

www.fundp.ac.be/biostats **Module 120**

120	GESTION DES RISQUES D'ERREUR	2
120.1	TYPES D'ERREUR	2
120.1.1	Erreur de type I (α)	2
120.1.2	Erreur de type II (β)	2
120.1.3	Exemple	2
120.1.4	Puissance	4
120.2	INTERPRETATION	4
120.2.1	Rejet de H_0	4
120.2.2	Acceptation de H_0	4
120.2.3	Paradoxe fondamental	5
120.3	RELATION ENTRE α ET β	6
120.3.1	Confiance et puissance d'un test	6
120.3.2	Planification expérimentale	8
120.4	LA PLUS PETITE DIFFERENCE SENSEE	9
120.5	DETERMINER N	10
120.6	EGALITE DES VARIANCES	12

120 Gestion des risques d'erreur

120.1 Types d'erreur

120.1.1 Erreur de type I (α)

La probabilité de se tromper en rejetant H_0 , α est connue et choisie (petite) par l'expérimentateur.

α est connu car si je me trompe en rejetant H_0 , c'est que H_0 est vraie. Si H_0 est vraie, $Mx(n)$ suit bien le modèle $x_{(ij)} = \mu + E_{(ij)}$ de paramètres connus : la probabilité d'obtenir par hasard une valeur Mx_{obs} entraînant le rejet de H_0 peut donc être déterminée : par définition, c'est α .

120.1.2 Erreur de type II (β)

Δ étant inconnu, β , la probabilité de se tromper en acceptant H_0 est inconnue et peut être grande.

Si je me trompe en AH_0 , elle est donc fausse. Si elle est fausse, $Mx(n)$ suit un modèle $x_{(ij)} = \mu + \Delta + E_{(ij)}$ de paramètre $\mu_1 = \mu + \Delta$. Si μ_1 est inconnu, la probabilité d'obtenir par hasard une valeur Mx_{obs} entraînant l'acceptation de H_0 ne peut donc pas être déterminée : cette probabilité s'appelle β .

Dans la plupart des situations expérimentales, Δ est inconnu, sinon pourquoi prendre des mesures pour tenter de le déterminer ?

Si Δ est fixé nous pouvons déterminer β

120.1.3 Exemple

Envisageons un scénario fictif dans lequel Δ est connu .

Supposons qu'une firme pharmaceutique commercialise un médicament dont l'activité par gélule doit être de 40 Unités, avec un écart -type de 5 Unités. Pour vérifier que l'activité moyenne des gélules ne dépasse pas 40 Unités, la production est régulièrement analysée sur 4 gélules par un contrôleur extérieur à la firme.

Pour réaliser son test, celui -ci prélève 4 gélules au hasard dans le stock, détermine l'activité du médicament par gélule, et en fait la moyenne.

Voyons tout d'abord comment il va déterminer s'il accepte ou non la production.

Si la production est bonne, l'activité du médicament dans une gélule, est

$H_0 : X : \text{v.a. } N(40;25).$

$H_0 : Mx(4) : \text{v.a. } N(40;25/4).$

$$\frac{\sigma}{\sqrt{n}} = \frac{5}{2} = 2.5$$

Pour $\alpha = 5\%$, $Z_{1-\alpha} = 1,64$ et $Mx_{\text{seuil}} = 1,65 * 2,5 + 40 = 44,11.$

En conséquence, chaque fois que la moyenne de ses 4 mesures dépassera 44,11, le contrôleur fera retirer la production de la commercialisation. En d'autres termes, cela signifie que si la production est toujours bonne, sur le long terme, 5% ne pourra pas être commercialisée. C'est le risque d'erreur α : probabilité de rejeter une H_0 vraie = probabilité de rejeter une bonne production.

Supposons à présent que lors de l'arrivée du contrôleur, le directeur sache très bien que la production en stock présente une activité moyenne de 45 Unités par gélule. Il se demande quelle est la probabilité que le contrôleur ne s'en rende pas compte.

