L'INFERENCE STATISTIQUE

1. DEFINTION

Induire les caractéristiques **inconnues** de la population à partir **d'observations partielles** avec une certaine **marge d'erreur**

2. RAPPELS

ENQUETES EPIDEMIOLOGIQUES :

- **Enquêtes générales ou exhaustives** : *Population entière*
- **Enquêtes partielles :** *Une partie réduite de la population*

Partie réduite :

- Echantillon représentatif

ENQUETES PAR SONDAGE

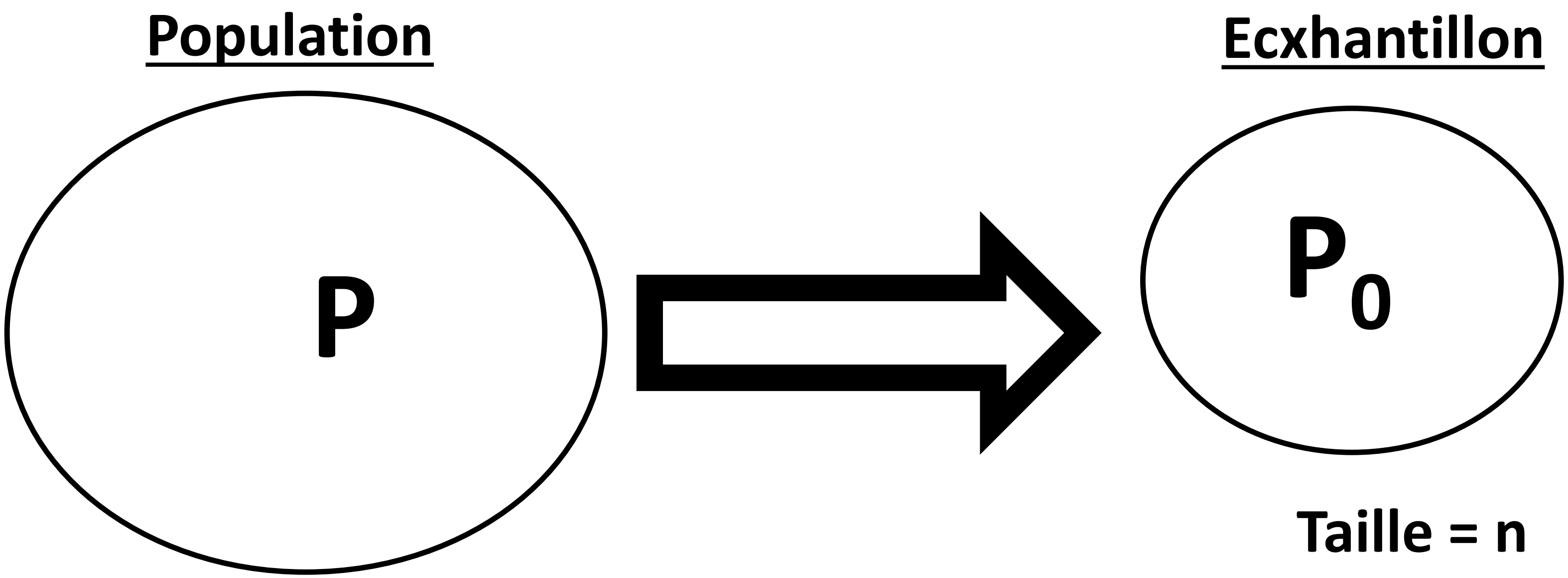
Statistique descriptive = Analyse numérique :

