

Licence mention Mathématiques - Semestre 3
Statistique

Puissance d'un test

1. Introduction

Reprenons l'exemple des rongeurs vu dans un précédent chapitre.

Sur un échantillon de $n = 400$ naissances, on a observé 206 mâles, soit une fréquence de mâles de $f = \frac{206}{400} = 0.515$.

On se demande s'il y a autant de mâles que de femelles dans la population ; autrement dit si la proportion de mâles dans la population est $p = 0.5$.

On peut effectuer le test statistique de $H_0 : p = p_0$ contre $H_1 : p \neq p_0$, avec $p_0 = 0.5$.

Si H_0 est vraie, on a $p = p_0 = 0.5$, $np_0 = 200 \geq 10$ et $n(1 - p_0) = 200 \geq 10$ et $U = \frac{f - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}}$ suit

approximativement la loi normale $\mathcal{N}(0; 1)$.

On détermine comme le réel u_α tel que $P(-u_\alpha < U < u_\alpha) = 1 - \alpha$ grâce à la table 2. Pour $\alpha = 0.05$, on a $u_\alpha = 1.96$.

On calcule $u = \frac{f - p_0}{\sqrt{\frac{p_0(1 - p_0)}{n}}} = \frac{0.515 - 0.5}{\sqrt{\frac{0.5(1 - 0.5)}{400}}} = 0.6$.

Comme $u \in]-u_\alpha, u_\alpha[$, alors on ne peut rejeter H_0 : il est donc possible que $p = 0.5$.

Remarquons qu'avec $p_0 = 0.51$, on a $u = \frac{0.515 - 0.51}{\sqrt{\frac{0.51(1 - 0.51)}{400}}} \simeq 0.2$ et que la décision est la même.

Ainsi, de nombreuses valeurs de p_0 conduisent à la même décision. C'est pourquoi on s'exprime avec beaucoup de précautions oratoires et on dit « on ne rejette pas l'hypothèse nulle H_0 ». On n'affirme donc pas « l'hypothèse nulle est vraie ». On retient que les tests ne sont pas faits pour « démontrer » H_0 , mais pour la « rejeter ».

Quatre situations sont possibles, résumées dans le tableau suivant :

Décision du test \ Réalité	H_0 est vraie	H_0 est fausse
Non rejet de H_0	Pas d'erreur	Erreur de deuxième espèce
Rejet de H_0	Erreur de première espèce	Pas d'erreur

Cas où H_0 est vraie.

Il est clair que $\alpha = P^{H_0}(U \notin]-u_\alpha, u_\alpha[)$. C'est donc la probabilité de rejeter H_0 alors que H_0 est vraie, appelée probabilité d'erreur de première espèce.

De façon complémentaire, $1 - \alpha = P^{H_0}(U \in]-u_\alpha, u_\alpha[)$. C'est donc la probabilité de ne pas rejeter H_0 alors que H_0 est vraie, donc la probabilité de ne pas commettre d'erreur.

Cas où H_0 est fausse.

On s'intéresse la probabilité β de ne pas rejeter H_0 , et donc de commettre une erreur puisque H_0 est fausse. On parle de probabilité d'erreur de deuxième espèce.

A l'inverse, on s'intéresse à la probabilité de rejeter H_0 , et donc de ne pas se tromper puisque H_0 est fausse. Cette probabilité s'appelle la *puissance du test* et est égale à $1 - \beta$.

Dans ce cas, la difficulté est que l'on ne connaît pas la valeur de $p = p_1 \neq p_0$. Pour s'en rendre compte, cherchons à déterminer β .

