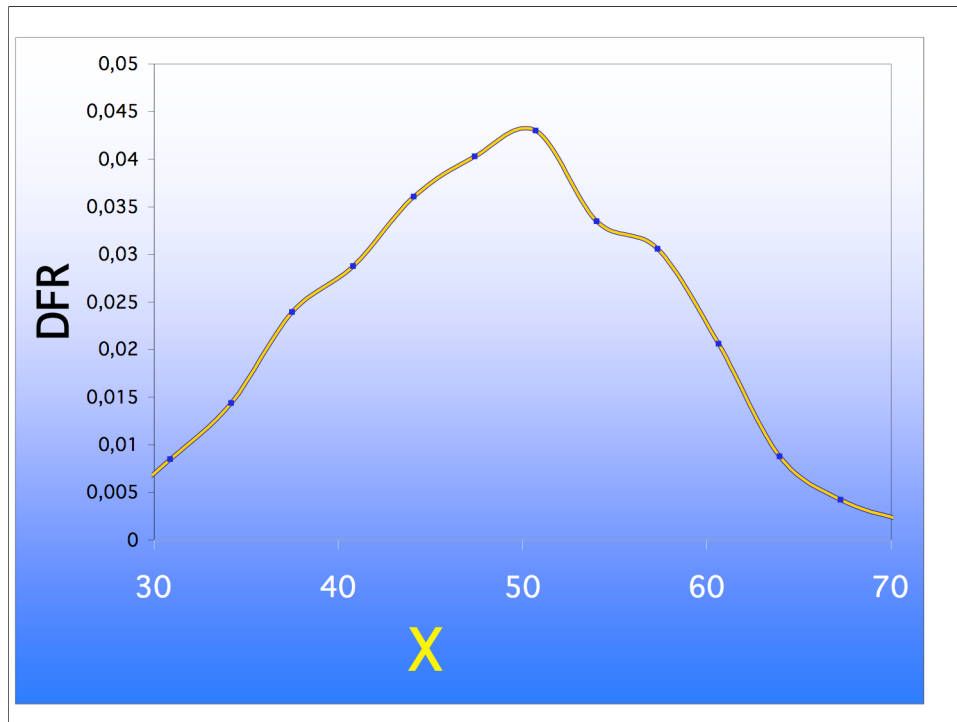
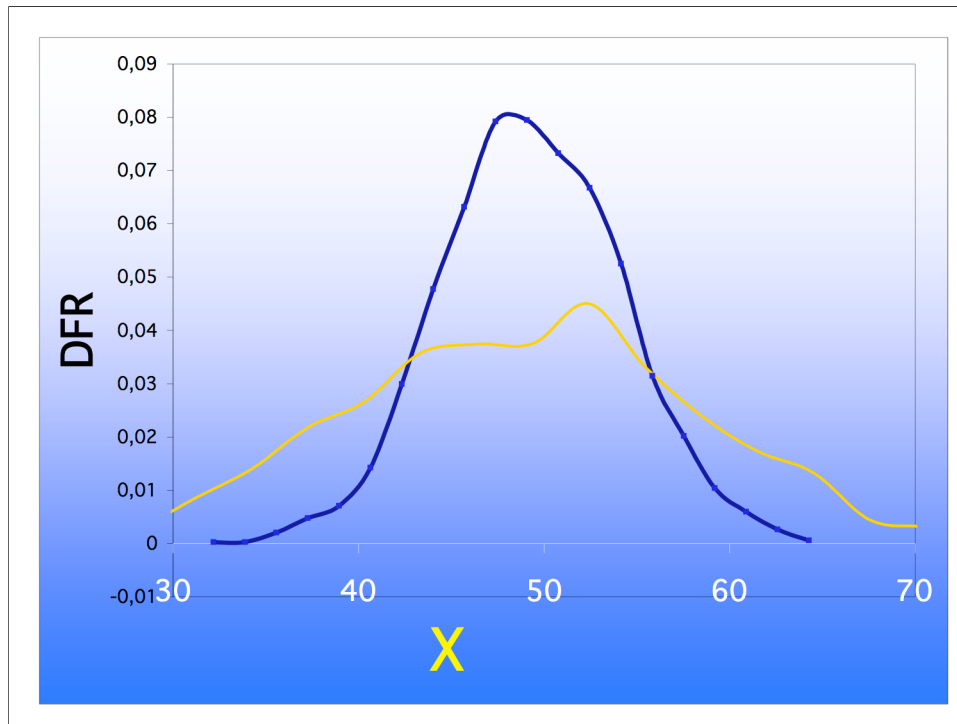




