

Licence de Psychologie - 3^{ème} année - EC PSY54AA

Exercices de Monitorat - fiche N°2

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 5 (données "Pièces de monnaie")

Les personnes qui éprouvent un grand besoin d'argent surestiment les tailles des pièces de monnaie, plus que celles qui n'en éprouvent qu'un besoin modéré. C'est ce qu'a montré une expérience réalisée en 1947. Cependant, certains chercheurs, qui n'ont pas pu reproduire ces résultats, contestent la validité de l'expérience.

On mène une expérience analogue à celle de 1947. On constitue donc un groupe de 20 personnes que l'on place en situation d'éprouver un grand besoin d'argent, et un groupe contrôle de 18 personnes. Les participants doivent, pour 5 pièces (10 centimes, 20 centimes, 50 centimes, 1 Euro, 2 Euros) identifier la taille de la pièce dans un ensemble de 7 tailles, de 85% à 115% de la taille réelle. Ces tailles sont codées de 1 à 7, la taille normale correspondant au code 4.

Le classeur Excel Pieces-Monnaie.xls contient trois feuilles de données rassemblant l'ensemble des valeurs observées. Les feuilles de données "Pieces.Monnaie.Exp" et "Pieces.Monnaie.Ctrl" donnent le résultat de l'évaluation de la taille de chacune des pièces pour les sujets de chacun des deux groupes ainsi qu'un score moyen par sujet, calculé sur l'ensemble des 5 pièces. La feuille de données "Pieces.Monnaie.Ensemble" rassemble l'ensemble des observations. On pourra, selon les questions, utiliser l'une ou l'autre de ces feuilles.

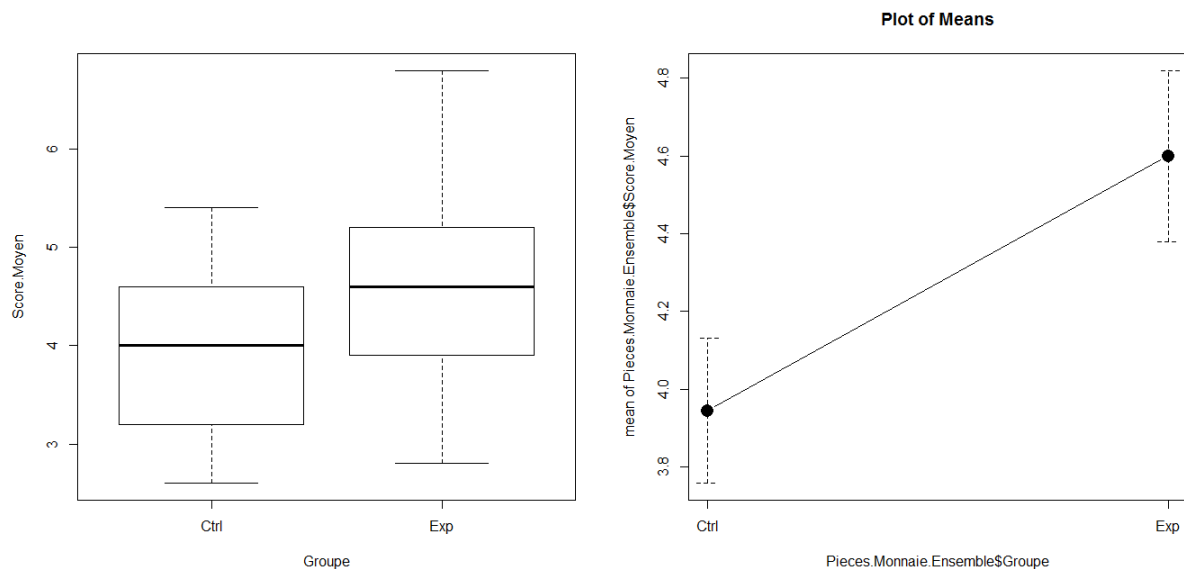
Questions portant sur les notions vues dans la fiche de TD N° 1

- 1) Importez les trois feuilles Excel dans des jeux de données de même nom. Sauvegardez ces jeux de données.
- 2) a) Calculez la moyenne, et l'écart type de la variable "Score Moyen" dans les deux groupes "Exp" et "Ctrl".

	mean	sd	IQR	0%	25%	50%	75%	100%	data:n
Ctrl	3.944444	0.7935195	1.25	2.6	3.25	4.0	4.5	5.4	18
Exp	4.600000	0.9862209	1.15	2.8	4.05	4.6	5.2	6.8	20

b) Complétez le rapport contenu dans le classeur en rédigeant une phrase commentant les paramètres descriptifs obtenus à la question précédente.

3) On veut étudier si le score moyen dépend significativement de la condition expérimentale à laquelle le sujet a été soumis. Réalisez un (ou des) graphique (de type histogramme, boîte à moustaches, ou graphique de moyennes) comparant les scores moyens des deux groupes.



4) Définissez une nouvelle variable calculée dans le jeu de données "Pieces.Monnaie.Ensemble" en composant une formule de calcul permettant de retrouver les scores moyens qui ont été saisis dans la variable Score.Moyen.

$(\text{Piece.1Eu} + \text{Piece.2Eu} + \text{Piece.10c} + \text{Piece.20c} + \text{Piece.50c})/5$

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

5) On veut étudier si le score moyen dépend significativement de la condition expérimentale à laquelle le sujet a été soumis.

a) Comparer les deux scores moyens des deux groupes à l'aide d'un test de comparaison de moyennes.

