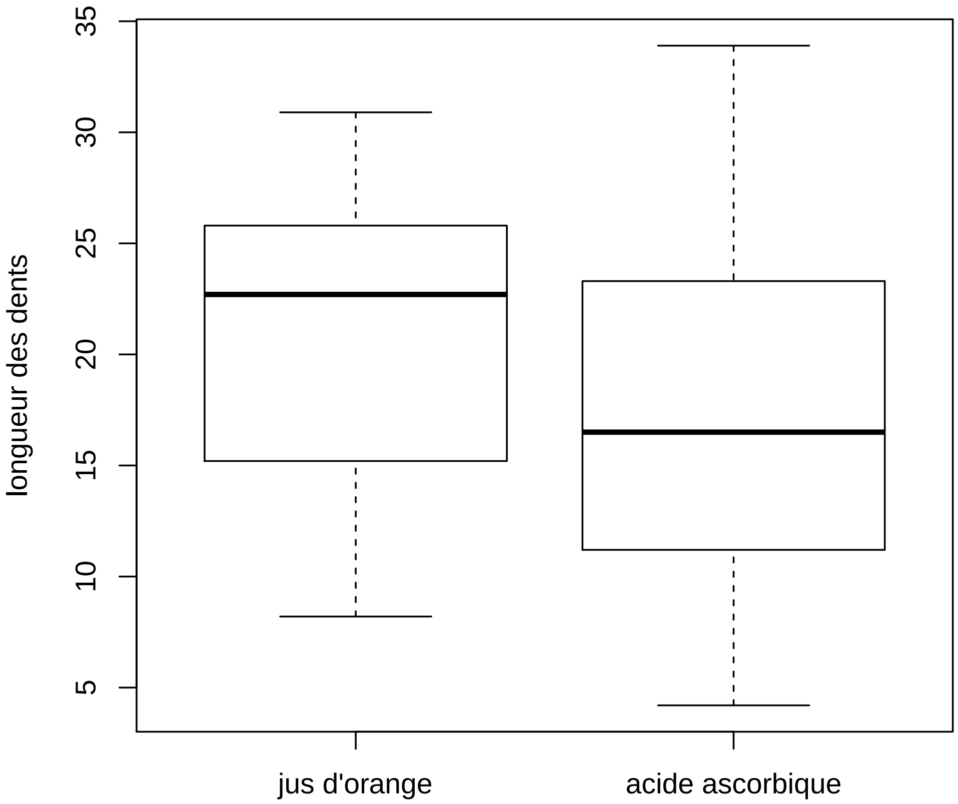
U.E.V.E.
Master 1 SGO Démarche Statistique 1

2013-2014

Devoir Surveillé : 21 novembre 2013
Durée : 2h
Corrigé du premier exercice

Consignes :
- Aucun document n'est autorisé
- Les tests demandés seront effectués au niveau $\alpha = 5\%$.
- Une bonne rédaction est un élément d'appréciation très important pour les copies.
- **Justifier soigneusement toutes vos réponses**.

Exercice 1 (Cochons d'inde)**.** On s'intéresse à l'effet de l'absorption de vitamine C sur la croissance des dents de cochons d'inde (). On considère deux modes d'administration de la vitamine C (jus d'orange ou acide ascorbique). Pour chaque mode d'administration, on mesure la longueur des dents de 30 cochons d'inde.



1. Commenter le graphique ci-dessus (boîte à moustaches).

   On observe les longueurs des dents dans deux échantillons de cochons d'inde ayant reçu de la vitamine C au moyen de jus d'orange (JO) ou d'acide ascorbique (AA). Les boxplots permettent de comparer la longueur des dents dans les échantillons JO et AA.

   On constate que la médiane (trait horizontal gras) est supérieure pour l'échantillon traité au jus d'orange. Par ailleurs, la dispersion de l'échantillon traité à l'acide ascorbique est *un peu plus importante que celle de l'échantillon traité au jus d'orange (largeur de la "boîte" et distance entre les "moustaches" plus importantes).*

2. A l'aide des graphiques ci-dessous (histogrammes) :

   (a) donner le nombre de  dont les dents mesurent moins de 15 mm

   Nombre de  *dont les dents mesurent moins de 15mm : 5+2=7 pour ceux traités au jus d'orange d'après l'histogramme de gauche, et 5+6+1=12 d'après l'histogramme de droite. En tout, 7+12=19*  *ont des dents de moins de 15mm dans l'échantillon.*

