MAT 2777
Probabilités et statistique pour ingénieur.e.s

Chapitre 6
Les tests d'hypothèses

P. Boily (uOttawa)

Hiver 2023

Aperçu

Scénario – Allégations et suspicions

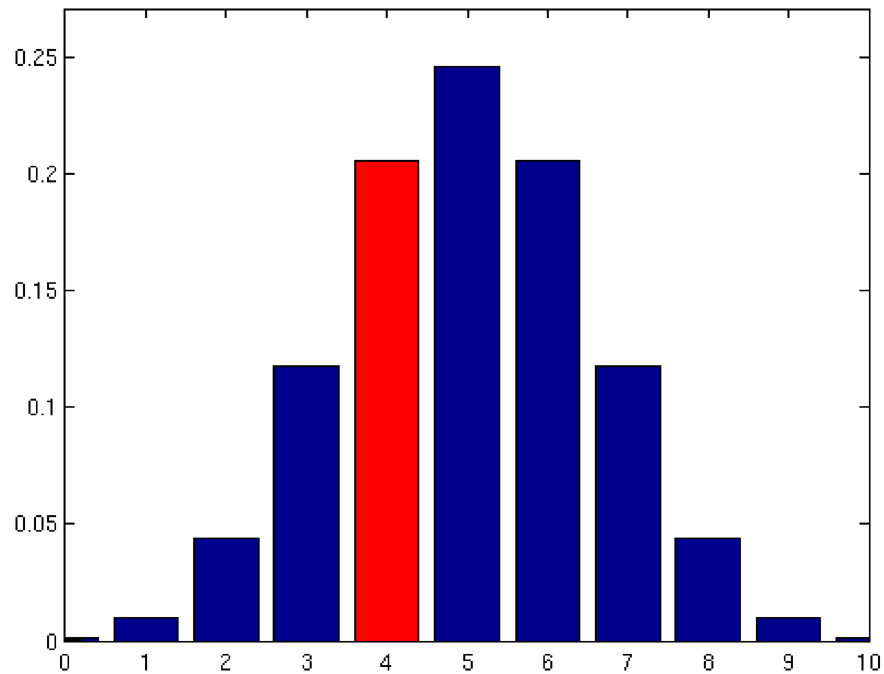
Un individu A affirme avoir une pièce de monnaie juste (sans biais), mais pour une raison quelconque, l'individu B se méfie de cette affirmation, estimant que la pièce est biaisée en faveur de pile (P).

B lance la pièce à 10 reprises; elle s'attend à un faible nombre de Faces, ce qu'elle a l'intention d'utiliser comme **évidence** contre l'affirmation de A (à savoir, que la pièce est juste). Soit $X = \#$ de Faces (F).

Supposons que $X = 4$. C'est moins que prévu pour une v.a. binomiale $X \sim \mathcal{B}(10, 0.5)$ puisque $\mathrm{E}[X] = 5$ dans ce cas ; les résultats seraient plutôt conformes à une pièce de monnaie pour laquelle $P(F) = 0.4$.

Est-ce le début de la fin pour l'allégation $P(F) = 0.5$?

Si la pièce est juste, alors $X \sim \mathcal{B}(10, 0.5)$; dans ce cas, $X = 4$ est quand même assez proche de $\mathrm{E}[X]$; en fait, $P(X = 4) = 0.205$ (par opposition à $P(X = 5) = 0.246$) ; l'événement $X = 4$ n'est pas **improbable**. Il semblerait donc ne pas y avoir d'évidence contre l'affirmation $P(F) = 0.5$.



La formulation de la phrase est **très** importante : "*Il semblerait donc ne pas y avoir d'évidence contre l'affirmation* $P(F) = 0.5$".

Nous n'avons pas rejeté l'allégation que $P(F) = 0.5$, mais cela **ne signifie pas nécéssairement que** $P(F) = 0.5$.

Le fait d'**accepter**, ou plutôt, de **ne pas rejeter** une affirmation est une déclaration **faible**.

Pour voir pourquoi, considérons un troisième individu qui affirme que la pièce est telle que $P(F) = 0.3$. Sous $X \sim \mathcal{B}(10.0.3)$, l'événement $X = 4$ est encore relativement probable, avec $P(X = 4) = 0.22$; nous **n'avons pas assez d'évidence pour rejeter** soit $P(F) = 0.5$, soit $P(F) = 0.3$.