Two Sample t-test

```
data: Score.Moyen by Groupe
t = -2.241, df = 36, p-value = 0.03129
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-1.24882652 -0.06228459
sample estimates:
mean in group Ctrl mean in group Exp
3.944444 4.600000
```

b) Compléter le rapport contenu dans le classeur en rédigeant une phrase de conclusion relative au test précédent.

6) a) Pour les sujets du groupe expérimental, les scores observés pour les pièces de 20 centimes et ceux relatifs aux pièces de 2 Euros sont-ils équivalents ? Répondez à cette question en réalisant un test de comparaison de moyennes.

Paired t-test

```
data: Pieces.Monnaie.Exp$Piece.20c and Pieces.Monnaie.Exp$Piece.2Eu
t = 2.602, df = 19, p-value = 0.01751
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.1662807 1.5337193
sample estimates:
mean of the differences
                0.85
```

b) Compléter le rapport contenu dans le classeur en rédigeant une phrase de conclusion relative à ce deuxième test.

7) a) Dans une nouvelle variable du jeu de données "Pieces.Monnaie.Exp", calculer le protocole des différences individuelles entre les scores observés pour les pièces de 2 Euros et les pièces de 20 centimes.

Piece.2Eu- Piece.20c

À noter que le contraire (20c-2Eu) est valable aussi. Les résultats seront les mêmes, modulo le signe.

b) Donner un intervalle de confiance, avec un degré de confiance de 95% pour la moyenne de cette variable. Les deux bornes de l'intervalle sont-elles de même signe ?

One Sample t-test

```
data: Pieces.Monnaie.Exp$D2R20
t = -2.602, df = 19, p-value = 0.01751
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 -1.5337193 -0.1662807
sample estimates:
mean of x
    -0.85
```

OU

One Sample t-test

```
data: Pieces.Monnaie.Exp$Diff.20c.2Eu
t = 2.602, df = 19, p-value = 0.01751
alternative hypothesis: true mean is not equal to 0
95 percent confidence interval:
 0.1662807 1.5337193
sample estimates:
mean of x
    0.85
```

8) a) Les scores moyens observés dans le groupe expérimental sont-ils significativement différents de la valeur 4, valeur correspondant à la taille réelle de la pièce ? Répondez à cette question à l'aide d'un test approprié.

```

data: Pieces.Monnaie.Exp$Score.Moyen
t = 2.7208, df = 19, p-value = 0.01357
alternative hypothesis: true mean is not equal to 4
95 percent confidence interval:
 4.138434 5.061566
sample estimates:
mean of x
 4.6

```

b) Complétez le rapport en rédigeant une phrase de conclusion.

9) Reprendre la même question en remplaçant le groupe expérimental par le groupe contrôle.

```

data: Pieces.Monnaie.Ctrl$Score.Moyen
t = -0.297, df = 17, p-value = 0.77
alternative hypothesis: true mean is not equal to 4
95 percent confidence interval:
 3.549836 4.339052
sample estimates:
mean of x
 3.944444

```

Exercice 6

Une étude réalisée par un médecin dans deux services hospitaliers spécialisés dans le traitement et l'accompagnement d'enfants handicapés a été menée sur une dizaine d'années auprès de 48 enfants dans chacun des établissements (soit 96 en tout) qui ont été traités au moins pendant 2 ans par ce service. Pour chacun de ces enfants, on a notamment relevé les variables suivantes :

- educ : Educatif . Nombre de séances par an (de 2 heures chacune) de prise en charge par un éducateur spécialisé ou par un psychologue.
- reed : Rééducatif . Nombre de séances par an (de 2 heures chacune) de prise en charge par un kinésithérapeute, un psychomotricien, un ergothérapeute ou un orthophoniste.
- barr : Barrières. Cette variable mesure l'importance des barrières d'ordre géographique (trottoirs, escaliers. . .) et architecturales (organisation de l'habitat) qui peuvent gêner l'enfant dans ses déplacements. Cette variable a été codée par le médecin : une forte valeur de la variable traduit de fortes barrières au déplacement ; une faible valeur au contraire traduit un environnement géographique amical.
- soc : Sociabilité. Variable codée par le médecin qui traduit l'attitude de l'enfant vis-à-vis des autres (communication et sociabilité). Une forte valeur de la variable traduit un enfant renfermé sur lui-même et peu communicatif ; une faible valeur au contraire est le fait d'un enfant sociable et communicatif.
- auto : Autonomie. Variable qui mesure, du point de vue du médecin, l'autonomie de l'enfant à la fois dans ses possibilités de déplacement dans un environnement amical (pas de trottoirs, pas d'escalier. . .) et dans ses rapports avec les centres (élocution,. . .). Là encore, une forte valeur de la variable révèle un enfant peu autonome, et une faible valeur un enfant autonome.
- grpe : Groupe. Variable qui désigne l'hôpital auquel appartient l'enfant (variable qui ne prend que 2 modalités 1 et 2)

Le classeur Excel [Hopitaux.xls](#) contient deux feuilles de données nommées Hopital.1 et Hopital.2 rassemblant les données observées dans chacun des deux hôpitaux.

Les feuilles de données "Hopital-1" et "Hopital-2" donnent les valeurs observées de ces variables pour les sujets de chacun des deux groupes. La feuille de données "Ensemble" rassemble l'ensemble des observations. On pourra, selon les questions, utiliser l'une ou l'autre de ces feuilles.

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

1) Ouvrez le classeur Excel et définissez une troisième feuille nommée Hopitaux.Ensemble, regroupant l'ensemble des données, celles relatives à l'hôpital 2 se trouvant en dessous de celles relatives à l'hôpital 1. Veillez à ce que les noms des variables issues de la feuille Hopital.2 ne soient pas collées dans la nouvelle feuille, comme s'il s'agissait d'une observation supplémentaire. Sauvegardez le classeur ainsi obtenu.