- Paramètres de réduction

  - Moyenne
  - Ecart-type

3. CARACTERE QUALITATIF :

3.1. Position du problème :

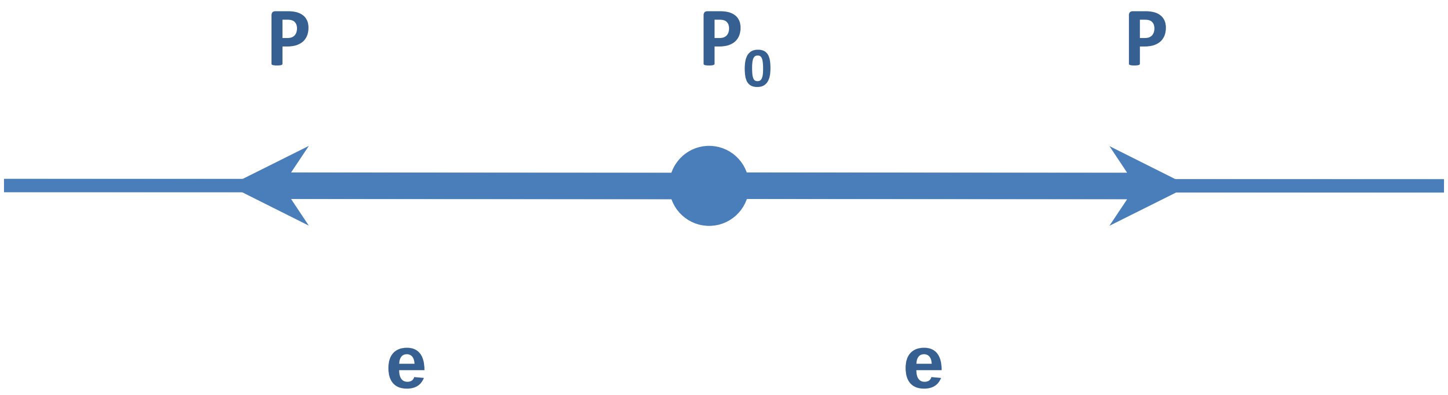


P_o Connue
P Inconnue ?

Quelle est la valeur que peut prendre P ?

3.2. Hypothèse :

- P proche de P_0
- P est dans un intervalle autour de P_0



$$P = P_0 \pm e$$

3.3. Intervalle de confiance :

Intervalle de confiance allant de $P_0 - e$ à $P_0 + e$

Intervalle de confiance = 2 e

Intervalle de confiance symétrique autour de P_0

Intervalle de confiance associé à un risque d'erreur α
Risque d'erreur α associé à l'écart réduit z

e = Ecart-type (s) x Ecart-réduit (z)

➢**Ecart-type =** Ecart-type de la proportion P_0

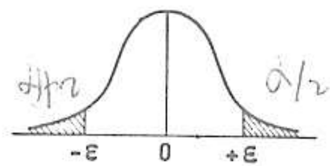
$$\textbf{Ecart-type} = \mathbf{s} = \sqrt{P_0 Q_0 / n}$$

$$(Q_0 = 1 - P_0)$$

➢**Ecart-Réduit = *z* :** Table de l'écart-réduit *(loi normale)*

Table de l'écart-réduit (loi normale) (*)

La table donne la probabilité α pour que l'écart-réduit égale ou dépasse, en valeur absolue, une valeur donnée ε, c'est-à-dire la probabilité extérieure à l'intervalle $(-\varepsilon, +\varepsilon)$.



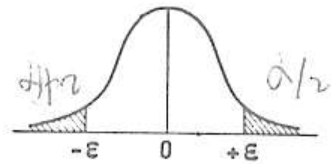
α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	∞	2,576	2,326	2,170	2,054	1,960	1,881	1,812	1,751	1,695
0,10	1,645	1,598	1,555	1,514	1,476	1,440	1,405	1,372	1,341	1,311
0,20	1,282	1,254	1,227	1,200	1,175	1,150	1,126	1,103	1,080	1,058
0,30	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,860
0,40	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,706	0,690
0,50	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,553	0,539
0,60	0,524	0,510	0,496	0,482	0,468	0,454	0,440	0,426	0,412	0,399
0,70	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,279	0,266
0,80	0,253	0,240	0,228	0,215	0,202	0,189	0,176	0,164	0,151	0,138
0,90	0,126	0,113	0,100	0,088	0,075	0,063	0,050	0,038	0,025	0,013

La probabilité α s'obtient par addition des nombres inscrits en marge.

Exemple : Pour $\varepsilon = 1{,}960$ la probabilité est $\alpha = 0{,}00 + 0{,}05 = 0{,}05$.

Table de l'écart-réduit (loi normale) (*)

La table donne la probabilité α pour que l'écart-réduit égale ou dépasse, en valeur absolue, une valeur donnée ε, c'est-à-dire la probabilité extérieure à l'intervalle $(-\varepsilon, +\varepsilon)$.



