

Chapitre 7: Echantillonnage et Estimation

I) Rappels sur les lois :

loi	Notation	Formule	Espérance	Variance
Bernoulli	$\mathcal{B}(p)$	$P(X=1)=p; P(X=0)=1-p$	$E(X)=p$	$V(X)=p(1-p)$
Binomiale	$\mathcal{B}(n; p)$	$P(X=k) = \binom{n}{k} p^k (1-p)^{n-k}$ $k \in \llbracket 0..n \rrbracket$	$E(X)=np$	$V(X)=np(1-p)$
Poisson	$\mathcal{P}(\lambda)$	$P(X=k) = \frac{\lambda^k}{k!} e^{-\lambda}$ $k \in \mathbb{N}$	$E(X)=\lambda$	$V(X)=\lambda$
Normale	$\mathcal{N}(m; \sigma)$ moyenne $\uparrow$ écart-type $\uparrow$	$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-m}{\sigma}\right)^2}$	$E(X)=m$	$V(X)=\sigma^2$
Normale centrée réduite	$\mathcal{N}(0; 1)$	$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$	$E(X)=0$	$V(X)=1$

II) Approximation de la loi binomiale

1) par la loi de Poisson:

Propriété: Pour n assez grand ($n > 30$) et p proche de 0 ($p \leq 0,1$) tels que $np(1-p)$ ne soit pas trop grand ($np(1-p) \leq 10$), on peut alors approximer une $\mathcal{B}(n; p)$ par une loi de Poisson $\mathcal{P}(\lambda)$ avec $\lambda = np$.

Exemple: Dans une chaîne de fabrication, 5% des pièces sont défectueuses. On prélève une pièce, on regarde si elle est défectueuse, on la remet dans la production. On répète 120 fois cette expérience. On désigne par X la variable aléatoire qui compte le nombre de pièces défectueuses.

- 1) Justifier la loi de X et ses paramètres.
- 2) Calculer $P(X=5)$

3) Montrer que l'on peut approcher cette loi par une loi de Poisson dont on précisera le paramètre (Y suit cette loi)

4) Calculer $P(Y=5)$

5) Comparer les 2 méthodes de calcul

1) Epreuve de Bernoulli: On tire une pièce dans la production on regarde si elle est défectueuse (Succès) : $P(\text{Succès}) = p = 5\% = 0,05$.
On répète 120 fois cette épreuve ($n=120$) de manière indépendante (tirage avec remise)

X compte le nombre de succès, donc $X \sim \mathcal{B}(120; 0,05)$

2) $P(X=5) \approx 0,163$ à 10^{-3} près.

3) $n > 30$; $p = 0,05 < 0,1$; $np(1-p) = 5,7 < 10$

donc $\mathcal{B}(120; 0,05)$ peut être approchée par une $\mathcal{P}(\lambda)$ avec $\lambda = np$
 $\lambda = 120 \times 0,05 = 6$

4) $Y \sim \mathcal{P}(6)$; $P(Y=5) \approx 0,161$ à 10^{-3} près.

5) $0,163 - 0,161 \approx 0,002$ les approximations sont très proches -

2) par la loi normale

Propriété: Pour n "assez grand" ($n \geq 50$) et pour p "ni proche de 0 ni proche de 1" ($p \approx 0,5$) tels que $np(1-p)$ "pas trop grand" ($10 \leq np(1-p) \leq 100$)

on peut approcher une $\mathcal{B}(n; p)$ par la loi $\mathcal{N}(n; \sigma)$

où $m = np$ et $\sigma = \sqrt{np(1-p)}$

Exemple: On lance 300 fois une pièce de monnaie truquée - (une partie)
La probabilité d'obtenir face est de $\frac{2}{3}$.