2) Importez ces trois feuilles de données dans des jeux de données de même nom. Sauvegardez ces jeux de données.

3) Dans le jeu de données Hopitaux.Ensemble, convertissez la variable groupe en facteur.

4) a) Calculer la moyenne et l'écart type de la variable "educ" dans chacun des deux groupes "Hopital 1" et "Hopital 2".

	mean	sd	IQR	0%	25%	50%	75%	100%	data:n
Hopital 1	88.00000	39.97499	36.00	5	61	90	97.00	187	48
Hopital 2	68.45833	29.77644	30.75	3	48	68	78.75	141	48

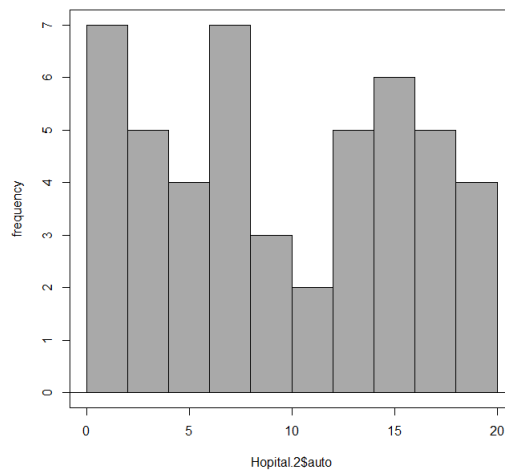
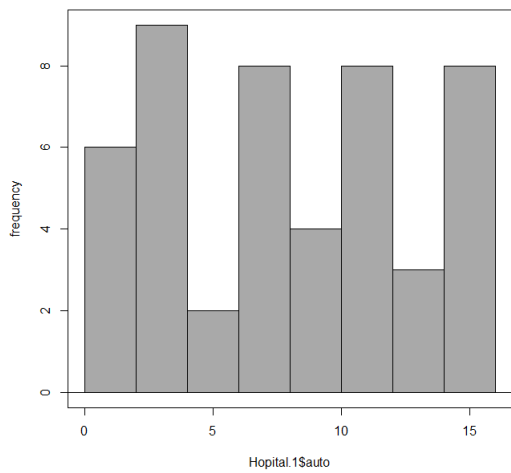
b) Compléter le rapport contenu dans le classeur en rédigeant une phrase commentant les paramètres descriptifs obtenus à la question précédente.

5) Même question pour la variable "barr".

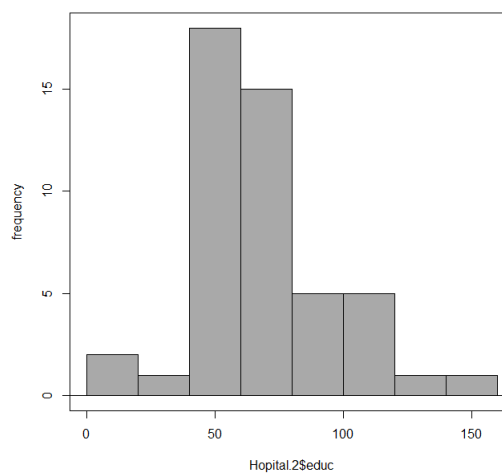
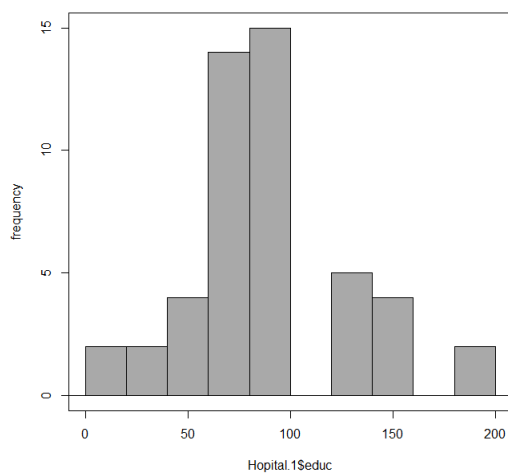
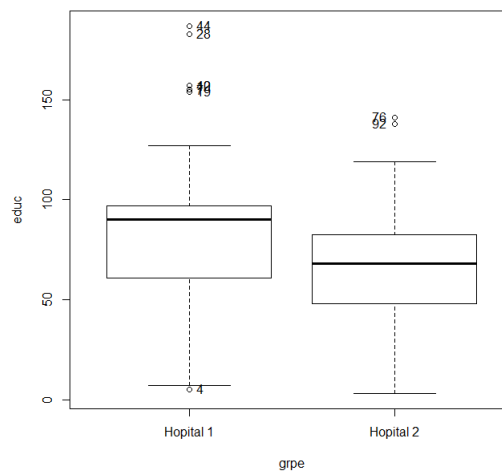
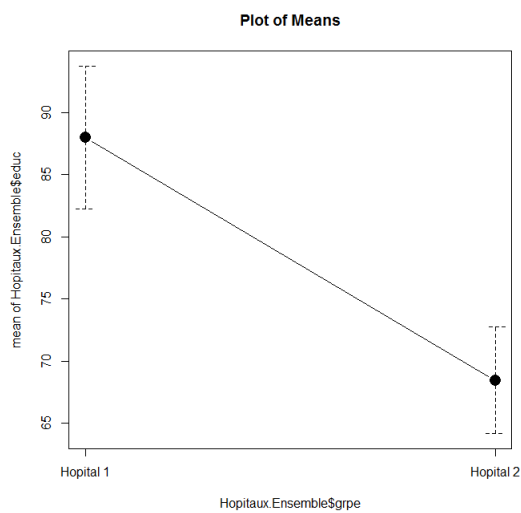
	mean	sd	IQR	0%	25%	50%	75%	100%	data:n
Hopital 1	9.166667	3.096555	4.25	4	6.75	9.5	11	15	48
Hopital 2	11.00000	3.596689	5.00	5	8.00	11.0	13	18	48

6) On veut étudier si la mesure de l'autonomie dépend significativement de l'hôpital dans lequel le sujet a été traité.

