Chapitre 6 : Tests d'hypothèses

6.1 Construction d'un test et règle de décision 2
6.2 Puissance d'un test 3
6.3 Quelques tests d'hypothèses....... 4
6.3.1 Test sur la moyenne d'une distribution normale de variance connue 4
6.3.2 Test sur la moyenne d'une distribution normale de variance inconnue 5
6.3.3 Test sur la variance d'une distribution normale de moyenne connue 6
6.3.4 Test sur la variance d'une distribution normale de moyenne inconnue 7
6.3.5 Test d'égalité de moyennes de deux populations normales de variances connues 7
6.3.6 Test d'égalité de moyennes de deux populations normales de variances inconnues 7
6.3.6.1 Les variances sont supposées égales 7
6.3.6.2 Les variances sont inégales 8
6.3.7 Test d'égalité de variances de deux populations normales de moyennes connues 9
6.3.8 Test d'égalité de variances de deux populations normales de moyennes inconnues 9
6.3.9 Test sur des données par paires (appariées) 9
6.4 Exemple de puissance d'un test 10

Les tests d'hypothèses servent à déterminer si une série d'observations (formant l'échantillon) permettent d'invalider ou non une hypothèse que l'on formule sur la population.

Exemple : Un ciment est fabriqué pour présenter une résistance de 30 MPa (valeur de design). Un échantillon de 5 éprouvettes a fourni les valeurs suivantes : 30.1, 29.5, 29.6, 28.4, 28.9. Peut-on rejeter l'hypothèse que le ciment dans son ensemble a une résistance de 30 MPa sur la seule foi de ces 5 échantillons?

Exemple : On installe un feu de circulation à une intersection jugée dangereuse. Les trois mois précédant l'installation du feu, 4 accidents majeurs avaient été enregistrés. Les trois mois suivants, seulement 2 accidents majeurs sont survenus. Peut-on affirmer que le feu de circulation a permis une réduction significative de la moyenne du nombre d'accidents?

Les tests d'hypothèses peuvent porter sur divers aspects. Les plus fréquemment utilisés sont :

- la moyenne d'une population est-elle différente d'une moyenne connue?
- deux populations ont-elles même moyenne?
- deux variables sont-elles indépendantes?
- la corrélation entre deux variables est-elle significative?
- la variance d'une population est-elle différente d'une variance connue?
- deux populations ont-elles même variance?
- une loi de distribution donnée s'applique-t-elle à une population donnée?
- deux populations ont-elles même distribution?

Note : Remarquez comment les énoncés précédents portent tous sur la population. Toutefois, c'est l'échantillon, sensé représenter la population, qui permettra de répondre à ces questions.

Note : Un test compare la valeur d'une statistique calculée sur un échantillon à la valeur théorique découlant de la distribution de cette statistique obtenue lorsqu'on considère l'hypothèse formulée vraie. Si la valeur de la statistique ne présente rien de particulier par rapport à la distribution théorique, on accepte l'hypothèse. Lorsque la valeur de la statistique apparaît clairement atypique, on se doit de rejeter l'hypothèse formulée. Cependant cette valeur atypique possède une certaine probabilité d'être observée même lorsque l'hypothèse est vraie. Donc un test statistique ne permet jamais d'être certain qu'une hypothèse est vraie ou fausse. Tout au plus s'agit-il d'une règle opérationnelle permettant de documenter statistiquement les décisions prises.

6.1 Construction d'un test et règle de décision

Reprenons l'exemple du ciment et des cinq éprouvettes évoqué précédemment. Supposons de plus que l'on sait que la distribution de la résistance des éprouvettes est N(30,2) pour un ciment conçu pour présenter une résistance de 30 MPa. On veut vérifier si la moyenne observée dans l'échantillon (1/5*(30.1+29.5+29.6+28.4+28.9)=29.3) permet d'affirmer que la valeur de conception (30 MPa) n'est pas respectée. On formule l'hypothèse :

H_0 $\mu = 30$ vs H_1 $\mu < 30$.