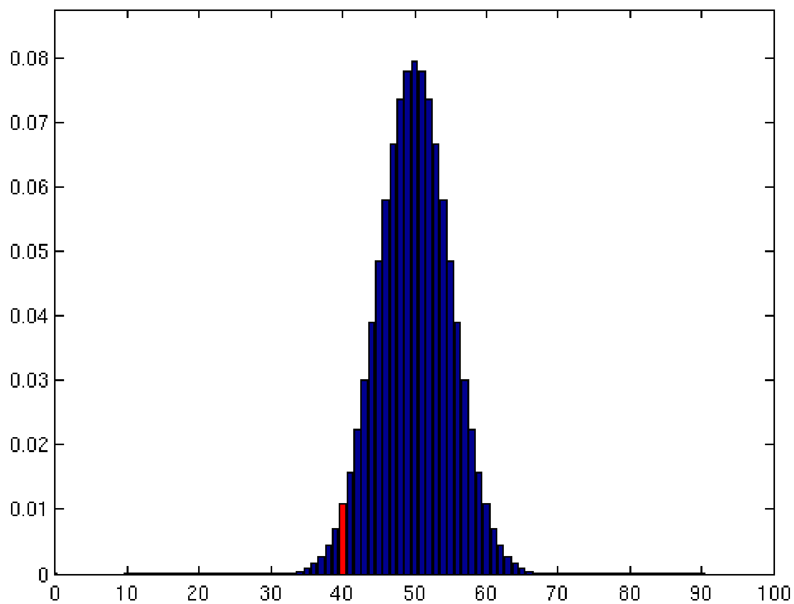
Cependant, **rejeter** une affirmation est une déclaration **forte** !

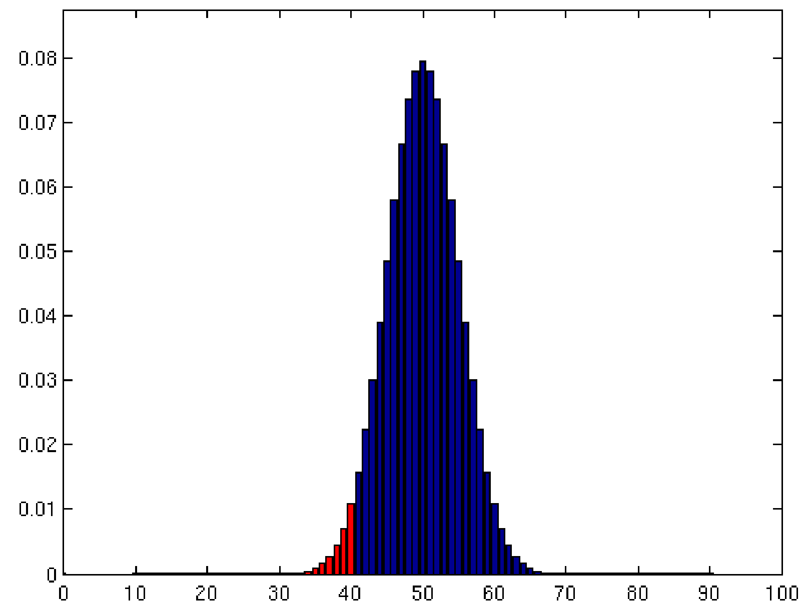
Supposons que B tire à pile ou face à 90 reprises supplémentaires, et qu'elle obtienne 36 faces, pour un total (Y) de 40 faces sur 100 lancers. Dans quelle mesure cela constitue-t-il une preuve contre l'affirmation ?

Supposons que $Y \sim \mathcal{B}(100, 0.5)$; nous avons $Y = 40$. C'est moins que prévu pour une v.a. binomiale puisque $\mathrm{E}[Y] = 50$; les résultats seraient à nouveau plutôt conformes à une pièce de monnaie pour laquelle $P(F) = 0.4$.

Cet événement ne se situe pas au **centre de la masse de probabilité** de la distribution de Y ; au contraire, il tombe dans la **queue** de la distribution (une zone de plus faible probabilité). Si $Y \sim \mathcal{B}(100, 0.5)$, alors $P(Y = 40) = 0.011$ (comparez avec la valeur précédente de 0.205). Ainsi, si la pièce est juste, l'événement $Y = 40$ est assez **peu probable**.

Une valeur dans la queue inférieure fournit une **preuve** contre l'affirmation: peut-on la **quantifier** ?





Les valeurs qui se trouvent "plus bas dans la queue de gauche" fournissent des preuves contre l'affirmation que $P(F) = 0.5$ (en faveur d'une pièce qui favorise F); nous utilisons la région de la queue sous l'observation : plus cette zone est **petite**, plus la preuve contre l'affirmation est **forte**.

Avec 4 F sur 10 lancers, l'évidence est la valeur$-p$ $P(X \leq 4)$, **en supposant que l'affirmation est véridique**, c-à-d lorsque $X \sim \mathcal{B}(10, 0.5)$, soit 0.377.

L'événement $X \leq 4$ demeure donc **assez probable** : on s'attend à observer des preuves aussi extrêmes (ou plus encore) par hasard, $\approx 38\%$ du temps.

Avec 40 F sur 100 lancers, la valeur$-p$ est $P(Y \leq 40)$, **en supposant que** $Y \sim \mathcal{B}(100, 0.5)$, soit 0.028.

Ainsi, si $P(F) = 0.5$, l'événement $Y \leq 40$ est **fort peu probable** : on s'attend à observer des preuves aussi extrêmes par hasard, $\approx 3\%$ du temps.

