Beimpact

Projet transdisciplinaire

Traitement des données

IPB
ENSC
BORDEAUX

Encadré par E. Kijewski

Ravat Anita - Léré Gwladys
Unrein Hélène - Baret Loïc

Traitement des données

I. Introduction

1. Participants :

Les données de l'expérience proviennent de 16 participants répartis en deux groupes, de sexe et de dépendance au champ différents. Les tableaux ci-dessous permet de comprendre leur répartition :

Carte	Femme	Homme	Total
Dépendant	1	1	2
Indépendant	3	4	7
Total	4	5	9

GPS	Femme	Homme	Total
Dépendant	0	2	2
Indépendant	2	3	5
Total	2	5	7

2. Données :

D'après les données provenant des différentes tâches (reconnaissance, de série et de positionnement) nous avons calculé différents scores de réussite. Ainsi nous obtenons les tableaux suivants :

Score tâche de reconnaissance	
Carte	GPS
-2	0
1	3
-1	-3
3	0
1	3
4	8
3	0
6	
9	

Score tâche de série	
Carte	GPS
0	0,70
0,62	0,67
0	0,61
0,75	0,72
0,75	1,00
0,67	0,87
0,72	0,58
0,53	
0,53	

Score tâche de positionnement	
Carte	GPS
0	0,36
0,43	0,44
0	0,50
0,40	0,57
0	0,00
0	0,00
0	0,44
0	
0	

II. Traitement des données

1. Programme

Nous utiliserons ici le logiciel R afin de traiter les données. Les données sont rentrées selon les codes suivants :

```
Carte_reconnaissance<-c(-2,1,-1,3,1,4,3,6,9)
GPS_reconnaissance<-c(0,3,-3,0,3,8,0)

Carte_serie<-c(0,0.62,0,0.75,0.75,0.67,0.72,0.53,0.53)
GPS_serie<-c(0.70,0.67,0.61,0.72,1,0.87,0.58)

Carte_position<-c(1,0.43,0,0.40,0.44,0.4,0.14,0.09,0.18)
GPS_position<-c(0.36,0.44,0.50,0.57,0.50,0.50,0.44)
```

2. Tâche de reconnaissance

a. Hypothèse

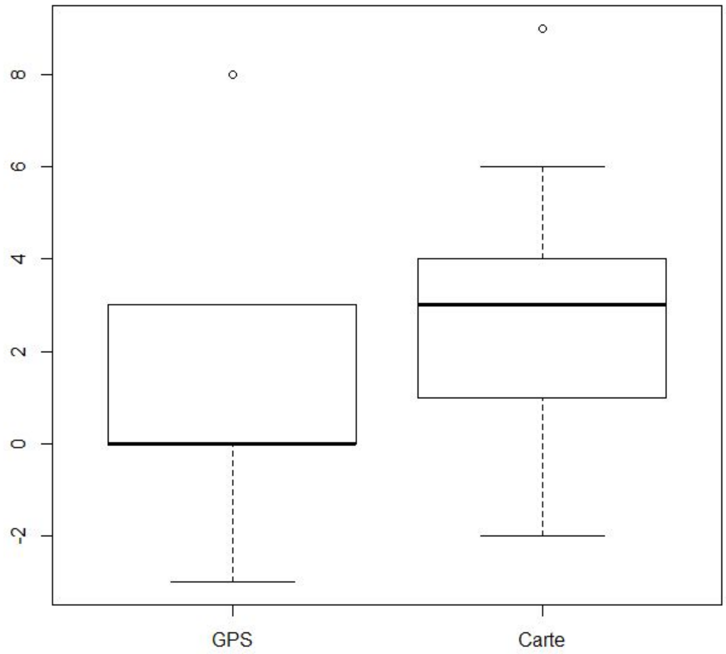
Ces deux échantillons sont indépendants l'un de l'autre. L'hypothèse que l'on se pose alors est : les participants "cartes" ont ils un meilleur score que les participants "GPS"? Soit l'hypothèse nulle tel qu'il n'y a pas de différence significative entre les deux échantillons; contre l'hypothèse H1 tel que les participants "cartes" retient significativement plus de repères que les participants "GPS".