On désigne par X la variable aléatoire qui à chaque partie associe le nombre de faces obtenues -

1) Quelle est la loi de X

2) Déterminer $P(X > 210)$

3) Montrer que l'on peut approcher la loi de X par une loi normale

4) Y suit cette loi, Calculer $P(Y > 210)$

Exercice:

1) Bernoulli + répétition + indépendance $\Rightarrow \mathcal{B}(300; \frac{2}{3})$

$$2) P(X > 210) = 1 - P(X \leq 210) \approx 0,098$$

$$3) n \geq 50 ; p = \frac{2}{3} \approx 0,5 ; np(1-p) \approx 67 \geq 10$$

$$\mathcal{B}(300; \frac{2}{3}) \rightarrow \mathcal{N}(m; \sigma)$$

$$\text{avec } m = 300 \times \frac{2}{3} = 200$$

$$\sigma = \sqrt{200 \times \frac{1}{3}} \approx 8,167$$

$$4) P(Y > 210) \approx 0,110$$

5) en comparant $0,110 - 0,098 \approx 12 \times 10^{-3}$. c'est très proche

III) Théorème de la limite centrale et échantillonnage

a) Propriété: Soient $X_1, X_2, \dots, X_n$, n variables aléatoires indépendantes ayant même espérance et même écart-type

$$\text{Soit } \overline{X}_n = \frac{X_1 + X_2 + \dots + X_n}{n}$$

Pour n suffisamment grand, $\overline{X}_n \sim \mathcal{N}\left(m; \frac{\sigma}{\sqrt{n}}\right)$

b) Lois d'échantillonnage:

i) Loi d'échantillonnage par rapport à la moyenne:

Etant donnée une population de taille N et X une variable aléatoire telle que $E(X) = m$ et $\sigma(X) = \sigma$

lorsqu'on prélève un échantillon de taille n , on obtient

n Variables aléatoires de même loi.

donc $\overline{X}_n = \frac{X_1 + \dots + X_n}{n}$ suit une loi qui peut être approchée par $\mathcal{N}\left(m; \frac{\sigma}{\sqrt{n}}\right)$

pour $n \geq 30$

Exemple: Une machine fabrique en grande série des pièces pour l'automobile.

A chaque pièce tirée au hasard, on associe sa longueur exprimée en mm.

On définit ainsi une variable aléatoire X .

On suppose que $X \sim \mathcal{N}(28,20; 0,027)$ $E(X) = 28,20$ $\sigma(X) = 0,027$

Soit M_m la variable qui à tout échantillon de 100 pièces associe la moyenne de leur longueur.

Quelle est la loi de M_m ?

$$m = 100 \geq 30, \quad M_m \sim \mathcal{N}\left(28,20; \frac{0,027}{\sqrt{100}}\right)$$

$$M_m \sim \mathcal{N}(28,20; 0,0027)$$

ii) loi d'échantillonnage des fréquences.

On étudie une population de taille N , un caractère X suivant une loi de Bernoulli $B(p)$.

Dans un échantillon de taille n , on obtient n variables aléatoires indépendantes on pose $f_n = \frac{X_1 + X_2 + \dots + X_n}{n}$

la variable aléatoire qui à tout échantillon de taille n associe la fréquence du succès.

pour n assez grand ($n \geq 50$)

$$f_n \sim \mathcal{N}\left(p; \sqrt{\frac{p(1-p)}{n}}\right)$$

Exercices: 1 - Echantillonnage des moyennes.

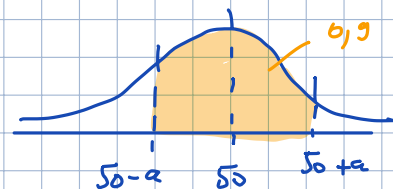
Ex 1: 1) $X_i \sim \mathcal{N}(50; 0,5)$, $i \in \llbracket 1 \dots 36 \rrbracket$ $n = 36$ $m = 50$ $\sigma = 0,5$

$$Y = \frac{X_1 + X_2 + \dots + X_{36}}{36} \sim \mathcal{N}\left(m; \frac{\sigma}{\sqrt{n}}\right)$$

$$Y \sim \mathcal{N}\left(50; \frac{0,5}{\sqrt{36}}\right) = \mathcal{N}\left(50; \frac{0,5}{6}\right)$$