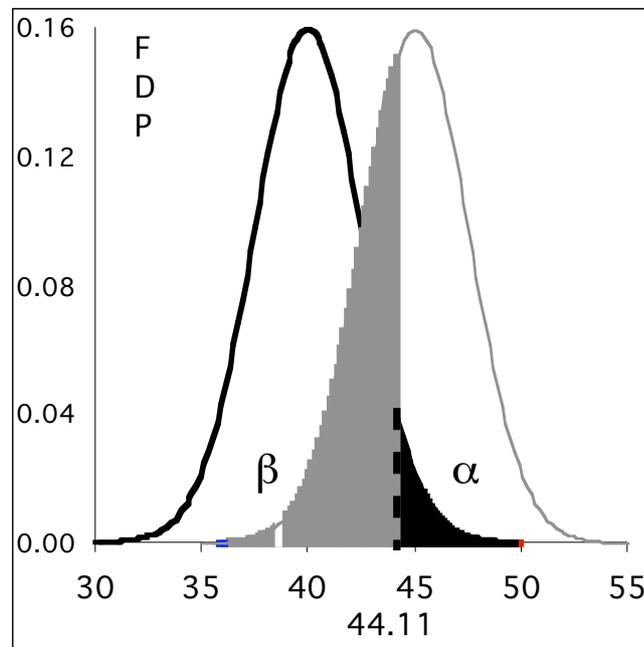


Figure 0 -1 Probabilités d'erreur α et β indiquées sur la distribution de $Mx(4)$ supposée par H_0 ($\mu = 40$) et la distribution de $Mx(4)$ réelle ($\mu = 45$).

Dans ce cas, l'hypothèse nulle est fautive. Si le contrôleur l'accepte, il se trompe en acceptant, mais, ignorant Δ , il ne peut pas déterminer son risque d'erreur β .

$H_1 : X : \text{v.a. } N(45;25).$

$H_1 : M_X(4) : \text{v.a. } N(45;25/4).$

$$\frac{\sigma}{\sqrt{n}} = \frac{5}{2} = 2.5$$

$$\beta = P(M_X(4) < 44,11) = 0,46$$

Le directeur, lui sait le déterminer, puisqu'il connaît la moyenne réelle de la population. (NDLR : toute ressemblance avec des personnes réelles serait pure coïncidence).

Le contrôleur accepte s'il trouve une valeur $M_X(4)$ inférieure à 44,11.

Sous H_0 $M_{X_{\text{seuil}}} = 1,65 * 2,5 + 40 = 44,11$

$$\beta = P(M_X(4) < 44,11) = P(Z \leq (44,11 - 45)/2,5) = P(Z \leq -0,36) = 0,36.$$

Attention : dans cet exemple, la valeur de la probabilité=0,36 et la valeur absolue de $Z = 0,36$, ce qui est une pure coïncidence.

120.1.4 Puissance

On peut donc se rendre compte que dans ces conditions le contrôleur a seulement un peu plus de 1 chance sur 2 (0,54) de rejeter cette production : on dira que son test est peu puissant car sa probabilité de mettre en évidence que l'hypothèse nulle est fautive est faible.

$1 - \beta$ est la <u>puissance</u> du test ; c'est la probabilité de mettre en évidence que l'hypothèse nulle est fautive, lorsqu'elle est effectivement fautive.
--

Si Δ est inconnu, β est inconnu et la puissance $1 - \beta$ est inconnue.

120.2 Interprétation

120.2.1 Rejet de H_0

L'erreur de type I, de probabilité α , consiste à rejeter une H_0 vraie. α est connu, petit, fixé par l'expérimentateur.

Lorsqu'il rejette une hypothèse nulle, l'expérimentateur encourt un petit risque α et pourra affirmer ses conclusions en connaissance de cause. Réciproquement, il aura confiance $(1 - \alpha)$ dans ses conclusions.

Il pourra conclure " mon échantillon me fournit une quasi certitude ($P < \alpha$) de l'in vraisemblance de H_0 ."

120.2.2 Acceptation de H_0

L'erreur de type II, de probabilité β consiste à accepter une H_0 fautive. Si Δ est

inconnu, β est inconnu, non fixé par l'expérimentateur et éventuellement grand. Lorsqu'il accepte une hypothèse nulle, l'expérimentateur encourt un risque β et il ne pourra rien affirmer.

Il pourra seulement conclure " mon échantillon ne m'a fourni aucune évidence de l'in vraisemblance de H_0 ."

120.2.3 Paradoxe¹ fondamental

Cette différence fondamentale entre l'acceptation et le rejet d'une hypothèse est un fait dans tous les types d'expérimentation.