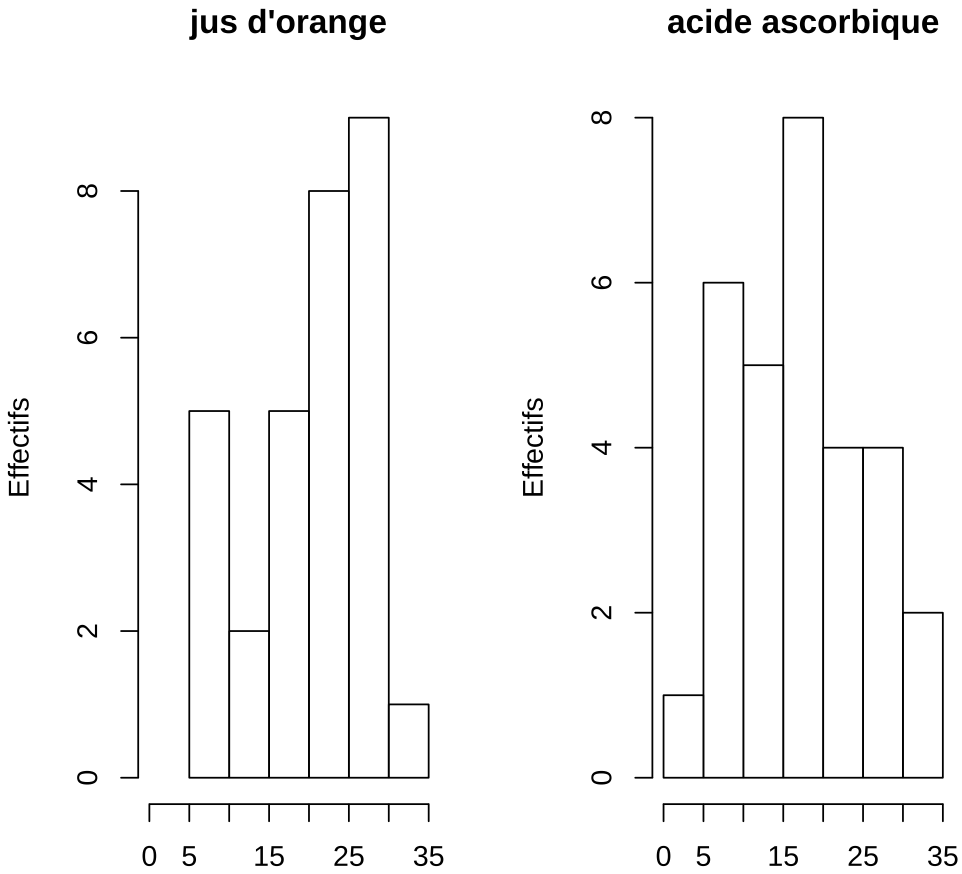
   (b) donner (sous forme de fraction) la proportion de  traités au jus d'orange parmi ceux dont les dents mesurent entre 20 et 30 mm

   Le nombre de  *dont les dents mesurent entre 20 et 30 mm est 8+9=17 pour ceux traités au jus d'orange, et 4+4=8 pour ceux traités à l'acide ascorbique. Donc le nombre total de*  *dont les dents mesurent entre 20 et 30 mm est 17+8=25, et la proportion cherchée est 17/25.*

   (c) la longueur de dents médiane des  dans l'échantillon est-elle supérieure pour ceux traités au jus d'orange ou à l'acide ascorbique ?

   D'après les histogrammes, la médiane de l'échantillon JO est située entre 20 et 25, et celle de l'échantillon AA est située entre 15 et 20.

   (d) la longueur de dents moyenne des  dans l'échantillon est-elle supérieure pour ceux traités au jus d'orange ou à l'acide ascorbique ?



3. On cherche à savoir si la longueur de dents moyenne des  dans la population est différente pour ceux traités au jus d'orange ou à l'acide ascorbique.

   (a) Préciser l'hypothèse nulle testée et l'alternative.

      – *hypothèse nulle : la longueur de dents moyenne des*  *dans la population est identique pour ceux traités au jus d'orange ou à l'acide ascorbique.*

      – *hypothèse alternative : la longueur de dents moyenne des*  *dans la population est différente pour ceux traités au jus d'orange ou à l'acide ascorbique.*

   (b) Expliciter l'erreur de première espèce et l'erreur de deuxième espèce dans ce contexte.