$$2) \quad \mathbb{P}(\mu - a \leq Y \leq \mu + a) = 0,9$$

$$\mathbb{P}(50 - a \leq Y \leq 50 + a) = 0,9$$

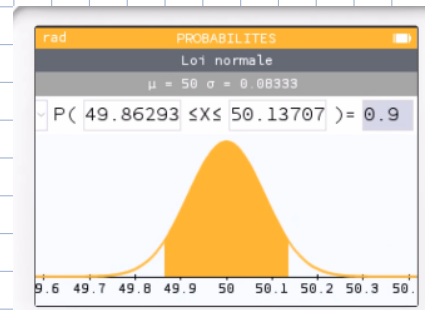


$$a \approx 50,1371 - 50$$

$$a \approx 0,1371 \text{ à } 10^{-4} \text{ près.}$$

idem avec $\mu = 40$ et $\sigma = 0,6$

$$a \approx 0,1645$$



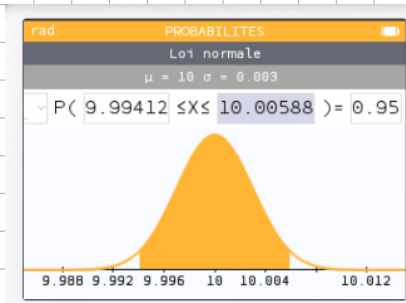
$$\mu = 50$$

$$\sigma = \frac{0,5}{6}$$

$$\sigma \approx 0,0833$$

Ex2: 1) $\bar{X}_{100} \sim \mathcal{N}\left(10; \frac{0,03}{\sqrt{100}}\right) = \mathcal{N}(10; 0,003)$

2) $P(10 - h \leq \bar{X}_{100} \leq 10 + h) = 0,95$



$$10 + h \approx 10,00588$$

$$h \approx 10,00588 - 10$$

$$h \approx 0,0059 \text{ à } 10^{-4}$$

$$X \sim \mathcal{N}(20; 0,1)$$

$$h \approx 0,0196$$

Ex3:

1) $\bar{X} \sim \mathcal{N}\left(4,2; \frac{0,2}{\sqrt{n}}\right)$

$$T = \frac{\bar{X} - m}{\sigma} \sim \mathcal{N}(0; 1)$$

2) Changement de variable : $\bar{T} = \frac{\bar{X} - m}{\sigma} = \frac{\bar{X} - 4,2}{\frac{0,2}{\sqrt{n}}}$

$$\bar{T} = \frac{\sqrt{n}}{0,2} \times (\bar{X} - 4,2) \sim \mathcal{N}(0; 1)$$

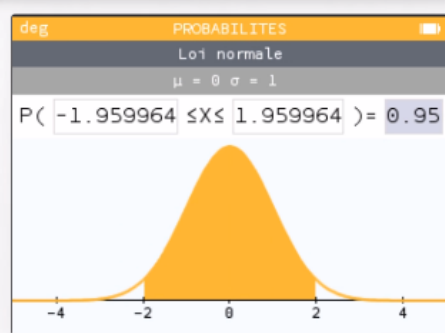
1ère méthode:

2) $P(4,17 \leq \bar{X} \leq 4,23) = 0,95$

$$P(4,17 - 4,2 \leq \bar{X} - 4,2 \leq 4,23 - 4,2) = 0,95$$

$$P(-0,03 \leq \bar{X} - 4,2 \leq 0,03) = 0,95$$

$$P\left(-0,03 \times \frac{\sqrt{n}}{0,2} \leq (\bar{X} - 4,2) \times \frac{\sqrt{n}}{0,2} \leq 0,03 \times \frac{\sqrt{n}}{0,2}\right) = 0,95$$



$$P\left(-0,03 \times \frac{\sqrt{n}}{0,2} \leq \bar{T} \leq 0,03 \times \frac{\sqrt{n}}{0,2}\right) = 0,95$$

$$0,03 \times \frac{\sqrt{n}}{0,2} \approx 1,96$$

$$\sqrt{n} \approx \frac{1,96 \times 0,2}{0,03}$$

$$n \approx \left(\frac{1,96 \times 0,2}{0,03}\right)^2$$

$$n \approx 171.$$

2^{ème} Méthode:

$$P(4,17 \leq \bar{X} \leq 4,23) = 0,95$$

Cours:

Si $X \sim \mathcal{N}(m; \sigma)$ on a:

$$P(m - \sigma \leq X \leq m + \sigma) \approx 0,68$$

$$P(m - 2\sigma \leq X \leq m + 2\sigma) \approx 0,95$$

$$P(m - 3\sigma \leq X \leq m + 3\sigma) \approx 0,997$$

$$m + 2\sigma = 4,23$$

$$4,2 + 2\sigma = 4,23$$

$$\sigma = \frac{0,03}{2} = 0,015$$

$$\sigma = \frac{0,2}{\sqrt{n}}$$

$$\text{donc } \frac{0,2}{\sqrt{n}} = \frac{0,015}{1}$$

$$\sqrt{n} = \frac{0,2}{0,015}$$

$$n = \left(\frac{0,2}{0,015}\right)^2 \approx 178.$$

3)