Prenons le cas d'un médecin qui reçoit un patient qui se plaint de fatigue persistante . Si le médecin réalise une analyse qui donne une évidence positive du syndrome de Lyme², il sera pratiquement certain que celui -ci est à l'origine du malaise. Si par contre il ne trouve rien de particulier, il ne peut certainement pas dire "vous n'avez rien". Il va poursuivre ses investigations, multiplier les analyses ... tant que les tests seront négatifs il lui sera impossible de dire si le patient n'a rien ou s'il n'a pas trouvé ce qu'il avait...

Prenons un exemple encore plus simple : je vous présente un sac de 1000 balles de ping -pong en affirmant qu'elles sont toutes blanches. Vous en prélevez une et elle est noire : vous êtes certain que mon affirmation est fausse (ici $\alpha = 0$). Si vous en prélevez une blanche : que pouvez -vous dire de mon affirmation? Rien, sinon que vous n'avez acquis aucune évidence qu'elle soit fausse... (β inconnu).

¹ Paradoxe : situation contraire à l'opinion commune.

² Syndrome de Lyme : infection causée un spirochète nommé *Borrelia* transmis par une piqûre de tique.

120.3 Relation entre α et β

120.3.1 Confiance et puissance d'un test.

Le meilleur test est évidemment celui qui permet de travailler simultanément avec une grande confiance ($1 - \alpha$) et une grande puissance ($1 - \beta$). Dans ces conditions, je suis presque certain d'accepter H_0 si elle est vraie et de la rejeter si elle est fautive. Cependant, sur un graphique on se rend très facilement compte que β dépend de α : plus α est petit, plus le seuil de rejet s'éloigne de μ , plus il est probable d'accepter H_0 si elle est fautive.

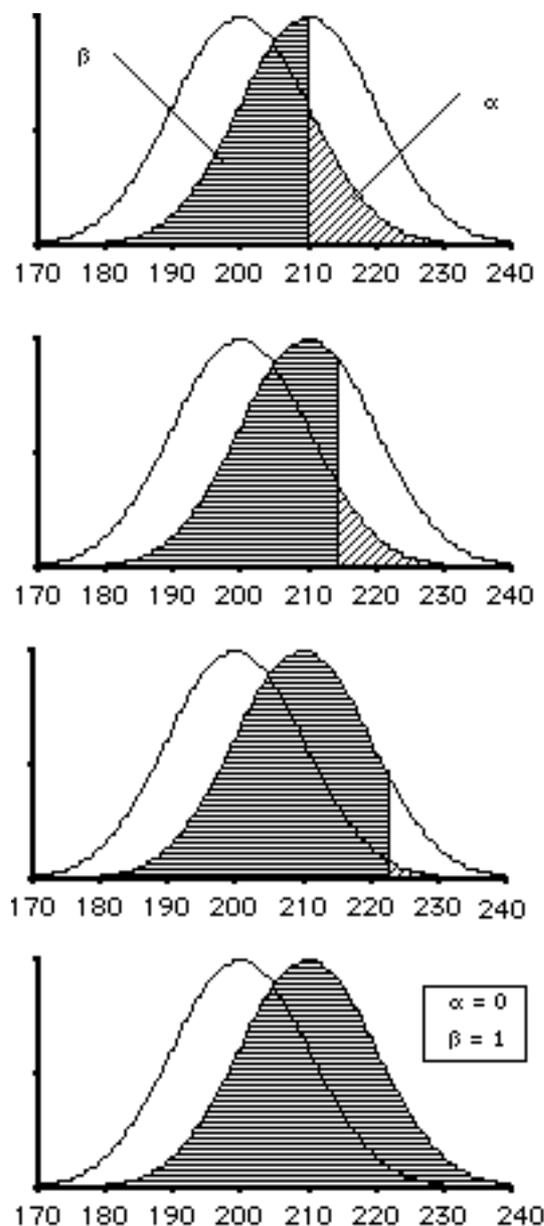


Figure 0 -2 Relation entre α (hachuré oblique) et β (hachuré horizontal) : si α diminue, β augmente, et réciproquement.

