Test statistique

David NERINI, oct. 2018

1 Introduction

Le problème d'effectuer un *test statistique* se pose dès lors l'on cherche à décider si une valeur arbitrairement choisie d'un paramètre populationnel (la moyenne, la médiane, l'écart-type,...) peut être valide au regard d'un nombre limité d'observations réalisées dans la population. Par exemple, on peut examiner la différence entre les moyennes de deux populations gaussiennes (disons des tailles). Notre première interrogation est de savoir si, à partir d'un échantillon de chacune des deux populations, les observations indiquent une vraie différence entre les moyennes des populations et non pas uniquement des fluctuations liées au hasard de l'échantillonnage.

Le type d'hypothèse que l'on teste en statistique est plus restreint que les hypothèses scientifiques en général. Supposer qu'il y a de la vie sur d'autres planètes de l'univers, que chaque particule de matière est composée d'un ensemble d'anneaux vibrants d'énergie sont des hypothèses scientifiques. Les hypothèses statistiques qui vont nous intéresser concernent le comportement d'un ensemble de k variables aléatoires $X^1, \cdots, X^k$ (la taille, le poids, $\cdots$) que l'on peut observer, ce qui n'est pas le cas des hypothèses précédentes. On peut voir ces variables comme les coordonnées d'un vecteur $\mathbf{X} \in \Omega \subset \mathbb{R}^k$ dans un espace Ω dit espace échantillon de dimension k dont les axes correspondent à chacune des variables. Les coordonnées de $\mathbf{X}$ sont aléatoires : on peut donc associer une densité de probabilité à $\mathbf{X}$ ainsi que la possibilité de construire un échantillon de taille n noté $E = \{\mathbf{X}_1, \cdots, \mathbf{X}_n\}$. Si l'on sélectionne une région $\omega \subset \Omega$ dans l'espace échantillon on peut en principe calculer la probabilité que le point $\mathbf{X}$ appartienne à ω en utilisant cette densité. Voici quelques exemples d'hypothèses dites statistiques :

- Une distribution gaussienne a une moyenne et une variance connues
- Une distribution gaussienne a une moyenne connue et une variance non spécifiée
- Une distribution est gaussienne de paramètres inconnus
- Deux distributions quelconques sont identiques.

Chacun de ces exemples implique des propriétés particulières de l'espace échantillon Ω. En d'autres termes, à chacune de ces hypothèses va correspondre un sous-espace de $\mathbb{R}^k$ particulier. On notera également que les trois premiers exemples supposent que la distribution des observations est connue a priori (loi normale), c'est à dire que leur forme est déterminée algébriquement et contrôlée par la connaissance de paramètres (moyennes et variances, covariances dans le cas gaussien). De telles hypothèses sont dites *paramétriques*. La quatrième hypothèse est de nature différente et nous n'en parlerons pas ici.

2 Région critique et hypothèse alternative

La première étape du test d'une hypothèse privilégiée (appelée aussi *hypothèse nulle* et notée H_0) se construit à partir d'un échantillon observé et consiste à partitionner l'espace échantillon Ω en deux régions RH_0 and $\overline{RH}_0$ telles que $RH_0 \cap \overline{RH}_0 = \emptyset$ et $RH_0 \cup \overline{RH}_0 = \Omega$ (ces deux régions ne sont pas vides et leur réunion constitue l'espace échantillon).

