Contrôle continu master 1 TVPS : Corrigé Sujet 1

1. Le tableau ci-dessous décrit la sensibilité ou la résistance de trois espèces de plante à un agent pathogène.

	X1	X2	X3
résistant	22	12	7
sensible	62	18	12

La résistance est-elle la même pour toutes les espèces ?

Test : Nous utilisons le test du Chi². H_0 : Indépendance entre l'espèce et la résistance

```
> .Test
        Pearson's Chi-squared test
X-squared = 2.353, df = 2, p-value = 0.3084

> .Test$expected # Expected Counts
         1         2          3
1 25.89474  9.24812  5.857143
2 58.10526 20.75188 13.142857
```

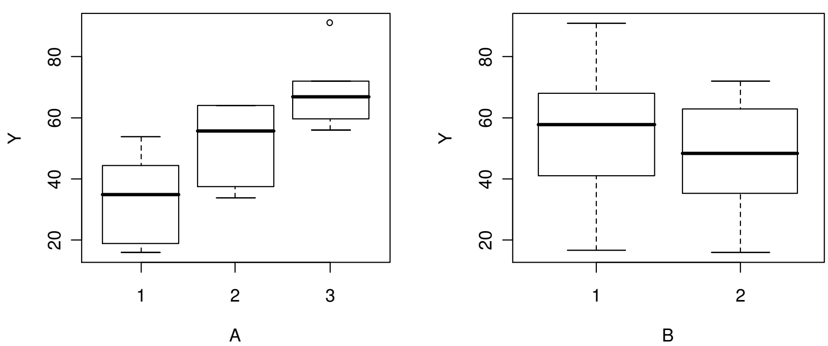
Condition : On observe que les effectifs théoriques sont tous supérieurs à 5. On peut utiliser le test.

Conclusion : p = 0.3 > 0.05, on accepte H_0.
L'espèce n'a pas d'influence significative sur la résistance.

2. On injecte à 24 lapins de l'insuline en leur donnant des doses notées A1, A2 et A3, préparées suivant deux protocoles différents notés B1 et B2. La réduction de sucre dans leur sang a été mesurée et elle a donné les résultats suivants

Réduction	B1				B2			
A1	17	21	49	54	33	37	40	16
A2	64	48	34	63	41	64	34	64
A3	62	72	61	91	56	62	57	72

Statistiques descriptives



Moyennes croisées

```
      B
A     1     2
  1 35.25 31.50
  2 52.25 50.75
  3 74.25 61.75
```

On observe peut de différence pour le facteur B.
Le facteur A présente des moyennes très différentes.
La variance ne semble pas stable avec A.

On envisage d'étudier l'influence des deux facteurs et leur interaction sur la réduction de sang. Pour cela on utilise une ANOVA.

Test : ANOVA à 2 facteurs

H0 : Les moyennes sont égales pour le facteur A, le facteur B et leur interaction.

Conditions :

Indépendance : supposée

Stabilité de la variance :

Test de Bartlett H0 : variance égale

```
> bartlett.test(Y ~ A, data=lapin)
      Bartlett's K-squared = 0.3806, df = 2, p-value = 0.8267


> bartlett.test(Y ~ B, data=lapin)
Bartlett's K-squared = 0.6795, df = 1, p-value = 0.4098
```

Conclusion : Dans les deux cas, on accepte H0 (p >0.05). Les variances ne sont pas significativement différentes en fonction du facteur A ou B.

Normalité des résidus :

Test de Shapiro

Ho : Distribution normale

```
> shapiro.test(lapin$LinearModel.1.residuals)
      data:  lapin$LinearModel.1.residuals
W = 0.9308, p-value = 0.1014
```

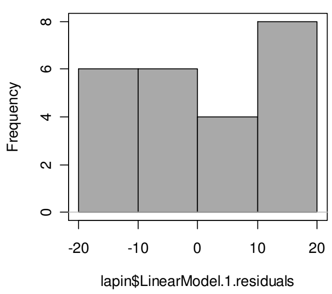
On accepte H0 (p>0.05). La distribution des résidus n'est pas significativement différente d'une loi normale.

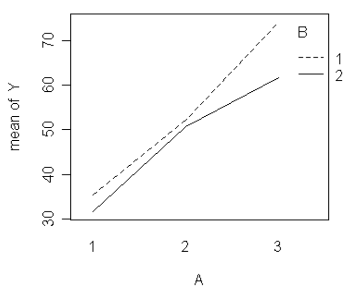
Sur l'histogramme des résidus ci-dessous, la distribution n'est pas franchement normale mais l'effectif est réduit.