On voit donc qu'il est exclu de prendre une valeur de α nulle : dans ces conditions, je n'obtiendrai jamais dans l'échantillon une valeur Mx_{obs} suffisamment éloignée de μ que pour entraîner un rejet. Par contre, il est évident que α doit rester assez petit : à quoi servirait le test si je ne peux rien affirmer ni en rejetant (α grand), ni en acceptant H_0 (β inconnu)? Le compromis entre une valeur de α qui ne soit ni trop grande, ni trop petite aboutit généralement à faire référence à des valeurs conventionnelles.

valeurs conventionnelles $\alpha = 5\%, 1\%$

Comment dès lors augmenter la puissance d'un test sans diminuer sa confiance? Examinons sur un graphique de quoi dépend β , pour α constant :

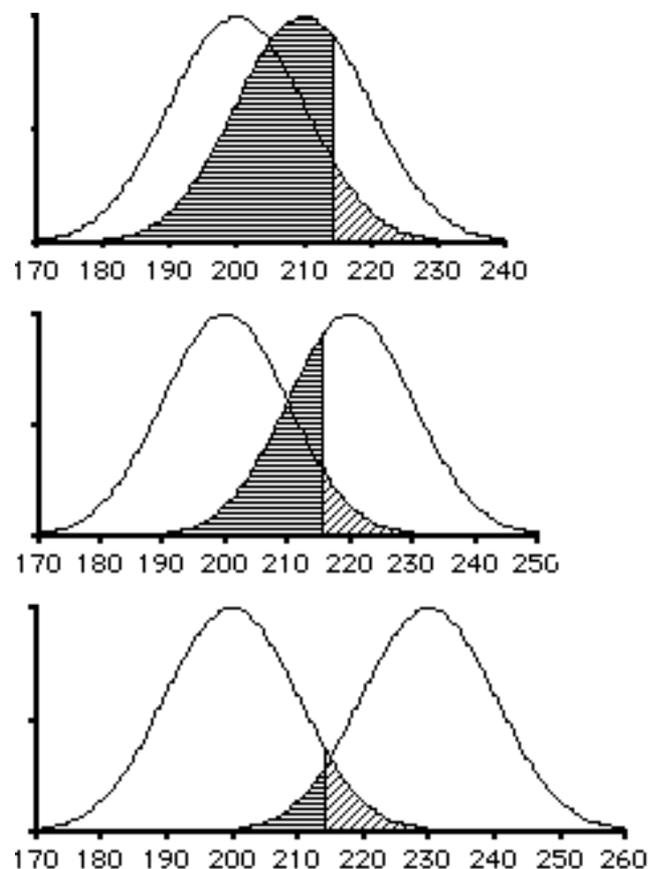


Figure 0 -3 Relation entre $1 - \beta$, puissance du test et Δ , le « degré de fausseté » de H_0 .

- β dépend de Δ la différence entre la moyenne vraie et la moyenne supposée par H_0 : plus Δ augmente, plus β diminue.

Ceci revient simplement à dire qu'il est plus facile de mettre en évidence une forte différence qu'une faible Δ est liée à la variabilité factorielle (information) qu'il faut rendre maximale par rapport au bruit (résiduelle).

Du point de vue expérimental, cela signifie qu'il faut essayer de rendre maximal l'effet que l'on cherche à mettre en évidence (par exemple en utilisant sur un cobaye³ des doses plus fortes que thérapeutiques : argument avancé pour justifier l'expérimentation animale).

- β dépend de σ^2/n : si la variance diminue, et/ou si n augmente, la variance de la distribution d'échantillonnage σ^2/n diminue, et β diminue.

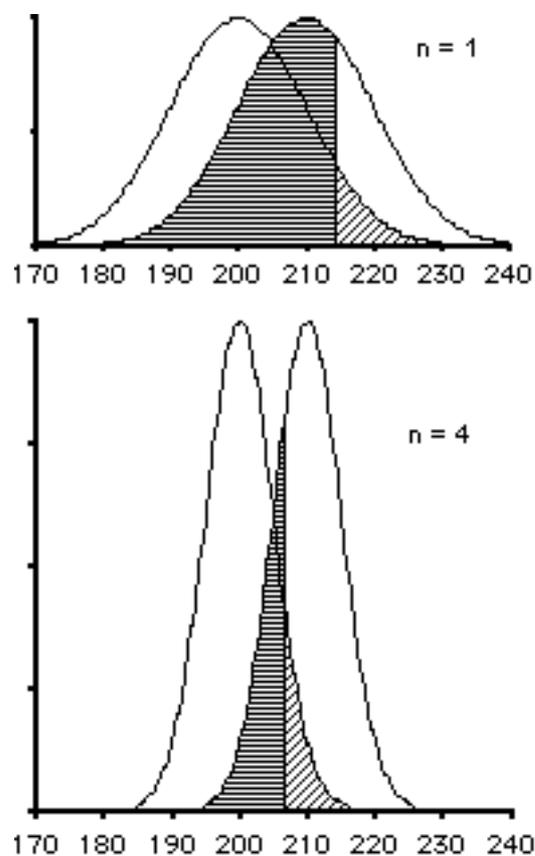


Figure 0 -4 Relation entre puissance et variance de la distribution d'échantillonnage, pour une même valeur de Δ .

