

Sauf indication contraire, on considèrera dans les exercices un risque de première espèce de 5%. Pensez à vérifier la cohérence des données et à éliminer s'il y a lieu et en le justifiant les mesures hors épure.

Exercice 1**The mexican hat****10 points**

Les observations de deux variables x et y ont données les résultats suivants :

x	y	x	y (suite)
0.0086	0.9218	0.6669	1.0555
0.0278	0.6985	0.6896	0.7939
0.0349	0.8236	0.7180	0.5309
0.0428	0.6890	0.7284	0.3656
0.0441	0.3969	0.7322	0.2657
0.0893	0.4636	0.7663	0.1466
0.1360	0.0463	0.7943	0.1428
0.1489	0.1235	0.8515	0.0630
0.1628	0.0772	0.8707	0.2168
0.1693	0.1889	0.8878	0.0342
0.1831	-0.1305	0.8907	0.0034
0.2185	0.0887	0.8953	0.1919
0.2469	0.2263	0.9005	0.1273
0.2478	0.4994	0.9598	0.7720
0.2729	0.8094	0.9786	0.5822
0.2808	0.5909	0.9856	0.7871
0.3424	0.8754	1.1069	1.0139
0.3720	1.3378	1.1653	0.9150
0.4105	1.5254	1.1838	0.8520
0.4138	1.7573	1.1937	1.1609
0.4248	2.0624	1.2135	1.1338
0.4745	1.9662	1.3720	1.2543
0.5352	2.0036	1.4483	1.0959
0.5363	2.0906	1.4531	0.9170
0.5434	1.7758	1.4725	0.9587
0.5772	1.7417	1.5479	1.1766
0.5901	1.5310	1.6564	0.9713
0.5909	1.8571	1.8226	0.7482
0.5952	1.8384	1.8795	0.8489
0.6009	1.4793	1.8868	1.0969
0.6046	1.7453	2.9232	3.3241

1. poser un modèle de régression linéaire entre x et y , identifier les paramètres du modèle, visualiser les observations et la droite de régression estimée,
2. quelle est la qualité du modèle de régression linéaire que vous avez posé,
3. y a t'il des point aberrants, si oui éliminez les et refaites les calculs de la régression,
4. peut on dire qu'il y a une relation entre ces deux variables x et y
5. un expert suggère d'utiliser le modèle continu suivant intégrant la fonction sinus

$$y_i = \begin{cases} a \sin(3\pi x_i) + b + \varepsilon_i & \text{si } x_i < 1 \\ c + \varepsilon_i & \text{si } x_i > 1 \end{cases} \quad i = 1, \dots, n$$

où a , b et c sont trois paramètres inconnus à estimer à partir des données et ε_i le bruit modélisé par une variable aléatoire distribuée suivant une loi normale centrée de variance σ^2 inconnue. Reformulez ce modèle sous la forme $\mathbf{y} = X\alpha + \varepsilon$ où \mathbf{y} est le vecteur de terme général y_i et X une matrice que l'on explicitera

6. identifier les paramètres du modèle et visualiser les observations et le modèle de régression
7. quelle est la qualité du modèle de régression linéaire que vous avez posé, que pensez vous des résidus.

Exercice 2**Chose promise, chose due****2 points**Expliquez la régression, du début jusqu'à la fin, comme nous l'avons vue en cours.