b. Statistiques descriptives

```
> summary(GPS_reconnaissance)
  Min. 1st Qu.  Median    Mean 3rd Qu. 
Max.
 -3.000   0.000   0.000   1.571   3.000 
 8.000
> var(GPS_reconnaissance)
[1] 12.28571
> sd(GPS_reconnaissance)
[1] 3.505098
> summary(Carte_reconnaissance)
   Min. 1st Qu.  Median    Mean 3rd Qu. 
Max.
 -2.000   1.000   3.000   2.667   4.000   9.000
> var(Carte_reconnaissance)
[1] 11.75
> sd(Carte_reconnaissance)
[1] 3.427827
```

D'après ces commandes on peut voir que la moyenne des scores pour l'échantillon Carte est de 2,6 contre 1,5 pour l'échantillon GPS. De plus la variance et l'écart type sont moins élevés pour l'échantillon Carte que pour l'échantillon GPS; respectivement 11,75 et 3,42 contre 12,2 et 3,5. Ce qui veut dire que l'échantillon Carte est moins étendu. Ainsi la question de la différence de score se pose justement.

> boxplot(GPS_reconnaissance, Carte_reconnaissance, names=c("GPS", "Carte"))



Sur ces boxplots on peut voir une distribution différente entre les deux échantillons : la médiane des Cartes est plus élevée que la médiane des GPS. Les intervalles de confiance autour de la médiane montrent une tendance : les cartes ont 95% de score entre 1 et 4, contre 0 et 3 pour les GPS. Ces deux constatations sont en faveurs de l'hypothèseque les participants “Cartes” reconnaissent plus de repères que les participants “GPS”.

c. *Statistiques inférentielles*

Normalité des échantillons

Nous allons tout d'abord tester la normalité des deux échantillons par le test de Shapiro. Pour ce faire on pose l'hypothèse nulle H0 tel que chaque échantillon suit une loi normale; contre l'hypothèse H1 que les échantillons ne suivent pas une loi normale.

```
> shapiro.test(GPS_reconnaissance)

	Shapiro-Wilk normality test

data:  GPS_reconnaissance
W = 0.90919, p-value = 0.3903
```

```
> shapiro.test(Carte_reconnaissance)

	Shapiro-Wilk normality test

data:  Carte_reconnaissance
W = 0.96618, p-value = 0.8603
```

Pour les “Cartes” p-value = 0.3903 > 0.05 = alpha. On ne rejette donc pas H0. L'échantillon des participants “Cartes” suit une loi normale avec une risque de première espèce, ou erreur alpha de 5% de se tromper. Pour les “GPS” p-value = 0.2685 > 0.05 = alpha. On ne rejette pas H0. L'échantillon des participants "GPS" suit une loi normale à un avec une risque de première espèce, ou erreur alpha de 5% de se tromper.

Test d'égalité des variances

Après avoir testé leur normalité nous allons tester l'égalité des variances. Pour ce faire on pose l'hypothèse H0 tel que les variances sont égales contre l'hypothèse H1 les variances sont différentes.

```
> var.test(GPS_reconnaissance, Carte_reconnaissance)

	F test to compare two variances

data:  GPS_reconnaissance and Carte_reconnaissance
F = 1.0456, num df = 6, denom df = 8, p-value = 0.9256
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.2247767 5.8549250
sample estimates:
ratio of variances
          1.045593
```

On peut voir que la p-value = 0.9256 >> 0.05 = alpha. Donc on ne rejette pas H0 donc les variances sont égales entre elle. Le ratio des deux variances est égale à 1.