Rappel du théorème central limite : le graphique représente la distribution de fréquence relative de 2000 valeurs individuelles suivant une distribution normale de moyenne 50 et de variance σ^2 100.



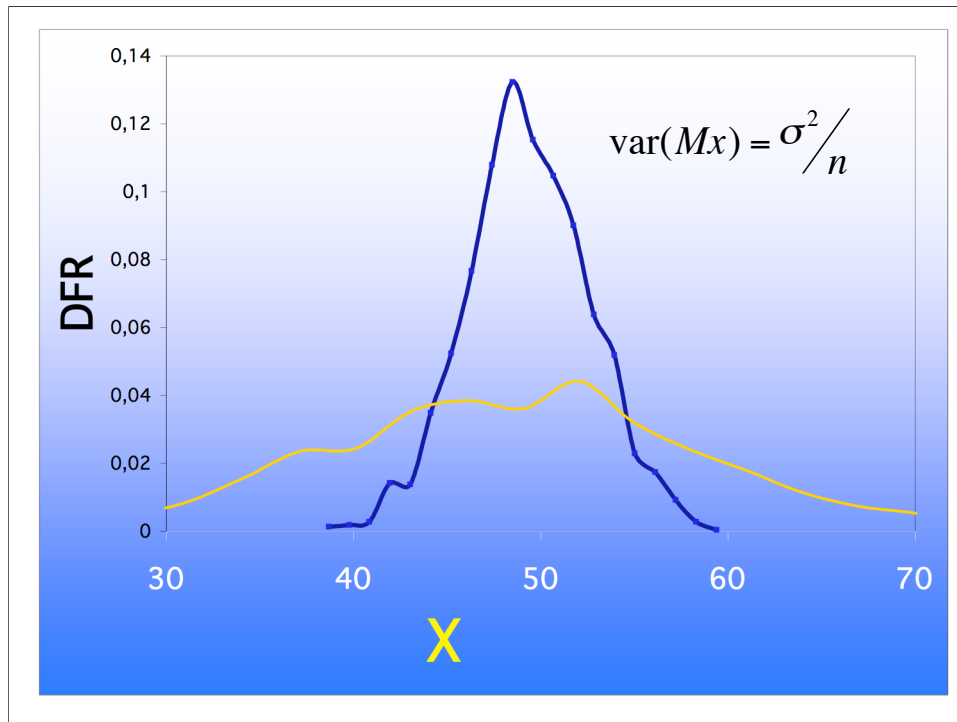
Rappel du théorème central limite :

la courbe jaune représente la distribution de fréquence relative de 2000 valeurs individuelles

suivant une distribution normale de variance $\sigma^2 = 100$; l'écart-type $\sigma = 10$;

la courbe bleue représente la distribution de fréquence relative de 2000 moyennes d'échantillons

de taille $n=4$. La variance de la moyenne est $\sigma^2/4$. L'écart-type = $\sigma / 2 = 5$



Rappel du théorème central limite :

la courbe jaune représente la distribution de fréquence relative de 2000 valeurs individuelles

suivant une distribution normale de variance $\sigma^2 = 100$; l'écart-type $\sigma = 10$;

la courbe bleue représente la distribution de fréquence relative de 2000 moyennes d'échantillons

de taille $n=9$. La variance de la moyenne est $\sigma^2/9$. L'écart-type $= \sigma / 3 = 3,3$.