Ex4: 1) $F \sim \mathcal{N}\left(p; \sqrt{\frac{p(1-p)}{n}}\right) = \mathcal{N}\left(0,04; \sqrt{\frac{0,04 \times 0,96}{64}}\right)$
 $F \sim \mathcal{N}(0,04; 0,0245)$

2) $P(0,03 \leq F \leq 0,05) \approx 0,317. \approx 10^{-3} \text{ pri.}$

3) $P(0,04 - \alpha \leq F \leq 0,04 + \alpha) = 0,95$
 $0,04 + \alpha \approx 0,088$
 $\alpha \approx 0,048$

4) Changement de variable: $T = \frac{F - m}{\sigma} = \frac{F - 0,04}{\sqrt{\frac{0,04 \times 0,96}{n}}}$
 $= \frac{F - 0,04}{\frac{\sqrt{0,04 \times 0,96}}{\sqrt{n}}} = \frac{F - 0,04}{\sqrt{0,04 \times 0,96}} \times \sqrt{n}$

$P(0,03 \leq F \leq 0,05) = 0,99$
 $\frac{0,01}{\sqrt{0,04 \times 0,96}} \times \sqrt{n} \approx 2,576$

$n \approx \left(\frac{2,576 \times \sqrt{0,04 \times 0,96}}{0,01} \right)^2$

$n \geq 2548.$

Ex5: $p = 0,28$ $n = 400$ $F \sim \mathcal{N}\left(0,28; \sqrt{\frac{0,28 \times 0,72}{400}}\right)$
 $F \sim \mathcal{N}(0,28; 0,02245)$

1) $P(0,26 \leq F \leq 0,30) \approx 0,627 \approx 10^{-3}$

2) $P(F \leq 0,22) \approx 0,004 \approx 10^{-3}$

IV Estimation: idée à l'aide d'un échantillon peut-on estimer un paramètre (moyenne, écart-type, fréquence...) avec quelle précision.

a) Estimation ponctuelle:

a) Propriété la valeur moyenne " m_e " d'un paramètre observé sur un échantillon de la population, dont la taille est fixée, fournit une estimation $\bar{x}$ de la moyenne réelle de ce paramètre dans la population.

Exemple: Une usine produit des vis cruciformes. On souhaite estimer la moyenne des longueurs des vis dans la production journalière qui s'élève à 10 000 pièces.

On choisit un échantillon de 150 vis et on calcule la moyenne $m_e = 4,57 \text{ mm}$

On peut dire que $\bar{x} = m_e = 4,57 \text{ mm}$ est une estimation ponctuelle de la moyenne des vis de la population

b) l'écart type σ_e d'un paramètre observé sur un échantillon de taille fixée permet de calculer σ une estimation de l'écart-type dans la population tel que :

$$\sigma = \sigma_e \times \sqrt{\frac{n}{n-1}}$$

Exemple: la mesure de la longueur des vis produites ($n=150$) permet de calculer $\sigma_e \approx 3 \text{ mm}$

une estimation ponctuelle de l'écart-type de la population est

$$\sigma = \sigma_e \times \sqrt{\frac{n}{n-1}} \approx 3 \times \sqrt{\frac{150}{149}} \approx 3,01 \text{ mm}$$

Remarque: $\lim_{n \rightarrow +\infty} \sqrt{\frac{n}{n-1}} = 1$, la correction de σ est de moins en moins grande plus la taille de l'échantillon augmente

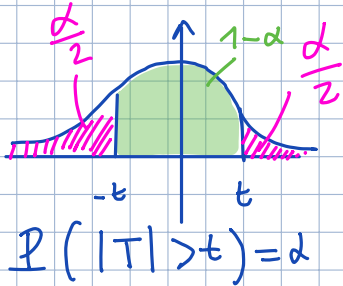
1) L'estimation par la fréquence est la même que pour la moyenne -
f_e est une estimation de f.