      – *erreur de première espèce : rejeter à tort l'hypothèse nulle, c'est-à-dire conclure que la longueur de dents moyenne des*  *dans la population est différente pour ceux traités au jus d'orange ou à l'acide ascorbique alors qu'en réalité elle ne l'est pas.*

      – *erreur de deuxième espèce : ne pas rejeter l'hypothèse nulle alors qu'elle est fausse, c'est-à-dire conclure que la longueur de dents moyenne des*  *dans la population est identique pour ceux traités au jus d'orange ou à l'acide ascorbique alors qu'en réalité elle est différente.*

   (c) Les longueurs des dents sont stockés dans la variable `x` pour le jus d'orange, et `y` pour l'acide ascorbique. Le(s)quel(s) des tests ci-dessous vous semble(nt) justifié(s) pour répondre à cette question ? *Les deux échantillons ne sont pas appariés : il s'agit de* *deux groupes de*  *qui n'ont rien à voir l'un avec l'autre. On peut donc faire un test de Wilcoxon sur échantillons indépendants :*

`wilcox.exact(x, y, paired=FALSE)`

On a 30 observations par groupe, on peut donc admettre (d'après le théorème de la limite centrale) que la statistique du test de Student suit sous l'hypothèse nulle une loi normale. Le test sur les variances (var.test) donne une p-valeur de 0.2331*, supérieure au seuil de significativité choisi (*$\alpha = 0.05$*), donc on peut supposer que les variances sont identiques dans les deux populations. Par conséquent, on peut également appliquer le test de Student sur échantillons indépendants avec variances identiques :*

`t.test(x, y, paired=FALSE, var.equal=TRUE)`

(d) Utiliser les sorties R ci-dessous pour donner et commenter le résultat de ce(s) test(s).
Pour les deux tests indiqués à la question précédente, la p-valeur est supérieure au seuil de significativité choisi ($\alpha = 0.05$*) :* $p = 0.06366$ *pour le test de Wilcoxon et* $p = 0.06039$ *pour le test de Student. Dans les deux cas on ne rejette donc pas l'hypothèse nulle. Ainsi, on conclut que les longueurs des dents moyennes (dans la population) des*  *traités au JO et à l'AA ne sont pas significativement différentes (au seuil de significativité de 5%).*

```
> wilcox.exact(x, y, paired=TRUE)

        Exact Wilcoxon signed rank test

data:  x and y
V = 326.5, p-value = 0.05291
alternative hypothesis: true mu is not equal to 0

> wilcox.exact(x, y, paired=FALSE)

        Exact Wilcoxon rank sum test

data:  x and y
W = 575.5, p-value = 0.06366
alternative hypothesis: true mu is not equal to 0

> t.test(x, y, paired=TRUE)

        Paired t-test

data:  x and y
t = 1.7489, df = 29, p-value = 0.09089
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.6269133  8.0269133
sample estimates:
mean of the differences
                    3.7

> t.test(x, y, paired=FALSE, var.equal=TRUE)
```

```
        Two Sample t-test

data:  x and y
t = 1.9153, df = 58, p-value = 0.06039
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1670064  7.5670064
sample estimates:
mean of x mean of y
 20.66333  16.96333

> t.test(x, y, paired=FALSE, var.equal=FALSE)

        Welch Two Sample t-test

data:  x and y
t = 1.9153, df = 55.309, p-value = 0.06063
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -0.1710156  7.5710156
sample estimates:
mean of x mean of y
 20.66333  16.96333

> var.test(x, y)

        F test to compare two variances

data:  x and y
F = 0.6386, num df = 29, denom df = 29, p-value = 0.2331
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.3039488 1.3416857
sample estimates:
ratio of variances
         0.6385951
```

4. Les  de l'expérience n'ont en fait pas tous eu la même dose de vitamine C. On s'intéresse maintenant au sous-groupe de 20  ayant reçu 1 mg de vitamine C : 10 ont été traités au jus d'orange, 10 à l'acide ascorbique. On cherche à savoir si la longueur de dents moyenne des  dans la population est différente pour ceux traités au jus d'orange ou à l'acide ascorbique pour cette dose.
   (a) Préciser l'hypothèse nulle testée et l'alternative.
      – *hypothèse nulle : la longueur de dents moyenne des*  *dans la population est identique pour ceux traités avec 1mg de jus d'orange ou 1 mg à l'acide ascorbique.*
      – *hypothèse alternative : la longueur de dents moyenne des*  *dans la population est différente pour ceux traités avec 1 mg jus d'orange ou 1 mg d'acide ascorbique.*
   (b) Expliciter l'erreur de première espèce et l'erreur de deuxième espèce dans ce contexte.

- *erreur de première espèce : rejeter à tort l'hypothèse nulle, c'est-à-dire conclure que la longueur de dents moyenne des*  *dans la population est différente pour ceux traités avec 1 mg de jus d'orange ou 1mg d'acide ascorbique alors qu'en réalité elle ne l'est pas.*
- *erreur de deuxième espèce : ne pas rejeter l'hypothèse nulle alors qu'elle est fausse, c'est-à-dire conclure que la longueur de dents moyenne des*  *dans la population est identique pour ceux traités avec 1 mg de jus d'orange ou 1 mg d'acide ascorbique alors qu'en réalité elle est différente.*

(c) Les longueurs des dents sont stockés dans la variable `x` pour le jus d'orange, et `y` pour l'acide ascorbique. Le(s)quel(s) des tests ci-dessous vous semble(nt) justifié(s) pour répondre à cette question ? *Les deux échantillons ne sont toujours pas appariés : il s'agit de deux groupes de*  *qui n'ont rien à voir l'un avec l'autre. On peut donc faire un test de Wilcoxon sur échantillons indépendants :*

```
wilcox.exact(x, y, paired=FALSE)
```

On a seulement 10 observations par groupe. On ne peut donc appliquer le test de Student que si l'on adment que les longueurs des dents des  *traités au JO ou à l'AA suivent chacune une loi normale. Si on l'admet. Le test sur les variances (var.test) donne une p-valeur de* 0.2046*, supérieure au seuil de significativité choisi (*$\alpha = 0.05$*), donc on peut supposer que les variances sont identiques dans les deux populations. Par conséquent, si on admet l'hypothèse de normalité*[1]*, on peut également appliquer le test de Student sur échantillons indépendants avec variances identiques :*

```
t.test(x, y, paired=FALSE, var.equal=TRUE)
```

(d) Utiliser les sorties `R` ci-dessous pour donner et commenter le résultat de ce(s) test(s). *Pour les deux tests indiqués à la question précédente, la p-valeur est supérieure au seuil de significativité choisi (*$\alpha = 0.05$*) :* $p = 0.06366$ *pour le test de Wilcoxon et* $p = 0.06039$ *pour le test de Student. Dans les deux cas on ne rejette donc pas l'hypothèse nulle. Ainsi, on conclut que les longueurs des dents moyennes (dans la population) des*  *traités avec 1 mg de JO sont significativement différentes de celles des*  *traités avec 1 mg d'AA (au seuil de significativité de* 5%*).*

```
> wilcox.exact(x, y, paired=TRUE)

        Exact Wilcoxon signed rank test

data:  x and y
V = 51, p-value = 0.01367
alternative hypothesis: true mu is not equal to 0

> wilcox.exact(x, y, paired=FALSE)

        Exact Wilcoxon rank sum test

data:  x and y
W = 88.5, p-value = 0.00223
alternative hypothesis: true mu is not equal to 0
```

1. Note du correcteur : je suis OK pour que vous fassiez cette hypothèse, du moment que vous le dites.

```
> t.test(x, y, paired=TRUE)

        Paired t-test

data:  x and y
t = 3.3721, df = 9, p-value = 0.008229
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 1.951911 9.908089
sample estimates:
mean of the differences
                   5.93

> t.test(x, y, paired=FALSE, var.equal=TRUE)

        Two Sample t-test

data:  x and y
t = 4.0328, df = 18, p-value = 0.0007807
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.840692 9.019308
sample estimates:
mean of x mean of y
    22.70     16.77

> t.test(x, y, paired=FALSE, var.equal=FALSE)

        Welch Two Sample t-test

data:  x and y
t = 4.0328, df = 15.358, p-value = 0.001038
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 2.802148 9.057852
sample estimates:
mean of x mean of y
    22.70     16.77

> var.test(x, y)

        F test to compare two variances

data:  x and y
F = 2.4176, num df = 9, denom df = 9, p-value = 0.2046
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.6004952 9.7332038
sample estimates:
ratio of variances
           2.41759
```