

**Exercices de Monitorat - fiche N°1**

*N.B. Tous les fichiers cités dans les énoncés sont accessibles en salle de TD, dans le répertoire : W:\PSY-L3\Monitorat-R. Ils peuvent également être téléchargés à partir de la page Web: <http://geai.univ-brest.fr/%7Ecarpentier/>*

**Exercice 1.**

Les données du jeu de données Parkinson.RData sont extraites d'une étude sur la maladie de Parkinson, maladie qui affecte, entre autres, l'élocution d'une personne. Dans cette étude, 8 personnes ont subi l'une des opérations les plus courantes pour traiter la maladie, pendant que 12 autres sujets formaient un groupe contrôle. Cette opération semble avoir amélioré l'état général des patients, mais comment a-t-elle affecté leur élocution ? Chaque patient a passé plusieurs tests. Les résultats de l'un de ces tests sont présentés. Plus le score est élevé, plus il y a de difficultés d'élocution.

*Questions portant sur les notions vues dans la fiche de TD N° 1*

- 1) Observer la structure de la feuille de données. Les données y ont en fait été saisies deux fois : soit "par groupe" (colonnes 1 et 2) soit "par variable" (colonnes 3 et 4). Calculer le maximum, le minimum, la moyenne et l'écart type corrigé dans chacun des deux groupes. Commenter les paramètres descriptifs obtenus à la question précédente.
- 2) Réaliser un graphique (par exemple des boîtes à moustache) comparant les distributions dans les deux groupes.

*Question portant sur les notions vues dans la fiche de TD N° 2*

- 3) On veut répondre à la question de l'énoncé à l'aide d'un test de comparaison. Comparer les deux groupes à l'aide d'un test de comparaison de moyennes. Rédiger une phrase de conclusion relative au test précédent.

**Exercice 2**

Des éducateurs utilisant des films pédagogiques ont voulu évaluer l'efficacité relative de divers supports visuels pour voir à quel point les observateurs apprennent et retiennent le sujet présenté. 179 participants ont suivi un bref exposé sur la manière dont les gens assument différents rôles dans un groupe. Le présentateur a montré 18 diapositives identifiant chaque rôle à un animal, une fois en style dessin animé, et une fois de manière réaliste. Tous les participants ont vu l'ensemble des 18 diapositives, mais en couleur pour un groupe (choisi au hasard) et en noir et blanc pour l'autre.

Immédiatement après l'exposé, les diapositives ont été montrées à nouveau aux participants, dans un ordre aléatoire. Il leur a alors été demandé de noter le rôle représenté par l'animal pour chaque diapositive. Les enquêteurs ont relevé et enregistré le nombre de personnages de

dessin animé et de personnages réalistes correctement identifiés. Quatre semaines plus tard, les participants ont repassé le même test; leurs scores ont été recalculés (les absents étant saisis en valeurs manquantes).

Les enquêteurs ont fait passer le test OTIS (épreuve d'aptitude mentale à notation rapide) à tous les participants pour évaluer rapidement leurs capacités naturelles. Au nombre des participants figuraient du personnel en formation pré-professionnelle et du personnel en place de trois hôpitaux de Pennsylvanie, engagé dans un programme interne de formation, ainsi qu'un groupe d'étudiants en licence de l'université de Penn State.

Ouvrez le classeur Excel Dessins.xls. Celui-ci comporte deux feuilles de données : les données complètes sont rassemblées dans la feuille "Dessin-Complet", les données relatives au groupe "noir et blanc" se trouvent également dans la feuille "Dessin-noir-et-blanc".

### *Questions portant sur les notions vues dans la fiche de TD N° 1*

1) On s'intéresse aux sujets qui ont visionné les diapositives en noir et blanc. Importez les données de la feuille correspondante dans un jeu de données R.

Les variables **Dessin1** et **Dessin2** sont définies de la manière suivante :

Dessin1 : Score du test sur dessin (immédiatement après présentation) : de 0 à 9.

Dessin2 : Score du test sur dessin (après 4 semaines) : de 0 à 9.

On veut comparer ces scores pour les 54 sujets qui ont subi les deux épreuves, avec des diapositives en noir et blanc.

a) Calculer le nombre d'observations valides, la moyenne, et l'écart type corrigé dans chacune des deux conditions.