2) Estimation par intervalle de confiance:

On souhaite, à partir des observations faites sur un échantillon, déterminer un intervalle de confiance dans lequel sera notre paramètre avec un risque (α) de se tromper.

a) Moyenne: On suppose que les conditions sont réunies pour faire une approximation que la loi d'échantillonnage de la moyenne $\bar{X}$ est une loi normale $\mathcal{N}(m; \frac{\sigma}{\sqrt{n}})$

$$\text{On pose } T = \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \sim \mathcal{N}(0; 1)$$



$$P(-t \leq T \leq t) = 1 - \alpha$$

$$P\left(-t \leq \frac{\bar{X} - m}{\frac{\sigma}{\sqrt{n}}} \leq t\right) = 1 - \alpha$$

$$P\left(-t \frac{\sigma}{\sqrt{n}} \leq \bar{X} - m \leq t \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

$$P\left(\bar{X} - t \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X} + t \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$$

Autrement dit $m \in \left[\bar{X} - t \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X} + t \frac{\sigma}{\sqrt{n}}\right]$ pour $100 \times (1 - \alpha) \%$ des échantillons.

Propriété: l'intervalle $\left[\bar{X} - t \frac{\sigma}{\sqrt{n}} \leq m \leq \bar{X} + t \frac{\sigma}{\sqrt{n}}\right]$ est l'intervalle de

confiance de la moyenne m de la population avec le

coefficient de confiance t tel que $P(-t \leq T \leq t) = 1 - \alpha$

α = risque ; $\bar{X}$ moyenne de l'échantillon ($\bar{x}$) ; n taille de l'échantillon ;

σ écart-type de la population ; $T \sim \mathcal{N}(0; 1)$

Remarque: les valeurs fréquentes du coefficient de confiance sont:

$$\text{pour } 1-\alpha = 0,99 \Rightarrow t \approx 2,576$$

$$1-\alpha = 0,95 \Rightarrow t \approx 1,96$$

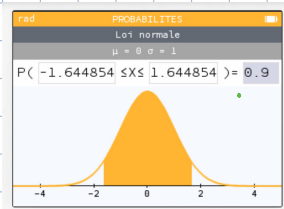
Exemple: On suppose que la durée de vie, exprimée en heures, d'une ampoule électrique suit la loi normale N de moyenne inconnue et d'écart-type $\sigma \approx 20$.

Une étude sur un échantillon de 16 ampoules donne une moyenne de durée de vie de 3000 h.

Déterminer un intervalle de confiance de la moyenne au seuil de risque de 10%

$$\sigma = 20 ; n = 16 ; \bar{X} = m_e = 3000 \text{ h} ; \alpha = 0,1$$

$$t \text{ est tel que } P(-t \leq T \leq t) = 1 - \alpha = 1 - 0,1 = 0,9$$



$$(T \pm) \text{invnormal}(0,9; 0; 1) \approx 1,645$$

α σ
alpha sigma

$$\text{on trouve } t \approx 1,645$$

$$\left[3000 - 1,645 \times \frac{20}{\sqrt{16}} ; 3000 + 1,645 \times \frac{20}{\sqrt{16}} \right] = [2992 ; 3008]$$

b) Fréquence: "idem" $F \sim N\left(p ; \sqrt{\frac{p(1-p)}{n}}\right)$

Propriété: l'intervalle $\left[F - t \sqrt{\frac{F(1-F)}{n-1}} ; F + t \sqrt{\frac{F(1-F)}{n-1}} \right]$

avec F la fréquence de l'échantillon

n la taille de l'échantillon

$$t \text{ tel que } P(-t \leq T \leq t) = 1 - \alpha$$

$$T \sim N(0; 1)$$

Exemple: Dans une commune on fait un sondage sur le contournement du centre-ville.

42% des 500 personnes interrogées sont contre.

Donner un intervalle de confiance de la proportion de gens "contre" ce projet au seuil de risque de 1%.

$$\bar{F} = 0,42 ; n = 500 ; \alpha = 0,01$$

$$t \text{ tel que } P(-t \leq T \leq t) = 1 - \alpha = 1 - 0,01 = 0,99$$

$$\left[0,42 - 2,576 \times \sqrt{\frac{0,42(1-0,42)}{499}} ; 0,42 + 2,576 \times \sqrt{\frac{0,42(1-0,42)}{499}} \right]$$

$$= [0,36 ; 0,48]$$

Dans la population il ya entre 36% et 48% de personnes "contre" le projet au seuil de risque de 1%.