α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	∞	2,576	2,326	2,170	2,054	1,960	1,881	1,812	1,751	1,695
0,10	1,645	1,598	1,555	1,514	1,476	1,440	1,405	1,372	1,341	1,311
0,20	1,282	1,254	1,227	1,200	1,175	1,150	1,126	1,103	1,080	1,058
0,30	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,860
0,40	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,706	0,690
0,50	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,553	0,539
0,60	0,524	0,510	0,496	0,482	0,468	0,454	0,440	0,426	0,412	0,399
0,70	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,279	0,266
0,80	0,253	0,240	0,228	0,215	0,202	0,189	0,176	0,164	0,151	0,138
0,90	0,126	0,113	0,100	0,088	0,075	0,063	0,050	0,038	0,025	0,013

La probabilité α s'obtient par addition des nombres inscrits en marge.

Exemple : Pour $\varepsilon = 1{,}960$ la probabilité est $\alpha = 0{,}00 + 0{,}05 = 0{,}05$.

α = 5 % : 0.05 : 0.00 + 0.05 → z = 1.96

α = 1 % : 0.01 : 0.00 + 0.01 → z = 2.576

α = 76 % : 0.76 : 0.70 + 0.06 → z = 0.305

$$P = P_0 \pm e$$

$$P = P_0 \pm z\, s$$

$$P = P_0 \pm z \sqrt{P_0\, Q_0 / n}$$

Risque d’erreur α inversement proportionnel à la largeur de l’intervalle de confiance

3.4. Exemple :

Sur un échantillon représentatif de 325 sujets, on observe 114 sujets malades.
Quelle est la proportion de sujets malades dans la population d'où est extrait cet échantillon?

$P_0 = 114 / 325 = 0.35$ (35 %)
$Q_0 = 1 – P_0 = 1 – 0.35 = 0.65$ (65 %)
$n = 325$

α = 5 % (0.05) $\longrightarrow$ **z = 1.96**

$$P = P_0 \pm e$$
$$P = P_0 \pm z\, s$$
$$P = P_0 \pm z \sqrt{P_0\, Q_0 / n}$$

$$P = 0.35 \pm 1.96 \times \sqrt{0.35 \times 0.65 / 325}$$
$$P = 0.35 \pm 1.96 \times 0.026$$
$$P = 0.35 \pm 0.05$$