Voici une autre façon de voir les choses : la valeur$-p$ est l'aire de la **queue** de la distribution de la v.a. lorsque l'on suppose que l'affirmation est vraie :

petite valeur$-p$ $\Longleftrightarrow$ preuves contre l'affirmation sont fortes

Nous avons évalué un vernaculaire et une notation traditionnels afin de décrire cette approche de ”test des hypothèses” :

1. “l'affirmation” est appelée l'**hypothèse nulle** et notée H_0 ;
2. la “suspicion” est appelée l'**hypothèse alternative** et notée H_1 ;
3. la quantité (aléatoire) que nous utilisons pour mesurer la preuve est appelée **statistique du test** – nous devons connaître sa distribution lorsque H_0 est vérifiée, et
4. la **valeur**$-p$ quantifie “la preuve contre H_0”.

Exemples: considérons la situation de pile ou face décrite précédemment. L'hypothèse nulle est

$$H_0 : P(F) = 0.5\,,$$

tandis que l'hypothèse alternative est

$$H_1 : P(F) < 0.5\,.$$

La pièce est lancée n fois : la statistique de test est $X = \#$ de F.

1. Si $n = 10$ et $X = 4$, la valeur$-p$ est $P(X \leq 4) = 0.377$, en supposant que $X \sim \mathcal{B}(10, 0.5)$.

2. Si $n = 100$ et $X = 40$, la valeur$-p$ est $P(X \leq 40) = 0.028$, en supposant que $X \sim \mathcal{B}(100, 0.5)$.

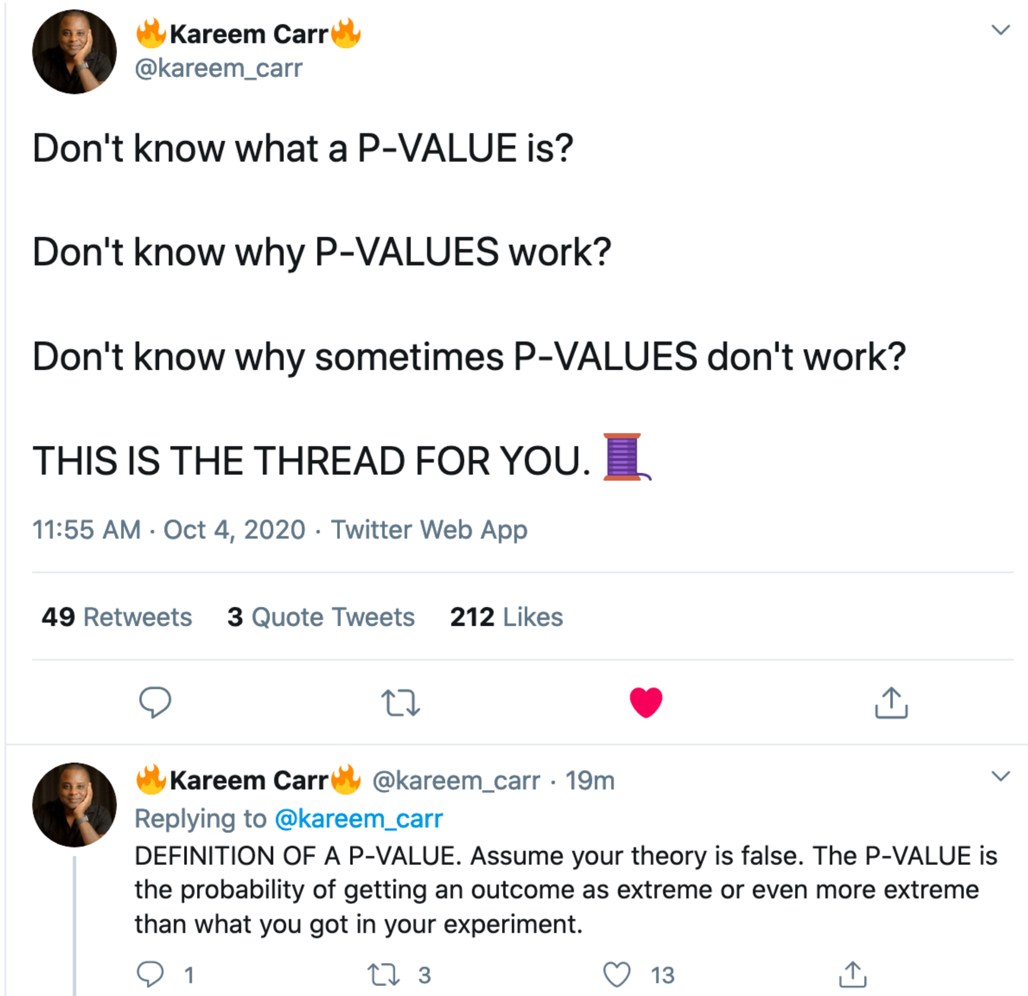
Quelle taille la valeur$-p$ doit-elle prendre ?

