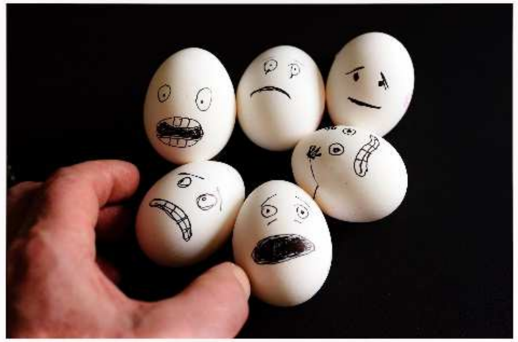
Correction exercices 4 à 8

Comparaison de 2 échantillons indépendants

4. Effet d'un fongicide sur la reproduction des oiseaux



On cherche à tester l'effet d'un fongicide au mercure sur la reproduction des oiseaux.

Un lot d'oiseaux est traité par voie orale et l'autre lot est non traité.

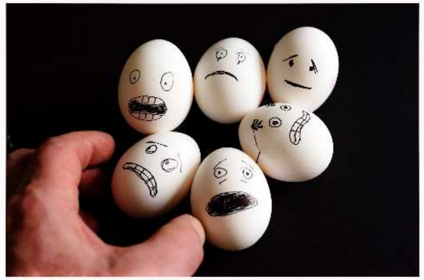
Les paramètres des épaisseurs de coquilles d'œufs sont les suivants :

- Lot traité : $n_1 = 30$; $\overline{x_1} = 284$; $S_1{}^* = 30$
- Lot non traité : $n_2 = 40$; $\overline{x_2} = 305$; $S_2{}^* = 31$

Le traitement au fongicide a-t-il un effet sur l'épaisseur moyenne des coquilles d'oeufs pondus par ces oiseaux ?

<u>NB: La condition d'homoscédasticité est considérée comme respectée.</u>

4. Effet d'un fongicide sur la reproduction des oiseaux



Les données :

- Lot traité : n_1 = 30 ; $\overline{x_1}$ = 284 ; $S_1{}^*$ = 30
- Lot non traité : n_2 = 40 ; $\overline{x_2}$ = 305 ; $S_2{}^*$ = 31

Les deux échantillons sont grands (n_1 et $n_2 \geq 30$) et homoscédasticité OK ➔ Test paramétrique

H_0 : $\overline{x_1} = \overline{x_2}$ (soit $\overline{x_1} - \overline{x_2} = 0$)
H_1 : $\overline{x_1} \neq \overline{x_2}$ (soit $\overline{x_1} - \overline{x_2} = 0$)

$$T = \frac{\overline{X_1} - \overline{X_2}}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$T_{obs} = \frac{284 - 305}{\sqrt{1077.6\left(\frac{1}{30}+\frac{1}{40}\right)}} = \frac{-21}{\sqrt{62.86}} = \frac{-21}{7.93} = -2.65$$

Avec $\hat{\sigma}^2 = \frac{(30-1) * 900 + (40-1) * 961}{30+31-2} = 1077.6$

$T_{théo}$ lu dans la table de Student (α = 0.05= : 1.99 pour ddl=80 et 2.021 ddl=40. Notre ddl= 59 se situe entre les 2.

$|T_{obs}| > T_{théo}$ quel que soit le ddl (= 80 ou =40) au seuil de 5%
On accepte H_1 au seuil de 5%

→ il y a une différence significative de l'épaisseur de coquille d'œufs entre le lot traité et le lot non traité.
On conclue à un effet du fongicide.

5. Dureté d'un alliage

Un alliage formé de zinc, cuivre et étain est produit à 2 températures différentes (H = haute et B = basse).

On produit 9 lots à température basse et 7 lots à haute température pour tester leur dureté.

On attribue ensuite un indice de dureté à chaque lot:

tf = très faible , **f =** faible , **m =** moyen , **F =** Fort , **TF =** Très Fort.

T° C	H	B	B	B	H	H	B	H	H	B	B	B	B	B	H	H
dureté	tf	f	m	m	f	f	TF	m	F	tf	F	F	TF	TF	tf	f

Est-ce que la température de production affecte la dureté de l'alliage?