On doit donc calculer $\beta = P^{H_1} \left(\frac{F - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \in]-u_\alpha, u_\alpha[\right)$, avec maintenant $U = \frac{F - p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}}$ qui suit approximativement la loi normale $\mathcal{N}(0; 1)$.

$$\begin{aligned} \text{On a } \beta &= P^{H_1} \left(-u_\alpha < \frac{F - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} < u_\alpha \right) \\ &= P^{H_1} \left(p_0 - u_\alpha \sqrt{\frac{p_0(1-p_0)}{n}} < F < p_0 + u_\alpha \sqrt{\frac{p_0(1-p_0)}{n}} \right) \\ \beta &= P^{H_1} \left(\frac{p_0 - u_\alpha \sqrt{\frac{p_0(1-p_0)}{n}} - p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}} < \frac{F - p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}} < \frac{p_0 + u_\alpha \sqrt{\frac{p_0(1-p_0)}{n}} - p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}} \right). \end{aligned}$$

On voit que cette probabilité β dépend de p_0 et p_1 (en particulier de $p_0 - p_1$ au numérateur) mais aussi de α (via u_α), et aussi de n .

Commençons par fixer $\alpha = 0.05$, $n = 400$ et faisons varier p_1 .

Pour $p_1 = 0.55$, on a $\beta \simeq 0.4839$.

Instructions R :

`p0=0.5`

`n=400`

`alpha=0.05`

`u_alpha=qnorm(1-alpha/2)`

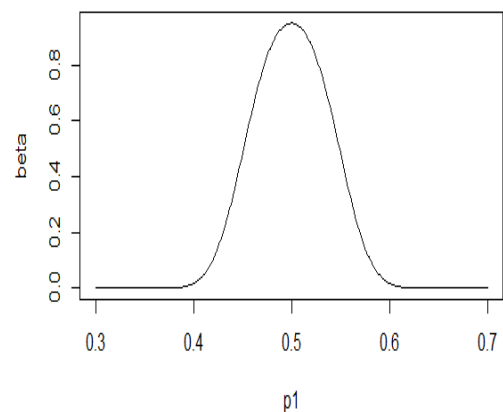
`p1=seq(0.3,0.7,0.001)`

`gauche=(p0-p1-u_alpha*sqrt(p0*(1-p0)/n))/sqrt(p1*(1-p1)/n)`

`droite=(p0-p1+u_alpha*sqrt(p0*(1-p0)/n))/sqrt(p1*(1-p1)/n)`

`beta=pnorm(droite)-pnorm(gauche)`

`plot(p1,beta,"l")`

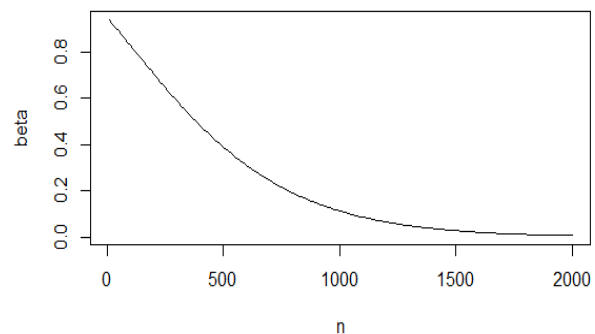
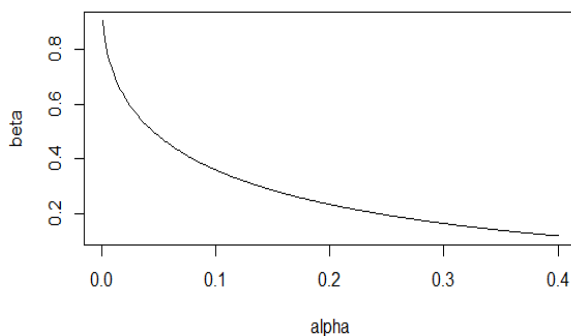


Poursuivons en fixant $p_1 = 0.55$, $n = 400$ et faisons varier α .

Pour $\alpha = 0.05$, on a $\beta \simeq 0.4839$.

Finissons en fixant $p_1 = 0.55$, $\alpha = 0.05$ et faisons varier n .

