



TP – Description des données. Estimation statistique

Exercice I.1.1

Dans une étude sur l’association entre le tabagisme et la maladie coronarienne, une équipe de chercheurs a obtenu la répartition suivante des valeurs de pression systolique (mm Hg) pour un groupe de 37 fumeurs.

La pression artérielle systolique	Les fréquences absolues	Les fréquences relatives (%)	Les fréquences cumulées (%)
89.5 - 109.5	5	13.5	13.5
109.5 - 129.5	15	40.5	54.0
129.5 - 149.5	A	27.0	C
149.5 - 169.5	3	8.1	89.1
169.5 - 189.5	2	B	94.5
189.5 - 209.5	D	5.4	99.9
TOTAL	37		100.0

1. La valeur **A** dans le tableau est la suivante:
2. La valeur **B** dans le tableau est la suivante:
3. La valeur **C** dans le tableau est la suivante:
4. La probabilité qu’un individu choisi au hasard à partir de cet échantillon de 37 fumeurs a une pression systolique entre 89.5 mm Hg et 129.5 mm Hg est :

✓ **Solution I.1.1**

1. La valeur **A** dans le tableau est la suivante:  $37 \times 0.27 = 9.99 \Rightarrow \mathbf{A=10}$

2. La valeur **B** dans le tableau est la suivante:  $\frac{2}{37} \times 100 = 5.4 \Rightarrow \mathbf{B=5.4}$

3. La valeur **C** dans le tableau est la suivante:  $54.0 + 27.0 = 81.0 \Rightarrow \mathbf{C=81.0}$

4. La probabilité qu’un individu choisi au hasard à partir de cet échantillon de 37 fumeurs a une pression systolique entre 89.5 mm Hg et 129.5 mm Hg est : **54.0%** (colonne des fréquences cumulées, ligne 2)



### Exercice I.1.2

En prévision d'un test national, un échantillon de 200 étudiants ont reçu un test simulé qui avaient répondu à un total de 100 questions. Chaque étudiant a répondu correctement à un certain nombre de questions entre 35 et 59. La moyenne du nombre de réponses correctes était 47, et l'écart type 4. On suppose que le nombre de réponses correctes suivent une distribution normale.

1. Le pourcentage d'étudiants ayant répondu correctement à un certain nombre de questions entre 43 et 51 est environ:
2. Le pourcentage d'étudiants ayant répondu correctement à au moins 55 questions est environ:

✓ **Solution I.1.2**

1. Si la moyenne est  $m = 47$  et l'écart-type  $s = 4$ , alors dans l'intervalle  $(47 \pm 4) = (43; 51)$  se trouve environ 68% des valeurs (dans ce cas, nous supposons que les valeurs suivent une distribution normale).

**Conclusion:** 68% des étudiants ont correctement répondu à un certain nombre de questions entre 43 et 51.

2. Si la moyenne est  $m = 47$  et l'écart-type  $s = 4$ , alors dans l'intervalle  $(47 \pm 2 * 4) = (39; 55)$  se trouve environ 95% des valeurs (dans ce cas, nous supposons que les valeurs suivent une distribution normale). Ceci signifie que seulement environ  $(1 - 0.95)/2 = 0.025$  des valeurs sont supérieures ou égales à 55.

**Conclusion:** environ 2.5% des étudiants ont répondu correctement au moins 55 questions.

### Exercice I.1.3

Le center des diabètes et de la nutrition de Timișoara possède un grand nombre de patients de qui viennent à notre attention qu'un échantillon de 32 patients. La glycémie a été déterminée avec un appareil portable (gluco-test, avec bande). Les données sont présentées dans le tableau suivant :



	Genre	GLICEMIE mg%
1	M	127
2	F	130
3	F	126
4	M	111
5	F	109
6	M	99
7	F	116
8	F	116
9	F	150
10	M	95
11	F	120
12	M	100
13	M	116
14	M	116
15	M	114
16	M	108

17	M	112
18	M	98
19	M	115
20	F	109
21	F	106
22	F	103
23	M	97
23	F	95
25	M	119
26	M	117
27	M	105
28	M	137
29	F	106
30	F	101
31	F	129
32	M	132

1. Pour un niveau de confiance de  $1 - \alpha = 0.68$ , déterminez l'intervalle de confiance pour les valeurs de glycémie. Interprétez le résultat (Formulez la conclusion).
2. Estimez la moyenne pour les valeurs de glycémie de ces patients,  $1 - \alpha = 0.95$ . Interprétez le résultat (Formulez la conclusion).

✓ **Solution I.1.3**

Par souci de simplicité, ci-dessous, nous supposons que les valeurs de glycémie suivent une distribution normale. Les deux versions sont généralement considérées correctes, la solution rigoureuse (en utilisant la distribution de t-Student) et la simplification (en utilisant la distribution normale et des scores Z).

Les valeurs moyennes de glucose dans le sang est  $m = 113.56$  et l'écart type  $s = 13.09$  (On peut les déterminées en utilisant *Microsoft Excel*, *EpiInfo* ou même par calcul manuel).



1. Pour  $1 - \alpha = 0.68$ ,  $z \approx 1$ , l'intervalle de confiance pour les différentes valeurs de glycémie sera  $(113.56 \pm 1 * 13.09) = (100.47; 126.65)$

**Conclusion:** on peut dire qu'environ 68% des valeurs de glycémie se situent entre 100.47 et 126.65

2. Pour  $1 - \alpha = 0.95$ ,  $z \approx 2$ , l'intervalle de confiance pour les valeurs moyennes de glycémie sera  $\left( 113.56 \pm 2 * \frac{13.09}{\sqrt{32}} \right) = (108.94; 118.19)$

**Conclusion:** on peut dire que les valeurs moyennes de glycémie (pour la population à partir de laquelle l'échantillon a été tiré) se situent entre 108.94 et 118.19, avec une probabilité de 95%.