5. Dureté d'un alliage

Comparaison 2 échantillons indépendants à petits effectifs avec des variables qualitatives ordinales

→ Test non paramétrique de Wilcoxon-Mann-Whitney pour les petits échantillons

H_0 : dureté non différentes selon la température ($U_1 = U_2$)
H_1 : dureté différentes selon la température ($U_1 \neq U_2$)

Dureté	tf	tf	tf	f	f	f	f	m	m	m	F	F	F	TF	TF	TF
température	haute	haute	basse	basse	haute	haute	haute	basse	basse	haute	haute	basse	basse	basse	basse	basse
rang	2	2	2	5.5	5.5	5.5	5.5	9	9	9	12	12	12	15	15	15
	*			*				*			*			*		

* *Rappel : rang moyen en cas d'ex-aequo!*

$R_H = 2 + 2 + 5.5 + 5.5 + 5.5 + 9 + 12 = 41.5$

$R_B = 2 + 5.5 + 9 + 9 + 12 + 12 + 15 + 15 + 15 = 94.5$

$$U_H = 41.5 - \frac{7(7+1)}{2} = 13.5$$

$$U_B = 94.5 - \frac{9(9+1)}{2} = 49.5$$

➔ **$U_{WMWobs} = 13.5$**

5. Dureté d'un alliage

Si $U_{WMWobs} \leq U_{WMWthéo}$

alors **on rejette H_0** au risque α

TABLE DE MANN-WHITNEY

Valeurs critiques (U_{crit}) à comparer avec la valeur observée (U_{obs}) à partir de vos 2 échantillons p
un test bilatéral au seuil α = 0.05 ou 0.01.

NB : n_1 et n_2 représentent le nombre d'observations dans chaque échantillon.

n_H = 7, n_B = 9 , U_{WMWobs} = 13.5

Table → α = 0.05 (bilatéral) on lit $U_{WMWthéo}$ = 12

n_B = 9

n_H = 7

n_2	α	n_1 = 3	4	5	6	7	8	9	10	11	12	13	14
3	.05	--	0	0	1	1	2	2	3	3	4	4	5
	.01	--	0	0	0	0	0	0	0	0	1	1	1
4	.05	--	0	1	2	3	4	4	5	6	7	8	9
	.01	--	--	0	0	0	1	1	2	2	3	3	4
5	.05	0	1	2	3	5	6	7	8	9	11	12	13
	.01	--	--	0	1	1	2	3	4	5	6	7	7
6	.05	1	2	3	5	6	8	10	11	13	14	16	17
	.01	--	0	1	2	3	4	5	6	7	9	10	11
7	.05	1	3	5	6	8	10	$\boxed{12}$	14	16	18	20	22
	.01	--	0	1	3	4	6	7	9	10	12	13	15

$U_{WMWobs} > U_{WMWthéo}$

➔ **On ne rejette pas H_0** : avec ce plan d'experience et ces tailles d'échantillons, on ne peut pas conclure à une différence de dureté de l'alliage selon la T° C.

Effectuer ce test sous R

NB : R ne comprendra pas la hiérarchie des variables semi-quantitatives allant de très faible (tf) à très forte (TF)…

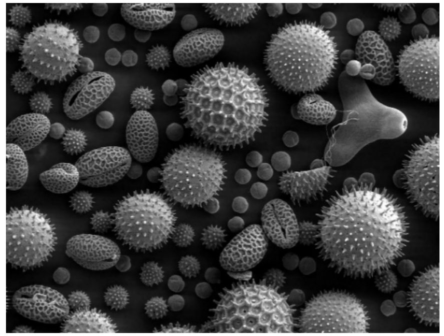
→ Astuce : pour contrer ce problème serait d'attribuer des valeurs pour chaque niveau de dureté.

```
# on recode la variable de dureté :
# tf=1, f=2, m=3, F=4, TF=5
Haute_T<-c(1, 2, 2, 3, 4, 1, 2)
Basse_T<-c(2, 3, 3, 5, 1, 4, 4, 5, 5)

wilcox.test(Haute_T,Basse_T)

        Wilcoxon rank sum test with continuity correction

  data:  Haute_T and Basse_T
  W = 13.5, p-value = 0.05891
  alternative hypothesis: true location shift is not equal to 0
```

→ Résultat non significatif…mais pas très loin de l'être…

#reflexionsurlapertinencedecettevaleurdeseuilalphaarbitrairementdéfinie!