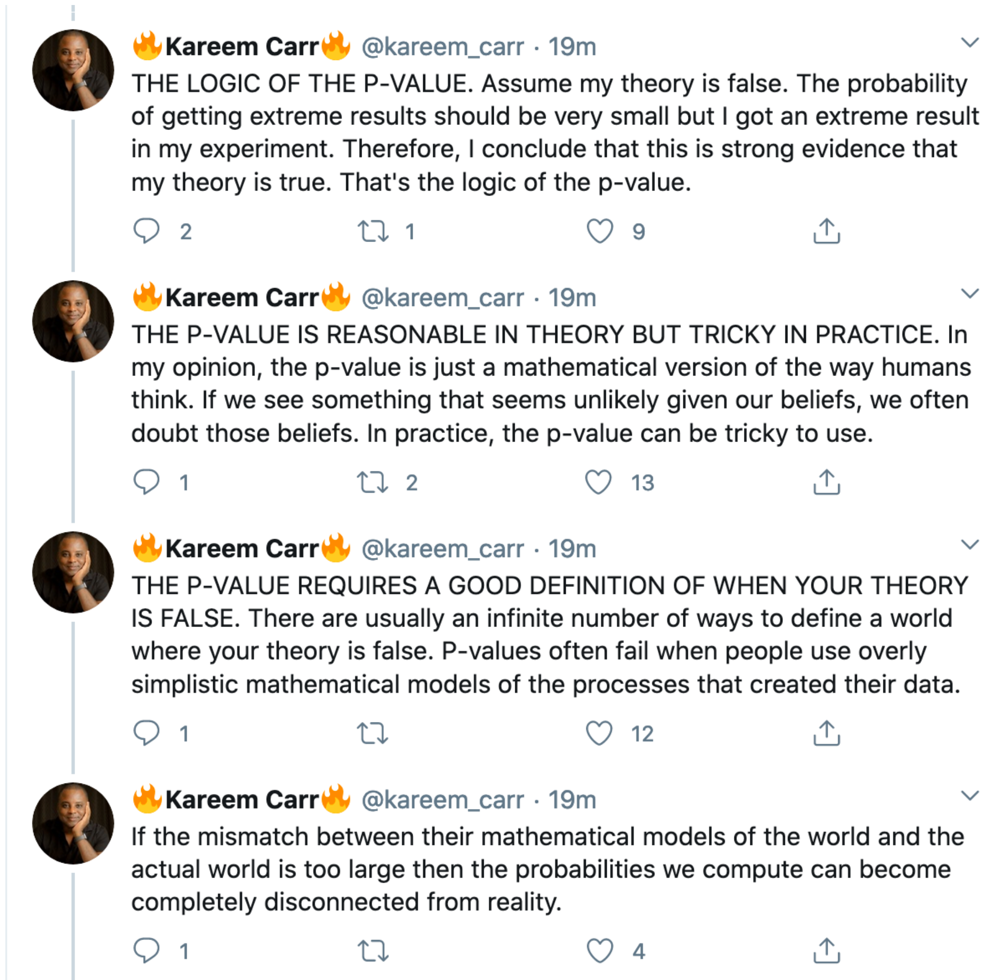
Nous avons conclu que 38% n'était "pas si petit que ça". À quel instant aurons-nous une "preuve irréfutable" contre H_0 ?

Il n'y a pas de réponse facile à cette question. Elle dépend de nombreux facteurs, notamment des pénalités que nous aurons à payer si nous nous sommes trompés. En règle générale, nous examinons la probabilité de commettre une **erreur de type I**, $\alpha = P(\text{rejeter } H_0 \mid H_0 \text{ est valide})$:

- si la valeur$-p \leq \alpha$, alors nous **rejetons** H_0 **en faveur de** H_1 ;
- si la valeur$-p > \alpha$, alors nous **ne pouvons pas rejeter** H_0 (ce qui **ne revient pas à accepter** H_0).

Par convention, nous utilisons souvent $\alpha = 0.01$ ou $\alpha = 0.05$.


Kareem Carr
@kareem_carr
Don't know what a P-VALUE is?
Don't know why P-VALUES work?
Don't know why sometimes P-VALUES don't work?
THIS IS THE THREAD FOR YOU.
11:55 AM · Oct 4, 2020 · Twitter Web App
49 Retweets 3 Quote Tweets 212 Likes
Kareem Carr @kareem_carr · 19m
Replying to @kareem_carr
DEFINITION OF A P-VALUE. Assume your theory is false. The P-VALUE is the probability of getting an outcome as extreme or even more extreme than what you got in your experiment.
1 3 13



Kareem Carr @kareem_carr · 19m

THE LOGIC OF THE P-VALUE. Assume my theory is false. The probability of getting extreme results should be very small but I got an extreme result in my experiment. Therefore, I conclude that this is strong evidence that my theory is true. That's the logic of the p-value.

2 1 9

Kareem Carr @kareem_carr · 19m

THE P-VALUE IS REASONABLE IN THEORY BUT TRICKY IN PRACTICE. In my opinion, the p-value is just a mathematical version of the way humans think. If we see something that seems unlikely given our beliefs, we often doubt those beliefs. In practice, the p-value can be tricky to use.

