Sauf indication contraire, on considérera dans les exercices un risque de première espèce de 5%.

Exercice 1 — Zé doux, thé — 2 points

Au niveau national, le score moyen sur les examens d'entrée au collège est 452 avec un écart type de 95. Un échantillon aléatoire de 152 étudiants de première année entrant au collège de Touffreville montre un score moyen de 502.

1. Y a t-il une différence significative entre le collège de Touffreville et le niveau national ?

Exercice 2 — T'es où, Zed ? — 2 points

L'alcalinité, en milligrammes par litre, de l'eau dans le cours supérieur et inférieur des rivières dans une région donnée est connu pour être distribué normalement. Dix lectures d'alcalinité sont faits dans le cours supérieur d'une rivière dans la région et quinze dans le cours inférieur de la même rivière avec les résultats suivants.

cours supérieur	91 75 91 88 94 63 86 77 71 69
cours inférieur	86 95 135 121 68 64 113 108 79 62 143 108 121 85 97

1. Peut on affirmer, avec une risque de première espèce de 1%, que l'alcalinité de l'eau dans le cours inférieur de cette rivière est supérieure à celle dans la partie supérieure ?

Exercice 3 — A Paris, hé ! — 3 points

Un aliment de commodité, connu comme « Quicknosh », a été introduit sur le marché en janvier 2010. Après une mauvaise année pour les ventes le fabricant a lancé une campagne de publicité intensive au cours de Janvier 2011. Le tableau ci-dessous enregistre les ventes, en milliers d'euros, pour une période d'un mois avant et d'un mois après la campagne de publicité, pour chacune des onze régions.

régions	nrd	sud	est	out	nes	ses	sse	nne	nno	sso	cen
ventes avant campagne	2,4	2,6	3,9	2,0	3,2	2,2	3,3	2,1	3,1	2,2	2,8
ventes après campagne	3,0	2,5	4,0	4,1	4,8	2,0	3,4	4,0	3,3	4,2	3,9

1. Peut on affirmer qu'une augmentation des ventes s'est produite ?

Exercice 4 — Les kids de Cold play — 4 points

Des études de marché amènent à penser que la musique de fond pourrait affecter le comportement d'achat des clients. Une étude dans un supermarché a comparé trois situations affectées au hasard : pas de musique, musique d'accordéon français, et une musique de mandoline italienne. Dans chaque situation, le nombre de bouteilles de vin français, italien, et autre achetés a été enregistré. Voici un tableau qui résume les résultats :

	Musique		
	aucune	accordéon	mandoline
vin français	30	39	30
vin italien	11	5	19
autre	84	75	84

1. Peut on affirmer que le type de musique influence les ventes ?

Exercice 5 — L'art est graisse ion — 9 points

On a observé deux variables : la température (notée t) et le temps nécessaire au démarrage (notée d). Les observations ont été résumées dans le tableau suivant :

i	t	d
1	9.15	0.53
2	5.08	1.44
3	11.06	3.13
4	9.82	1.24
5	3.93	3.23
6	-0.93	3.94
7	8.18	0.16
8	-3.18	3.94
9	3.19	3.83
10	5.79	0.79
11	-2.05	3.96
12	1.20	3.90
13	7.37	0.03
14	-0.36	3.89
15	2.48	3.97

et on donne les résultats suivant :

$$\sum_{i=1}^{15} t_i = 60.77 \qquad \sum_{i=1}^{15} d_i = 38.07 \qquad \sum_{i=1}^{15} t_i^2 = 532.33 \qquad \sum_{i=1}^{15} d_i^2 = 132.70 \qquad \sum_{i=1}^{15} t_i d_i = 79.18$$

1. la régression linéaire simple :
   a) posez un modèle de régression linéaire de d en fonction de t
   b) estimez les paramètres du modèle
   c) proposez une estimation de la variance de l'erreur
   d) quelle serait selon vous le temps nécessaire au démarrage pour une température de -4
   e) donnez un intervalle de confiance sur cette prédiction.
   f) que pensez vous de la qualité du modèle ?
2. L'expert du domaine prétend que pour une température inférieure à 3,5, le temps nécessaire au démarrage est constant, et que ce temps est un polynôme du second degré en fonction de la température lorsque celle ci dépasse 3,5.
   a) posez le modèle de régression associé aux dires de l'expert
   b) écrire ce modèle sous forme matricielle
   c) proposez une fonction de type Matlab permettant d'estimer sous forme vectorielle les paramètres du modèle.
   d) si ce modèle était exact, quelle serait selon vous le temps nécessaire au démarrage pour une température de -4 ?