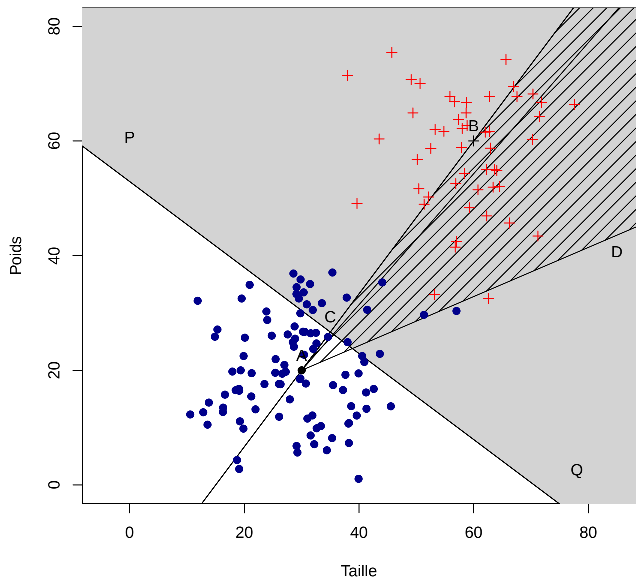

Figure 1: Zone de non rejet de H_0 (blanc) et deux zones critiques de forme différente (gris et hachuré). Dans ce cas, ces zones sont des surfaces.

Si une observation de $\mathbf{X}$ tombe dans la région $\overline{RH}_0$ appelée *région de non-rejet*, on dira que l'hypothèse nulle n'est pas rejetée. Si au contraire une observation de $\mathbf{X}$ tombe dans la zone RH_0, on dira que l'hypothèse nulle est rejetée. Cette zone RH_0 est appelée *région critique* ou *zone de rejet*. Si l'on connait la densité de probabilité du vecteur $\mathbf{X}$ sous H_0, on peut déterminer la région critique RH_0 telle que la probabilité de de rejeter l'hypothèse nulle (c'est à dire la probabilité que $\mathbf{X}$ tombe dans RH_0) soit identique à une valeur fixée α telle que

$$P\left(\mathbf{X} \in RH_0\right) | H_0) = \alpha.$$

Cette valeur α, qui est une probabilité, est appelée *niveau de significativité* du test : elle est volontairement choisie petite car on va toujours privilégier l'hypothèse nulle en évitant de la rejeter trop facilement. La construction d'un test, c'est à dire la forme de la région critique RH_0, va dépendre fortement du choix d'une autre hypothèse, celle dite *hypothèse alternative* notée H_1. Nous proposons ici de montrer comment un test est construit et comment on en mesure sa qualité. Une façon de procéder est de calculer la *puissance* du test. Nous allons proposer une démarche de calcul de cette quantité. Nous terminerons par une méthode permettant de choisir la taille adéquate n d'un échantillon de manière à construire un test de puissance désirée.

3 Elaboration d'un test

Nous allons placer la suite de cet exposé dans le cas où l'on travaille avec une variable réelle $X \in \mathbb{R}$. On suppose que l'on dispose d'un échantillon *i.i.d* $E = \{X_1, \ldots, X_n\}$ de cette variable X de loi fixée dont la forme dépend d'un paramètre θ. A partir d'une observation de l'échantillon E, notée $E_{obs} = \{x_1, \ldots, x_n\}$, on souhaite effectuer un test sur ce paramètre permettant de confronter une hypothèse nulle H_0 à une hypothèse alternative H_1.

3.1 Etape 1

On choisit d'abord les deux hypothèses H_0 et H_1 de la manière suivante.

Définition 3.1 *Soit une hypothèse nulle $H_0 : \theta = \theta_0$ que l'on souhaite tester contre une hypothèse alternative H_1. Cette dernière peut prendre les formes suivantes :*

- *$H_1 : \theta > \theta_0$ ou bien $\theta < \theta_0$ (tests unilatéraux)*
- *$H_1 : \theta = \theta_1$ ou bien $\theta \neq \theta_0$ (test unilatéral et test bilatéral pour le second cas)*

Le choix de H_1 est guidé par l'énoncé du problème à traiter.