1 2 13

Kareem Carr @kareem_carr · 19m

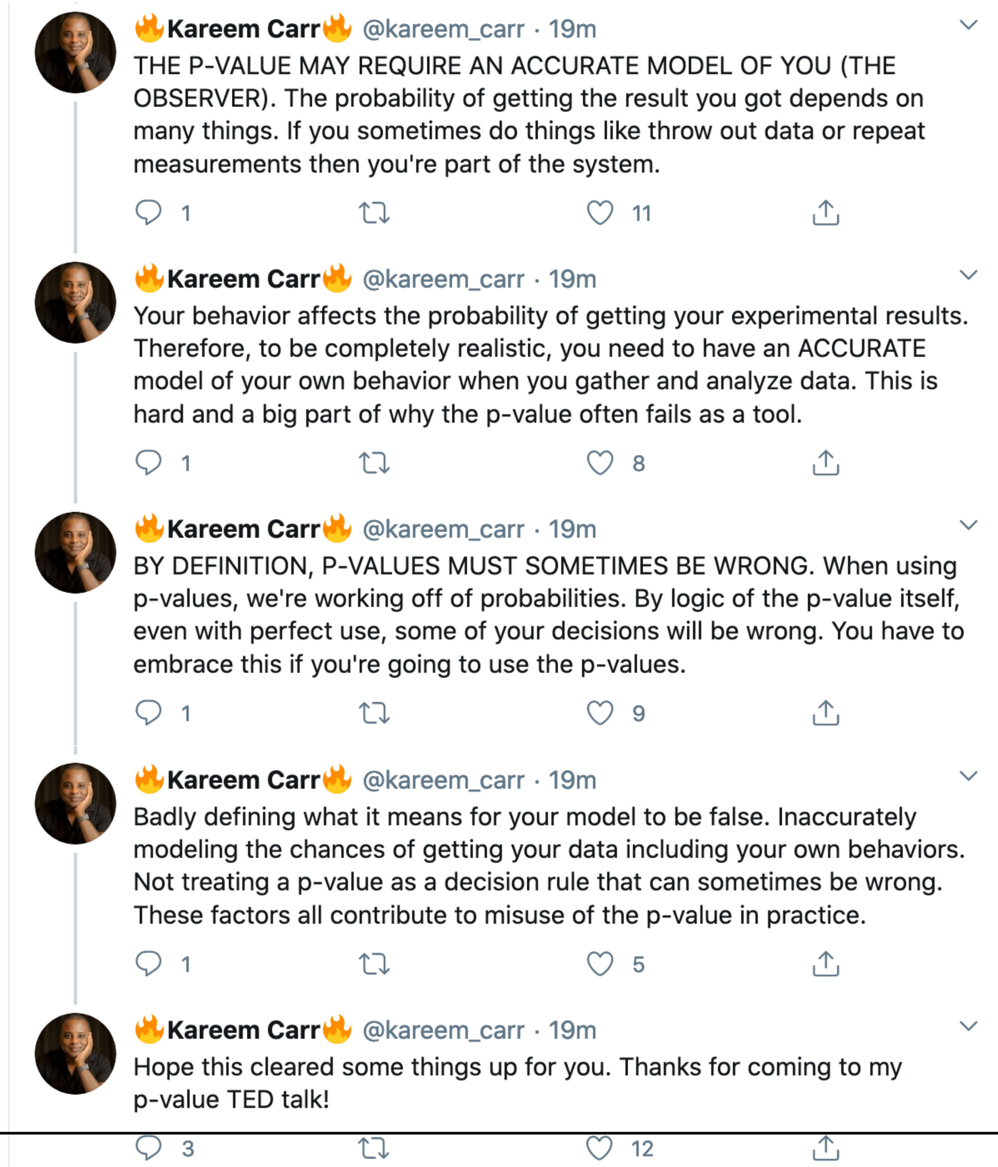
THE P-VALUE REQUIRES A GOOD DEFINITION OF WHEN YOUR THEORY IS FALSE. There are usually an infinite number of ways to define a world where your theory is false. P-values often fail when people use overly simplistic mathematical models of the processes that created their data.

1 12

Kareem Carr @kareem_carr · 19m

If the mismatch between their mathematical models of the world and the actual world is too large then the probabilities we compute can become completely disconnected from reality.

1 4



Kareem Carr @kareem_carr · 19m
THE P-VALUE MAY REQUIRE AN ACCURATE MODEL OF YOU (THE OBSERVER). The probability of getting the result you got depends on many things. If you sometimes do things like throw out data or repeat measurements then you're part of the system.
1 11

Kareem Carr @kareem_carr · 19m
Your behavior affects the probability of getting your experimental results. Therefore, to be completely realistic, you need to have an ACCURATE model of your own behavior when you gather and analyze data. This is hard and a big part of why the p-value often fails as a tool.
1 8

Kareem Carr @kareem_carr · 19m
BY DEFINITION, P-VALUES MUST SOMETIMES BE WRONG. When using p-values, we're working off of probabilities. By logic of the p-value itself, even with perfect use, some of your decisions will be wrong. You have to embrace this if you're going to use the p-values.
1 9

Kareem Carr @kareem_carr · 19m
Badly defining what it means for your model to be false. Inaccurately modeling the chances of getting your data including your own behaviors. Not treating a p-value as a decision rule that can sometimes be wrong. These factors all contribute to misuse of the p-value in practice.
1 5

Kareem Carr @kareem_carr · 19m
Hope this cleared some things up for you. Thanks for coming to my p-value TED talk!
3 12

6.1 – Les tests d'hypothèse

Une **hypothèse** est une conjecture concernant la valeur d'un paramètre.

Les tests d'hypothèses nécessitent deux hypothèses concurrentes :

- une **hypothèse nulle**, désignée par H_0 ;
- une **hypothèse alternative**, désignée par H_1 ou H_A.