Remarquer que R Commander calcule les paramètres de Dessin1 sur l'ensemble des 89 observations, alors qu'il ne prend en compte que les 54 observations valides de Dessin2 pour calculer les paramètres de cette seconde variable.

b) Définissez un sous-ensemble du jeu de données, en utilisant par exemple l'expression de sélection :

```
!(is.na(Dessin2))
```

N.B. : "!" est l'opérateur de négation, et is.na() est une fonction retournant la valeur TRUE si la valeur de la variable est NA (valeur manquante), FALSE sinon.

Calculer de nouveau le nombre d'observations valides, la moyenne, et l'écart type corrigé dans chacune des deux conditions.

Commenter les paramètres descriptifs obtenus à la question précédente.

Réaliser des graphiques (de type histogramme ou boîte à moustaches) comparant les distributions des variables Dessin1 et Dessin2.

2) On veut comparer les scores observés immédiatement après la présentation pour les sujets qui ont visionné les dessins en noir et blanc et pour ceux qui ont visionné les dessins en couleurs. On utilise donc ici le jeu de données de la feuille "Dessin-Complet".

Calculer le nombre d'observations, la moyenne, et l'écart type corrigé dans chacune des deux conditions.

Rédiger une phrase commentant les paramètres descriptifs obtenus à la question précédente.

Réaliser un graphique (par exemple une boîte à moustaches) comparant les distributions dans les deux groupes.

*Questions portant sur les notions vues dans la fiche de TD N° 2*

3) On veut comparer les scores enregistrés dans les variables Dessin1 et Dessin2. Comparer les deux séries d'observations à l'aide d'un test de comparaison de moyennes. Rédiger une phrase de conclusion relative au test précédent.

4) On veut comparer les scores (observés immédiatement après la présentation) pour les sujets qui ont visionné les dessins en noir et blanc et pour ceux qui ont visionné les dessins en couleurs. On utilise donc ici la feuille de données "Dessin-Complet". Comparer les deux séries d'observations à l'aide d'un test de comparaison de moyennes. Rédiger une phrase de conclusion relative au test précédent.

*Questions portant sur les notions vues dans la fiche de TD N° 3*

5) A l'aide d'un test de la médiane, on veut tester si les deux groupes définis par la variable "Couleur" sont homogènes du point de vue "Score OTIS". Répondre à cette question à l'aide d'un test de comparaison approprié. Rédiger une phrase de conclusion relative à ce deuxième test.

### **Exercice 3 (données Lambert)**

Pour évaluer l'attitude des gens à l'égard des autres, on peut leur demander d'évaluer des individus parlant une autre langue. Mais si les évaluations de ces individus diffèrent, on ignore s'il faut attribuer la variabilité des résultats à l'individu ou à sa langue. Pour tenir compte de ce facteur, Wallace Lambert, de l'université McGill, a été le pionnier de la technique du style apparié. Le sujet écoute divers extraits de langues étrangères, puis note certaines qualités de l'individu enregistré comme amical, honnête, digne de confiance. Mais trois enregistrements proviennent en fait du même individu utilisant différents styles de langage. La différence de notation peut alors être attribuée à des attitudes à l'égard de certaines langues et des populations correspondantes.

Dans une reprise de cette expérience, 58 élèves d'un lycée américain ont évalué sept qualités sur une échelle d'évaluation bipolaire de 1 à 5, à partir d'extraits énoncés dans trois langues : une langue d'un groupe d'immigrants récents, l'anglais et le français.

Les élèves ont aussi rempli des questionnaires complémentaires sur leur expérience d'autres langues et leur sexe.

Cette étude a cherché à déterminer si des élèves en âge de lycée avaient des attitudes négatives significatives à l'égard de la langue de ces immigrants. On avait anticipé qu'ils auraient des attitudes positives à l'égard de l'anglais et du français.

Ouvrez le classeur Excel [Langues.xls](#) ou le jeu de données [Langues.RData](#).

La feuille Langues comporte les variables suivantes :

Sujet : identifiant du sujet

Imm-Qual : Score synthétique observé sur l'extrait de langue parlé par un groupe d'immigrants récents (somme des évaluations pour les 7 qualités envisagées) ;

Ang-Qual : Score synthétique observé sur l'extrait en langue anglaise ;

Fr-Qual : Score synthétique observé sur l'extrait en langue française ;  
Connaissance Français : connaissance de la langue française (oui/non)  
Sexe : sexe du sujet (M/F).