Note : Le test est dit *unilatéral* ici car on ne s'intéresse qu'au cas où la moyenne ne rencontre pas la valeur de conception (le cas $\mu > 30$ constitue en fait un « bonus » dont il n'y a pas lieu de s'inquiéter). Lorsque l'hypothèse alternative est de la forme $\mu \neq 30$ on dit alors que le test est *bilatéral*).

Note : Seule la distribution sous H_0 ici est entièrement spécifiée (on parle alors d'hypothèse *simple*). L'hypothèse *alternative* H_1 est dite hypothèse *composée* car elle ne spécifie pas une distribution unique lorsque H_1 vraie. C'est le cas le plus fréquemment rencontré en pratique bien que parfois on puisse aussi avoir une alternative simple.

On sait que la moyenne d'un échantillon dont les observations sont indépendantes et issues d'une distribution normale est une v.a. et : $\overline{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$. On va calculer $P(\overline{X} < 29.3)$. Si cette probabilité est inférieure à un seuil raisonnable que l'on s'est fixé (souvent $\alpha = 0.05$) alors on rejettera l'hypothèse H_0 car la valeur $\overline{x} = 29.3$ apparaîtra comme atypique pour la distribution de $\overline{X}$.
On calcule :

$$P(\overline{X} < 29.3) = P\left(\frac{(\overline{X} - \mu)}{\sigma / \sqrt{n}} < \frac{29.3 - \mu}{\sigma / \sqrt{n}}\right) = P\left(\frac{\overline{X} - 30}{\sqrt{2/5}} < \frac{29.3 - 30}{\sqrt{2/5}}\right) = P(N(0,1) < -1.11) = 0.134$$

Comme 0.134>0.05, on ne peut rejeter H_0, i.e. on n'a pas d'évidences suffisantes pour affirmer que la valeur de conception n'est pas respectée.

Note : Le nombre d'observations joue un rôle clé dans cette décision. Si l'échantillon avait par exemple contenu 20 observations au lieu de 5, toujours présentant la même moyenne, on aurait calculé P=0.0136 et on serait arrivé à la décision contraire.

Un test statistique consiste donc en :

- Identifier l'hypothèse simple H_0 et l'hypothèse alternative H_1;
- trouver une statistique liée à H_0 dont on connaît la loi de distribution;
- identifier le niveau de signification du test, α ; habituellement, α est choisi égal à 0.1, 0.05 ou 0.01 mais ceci peut varier selon l'application;
- évaluer la valeur de la statistique sur l'échantillon;
- calculer la probabilité[1] que l'on observe une valeur aussi ou plus atypique que la valeur obtenue sur l'échantillon;
- rejeter l'hypothèse H_0 si la probabilité obtenue est inférieure à α .