Réaliser un/des graphique(s) (de type histogramme, boîte à moustaches ou graphe de moyennes) comparant les scores d'autonomie des deux groupes.



7) On veut étudier si le nombre de séances de prise en charge par un éducateur ou un psychologue dépend significativement de l'hôpital dans lequel le sujet a été traité. Réaliser un/des graphique(s) (de type histogramme, boîte à moustaches ou graphe de moyennes) comparant le nombre de séances dans les deux groupes.



8) Affectez à une nouvelle variable du jeu de données "Hopitaux.Ensemble" une formule de calcul permettant de retrouver pour chaque enfant, le nombre total de séances de prise en charge par des personnels para-médicaux (somme des variables educ et reed).

```
educ+ reed
```

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

9) On veut étudier si le nombre de séances de prise en charge par un éducateur ou un psychologue dépend significativement de l'hôpital dans lequel le sujet a été traité.

a) Comparer les nombres de séances dans les deux groupes à l'aide d'un test de comparaison de moyennes.

```
Two Sample t-test
```

```
data: educ by grpe
t = 2.7161, df = 94, p-value = 0.007862
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 5.256483 33.826850
sample estimates:
mean in group Hopital 1 mean in group Hopital 2
      88.00000           68.45833
```

b) Compléter le rapport contenu dans le classeur en rédigeant une phrase de conclusion relative au test précédent.

10) a) Pour les sujets traités dans l'hôpital 1, le nombre de séances de prise en charge par un éducateur ou un psychologue et le nombre de séances de prise en charge par un kinésithérapeute, un psychomotricien, un ergothérapeute ou un orthophoniste sont-ils significativement différents ?

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

```
Paired t-test
```

```
data: Hopital.1$educ and Hopital.1$reed
t = -5.7667, df = 47, p-value = 6.058e-07
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
-84.44386 -40.76448
sample estimates:
mean of the differences
      -62.60417
```

b) Compléter le rapport contenu dans le classeur en rédigeant une phrase de conclusion relative à ce deuxième test.

11) a) On estime que le nombre de séances de prise en charge par un éducateur ou un psychologue devrait être de 80 par an en moyenne. Au vu de l'échantillon proposé, cette moyenne est-elle atteinte dans l'hôpital 1 ?

Répondez à cette question à l'aide d'un test approprié.

```
One Sample t-test
```

```
data: Hopital.1$educ
```

```
t = 1.3865, df = 47, p-value = 0.1721
alternative hypothesis: true mean is not equal to 80
95 percent confidence interval:
 76.39247 99.60753
sample estimates:
mean of x
      88
```

b) Complétez le rapport en rédigeant une phrase de conclusion.

12) On veut étudier si la mesure de l'autonomie dépend significativement de l'hôpital dans lequel le sujet a été traité.

a) Comparer les scores d'autonomie des deux groupes à l'aide d'un test de comparaison de moyennes.

```
Two Sample t-test

data:  auto by grpe
t = -1.3684, df = 94, p-value = 0.1744
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.8296602  0.7046602
sample estimates:
mean in group Hopital 1 mean in group Hopital 2
      8.395833          9.958333
```

b) Compléter le rapport contenu dans le classeur en rédigeant une phrase de conclusion relative au test précédent.

13) a) Pour les sujets traités dans l'hôpital 1, le score sur l'échelle de sociabilité et le score sur l'échelle d'autonomie sont-ils significativement différents ?

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

```
Paired t-test

data:  Hopital.1$soc and Hopital.1$auto
t = 2.1413, df = 47, p-value = 0.03746
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 0.06933919 2.22232747
sample estimates:
mean of the differences
      1.145833
```

b) Compléter le rapport contenu dans le classeur en rédigeant une phrase de conclusion relative à ce deuxième test.

14) a) On estime que, si le score moyen d'autonomie des patients accueillis dans l'hôpital 2 est significativement inférieur à 12, cet hôpital devrait bénéficier de personnels d'encadrement supplémentaires. Au vu de l'échantillon proposé, cette moyenne est-elle atteinte dans l'hôpital 2?

Répondez à cette question à l'aide d'un test approprié.

One Sample t-test

```
data: Hopital.2$auto
t = -2.3011, df = 47, p-value = 0.01293
alternative hypothesis: true mean is less than 12
95 percent confidence interval:
  -Inf 11.44706
sample estimates:
mean of x
  9.958333
```

b) Complétez le rapport en rédigeant une phrase de conclusion.

Exercice 7

Une société d'études en marketing a collecté des données sur les clientes ayant récemment acheté des jeans dans des grands magasins.

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

1) On a notamment établi un tableau de contingence en croisant la variable Age (4 classes ont été définies : de 10 à 34 ans, de 35 à 44 ans, de 45 à 54 ans, 55 ans et plus) et la variable "Magasin" (4 groupes de magasins ont été définis). On a obtenu le tableau suivant :

	Magasins A	Magasins B	Magasins C	Magasins D
de 10 à 34 ans	46	24	36	73
de 35 à 44 ans	29	35	14	32
de 45 à 54 ans	14	17	10	15
55 ans et plus	12	18	8	8

Le type de magasin fréquenté est-il lié à l'âge ? Répondez à cette question en réalisant un test du khi-2.

```
> .Table # Counts
      A B C D
10 a 34  46 24 36 73
35 a 44  29 35 14 32
45 a 54  14 17 10 15
55 et plus 12 18  8  8

> .Test <- chisq.test(.Table, correct=FALSE)
Pearson's Chi-squared test
data: .Table
X-squared = 26.5464, df = 9, p-value = 0.001662
```

Rédigez une phrase de conclusion relative à ce test. Examinez les tableaux des effectifs observés et des effectifs théoriques et faites un commentaire.

```
> .Test$expected # Expected Counts
      A B C D
10 a 34  46.23785 43.03325 31.13043 58.59847
35 a 44  28.41432 26.44501 19.13043 36.01023
45 a 54  14.46547 13.46292  9.73913 18.33248
55 et plus 11.88235 11.05882  8.00000 15.05882
```

2) On a établi un tableau de contingence en croisant la variable PrixJean (prix du dernier jean acheté, 4 classes ont été définies : jusqu'à 15\$, plus de 15\$ et jusqu'à 20\$, plus de 20\$ et jusqu'à 30\$, plus de 30\$) et la variable "Dépenses" (Cette variable représente le degré d'adhésion à l'affirmation "Je dépense beaucoup en vêtements et accessoires"; la réponse est codée sur une échelle de 1 à 6 où 1 = absolument pas et 6 = tout à fait). On a obtenu les résultats suivants :

	Opinion relative aux dépenses en vêtements					
	1	2	3	4	5	6
jusqu'à 15\$	28	14	16	6	9	5
jusqu'à 20\$	20	18	16	12	4	4
jusqu'à 30\$	11	16	24	37	16	7
plus de 30\$	7	10	15	31	33	19

Le prix du dernier jeans acheté et l'opinion relative aux dépenses de vêtements sont-ils liés ? Répondez à cette question en réalisant un test du khi-2.

```
> .Table # Counts
      1  2  3  4  5  6
15$ 28 14 16  6  9  5
20$ 20 18 16 12  4  4
30$ 11 16 24 37 16  7
plus  7 10 15 31 33 19

> .Test <- chisq.test(.Table, correct=FALSE)

      Pearson's Chi-squared test
data:  .Table
X-squared = 84.1846, df = 15, p-value = 1.19e-11
```

Rédigez une phrase de conclusion relative à ce test. Examinez les tableaux des effectifs observés et des effectifs théoriques et faites un commentaire.

```
> .Test$expected # Expected Counts
      1          2          3          4          5          6
15$ 13.61905 11.96825 14.65079 17.74603 12.79365  7.222222
20$ 12.92063 11.35450 13.89947 16.83598 12.13757  6.851852
30$ 19.38095 17.03175 20.84921 25.25397 18.20635 10.277778
plus 20.07937 17.64550 21.60053 26.16402 18.86243 10.648148
```

Exercice 8

Une enquête a été réalisée aux Etats-Unis pour étudier l'incidence de l'alcoolisme dans différentes catégories socio-professionnelles. On a obtenu le tableau suivant :

	Alcooliques	Non alcooliques
Clergé	32	268
Educateurs	51	199
Chefs d'entreprises	67	233
Commerçants	83	267

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

Le taux d'alcoolisme est-il le même pour toutes les catégories socioprofessionnelles ? Répondez à cette question en réalisant un test du khi-2.

```
> .Table # Counts
                Alcohol Non alcohol
Clergé           32         268
Educateurs       51         199
Chefs d entreprise 67         233
Commerçants      83         267

> .Test <- chisq.test(.Table, correct=FALSE)
> .Test
      Pearson's Chi-squared test
data:  .Table
X-squared = 20.5967, df = 3, p-value = 0.0001277
```

Rédigez une phrase de conclusion relative à ce test.

Exercice 9

3025 auditeurs d'un centre de formation pour adultes ont répondu à une enquête sur leurs difficultés à suivre les enseignements en cours du soir qui leur sont proposés.

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

1) On a extrait de cette enquête les 2 questions suivantes :

Libellé	Réponses proposées	Code Question	Code Modalité
Le contenu de cet enseignement est-il nouveau pour vous ?	1. Oui, tout à fait 2. En partie 3. Pas du tout	CONTENU	1. nouveau 2. en partie nouveau 3. pas nouveau
Avez-vous des difficultés à travailler après le cours ?	1. Oui beaucoup 2. En partie 3. Pas du tout	TRAVAIL	1. très difficile 2. assez difficile 3. pas difficile

Le croisement des réponses obtenues aux deux questions conduit au tableau suivant :

		CONTENU		
		nouveau	en part. nouv.	pas nouveau
TRAVAIL	tres difficile	432	340	97
	assez difficile	412	482	97
	pas difficile	485	492	188

a) Y a-t-il un lien entre les réponses aux deux questions traitées ? Répondez à cette question en réalisant un test du khi-2.

```
> .Table # Counts
                Nouv En partie nouv Pas nouv
Tres diff      432           340         97
Assez diff     412           482         97
```

Pas diff 485 492 188

```
> .Test <- chisq.test(.Table, correct=FALSE)
      Pearson's Chi-squared test
data:  .Table
X-squared = 38.7153, df = 4, p-value = 7.977e-08
```

c) Rédigez une phrase de conclusion relative à ce test.

2) On a également extrait de cette enquête les 2 questions suivantes :

Libellé	Réponses proposées	Code Question	Code Modalité
Le contenu de cet enseignement est-il nouveau pour vous ?	1. Oui, tout à fait 2. En partie 3. Pas du tout	CONTENU	1. nouveau 2. en partie nouveau 3. pas nouveau
Vous trouvez que le rythme des enseignements est :	1. Trop rapide 2. Comme il faut 3. Trop lent	RYTHME	1. trop rapide 2. comme il faut 3. trop lent

Le croisement des réponses obtenues aux deux questions conduit au tableau suivant :

		RYTHME		
		trop rapide	comme il faut	trop lent
CONTENU	nouveau	338	934	57
	en part. nouv.	252	984	78
	pas nouveau	57	279	46

a) Y a-t-il un lien entre les réponses aux deux questions traitées ? Répondez à cette question en réalisant un test du khi-2.

```
> .Table # Counts
      Trop rapide Comme il faut Trop lent
Nouveau          338          934         57
En partie nouv.    252          984         78
Pas nouveau        57           279         46

> .Test <- chisq.test(.Table, correct=FALSE)
      Pearson's Chi-squared test
data:  .Table
X-squared = 52.3908, df = 4, p-value = 1.143e-10
```

b) Insérez un rapport dans le classeur précédent et rédigez une phrase de conclusion relative à ce test.

Exercice 10

Des chercheurs se sont intéressés à l'effet des panneaux publicitaires placés le long des routes sur l'attention des conducteurs de véhicules. Pour cela, ils ont observé la conduite de 48 sujets sur un simulateur, dans diverses conditions : avec ou sans publicité, en circulation urbaine, autoroutière ou sur route secondaire.

A la fin de chaque séquence de conduite, on demandait notamment à chaque sujet quel était le dernier panneau routier rencontré et, dans la condition "avec publicités", quel était le dernier

panneau publicitaire rencontré. On notait si la réponse était correcte ou incorrecte. En circulation urbaine, la comparaison des rappels conduit aux résultats suivants :

		Publicité	
		Rappel correct	Rappel incorrect
Panneau routier	Rappel correct	15	21
	Rappel incorrect	6	6

1) Générez un nouveau jeu de données nommé "Conduite", structuré de façon à pouvoir réaliser le test envisagé à la question suivante.

Sujet	Rappel.panneau	Rappel.publicite
s1	1	1
s2	1	1
s3	1	1
s4	1	1
s5	1	1
s6	1	1
s7	1	1
s8	1	1
s9	1	1
s10	1	1
s11	1	1
s12	1	1
s13	1	1
s14	1	1
s15	1	1
s16	0	1
s17	0	1
s18	0	1
s19	0	1
s20	0	1
s21	0	1
s22	1	0
s23	1	0
s24	1	0
s25	1	0
s26	1	0
s27	1	0
s28	1	0
s29	1	0
s30	1	0
s31	1	0
s32	1	0
s33	1	0
s34	1	0
s35	1	0
s36	1	0
s37	1	0
s38	1	0
s39	1	0
s40	1	0
s41	1	0
s42	1	0
s43	0	0
s44	0	0
s45	0	0
s46	0	0
s47	0	0
s48	0	0

ou

Sujet	Panneau	Publicite
s1	0	1
s2	0	1
s3	0	1
s4	0	1

s5	0	1
s6	0	1
s7	1	0
s8	1	0
s9	1	0
s10	1	0
s11	1	0
s12	1	0
s13	1	0
s14	1	0
s15	1	0
s16	1	0
s17	1	0
s18	1	0
s19	1	0
s20	1	0
s21	1	0
s22	1	0
s23	1	0
s24	1	0
s25	1	0
s26	1	0
s27	1	0

2) a) Les performances de rappel sont-elles les mêmes dans les deux conditions ? Répondez à cette question en réalisant un test approprié.

```
> friedman.test(.Responses)
      Friedman rank sum test
data:  .Responses
Friedman chi-squared = 8.3333, df = 1, p-value = 0.003892
```

b) Rédigez une phrase de conclusion relative à ce test.

Exercice 11 - Enquete-Fraude

Une enquête a été menée en 2007 auprès d'un échantillon d'étudiants d'une Université parisienne. Il s'agissait d'un questionnaire portant sur la fraude aux examens universitaires.

Les personnes interrogées répondaient à quelques questions d'ordre général (âge, sexe, niveau d'études, UFR, redoublement éventuel, mention au Bac), puis à plusieurs questions relatives à la fraude. Dix techniques de fraude étaient envisagées : copier en regardant la copie d'un autre, communiquer avec un voisin, échanger les brouillons, etc. Pour chaque technique, les sujets interrogés devaient répondre sur une échelle de Likert à 5 niveaux : jamais, rarement, parfois, souvent, toujours. Les réponses recueillies ont été codées de 0 (pour "jamais") à 4 (pour "toujours") et un score de synthèse (variable ScoreTricheTotal) est calculé en additionnant les valeurs correspondant à chaque méthode.

Enfin, on demandait aux sujets s'ils avaient utilisé l'une de ces techniques au Lycée et lors des épreuves du Bac. Ces dernières données sont saisies à la fois sous forme de facteurs (Oui/Non : variables LyceeTriche et BacTriche) et sous forme de variables numériques dichotomiques (1/0 : variables LyceeTricheNum et BacTricheNum).