L'hypothèse est testée en évaluant les preuves expérimentales :

- nous **rejetons** H_0 si les preuves contre H_0 sont **fortes** ;
- nous **ne rejetons pas** H_0 si les preuves contre H_0 sont **insuffisantes**.

Si les preuves à l'encontre de H_0 sont **suffisamment fortes**, nous **rejetons** H_0 **en faveur de** H_1 ; on dira de l'évidence **en faveur de** H_1 qu'elle est **significative**. Si les preuves à l'encontre de H_0 sont **faibles**, nous **ne pouvons pas rejeter** H_0 ; on dira de l'évidence contre H_0 qu'elle est **non significative**.

Lorsque nous **ne parvenons pas à rejeter** H_0, cela ne veut pas dire que nous **acceptons** H_0 : nous n'avons tout simplement **pas assez** d'évidence pour **rejeter** H_0.

Principe général: les hypothèses doivent être formulées **avant l'expérience** ou **l'étude**, qui est menée afin d'évaluer les preuves allant à l'encontre de l'hypothèse **nulle**.

Autrement dit, il est crucial de formuler H_1 **avant** avoir examiné les données que l'on utilisera dans l'analyse.

Les hypothèses scientifiques peuvent souvent être exprimées en termes de la présence ou de l'absence d'un effet dans les données.

Dans ce cas, nous utilisons l'hypothèse nulle suivante :

$$H_0 : \textbf{ il n'y a pas d'effet}$$

à l'encontre de l'hypothèse alternative :

$$H_1 : \textbf{ il y a un effet}$$

Avez-vous entendu parler de la crise de réplicabilité en psychologie ?

Les erreurs dans les tests d'hypothèses

On peut commettre deux types d'erreurs en testant H_0 par rapport à H_1.

	Décision: nous rejetons H_0	**Décision:** nous ne rejetons pas H_0
Réalité: H_0 est valide	Erreur de type I	Pas d'erreur
Réalité: H_0 n'est pas valide	Pas d'erreur	Erreur de type II

- Si nous rejetons H_0 alors que H_0 est valide $\Longrightarrow$ nous avons commis une **erreur de type I**.
- Si nous ne parvenons pas à rejeter H_0 lorsque H_0 est faux $\Longrightarrow$ nous avons commis une **erreur de type II**.

Exemples:

1. Si nous concluons d'un traitement qu'il est utile pour traiter une maladie particulière mais que ce n'est pas le cas en réalité, nous avons commis une **erreur de type I**.

2. Si nous ne pouvons pas conclure d'un traitement qu'il est utile pour traiter la maladie, mais qu'en réalité le traitement est efficace, nous avons commis une **erreur de type II**.

Quel est le pire de ces deux types d'erreur ?

La puissance d'un test

La probabilité de commettre une erreur de type I est désignée par

$$\alpha = P(\text{rejeter } H_0 \mid H_0 \text{ est valide}).$$

La probabilité de commettre une erreur de type II est

$$\beta = P(\text{ne pas rejeter } H_0 \mid H_0 \text{ n'est pas valide}).$$

La **puissance d'un test** est la probabilité de rejeter correctement H_0 :

$$\text{puissance} = P(\text{rejeter } H_0 \mid H_0 \text{ n'est pas valide}) = 1 - \beta.$$

Les valeurs conventionnelles de α, β, et de la puissance sont respectivement de 0.05, 0.2, et 0.8.

6.2 – Les types d'hypothèses nulles et d'hypothèses alternatives

Soit μ le paramètre de population d'intérêt. Les hypothèses sont exprimées en termes de valeurs de ce paramètre.

L'hypothèse nulle est une hypothèse **simple**, c-à-d qu'elle prend la forme :

$$H_0 : \mu = \mu_0.$$

où μ_0 est une certaine valeur candidate ("simple" signifie que l'on suppose qu'il s'agit d'une valeur **unique**).

L'hypothèse alternative H_1 est une hypothèse **mixte**, c'est-à-dire qu'elle contient **plus d'une valeur candidate**.

Selon le contexte, le test d'hypothèse prend l'une des trois formes suivantes :

$$H_0 : \mu = \mu_0. \text{ où } \mu_0 \text{ est un nombre,}$$

- par rapport à une alternative **bilatérale** : $H_1 : \mu \neq \mu_0$;
- par rapport à une alternative **à gauche** : $H_1 : \mu < \mu_0$;
- par rapport à une alternative **à droite** : $H_1 : \mu > \mu_0$.

Rappel : la formulation de l'hypothèse alternative dépend de notre hypothèse de recherche et est déterminée **avant** l'expérience ou l'étude.

Exemple : on cherche souvent à vérifier si de nouvelles conditions expérimentales entraînent la modification d'une caractéristique spécifique. Un enquêteur affirme que l'utilisation d'un nouveau type de sol produra des plantes plus grandes, en moyenne, par rapport à celle d'un sol traditionnel. La hauteur moyenne des plantes sous cette dernière est de 20 cm.