6.2 Puissance d'un test

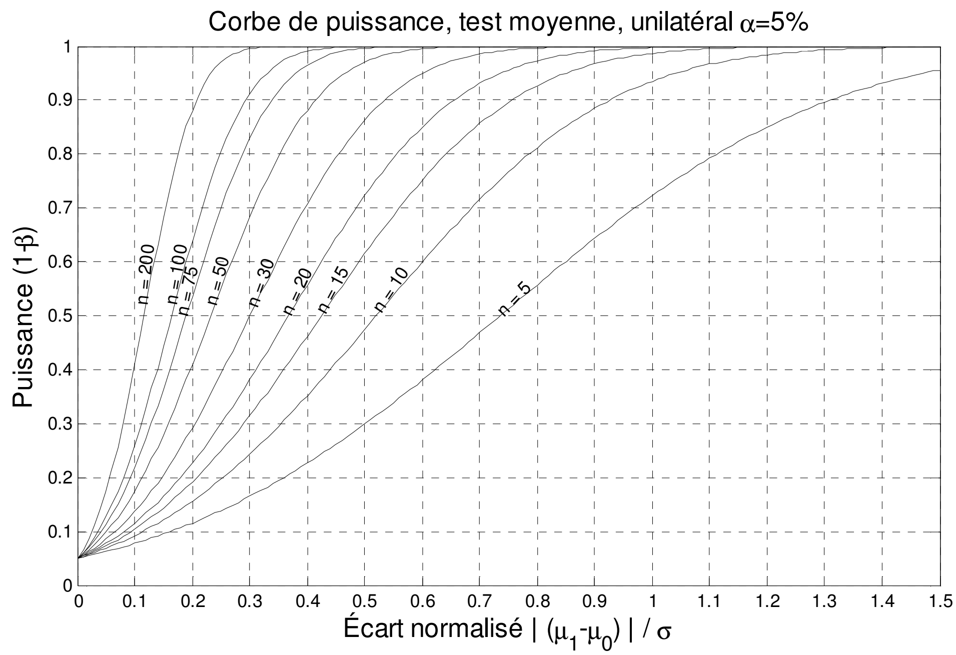
Supposons que l'alternative H_1 soit aussi une hypothèse simple. On peut alors évaluer
P(Accepter $H_0|H_1$ est vraie). Cette probabilité est appelée l'erreur de seconde espèce (β ; α étant l'erreur de première espèce). La probabilité complémentaire P(Rejeter $H_0|H_1$ est vraie)=1- β est appelée la puissance du test. On cherche normalement à avoir un test le plus puissant possible car ceci assure que l'on pourra éventuellement détecter de subtiles variations de moyenne, de variance, etc. La puissance d'un test donné dépend de deux facteurs : le nombre d'observations et l'écart entre les deux hypothèses simples H_0 et H_1. Plus on a d'observations, pour un même écart, plus le test est puissant. Plus l'écart est grand entre H_0 et H_1, pour un même nombre d'observations, plus il est facile de détecter cette différence et donc plus le test est puissant.

Cas de H_1 hypothèse composée : Dans ce cas, on considère plusieurs valeurs possibles de l'alternative et pour chacune on calcule la puissance correspondante. On trace ensuite la courbe de puissance du test en fonction de la valeur sous l'alternative. Lorsque la valeur considérée sous H_1 tend vers celle sous H_0, alors la puissance du test tend vers le niveau α du test (i.e. le test est très peu puissant).

Courbe caractéristique : La courbe caractéristique est la courbe $\beta(H_1)$ alors que la courbe de puissance est la courbe complémentaire, 1- $\beta(H_1)$.

Exemple : La figure suivante montre la puissance du test unilatéral sur la moyenne (avec variance connue) au niveau α = 5%, H_0 $\mu = \mu_0$ vs H_1 $\mu > \mu_0$ en fonction de l'écart normalisé entre les moyennes sous H_0 et H_1 pour différentes valeurs de n. Ainsi, 200 observations permettent de détecter un écart normalisé de 0.13 environ avec une puissance de 80% alors que 5 observations nécessitent un écart de 1.12 à cette même puissance. Pour un même écart normalisé de 0.3, la puissance du test n'est que de 17% environ avec 5 observations, 50% avec 30 observations et presque 100% avec 200 observations. Remarquez comme la puissance augmente rapidement aux faibles « n » puis plus lentement. Remarquez également comme la puissance est faible pour les petits écarts normalisés. Même avec 200 observations, la puissance n'est que de 40% pour détecter un écart de 0.1.

[1] Alternativement, on peut aussi calculer la valeur critique dans la distribution de la statistique correspondant au niveau α désiré et on rejette H_0 si la valeur de la statistique excède ce seuil.



Note importante : Il existe un lien <u>très étroit</u> entre un test d'hypothèse simple sur la valeur d'un paramètre et l'estimation par intervalle de confiance de ce même paramètre. Si la valeur du paramètre testée sous H_0 est incluse dans l'intervalle de confiance construit autour de l'estimateur alors on ne peut rejeter H_0 pour un test de niveau de confiance α correspondant à un intervalle de niveau 1-α . De même, si la valeur sous H_0 se trouve à l'extérieur de l'intervalle de confiance, alors on rejette H_0.