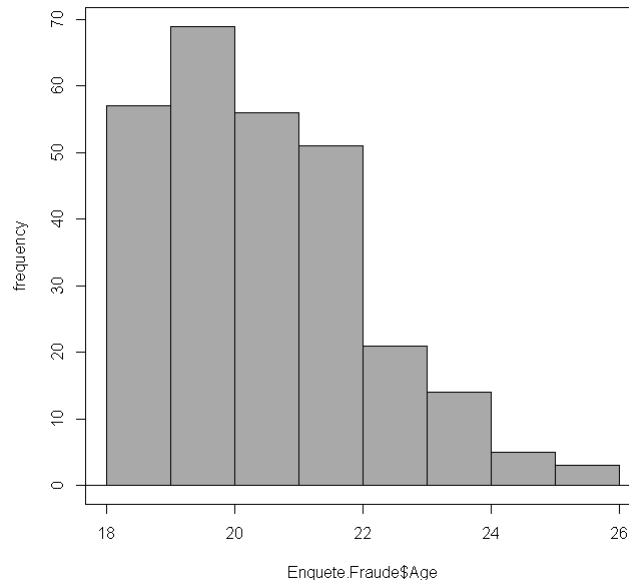
Les données sont rassemblées dans le jeu de données Enquete.Fraude.RData.

1) On s'intéresse à la variable "Age".

a) Calculez la moyenne et l'écart type de cette variable.

```
> numSummary(Enquete.Fraude[,"Age"], statistics=c("mean", "sd"),
+   quantiles=c(0,.25,.5,.75,1))
      mean      sd   n NA
20.91667 1.663603 276 10
```

b) Réalisez un graphique représentant la distribution de cette variable dans l'échantillon.

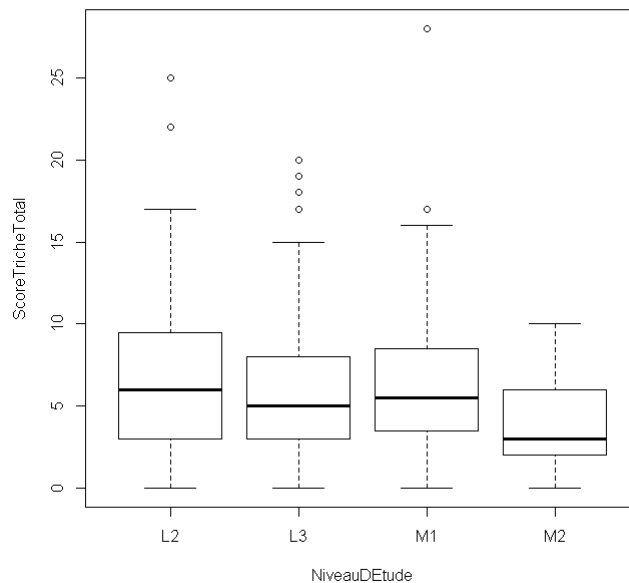


2) a) Calculez la moyenne et l'écart type de la variable ScoreTricheTotal dans les groupes définis par la variable "NiveauDEtudes".

```
> numSummary(Enquete.Fraude[,"ScoreTricheTotal"],
+   groups=Enquete.Fraude$NiveauDEtude, statistics=c("mean", "sd"),
+   quantiles=c(0,.25,.5,.75,1))
      mean      sd data:n
L2 6.728155 5.010174   103
L3 6.235772 4.244729   123
M1 6.250000 5.753881    36
M2 4.000000 2.796101    23
```

b) Rédigez une phrase commentant les paramètres descriptifs obtenus à la question précédente.

c) Réalisez un graphique de type boîte à moustaches comparant les différents groupes définis par le niveau d'études du point de vue de la variable ScoreTricheTotal.



3) La technique "Envoyer et recevoir des SMS" (variable "SMS") est-elle pratiquée moins couramment que la technique "Fabriquer ou utiliser des antisèches" (variable "Antiseche") ?
 a) Répondez à cette question à l'aide d'un test non paramétrique.

```
> wilcox.test(Enquete.Fraude$SMS, Enquete.Fraude$Antiseche,
+ alternative='less', paired=TRUE)
```

Wilcoxon signed rank test with continuity correction

```
data: Enquete.Fraude$SMS and Enquete.Fraude$Antiseche
V = 789.5, p-value < 2.2e-16
alternative hypothesis: true location shift is less than 0
```

b) Interprétez les résultats de ce test.

4) La technique "Envoyer et recevoir des SMS" (variable "SMS") est-elle pratiquée plus couramment que la technique "Voler les sujets avant l'examen" (variable "VolerSujet") ?
 a) Répondez à cette question à l'aide d'un test de comparaison de moyennes.

```
> t.test(Enquete.Fraude$SMS, Enquete.Fraude$VolerSujet,
alternative='greater',
+ conf.level=.95, paired=TRUE)
```

Paired t-test

```
data: Enquete.Fraude$SMS and Enquete.Fraude$VolerSujet
t = 2.3033, df = 285, p-value = 0.01099
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 0.02776039      Inf
sample estimates:
mean of the differences
      0.0979021
```

b) Interprétez les résultats de ce test.

5) Y a-t-il un lien entre le fait d'avoir triché au lycée et le fait d'avoir triché lors des épreuves du Bac (variables "LyceeTriche" et "BacTriche").

- a) Le test pertinent est là le test du khi-2 sur un tableau de contingence. Justifiez.
- b) Réalisez le test.

```
> .Table
      BacTriche
LyceeTriche Non  Oui
      Non   39   2
      Oui  189  55

> .Test <- chisq.test(.Table, correct=FALSE)
> .Test
      Pearson's Chi-squared test
data:   .Table
X-squared = 6.8444, df = 1, p-value = 0.008892
```

c) Rédigez une phrase de conclusion relative à ce test.