Test d'égalité des moyennes

Après avoir testé l'égalité des variances nous allons tester l'égalité des moyennes. Pour ce faire on pose l'hypothèse H0 tel que les moyennes sont égales contre l'hypothèse H1 la moyenne des scores des "Cartes" est plus grandes que la moyenne des scores des "GPS".

```
>reconnaissance, GPS_reconnaissance, var.equal=TRUE, alternative="greater")

	Two Sample t-test

data:  Carte_reconnaissance and GPS_reconnaissance
t = 0.62791, df = 14, p-value = 0.2701
alternative hypothesis: true difference in means is greater than 0
95 percent confidence interval:
 -1.976941       Inf
```

sample estimates:

mean of x mean of y

2.666667 1.571429

Ici la p-value =0,2701 > 0,05. Donc on ne rejette pas H0 : les moyennes ne sont pas différentes de façon significative.

d. Conclusion

Ainsi pour répondre à la question posée, nous avons rejeté l'hypothèse nulle H0 et accepté H1 tel que l'échantillon "Cartes" n'a pas significativement un score plus élevé que l'échantillon GPS. On peut donc en conclure qu'il n'y a pas de différence significative de score entre les deux échantillons, que la moyenne de score des "Cartes" n'est pas plus élevée que celle des "GPS", avec un risque de second espèce, ou erreur alpha de se tromper.

3. Tâche de série

Le nombre de bon enchainement est liés au nombre de bon repère trouvé. Ce qui veut dire que cette série est dépendant de la série précédente. Or nos calculs ont permis de faire un ratio sur le nombre total de bonne réponse qu'aurait dû trouver le participant (voir fiche explication de calcul. Ce qui veut dire que les données des séries sont traités de la même façon que les données précédentes.

a. Hypothèse

Ces deux échantillons sont indépendants l'un de l'autre. L'hypothèse que l'on se pose alors est : les participants "cartes" ont ils un meilleur score que les participants "GPS"? Soit l'hypothèse nulle tel qu'il n'y a pas de différence significative entre les deux échantillons; contre l'hypothèse H1 tel que les participants "cartes" retient significativement plus l'ordre des repères que les participants "GPS".

b. Statistiques descriptives

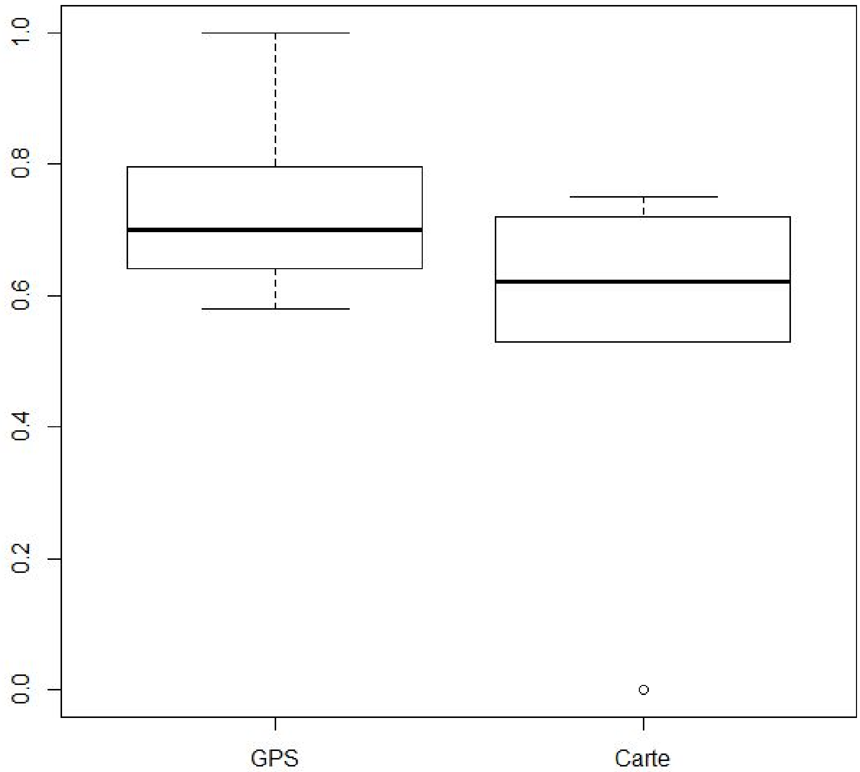
```
> summary(GPS_serie)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
 0.5800  0.6400  0.7000  0.7357  0.7950  1.0000
> var(GPS_serie)
[1] 0.02229524
> sd(GPS_serie)
[1] 0.1493159
```

```
> summary(Carte_serie)
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.0000  0.5300  0.6200  0.5078  0.7200  0.7500
> var(Carte_serie)
[1] 0.08974444
```

```
> sd(Carte_serie)
[1] 0.2995738
```