Conclusion :

```
Anova Table (Type II tests)
Response: Y
          Sum Sq Df F value    Pr(>F)
A         4799.1  2 12.8748 0.0003378 ***
B          210.0  1  1.1270 0.3024563
A:B        135.1  2  0.3624 0.7009697
Residuals 3354.7 18
```

Seul le facteur A a un effet moyen significatif avec un risque de 1ere espèce de 0.0003, le facteur B n'a pas d'influence significative ni en moyenne ni en interaction.






```
> attach(lapin)
> interaction.plot(A,B,Y)
```

Le diagramme des interaction confirme ce résultat. Seul B a un petit effet combiné avec A3.

```
> pairwise.t.test(Y,A)
 Pairwise comparisons using t tests with pooled SD
data:  Y and A
  1       2
2 0.02503 -
3 0.00011 0.02503
```

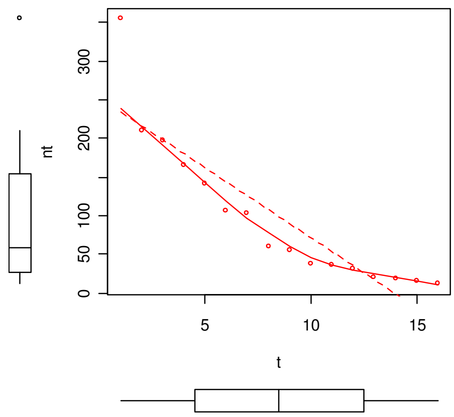
Les comparaisons multiples montrent que les 3 modalités de A sont différentes significativement 2 à 2.
3 > 2 > 1

3. Des bactéries marines sont soumises à un rayon X de 200kV pendant une période de t=1 à 16 intervalles de 3 min. Le nombre de colonies survivantes, nt, est étudiée en fonction de t :

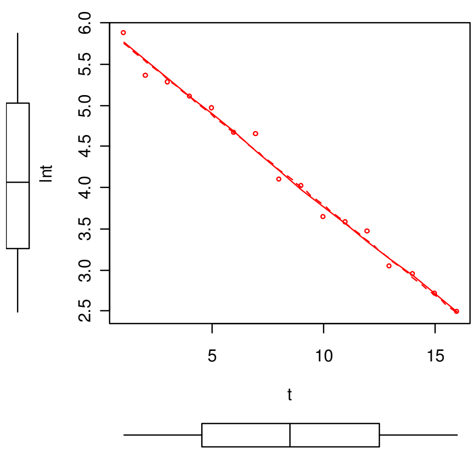
t	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**	**12**	**13**	**14**	**15**	**16**
nt	**355**	**211**	**197**	**166**	**142**	**106**	**104**	**60**	**56**	**38**	**36**	**32**	**21**	**19**	**15**	**12**

Existe-t-il une relation entre t et nt? Peut-on expliquer nt par t à l'aide d'un modèle simple comme la régression linéaire?

Nuage de points



nt en fonction de t



log(nt) en fonction de t

La relation n'étant pas linéaire nous allons étudier log(nt) en fonction de t.
La seconde relation présente une relation linéaire. Testons sa signification en testant la nullité du coefficient de corrélation.

H0 : "corrélation nulle

```
        Pearson's product-moment correlation

data:  bacteries$lnt and bacteries$t
t = -37.9617, df = 14, p-value = 1.606e-15
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
 -0.9983715 -0.9857648
sample estimates:
       cor
-0.9951777
```

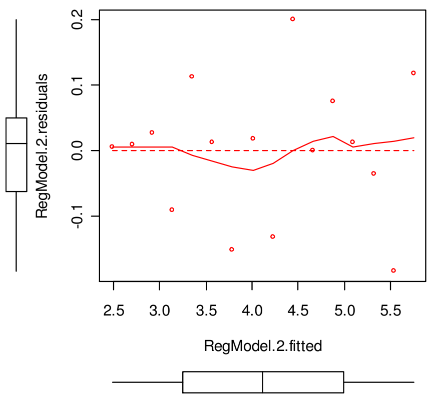
Conclusion: On rejette H0 avec un risque de 1ere espèce de 10^-15. Le coefficient de corrélation est significativement différent de 0. Il existe une relation significative entre t et log(nt)