120.3.2 Planification expérimentale

³ Cobaye : Sujet d'expérience, en général

On a trop rapidement tendance à en conclure que la taille de l'échantillon doit nécessairement être élevée. En fait, dans une bonne expérience, il faut tout d'abord rendre minimale la variabilité résiduelle (bruit). Différents moyens sont possibles :

- diminuer la variabilité résiduelle expérimentale (en prenant des précautions, en fixant l'heure de la mesure, en choisissant un appareil plus performant, voire en choisissant une autre variable qui permette de rendre compte plus précisément d'un phénomène : pour étudier la croissance, la taille est plus stable que le poids...)

- diminuer la variabilité résiduelle individuelle en définissant une population plus restrictive (par exemple en étudiant séparément une tranche d'âge, une seule race ou un seul sexe).

Dans une bonne expérience, on fixe un α qui ne soit pas inutilement petit, on rend Δ le plus grand possible, on limite au maximum la variabilité résiduelle, puis enfin, on choisit la taille d'échantillon minimale. Celle-ci dépendra de la puissance que l'on veut obtenir et de la plus petite valeur de Δ qui ait un sens.

120.4 La plus petite différence sensée

Fixer Δ à une valeur minimale revient à préciser le niveau à partir duquel la différence entre les moyennes μ_1 et μ a une signification pour l'expérimentateur. Pour éviter la confusion entre signification (différence interprétable) et significatif (différence non fortuite) nous parlerons d'une différence sensée⁴.

Lorsque l'expérimentateur est amené à accepter H_0 , une de ses attitudes possibles est de recommencer l'expérience en augmentant la taille de l'échantillon. Remarquons cependant qu'il est impossible de prouver que H_0 est vraie : en effet, plus Δ est petit, plus la puissance est faible, et plus le nombre d'observations nécessaires pour pouvoir mettre cette différence en évidence est élevé. A la limite, pour prouver que H_0 est vraie, il faudrait pouvoir prélever toute la population. L'expérimentateur va donc rencontrer une limite matérielle qui l'obligera à renoncer à augmenter encore son échantillonnage, et à accepter H_0 avec un risque d'erreur inconnu.

En cas d'acceptation de H_0 , l'intervalle de confiance fournit une information importante.

⁴ Sensé : Qui a du bon sens, raisonnable, pertinent.

Supposons qu'un médecin teste l'effet d'un médicament susceptible d'augmenter la pression sanguine. Dans la population étudiée, sans traitement, la pression sanguine moyenne est de 140 mm Hg. Sur un échantillon de 50 patients ayant reçu le médicament, on calcule une moyenne $Mx(50) = 141,2$ mm Hg, qui entraîne une acceptation de H_0 .

Conclusion de cette expérience :

- soit le médicament n'a pas d'effet*
- soit l'échantillon prélevé n'a pas permis de montrer cet effet.*

Doit-il recommencer l'expérience avec un nombre de patients plus élevé? Examinons l'intervalle de confiance de la moyenne. Supposons les valeurs suivantes : $P(137,2 \leq \mu \leq 145,2) \geq 0,95$.

On se rend compte que si le médicament a un effet, il est peu probable que la moyenne de la population traitée soit supérieure à 145,2.

La question qui se pose est donc de savoir si cet effet éventuel est sensé (est intéressant ou non ?)

Si l'expérimentateur recherche un médicament qui doit provoquer une augmentation d'au moins 10 mm Hg, l'expérimentateur pourra accepter H_0 définitivement. Il n'aura pas prouvé que H_0 est vraie, mais il pourra considérer qu'il n'a vraisemblablement pas l'effet attendu. La différence, si elle existe, n'a pas de sens dans ce contexte.