Tables de la loi normale : Cette table nous donne les valeurs de t telle que $P(X \leq t)$ lorsque X suit une loi nomale centrée réduite

$$\Pi(t) = P(X \leq t) = \int_{-\infty}^{t} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx \quad \text{et} \quad \Pi(-t) = 1 - \Pi(t).$$

t	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0.0	0.5000	0.5040	0.5080	0.5120	0.5160	0.5199	0.5239	0.5279	0.5319	0.5359
0.1	0.5398	0.5438	0.5478	0.5517	0.5557	0.5596	0.5636	0.5675	0.5714	0.5753
0.2	0.5793	0.5832	0.5871	0.5910	0.5948	0.5987	0.6026	0.6064	0.6103	0.6141
0.3	0.6179	0.6217	0.6255	0.6293	0.6331	0.6368	0.6406	0.6443	0.6480	0.6517
0.4	0.6554	0.6591	0.6628	0.6664	0.6700	0.6736	0.6772	0.6808	0.6844	0.6879
0.5	0.6915	0.6950	0.6985	0.7019	0.7054	0.7088	0.7123	0.7157	0.7190	0.7224
0.6	0.7257	0.7291	0.7324	0.7357	0.7389	0.7422	0.7454	0.7486	0.7517	0.7549
0.7	0.7580	0.7611	0.7642	0.7673	0.7704	0.7734	0.7764	0.7794	0.7823	0.7852
0.8	0.7881	0.7910	0.7939	0.7967	0.7995	0.8023	0.8051	0.8078	0.8106	0.8133
0.9	0.8159	0.8186	0.8212	0.8238	0.8264	0.8289	0.8315	0.8340	0.8365	0.8389
1.0	0.8413	0.8438	0.8461	0.8485	0.8508	0.8531	0.8554	0.8577	0.8599	0.8621
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177
1.4	0.9192	0.9207	0.9222	0.9236	0.9251	0.9265	0.9279	0.9292	0.9306	0.9319
1.5	0.9332	0.9345	0.9357	0.9370	0.9382	0.9394	0.9406	0.9418	0.9429	0.9441
1.6	0.9452	0.9463	0.9474	0.9484	0.9495	0.9505	0.9515	0.9525	0.9535	0.9545
1.7	0.9554	0.9564	0.9573	0.9582	0.9591	0.9599	0.9608	0.9616	0.9625	0.9633
1.8	0.9641	0.9649	0.9656	0.9664	0.9671	0.9678	0.9686	0.9693	0.9699	0.9706
1.9	0.9713	0.9719	0.9726	0.9732	0.9738	0.9744	0.9750	0.9756	0.9761	0.9767
2.0	0.9772	0.9778	0.9783	0.9788	0.9793	0.9798	0.9803	0.9808	0.9812	0.9817
2.1	0.9821	0.9826	0.9830	0.9834	0.9838	0.9842	0.9846	0.9850	0.9854	0.9857
2.2	0.9861	0.9864	0.9868	0.9871	0.9875	0.9878	0.9881	0.9884	0.9887	0.9890
2.3	0.9893	0.9896	0.9898	0.9901	0.9904	0.9906	0.9909	0.9911	0.9913	0.9916
2.4	0.9918	0.9920	0.9922	0.9925	0.9927	0.9929	0.9931	0.9932	0.9934	0.9936
2.5	0.9938	0.9940	0.9941	0.9943	0.9945	0.9946	0.9948	0.9949	0.9951	0.9952
2.6	0.9953	0.9955	0.9956	0.9957	0.9959	0.9960	0.9961	0.9962	0.9963	0.9964
2.7	0.9965	0.9966	0.9967	0.9968	0.9969	0.9970	0.9971	0.9972	0.9973	0.9974
2.8	0.9974	0.9975	0.9976	0.9977	0.9977	0.9978	0.9979	0.9979	0.9980	0.9981
2.9	0.9981	0.9982	0.9982	0.9983	0.9984	0.9984	0.9985	0.9985	0.9986	0.9986
3.0	0.9987	0.9987	0.9987	0.9988	0.9988	0.9989	0.9989	0.9989	0.9990	0.9990
3.1	0.9990	0.9991	0.9991	0.9991	0.9992	0.9992	0.9992	0.9992	0.9993	0.9993
3.2	0.9993	0.9993	0.9994	0.9994	0.9994	0.9994	0.9994	0.9995	0.9995	0.9995
3.3	0.9995	0.9995	0.9995	0.9996	0.9996	0.9996	0.9996	0.9996	0.9996	0.9997
3.4	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9997	0.9998
3.5	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998	0.9998
3.6	0.9998	0.9998	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.7	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.8	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999	0.9999
3.9	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000