$$P = 0.30 – 0.40$$

P varie entre 30 % et 40 % avec un risque d'erreur de 5 %

P varie entre 30 % et 40 % dans 95 % des cas

P = 0.35 : IC 95 % = 30 % - 40 %

$\alpha = 1\ \%\ (0.01) \longrightarrow z = 2.57$

$$P = P_0 \pm e$$

$$P = P_0 \pm z\, s$$

$$P = P_0 \pm z \sqrt{P_0\, Q_0 / n}$$

$$P = 0.35 \pm 2.57 \times \sqrt{0.35 \times 0.65 / 325}$$

$$P = 0.35 \pm 2.57 \times 0.026$$

$$P = 0.35 \pm 0.068$$

$$P = 0.282 – 0.418$$

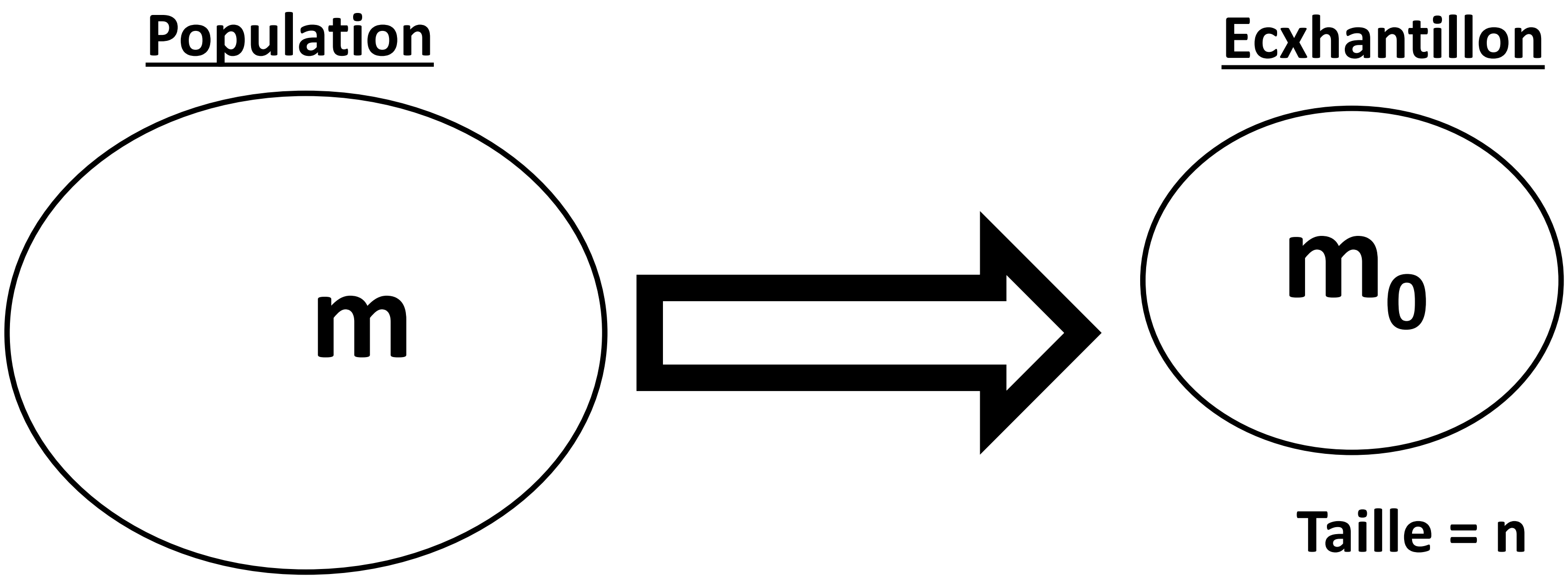
P varie entre 28.2 % et 41.8 % avec un risque d'erreur de 1 %

P varie entre 28.2 % et 41.8 % dans 99 % des cas

P = 0.35 : IC 99 % = 28.2 % - 41.8 %

4. CARACTERE QUANTITATIF :

4.1. Position du problème :



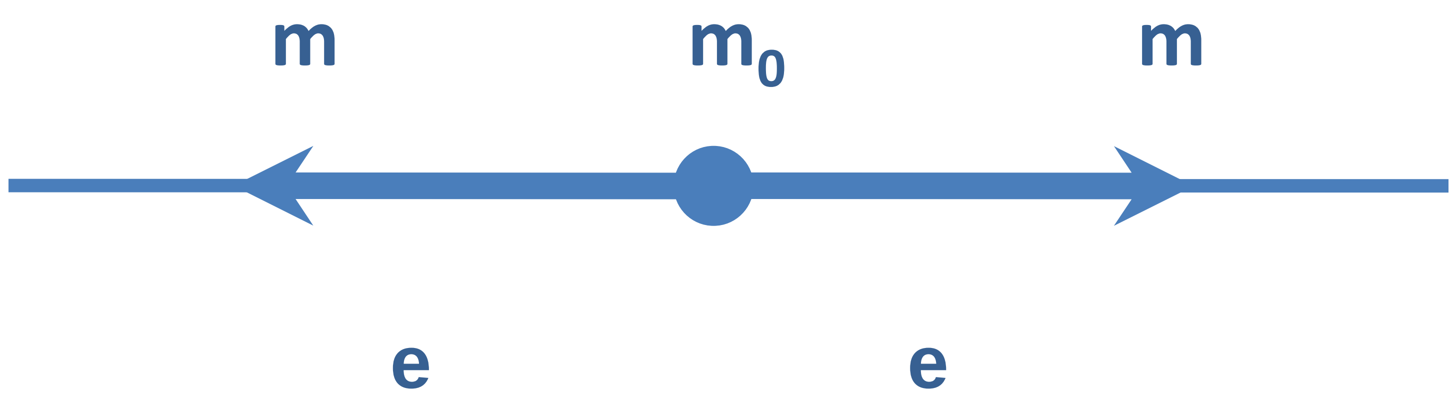
m_o Connue

m Inconnue ?

Quelle est la valeur que peut prendre m ?