$$\text{var}(x) = \sigma^2 + \sigma_A^2$$

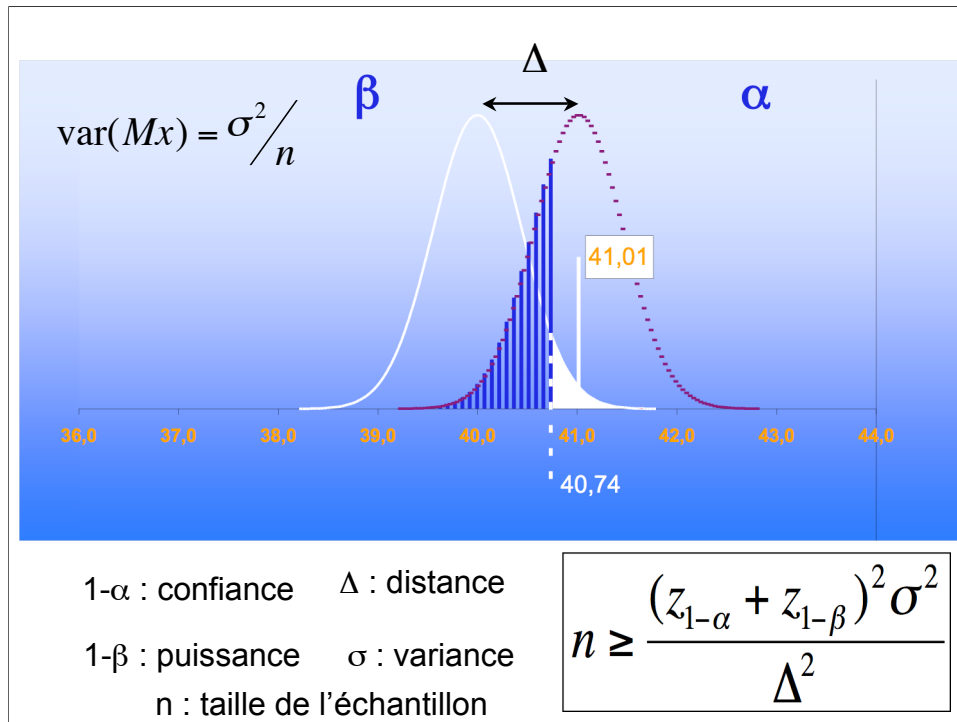
$$\text{var}(Mx_i) = \sigma^2/n + \sigma_A^2$$

$$\text{var}(Mx) = \frac{\sigma^2/n + \sigma_A^2}{n_A} = \sigma^2/n_A n + \sigma_A^2/n_A$$

Dans un échantillonnage à deux niveaux, le théorème central limite joue deux fois.

La variance de la moyenne par individu $Mx(i)$ dépend du nombre de duplicats (n),

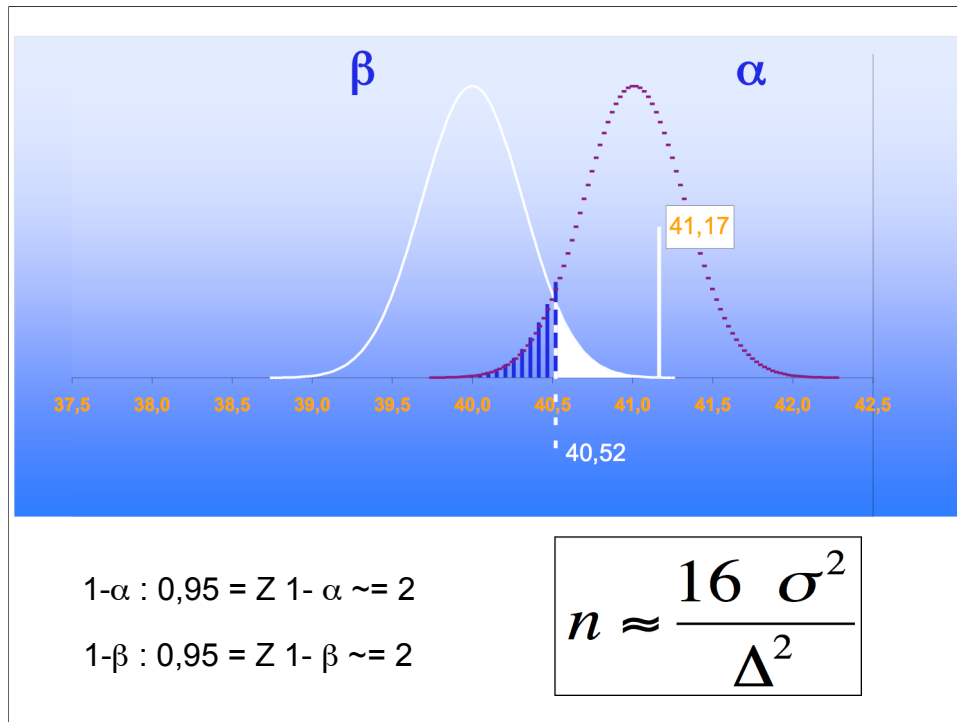
et la variance de la moyenne général MX dépend en plus du nombre de réplcats (n_A)