1. Formulez les hypothèses qui seront testées pour vérifier l'affirmation.

2. Si un autre enquêteur soupçonne le contraire, c'est-à-dire que la hauteur moyenne des plantes lors de l'utilisation du nouveau sol sera plus petite que la hauteur moyenne des plantes avec l'ancien sol. Quelles sont les hypothèses à formuler ?

3. Une 3ème chercheuse pense qu'elle aura un effet, mais elle n'est pas sûr que l'effet sera de produire des plantes plus courtes ou plus grandes. Quelles hypothèses faut-il alors formuler ?

Solution : soit μ la hauteur moyenne des plantes dans le nouveau type de sol. Dans les trois cas, l'hypothèse nulle est $H_0 : \mu = 20$. L'hypothèse alternative dépend de la situation :

1. $H_1 : \mu > 20$;

2. $H_1 : \mu < 20$;

3. $H_1 : \mu \neq 20$.

Pour chaque H_1, les valeurs$-p$ correspondantes sont calculées **différemment** lors du test de H_0 contre H_1.

6.3 – Les statistiques des tests et les régions critiques

Nous utilisons une **statistique de test** afin de tester une hypothèse statistique. C'est une fonction de l'échantillon aléatoire et du paramètre de population d'intérêt.

Nous rejetons H_0 si la valeur de la statistique de test se trouve dans la **région critique** (ou la **zone de rejet**) ; c'est un **intervalle** de $\mathbb{R}$.

Elle est obtenue en utilisant la définition des erreurs dans les tests d'hypothèse. On la choisit de sorte à ce que

$$\alpha = P(\textbf{rejeter } H_0 \mid H_0 \textbf{ est valide})$$

soit égal à une certaine valeur prédéterminée, comme 0.05 ou 0.01.

Exemples : on met au point un nouveau procédé de durcissement pour un certain type de ciment ; il donne une résistance moyenne à la compression de 5000 kg/cm^2, avec un écart-type de 120 kg/cm^2. On test $H_0 : \mu = 5000$ par rapport à $H_1 : \mu < 5000$, avec un échantillon aléatoire de 49 morceaux de ciment. Supposons que la région critique dans ce cas précis soit $\overline{X} < 4970$, c'est-à-dire que nous rejetterions H_0 si $\overline{X} < 4970$.

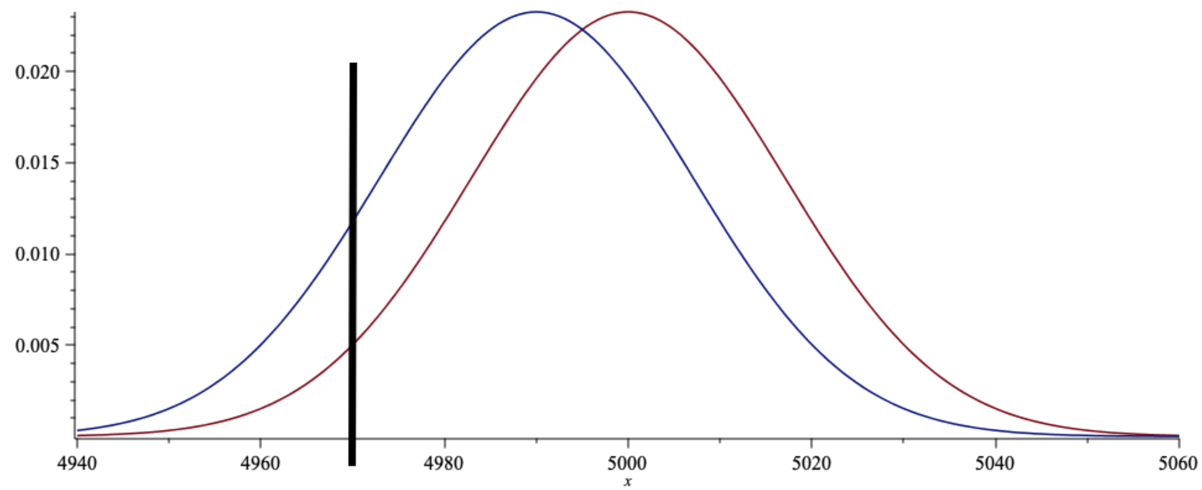
1. Quelle est la probabilité de commettre une erreur de type I ?

   Solution : par définition, nous avons

   $$\alpha = P(\text{type I}) = P(\text{rejeter } H_0 \mid H_0 \textbf{ valide}) = P(\overline{X} < 4970 \mid \mu = 5000).$$