Pour $n = 400$, on a $\beta \simeq 0.4839$.



Sur cet exemple, on constate que β décroît lorsque la différence $p_0 - p_1$ augmente (p_1 s'éloigne de p_0), lorsque α augmente et lorsque n augmente.

A l'inverse, la puissance $1 - \beta$ du test s'accroît dans les mêmes conditions.

Dans l'exemple précédent, on a calculé la statistique de test F puis $U = \frac{F - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$.

Si $|U| < u_\alpha$, on décide de conserver l'hypothèse nulle H_0 .

Si $|U| \geq u_\alpha$, on décide de rejeter l'hypothèse nulle H_0 et on dit que le résultat est *statistiquement significatif* au seuil α .

Si nous sommes effectivement dans le cadre de l'hypothèse nulle H_0 , nous savons que nous risquons de nous tromper dans $\alpha \times 100$ % des cas, c'est le risque que nous avons pris.

Mais que se passe-t-il si nous sommes effectivement dans le cadre de l'hypothèse alternative H_1 ? Quel risque prenons-nous ?

Il faut choisir dans quelle mesure on s'écarte de l'hypothèse nulle, c'est ce qu'on appelle la *taille d'effet*.

C'est une décision qui se prend à partir de considérations scientifiques. Le statisticien ne peut ici se substituer au praticien. Il lui demande en particulier à partir de quelle taille un effet constitue une différence scientifiquement significative.

On supposera qu'il nous répond qu'en dessous de 0.44 ou au dessus de 0.56, la différence peut être considéré comme importante par rapport à $p_0 = 0.5$.

Ce qui nous intéresse est l'écart à l'hypothèse nulle, ici 0.06. Cela nous ramène à une écriture où l'hypothèse nulle est équivalente à ce que la taille d'effet soit nulle aussi.

Définitions

Une taille d'effet d mesure combien on s'écarte de l'hypothèse nulle.

La puissance d'un test est la probabilité de rejeter l'hypothèse nulle à raison, c'est-à-dire lorsqu'on est en vérité dans le cadre de l'hypothèse alternative.

La puissance du test est donc le complément de l'erreur de deuxième espèce, on la note $1 - \beta$.

Remarque

On considère généralement que la puissance doit au moins être égale à 0.80 pour être satisfaisante.

Les types d'analyse de puissance

Une analyse de puissance peut prendre plusieurs faces :

- on connaît le niveau du test, la taille d'échantillon et la taille d'effet et on cherche à calculer la puissance. Ceci permet de voir si notre dispositif expérimental est bien calibré.
- on connaît la puissance voulue, le niveau du test et la taille d'effet à détecter. On cherche alors à calculer la taille d'échantillon nécessaire pour monter un dispositif expérimental efficace.
- une fois les données collectées (post-mortem donc), il est possible également de calculer la puissance pour certains effets, et il est aussi possible de calculer la taille de l'effet mesuré sur l'échantillon. On arrive ainsi parfois à des contradictions entre la significativité statistique et la significativité scientifique.

2. Etude des risques dans les tests d'hypothèses avec R

2.1. Etude du risque de première espèce α

a) Simuler $M = 500$ échantillons de taille $n = 20$ issus d'une population gaussienne de loi normale de moyenne $\mu = 4$ et d'écart-type $\sigma = 1.2$.

```
n <- 20
```

```
Nsimu <- 500
```

```
ech <- rnorm(n, mean=4, sd=1.2) # seulement un échantillon
```

```
simu <- replicate(Nsimu, rnorm(n, mean=4, sd=1.2), simplify=TRUE)
```

b) Sur chaque échantillon, effectuer au risque $\alpha = 0.05$ le test de Student de $H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$, avec $\mu_0 = 4$.

```
ttest <- function(x) t.test(x, mu=4)$p.value
```

```
res <- apply(simu, MARGIN=2, FUN=ttest)
```

c) Compter le nombre de rejets de H_0 . Quel résultat attendait-on ?

```
compt <- sum((res < 0.05))
```

```
compt
```