6.3 Quelques tests d'hypothèses

6.3.1 Test sur la moyenne d'une distribution normale de variance connue

Test : $H_0: \mu = \mu_0$ vs $H_1: \mu < \mu_0$, ou $H_1: \mu > \mu_0$, ou $H_1: \mu \neq \mu_0$

Statistique : $Z = \dfrac{(\bar{X} - \mu_0)}{\sigma / \sqrt{n}} \sim N(0,1)$

Règle de décision :

$H_1: \mu < \mu_0$,

$$\text{Si } P\left(Z < \frac{(\bar{x} - \mu_0)}{\sigma / \sqrt{n}}\right) < \alpha \text{, rejeter } H_0.$$

$H_1: \mu > \mu_0$

$$\text{Si } P\left(Z > \frac{(\bar{x} - \mu_0)}{\sigma / \sqrt{n}}\right) < \alpha \text{, rejeter } H_0.$$

$H_1 : \mu \neq \mu_0$

$$\text{Si } P\left(Z < \frac{(\bar{x} - \mu_0)}{\sigma / \sqrt{n}}\right) < \alpha/2\text{, ou } P\left(Z > \frac{(\bar{x} - \mu_0)}{\sigma / \sqrt{n}}\right) < \alpha/2\text{, rejeter } H_0.$$

Note : En raison du théorème central limite, lorsque « n » est suffisamment grand, le test demeure approximativement valide même si la population n'est pas normale au départ.

Puissance du test : P(Rejeter H_0 | H_1 vraie)

Soit $Z \sim N(0,1)$ et z_α tel que $P(Z > z_\alpha) = \alpha$ et $-z_\alpha$ tel que $P(Z < -z_\alpha) = \alpha$

$H_1 : \mu < \mu_0$

$$P\left(\frac{(\bar{x} - \mu_0)}{\sigma / \sqrt{n}} < -z_\alpha \mid \mu = \mu_1\right) = P\left(\frac{(\bar{x} - \mu_0)}{\sigma / \sqrt{n}} + \frac{\mu_0}{\sigma / \sqrt{n}} - \frac{\mu_1}{\sigma / \sqrt{n}} < -z_\alpha + \frac{\mu_0}{\sigma / \sqrt{n}} - \frac{\mu_1}{\sigma / \sqrt{n}}\right)$$
$$= P\left(Z < -z_\alpha + \frac{\mu_0 - \mu_1}{\sigma / \sqrt{n}}\right)$$

$H_1 : \mu > \mu_0$

$$P\left(\frac{(\bar{x} - \mu_0)}{\sigma / \sqrt{n}} > z_\alpha \mid \mu = \mu_1\right) = P\left(\frac{(\bar{x} - \mu_0)}{\sigma / \sqrt{n}} + \frac{\mu_0}{\sigma / \sqrt{n}} - \frac{\mu_1}{\sigma / \sqrt{n}} > z_\alpha + \frac{\mu_0}{\sigma / \sqrt{n}} - \frac{\mu_1}{\sigma / \sqrt{n}}\right)$$
$$= P\left(Z > Z_\alpha + \frac{\mu_0 - \mu_1}{\sigma / \sqrt{n}}\right)$$

Par la symétrie de la loi normale, les deux derniers résultats peuvent s'écrire comme :

$$(1 - \beta) = P\left(Z < -z_\alpha + \frac{|\mu_0 - \mu_1|}{\sigma / \sqrt{n}}\right)$$

$H_1 : \mu \neq \mu_0$

$$(1 - \beta) = P\left(Z < -z_{\alpha/2} + \frac{\mu_0 - \mu_1}{\sigma / \sqrt{n}}\right) + P\left(Z > z_{\alpha/2} + \frac{\mu_0 - \mu_1}{\sigma / \sqrt{n}}\right)$$

6.3.2 Test sur la moyenne d'une distribution normale de variance inconnue

Test : $H_0 : \mu = \mu_0$ vs $H_1 : \mu < \mu_0$, ou $H_1 : \mu > \mu_0$:, ou $H_1 : \mu \neq \mu_0$

Statistique : $t = \dfrac{(\bar{X} - \mu_0)}{s / \sqrt{n}} \sim t_{n-1}$

Règle de décision :

$H_1 : \mu \neq \mu_0$

$$\text{Si } P\left(t < \frac{(\bar{x}-\mu_0)}{s/\sqrt{n}}\right) < \alpha/2, \text{ ou } P\left(t > \frac{(\bar{x}-\mu_0)}{s/\sqrt{n}}\right) < \alpha/2, \text{ rejeter } H_0.$$

Les cas unilatéraux s'obtiennent de façon similaire.

