

Analyse complète d'un problème sous R

Thème de l'étude :

Projet d'enfants chez des patients atteints de cancer à 2 ans du diagnostic
Facteurs associés au projet d'enfant chez des femmes de moins de 45 ans

Fichier fourni :

...\data44.xls

Descriptif des variables :

n_ident	= numéro identifiant
cancer	= localisation cancéreuse
age	= âge
sexe	= H, F
enfant	= Avez-vous des enfants ?
n_enfants	= Combien au total ? (<i>y compris les enfants décédés</i>)
menopause	= Etes-vous ménopausée ?
sterilite	= Etes-vous stérile ?
fertilite	= Fertilité (oui/non) : variable reconstruite
projet	= Avez-vous le projet d'avoir un enfant ou un autre enfant ?
PCS	= Score de qualité de vie Physique (échelle SF36, plus haut=meilleure QDV)
MCS	= Score de qualité de vie Mentale (échelle SF36, plus haut=meilleure QDV)

////////////////////
QUESTIONS
////////////////////

1) Fichier Excel

Importez sur votre disque dur le fichier « data44.xls » (qui est mis à votre disposition à partir du lien « Fichiers de données » dans la partie « Nota Bene / Boîte à outils »).

- Quelle est la nature des variables étudiées ?
- Que remarquez vous en survolant très rapidement le fichier .xls ?

2) Vérification des données dans R

- Ouvrez le logiciel R et importez le fichier « data44 ». En affichant le tableau de données dans R que remarquez vous en le survolant très rapidement ?
- Vérifiez que la sélection des patientes à étudier selon les critères prévus (âge < 45 ans, sexe féminin) a été convenablement réalisée, à l'aide de calcul et par méthode graphique ?
- Recherchez la ligne de l'individu trop âgé pour être inclus et affichez ses caractéristiques.
- Sélectionnez un nouveau sous-tableau sans homme et vérifiez que les variables âge et sexe respectent maintenant les critères fixés.
- Vérifiez la cohérence entre les variables enfant (Avez-vous des enfants ?) et n_enfants (nombres d'enfants).

3) Statistiques Descriptives dans R

- a) Calculez la proportion de nullipares (0 enfant) et de primipares (1 seul enfant) dans l'échantillon.
- b) Quel est le nombre moyen d'enfants dans l'échantillon ? Calculez la variance et proposez une représentation graphique.
- c) Quelle est la localisation cancéreuse la plus fréquente ? Proposez un graphique de l'ensemble des localisations.
- d) Combien de femmes expriment un projet de maternité future ? Combien ont répondu à cette question ?
- e) Parmi les femmes qui ont répondu à la question sur le projet de maternité, quelle proportion exprime un projet futur ?
- f) Combien y a-t-il de données manquantes pour la variable projet ("projet d'enfants") ? La variable fertilité permet-elle d'expliquer ces données manquantes ?

4) Analyse finale avec tests statistiques dans R¹

- a) Créez un nouveau sous-tableau des 175 femmes de moins de 45 ans ayant une fertilité préservée.

Normalité des distributions :

- b) L'âge est-il « normalement » distribué (test + illustration graphique) ?

Scores de qualité de vie physique et mentale (PCS/MCS) :

- c) Les 2 composantes de la qualité de vie (physique et mentale) sont-elles liées ?
- d) A 2 ans du diagnostic de cancer, ces femmes ont-elles une composante de qualité de vie plus altérée que l'autre (test apparié : « *paired=TRUE* ») ?
- e) La qualité de vie des patientes est-elle différente selon l'existence d'un projet futur d'enfant ? (2 séries d'analyse pour les 2 composantes physique et mentale de la QDV)

Maternité préalable et souhaitée :

- f) Existe-t-il un lien entre le fait d'avoir déjà des enfants et un projet futur de maternité ?
- g) Les femmes qui expriment un projet futur de maternité ont-elles en moyenne le même âge que celles qui ne souhaitent plus d'enfants ?
- h) Existe-t-il un lien entre l'âge des femmes le fait d'avoir déjà un(des) enfant(s) ?
- i) En étudiant séparément les femmes n'ayant pas d'enfant et chez celles qui en ont au moins un. Les femmes qui expriment un projet futur de maternité ont-elles en moyenne le même âge que celles qui ne souhaitent plus d'enfants ?

¹ Pour les différents tests de comparaison des moyennes, on utilisera des tests paramétriques car les effectifs comparés sont supérieurs à 30 ; sauf pour la dernière question, pour laquelle on utilisera systématiquement des tests non paramétriques.