6) Du point de vue des techniques de fraude, les sujets ont-ils eu le même comportement au Lycée et lors des épreuves du Bac.

- a) Le test pertinent est là le test du khi-2 de MacNemar. Justifiez.
- a) Réalisez le test en utilisant les variables LyceeTricheNum et BacTricheNum.

```
> friedman.test(.Responses)

      Friedman rank sum test

data:   .Responses
Friedman chi-squared = 183.0838, df = 1, p-value < 2.2e-16
```

b) Rédigez une phrase de conclusion relative à ce test.

7) Le score de synthèse "ScoreTricheTotal" est-il significativement différent selon le sexe de la personne ?

- a) Comparez les scores observés pour les deux sexes à l'aide d'un test paramétrique.

```
> t.test(ScoreTricheTotal~Sexe, alternative='two.sided', conf.level=.95,
+ var.equal=TRUE, data=Enquete.Fraude)

      Two Sample t-test

data:   ScoreTricheTotal by Sexe
t = -3.5815, df = 284, p-value = 0.0004018
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -3.0921911 -0.8987728
sample estimates:
mean in group Femme mean in group Homme
      5.125000          7.120482
```

b) Rédigez une phrase de conclusion relative à ce test.

c) Reprenez le test précédent en vous limitant aux étudiants qui ont déclaré avoir utilisé au moins une des techniques de triche (variable ScoreTricheTotal strictement positive) et rédigez une phrase de conclusion.

Sous-ensemble défini par : ScoreTricheTotal > 0

```
> t.test(ScoreTricheTotal~Sexe, alternative='two.sided', conf.level=.95,  
+ var.equal=TRUE, data=Enquete.Fraude.1)
```

Two Sample t-test

```
data: ScoreTricheTotal by Sexe  
t = -3.3695, df = 266, p-value = 0.0008647  
alternative hypothesis: true difference in means is not equal to 0  
95 percent confidence interval:  
-2.9945675 -0.7856396  
sample estimates:  
mean in group Femme mean in group Homme  
5.590909 7.481013
```

8) Pour chacune des dix techniques de fraude, on souhaite étudier si le score est significativement différent selon le sexe de la personne.

a) Comparez les scores observés pour les deux sexes, pour chacune des variables Copier, Communiquer, EchangeBrouillon, Antiseche, SMS, CoursGenoux, GarderCopie, PreparerSalle, VolerSujet, Autres à l'aide de tests non paramétriques.

```
> wilcox.test(Copier ~ Sexe, alternative="two.sided", data=Enquete.Fraude)
```

```
data: Copier by Sexe  
W = 8162.5, p-value = 0.006678  
alternative hypothesis: true location shift is not equal to 0
```

```
data: Communiquer by Sexe  
W = 8543, p-value = 0.03045  
alternative hypothesis: true location shift is not equal to 0
```

```
data: EchangeBrouillon by Sexe  
W = 8184.5, p-value = 0.006032  
alternative hypothesis: true location shift is not equal to 0
```

```
data: Antiseche by Sexe  
W = 8051.5, p-value = 0.003501  
alternative hypothesis: true location shift is not equal to 0
```

```
data: SMS by Sexe  
W = 9154, p-value = 0.05281  
alternative hypothesis: true location shift is not equal to 0
```

```
data: CoursGenoux by Sexe  
W = 9215, p-value = 0.1403  
alternative hypothesis: true location shift is not equal to 0
```

```
data: GarderCopie by Sexe  
W = 9500.5, p-value = 0.1498  
alternative hypothesis: true location shift is not equal to 0
```

```
data: PreparerSalle by Sexe  
W = 9260, p-value = 0.08556
```

alternative hypothesis: true location shift is not equal to 0

data: VolerSujet by Sexe

W = 9166, p-value = 0.0189

alternative hypothesis: true location shift is not equal to 0

data: Autres by Sexe

W = 9341.5, p-value = 0.02046

alternative hypothesis: true location shift is not equal to 0

b) Rédigez des phrases de conclusion relatives à ces tests.

9) Y a-t-il un lien entre la mention obtenue au Bac et le fait d'avoir redoublé au cours de la scolarité ? (variables "Redoublement" et "MentionBac").

a) Quel test permet d'apporter une réponse à cette question ?

b) Réalisez le test indiqué.

```
> .Table
      MentionBac
Redoublement Assez Bien Bien ou TB Passable Rattrapage
      Non      46      19      64      14
      Oui      40       5      70      28
```

```
> .Test <- chisq.test(.Table, correct=FALSE)
```

```
      Pearson's Chi-squared test
```

```
data: .Table
```

```
X-squared = 13.5206, df = 3, p-value = 0.003636
```

c) Rédigez une phrase de conclusion relative à ce test.

10) Le score de synthèse "ScoreTricheTotal" est-il significativement différent selon l'UFR de rattachement ?

a) Comparez les scores observés dans les trois UFR de rattachement. Puisqu'il s'agit de comparer plus de deux groupes, on utilisera une analyse de variance à un facteur.

```
> AnovaModel.1 <- aov(ScoreTricheTotal ~ UFR, data=Enquete.Fraude)
```

```
> summary(AnovaModel.1)
```

```
      Df Sum Sq Mean Sq F value Pr(>F)
UFR      2    153   76.72   3.466 0.0326 *
Residuals 283   6265   22.14
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

b) Rédigez une phrase de conclusion relative à ces tests.