

## TP3 TESTS DE COMPARAISONS DE MOYENNES SOUS R

### Exercice 1 : Efficacité d'un régime.

#### TEST de comparaison de moyennes

Le service des études d'un laboratoire de fabrication de régimes amaigrissants a commandé une enquête en vue d'améliorer les performances d'un nouveau régime d'une durée de 6 semaines. Un échantillon de 8 femmes et 12 hommes a été sélectionné. On a noté le poids des différentes personnes au début et à la fin du régime.

Les résultats sont les suivants :

-Femmes

Début	79	80	75	74	76	71	69	65
Fin	75	72	74	71	71	66	65	63

-Hommes

Début	78	85	88	84	82	79	96	102	95	83	81	92
Fin	71	80	84	80	80	73	87	91	90	80	77	88

Le régime pour les hommes est-il plus efficace que celui pour les femmes ?

Pour répondre à cette question nous devons utiliser un test de comparaison de moyenne (pensez à vous référer au fichier récapitulatif des tests), voici les étapes intermédiaires pour y arriver :

- 1) Rentrez les données sous quatre vecteurs Femmes(début,fin) et Hommes(début,fin)
- 2) A partir de vos vecteurs initiaux, calculer deux vecteurs que vous nommerez 'diff\_femmes' et 'diff\_hommes' car nous devons obtenir qu'une unique série de valeur pour les femmes et pour les hommes afin de réaliser le futur test.

**! Présentation à appliquer valable pour tous les tests statistiques !**

*1-énoncez clairement vos hypothèses H0 et H1*

*2-réalisez le test sur R et copier le dans votre CR*

*3-concluez en fonction de la p-value obtenue*

- 3) *Remarque : il faudrait tout d'abord se demander si les données sont appariées ou non -> ici les données sont non-appariées = indépendantes (nous verrons un autre exemple dans l'exercice suivant).*

Premièrement, testez la normalité de chacun des échantillons avec un test de Shapiro.Wilk `shapiro.test()` et concluez.

- 4) Deuxièmement, testez l'égalité des variances des deux échantillons à l'aide d'un test de Fisher-Snedecor : `var.test(diff_femmes,diff_hommes)` et concluez.
- 5) Utiliserez-vous un test bilatéral (on teste s'il y a une **différence** entre les deux séries) ou unilatéral (on teste si une série est **supérieure** à l'autre série ou, au choix, on teste si une série est **inférieure** à l'autre série) ?
- 6) En fonction des résultats aux questions 3) et 4) quel test utiliseriez-vous ?
- 7) Effectuez le test choisi et concluez.

(écrire `>help(t.test)` par exemple pour avoir de l'aide sur la manière d'écrire la fonction)

*Vous trouverez la correction de cet EXERCICE en bas de page.*

## Exercice 2 : METEOROLOGIE

On dispose de deux relevés hydrologiques indiquant les hauteurs de précipitations sur 8 huit zones lors de deux épisodes pluvio-orageux successifs.

Ces deux relevés sont donnés par le tableau suivant :

ZONE	1	2	3	4	5	6	7	8
Episode1	125	130	132	135	136	138	140	145
Episode2	127	132	133	136	139	141	145	148

On souhaiterait comparer les moyennes des précipitations d'un épisode orageux à l'autre pour voir si l'on constate une différence en fonction des zones.

Pour répondre à la question posée, suivez les étapes suivantes :

1) IMPORTANT : Les données sont-elles appariées (par paires) ? (autrement dit, dans notre contexte, la 1<sup>ère</sup> valeur de la série 1 est-elle liée avec la 1<sup>ère</sup> valeur de la série 2, etc.)

2) Puisque les relevés hydrologiques sont effectués dans des zones respectivement identiques et que la problématique nous spécifie bien une comparaison « en fonction des zones » : les données sont considérées appariées.

Nous allons réaliser les tests préalables habituels (Shapiro puis Fisher), *mais attention lorsque les données sont appariées c'est un cas particulier pour Shapiro : il faut étudier la normalité des différences entre chaque échantillons appariés. Pour cela, il vous faudra utiliser un nouveau vecteur =episode1-episode2.*

3) Selon vos résultats obtenus aux tests préalables, quel test de comparaison de moyennes utiliseriez-vous ?

4) Réalisez le test en veillant bien à spécifier tous les arguments nécessaires au test (*cf Guide R*).

### Exercice 3 : Industrie

Deux entreprises industrielles souhaitent comparer les rendements de deux de leurs machines exprimés en pièces fabriquées par heure, on obtient les résultats suivants :

Machine 1	108	107	110	105	109	108	103
Machine 2	107	111	112	109	108	116	104

a) Réaliser les tests préalables, puis selon vos résultats choisissez le test approprié.

b) Réaliser le test permettant de déterminer si une des deux machines est plus performante que l'autre.