Note : Lorsque « n » est suffisamment grand, $t_{n-1} \approx N(0,1)$, de plus, en raison du théorème central limite, le test demeure approximativement valide même si la population n'est pas normale au départ.

6.3.3 Test sur la variance d'une distribution normale de moyenne connue

Si la moyenne est connue, on estime la variance par :

$$\hat{\sigma}^2 = \frac{1}{n}\sum (X_i - \mu)^2$$

Clairement, si $X_i \sim N(\mu, \sigma^2) \Leftrightarrow Z_i = \frac{(X_i - \mu)}{\sigma} \sim N(0,1) \Leftrightarrow Z_i^2 \sim \chi_1^2 \Leftrightarrow \sum Z_i^2 \sim \chi_n^2$

Or $\sum Z_i^2 = \frac{n\hat{\sigma}^2}{\sigma^2}$

Test : $H_0 : \sigma = \sigma_0$ vs $H_1 : \sigma < \sigma_0$, ou H_1 $\sigma > \sigma_0$:, ou H_1 $\sigma \neq \sigma_0$

Statistique : $\frac{n\hat{\sigma}^2}{\sigma_0^2} \sim \chi_n^2$

Règle de décision :

H_1 : $\sigma \neq \sigma_0$

$$\text{Si } P\left(\chi_n^2 < \frac{n\hat{\sigma}^2}{\sigma_0^2}\right) < \alpha/2, \text{ ou } P\left(\chi_n^2 > \frac{n\hat{\sigma}^2}{\sigma_0^2}\right) < \alpha/2, \text{ rejeter } H_0.$$

Les cas unilatéraux s'obtiennent de façon similaire.

Note : Lorsque « n » est grand, le test demeure approximativement valide même si la distribution de X n'est pas normale.

6.3.4 Test sur la variance d'une distribution normale de moyenne inconnue

Dans ce cas, on estime la variance par :

$$S^2 = \frac{1}{n-1}\sum\left(X_i - \overline{X}\right)^2$$

On a :

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$$

Le test découle directement (voir section précédente).

6.3.5 Test d'égalité de moyennes de deux populations normales de variances connues

Soit $X \sim N(\mu_X, \sigma_X^2)$ et $Y \sim N(\mu_Y, \sigma_Y^2)$. Soit n_X et n_Y le nombre d'observations dans chaque population. Considérons la distribution de la différence des moyennes des échantillons. On a :

$$\overline{X} - \overline{Y} \sim N\left(\mu_X - \mu_Y, \frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}\right) \Leftrightarrow \frac{\left(\overline{X} - \overline{Y}\right) - \left(\mu_X - \mu_Y\right)}{\left(\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}\right)^{1/2}} \sim N(0,1)$$

Sous H$_0$ $\mu_X = \mu_Y$

$$\frac{\left(\overline{X} - \overline{Y}\right)}{\left(\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}\right)^{1/2}} \sim N(0,1)$$

Le test découle directement.

6.3.6 Test d'égalité de moyennes de deux populations normales de variances inconnues

6.3.6.1 Les variances sont supposées égales

Soit $X \sim N(\mu_X, \sigma^2)$ et $Y \sim N(\mu_Y, \sigma^2)$. Soit n_X et n_Y le nombre d'observations dans chaque population.

Comme la variance commune est inconnue, on va l'estimer en utilisant l'ensemble des observations par :

$$S_p^2 = \frac{(n_X - 1)S_X^2 + (n_Y - 1)S_Y^2}{n_X + n_Y - 2}$$

On peut montrer que cet estimateur est sans biais pour σ^2 et $S_p^2 \sim \chi^2_{n_X + n_Y - 2}$

Sous H$_0$ $\mu_X = \mu_Y$

$$\frac{\left(\overline{X}-\overline{Y}\right)}{S_p\left(\frac{1}{n_X}+\frac{1}{n_Y}\right)^{1/2}} \sim t_{n_X+n_Y-2}$$

Le test découle directement.