Cinq éléments sont liés dans un test d'hypothèse : confiance, puissance, distance entre les moyennes vraies

Variance de X , et n , la taille de l'échantillon. (la variance de la moyenne est $\sigma^2/9$).

Déterminer n demande de fixer tous les autres paramètres.



La formule se simplifie pour environs 95% de confiance et de puissance. $((Z_{1-\alpha}) + (Z_{1-\beta}))^2 \approx 16$

$$\text{var}(Mx) = \frac{\sigma^2/n + \sigma_A^2}{n_A}$$

$$n_A \approx \frac{16 (\sigma^2/n + \sigma_A^2)}{\Delta^2}$$

Pour déterminer n_A , on suit le même principe, en remplaçant σ^2 par sa valeur

$$\text{var}(M_x) = \sigma^2 / n + \sigma_A^2 / n_A$$

si $n = 1$ et $n_A = N$

$$\text{var}(M_x) = \sigma^2 / N + \sigma_A^2 / N = \text{valeur minimale}$$

si $n = N$ et $n_A = 1$

$$\text{var}(M_x) = \sigma^2 / N + \sigma_A^2 = \text{valeur maximale}$$

Suivant la logique mathématique, le plan expérimental qui fournit la meilleure estimation de la moyenne implique que $N=n_A$ et donc que $n=1$. C'est la combinaison qui génère la variance de M_x minimale.

$$n \approx \sqrt{\frac{C_A \sigma^2}{C \sigma_A^2}}$$

n : nombre de duplicats

σ^2 : variance instrumentale C : coût du duplicat

σ_A^2 : variance biologique C_A : coût du réplicat

Pour sortir de cette logique, il faut prendre en compte le coût du réplicat (C_A) et le coût du duplicat (C).

Le réplicat est souvent un être vivant, souvent un animal auquel on peut attribuer un coût objectif et

Un coût subjectif (moral ou éthique). Le duplicat est souvent une mesure de laboratoire (dosage, comptage).

La prise en compte du rapport des coûts et du rapport de variances fournit une valeur optimale pour n .

$$n \approx \sqrt{\frac{C_A \sigma^2}{C \sigma_A^2}}$$

$$n_A \approx \frac{16 (\sigma^2/n + \sigma_A^2)}{\Delta^2}$$

En assemblant les deux formules on trouve la combinaison optimale de n et nA pour une confiance

Et une puissance de 95%, à condition de fixer Δ^2 et de déterminer les coûts respectifs.

Une expérience préliminaire devra permettre d'estimer σ^2 et σ_A^2 ce qui peut être réalisé par une ANOVA I

Modèle : $x(i)j = \mu + A_i + E_{(i)j}$

Equation : SCET = SCEA + SCER

Espérance des carrés moyens :

		n	nA	E(CM)	
		i	j		
σ^2_A	A_i	1	n	$n\sigma^2_A +$	σ^2
σ^2	$E_{(i)j}$	1	1		σ^2

$E(CMA) =$	$n\sigma^2_A +$	σ^2
$E(CMR) =$		σ^2

$$n\sigma_A^2 = E(CMA) - E(CMR)$$

$$S_A^2 = \frac{CMA - CMR}{n} \quad S^2 = CMR$$

CQFD