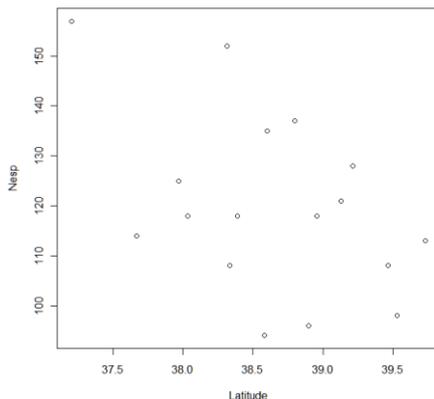


I. Un des résultats bien connus en écologie est que la richesse en espèce diminue en s'éloignant de l'équateur. On a recueilli le nombre d'espèces d'oiseaux en divers site aux USA (dans le Delaware) à différentes latitudes (tableau ci-dessous).

Lieu	Latitude	Nesp
Bombay Hook	39.217	128
Cape Henlopen	38.800	137
Middle-town	39.467	108
Milford	38.958	118
Rehoboth	38.600	135
Seaford	38.583	94
Wilmington	39.733	113
Crisfield	38.033	118
Denton	38.900	96
Elkton	39.533	98
Lower Kent County	39.133	121
Ocean City	38.317	152
Salisbury	38.333	108
Dorchester County	38.387	118
Cape Charles	37.200	157
Chincoteague	37.967	125
Wachapreague	37.667	114

Le nuage de dispersion obtenu puis les résultats de régression linéaire du nombre d'espèces (Nesp) sur la latitude sont les suivants :



Call:

lm(formula = Nesp ~ Latitude)

Residuals:

Min 1Q Median 3Q Max  
-26.651 -11.193 -1.989 14.555 28.141

Coefficients:

	Estimate	Std. Error	t value	Pr(> t )
(Intercept)	586.002	230.118	2.547	0.0224 *
Latitude	-12.061	5.955	-2.025	0.0610 .

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 16.37 on 15 degrees of freedom  
Multiple R-squared: 0.2147, Adjusted R-squared: 0.1624  
F-statistic: 4.102 on 1 and 15 DF, p-value: 0.06101

1. Que pensez-vous du graphique et que présage-t-il ? La régression est-elle significative ? Quelle part de la variance totale explique-t-elle ?
2. Ecrivez la formule de la droite de régression et calculez-en 2 points pour la tracer ensuite. Que pensez-vous à propos des résidus ?
3. Ce résultat contredit-il la théorie ? Pouvez-vous émettre une hypothèse qui expliquerait ce résultat et suggérer un complément d'échantillonnage ?

II. On a analysé l'influence du temps et de trois espèces ligneuses d'arbre sur la décomposition de la masse d'une litière constituée de feuilles de Lierre. Pour ce faire, 24 sachets d'une masse identique de feuilles de lierre ont été constitués, sachets permettant une décomposition naturelle. Puis une première série de 8 sachets, choisis au hasard, a été déposée sous un chêne, une deuxième sous un peuplier, et la dernière série sous un frêne. Après 2, 7, 10 et 16 semaines respectivement, deux sachets sont prélevés au hasard sous chaque arbre et la masse résiduelle est déterminée pour chacun d'eux. Cette masse est exprimée en pourcentage de la masse initiale.

Sachet	Semaine	Arbre	Masse
1	2	Chêne	85,10
2	2	Chêne	87. 60
3	2	Peuplier	85. 20
4	2	Peuplier	84. 90
5	2	Frêne	84. 30
6	2	Frêne	85. 75
7	7	Chêne	75. 90
8	7	Chêne	72. 85
9	7	Peuplier	73. 00
10	7	Peuplier	75. 70

11	7	Frêne 72. 80
12	7	Frêne 70. 80
13	10	Chêne 71. 60
14	10	Chêne 66. 95
15	10	Peuplier 74. 15
16	10	Peuplier 71. 85
17	10	Frêne 67. 10
18	10	Frêne 64. 95
19	16	Chêne 62. 10
20	16	Chêne 64. 30
21	16	Peuplier 67. 25
22	16	Peuplier 60. 25
23	16	Frêne 58. 75
24	16	Frêne 59. 00

Sum Sq	Df	F	P	
Semaine	1741.31	3	121.6927	3.004E-09
Arbre	58.08	2	6.0881	0.01495
Interaction	30.22	6	1.0559	0.43853
Résiduelle	57.24	12		
Totale	1886.84	23		

Quelle analyse a été effectuée ? Expliquez le tableau de résultats ci-dessus (signification des lignes et colonnes).  
Que conclure de cette analyse ?