### Exercice 4

#### TEST de comparaison de moyennes

Comparaison de l'efficacité de deux médicaments A et B.

Avec le médicament A la durée moyenne de disparition de la douleur est de 30mn.

On a administré le médicament B à 12 patients et on a relevé les durées de disparition de la douleur.

Durée(mn)    25   28   20   32   17   24   41   28   25   30   27   24

Les variances de A et de B sont supposées égales.

On décide de commercialiser le médicament B seulement si on est certain à 95 % qu'il est plus efficace (douleur qui disparaît plus rapidement) que le médicament A.

Construire un test permettant de décider ou non de la commercialisation de B.

- 1) Créer un vecteur « tempsB » des données expérimentales
- 2) Calculer la moyenne et la variance de la série B.
- 3) Vérifier la normalité de la distribution avec un test de Shapiro-Wilk
- 4) Réaliser le test de comparaison de moyennes adapté pour répondre à la problématique.

### CORRECTION EXERCICE 1

1) Rentrez les données sous deux vecteurs Femmes et Hommes

```
> #enregistrer les donnees femmes debut
> #enregistrer les donnees femmes fin
> debutF<-c(79,80,75,74,76,71,69,65)
> finF<-c(75,72,74,71,71,66,65,63)
```

```
> debutH<-c(78,85,88,84,82,79,96,102,95,83,81,92)
> finH<-c(71,80,84,80,80,73,87,91,90,80,77,88)
```

2) Déterminez deux vecteurs que vous nommerez « diff\_femmes » et « diff\_hommes ».

```
> diff<-debutF-finF
> diff
[1] 4 8 1 3 5 4 2
```

```
> diffH<-debutH-finH
> diffH
[1] 7 5 4 4 2 6 9 11 5 3 4 4
```

3) Testez la normalité de chacun des échantillons et concluez.

Test de Shapiro-Wilk.

H0 : Les données suivent une loi normale

H1 : Les données ne suivent pas une loi normale

Risque alpha fixé à  $\alpha=5\%$

```
> shapiro.test(diff)

Shapiro-Wilk normality test

data: diff
W = 0.95723, p-value = 0.7833
```

P-value > 5%, on admet H0 : Les données suivent une loi normale, avec un risque de seconde espèce (risque bêta).

```
> shapiro.test(diffH)

Shapiro-Wilk normality test

data: diffH
W = 0.89544, p-value = 0.1385
```

P-value > 5%, on admet H0 : Les données suivent une loi normale, avec un risque de seconde espèce.

4) Utilisez-vous un test unilatéral ou bilatéral ?

On utilisera un test unilatéral puisque l'on veut mettre en évidence une meilleure efficacité du régime pour les hommes.

5) Testez les variances des deux populations à l'aide d'un test de Fisher-Snedecor.

Test de Fisher

H0 : Les variances sont égales

H1 : Les variances sont différentes

Risque alpha fixé à  $\alpha=5\%$

```
> var.test(diff,diffH)

F test to compare two variances

data: diff and diffH
F = 0.69201, num df = 7, denom df = 11, p-value = 0.6428
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.1841106 3.2589778
sample estimates:
ratio of variances
 0.6920052
```

P-value > 5%, on admet H0 : Les variances sont égales, avec un risque de seconde espèce.

6) En fonction des résultats des questions 4) et 5) quel type de test utiliserez-vous ?

Hypothèse de normalité pour chacune des variables + variances égales -> conditions d'application du Test de STUDENT

7) Effectuez le test choisi et concluez.

Test de Student

H0 : Les moyennes des régimes Hommes et Femmes sont égales.

H1 : Le régime Hommes est PLUS efficace que le régime Femmes.

Risque alpha fixé à  $\alpha=5\%$

```
> t.test(diffH,diff, alternative="greater",var.equal=T)

Two Sample t-test

data: diffH and diff
t = 1.2114, df = 18, p-value = 0.1207
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -0.5752554      Inf
sample estimates:
mean of x mean of y
 5.333333  4.000000
```

P-value > 5%, TEST NON SIGNIFICATIF, on ne peut pas conclure que le régime des Hommes est plus efficace que celui des Femmes au seuil  $\alpha=5\%$ .

Remarque :

Dans le cas d'un test de Student unilatéral, il faut faire attention à l'ordre des variables sur R dans la commande `t.test()`. En choisissant « greater » en alternative hypothesis, on teste le fait que la série1 soit supérieure à la série2.