

Contrôle continu L2 Pharmacie TD Statistique

BETTEGA Francois

11 septembre, 2020

Introduction

Le 17 mars 2020 *Hydroxychloroquine and azithromycin as a treatment of COVID-19: results of an open label non-randomized clinical trial* était en ligne, l'article marseillais d'évaluation de l'efficacité de l'hydroxychloroquine et azithromycine dans le traitement du COVID-19. Je vous propose dans ce contrôle continu de reproduire une partie des traitements statistiques présentés dans cet article.

Pour des questions de simplicité, les jeux de données fournis sont nettoyés. Si vous êtes intéressés par les transformations effectuées sur les jeux de données, vous pourrez retrouver le code R à l'adresse suivante https://github.com/fbettega/TD_L2_pharmacie/ dans [ce fichier](#).

Vous trouverez un fichier contenant les données pour chacune des questions, les fichiers étant nommés à partir du numéro de la question.

Vous pouvez me contacter à l'adresse suivante pour toutes questions : francois.bettega+L2_pharma@gmail.com

“Je n’ai pas réussi à ouvrir les jeux de données” ne sera pas une excuse acceptable en l’absence de sollicitation.

Modalités d’évaluation

Ce devoir sera à rendre pour le 30 septembre à l'adresse mail suivante : francois.bettega+L2_pharma@gmail.com.

Vous serez évalués sur :

- le choix justifié des tests statistiques (40%) ainsi que sur les hypothèses qui doivent être vérifiées avant de pratiquer le test
- votre rédaction et vos interprétations des résultats de ces tests (50%)
- les résultats numériques obtenus (10%).

Je vous encourage à coopérer et échanger entre les groupes tout en vous prévenant qu’une attention particulière sera portée à la détection du plagiat y compris dans votre code R.

La forme du compte rendu attendue est constituée de deux fichiers : le rapport au format compilé (knit) (HTML,PDF,Word...) et le fichier Rmarkdown (.Rmd). Le rapport contiendra pour chaque question : - le test effectué et sa motivation - Le résultat du test, ainsi que votre interprétation.

En cas de problème avec la création de votre rapport à partir du Rmarkdown, ne paniquez pas. Pour l'évaluation, il faut juste un fichier PDF¹/Word/Html contenant votre rapport ainsi que votre code au format R ou Rmd. Par conséquent les fichiers Rmd qui ne compilent pas ainsi que rapport + code sont recevables.

Attention les P-values et tests absents du code R mais commentés ou les hypothèses non vérifiées dans le code R conduiront à une perte de 0.5 point par occurrence

¹Pour rappel vous pouvez facilement créer un document pdf à partir de word ou google doc en utilisant *enregistrer sous*.

Résumé des documents à rendre

- Un fichier Rmd + votre rapport knit à partir du fichier (format PDF/word/html) (0.5 point bonus)
- Un fichier R ou Rmd + votre rapport au format PDF/Word

TD noté sur 10 points (ramené sur 20) 6 points pour les questions + 1 point si vos code R tournent sans erreur + 1 point sur la qualité générale du rapport + 2 points sur l'intégration de graphiques pour visualiser les données (0.5 pt par graphique pertinent) **en cas d'abus, par exemple si vous produisez un grand nombres de graphiques par question sans réfléchir à leur pertinence, des points pourront être retirés.**

A cela s'ajoute un point bonus sur les interprétations biologiques et pharmaceutiques un points bonus sur une représentation graphique non présentée en TD 0.5 point bonus sur le fait de commenter votre code R en utilisant : # suivi de texte pour expliquer ce que vous essayez de faire 0.5 point pour avoir knit votre rapport depuis votre Rmd