3.2 Etape 2

En se plaçant sous H_0, on va chercher à construire une *statistique de test*, c'est à dire une variable aléatoire dont la densité dépend de l'échantillon $E = \{X_1, \cdots, X_n\}$ des n variables aléatoires X_i supposées indépendantes et identiquement distribuées ($i.i.d.$), de loi mère celle de X , tel que

$$Z = g\left(X_1, \cdots, X_n\right).$$

Le choix de cette *variable décisionnelle* est une étape délicate car il s'agit dans le meilleur des cas d'utiliser les théorèmes des probabilités et des statistiques pour déterminer la loi de Z, ce qui n'est pas toujours possible! En général cette statistique de test sera dépendante du paramètre θ. On va ensuite déterminer le comportement de Z quand on s'éloigne de l'hypothèse nulle pour déterminer une zone de rejet. L'idée d'un test est de considérer faible la probabilité d'observer une valeur de Z notée $z_{obs} = g\left(x_1, \cdots, x_n\right)$ (et calculée avec E_{obs}) qui s'éloigne trop de son espérance, si l'hypothèse nulle est la bonne.

Définition 3.2 *On appelle zone de rejet de l'hypothèse nulle H_0, notée RH_0 un intervalle ouvert de $\mathbb{R}$. La forme de cet intervalle dépend de l'hypothèse H_1 :*

- *Si $H_1 : \theta < \theta_0$ ou si $H_1 : \theta = \theta_1, \theta_1 \leq \theta_0$, alors nous sommes dans le cas d'un test unilatéral avec zone de rejet à gauche et $RH_0 =]-\infty; z_\alpha[$ est la première demi droite ouverte.*
- *Si $H_1 : \theta > \theta_0$ ou si $H_1 : \theta = \theta_1, \theta_1 \geq \theta_0$, alors nous sommes dans le cas d'un test unilatéral avec zone de rejet à droite et $RH_0 =]z_{1-\alpha}; +\infty[$ est la seconde demi droite ouverte.*
- *Si $H_1 : \theta \neq \theta_0$, alors nous sommes dans le cas d'un test bilatéral avec zone de rejet à droite et à gauche et $RH_0 =]-\infty; z_{\alpha/2}[\cup]z_{1-\alpha/2}; +\infty[$ est l'union de deux intervalles disjoints.*

Les zones de rejet ou de non rejet de ces hypothèses seront des intervalles de la droite des réels. La taille de la zone de rejet RH_0 dépend du niveau de significativité $0 \leq \alpha \leq 1$, choisit arbitrairement petit et signifiant que le résultat observé à une probabilité faible d'être le fruit du hasard. Les valeurs z_α et $z_{1-\alpha}$ sont les quantiles d'ordre α et d'ordre $1-\alpha$ de la variable aléatoire Z que l'on est en train de tester.

3.3 Etape 3

Grâce à l'échantillon observé E_{obs}, on établit une règle de décision qui permet, après calcul de la valeur observée z_{obs} de Z, de savoir si l'on rejette où si l'on ne rejette pas l'hypothèse privilégiée H_0.

Définition 3.3 *A l'issu d'un test statistique, on prendra la décision de rejeter H_0, notée $D(RH_0)$ si $z_{obs} \in RH_0$ ou la décision de ne pas rejeter H_0, notée $D(\overline{RH_0})$ sinon.*

La conclusion du test s'énonce ensuite de manière littérale, en n'oubliant pas de faire apparaître son niveau de significativité. Dans le cas d'un test unilatéral, on peut calculer sa *p-value* notée $\alpha^\star$ qui correspond à la probabilité de rejeter H_0 en prenant comme quantile z_{obs}

$$\alpha^\star = P(X \in]-\infty; z_{obs}[\,|H_0) \text{ pour une zone de rejet à gauche,}$$
$$\alpha^\star = P(X \in]z_{obs}; +\infty[\,|H_0) \text{ pour une zone de rejet à droite.}$$

3.4 Etape 4

La procédure décisionnelle peut être complétée en calculant la puissance du test effectué. C'est l'objet de la section suivante.

4 Puissance de test

Rejeter une hypothèse ne signifie pas qu'elle soit fausse. En effet, en supposant H_0 vraie, la probabilité que la valeur observée de Z ne se situe pas dans l'intervalle $\overline{RH_0}$ existe.

Définition 4.1
- *On appelle* ***erreur de première espèce*** *(ou erreur de type I) le fait de rejeter l'hypothèse nulle H_0 en faveur d'une hypothèse alternative H_1 quand cette hypothèse H_0 est vraie. On notera par la suite $\alpha = P(Z \in RH_0|H_0)$ la probabilité de cette erreur.*
- *On appelle* ***erreur de deuxième espèce*** *ou erreur de type II le fait de ne pas rejeter l'hypothèse nulle H_0 alors qu'une hypothèse alternative H_1 est vraie. On notera $\beta = P(Z \in \overline{RH_0}|H_1)$ la probabilité de cette erreur.*

Nous allons considérer l'exemple suivant pour illustrer ces notions. Supposons que l'on souhaite tester l'effet d'un régime alimentaire sur une population de bars juvéniles dans une pisciculture. Notons X (mm) la taille des bars et nous savons qu'elle est distribuée normalement avec une moyenne $\mu = 170\ mm$ et un écart-type $\sigma = 10\ mm$. A partir d'un échantillon de taille $n = 25$, l'ingénieur en charge de l'étude décide de rejeter l'hypothèse nulle ($H_0 : \mu = 170\ mm$) si la moyenne de son échantillon $\overline{x}$ est supérieure ou égale à 172 mm : c'est l'hypothèse alternative $H_1\ : \mu \geq 172\ mm$. Nous sommes dans le cas d'un test unilatéral avec zone de rejet à droite. Si H_0 est rejetée (H_1 acceptée), l'ingénieur conclura alors en faveur du régime alimentaire. Quelle est alors la probabilité que l'ingénieur commette une erreur de type I?

Une erreur de première espèce est commise si l'ingénieur observe une valeur de $\overline{X}$ qui tombe dans la région RH_0 de rejet de H_0 et qui correspond à la demi droite sous la zone rouge de la figure 1, graphe du haut. En vertu du théorème central limite, la statistique de test $\overline{X}$ suit une loi normale de moyenne $\mu = 170\ mm$ et de variance σ^2/n sous H_0. La probabilité α peut donc être déduite en utilisant la version centrée-réduite de $\overline{X}$

$$Z = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \tag{1}$$

que l'on sait suivre une gaussienne de moyenne $\mu = 0$ et de variance $\sigma^2 = 1$ sous H_0. De cette manière, on en déduit que $\alpha = P(\overline{X} \geq 172|H_0) = P(Z \geq 1.00) = 0.1587$. On note $\overline{x}_{seuil} = \overline{x}_{1-\alpha} = 172$ la borne inférieure de RH_0. Or, une probabilité de $\alpha = 0.1587$ est élevée pour prendre une décision. Nous allons voir comment procéder pour réduire cette probabilité.

Imaginons maintenant que la vraie moyenne de la population soit $\mu = 173\ mm$ sans que l'ingénieur n'en sache rien : c'est l'hypothèse H_1 qui est vraie dans ce cas. On suppose que l'écart type est toujours le même. Cette information supplémentaire ne change pas la nature unilatérale du test avec zone de rejet de H_0 à droite. Quelle est dans ce cas la probabilité que l'ingénieur commette une erreur de type II?

Une erreur de deuxième espèce est commise si celui-ci observe une valeur moyenne de l'échantillon $\overline{X}$ qui tombe dans la région de non-rejet de H_0, notée $\overline{RH_0}$ (ou zone de rejet de H_1, RH_1) et qui correspond à la demi droite qui supporte la zone bleue sur la figure 1, c'est à dire à l'intervalle $\overline{x} \leq 172\ mm$. Graphiquement, la probabilité β de commettre une erreur de type II peut être calculée en utilisant, comme précédemment, la distribution normale centrée-réduite mais sous $H_1 : \mu = 173\ mm$. De cette manière, on obtient

$$\beta = P(\overline{X} < 172|H_1) = P(Z < -0.50) = 0.3085. \tag{2}$$

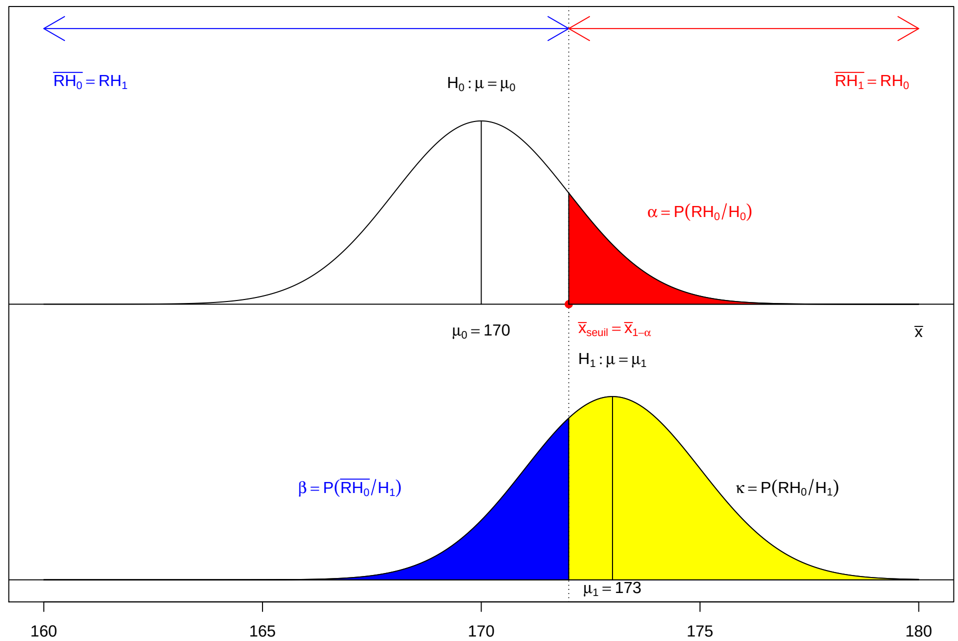

Figure 2: Représentation des distributions gaussiennes sous H_0 et sous H_1. Les zones colorées représentent les probabilités notées de la même couleur. La variable étudiée est ici $\overline{X}$, la moyenne empirique de l'échantillon.

Cette probabilité de commettre une erreur de type II est également élévée. Comment faire pour la réduire? Au lieu de considérer que l'ingénieur commette une erreur, nous allons considérer la probabilité qu'il prenne la bonne décision. C'est de cette manière qu'on peut évaluer la puissance du test.

Définition 4.2 *La* ***puissance d'un test*** *est la probabilité de prendre une décision correcte si l'hypothèse alternative H_1 est vraie. De la même manière, la puissance d'un test est la probabilité de rejeter l'hypothèse nulle H_0 quand H_1 est vraie. On note cette puissance :*

$$\kappa = 1 - \beta = P(Z \in RH_0 | H1)$$

Reformulons la question de notre problème de la manière suivante : quelle est la probabilité que l'ingénieur prenne la bonne décision de rejeter l'hypothèse nulle en faveur de l'alternative $H_1 : \mu = 173\ mm$ que nous savons vraie?

Dans ce cas, la bonne décision est prise si $\overline{X} \geq 172\ mm$ ce qui nous donne la puissance du test

$$\kappa = P(\overline{X} \geq 172 | \mu = 173) = P(Z \geq -0.5) = 0.6915 \tag{3}$$

probabilité évidemment proportionnelle à ce que l'on a déjà calculé :

$$\kappa = 1 - \beta = 1 - 0.3085 = 0.6915. \tag{4}$$

Ce résultat pose des questions. La puissance du test est légèrement meilleure (69.15%) que ce que l'ingénieur aurait obtenu en tirant une pièce au hasard, soit 50% de chance de choisir la bonne hypothèse.

Ceci nous amène à réfléchir à ce qu'il faudrait faire lorsqu'on élabore un test :

1. **Minimiser la probabilité de commettre une erreur de type I**, c'est à dire minimiser $\alpha = P(Z \in RH_0|H_0)$. Classiquement, un niveau de significativité $\alpha \leq 0.1$ est requit.

2. **Maximiser la puissance du test**, ce qui revient à minimiser

$$\beta = P(Z \in RH_1|H_1) = P(Z \in \overline{RH_0}|H_1).$$

Typiquement, on souhaiterait une puissance de test $\kappa \geq 0.8$ soit $\beta \leq 0.2$.

Le tableau 1 permet de synthétiser l'ensemble des probabilités calculées.

Zone de rejet \ Hyp. vraie	H_0	H_1
$\overline{RH}_0$ ($RH1$)	$1-\alpha$	β
RH_0 ($\overline{RH}_1$)	α	$\kappa = 1-\beta$

Table 1: Tableau de probabilités conditionnelles

Or, le plus souvent, on se contente juste de la première étape, sans se soucier de valider la décision prise en s'intéressant à la puissance du test. Un autre problème apparaît également lorsque l'hypothèse alternative s'énonce sous la forme d'un intervalle. Ceci veut dire qu'il y a une infinité de valeur pouvant être prise sous H_1, ce qui implique que la valeur de β n'est pas unique. Nous allons construire dans ce cas une généralisation de la puissance d'un test.

5 Fonction de puissance

Reprenons l'exemple précédent. On suppose que l'on souhaite tester l'hypothèse $H_0 : \mu = 170\ mm$ contre l'hypothèse alternative $H_1 : \mu > 170\ mm$. Après avoir fixé une valeur raisonnable pour la probabilité de faire une erreur de type I, disons $\alpha = 0.05$, on prend un échantillon aléatoire $\{X_1, \cdots, X_n\}$ de taille $n = 25$ et on en calcule la moyenne empirique $\overline{X}$. Quelle serait la puissance de ce test d'hypothèse si la vraie moyenne de la population était $\mu = 175\ mm$ sous H_1? Et $\mu = 180\ mm$? Qu'en serait-il pour toute autre valeur de $\mu > 170\ mm$?

On voit bien dans ce cas la nécessité de quantifier l'évolution de la puissance $\kappa = 1 - \beta$ en fonction de μ_1 (Fig. 3). Nous constatons que la puissance de test dépend de cette valeur populationnelle μ_1 que l'on fixe sous H_1 et que lorsque cette valeur s'éloigne de $H_0 : \mu = \mu_0$, la puissance du test augmente. Mais cette fonction dépend également du risque α de rejeter H_0 à tort. La fonction puissance est donc donnée par :

$$K(\mu_1; \alpha) = P(\overline{X} \geq \overline{x}_{1-\alpha}|\mu = \mu_1) = 1 - F_{(\mu_1, \sigma^2/n)}(\overline{x}_{1-\alpha})$$

où $\overline{x}_{1-\alpha}$ est le quantile (valeur seuil) au risque α de rejet de H_0 et $F_{(\mu_1, \sigma^2/n)}$ est la fonction de répartition de la loi normale de moyenne $\mu = \mu_1$ et de variance σ^2/n. On peut donc tracer cette fonction pour différentes valeurs de μ_1 mais également lorsque l'on change le risque α. Dans notre cas. la taille de l'échantillon est fixe à $n = 25$.