6.3.6.2 Les variances sont inégales

Dans ce cas, il n'y a pas de test exact. Il y a un test approximatif toutefois.

Sous H_0 $\mu_X = \mu_Y$

$$\frac{\left(\overline{X}-\overline{Y}\right)}{\left(\frac{S_X^2}{n_X}+\frac{S_Y^2}{n_Y}\right)^{1/2}} \approx t_v$$

avec

$$v = \frac{\left(\frac{S_X^2}{n_X}+\frac{S_Y^2}{n_Y}\right)^2}{\frac{\left(S_X^2/n_X\right)^2}{n_X+1}+\frac{\left(S_Y^2/n_Y\right)^2}{n_Y+1}} - 2$$

Le test découle directement.

Exemple : Lafarge produit du ciment à St-Constant à l'aide de deux circuits indépendants. Les teneurs moyennes et les variances en C_3S des deux circuits sont mesurées sur deux échantillons :

	n	$\overline{x}$	s^2
Circuit 1	30	122	32
Circuit 2	20	125	20

Les deux circuits présentent-ils une moyenne significativement différente ?

Sous H_0 : $\mu_X = \mu_Y$

vs H_1: $\mu_X \neq \mu_Y$

On calcule la statistique t_0 : $\frac{(122-125)}{(32/30+20/20)^{0.5}} = -2.08$

Les degrés de liberté sont : $v = \frac{(32/30+20/20)^2}{(32/30)^2/31+(20/20)^2/21} - 2 = 48.6 \rightarrow 48$

On lit $t_{48;.05/2} = -2.31$

Comme –2.08>-2.31, on ne peut rejeter H_0.

6.3.7 Test d'égalité de variances de deux populations normales de moyennes connues

On estime chaque variance par : $\hat{\sigma}^2 = \frac{1}{n}\sum (X_i - \mu)^2$

On a : $\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_n$

Sous H_0 : $\sigma_X^2 = \sigma_Y^2$, on a, utilisant la définition de la loi F :

$$\frac{\hat{\sigma}_X^2}{\hat{\sigma}_Y^2} \sim F_{n_X, n_Y}$$

Le test découle directement.

6.3.8 Test d'égalité de variances de deux populations normales de moyennes inconnues

On estime chaque variance par : $S^2 = \frac{1}{(n-1)}\sum (X_i - \overline{X})^2$

On a : $\frac{(n-1)S^2}{\sigma^2} \sim \chi^2_{n-1}$

Sous H_0 : $\sigma_X^2 = \sigma_Y^2$, on a, utilisant la définition de la loi F :

$$\frac{S_X^2}{S_Y^2} \sim F_{n_X - 1, n_Y - 1}$$

Le test découle directement.

6.3.9 Test sur des données par paires (appariées)

Imaginons « n » données provenant par paires. Ce pourrait être par exemple « n » bétons différents pour lesquels on prépare deux éprouvettes chacun et on fait subir à chacune des deux éprouvettes un test de résistance différent. On voudrait tester si les deux tests de résistance donnent les mêmes valeurs moyennes. Ce pourrait aussi être « n » échantillons de sols auxquels on fait subir un certain traitement. On calcule la moyenne avant et après traitement. La question se pose alors : le traitement a-t-il eu un effet significatif, i.e. les moyennes sont-elles significativement différentes ?

On pourrait penser utiliser le test d'égalité des moyennes pour deux populations. Toutefois ce serait une erreur car il s'agit des mêmes données. On ne peut sûrement pas les considérer comme indépendantes, ce qui est une condition essentielle pour pouvoir appliquer le test d'égalité des moyennes vu précédemment.