Tables de la loi de Student : Cette table nous donne les valeurs de t telle que $\mathbb{P}(T > t)$ lorsque T suit une loi de student à ν degrés de liberté

ν	0,10	0,05	0,025	0,01	0,005	0,001
1	3,078	6,314	12,706	31,821	63,657	318,313
2	1,886	2,920	4,303	6,965	9,925	22,327
3	1,638	2,353	3,182	4,541	5,841	10,215
4	1,533	2,132	2,776	3,747	4,604	7,173
5	1,476	2,015	2,571	3,365	4,032	5,893
6	1,440	1,943	2,447	3,143	3,707	5,208
7	1,415	1,895	2,365	2,998	3,499	4,782
8	1,397	1,860	2,306	2,896	3,355	4,499
9	1,383	1,833	2,262	2,821	3,250	4,296
10	1,372	1,812	2,228	2,764	3,169	4,143
11	1,363	1,796	2,201	2,718	3,106	4,024
12	1,356	1,782	2,179	2,681	3,055	3,929
13	1,350	1,771	2,160	2,650	3,012	3,852
14	1,345	1,761	2,145	2,624	2,977	3,787
15	1,341	1,753	2,131	2,602	2,947	3,733
16	1,337	1,746	2,120	2,583	2,921	3,686
17	1,333	1,740	2,110	2,567	2,898	3,646
18	1,330	1,734	2,101	2,552	2,878	3,610
19	1,328	1,729	2,093	2,539	2,861	3,579
20	1,325	1,725	2,086	2,528	2,845	3,552
21	1,323	1,721	2,080	2,518	2,831	3,527
22	1,321	1,717	2,074	2,508	2,819	3,505
23	1,319	1,714	2,069	2,500	2,807	3,485
24	1,318	1,711	2,064	2,492	2,797	3,467
25	1,316	1,708	2,060	2,485	2,787	3,450
26	1,315	1,706	2,056	2,479	2,779	3,435
27	1,314	1,703	2,052	2,473	2,771	3,421
28	1,313	1,701	2,048	2,467	2,763	3,408
29	1,311	1,699	2,045	2,462	2,756	3,396
30	1,310	1,697	2,042	2,457	2,750	3,385
31	1,309	1,696	2,040	2,453	2,744	3,375
32	1,309	1,694	2,037	2,449	2,738	3,365
33	1,308	1,692	2,035	2,445	2,733	3,356
34	1,307	1,691	2,032	2,441	2,728	3,348
35	1,306	1,690	2,030	2,438	2,724	3,340
36	1,306	1,688	2,028	2,434	2,719	3,333
37	1,305	1,687	2,026	2,431	2,715	3,326
38	1,304	1,686	2,024	2,429	2,712	3,319
39	1,304	1,685	2,023	2,426	2,708	3,313
40	1,303	1,684	2,021	2,423	2,704	3,307
50	1,299	1,676	2,009	2,403	2,678	3,261
100	1,290	1,660	1,984	2,364	2,626	3,174
∞	1,282	1,645	1,960	2,326	2,576	3,090

Tables de la loi du χ^2 : Cette table nous donne les valeurs de t telle que $\mathbb{P}(X > t)$ lorsque X suit une loi du chi 2 à ν degrés de liberté