D'après ces commandes on peut voir que la moyenne des scores pour l'échantillon Carte est de 0,5 contre 0,7 pour l'échantillon GPS. De plus la variance et l'écart type sont plus élevés pour l'échantillon Carte que pour l'échantillon GPS; respectivement 0,08 et 0,29 contre 0,02 et 0,14. Ce qui veut dire que l'échantillon Carte est plus étendu. Ainsi la question de la différence de score ne se pose pas dans le même sens : on peut supposer ici que l'échantillon GPS retient significativement plus l'ordre des repères que les participants Carte.

```
> boxplot(GPS_serie, Carte_serie, names=c("GPS", "Carte"))
```



Sur ces boxplots on peut voir une distribution différente entre les deux échantillons : la médiane des Cartes est moins élevée que la médiane des GPS. Les intervalles de confiance autour de la médiane montrent une tendance : les cartes ont 95% de score entre 0.5 et 0.7, contre 0.65 et 0.8 pour les GPS. Ces deux constatations ne sont pas en faveurs de l'hypothèse que les participants “Cartes” retiennent plus l'ordre des repères que les participants “GPS”.

c. Statistiques inférentielles

Normalité des échantillons

Nous allons tout d'abord tester la normalité des deux échantillons par le test de Shapiro. Pour ce faire on pose l'hypothèse nulle H0 tel que chaque échantillon suit une loi normale; contre l'hypothèse H1 que les échantillons ne suivent pas une loi normale.

```
> shapiro.test(GPS_serie)

	Shapiro-Wilk normality test

data:  GPS_serie

W = 0.90295, p-value = 0.3492
```

```
> shapiro.test(Carte_serie)

	Shapiro-Wilk normality test

data:  Carte_serie

W = 0.75384, p-value = 0.006007
```

Pour les “Cartes” p-value = 0.006007 < 0.05 = alpha. On rejette donc H0. L'échantillon des participants “Cartes” ne suit pas une loi normale avec une risque de seconde espèce, ou erreur betha de se tromper. Ce qui veut dire qu'il y a une valeur prédominante. Pour les “GPS” p-value = 0.3492 > 0.05 = alpha. On ne rejette pas H0. L'échantillon des participants “GPS” suit une loi normale à un avec une risque de première espèce, ou erreur alpha de 5% de se tromper.

Test d'égalité des moyennes

Après avoir testé l'égalité des variances nous allons tester l'égalité des moyennes. Pour ce faire on pose l'hypothèse H0 tel qu'il n'y a pas de différence entre les moyennes contre l'hypothèse H1 il y a une différence entre les moyennes.

```
> wilcox.test(Carte_serie, GPS_serie, alternative="greater")

	Wilcoxon rank sum test with continuity correction

data:  Carte_serie and GPS_serie

W = 19, p-value = 0.9164

alternative hypothesis: true location shift is greater than 0

Warning message:

In wilcox.test.default(Carte_serie, GPS_serie, alternative = "greater") :
```

impossible de calculer la p-value exacte avec des ex-aequos

Ici la p-value =0,9164 > 0,05. Donc on ne rejette pas H0 : les moyennes ne sont pas différentes de façon significative avec un risque de première espèce de 5% de se tromper.

d. Conclusion

Ainsi pour répondre à la question posée, nous avons rejeté l'hypothèse nulle H0 et accepté H1 tel que l'échantillon "Cartes" n'a pas significativement un score plus élevé que l'échantillon GPS. On peut donc en conclure qu'il n'y a pas de différence significative de score entre les deux échantillons, que la moyenne de score des "Cartes" n'est pas plus élevée que celle des "GPS", avec un risque de second espèce, ou erreur betha de se tromper.

