I Quiz: TD_CA

The following table represents the type of studies pursued (university, preparatory classes, others) according to the high school pathway followed (1970 nomenclature):

A: Literature,

B: Economics,

C: Sciences (mathematics, physics),

D: Natural Sciences (biology, chemistry, ...),

	Univ	Prep Classes	Others
A	13	2	5
B	20	2	8
C	10	5	5
D	7	1	22

Question 1

What do the values in the table represent ?

Question 2

What is their nature?

Question 3

What are the main objectives of performing an CA on this dataset?

Question 4

How do we check if there is a significant ***association*** between high school pathway and type of study?

Question 5

What are the conditions of perform a `χ^2-Test of Independence

Question 6

1. Calculate the Chi-square distance between the observed table and the table under independence.

Hint:

State the null and alternative hypotheses of the test.

Provide the table of theoretical frequencies.

Specify the conditions for applying the test. Are they met?

Give the Chi-Square test statistic and its distribution under the null hypothesis.

Verify that the observed value of the Chi-Square statistic. State the decision rule for the test.

We aim to calculate the results of the CA by answer the following questions

Question 7

Deduce the inertia, interpret the results

Question 8

Calculate row profiles (relative frequencies for each row) and column profiles (relative frequencies for each column).

Question 9

Provide the centers of gravity, G_r and G_c, associated with the row profiles and column profiles.

Question 10

Provide the diagonal matrices of row profiles D_r and column profiles .

Question 11

Provide the matrices of row profiles L and column profiles C.

Question 12

Calculate the two matrices $A_r := L^tC$ and $A_c := C^tL$.

Question 13

Calculate the eigenvalues and eigenvectors of A_r and A_c.

Question 14

What does G_r (respectively G_c) represent for the matrix A_r (respectively A_c)?

Let $Y_r := L - 1_3 g_r^t$ and $Y_c := C - 1_4 g_c^t$ be the centered matrices of row profiles and column profiles, respectively. The ***covariance matrices*** of the row profiles and column profiles are defined as:

$$\Sigma_r := Y_r^t D_r Y_r \quad \text{and} \quad \Sigma_c := Y_c^t D_c Y_c$$

The correspondence analysis is essentially based on the two matrices $\Sigma_r D_c^{-1}$ and $\Sigma_c D_r^{-1}$,

Question 15

Deduce the eigenvalues and eigenvectors of and $\Sigma_c D_r^{-1}$.

Question 16

Compute the total inertia of the row profiles L with respect to their center of gravity g_r and deduce the total inertia of the column profiles with respect to their center of gravity g_c.

Question 17

What are the percentages of inertia relative to the principal axes for both profiles?

Exercise 2 :Chi-Square Test of Independence

We survey 1587 second-year master's (M2) students about their parents' socio-professional category. The students are enrolled in different programs: engineering schools, business schools, and scientific universities. The results are as follows:

	Workers	Employees	Executives	Liberal Professions
Engineering Schools	50	280	120	20
Business Schools	8	29	210	350
Scientific Universities	150	230	100	40

We aim to study the influence of the socio-professional background of parents on the type of education chosen by their children.

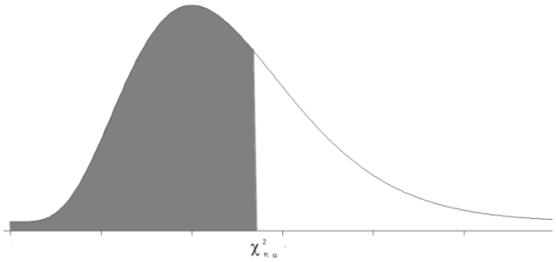
Question 18

1.What are the variables being studied, and what is their nature?