On va plutôt considérer la v.a $Y_i = X_{i1} - X_{i2}$ où les indices 1 et 2 distinguent les données de chaque paire, (ex. avant et après traitement). Sous hypothèse que Y est normale, on posera

H_0 : $\mu_Y = 0$

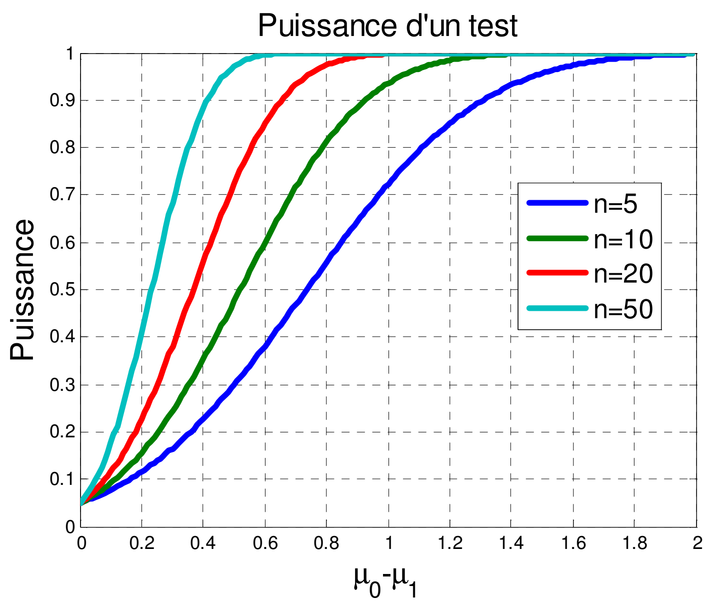
H_1 : $\mu_Y \neq 0$

La statistique $t = \dfrac{\bar{Y}}{S_Y / \sqrt{n}} \sim t_{n-1}$.

Le test découle directement.

6.4 Exemple de puissance d'un test

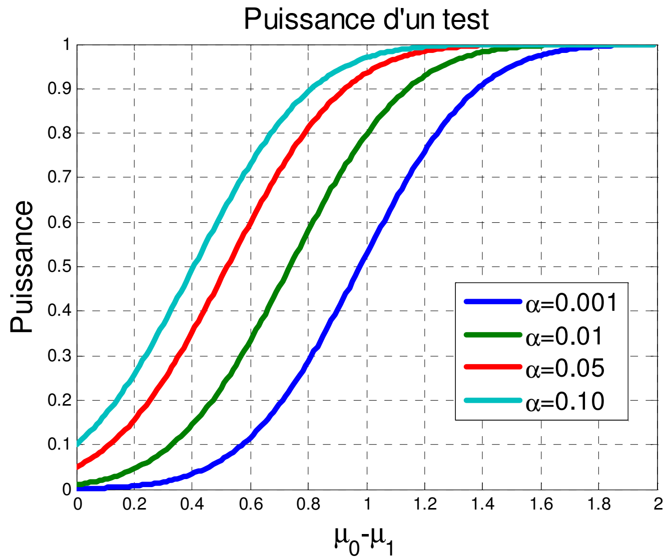
Considérons le test sur la moyenne d'une population normale lorsque la variance est connue. Sans perte de généralité, on pose $\sigma^2 = 1$. La figure suivante montre la puissance du test $H_0 : \mu = \mu_0$ vs $H_1 : \mu < \mu_0$ pour différentes valeurs de l'alternative et en fonction du nombre d'observations dans l'échantillon.



Note : Lorsque l'écart entre les hypothèses s'approche de 0, la puissance du test s'approche du niveau α du test (ici 0,05).

Note : Le graphe précédent permet de déterminer le nombre d'observations requis pour avoir une puissance suffisante en fonction de l'écart anticipé par rapport à μ_0 pour l'alternative. Ainsi, si l'on vise une puissance de 0,8 un échantillon de taille n=5 ne permet de détecter qu'un écart de 1,1 alors qu'un échantillon de 50 permet de détecter un écart aussi faible que 0,35 approximativement.

Note : Si l'on connaît la probabilité à priori que se réalise l'hypothèse H_0 (disons p_0) et l'alternative simple H_1 (disons p_1), alors on peut calculer la probabilité de prendre la bonne décision avec le test : $p_0(1-\alpha)+p_1(1-\beta)$. Plus α est grand, plus β est petit (et plus la puissance du test est grande).



Puissance du test unilatéral sur la moyenne (n=10) en fonction de α .