ν	0.3	0.2	0.1	0.05	0.025	0.01	0.005	0.001
1	1.0742	1.6424	2.7055	3.8415	5.0239	6.6349	7.8794	10.8276
2	2.4079	3.2189	4.6052	5.9915	7.3778	9.2103	10.5966	13.8155
3	3.6649	4.6416	6.2514	7.8147	9.3484	11.3449	12.8382	16.2662
4	4.8784	5.9886	7.7794	9.4877	11.1433	13.2767	14.8603	18.4668
5	6.0644	7.2893	9.2364	11.0705	12.8325	15.0863	16.7496	20.5150
6	7.2311	8.5581	10.6446	12.5916	14.4494	16.8119	18.5476	22.4577
7	8.3834	9.8032	12.0170	14.0671	16.0128	18.4753	20.2777	24.3219
8	9.5245	11.0301	13.3616	15.5073	17.5345	20.0902	21.9550	26.1245
9	10.6564	12.2421	14.6837	16.9190	19.0228	21.6660	23.5894	27.8772
10	11.7807	13.4420	15.9872	18.3070	20.4832	23.2093	25.1882	29.5883
11	12.8987	14.6314	17.2750	19.6751	21.9200	24.7250	26.7568	31.2641
12	14.0111	15.8120	18.5493	21.0261	23.3367	26.2170	28.2995	32.9095
13	15.1187	16.9848	19.8119	22.3620	24.7356	27.6882	29.8195	34.5282
14	16.2221	18.1508	21.0641	23.6848	26.1189	29.1412	31.3193	36.1233
15	17.3217	19.3107	22.3071	24.9958	27.4884	30.5779	32.8013	37.6973
16	18.4179	20.4651	23.5418	26.2962	28.8454	31.9999	34.2672	39.2524
17	19.5110	21.6146	24.7690	27.5871	30.1910	33.4087	35.7185	40.7902
18	20.6014	22.7595	25.9894	28.8693	31.5264	34.8053	37.1565	42.3124
19	21.6891	23.9004	27.2036	30.1435	32.8523	36.1909	38.5823	43.8202
20	22.7745	25.0375	28.4120	31.4104	34.1696	37.5662	39.9968	45.3147
21	23.8578	26.1711	29.6151	32.6706	35.4789	38.9322	41.4011	46.7970
22	24.9390	27.3015	30.8133	33.9244	36.7807	40.2894	42.7957	48.2679
23	26.0184	28.4288	32.0069	35.1725	38.0756	41.6384	44.1813	49.7282
24	27.0960	29.5533	33.1962	36.4150	39.3641	42.9798	45.5585	51.1786
25	28.1719	30.6752	34.3816	37.6525	40.6465	44.3141	46.9279	52.6197
26	29.2463	31.7946	35.5632	38.8851	41.9232	45.6417	48.2899	54.0520
27	30.3193	32.9117	36.7412	40.1133	43.1945	46.9629	49.6449	55.4760
28	31.3909	34.0266	37.9159	41.3371	44.4608	48.2782	50.9934	56.8923
29	32.4612	35.1394	39.0875	42.5570	45.7223	49.5879	52.3356	58.3012
30	33.5302	36.2502	40.2560	43.7730	46.9792	50.8922	53.6720	59.7031
31	34.5981	37.3591	41.4217	44.9853	48.2319	52.1914	55.0027	61.0983
32	35.6649	38.4663	42.5847	46.1943	49.4804	53.4858	56.3281	62.4872
33	36.7307	39.5718	43.7452	47.3999	50.7251	54.7755	57.6484	63.8701
34	37.7954	40.6756	44.9032	48.6024	51.9660	56.0609	58.9639	65.2472
35	38.8591	41.7780	46.0588	49.8018	53.2033	57.3421	60.2748	66.6188
36	39.9220	42.8788	47.2122	50.9985	54.4373	58.6192	61.5812	67.9852
37	40.9839	43.9782	48.3634	52.1923	55.6680	59.8925	62.8833	69.3465
38	42.0451	45.0763	49.5126	53.3835	56.8955	61.1621	64.1814	70.7029
39	43.1053	46.1730	50.6598	54.5722	58.1201	62.4281	65.4756	72.0547
40	44.1649	47.2685	51.8051	55.7585	59.3417	63.6907	66.7660	73.4020

Soit $Y \sim \mathcal{N}(0,1)$ une variable aléatoire normale centrée réduite. Soit $Y_1, Y_2, ..., Y_n$ un échantillon de n réalisation i.i.d. de cette variable aléatoire.

La loi du χ^2 On appelle loi du χ^2 à n degrés de libertés la loi de la variable aléatoire $Z_n = \sum_{i=1}^n Y_i^2$

La loi de student On appelle loi de student à n degrés de libertés la loi de la variable aléatoire T_n

$$T_n = \frac{N}{\sqrt{\frac{X_n}{n}}} \qquad \begin{array}{l} N \sim \mathcal{N}(0,1) \\ \\ X_n \sim \chi^2_n \end{array}$$