4. Tâche de positionnement

a. Hypothèse

Ces deux échantillons sont indépendants l'un de l'autre. L'hypothèse que l'on se pose alors est : les participants "cartes" ont ils un meilleur score que les participants "GPS"? Soit l'hypothèse nulle tel qu'il n'y a pas de différence significative entre les deux échantillons; contre l'hypothèse H1 tel que les participants "cartes" placent significativement plus de repère sur la carte que les participants "GPS".

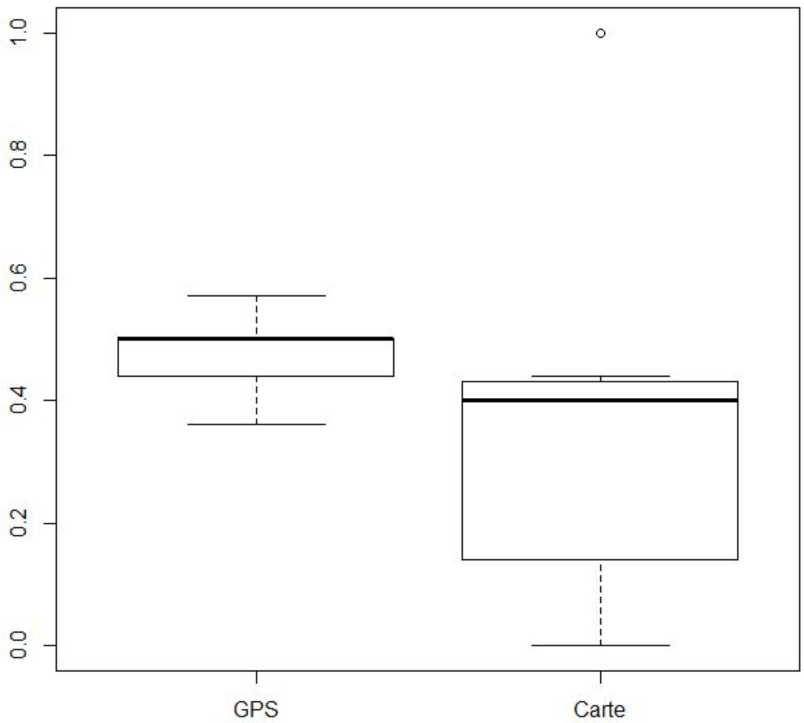
b. Statistiques descriptives

```
> summary(GPS_position)
  Min. 1st Qu.  Median    Mean 3rd Qu. 
Max.
 0.3600  0.4400  0.5000  0.4729  0.5000  0.5700
> var(GPS_position)
[1] 0.00442381
> sd(GPS_position)
[1] 0.06651172
```

```
> summary(Carte_position)
   Min. 1st Qu.  Median    Mean 3rd Qu. 
Max.
 0.0000  0.1400  0.4000  0.3422  0.4300 
1.0000
> var(Carte_position)
[1] 0.08806944
> sd(Carte_position)
[1] 0.296765
```

D'après ces commandes on peut voir que la moyenne des scores pour l'échantillon Carte est de 0,4 contre 0,5 pour l'échantillon GPS. De plus la variance et l'écart type sont plus élevés pour l'échantillon Carte que pour l'échantillon GPS; respectivement 0,08 et 0,29 contre 0,004 et 0,066. Ce qui veut dire que l'échantillon Carte est plus étendu. Ainsi la question de la différence de score ne se pose pas dans le même sens : on peut supposer ici que l'échantillon GPS retient significativement plus la position dans l'espace des repères que les participants Carte.

```
> boxplot(GPS_position, Carte_position, names=c("GPS", "Carte"))
```