III On a étudié l'effet de la dose d'un poison sur des groupes de cafards :

dose	effectif	morts
1	1.691	59
2	1.724	60
3	1.755	62
4	1.784	56
5	1.811	63
6	1.837	59
7	1.861	62
8	1.884	60

Les résultats d'analyse avec R sont les suivants. Quel type d'analyse a été réalisé ? Est-ce justifié ? Que conclure ?

```
Call:
glm(formula = morts ~ ldose, family = binomial(link = "logit"))
Deviance Residuals:
Min 1Q Median 3Q Max
-1.5878 -0.4085 0.8442 1.2455 1.5860
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -60.740   5.182  -11.72 <2e-16 ***
ldose        34.286   2.913   11.77 <2e-16 ***
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 284.202 on 7 degrees of freedom
Residual deviance: 11.116 on 6 degrees of freedom
AIC: 41.314
Number of Fisher Scoring iterations: 4
```

IV Afin de faire des estimations d'intentions de vote pour les prochaines élections présidentielles, un institut de sondage national mène une enquête dans plusieurs régions de France et auprès de plusieurs catégories socioprofessionnelles. Les résultats suivants sont obtenus (données fictives donnant le nombre d'intentions de votre à droite et à gauche, dans cinq régions françaises et pour cinq catégories socioprofessionnelles) :

région	catégorie	votes à droite	votes à gauche
nord	ouvriers	255	245
nord	artisans	268	232
nord	agriculteurs	250	250
nord	employés	330	170
nord	cadres	360	140
centre	ouvriers	263	237
centre	artisans	317	183
centre	agriculteurs	306	194
centre	employés	270	230
centre	cadres	360	140
sud-est	ouvriers	350	150
sud-est	artisans	192	308
sud-est	agriculteurs	335	165
sud-est	employés	180	320
sud-est	cadres	377	123
sud-ouest	ouvriers	98	402
sud-ouest	artisans	159	341
sud-ouest	agriculteurs	252	248
sud-ouest	employés	325	175
sud-ouest	cadres	348	152
Bretagne	ouvriers	195	305
Bretagne	artisans	226	274
Bretagne	agriculteurs	301	199
Bretagne	employés	98	402
Bretagne	cadres	410	90

Quelle(s) méthode(s) statistique(s) proposeriez-vous pour analyser ce tableau de données, et quelle sorte d'informations serait-on en droit d'en tirer ? Justifier votre réponse, notamment en insistant sur les raisons de votre choix.

V. Un pêcheur statisticien, grand amateur de brochets, se procure des appâts (ablettes et gardons) au moyen de nasses qu'il plonge **dans son étang favori**. Il a noté soigneusement ses prises et les conditions dans lesquelles il les a obtenues (<http://www.unice.fr/coquillard/UE7/peche.txt>) sur plusieurs années.

1. A quelle méthode pensez-vous pour analyser l'ensemble de ces données ? Justifiez soigneusement votre réponse.
2. Pour commencer, il décide d'examiner les relations (*dont il pense qu'elles sont linéaires*) entre ses prises et le nombre de nasses qu'il a disposé ainsi qu'entre ses prises et l'oxymétrie. Au moyen de R vérifiez si son intuition est correcte ou non. Quel pourcentage de la variance totale des prises est expliqué par le nombre de nasses ? la teneur en oxygène ?
3. Enfin, il décide de trouver un modèle qui explique bien ses prises... Au moyen de R, proposez-lui un modèle. La mesure de l'oxygène a-t-elle été judicieuse ? A-t-il intérêt à laisser en place ses nasses un temps proportionnel à leur nombre ?
4. A la réflexion, quelles remarques faites-vous à propos de son échantillonnage (i.e. est-il critiquable ?)
5. Pour finir, contactant 2 autres passionnés de pêche, il rassemble leurs données en un seul fichier ([http://www.unice.fr/coquillard/UE7/peche\\_2.txt](http://www.unice.fr/coquillard/UE7/peche_2.txt)). Proposez un modèle expliquant le nombre de prises, en expliquant votre choix d'analyse.