Ex6: 1) une estimation ponctuelle de μ est $\mu_e \approx 9,99$

$$2) \left[\mu_e - t \frac{\sigma}{\sqrt{n}} ; \mu_e + t \frac{\sigma}{\sqrt{n}} \right] = \left[9,99 - 1,96 \times \frac{9,19}{\sqrt{50}} ; 9,99 + 1,96 \times \frac{9,19}{\sqrt{50}} \right]$$

$$t \text{ tel que } P(-t \leq T \leq t) = 0,95 \Rightarrow t \approx 1,96$$

$$\text{donc } I = [9,937 ; 10,043] \text{ à } 95\% \text{ de confiance.}$$

3) Non car il y a 5% de risque.

Ex7: 1) $m_e \approx 150,1$ $\sigma_e \approx 3,051$

$$2) \mu \approx m_e = 150,1 \quad \sigma \approx \sigma_e \sqrt{\frac{100}{99}} = 3,066$$

$$3) \bar{P} \sim N \left(150,1 ; \frac{3,066}{\sqrt{50}} \right) = N(150,1 ; 0,434)$$

$$4) \left[150,1 - 1,96 \times \frac{3,066}{\sqrt{50}} ; 150,1 + 1,96 \times \frac{3,066}{\sqrt{50}} \right]$$

$$= [149,615 ; 150,585] \text{ avec } \alpha = 5\%$$

$$\text{si } \alpha = 10\% \quad [149,693 ; 150,507]$$

$$\text{si } \alpha = 1\% \quad [149,462 ; 150,738]$$

$$\alpha = 10\% \quad 1 - \alpha = 90\% \quad t \approx 1,645$$

$$\alpha = 5\% \quad 1 - \alpha = 95\% \quad t \approx 1,96$$

$$\alpha = 1\% \quad 1 - \alpha = 99\% \quad t \approx 2,576$$

Ex 8:

1) a) $p \approx 0,8$

b) $F \sim \mathcal{N}\left(p; \sqrt{\frac{p(1-p)}{n}}\right) = \mathcal{N}\left(0,8; \sqrt{\frac{0,8 \times 0,2}{400}}\right)$
 $= \mathcal{N}(0,8; 0,02)$

c) $P(F < 0,78) \approx 0,159$

d) $P(0,78 < F < 0,82) \approx 0,683$

e) $\left[p - t \sqrt{\frac{p(1-p)}{n-1}}; p + t \sqrt{\frac{p(1-p)}{n-1}} \right]$
 $= \left[0,8 - 1,96 \sqrt{\frac{0,8 \times 0,2}{399}}; 0,8 + 1,96 \sqrt{\frac{0,8 \times 0,2}{399}} \right]$
 $= [0,761; 0,839]$ à un risque de 5%

2) $F_n \sim \mathcal{N}\left(p; \sqrt{\frac{p(1-p)}{n}}\right) = \mathcal{N}\left(0,8; \sqrt{\frac{0,16}{n}}\right)$

$$P(0,8 - 0,02 \leq F_n \leq 0,8 + 0,02) = 0,99$$

$$T_n = \frac{F_n - 0,8}{\sqrt{\frac{0,16}{n}}} \sim \mathcal{N}(0; 1)$$

$$P\left(\frac{-0,02}{\sqrt{\frac{0,16}{n}}} \leq T_n \leq \frac{0,02}{\sqrt{\frac{0,16}{n}}}\right) = 0,99$$

$$\frac{0,02}{\sqrt{\frac{0,16}{n}}} \approx 2,576$$

$$0,02 \times \sqrt{\frac{n}{0,16}} \approx 2,576$$

$$0,02 \times \frac{\sqrt{n}}{0,4} \approx 2,576$$

$$\sqrt{n} \approx \frac{2,576 \times 0,4}{0,02}$$

$$n \approx \left(\frac{2,576 \times 0,4}{0,02} \right)^2$$

$$n \geq 2655$$

Ex BTS 2012: 1) $p \approx \frac{37}{50} = 0,74$

$$\begin{aligned} 2) \text{ a) } & \left[p - t \sqrt{\frac{p(1-p)}{n-1}} ; p + t \sqrt{\frac{p(1-p)}{n-1}} \right] \\ & = \left[0,74 - 1,96 \sqrt{\frac{0,74 \times 0,26}{49}} ; 0,74 + 1,96 \sqrt{\frac{0,74 \times 0,26}{49}} \right] \\ & \approx \left[0,617 ; 0,863 \right] \end{aligned}$$

b) Non on a un risque de 5%.

Ex BTS 2011: