

## INFORMATIQUE ET MODÈLES LINÉAIRES

Novembre 2011 - Session 1, Semestre 1

Durée de l'épreuve : 3h00

Tous documents autorisés.

(Les trois exercices sont indépendants. Un soin tout particulier sera apporté à la rédaction des réponses)

**Exercice 1**

1. (a) En vous aidant de la commande `seq()`, générez la séquence  

$$-1, -0.9, -0.8, \dots, 0, \dots, 0.8, 0.9, 1$$
et stockez la dans un vecteur  $x$ .  
(b) Comment extraire le sous-vecteur correspondant aux éléments situés aux positions 5 et 7 à 10 ?  
(c) Comment extraire les éléments négatifs et les affecter au vecteur `xmoins` ?  
(d) Comment retirer du vecteur `x` les éléments supérieurs à 0.5 et les affecter au vecteur `xnew` ?
2. (a) Générez une matrice  $M$  à l'aide de la fonction `matrix`, à 10 lignes et 5 colonnes, aléatoire (avec des valeurs réelles comprises entre 0 et 1).  
(b) Déterminez le nombre d'éléments supérieurs à 0.9.  
(c) Remplacez les éléments de  $M$  inférieurs à 0.5 par des 0.  
(d) Testez et vérifiez son type et la nature de ses éléments.  
(e) Créez un “data frame” nommé `MDF` à partir de  $M$ . Vérifiez que `MDF` a la structure voulue.  
(f) Extrayez le vecteur correspondant à la troisième colonne.  
(g) Extrayez la liste correspondant à la deuxième ligne.
3. Générez un data frame nommé `DF` tel que :
  - (a) les variables (colonnes) sont nommées : “Sexe”, “Âge” puis “Note 1”, “Note 2”, …, “Note 15” ;
  - (b) le nombre d’individus (lignes) est de 50, et les lignes sont nommées “Étudiant  $n$ ” où  $n$  est un numéro aléatoire unique ;
  - (c) le sexe de chaque individu est choisi au hasard parmi “Masculin” et “Féminin” ;
  - (d) les notes sont générées aléatoirement entre 0 et 20 et arrondies à 0.5 ;
  - (e) Les âges sont générés aléatoirement entre 18 et 24.

Extrayez le sous-ensemble des données correspondant aux variables “Note 3”, “Note 7”, “Note 8”, …, “Note 17”. Extrayez le sous-ensemble des données correspondant aux filles.

(On utilisera entre-autres les fonctions `sample`, `matrix`, `row.names` et `paste`)

**Exercice 2**

On s'intéresse au jeu de données stocké dans les fichiers “CO2.csv” ou “CO2.txt”, présents dans le dossier MASTER SIDE/Examen03-11-2011. Ce jeu présente le “Carbon Dioxide Uptake in Grass Plants” c'est-à-dire la consommation de dioxyde de carbone par des plantes herbeuses réfrigérées ou non (“Chilled” signifie “réfrigérée”).

1. Ouvrez la table dans R à partir du fichier “CO2.csv” ou “CO2.txt” et visualisez la table, affichez le nom des variables colonnes. Présentez les différents types de plantes.
2. Précisez les modalités pour chacune des variables qualitatives (au nombre de 3) de la table.
3. Résumez l'information contenue dans la table.

4. Représentez graphiquement à l'aide du graphique adéquat les variables qualitatives de la table dans une même fenêtre graphique.
5. On s'intéresse à la colonne "conc". Retrouvez les informations contenues dans cette variable sans utiliser la fonction `summary`. Représentez graphiquement cette variable à l'aide du graphique adéquat.

### Exercice 3

On se donne l'énoncé ci-dessous issu d'un examen sur les modèles linéaires ainsi que sa correction. Re trouvez à l'aide de R les réponses aux questions.

On dit que dans une famille, les aînés ont tendance à être plus indépendants que leurs cadets. Un chercheur élabore une échelle d'indépendance en 25 points et procède à l'évaluation de 20 aînés et du frère ou de la sœur qui suit directement chacun des aînés. Imaginons qu'il obtienne les résultats suivants :

$i$	Aîné	Cadet	$i$	Aîné	Cadet
1	8	9	11	17	13
2	13	15	12	12	8
3	8	10	13	2	7
4	5	7	14	13	8
5	12	10	15	19	14
6	15	13	16	18	12
7	5	8	17	14	8
8	15	12	18	17	11
9	16	13	19	18	12
10	5	9	20	20	10

Un collaborateur du premier chercheur suggère que la différence observée sur une paire dépend essentiellement du score de l'aîné

1. Pour chaque paire, on appelle  $x_i$  le score de l'aîné et  $y_i$  la différence (algébrique) des scores entre l'aîné et le cadet. Calculez le coefficient de corrélation  $R$  entre les deux séries étudiées.
2. À l'aide d'un test de Student-Fisher au seuil de  $\alpha = 5\%$ , montrer que l'hypothèse  $H_0 : \ll R = 0 \gg$  n'est pas acceptable et que par conséquent, la corrélation est significative. On utilisera pour cela la statistique

$$T = \sqrt{n-2} \frac{R}{\sqrt{1-R^2}}$$

qui suit une loi de Student-Fisher à  $\nu = n - 2$  degrés de liberté, ainsi que la table de l'annexe.

3. Déterminez une équation de la droite de régression des  $(y_i)$  par rapport aux  $(x_i)$ .
4. Représentez sur un graphique le nuage de points  $(x_i; y_i)$  et la droite déterminée au 3.
5. Donnez une estimation du niveau d'indépendance d'un cadet dont le frère (aîné) a obtenu une note de 22

Correction :

1. On a  $\bar{x} = \frac{252}{20} = 12,6$ ,  $\bar{y} = \frac{43}{20} = 2,15$ ,  $\sigma_X^2 = \frac{37222}{20} - (12,6)^2 = 27,34$  (donc  $\sigma_X = 5,23$ ),  $V(Y) = \sigma_Y^2 = \frac{415}{20} - (2,15)^2 = 16,13$  (donc  $\sigma_Y = 4,015$ ) et  $Cov(X, Y) = \frac{920}{20} - (12,6 \times 2,15) = 18,91$ . Finalement,  $R = \frac{18,91}{5,23 \times 4,015} = 0,9002$ .

2. On détermine ensuite la valeur observée de la statistique  $T$  soit  $T_{obs} = \sqrt{n-2} \frac{R}{\sqrt{1-R^2}} = 8,76$ . Pour un nombre de degrés de liberté (d.d.l.) égal à 18, un seuil  $\alpha = 5\%$ , et un test bilatéral, on obtient grâce à la table de la loi de Student Fisher,  $T_{lu} = 2,101$ . On remarque que  $T_{lu} < T_{obs}$  donc on rejette l'hypothèse  $H_0$  selon laquelle  $R$  est négligeable. Le coefficient  $R$  est donc significatif (la corrélation est significative).

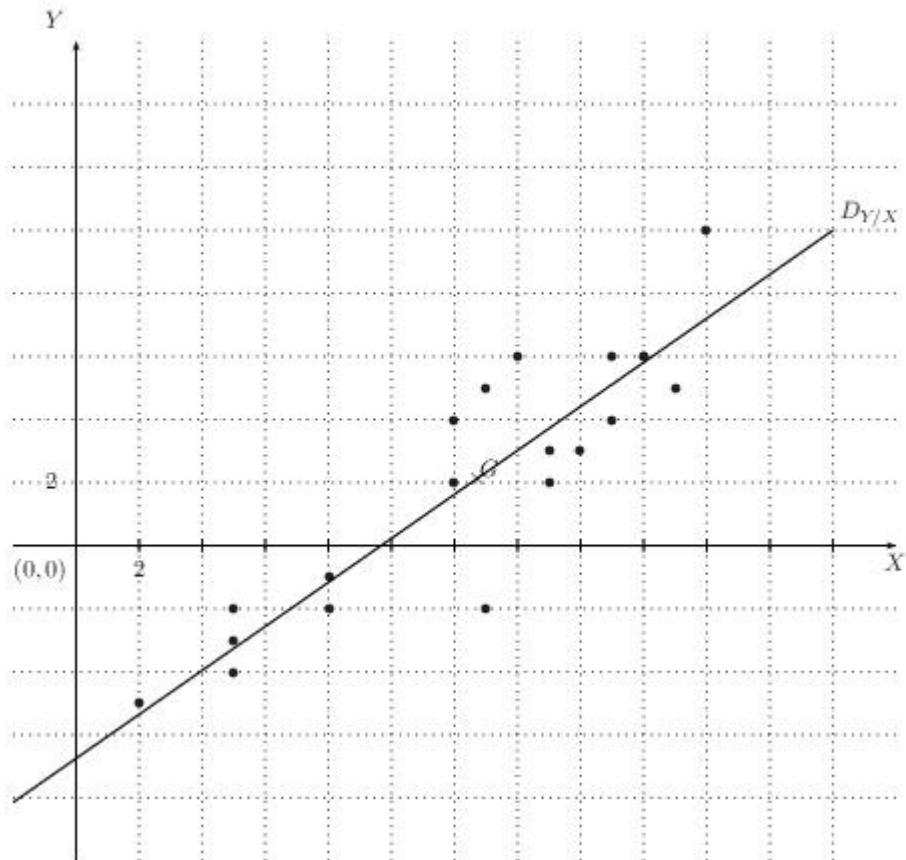
3. Montrons que la droite de régression admet pour équation

$$D_{Y/X} : Y = 0,69X - 6,56.$$

Les coefficients de l'équation  $Y = bX + a$  de la droite des moindres carrés sont donnés par

$$\begin{cases} b &= \frac{Cov(X, Y)}{V(X)} = \frac{18,91}{27,34} = 0,69 \\ a &= \bar{y} - b\bar{x} = 2,15 - 0,69 \times 12,6 = -6,56 \end{cases}.$$

4. On a le nuage de point suivants.



5. Si on pose  $X = 22$  dans l'équation trouvée précédemment, on trouve  $Y = 0,69 \times 22 - 6,56 = 8,62$ . Une estimation du niveau d'indépendance d'un cadet dont le frère (ainé) a obtenu une note de 22 est de  $22 - 8,62 = 13,38$ .

## ANNEXE A - Probabilités individuelles et cumulées de la loi de Student-Fisher $t_{\nu,\alpha}$ .

Cette table donne les valeurs  $t_{\nu,\alpha}$  telles que  $p(\{t_{\nu,\alpha} < t_{\nu} < +t_{\nu,\alpha}\}) = 1 - \alpha$  :

$\nu \backslash \alpha$	0, 90	0, 80	0, 70	0, 60	0, 50	0, 40	0, 30	0, 20	0, 10	0, 05	0, 02	0, 01	0, 001
1	0, 158	0, 325	0, 510	0, 727	1, 000	1, 376	1, 963	3, 078	6, 314	12, 706	31, 821	63, 657	636, 619
2	0, 142	0, 289	0, 445	0, 617	0, 816	1, 061	1, 386	1, 886	2, 920	4, 303	6, 965	9, 925	31, 598
3	0, 137	0, 277	0, 424	0, 584	0, 765	0, 978	1, 250	1, 638	2, 353	3, 182	4, 541	5, 841	12, 929
4	0, 134	0, 271	0, 414	0, 569	0, 741	0, 941	1, 190	1, 533	2, 132	2, 776	3, 747	4, 604	8, 610
5	0, 132	0, 267	0, 408	0, 559	0, 727	0, 920	1, 156	1, 476	2, 015	2, 571	3, 365	4, 032	6, 869
6	0, 131	0, 265	0, 404	0, 553	0, 718	0, 906	1, 134	1, 440	1, 943	2, 447	3, 143	3, 707	5, 959
7	0, 130	0, 263	0, 402	0, 549	0, 711	0, 896	1, 119	1, 415	1, 895	2, 365	2, 998	3, 499	5, 408
8	0, 130	0, 262	0, 399	0, 546	0, 706	0, 889	1, 108	1, 387	1, 860	2, 306	2, 896	3, 355	5, 041
9	0, 129	0, 261	0, 398	0, 543	0, 703	0, 883	1, 100	1, 383	1, 833	2, 262	2, 821	3, 250	4, 781
10	0, 129	0, 260	0, 397	0, 542	0, 700	0, 879	1, 093	1, 372	1, 812	2, 228	2, 764	3, 169	4, 587
11	0, 129	0, 260	0, 396	0, 540	0, 697	0, 876	1, 088	1, 363	1, 796	2, 201	2, 718	3, 106	4, 437
12	0, 128	0, 259	0, 395	0, 539	0, 695	0, 873	1, 083	1, 356	1, 782	2, 179	2, 681	3, 055	4, 318
13	0, 128	0, 259	0, 394	0, 538	0, 694	0, 870	1, 079	1, 350	1, 771	2, 160	2, 650	3, 012	4, 221
14	0, 128	0, 258	0, 393	0, 537	0, 692	0, 868	1, 076	1, 345	1, 761	2, 145	2, 624	2, 977	4, 140
15	0, 128	0, 258	0, 393	0, 536	0, 691	0, 866	1, 074	1, 341	1, 753	2, 131	2, 602	2, 947	4, 073
16	0, 128	0, 258	0, 392	0, 535	0, 690	0, 865	1, 071	1, 337	1, 745	2, 120	2, 583	2, 921	4, 015
17	0, 128	0, 257	0, 392	0, 534	0, 689	0, 863	1, 069	1, 333	1, 740	2, 110	2, 567	2, 898	3, 965
18	0, 127	0, 257	0, 392	0, 534	0, 688	0, 862	1, 067	1, 330	1, 734	2, 101	2, 552	2, 878	3, 922
19	0, 127	0, 257	0, 391	0, 533	0, 688	0, 861	1, 066	1, 328	1, 729	2, 093	2, 539	2, 861	3, 883
20	0, 127	0, 257	0, 391	0, 533	0, 687	0, 860	1, 064	1, 325	1, 725	2, 086	2, 528	2, 845	3, 850
21	0, 127	0, 257	0, 391	0, 532	0, 686	0, 859	1, 063	1, 323	1, 721	2, 080	2, 518	2, 831	3, 819
22	0, 127	0, 256	0, 390	0, 532	0, 686	0, 858	1, 061	1, 321	1, 717	2, 074	2, 508	2, 819	3, 792
23	0, 127	0, 256	0, 390	0, 532	0, 685	0, 858	1, 060	1, 319	1, 714	2, 069	2, 500	2, 807	3, 767
24	0, 127	0, 256	0, 390	0, 531	0, 685	0, 857	1, 059	1, 318	1, 711	2, 064	2, 492	2, 797	3, 745
25	0, 127	0, 256	0, 390	0, 531	0, 684	0, 856	1, 058	1, 316	1, 708	2, 060	2, 485	2, 787	3, 725
26	0, 127	0, 256	0, 390	0, 531	0, 684	0, 856	1, 058	1, 315	1, 706	2, 056	2, 479	2, 779	3, 707
27	0, 127	0, 256	0, 389	0, 531	0, 684	0, 855	1, 057	1, 314	1, 703	2, 052	2, 473	2, 771	3, 690
28	0, 127	0, 256	0, 389	0, 530	0, 683	0, 855	1, 056	1, 313	1, 701	2, 048	2, 467	2, 763	3, 674
29	0, 127	0, 256	0, 389	0, 530	0, 683	0, 854	1, 055	1, 311	1, 699	2, 045	2, 462	2, 756	3, 649
30	0, 127	0, 256	0, 389	0, 530	0, 683	0, 854	1, 055	1, 310	1, 697	2, 042	2, 457	2, 750	3, 656
40	0, 126	0, 255	0, 388	0, 529	0, 681	0, 851	1, 050	1, 303	1, 684	2, 021	2, 423	2, 704	3, 551
80	0, 126	0, 254	0, 387	0, 527	0, 679	0, 848	1, 046	1, 296	1, 671	2, 000	2, 390	2, 660	3, 460
120	0, 126	0, 254	0, 386	0, 526	0, 677	0, 845	1, 041	1, 289	1, 658	1, 980	2, 358	2, 617	3, 373
$\infty$	0, 126	0, 253	0, 385	0, 524	0, 674	0, 842	1, 036	1, 282	1, 645	1, 940	2, 326	2, 576	3, 291