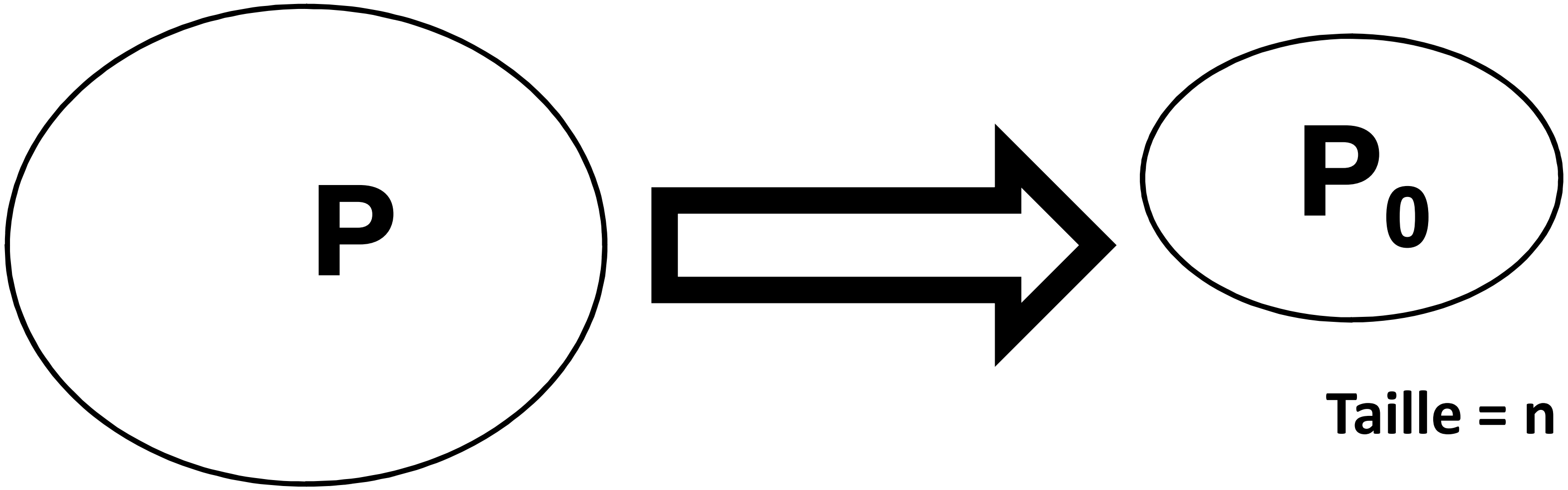
LES FLUCTUATIONS D'ECHANTILLONNAGE

1. INTRODUCTION

Etudier la variabilité des résultats partiels observés sur un échantillon par rapport au résultats globaux de la population totale

2. CARACTERE QUALITATIF

2.1. POSITION DU PROBLEME



P Connue

P_o Inconnue ?

Quelle est la valeur que peut prendre P_0 ?

2.2. RAPPELS

B (n,P)

