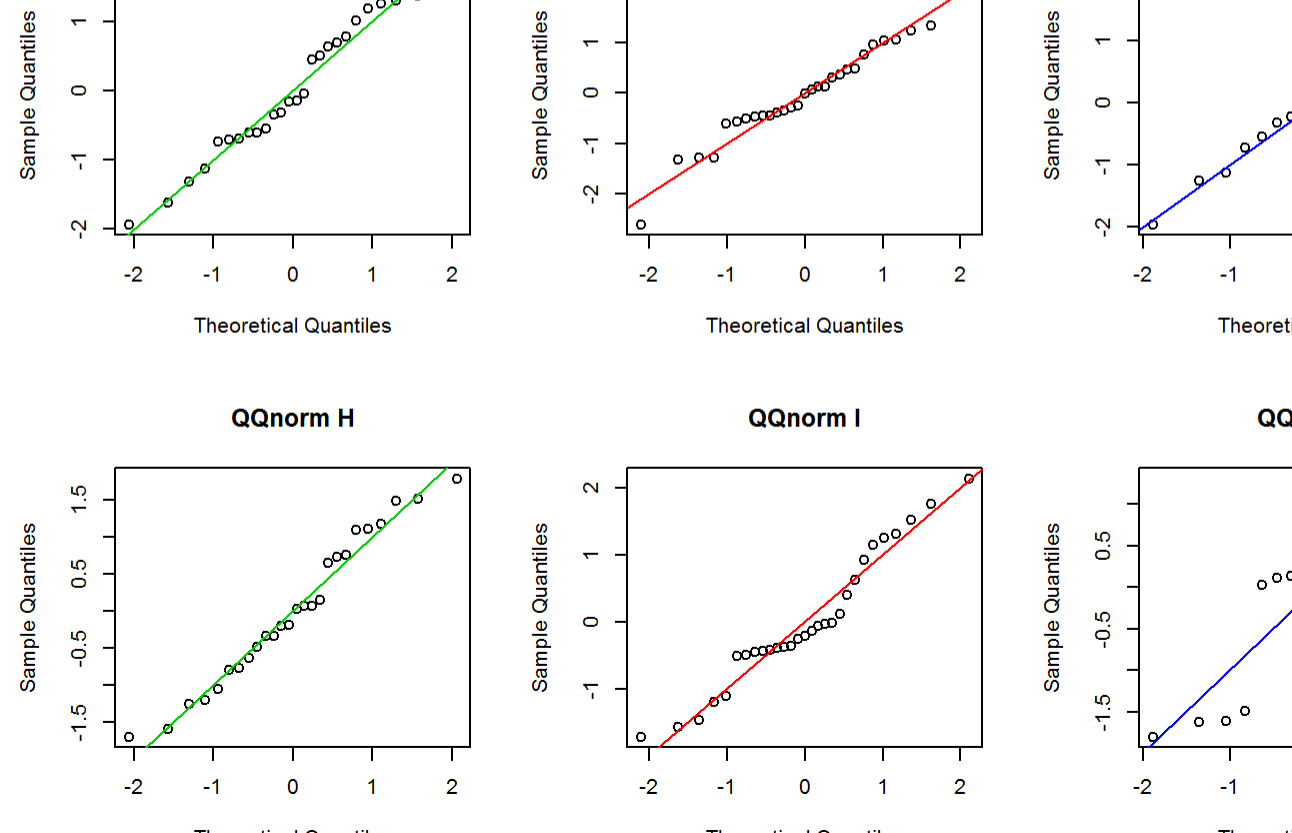
Sur les données du boxplot liée à la prise/perde de poids des patients on observe des moyennes de 1.3023, -0.2039 et 3.291 pour les groupes CBT, Cont et FT respectivement, donc en moyenne les patients du groupe FT gagnent plus de poids que ceux des groupes Cont et CBT, de plus en moyenne ceux du groupe Cont perdent du poids. Les boxplots des traitements FT et CBT ne se chevauchent pas. Par conséquent, nous pouvons conclure qu'il existe une différence statistiquement significative entre ces deux groupes de traitement sans même effectuer de test de signification.

On en déduit de tout cela que la prise de poids semble plus forte avec le traitement "FT" on a même une différence significative par rapport aux 2 autres traitements, il semble être le traitement le plus efficace pour la prise de poids des patients anorexiques dans cette étude

### Question 5

Analysar la distribution des différentes variables liées aux poids par type de traitement. Est-ce envisageable de les modéliser par une loi normale ?

Considérons ces 3 Boxplots regroupant chacun des groupes pour analyser la distribution des différentes variables liées aux poids par type de traitement



Résumé Boxplot de la variation du poids

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	-4.1223	-0.3371	0.6342	1.3023	1.7607	9.4877

Résumé Boxplot de CBT

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	-5.5266	-3.3718	-0.1585	-0.2039	1.6306	7.2827

Résumé Boxplot de FT

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	-2.401	1.767	4.077	3.291	5.164	9.739

On remarque que le boxplot de CBT est presque symétrique la distribution de la variation du poids est à peu près symétrique la médiane et la moyenne sont légèrement différentes à égale distance des quartiles 1 et 3. Les 3 boxplots n'ont pas les mêmes médiane celle du groupe Cont est même inférieure à zéro, on a quand même une différence significative entre la distribution de la variation du poids dans le groupe FT et celles observées dans les autres pour les traitements Cont et FT que pour le traitement CBT groupe auquel on observe aussi des valeurs aberrantes (petits points isolés) dans la distribution de la variation du poids.