Question 19

2. Chi-Square test of independence χ^2 is performed between the two variables.

TABLE DE LA LOI KHI-DEUX



n\α	0.5 %	1 %	2.5 %	5 %	10 %	50 %	90 %	95 %	97.5 %	99 %	99.5 %
1	0.000	0.000	0.001	0.004	0.016	0.455	2.706	3.841	5.024	6.635	7.879
2	0.010	0.020	0.051	0.103	0.211	1.386	4.605	5.991	7.378	9.210	10.597
3	0.072	0.115	0.216	0.352	0.584	2.366	6.251	7.815	9.348	11.345	12.838
4	0.207	0.297	0.484	0.711	1.064	3.357	7.779	9.488	11.143	13.277	14.860
5	0.412	0.554	0.831	1.145	1.610	4.351	9.236	11.070	12.833	15.086	16.750
6	0.676	0.872	1.237	1.635	2.204	5.348	10.645	12.592	14.449	16.812	18.548
7	0.989	1.239	1.690	2.167	2.833	6.346	12.017	14.067	16.013	18.475	20.278
8	1.344	1.646	2.180	2.733	3.490	7.344	13.362	15.507	17.535	20.090	21.955
9	1.735	2.088	2.700	3.325	4.168	8.343	14.684	16.919	19.023	21.666	23.589
10	2.156	2.558	3.247	3.940	4.865	9.342	15.987	18.307	20.483	23.209	25.188
11	2.603	3.053	3.816	4.575	5.578	10.341	17.275	19.675	21.920	24.725	26.757
12	3.074	3.571	4.404	5.226	6.304	11.340	18.549	21.026	23.337	26.217	28.300
13	3.565	4.107	5.009	5.892	7.042	12.340	19.812	22.362	24.736	27.688	29.819
14	4.075	4.660	5.629	6.571	7.790	13.339	21.064	23.685	26.119	29.141	31.319
15	4.601	5.229	6.262	7.261	8.547	14.339	22.307	24.996	27.488	30.578	32.801
16	5.142	5.812	6.908	7.962	9.312	15.338	23.542	26.296	28.845	32.000	34.267
17	5.697	6.408	7.564	8.672	10.085	16.338	24.769	27.587	30.191	33.409	35.718
18	6.265	7.015	8.231	9.390	10.865	17.338	25.989	28.869	31.526	34.805	37.156
19	6.844	7.633	8.907	10.117	11.651	18.338	27.204	30.144	32.852	36.191	38.582
20	7.434	8.260	9.591	10.851	12.443	19.337	28.412	31.410	34.170	37.566	39.997
21	8.034	8.897	10.283	11.591	13.240	20.337	29.615	32.671	35.479	38.932	41.401
22	8.643	9.542	10.982	12.338	14.041	21.337	30.813	33.924	36.781	40.289	42.796
23	9.260	10.196	11.689	13.091	14.848	22.337	32.007	35.172	38.076	41.638	44.181
24	9.886	10.856	12.401	13.848	15.659	23.337	33.196	36.415	39.364	42.980	45.559
25	10.520	11.524	13.120	14.611	16.473	24.337	34.382	37.652	40.646	44.314	46.928
26	11.160	12.198	13.844	15.379	17.292	25.336	35.563	38.885	41.923	45.642	48.290
27	11.808	12.879	14.573	16.151	18.114	26.336	36.741	40.113	43.195	46.963	49.645
28	12.461	13.565	15.308	16.928	18.939	27.336	37.916	41.337	44.461	48.278	50.993
29	13.121	14.256	16.047	17.708	19.768	28.336	39.087	42.557	45.722	49.588	52.336
30	13.787	14.953	16.791	18.493	20.599	29.336	40.256	43.773	46.979	50.892	53.672
31	14.458	15.655	17.539	19.281	21.434	30.336	41.422	44.985	48.232	52.191	55.003
32	15.134	16.362	18.291	20.072	22.271	31.336	42.585	46.194	49.480	53.486	56.328
33	15.815	17.074	19.047	20.867	23.110	32.336	43.745	47.400	50.725	54.776	57.648
34	16.501	17.789	19.806	21.664	23.952	33.336	44.903	48.602	51.966	56.061	58.964
35	17.192	18.509	20.569	22.465	24.797	34.336	46.059	49.802	53.203	57.342	60.275

$\chi^2_{n;\alpha} \simeq \frac{\left((z_\alpha+\sqrt{2n-1})\right)^2}{2}$ si n grand