x : Nombre de sujets présentant la modalité

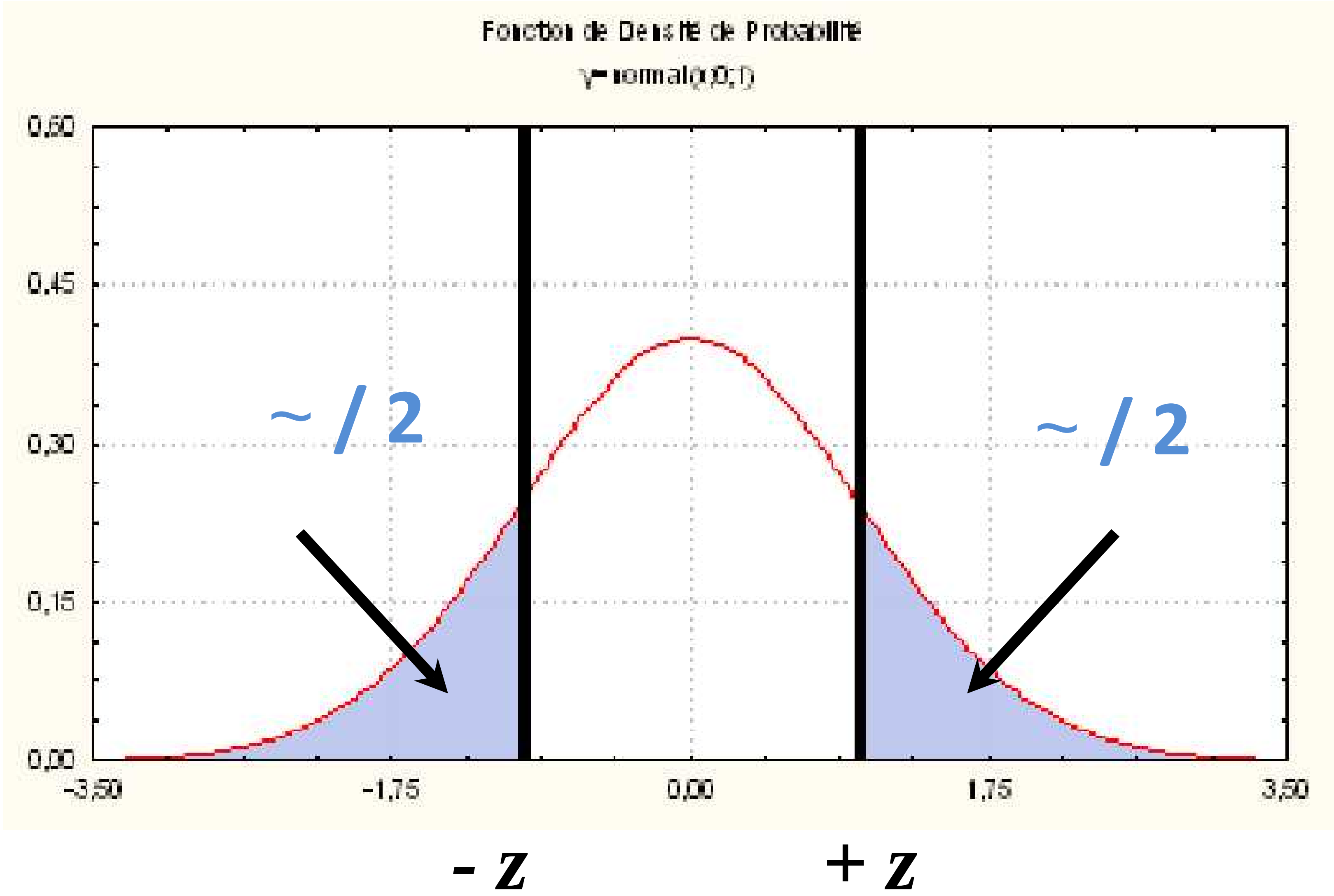
p = x / n

Plusieurs tirages : **p suit une loi binomiale de :**

- **Moyenne p (np)**
- **Variance pq/n (npq)**
- **Ecart-type : $\sqrt{pq/n}$ ($\sqrt{npq}$)**

np & $nq \geq 5$

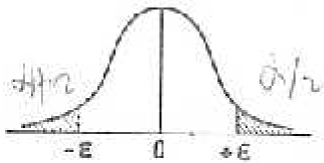
Approximation par la loi normale



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

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 $\alpha = 0,00 \quad 0,05 = 0,05$.