Boxplot poids avant traitement

Résumé Boxplot de CBT

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	31.71	36.42	37.42	37.46	38.51	42.99

Résumé Boxplot de Cont

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	31.04	35.21	36.53	36.95	38.80	41.50

Résumé Boxplot de "FT"

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	33.25	36.47	37.73	37.70	38.96	42.67

On remarque avec les boxplots de Cont et FT que les distributions du poids avant traitement sont symétriques pour tous les 2 moyenne(Mean) et médiane sont quasiment identiques et un peu différentes pour Cont mais quand même proche on a aussi beaucoup plus de variabilité de la distribution poids avant pour les traitements Cont et FT que pour le traitement CBT, on observe aussi des valeurs aberrantes dans la distribution du poids avant pour le groupe CBT

Boxplot poids après traitement

Résumé Boxplot de CBT

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	32.38	37.18	38.81	38.82	41.18	46.93

Résumé Boxplot de Cont

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	33.07	35.14	36.96	36.74	38.35	49.59

Résumé Boxplot de FT

##	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
##	34.07	42.09	42.90	40.99	43.13	46.92

On remarque que seul le boxplot de Cont nous montre une distribution symétrique du poids après le traitement pas ceux de FT et CBT, on a beaucoup plus de variabilité de la distribution du poids après traitement pour le groupe Cont et beaucoup moins pour le groupe FT et CBT, on observe aussi des valeurs aberrantes pour la distribution du poids après le traitement

En ce qui concerne la normalité

Soit:

A le vecteur des variations du poids pour thérapie pour Cont

B le vecteur des variations du poids pour thérapie pour CBT

C le vecteur des variations du poids pour thérapie pour FT

D le vecteur des poids avant la thérapie pour Cont

E le vecteur des poids avant la thérapie pour CBT

F le vecteur des poids avant la thérapie pour FT

H le vecteur des poids après la thérapie pour Cont

I le vecteur des poids après la thérapie pour CBT

J le vecteur des poids après la thérapie pour FT

Voici les histogrammes de ces vecteurs superposé à la densité de la loi théorique:



On voit que tous les vecteurs semblent avoir une loi normale (car le point semble s'aligner sur la bissectrice dans les graphiques) sauf pour le vecteur J pour les vecteurs B et I et E ou groupes CBT certains points s'éloignent de la bissectrice sans pas au point de rejeter l'hypothèse gaussienne

Donc l'ensemble des variables liées aux poids par type de traitement peuvent être modéliser par une loi normales sauf pour le vecteur du poids après traitement pour le groupe FT

On aurait pu regarder les fonctions de répartition ou l'ont voit que l'adaptation entre la fonction de répartition empirique et la fonction de répartition théorique est très bonne pour l'ensemble des vecteurs (y compris E un peu moins bonne pour B et I - 135mais correcte) sauf pour J.



Les p-values sont toutes supérieures à 0.05 donc on ne rejette pas l'hypothèse de normalité au seuil de 5% pour aucun vecteur ni à celui de 10% mais de jute pour le vecteur J

Observons maintenant les qqnorm pour avoir plus de certitudes



On voit que tous les qqplots ont une loi normale (car le point semble s'aligner sur la bissectrice dans les graphiques) sauf pour le vecteur J pour les vecteurs B et I et E ou groupes CBT certains points s'éloignent de la bissectrice sans pas au point de rejeter l'hypothèse gaussienne

Donc l'ensemble des variables liées aux poids par type de traitement peuvent être modéliser par une loi normales sauf pour le vecteur du poids après traitement pour le groupe FT

On aurait pu regarder les fonctions de répartition ou l'ont voit que l'adaptation entre la fonction de répartition empirique et la fonction de répartition théorique est très bonne pour l'ensemble des vecteurs (y compris E un peu moins bonne pour B et I - 135mais correcte) sauf pour J.



### Question 6

Donner votre conclusion.

Au vu de l'analyse des données sur l'anorexie lors de cette étude clinique nous avons constaté que les 3 groupes n'étaient pas identiques au début notamment par leurs tailles on aurait pu tester les 3 traitements sur un groupe de même personnes par exemples à différentes périodes même pendant plusieurs années si les distributions ne sont pas bien équilibré le résultat final n'est pas très concret. Toutefois durant l'étude nous avons pu constater qu'en moyenne 2 traitements étaient efficaces FT et CBT mais même entre ces 2 traitements il y avait des différences significatives sur les distributions des variables liées au poids surtout après le traitement. FT était donc le plus efficace et bien plus que les deux autres pour la prise de poids de ses patients anorexiques en effet le traitement Cont qui n'est pas vraiment un traitement apporte des résultats variés sur la prise de poids, en moyenne il n'est pas efficace sur ses patients qui sont 54% à avoir perdu du poids ce qui paraît beaucoup trop important nous avons aussi pu observer des différences significatives de gain de poids entre les 3 traitements, mais il y a aussi beaucoup de variations individuelles comme on peut le voir sur les graphiques ci-dessous qui nous montre un lien entre le poids au début de l'étude et la prise/perte de poids en générale, il semblerait que les patients plus lourds, toujours sous un poids normal, perdent moins de poids que les patients plus légers pour chaque traitement dans tous les cas les 3 traitements n'auront pas d'impact négatifs sur la santé d'un patient très léger de poids inférieur à 34 kg par exemple mais pour le groupe Cont on a un impact plutôt négatif pour les patients plus lourds sur le premier graphique on voit qu'ils perdent du poids au début de l'étude (contrairement à ceux des groupes Cont et FT qui eux en gagnent), ce qui est très négatif pour leurs santé sachant qu'ils sont déjà anorexiques.

C'est le traitement le plus réussi pour les patients anorexiques plus lourds