Exercice 3**Cusé aime****3 points**

A chacune des questions indiquez la ou les réponses exactes

- On se demande si un nouveau médicament M est plus efficace que le traitement traditionnel T. On réalise un essai thérapeutique en double aveugle où la moitié des patients (tirés au sort) reçoit M, et l'autre reçoit T. Ces deux groupes de patients sont notés M et T. Malheureusement, la pharmacie s'est trompée dans la distribution des médicaments : tous les malades (T et M) ont reçu T, ce dont ni les malades ni les médecins ne se sont aperçus car l'essai est en double aveugle. Le statisticien effectue le test de comparaison entre les groupes M et T. Lui non plus ne connaît pas l'erreur de la pharmacie. Le test est effectué, comme toujours, au risque de première espèce 5 %. La puissance de l'étude a été établie à 90 %.
 - Il y a 5 chances sur 100 pour que le test conduise à déclarer que T et M ont des effets différents
 - Il y a 95 chances sur 100 pour que le test conduise à déclarer que T et M ont des effets différents
 - Il y a 10 chances sur 100 pour que le test conduise à déclarer que T et M ont des effets différents
 - Il y a 90 chances sur 100 pour que le test conduise à déclarer que T et M ont des effets différents
 - Il y a 0 chance sur 100 pour que le test conduise à déclarer que T et M ont des effets différents
- On se demande si un traitement T modifie la glycémie des malades qui le reçoivent. On mesure la glycémie des sujets de deux groupes de 49 patients : les patients du premier groupe (groupe T) sont traités par T ; ceux de l'autre groupe ne sont pas traités (groupe NT). Les groupes sont constitués par tirage au sort, et on compare leur moyenne. Les moyennes et variances observées dans les deux groupes sont : $m_T = 5,9$ mmol/ml, $m_{NT} = 5,5$ mmol/ml, $s^2_T = 0,4$, $s^2_{NT} = 0,6$.
 - L'hypothèse nulle testée est que les moyennes observées m_T et m_{NT} sont identiques
 - les moyennes observées m_T et m_{NT} diffèrent significativement
 - la p valeur est comprise entre 1 % et 1 ‰
 - la p valeur est comprise entre 5 % et 1 %
 - Le test est valide car les tailles des groupes sont suffisantes
- Un nouvel examen biologique permet de distinguer trois formes différentes, A, B et C, d'une maladie intestinale connue depuis longtemps et que l'on diagnostique facilement. On se place dans la population des sujets présentant cette maladie ; on les appelle malades. On considère 219 malades. Après avoir recherché la forme de la maladie et interrogé ces malades sur leurs douleurs, vous disposez du tableau d'effectifs suivant :

	A	B	C
Présence de douleurs	38	50	56
Absence de douleurs	34	22	19

Pour répondre à la question précédente, on réalise un test d'hypothèses.

- L'hypothèse nulle peut s'exprimer ainsi : forme de la maladie et douleur sont liées
 - L'hypothèse nulle peut s'exprimer ainsi : forme de la maladie et douleur sont indépendantes
 - L'égalité $\Pr(\text{forme de la maladie}=A \text{ et douleur}=présente) = \Pr(\text{forme de la maladie}=A)$ est toujours juste sous H_0
 - Le paramètre du test suit, sous H_0 , un χ^2 à 2 degrés de liberté. On sait déjà que le test peut être réalisé car tous les effectifs du tableau sont supérieurs à 5.
-

Exercice 4**T'es Steu Duqui 2****2 points**

D'après le tableau d'effectifs vu dans la question précédente :

	A	B	C
Présence de douleurs	38	50	56
Absence de douleurs	34	22	19

- forme de la maladie et douleur sont indépendantes ?
-

Exercice 5**Chtu dent****3 points**

Le tableau suivant montre les résultats de 15 étudiants à un examen avant et après avoir suivi un cours.

étudiant	Note avant	Note après	Différence
1	18	22	+4
2	21	25	+4
3	16	17	+1
4	22	24	+2
5	19	16	-3
6	24	29	+5
7	17	20	+3
8	21	23	+2
9	23	19	-4
10	18	20	+2
11	14	15	+1
12	16	15	-1
13	16	18	+2
14	18	18	0
15	17	16	-1

1. En faisant l'hypothèse que les notes avant et après sont deux variables aléatoires distribuées suivant une loi normale dont on précisera les paramètres, quelle serait la loi suivie par leur différence ?
2. Peut on affirmer que le cours à permis aux étudiants d'améliorer leur scores ?