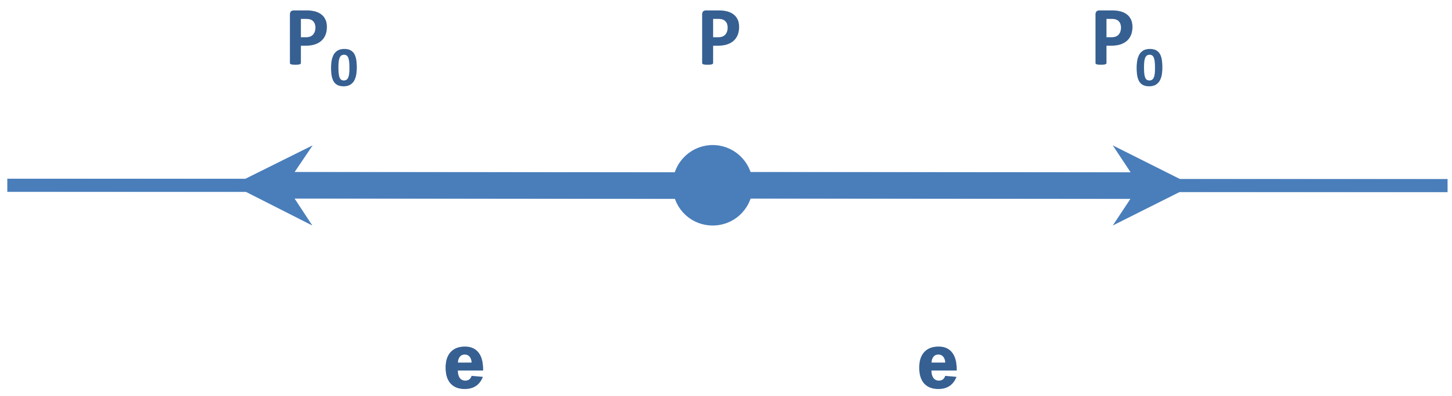
Table pour les petites valeurs de la probabilité

α	0,001	0,000 1	0,000 01	0,000 001	0,000 000 1	0,000 000 01	0,000 000 001
ε	3,29053	3,89059	4,41717	4,89164	5,32672	5,73073	6,10941

(*) D'après Fisher et Yates, Statistical tables for biological, agricultural, and medical research (Oliver and Boyd, Edinburgh) avec l'aimable autorisation des auteurs et des éditeurs.

2.3. HYPOTHESE

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



P_0 = P ï e

2.4. INTERVALLE DE PARI

Intervalle de pari allant de P – e à P + e
Intervalle de pari= 2 e
Intervalle de pari associé à un risque d'erreur α
Risque d'erreur α associé à l'écart réduit z

$$e = z\,s$$

$$e = z\sqrt{pq/n}$$

$$P_0 = P \pm e$$

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

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

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

2.5. APPLICATION

Fréquence de sujets du groupe sanguin O dans une population = 40 %
Echantillon de 250 sujets
Quelle est cette fréquence au sein de l'échantillon ?

2.5. 1. Les données :

P = 40 % = 0,4
Q = 1 – 0,4 = 0,6
n = 250
P_0 = ?

2.5. 2. Position du problème :

Il s'agit d'un problème de fluctuation d'echantillonnage d'un pourcentage

$P_0 = P$ ï e
$P_0 = P$ ï $z\ s$
$P_0 = P$ ï z Ě$pq\ /n$

2.5. 3. Vérification des conditions d'application :

np et nq ≥ 5

$np = 250 \times 0{,}4 = 100$

$nq = 250 \times 0{,}6 = 150$

Les conditions d'application sont réunies

2.5. 4. Application :

- ***Risque d'erreur*** $\alpha = 5\ \%\ (0{,}05) \Longrightarrow z = 1{,}96$

$$P_0 = P \pm e$$

$$e = z\, s$$

$$s = \sqrt{pq/n} = \sqrt{0{,}4 \times 0{,}6 / 250} = 0{,}031$$

$$e = 1{,}96 \times 0{,}031 = 0{,}06$$

$$P_0 = 0{,}4 \pm 0{,}06$$

2.5. 5. Résultat :

P_0 = 0,4 ï 0,06

P_0 = 0,34 - 0,46

2.5. 6. Conclusion :

La fréquence de sujets du groupe sanguin O au sein de l’échantillon varie entre 34 % et 46 % (p = 0,05)

2.6. UTILISATION DES FREQUENCES ABSOLUES

e = z s

s = Ě *npq* = Ě *250 x 0,4 x 0,6 = 7,7*

e = 1,96 x 7,7 = 15

X_0 = 100 ï 15

X_0 = 85 - 115

2.7. VARIATION EN FONCTION DU RISQUE D'ERREUR

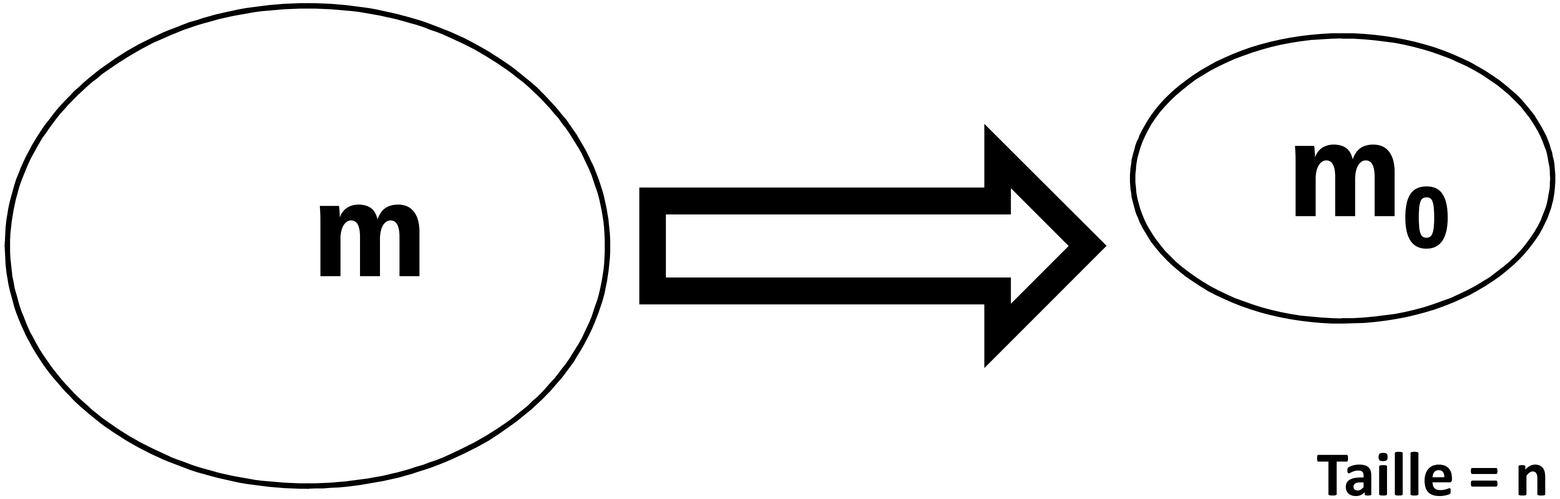
P = 40 %

n = 250

~	**z**	**s = Ěpq/n**	**e = z s**	**Intervalle de pari**
0,001	3,29		0,1 (10 %)	30 % - 50 %
0,01	2,57		0,08 (8 %)	32 % - 48 %
0,05	1,96	0,031	0,06 (6 %)	34 % - 46 %
0,2	1,28		0,04 (4 %)	36 % - 44 %
0,5	0,67		0,02 (2 %)	38 % - 42 %

3. CARACTERE QUANTITATIF

3.1. POSITION DU PROBLEME



Taille = n

m Connue

m_o Inconnue ?

Quelle est la valeur que peut prendre m_0 ?