- d) Augmenter le nombre M de simulations.

```
Nsimu <- 5000
simu <- replicate(Nsimu, rnorm(n, mean=4, sd=1.2), simplify=TRUE)
res <- apply(simu, MARGIN=2, FUN=ttest)
proportion <- mean((res < 0.05))
proportion
```

- e) Reprendre cette procédure pour la valeur $\mu_0 = 1$ avec des échantillons de taille $n = 100$ provenant d'une loi de khi-deux à 1 degré de liberté.

```
n <- 100
Nsimu <- 5000
simu <- replicate(Nsimu, rchisq(n, df=1), simplify=TRUE)
ttest <- function(x) t.test(x, mu=1)$p.value
res.chi2 <- t(apply(simu, MARGIN=2, FUN=ttest))
proportion <- mean((res.chi2 < 0.05))
proportion
```

- f) Même travail que dans la question e) avec $n = 10$. Que remarque-t-on ? Comment l'expliquer ?

```
n <- 10
Nsimu <- 5000
simu <- replicate(Nsimu, rchisq(n, df=1), simplify=TRUE)
ttest <- function(x) t.test(x, mu=1)$p.value
res.chi2 <- t(apply(simu, MARGIN=2, FUN=ttest))
proportion <- mean((res.chi2 < 0.05))
proportion
```

g) En fait, pour chacun des M échantillons, on prend la décision de rejeter ou non H_0 . Notons D_i la variable aléatoire qui pour le i -ème échantillon prend la valeur 1 si on rejette H_0 et la valeur 0 sinon. Les variables D_i suivent toutes la même loi de Bernoulli de paramètre $p = P(D_i = 1) = P(\text{rejet de } H_0)$ et sont indépendantes. Ainsi, la variable $D = \sum_{i=1}^M D_i$ suit la loi binomiale $B(M; p)$ et compte le nombre de rejets de H_0 sur les M échantillons.

Si H_0 est vraie et que le test est bien construit, on devrait avoir $p = P^{H_0}(\text{rejet de } H_0) = \alpha$.

Déterminer un intervalle de confiance de p au niveau de confiance 95% et conclure.

```
IC.p <- t.test((res.chi2 < 0.05) * 1)$conf
IC.p
```

2.2. Etude de la puissance $1 - \beta$

- a) Simuler $M = 500$ échantillons de taille $n = 20$ issus d'une population gaussienne de loi normale de moyenne $\mu = 5$ et de variance d'écart-type $\sigma = 1.2$.

```
n <- 20
Nsimu <- 500
simu <- replicate(Nsimu, rnorm(n, mean=5, sd=1.2), simplify=TRUE)
```

- b) Sur chaque échantillon, effectuer au risque $\alpha = 0.05$ le test de Student de $H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$, avec $\mu_0 = 4$.

```
ttest <- function(x) t.test(x, mu=4)$p.value
res <- apply(simu, MARGIN=2, FUN=ttest)
```

- c) Compter le nombre de non-rejets de H_0 . Donner alors une estimation de la puissance de ce test pour la situation de H_1 où $\mu = 5$.

```
compt <- sum((res < 0.05))
puissance <- mean((res < 0.05))
```