4.2. Hypothèse :

- m proche de m_0
- m est dans un intervalle autour de m_0



$$m = m_0 \pm e$$

4.3. Intervalle de confiance :

Intervalle de confiance allant de $m_0 - e$ à $m_0 + e$

Intervalle de confiance = 2 e

Intervalle de confiance symétrique autour de m_0

Intervalle de confiance associé à un risque d'erreur α
Risque d'erreur α associé à l'écart réduit z

e = Ecart-type (s) x Ecart-réduit (z)

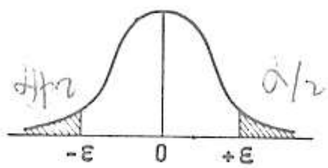
- **Ecart-type =** Ecart-type de la moyenne m_0

$$\textbf{Ecart-type} = \mathbf{s} / \sqrt{n}$$

- **Ecart-Réduit = *z* :** Table de l'écart-réduit *(loi normale)*

Table de l'écart-réduit (loi normale) (*)

La table donne la probabilité α pour que l'écart-réduit égale ou dépasse, en valeur absolue, une valeur donnée ε, c'est-à-dire la probabilité extérieure à l'intervalle $(-\varepsilon, +\varepsilon)$.



α	0,00	0,01	0,02	0,03	0,04	0,05	0,06	0,07	0,08	0,09
0,00	∞	2,576	2,326	2,170	2,054	1,960	1,881	1,812	1,751	1,695
0,10	1,645	1,598	1,555	1,514	1,476	1,440	1,405	1,372	1,341	1,311
0,20	1,282	1,254	1,227	1,200	1,175	1,150	1,126	1,103	1,080	1,058
0,30	1,036	1,015	0,994	0,974	0,954	0,935	0,915	0,896	0,878	0,860
0,40	0,842	0,824	0,806	0,789	0,772	0,755	0,739	0,722	0,706	0,690
0,50	0,674	0,659	0,643	0,628	0,613	0,598	0,583	0,568	0,553	0,539
0,60	0,524	0,510	0,496	0,482	0,468	0,454	0,440	0,426	0,412	0,399
0,70	0,385	0,372	0,358	0,345	0,332	0,319	0,305	0,292	0,279	0,266
0,80	0,253	0,240	0,228	0,215	0,202	0,189	0,176	0,164	0,151	0,138
0,90	0,126	0,113	0,100	0,088	0,075	0,063	0,050	0,038	0,025	0,013

La probabilité α s'obtient par addition des nombres inscrits en marge.

Exemple : Pour $\varepsilon = 1,960$ la probabilité est $\alpha = 0,00 + 0,05 = 0,05$.

$$m = m_0 \pm e$$

$$m = m_0 \pm z\, s/\sqrt{n}$$

- Intervalle de confiance de la variable :

$$x = m_0 \pm e$$

$$x = m_0 \pm z\, s$$

Risque d'erreur α inversement proportionnel à la largeur de l'intervalle de confiance

4.4. Exemple :

Dans une enquête visant à établir les valeurs normales de la cholestérolémie chez une population, on procède au tirage au sort d'un échantillon de 500 sujets qui donne une moyenne de 1.15 g/l avec un écart-type de 0.35 g/l

Que peut conclure à propos des normes du cholestérol sanguin chez la population en question ?

- **Tirage au sort = Echantillon représentatif**

- **Valeurs normales ou normes = 95 % de la population**

m_0 = 1.15 g/l
S =0.35 g/l
n = 500

α = 5 % (0.05) $\longrightarrow$ z = 1.96

$$m = m_0 \pm e$$

$$m = m_0 \pm z\, s/\sqrt{n}$$

$$m = 1.15 \pm 1.96 \times (0.35/\sqrt{500})$$

$$m = 1.15 \pm 1.96 \times 0.015$$

$$m = 1.15 \pm 0.03$$

m = 1.12 g/l – 1.18 g/l

La cholestérolémie moyenne de la population varie entre 1.12 g/l et 1.18 g/l avec un risque d'erreur de 5 %