3.2. RAPPELS

m : Moyenne de la variable x : $\sum x_i /n$

s^2 : Variance de la variable x : $\sum (x_i - m)^2/n$

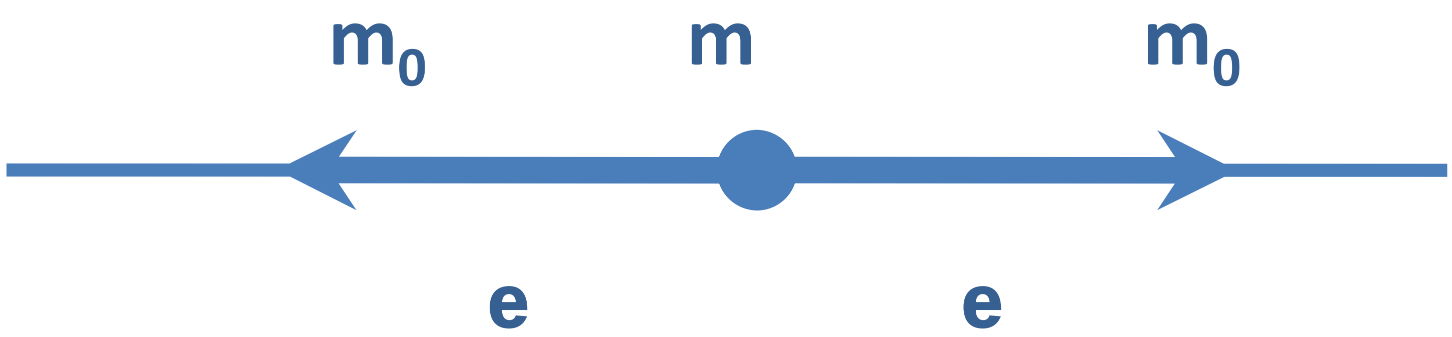
s : Ecart-type de la variable x : $\sqrt{\sum (x_i - m)^2/n}$

Variance de la moyenne m = $\boldsymbol{s^2 / n}$

Ecart-type de la moyenne m = $\mathbf{s} / \sqrt{\boldsymbol{n}}$

3.3. HYPOTHESE

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



$m_0 = m$ ï e

Conditions d'application :

- ❑ Loi normale
- ❑ Effectif ≥ 30

3.4. INTERVALLE DE PARI

Intervalle de pari allant de m – e à m + e
Intervalle de par = 2 e
Intervalle de pari associé à un risque d'erreur α
Risque d'erreur α associé à l'écart réduit z

$$e = z\, s/\sqrt{n}$$

$$m_0 = m \pm e$$

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

Pour les fluctuations de la variable x :

$$e = z\, s$$

$$x_i = m \pm e$$

$$x_i = m \pm z\, s$$

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

3.5. APPLICATION

La glycémie moyenne d'une population est estimée à 0,95 g/l avec une variance de 0,09 g^2/l^2

Echantillon de 350 sujets

Quelle est la glycémie moyenne au sein de l'échantillon ?

3.5. 1. Les données :

$m = 0,95$ g/l

$s^2 = 0,09\ g^2/l^2$

$s = \sqrt{0,09} = 0,3$ g/L

$n = 350$

$m_0 = ?$

3.5. 2. Position du problème :

Il s'agit d'un problème de fluctuation d'echantillonnage d'une moyenne

$$m_0 = m \pm \varepsilon$$

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

3.5. 3. Vérification des conditions d’application :

- La glycémie est une variable suivant une loi normale
- n = 350 > 30

Les conditions d’application sont réunies

3.5. 4. Application :

➢ ***Risque d’erreur*** *~ = 5 % (0,05)* $\Longrightarrow$ *z = 1,96*

m_0 = m ï e

e = z s/Ĕn

s = 0,3/Ĕ350 = 0,016

e = 1,96 x 0,016 = 0,031

m_0 = 0,95 ï 0,03

3.5. 5. Résultat :

m_0 = 0,95 ï 0,03

m_0 = 0,92 - 0,98

3.5. 6. Conclusion :

La glycémie moyenne au sein de l'échantillon
varie entre 0,92 et 0,98 g/l (~ = 0,05)

3.5. 7. Fluctuation de la variable :

- ***Risque d'erreur ~ = 5 % (0,05)*** $\Longrightarrow$ ***z = 1,96***

x_i = m ï e

e = z s

e = 1,96 x 0,3 = 0,6

x_i = 0,95 ï 0,6

Pour 95 % des sujets de l'échantillon la glycémie est comprise entre 0,35 et 1,55 g/l