- d) Augmenter la taille de l'échantillon à $n = 100$ et estimer la puissance de ce test. Que remarque-t-on ?

```
n <- 100
Nsimu <- 500
simu <- replicate(Nsimu, rnorm(n, mean=5, sd=1.2), simplify=TRUE)
res <- apply(simu, MARGIN=2, FUN=ttest)
puissance <- mean((res < 0.05))
puissance
```

e) Reprendre cette procédure pour le test de Student de $H_0 : \mu = \mu_0$ **contre** $H_1 : \mu \neq \mu_0$, avec $\mu_0 = 1$. avec des échantillons de taille $n = 100$ provenant d'une loi de khi-deux à 2 degrés de liberté.

```
n <- 100
Nsimu <- 500
simu <- replicate(Nsimu,rchisq(n,df=2),simplify=TRUE)
ttest <- function(x) t.test(x,mu=1)$p.value
res.chi2 <- t(apply(simu,MARGIN=2,FUN=ttest))
proportion <- mean((res.chi2 < 0.05))
proportion
```

f) Même travail que dans la question e) avec $n = 10$. Que remarque-t-on ? Comment l'expliquer ?

```
n <- 10
Nsimu <- 500
simu <- replicate(Nsimu,rchisq(n,df=2),simplify=TRUE)
ttest <- function(x) t.test(x,mu=1)$p.value
res.chi2 <- t(apply(simu,MARGIN=2,FUN=ttest))
proportion <- mean((res.chi2 < 0.05))
proportion
```

3. Cas de tests sur une moyenne

D'après Stéphane CHAMPELY, Université Lyon 1, France.

3.1. Rappels sur le test de Student à un échantillon

Nous sommes dans le cadre d'un échantillon issu d'une loi $N(\mu, \sigma)$.

On veut tester $H_0 : \mu = \mu_0$ **contre** $H_1 : \mu > \mu_0$.

Dans une étude concernant les souhaits d'équipements des usagers des piscines lyonnaises, était demandé sur une échelle de Likert à 5 niveaux (1, 2, 3, 4 et 5) si un sauna était désiré.

On veut voir si le niveau de souhait est supérieur à la position neutre $\mu_0 = 3$.

On va supposer que $n = 25$ individus ont accepté de répondre.

Le test repose sur le calcul de la statistique $T = \frac{\bar{X} - \mu}{\frac{S_c}{\sqrt{n}}}$ qui, sous l'hypothèse H_0 , suit la loi de Student à

$n - 1$ degrés de liberté.

On détermine t'_α tel que $P(T < t'_\alpha) = 1 - \alpha$, i.e. $t'_\alpha = t_{2\alpha}$, et on décide que :

- si $t < t'_\alpha$, alors on ne peut rejeter H_0 ;

- si $t \geq t'_\alpha$, alors on rejette H_0 avec une probabilité α de se tromper.

On a par exemple pour $n = 25$ et le risque conventionnel $\alpha = 0.05$, on a $t'_\alpha = 1.711$ et on rejette H_0 si $T > 1.711$.

Instruction R : `qt(0.05,df=25-1,lower.tail=FALSE)`

3.2. La taille d'effet

La taille d'effet est définie par $d = \frac{\mu - \mu_0}{\sigma}$.

On voit que pour l'hypothèse nulle H_0 , la taille d'effet correspondante est nulle, elle mesure donc bien l'éloignement à cette hypothèse. De plus, cette quantité est indépendante des unités de mesure (on exprime les écarts en unités de variabilité).

Comment choisir d ?

Il est préférable de le faire en partant de considérations scientifiques basées sur l'expérience du praticien, sur la base d'une étude pilote (de petite taille) ou bien en utilisant des données historiques ou de la littérature.

Il faut bien comprendre que la taille d'effet conditionne notre capacité à obtenir des résultats significatifs. Sa détermination, au moins approximative, est donc fondamentale.

Dans l'étude de souhait qui nous intéresse, on va supposer que nous allons prendre un écart-type probablement trop grand, qui correspond à la loi uniforme sur les cinq niveaux soit $\sigma = \sqrt{2}$

On va considérer comme effet intéressant que le niveau moyen de souhait soit de 3.5. Ceci conduit donc à une valeur de taille d'effet de $d = \frac{3.5 - 3}{\sqrt{2}} \simeq 0.35$.