120.5 Déterminer n

Définir la plus petite différence significative pour l'expérimentateur permet de déterminer le nombre d'observations nécessaires pour réaliser le test avec une puissance fixée. Imaginons qu'au laboratoire, j'achète une pipette automatique qui prélève un volume de 5 ml avec un écart -type de 0.01 ml. Avant de l'utiliser, je désire m'assurer qu'elle ne prélève pas systématiquement plus de 5 ml. Les hypothèses sont :

$$H_0 : \mu = 5 \text{ et } H_1 : \mu > 5$$

Une série de n pipetages fournira une moyenne $Mx(n)$, qui me permettra d'accepter ou de rejeter H_0 avec une probabilité α .

Il y aura donc une probabilité α d'écarter une bonne pipette. Mais mon objectif est ici de pouvoir accepter H_0 avec une certitude suffisamment grande qu'elle soit vraie. En effet, je dois pouvoir utiliser la pipette dans laquelle j'ai suffisamment confiance.

Définissons ce que je vais considérer comme une "mauvaise pipette" (si la pipette prélève 5,0001 ml, cette différence n'a pas réellement de sens).

Fixons la plus petite différence sensée à 0,01 ml. Je veux dès lors être pratiquement certain ($P = 1 - \beta$) de mettre en évidence que la pipette prélève plus de 5 ml à partir du moment où elle prélève en réalité un volume supérieur ou égal à 5.01 ml.

Quel est le nombre de pipetages à réaliser pour que les conditions suivantes soient réalisées lors du test :

$\alpha = 5\%$	$\sigma = 0,01$
$\beta = 5\%$	$\Delta = 0,01$

Tableau 0 -0-1 Paramètres fixés, permettant la détermination de n .

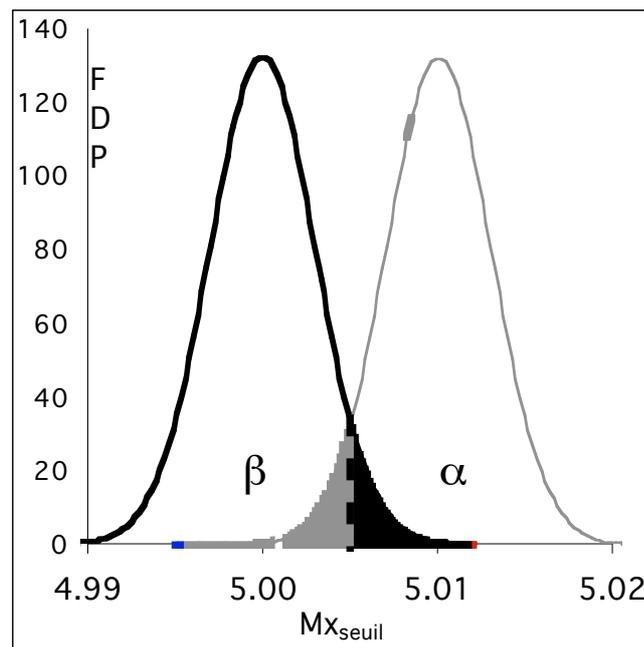


Figure 0 -5 Distribution d'échantillonnage réelle et supposée pour l'erreur de n pipetages.

Remarquons qu'une même valeur Mx_{seuil} délimite la probabilité α dans la distribution centrée sur μ (ici : $\mu = 5$) la probabilité β dans la distribution centrée sur μ_1 (ici : $\mu_1 = 5,01$)

$$z = \frac{x - \mu}{\sigma}$$

$$x = \mu + z\sigma$$

en particulier :

$$Mx_{seuil} = \mu + z_{1-\alpha} \frac{\sigma}{\sqrt{n}}$$

$$Mx_{seuil} = \mu_1 + z_{\beta} \frac{\sigma}{\sqrt{n}}$$

$$Mx_{seuil} = \mu_1 - z_{1-\beta} \frac{\sigma}{\sqrt{n}}$$

$$\frac{(z_{1-\alpha} + z_{1-\beta})\sigma}{\sqrt{n}} = \mu_1 - \mu = \Delta$$

$$\frac{(z_{1-\alpha} + z_{1-\beta})\sigma}{\Delta} = \sqrt{n}$$

$$n = \frac{(z_{1-\alpha} + z_{1-\beta})^2 \sigma^2}{\Delta^2}$$

Équation 0-1

pour simplifier, nous pouvons retenir que pour α et $\beta = 5\%$, (z proche de 2) on obtient approximativement n par la relation :

$$n \approx \frac{16\sigma^2}{\Delta^2} \quad \text{Équation 0-2}$$

En effectuant le test d'hypothèse sur un échantillon de 3 pipetages (il faut arrondir à l'unité supérieure), je pourrai tester la pipette en ayant :

- 95% de confiance, soit moins de 5% de chance d'écarter une bonne pipette
- 95% de puissance, soit moins de 5% de chance d'accepter une pipette dont le volume est \geq à 5.01 ml.