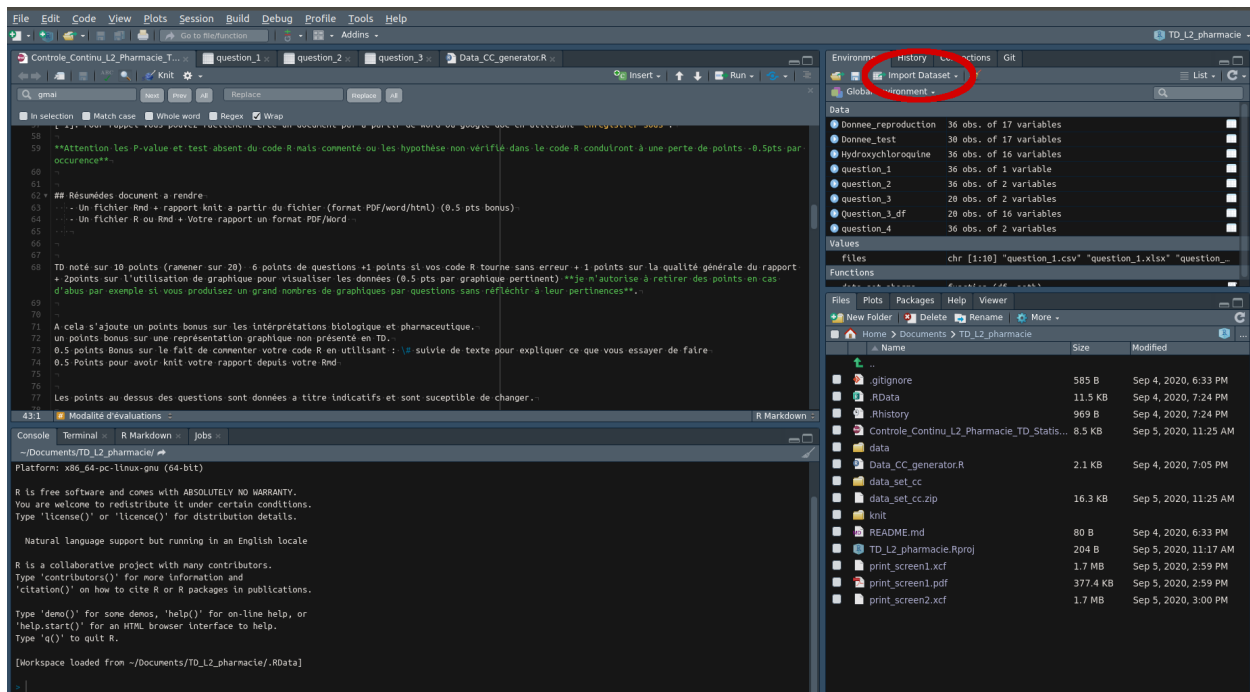
Les points mentionnés en regard des questions sont donnés à titre indicatif et sont susceptibles de changer.

Aide à l'ouverture des fichiers de données

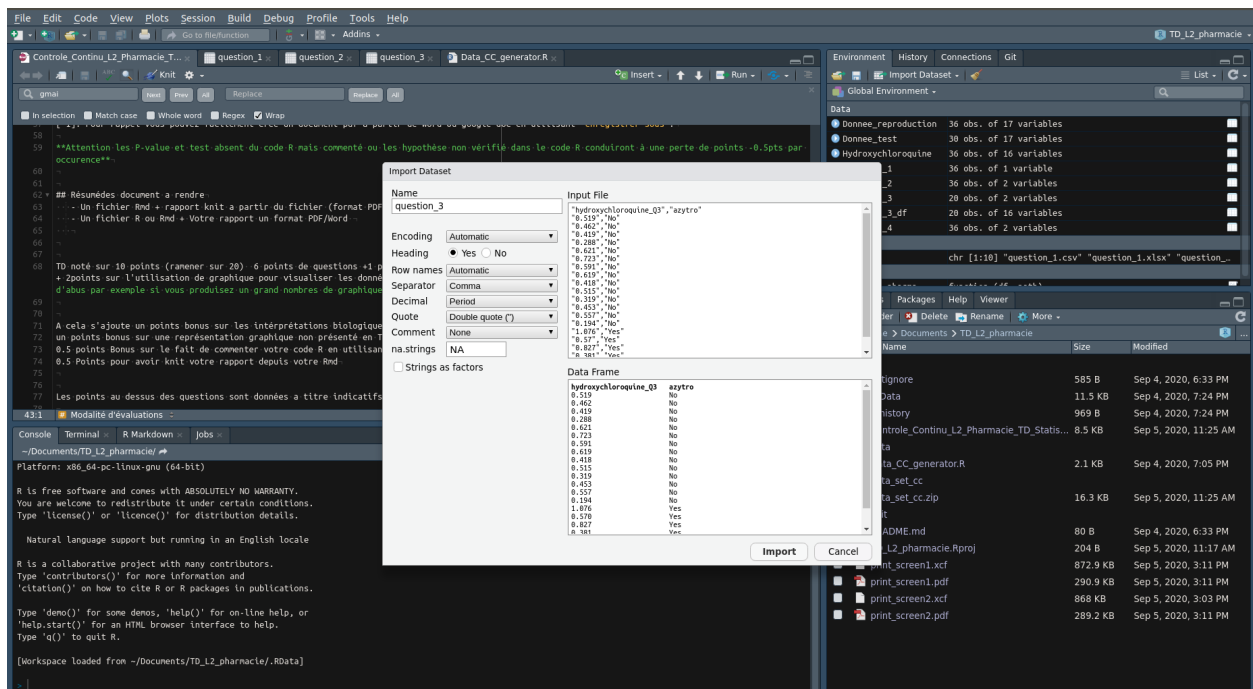
Vous avez un jeu de données par question il s'appelle **question_x** et il est fourni au format xlsx et csv.

Si vous mettez les fichiers dans le même dossier que votre script Rmd vous pouvez utiliser la ligne de code présentée au début de chaque question pour ouvrir le fichier approprié.

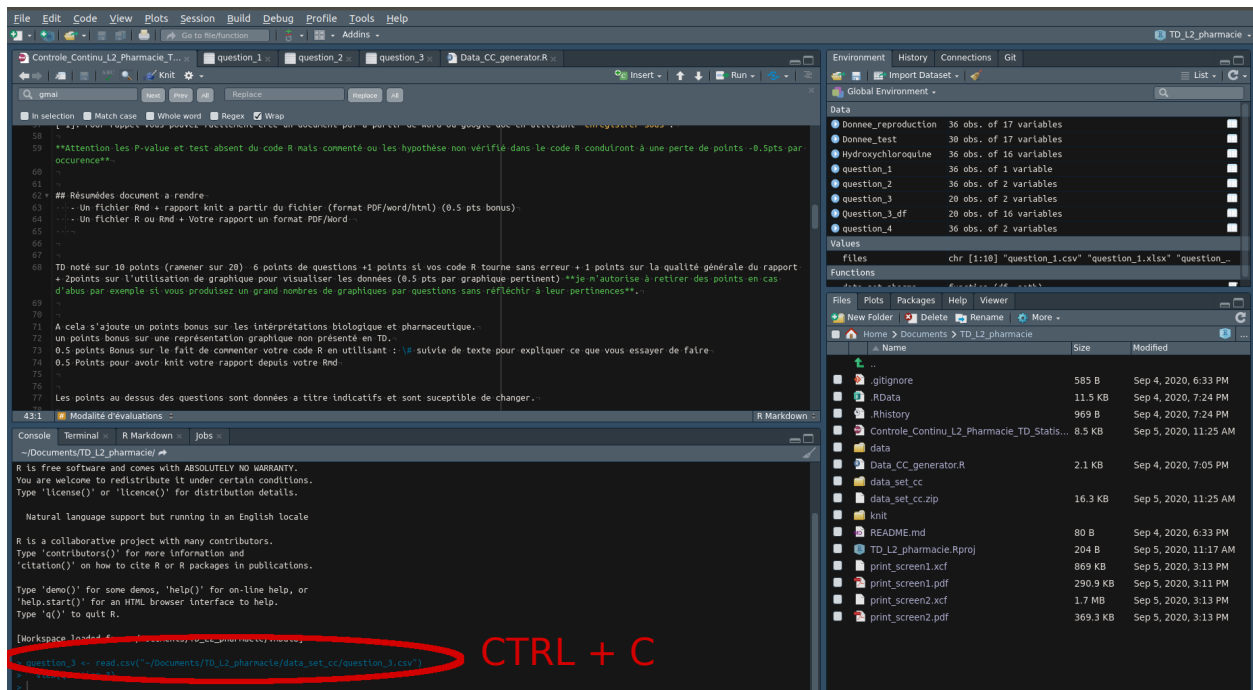
Sinon vous pouvez utiliser le bouton en haut à droite de l'interface de R studio