6 Taille de l'échantillon

Comme on le voit ci-dessus, une autre manière de contrôler la puissance du test consiste à modifier la taille de l'échantillon. En effet, on sait que $\overline{X} \rightsquigarrow \mathcal{N}(\mu, \frac{\sigma^2}{n})$ en vertu du théorème central limite. Il peut donc être intéressant dans certaines études d'augmenter la puissance du test en imposant une valeur de

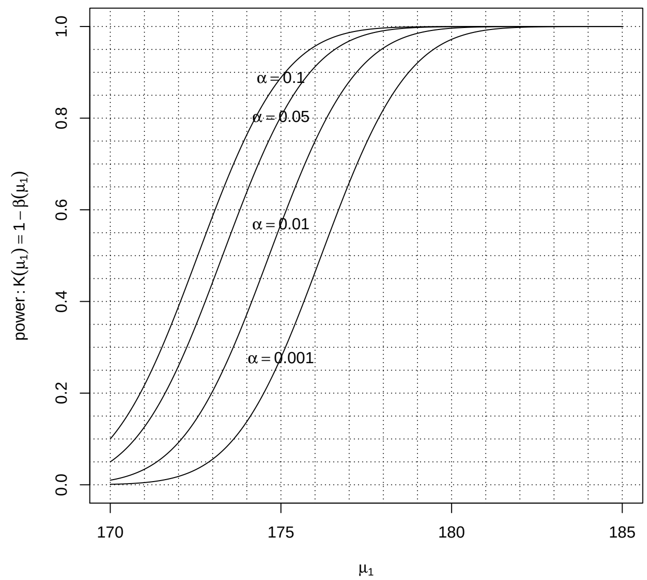

Figure 3: Tracé de la fonction de puissance. Cette fonction dépend de μ_1 mais également de α. Pour chaque valeur de μ_1 on peut lire la probabilité de rejeter l'hypothèse nulle alors que l'alternative est la bonne.

κ et en cherchant l'effectif n de l'échantillon qui permette d'atteindre cette puissance pour une hypothèse alternative $H_1 : \mu = \mu_1$.

On peut tracer

$$K(\mu_1; n) = P(\overline{X} \geq \overline{x}_{1-\alpha} | \mu = \mu_1) = 1 - F_{(\mu_1, \sigma^2/n)}(\overline{x}_{1-\alpha})$$

pour une valeur fixée de α mais en augmentant le nombre d'observations n, ce qui aura pour effet de resserrer la fonction de répartition $F_{(\mu_1,\sigma^2/n)}$ autour de μ_1 (Fig. 4). L'examen de la figure 4 montre qu'un échantillon de taille $n = 36$ est nécessaire si l'on veut atteindre une valeur de puissance $\kappa = 0.9$, sous l'hypothèse alternative $H_1 : \mu_1 = 175\ mm$.

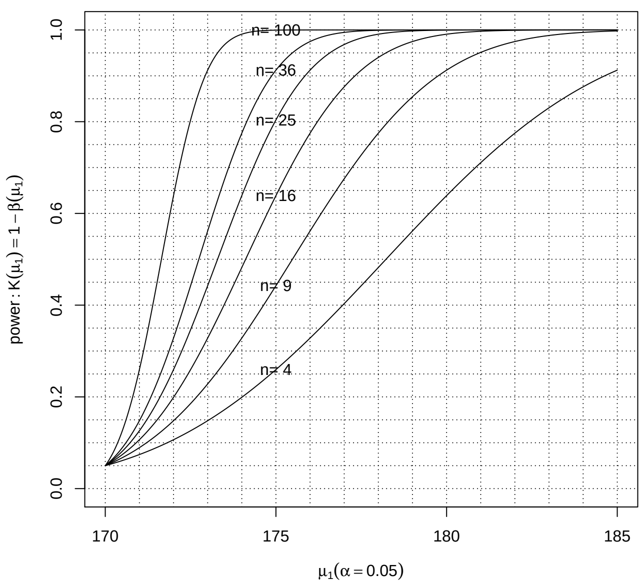

Figure 4: Tracé de la fonction de puissance pour $\alpha = 0.05$. Cette fonction dépend de μ_1 mais également de n. Pour une valeur μ_1 fixée, on peut lire la puissance associée à diverses valeurs de l'effectif n.