Theorem 0.1 (Théorème du χ^2 (Pearson)) *pour N_{ij} effectif observés et pour n_{ij} effectif théorique*

$$X_{ij} = \frac{N_{ij} - n\,\widehat{p}_{ij}}{\sqrt{n\,\widehat{p}_{ij}}} \qquad \sum_{i=1}^{I}\sum_{j=1}^{J} X_{ij}^2 \longrightarrow \chi^2_{(I-1)(J-1)}$$

la variable $T_{n_x+n_y-2}$ suit une loi de student à $n_x + n_y - 2$ degrés de liberté :

$$T_{n_x+n_y-2} = \sqrt{n_x + n_y - 2}\,\frac{\overline{X} - \overline{Y} - (\mu_x - \mu_y)}{\sqrt{\left(\frac{1}{n_x} + \frac{1}{n_y}\right) S_{xy}^2}}$$

avec $S_{xy}^2 = S_x^2 + S_y^2 = \sum_{i=1}^{n_x}(X_i - \overline{X})^2 + \sum_{i=1}^{n_y}(Y_i - \overline{Y})^2$

Mise en œuvre du test du χ^2

1. on construit un tableau de contingence O des observations (2 variables qualitatives de respectivement I et J modalités)
2. on calcule les marginales $p_i = \frac{1}{n}\sum_{j=1}^J O_{ij}$
3. on calcule pour chaque case du tableau des effectifs théoriques $T_{ij} = np_ip_j$ (en supposant l'indépendance)
4. on calcule la distance du χ^2 $D(O,T) = \sum_{i=1}^{I}\sum_{j=1}^{J}\frac{(O_{ij} - T_{ij})^2}{T_{ij}}$
5. on calcule le nombre de degrés de liberté du χ^2 : $d = (I-1)(J-1)$
6. on regarde dans les tables d'une variable aléatoire Z distribué suivant une loi χ^2 à d degrés de liberté la p-valeur de $D(O,T)$: pval $= \mathbb{P}\big(Z \geq D(O,T)\big)$
7. on décide qu'on ne peut pas conclure à la dépendance si la p-valeur est supérieure à 0,05, si $pval \geq 0,05$

Mise en œuvre du test de comparaison des moyennes (T test ou test de student)

1. la question : les deux groupes sont ils des réalisations de la même loi
2. le modèle : gaussien
3. les hypothèses : même variance σ^2 inconnue
4. caclul de $t = \frac{\overline{x}_t - \overline{x}_p}{\sqrt{\widehat{\sigma}^2\left(\frac{1}{n_t} + \frac{1}{n_p}\right)}}$

   $\overline{x}_t$ moyenne avec traitement
   $\overline{x}_p$ moyenne sans traitement
   $\widehat{\sigma}^2 = \frac{1}{n_t+n_p-2}\left(\sum_{i=1}^{n_t}(x_{ti} - \overline{x}_t)^2 + \sum_{i=1}^{n_p}(x_{pi} - \overline{x}_p)^2\right)$
   n_t nombre de cas avec traitement
   n_p nombre de cas sans traitement
5. calcul de la p-valeur $T \sim \mathcal{T}_{n_t+n_p-2}$ (ou lecture sur les tables) $pval = \mathbb{P}(T \leq t)$ ou $pval = \mathbb{P}(T \geq t)$ ou $pval = \mathbb{P}(-|t| \leq T \leq |t|)$
6. on décide que les deux groupes sont ils des réalisation de la même loi si la p-valeur est supérieure à 0,05, si $pval \geq 0,05$

Mise en œuvre du test sur la pente de la régression

1. les hypothèses : $\begin{cases} \mathcal{H}_0 : \text{indépendance} & a = 0 \\ \mathcal{H}_1 : \text{dépendance} & a \neq 0 \end{cases}$
2. caclul de $\widehat{a} = \frac{\sum_{i=1}^n (x_i - \overline{x})(y_i - \overline{y})}{\sum_{i=1}^n (x_i - \overline{x})^2}$
3. calcul de $\widehat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n \big(y_i - (\widehat{a}x_i + \widehat{b})\big)^2$ et de $S_x^2 = \sum_{i=1}^n (x_i - \overline{x})^2$
4. caclul de $t = \frac{\widehat{a}}{\sqrt{\frac{\widehat{\sigma}^2}{S_x^2}}}$
5. calcul du nombre de degrés de liberté $d = n - 2$
6. calcul de la p-valeur $T \sim \mathcal{T}_d$ (ou lecture sur les tables)

$$pval = \mathbb{P}(|T| \geq t)$$

7. on décide de garder $\mathcal{H}_0$ $(a = 0)$ si la p-valeur est supérieure à 0,05, si $pval \geq 0,05$