L'autre méthode, proposée par Cohen, est d'utiliser des niveaux conventionnels de taille d'effet. Mais il faut bien comprendre, et il le souligne, qu'ils devraient être conventionnels au domaine d'étude. Il propose d'employer les trois niveaux suivants :

- $d = 0.2$ faible effet (qui correspond par exemple à la différence de taille entre des filles de 15 et 16 ans)
- $d = 0.5$ effet moyen (différence de taille entre filles de 14 et 18 ans ou différence de QI entre employés et managers...)
- $d = 0.8$ effet fort (différence de taille entre filles de 13 et 18 ans ou différence de QI entre lycéens et titulaires d'un Doctorat)

C'est donc un effet de faible à moyen que nous souhaitons étudier.

3.3. La puissance

On doit donc calculer la probabilité de la région de rejet de H_0 dans le cadre de l'hypothèse alternative H_1 où la taille d'effet est d .

$$1 - \beta = P^{H_1}(T \geq t'_\alpha)$$

On montre après quelques étapes que

$$1 - \beta = P \left(\sqrt{n-1} \frac{\left(\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} + \frac{\mu - \mu_0}{\frac{\sigma}{\sqrt{n}}} \right)}{\sqrt{\frac{(n-1)S_c^2}{\sigma^2}}} \geq t'_\alpha \right)$$

que l'on peut ré-écrire

$$1 - \beta = P \left(\sqrt{df} \frac{(U + "ncp")}{\sqrt{Y^2}} \geq t'_\alpha \right)$$

$$\text{avec } df = n - 1, U = \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}}, "ncp" = \sqrt{n} \frac{\mu - \mu_0}{\sigma} \text{ et } Y^2 = \frac{(n-1)S_c^2}{\sigma^2}.$$

Or la loi de $\sqrt{df} \frac{(U + "ncp")}{\sqrt{Y^2}}$ est une loi qui est connue et tabulée. C'est la loi de Student décentrée avec

$df = n - 1$ degrés de liberté et comme paramètre de non-centralité " ncp " = $\sqrt{n} \frac{\mu - \mu_0}{\sigma}$

Dans R, les fonctions `pt` ou `qt` avec les options `ncp` et `df` permettent de réaliser ces calculs.

Nous pouvons aussi utiliser la fonction `pwr.t.test` qui permet d'obtenir directement le résultat avec l'option `type="one.sample"`.

Pour notre exemple d'étude de souhaits basé sur le test unilatéral avec le niveau de signification conventionnel $\alpha = 0.05$, la taille d'échantillon $n = 25$ et la taille d'effet $d = 0.35$, on obtient $1 - \beta = 0.52$, ce qui n'est pas satisfaisant.

```
power.t.test(n=25,d=0.35,sig.level=0.05,type="one.sample", alternative="one.sided")
```

Remarque

On généralise facilement le cas de test unilatéral précédent à une situation bilatérale en calculant la puissance comme

$$1 - \beta = P^{H_1}(|T| \geq t_\alpha) = P^{H_1}(T \leq -t_\alpha) + P^{H_1}(T \geq t_\alpha)$$

```
power.t.test(n=25,d=0.35,sig.level=0.05,type="one.sample",alternative="two.sided")
```

3.4. La taille d'échantillon

Si l'on souhaite calculer la taille d'échantillon pour une puissance donnée, il "suffit" de résoudre l'équation $1 - \beta = P^{H_1}(T \geq t'_\alpha)$ où l'inconnue est n .

Cela peut être réalisé numériquement avec la fonction `power.t.test`.

En ce qui concerne notre étude de souhaits avec une taille d'effet de $d = 0.35$, un niveau de signification conventionnel pour un test unilatéral, si l'on souhaite obtenir une puissance de $1 - \beta = 0.80$, il faut $n = 52$ unités statistiques.

```
power.t.test(power=0.8,d=0.35,sig.level=0.05,type="one.sample",alternative="one.sided")
```