6. Le pollen dans les sédiments



Soit deux prélèvements de sédiments provenant de deux zones continentales différentes, on étudie la concentration en pollen dans les sédiments (µg L^{-1}) :

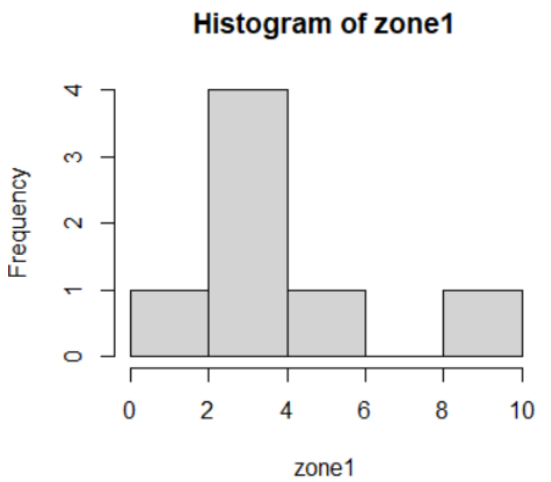
Zone 1 : 2.1, 3.6, 3.1, 4.6, 1.9, 10.0, 2.1

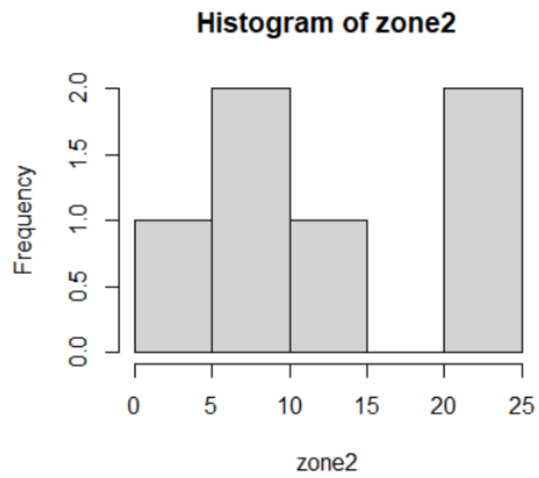
Zone 2 : 4.2, 8.1, 20.3, 9.2, 12.7, 20.7

Les concentrations en pollen sont elles significativement différentes entre les deux zones ?

On dispose des informations suivantes :

```
        Shapiro-Wilk normality test

data:  zone1
W = 0.7401, p-value = 0.009987


        Shapiro-Wilk normality test

data:  zone2
W = 0.9052, p-value = 0.4056
```





6. Le pollen dans les sédiments

2 très petits échantillons indépendants (n_1 et $n_2 < 9$) qui ne suivent pas une loi normale

→ test non paramétrique de Wilcoxon Mann-Whitney pour petits échantillons

$H_0 : U_1 = U_2$
$H_1 : U_1 \neq U_2$

Ordre croissant	1.9	2.1	2.1	3.1	3.6	4.2	4.6	8.1	9.2	10	12.7	20.3	20.7
Reference	zone 1	zone 1	zone 1	zone 1	zone 1	zone 2	zone 1	zone 2	zone 2	zone 1	zone 2	zone 2	zone 2

Rangs zone1	1	2.5	2.5	4	5		7			10			
Rangs zone2						6		8	9		11	12	13

$R_1 = 1+2.5+2.5+4+5+7+10 = 32$

$R_2 = 6 + 8 + 9 + 11 + 12 + 13 = 59$

$U_1 = 32 - \frac{7(7+1)}{2} = 4$

$U_2 = 59 - \frac{6(6+1)}{2} = 38$

→ **Vérification de la propriété :** $\mathbf{U_1 + U_2 = n_1 * n_2}$: *4 + 38 = (7 * 6) = 42*

$U_{obs} = U_1 = 4$ (plus petite valeur entre $\mathbf{U_1}$ et $\mathbf{U_2}$)

6. Le pollen dans les sédiments

Si $U_{WMWobs} \leq U_{WMWthéo}$

alors **on rejette H_0** au risque α

TABLE DE MANN-WHITNEY

Valeurs critiques (U_{crit}) à comparer avec la valeur observée (U_{obs}) à partir de vos 2 échantillons p
un test bilatéral au seuil α = 0.05 ou 0.01.