m varie entre 1.12 g/l et 1.18 g/l dans 95 % des cas

m = 1.15 g/l : IC 95 % = 1.12 g/l – 1.18 g/l

- **Intervalle de confiance de la variable :**

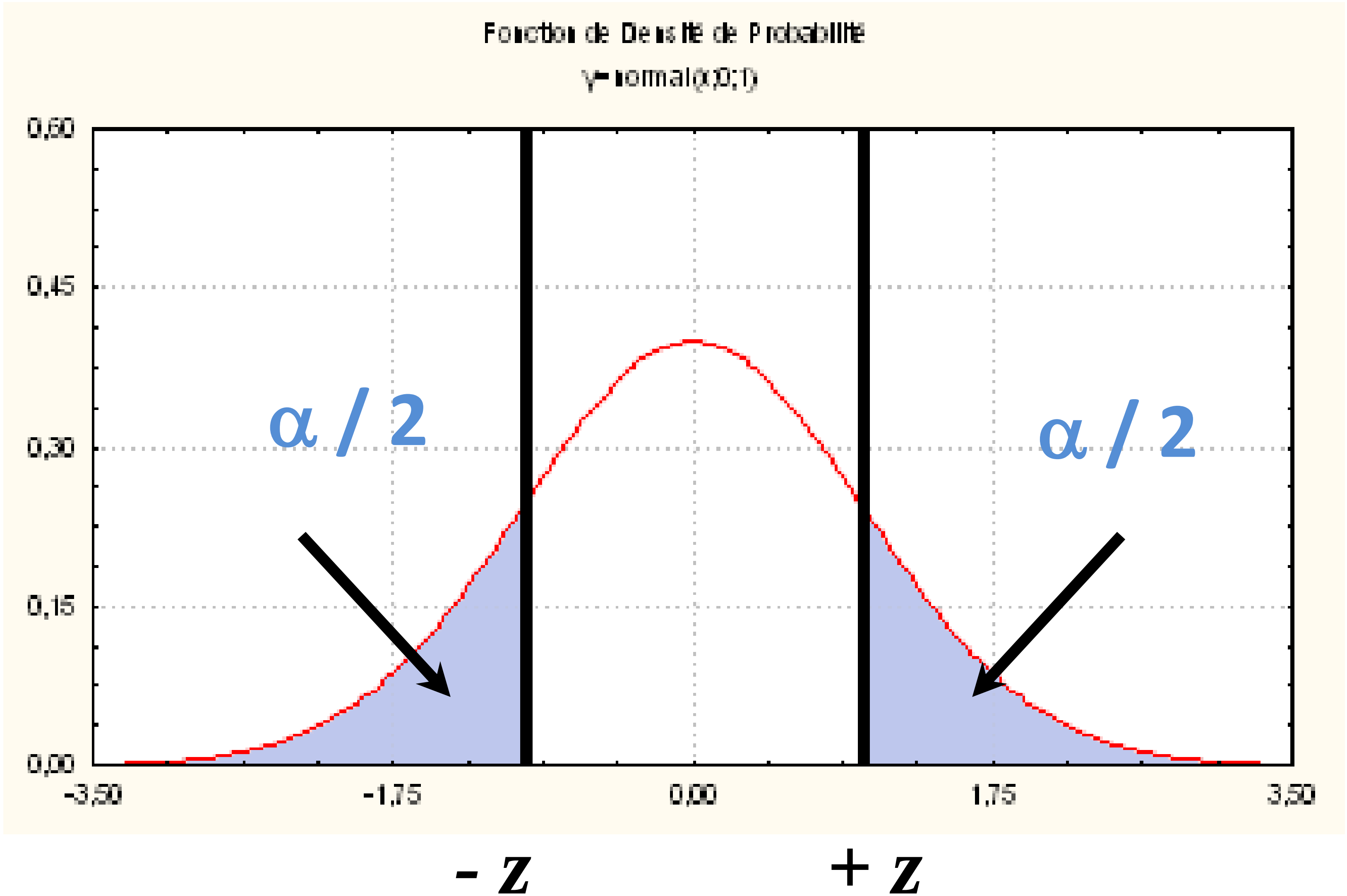
$x = m_0 \pm e$

$x = m_0 \pm z\, s$

$x = 1.15 \pm 1.96 \times 0.35$

$x = 1.15 \pm 0.7$

95 % des sujets de la population ont une cholestérolémie comprise entre 0.45 g/l et 1.85 g/l



α = Probabilité d'avoir des valeurs à l'extérieur d'un intervalle (- z , + z)