Sur ces boxplots on peut voir une distribution différente entre les deux échantillons : la médiane des Cartes est moins élevée que la médiane des GPS. Les intervalles de confiance autour de la médiane montrent une tendance : les cartes ont 95% de score entre 0.1 et 0.43, contre 0.44 et 0.5 pour les GPS. Ces deux constatations ne sont pas en faveurs de l'hypothèse que les participants “Cartes” retiennent plus la position des repères dans l'espace que les participants “GPS”.

c. *Statistiques inférentielles*

Normalité des échantillons

Nous allons tout d'abord tester la normalité des deux échantillons par le test de Shapiro. Pour ce faire on pose l'hypothèse nulle H0 tel que chaque échantillon suit une loi normale; contre l'hypothèse H1 que les échantillons ne suivent pas une loi normale.

```
> shapiro.test(GPS_position)
        Shapiro-Wilk normality test
data:  GPS_position
W = 0.93459, p-value = 0.5906
```

```
> shapiro.test(Carte_position)
        Shapiro-Wilk normality test
data:  Carte_position
W = 0.86475, p-value = 0.108
```

Pour les “Cartes” p-value = 0.5906 > 0.05 = alpha. On ne rejette donc pas H0. L’échantillon des participants “Cartes” suit une loi normale avec une risque de première espèce, ou erreur alpha de 5% de se tromper. Pour les “GPS” p-value = 0.108 > 0.05 = alpha. On ne rejette pas H0. L’échantillon des participants “GPS” suit une loi normale à un avec une risque de première espèce, ou erreur alpha de 5% de se tromper.

Test d’égalité des variances

Après avoir testé leur normalité nous allons tester l’égalité des variances. Pour ce faire on pose l’hypothèse H0 tel que les variances sont égales contre l’hypothèse H1 les variances sont différentes.

```
> var.test(GPS_position, Carte_position)

	F test to compare two variances

data:  GPS_position and Carte_position
F = 0.050231, num df = 6, denom df = 8, p-value = 0.001762
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.01079841 0.28127423
sample estimates:
ratio of variances
	0.05023092
```

On peut voir que la p-value = 0.001 << 0.05 = alpha. Donc on rejette H0 donc les variances ne sont pas égales entre elle. Le ratio des deux variances n’est pas égale à 1.

Test d’égalité des moyennes

Après avoir testé l'égalité des variances nous allons tester l’égalité des moyennes. Pour ce faire on pose l’hypothèse H0 tel que les moyennes sont égales contre l’hypothèse H1 la moyenne des scores des “Cartes” est plus grandes que la moyenne des scores des “GPS”.

```
> t.test(Carte_reconnaissance, GPS_reconnaissance, var.equal=FALSE)

	Welch Two Sample t-test
```

data: Carte_reconnaissance and GPS_reconnaissance

t = 0.62604, df = 12.895, p-value = 0.5422

alternative hypothesis: true difference in means is not equal to 0

95 percent confidence interval:

-2.687402 4.877878

sample estimates:

mean of x mean of y

2.666667 1.571429

Ici la p-value =0,5422 > 0,05. Donc on ne rejette pas H0 : les moyennes ne sont pas différentes de façon significative.

d. Conclusion

Ainsi pour répondre à la question posée, nous avons rejeté l'hypothèse nulle H0 et accepté H1 tel que l'échantillon "Cartes" n'a pas significativement un score plus élevé que l'échantillon GPS. On peut donc en conclure qu'il n'y a pas de différence significative de score entre les deux échantillons, que la moyenne de score des "Cartes" n'est pas plus élevée que celle des "GPS", avec un risque de second espèce, ou erreur alpha de se tromper.

III. Conclusion

Ainsi il a été montré que la carte ne permet pas de retenir mieux les repères que le GPS que ce soit dans leur forme, dans leur ordre ou dans leur position. La carte et le GPS sont deux outils qui permettent de se déplacer tout en apprenant un nouvel espace. La carte n'est pas un outil qui permet de mieux retenir les repères contrairement à ce que nous avions pensé.