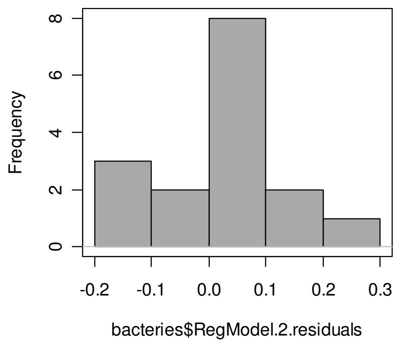
Modèle de régression linéaire
Call:
lm(formula = lnt ~ t, data = bacteries)
Coefficients:

	Estimate	Std. Error	t value	Pr(>|t|)	
(Intercept)	5.97234	0.05560	107.42	< 2e-16	***
t	-0.21828	0.00575	-37.96	1.61e-15	***

Residual standard error: 0.106 on 14 degrees of freedom
Multiple R-Squared: 0.9904, Adjusted R-squared: 0.9897
F-statistic: 1441 on 1 and 14 DF, p-value: 1.606e-15

Vérification des hypothèses :





Indépendance:
Test de durbin-Watson H0 pas d'autocoorélation

data: bacteries$RegModel.2.residuals
W = 0.9554, p-value = 0.5801

On accepte H0

Normalité :
Test de Shapiro H0 Normalité

W = 0.9554, p-value = 0.5801
Shapiro-Wilk normality test
On accepte H0

Stabilité de la variance :
Test de Breusch H0 pas de relation entre variance et $\hat{y}$.

data: bacteries$RegModel.2.residuals
W = 0.9554, p-value = 0.5801

On accepte H0

Les représentations graphiques confirment ces remarques.

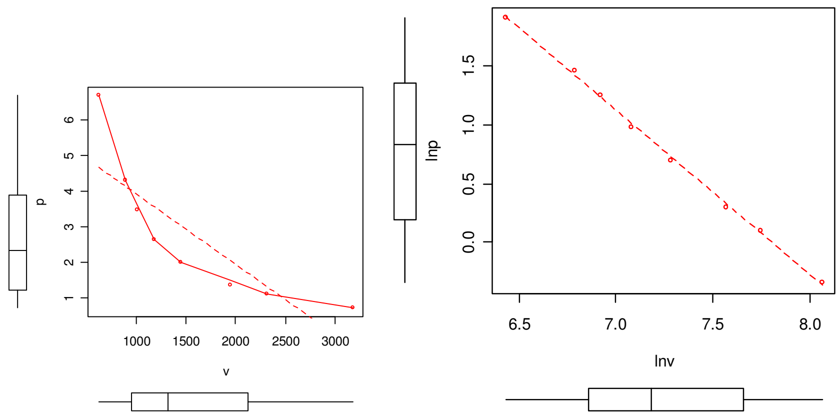
On peut donc exploiter statistiquement le modèle.
Les deux coefficients sont significativement non nulles (p<10^-15).

Sujet 2

1.

Exercice 5 Le tableau ci-dssous donne les valeurs expérimentales du volume V (en cm^3) et de la pression P (en Kg par cm^3) d'un gaz. D'après les lois de la thermodynamique de Laplace pour un gaz parfait, on a la relation $PV^\gamma = C$ où γ et C sont des constantes.

Volume (v_i)	620	890	1013	1186	1454	1944	2313	3179
Pression (p_i)	6.7	4.3	3.48	2.644	1.997	1.35	1.1	0.71



La relation n'est pas linéaire ne op et v mais le devient entre log(v) et log(p). Le rôle des variables est a priori symétrique. Elle le devient après passage au logarithme.