*Questions portant sur les notions vues dans la fiche de TD N° 1*

- 1) a) Calculer la moyenne, et l'écart type des scores observés pour chacune des trois langues.  
b) Rédiger une phrase commentant les paramètres descriptifs obtenus à la question précédente.
- 2) On veut comparer les scores pour la langue du groupe d'immigrants et pour le français. Réaliser des graphiques (de type histogramme ou boîte à moustaches) permettant de comparer les deux distributions.
- 3) On veut comparer les scores pour le français et pour l'anglais. Réaliser des graphiques (de type histogramme ou boîte à moustaches) permettant de comparer les deux distributions.

*Questions portant sur les notions vues dans la fiche de TD N° 2*

- 4) On veut comparer les scores pour la langue du groupe d'immigrants et pour l'anglais. Comparer les deux séries d'observations à l'aide d'un test de comparaison de moyennes. Rédiger une phrase de conclusion relative au test précédent.
- 5) Pour la langue du groupe d'immigrants, les scores attribués par les sujets masculins et ceux attribués par les sujets féminins sont-ils équivalents ? Répondez à cette question à l'aide d'un test de comparaison de moyennes. Rédiger une phrase de conclusion relative à ce deuxième test.
- 6) On veut comparer les scores pour la langue du groupe d'immigrants et pour le français. Comparer les deux séries d'observations à l'aide d'un test de comparaison de moyennes. Rédiger une phrase de conclusion relative au test précédent.
- 7) Les scores attribués pour le français par les sujets déclarant connaître le français et ceux attribués par les autres sujets sont-ils équivalents ? Répondez à cette question à l'aide d'un test de comparaison de moyennes.
- 8) On veut comparer les scores pour le français et pour l'anglais. Comparer les deux séries d'observations à l'aide d'un test de comparaison de moyennes. Rédiger une phrase de conclusion relative au test précédent.

*Question portant sur les notions vues dans la fiche de TD N° 3*

- 9) La connaissance du français est-elle la même pour les garçons et les filles. Répondez à cette question à l'aide d'un test du khi-2.

#### **Exercice 4 (données Assurance)**

Une compagnie d'assurances désire évaluer le comportement des automobilistes qu'elle assure. Les statisticiens de la compagnie choisissent 18 conducteurs qu'ils répartissent en deux groupes en fonction de leurs antécédents d'assurés : "groupe à fort risque d'accident" et "groupe à faible risque d'accident". Ils décident d'étudier les façons de conduire de ces automobilistes sur autoroute, en ville et sur un circuit spécialement aménagé. Ils enregistrent, pour chaque conducteur, le nombre d'erreurs de conduite.

Ouvrez le classeur Excel Assurances.xls. La feuille de données Assurances rassemble les données observées. Importez ou copiez-collez ces données dans un jeu de données R.

*Questions portant sur les notions vues dans la fiche de TD N° 1*

1) Calculer la moyenne et l'écart type du nombre d'erreurs commises pour chacune des trois conditions de circulation.

Rédigeant une phrase commentant les paramètres descriptifs obtenus à la question précédente.

2) On veut comparer le nombre d'erreurs commises par les sujets sur autoroute et en ville. Réaliser des graphiques (de type histogramme ou boîte à moustaches) comparant les deux distributions.

*Questions portant sur les notions vues dans la fiche de TD N° 2*

3) On veut comparer le nombre d'erreurs commises par les sujets sur autoroute et en ville. Comparer les deux séries d'observations à l'aide d'un test de comparaison de moyennes. Rédiger une phrase de conclusion relative au test précédent.

4) Les nombres d'erreurs commises sur autoroute par les conducteurs "à faible risque" et par les conducteurs "à risque fort" sont-ils significativement différents ?

Répondez à cette question en réalisant un test de comparaison de moyennes.

Rédiger une phrase de conclusion relative à ce deuxième test.

5) On veut comparer le nombre d'erreurs commises par les sujets sur circuit et en ville. Comparer les deux séries d'observations à l'aide d'un test de comparaison de moyennes. Rédiger une phrase de conclusion relative au test précédent.

6) Les nombres d'erreurs commises sur circuit par les conducteurs "à faible risque" et par les conducteurs "à risque fort" sont-ils significativement différents ?

Répondre à cette question en réalisant un test de comparaison de moyennes.

Rédiger une phrase de conclusion relative à ce deuxième test.