Tables de la loi normale : Cette table nous donne les valeurs de t telle que $P(X \leq t)$ lorsque X suit une loi normale 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	0.5	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	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0

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

Calcul matriciel

- vecteur \mathbf{v} est un vecteur colonne de taille n
- transposée $\mathbf{v}^\top = (v_1, \dots, v_n)$ transforme un vecteur colonne en un vecteur ligne
- produit scalaire $p = \mathbf{x}^\top \mathbf{y} = \sum_{i=1}^n x_i y_i$
- produit extérieur $M = \mathbf{xy}^\top$ est une matrice
- norme d'un vecteur $\|\mathbf{v}\|^2 = \mathbf{v}^\top \mathbf{v} = \sum_{i=1}^n v_i^2$
- matrice M : n lignes et p colonnes
- norme matricielle $\|M\|_F^2 = \sum_{i=1}^n \sum_{j=1}^p M_{ij}^2$
- transposée M^\top la matrice telle que $M_{ij}^\top = M_{ji}$
- produit matrice vecteur $\mathbf{u} = M\mathbf{v}$ est un vecteur avec $u_i = \sum_{j=1}^p M_{ij} v_j$
- produit matrice matrice $C = AB$, $C_{ij} = \sum_{k=1}^p A_{ik} B_{kj}$
- gradient d'une forme linéaire : $\nabla_{\mathbf{x}} (\mathbf{a}^\top \mathbf{x}) = \mathbf{a}$
- gradient d'une forme quadratique : $\nabla_{\mathbf{x}} (\mathbf{x}^\top A \mathbf{x}) = (A + A^\top) \mathbf{x}$

Variables aléatoires et lois

- Soit $Y \sim \mathcal{N}(0, 1)$ une variable aléatoire normale centrée réduite.
- Soit Y_1, Y_2, \dots, 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

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

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

$$X_{ij} = \frac{N_{ij} - n \hat{p}_{ij}}{\sqrt{n \hat{p}_{ij}}} \quad \sum_{i=1}^I \sum_{j=1}^J X_{ij}^2 \rightarrow \chi_{(I-1)(J-1)}^2$$

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

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

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

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 $T_{ij} = np_i p_j$ (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 $D(O, T)$: $pval = \mathbb{P}(Z \geq D(O, T))$
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. calcul de

$$t = \frac{\bar{x}_t - \bar{x}_p}{\sqrt{\hat{\sigma}^2 \left(\frac{1}{n_t} + \frac{1}{n_p} \right)}} \quad \hat{\sigma}^2 = \frac{1}{n_t + n_p - 2} \left(\sum_{i=1}^{n_t} (x_{ti} - \bar{x}_t)^2 + \sum_{i=1}^{n_p} (x_{pi} - \bar{x}_p)^2 \right)$$

\bar{x}_t moyenne avec traitement
 \bar{x}_p moyenne sans traitement
 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. calcul de $\hat{a} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$

3. calcul de $\hat{\sigma}^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - (\hat{a}x_i + \hat{b}))^2$ et de $S_x^2 = \sum_{i=1}^n (x_i - \bar{x})^2$

4. calcul de $t = \frac{\hat{a}}{\sqrt{\frac{\hat{\sigma}^2}{S_x^2}}}$

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 $pval \geq 0,05$

La régression linéaire :

1. modèle linéaire : $y = \sum_{j=1}^p x_j \alpha_j + \alpha_0 + \varepsilon = X\alpha + \varepsilon; \quad \varepsilon \sim \mathcal{N}(0, \sigma^2 I)$

2. estimateur des moindres carrés : $\hat{\alpha} = (X^\top X)^{-1} X^\top y$

3. estimateur des résidus : $\hat{\varepsilon}_i = y_i - \hat{y}_i$ avec $\hat{y} = X\hat{\alpha}$

4. estimateur de la variance des résidus : $\hat{\sigma}^2 = s^2 = \frac{1}{n-p} \hat{\varepsilon}^\top \hat{\varepsilon}$

5. $R^2 = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$

6. la matrice d'influence : $H = X (X^\top X)^{-1} X^\top$

7. les contributions (ou distance de Cook) : $c_i = \frac{H_{ii}}{p(1-H_{ii})^2} \frac{\hat{\varepsilon}_i^2}{s^2}$