Testons sa signification en testant la nullité du coefficient de corrélation. H0 : "corrélation nulle

Pearson's product-moment correlation
data: gaz$lnp and gaz$lnv
t = -59.9958, df = 6, p-value = 1.441e-09
alternative hypothesis: true correlation is not equal to 0
95 percent confidence interval:
-0.9998557 -0.9952047
sample estimates:
cor
-0.9991676

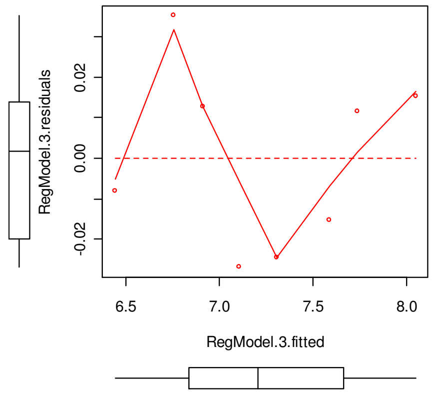
Conclusion: On rejette H0 avec un risque de 1ere espèce de 10-9.
Le coefficient de corrélation est significativement différent de 0. Il existe une relation significative entre t et log(nt)

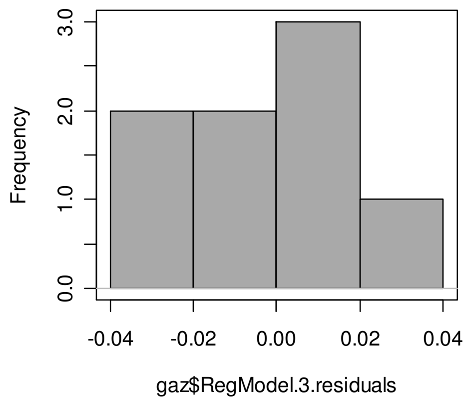
Modèle de régression linéaire
Coefficients:

```
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  7.80316    0.01265   616.9 1.22e-15 ***
lnp         -0.71785    0.01197   -60.0 1.44e-09 ***
```

Residual standard error: 0.02375 on 6 degrees of freedom
Multiple R-Squared: 0.9983, Adjusted R-squared: 0.9981
F-statistic: 3599 on 1 and 6 DF, p-value: 1.441e-09

Vérification des hypothèses :





Indépendance:
Test de durbin-Watson H0 pas d'autocoorélation
data: lnv ~ lnp
DW = 1.4042, p-value = 0.1351
On accepte H0

Normalité :
Test de Shapiro H0 Normalité
data: gaz$RegModel.3.residuals
W = 0.9304, p-value = 0.5195
On accepte H0

Stabilité de la variance :
Test de Breusch H0 pas de relation entre variance et $\hat{y}$.
data: lnv ~ lnp
BP = 0.2291, df = 1, p-value = 0.6322

On accepte H0

Les représentations graphiques confirment ces remarques.

On peut donc exploiter statistiquement le modèle.
Les deux coefficients sont significativement non nulles (p<10^-9).

2. Le tableau ci-dessous décrit la mortalité à la suite d'une opération à cœur ouvert en fonction du type d'anesthésie (halothane ou morphine) :

	vivants	décédés	…..
halothane	53	8	
morphine	57	10	

L'anesthésie a-t-elle une influence sur le taux de survie ?

Test : Nous utilisons le test du Chi². H_0 : Indépendance entre l'espèce et la résistance

```
> rowPercents(.Table) # Row Percentages
     1    2 Total Count
1 86.9 13.1   100    61
2 85.1 14.9   100    67

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

Pearson's Chi-squared test
data: .Table
X-squared = 0.0866, df = 1, p-value = 0.7685

> .Test$expected # Expected Counts

```
         1        2
1 52.42188 8.578125
2 57.57812 9.421875
```

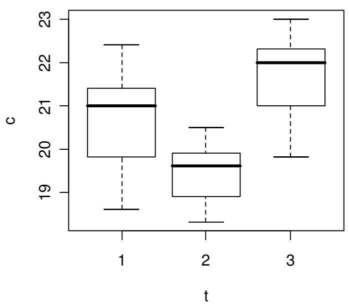
Condition : On observe que les effectifs théoriques sont tous supérieurs à 5. On peut utiliser le test.

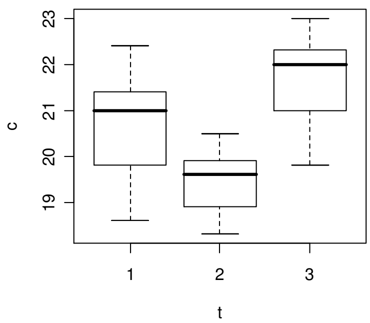
Conclusion : p = 0.77 > 0.05, on accepte H_0.
L'anesthésie n'a pas d'influence significative sur la mortalité.

3.On étudie la consommation d'un véhicule en fonction du type de véhicule (5 types A,B,C,D,E) et du type de ville (3 types a,b,c). On dispose de 3 répétitions pour chaque modalité.

Consommation	a			b			c		
A	20.3	19.8	21.4	21.6	22.4	21.3	19.8	18.6	21.0
B	19.5	18.6	18.9	20.1	19.9	20.5	19.6	18.3	19.8
C	22.1	23.0	22.4	20.1	21.0	19.8	22.3	22.0	21.6

Statistiques descriptives