Sélectionner les paramètres appropriés



Ensuite faite un copier coller de la seconde ligne de commande dans votre terminal R



Question 1 (1 pt)

```
question_1 <- read.csv("data_set_cc/question_1.txt")
```

Ce jeu de données contient l'âge de tous les patients de l'étude mentionnée plus haut.

On vous affirme que l'âge des patients n'est pas représentatif de l'âge moyen de la population française qui

est de 41.2 ans. Testez cette affirmation.

Question 2 (1 pt)

```
question_2 <- read.csv("data_set_cc/question_2.txt")
```

Ce jeu de données contient : - l'âge des patients - la présence ou l'absence de traitement par hydroxychloroquine
L'âge des patients est-il différent dans les 2 groupes selon la prise ou non du traitement par hydroxychloroquine ?

Pour sélectionner les patients par groupe vous pouvez utiliser le code ci-dessous.

```
age_tt <- question_2$age[question_2$hydroxychloroquine == "Yes"]
```

```
age_non_tt <- question_2$age[question_2$hydroxychloroquine == "No"]
```

Question 3 (1 pt)

```
question_3 <- read.csv("data_set_cc/question_3.txt")
```

Ce jeu de données contient : - hydroxychloroquine_Q3 : Concentration d'Hydroxychloroquine dans le serum des patients en µg/ml - azytro: traitement par azithromycine oui ou non

Pour sélectionner les patients par groupe vous pouvez utiliser le code ci-dessous

```
serum_hydroxy_azitro <- question_3$hydroxychloroquine_Q3[question_3$azytro == "Yes"]
```

```
serum_hydroxy_non_azitro <- question_3$hydroxychloroquine_Q3[question_3$azytro == "No"]
```

On vous affirme que la moyenne de la concentration d'hydroxychloroquine dans le serum des patients traités par azithromycine est différente de la moyenne de la concentration d'hydroxychloroquine dans le serum des patients non traités par azithromycine. Testez cette affirmation en utilisant **un test non paramétrique**.

Question 4 (4 pts)

L'article tente d'évaluer si les patients traités par hydroxychloroquine ont des résultats de PCR au jour 6 différents des patients non traités. Pour les questions suivantes vous devez reproduire cette analyse selon les deux manières suivantes selon la méthode d'imputation des patients qui n'ont pas eu de PCR au jour 6.

La première manière (question 4.1), comme dans l'article, considère que les patients traités ont une PCR négative et les patients non traités ont une PCR positive

La seconde manière (question 4.2) considère que les patients n'ayant pas reçu de PCR au jour 6 sont simplement écartés.

Les variables dans les deux jeux de données sont : PCR_1 ou PCR_2 : La PCR au 6eme jour est elle oui ou non positive

hydroxychloroquine_1 ou hydroxychloroquine_2 : Le patient est il oui ou non traité par hydroxychloroquine.

Question 4.1 (1.25 pts)

Le code fourni plus haut doit être adapté pour charger les jeux de données et sélectionner les patients par groupe Avec le jeu de données question_4_1 tester l'hypothèse suivante : les patients traités par hydroxychloroquine ont-ils des résultats de PCR au jour 6 différents des patients non traité.

Question 4.2 (1.25 pts)

Avec le jeu de données question_4_2 tester l'hypothèse suivante : les patients traités par hydroxychloroquine ont-ils des résultats de PCR au jour 6 différents des patients non traité.

Normalement vous pouvez utiliser exactement le même code pour la question 4_1 et 4_2

Commenter et comparer ces résultats (0.5 pt)

Conclusion (1 pt bonus)

Discuter ces résultats dans un contexte pharmaceutique

Donner votre avis sur ces méthodes d'imputation et leurs conséquences potentielles

Surprenez moi !!