120.6 Egalité des variances

L'hypothèse d'égalité des variances est une des plus fondamentales (avec l'indépendance des observations) qui garantit la validité des tests statistiques.

Imaginons qu'un vétérinaire compare le niveau de progestérone⁵ de 4 brebis⁶ gravides⁷, par rapport aux brebis contrôles (non gravides).

Imaginons que la concentration de cette hormone soit de $5 U^8 \pm 1U$ (σ) dans la population des brebis contrôles.

Imaginons que la variabilité de la progestérone soit plus élevée chez les brebis gravides, sans changement de la moyenne. $\sigma = 3U$ $\mu = 5U$

Si un test de comparaison de moyenne est réalisé sans tenir compte de l'inégalité des variances, le seuil α réel est (parfois nettement) plus grand que prévu par le modèle.

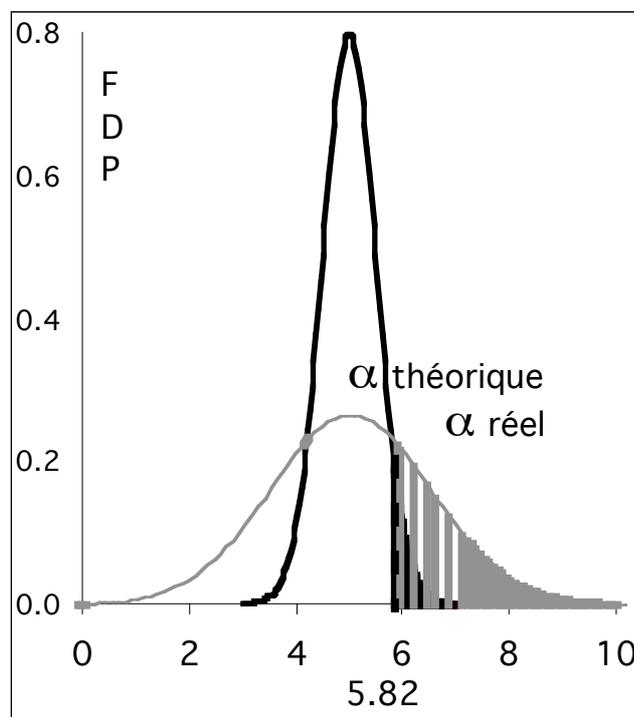


Figure 0 -6 Comparaison de la distribution d'échantillonnage de $Mx(4)$ pour la population contrôlée ($\sigma = 1U$) et la population gravide ($\sigma = 3U$) sous H_0 vraie $\mu = 5U$. La probabilité de rejeter à tort H_0 (α réel, en gris) est nettement plus grande que prévue par le modèle sous H_0 (α théorique, en noir).

⁵ Progestérone : hormone progestative sécrétée par le corps jaune de l'ovaire pendant la deuxième partie du cycle menstruel et pendant la grossesse.

⁶ Brebis : mouton femelle.

⁷ Gravide : qui porte un fœtus ou un embryon, en parlant d'une femelle ou d'un utérus.

⁸ U = unité d'activité de l'hormone

L'expérimentateur qui teste le modèle $\mu = 5U$ en supposant que les brebis gravides ont également un écart -type de $1U$ va rejeter H_0 lorsque $Mx(4) > Mx_{seuil}$.

Au seuil α 5%,

$$Mx_{seuil} = \mu + z_{1-\alpha} \frac{\sigma}{\sqrt{n}} =$$

$$5 + 1.64 \frac{1}{\sqrt{4}} = 5.82$$

En considérant $\sigma = 3U$ probabilité de rejet réelle est donc

$$P(Z \geq \frac{Mx_{seuil} - \mu}{\sigma/\sqrt{n}}) = P(Z \geq \frac{5.82 - 5}{3/\sqrt{4}}) = 0.29$$