```
   v
t    1        2        3
  1 20.5 21.76667 19.80000
  2 19.0 20.16667 19.23333
  3 22.5 20.30000 21.96667
```

On observe un effet de chaque facteur sur la consommation.
La variance semble stable.

On envisage d'étudier l'influence des deux facteurs et leur interaction sur la consommation. Pour cela on utilise une ANOVA.

Test : ANOVA à 2 facteurs
H0 : Les moyennes sont égales pour le facteur type, le facteur ville et leur interaction.

Conditions :

Indépendance : supposée

Stabilité de la variance :
Test de Bartlett   H0 : variance égale

> bartlett.test(c ~ t, data=essence)
Bartlett's K-squared = 1.7457, df = 2, p-value = 0.4178

> bartlett.test(c ~ v, data=essence)
Bartlett's K-squared = 2.6381, df = 2, p-value = 0.2674

Conclusion : Dans les deux cas, on accepte H0 (p >0.05). Les variances ne sont pas significativement différentes en fonction des facteurs v ou t.

Normalité des résidus :

Test de Shapiro

Ho : Distribution normale

```
      Shapiro-Wilk normality test
data:  essence$LinearModel.5.residuals
W = 0.9847, p-value = 0.9498
```

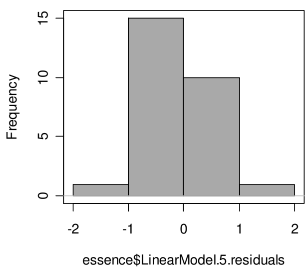
On accepte H0 (p>0.05). La distribution des résidus n'est pas significativement différente d'une loi normale.

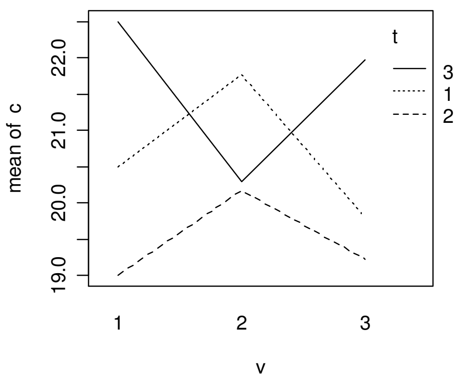
Sur l'histogramme des résidus ci-dessous, la distribution est proche d'une loi normale.

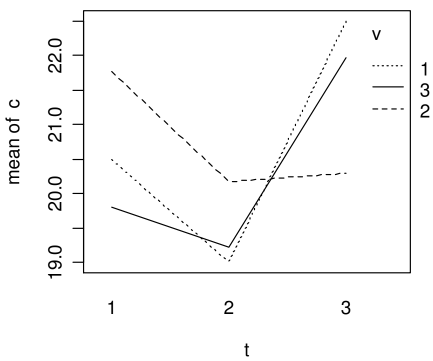
Conclusion :

```
Response: c
           Sum Sq Df F value    Pr(>F)
t         20.4230  2 22.2886 1.348e-05 ***
v          0.8585  2  0.9369 0.4101162
t:v       15.2926  4  8.3448 0.0005443 ***
Residuals  8.2467 18
```

Seul le facteur t a un effet moyen significatif avec un risque de 1ere espèce de 0.00001, le facteur v n'a pas d'influence significative en moyenne. Les interactions sont également significative avec p=0.0005.








```
> attach(essence)
> interaction.plot(v,t,c)
```

Le diagramme des interactions confirme ces résultats. l'effet de t est fort entre 3 et , celui de v est faible. Il existe de fortes interactions : t3 est supérieur à t2 dans les villes 1 et 3 mais pas 2.

```
> pairwise.t.test(c,t)
 Pairwise comparisons using t tests with pooled SD
data:  Y and A
  1       2
2 0.03349 -
3 0.07040 0.00049
```

Les comparaisons multiples montrent que les modalités de t 1 et 3 ne sont pas significativement différentes, mais 2 l'est à 1 et 3. 1=3 > 2