NB : n_1 et n_2 représentent le nombre d'observations dans chaque échantillon.

n_1 = 7, n_2 = 6 , U_{WMWobs} = 4

Table (pour α = 0.05 (bilatéral)) → $U_{WMWthéo}$ = 6

$U_{WMWobs} < U_{WMWthéo}$

→ Rejet de H_0 au risque α=0.05 :
la différence de concentration en pollen est significative entre les 2 zones

(mais ne le serait pas pour α=0.01 avec $U_{théo}$ = 3)

n_1 = 7

n_2 = 6

n_2	α	n_1											
		3	4	5	6	7	8	9	10	11	12	13	14
3	.05	--	0	0	1	1	2	2	3	3	4	4	5
	.01	--	0	0	0	0	0	0	0	0	1	1	1
4	.05	--	0	1	2	3	4	4	5	6	7	8	9
	.01	--	--	0	0	0	1	1	2	2	3	3	4
5	.05	0	1	2	3	5	6	7	8	9	11	12	13
	.01	--	--	0	1	1	2	3	4	5	6	7	7
6	.05	1	2	3	5	$\boxed{6}$	8	10	11	13	14	16	17
	.01	--	0	1	2	3	4	5	6	7	9	10	11
7	.05	1	3	5	6	8	10	12	14	16	18	20	22
	.01	--	0	1	3	4	6	7	9	10	12	13	15

Effectuer ce test sous R

```
zone1<-c(2.1, 3.6, 3.1,4.6,1.9,10,2.1)
zone2<-c(4.2,8.1,20.3,9.2,12.7,20.7)

wilcox.test(zone1,zone2)
```

```
        Wilcoxon rank sum test with continuity correction

data:  zone1 and zone2
W = 4, p-value = 0.01826
alternative hypothesis: true location shift is not equal to 0

Warning message:
In wilcox.test.default(zone1, zone2) :
  impossible de calculer la p-value exacte avec des ex-aequos
```

7. Oeufs de coucou

On a mesuré la longueur des œufs de coucou trouvés dans les nids de deux espèces d'oiseaux ; roitelet et fauvette.

Les tailles d'œufs sont données en cm, ci-dessous :

Nids de petite taille (roitelet) : 19.8, 22.1, 21.5, 20.9, 22.0, 21.0, 22.3, 21.0, 20.3, 22.0, 22.0, 20.9, 20.8, 21.2, 21.0, 21.3

Nids de grande taille (fauvette) : 22.0, 23.9, 20.9, 23.8, 25.0, 24.0, 23.8, 21.7, 22.8, 23.1, 23.5, 23.0, 23.0, 23.1, 23.2

Est-ce-que le coucou adapte la taille de ses œufs à l'espèce occupant le nid dans lequel il pond ?

On suppose ici que la taille des œufs de coucou suit une loi Normale dans les populations dont sont issus les échantillons.

7. Oeufs de coucou

Roitelet : n_1 = 16 ; $\bar{x}_1$ = 21.26 ; $S_1{}^*$ = 0.48

Fauvette: n_2 = 15 ; $\bar{x}_2$ = 23.12 ; $S_2{}^*$ = 1.02

Homoscédasticité ?

F_{obs} = 1.02/0.48 = 2.13

$F_{théo}$ = 2.89 (table Fisher pour α=0.05/2, et ddl =14, 15)

$F_{obs} < F_{théo}$ → H0 non rejeté

→ **Test t pour échantillons indépendants :**

Variance commune = 0.74

$\mathbf{t_{obs}}$ = -6.02 (ou +6.02 suivant le sens de la différence de moyennes au numérateur)

$\mathbf{t_{théo}}$ = 2.045 (table Student pour α=0.05/2, et ddl =16 + 15 – 2 = 29)

$| \mathbf{t_{obs}} | > \mathbf{t_{théo}} |$

→ On conclue que le coucou adapte la taille de ses œufs à l'espèce qui occupe le nid dans lequel il pond (H1)

Effectuer ce test sous R

```
roitelet<-c(19.8, 22.1, 21.5, 20.9, 22.0, 21.0, 22.3, 21.0,
            20.3, 22.0,22.0, 20.9,20.8, 21.2, 21.0, 21.3)
fauvette<-c(22.0, 23.9, 20.9, 23.8, 25.0, 24.0, 23.8, 21.7,
            22.8, 23.1, 23.5, 23.0,23.0, 23.1, 23.2)
```

```
> var.test(fauvette,roitelet)

        F test to compare two variances

data:  fauvette and roitelet
F = 2.1258, num df = 14, denom df = 15, p-value = 0.1596
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.7352041 6.2697467
sample estimates:
ratio of variances
          2.125827
```

```
> t.test(roitelet,fauvette, var.equal = T)

        Two Sample t-test

data:  roitelet and fauvette
t = -6.0166, df = 29, p-value = 1.522e-06
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -2.497292 -1.230208
sample estimates:
mean of x mean of y
 21.25625  23.12000
```

P-value hautement significative (p<0.001)

8. Les marguerites

On a mesuré la longueur des tiges (mm) de deux groupes de marguerites.

Le premier est situé à l'ombre le matin, le second est au soleil toute la journée.

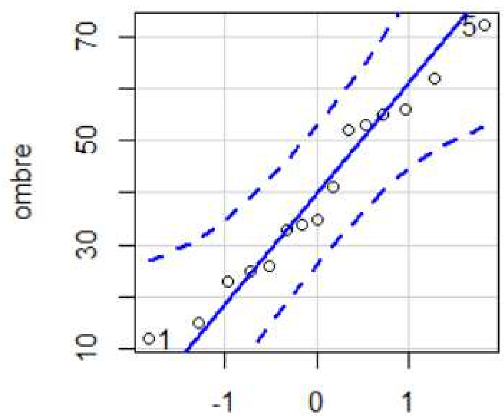
Ombre : 12, 23, 53, 62, 72, 25, 56, 33, 34, 15, 26, 35, 55, 52, 41

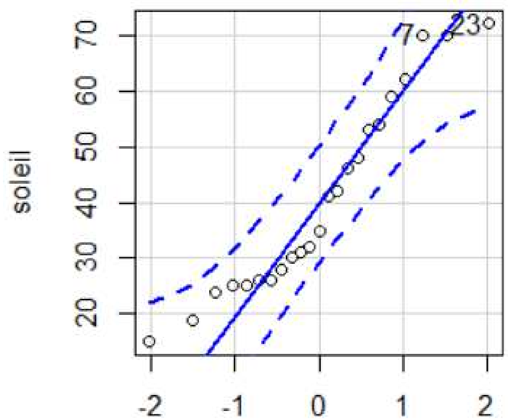
Soleil : 32, 54, 48, 46, 59, 62, 70, 15, 42, 41, 53, 31, 35, 26, 25, 30, 25, 26, 19, 24, 28, 70, 72

1) Est-ce que la croissance des tiges de marguerites diffère entre ces 2 modalités d'ensoleillement ?

On dispose des informations suivantes :

```
        Shapiro-Wilk normality test

data:  ombre
W = 0.95997, p-value = 0.6918


        Shapiro-Wilk normality test

data:  soleil
W = 0.9231, p-value = 0.07765
```





2) Faites le test de Wilcon-Mann-Whitney pour grands échantillons (même si $n_{Ombre} < 20$)

8. Les marguerites

Question 1 :

Homoscédasticité ?

- F_{obs} = 323.54/304.98 = 1.06
- $F_{théo}$ = 2.53 (table Fisher pour α=0.05/2, et ddl =14, 22)
- $F_{obs} < F_{théo}$ → H0 non rejeté

Test t pour échantillons indépendants :

Variance commune = 312.20

t_{obs} = -0.16 (ou + 0.16 suivant le sens de la différence de moyennes au numérateur)

$t_{théo}$ → table de Student pour α=0.05/2, et ddl =15 + 23 – 2 = 36

Dans la table on a : $t_{théo}$ = 2.042 pour ddl=30 et $t_{théo}$ = 2.021 pour ddl=40 , mais pas pour ddl=36.

On prend comme référence la valeur de $t_{théo}$ la plus conservative → 2.042

$| t_{obs} | < | t_{théo} |$

→ On conclue que, étant donné le design expérimental et la taille des échantillons, les observations n'ont pas permis de mettre en évidence une différence significative de croissance des tiges de marguerites entre les deux modalités d'ensoleillement (H0 n'est pas rejetée au seuil de 5%).

Effectuer ce test sous R

```
ombre<-c(12, 23, 53, 62, 72, 25, 56, 33, 34, 15, 26, 35, 55, 52, 41)
soleil<-c(32, 54, 48, 46, 59, 62, 70, 15, 42, 41, 53, 31, 35, 26, 25,
          30, 25, 26, 19, 24, 28, 70, 72)
```

```
> var.test(ombre,soleil)

        F test to compare two variances

data:  ombre and soleil
F = 1.0609, num df = 14, denom df = 22, p-value = 0.8754
alternative hypothesis: true ratio of variances is not equal to 1
95 percent confidence interval:
 0.4195606 2.9851724
sample estimates:
ratio of variances
          1.060851
```

```
> t.test(ombre,soleil, var.equal = T)

        Two Sample t-test

data:  ombre and soleil
t = -0.1646, df = 36, p-value = 0.8702
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -12.85812  10.92768
sample estimates:
mean of x mean of y
 39.60000  40.56522
```

8. Les marguerites

Question 2 : test de Wilcoxon-Mann Whitney pour grands échantillons (O=ombre, S=soleil)

$H_0 : U_1 = U_2$, tailles non différentes selon ensoleillement
$H_1 : U_1 \neq U_2$, tailles différentes selon ensoleillement

Taille	12	15	15	19	23	24	25	25	25	26	26	26	28	30	31	32	33	34	35
échantillon	O	O	S	S	O	S	O	S	S	O	S	S	S	S	S	S	O	O	O
rang	1	2.5	2.5	4	5	6	8	8	8	11	11	11	13	14	15	16	17	18	19.5

Taille	35	41	41	42	46	48	52	53	53	54	55	56	59	62	62	70	70	72	72
échantillon	S	O	S	S	S	S	O	O	S	S	O	O	S	O	S	S	S	O	S
rang	19.5	21.5	21.5	23	24	25	26	27.5	27.5	29	30	31	32	33.5	33.5	35.5	35.5	37.5	37.5

$R_{ombre} = 289$

$R_{soleil} = 452$

$$U_{ombre} = 289 - \frac{15(15+1)}{2} = 169$$

$$Usoleil = 452 - \frac{23(23+1)}{2} = 176$$

$\mathbf{U_{wmwobs} = 169}$

8. Les marguerites

Question 2 : test de Wilcoxon-Mann Whitney pour grands échantillons.

Grands échantillons donc U_{wmw} suit approximativement une loi normale de :

$$\mu_U = \frac{n_1 \times n_2}{2} = \frac{23 \times 15}{2} = 172.5 \quad et \quad \sigma_U = \sqrt{\frac{n_1 \times n_2 \times (n_1 + n_2 + 1)}{12}} = 33.48$$

Calcul de Z_{wmw} :

$$Zwmw = \frac{\mathrm{U}_{\mathrm{WMWobs}} - \mu_U}{\sigma_U} = \frac{169 - 172.5}{33.48} = -0.105$$

Comparaison de $|Z_{wmw}|$ avec $Z_{théo}$ = 1.96 (table de la loi normale pour un $\alpha = 0.05$)

$|Z_{wmw}| = 0.105 < 1.96$

Non rejet de H_0 ;

On ne peut pas mettre en évidence une différence dans la hauteur des marguerites.

Effectuer ce test sous R

Question 2 : test de Wilcoxon-Mann Whitney pour grands échantillons.

```
> wilcox.test(ombre,soleil)

        Wilcoxon rank sum test with continuity correction

data:  ombre and soleil
W = 169, p-value = 0.9286
alternative hypothesis: true location shift is not equal to 0
```