   Ainsi, selon le TLC,

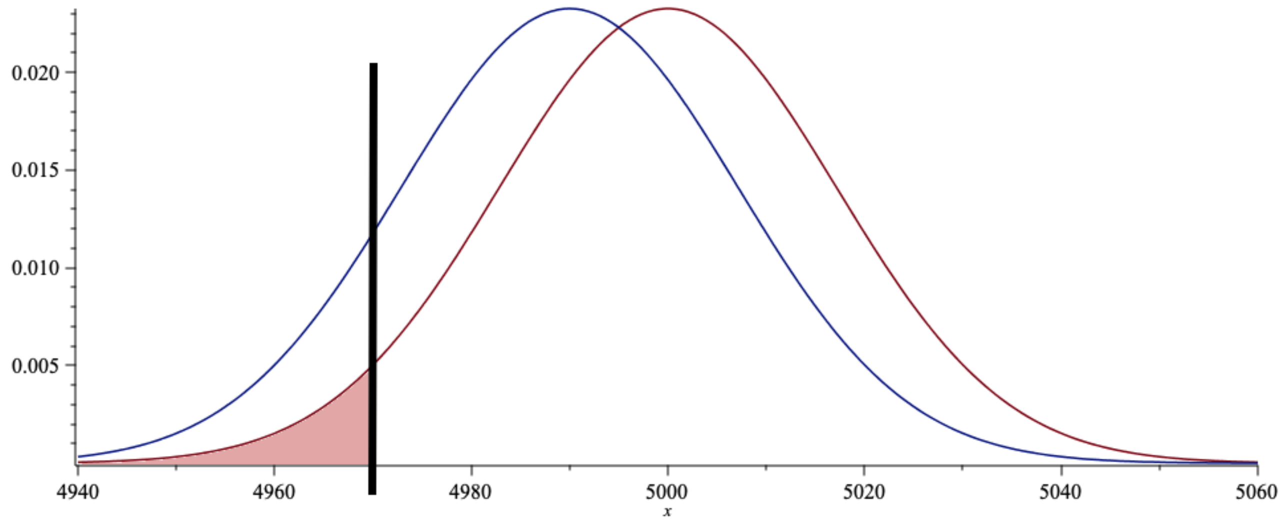
   $$\alpha \approx P\left(\frac{\overline{X} - \mu}{\sigma/\sqrt{n}} < \frac{4970 - 5000}{120/7}\right) \approx P(Z < -1.75) \approx 0.0401\,.$$



Dist. d'échantillonnage de $\overline{X}$ sous H_0 en **rouge** ($\mu = 5000$, $\sigma = 120/7$)

Dist. d'échantillonnage de $\overline{X}$ sous H_1 en **bleu** ($\underbrace{\mu = 4990}_{\text{disons}}$, $\sigma = 120/7$)

Région critique $\overline{X} < 4970$, à **gauche** de la droite verticale.



$$\alpha = P(\text{rejeter } H_0 \mid H_0 \text{ valide}) = P(\overline{X} < 4970 \mid \mu = 5000)$$

rejeter $H_0 \Longrightarrow$ à gauche de $\overline{X} = 4970$ (dans la région critique)

H_0 est valide $\Longrightarrow$ sous la dist. en **rouge** ($\mu = 5000$)

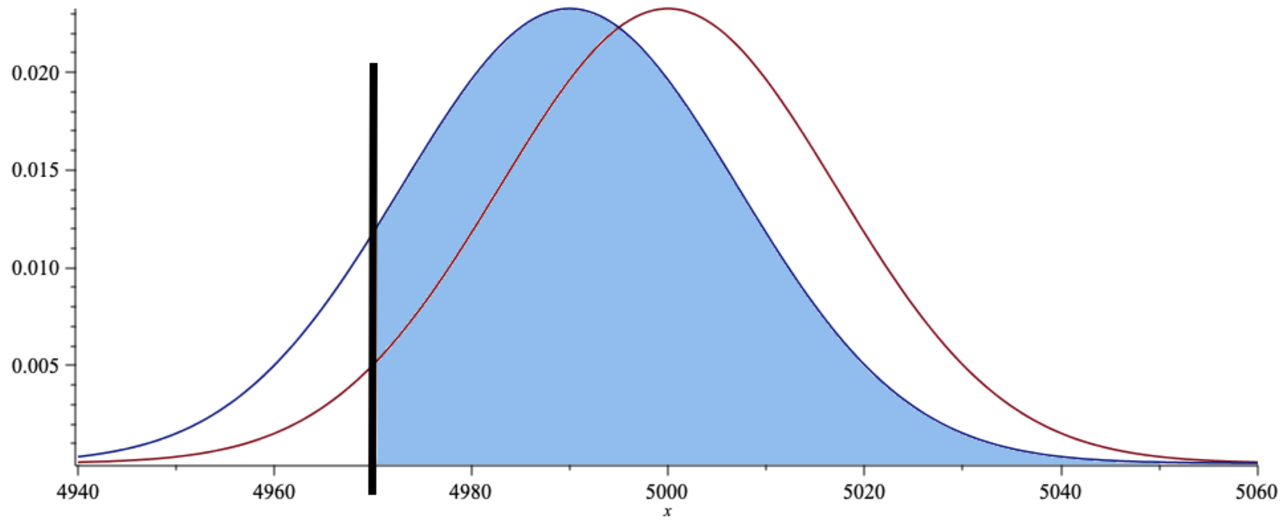
2. Evaluez la probabilité de commettre une erreur de type II si μ est en réalité 4990, disons (et non 5000, comme le prétend H_0).

Solution : par définition, nous avons

$$\begin{aligned}\beta = P(\text{type II}) &= P(\text{ne pas rejeter } H_0 \mid H_0 \textbf{ non valide}) \\ &= P(\overline{X} > 4970 \mid \mu = 4990).\end{aligned}$$

Ainsi, selon le TLC,

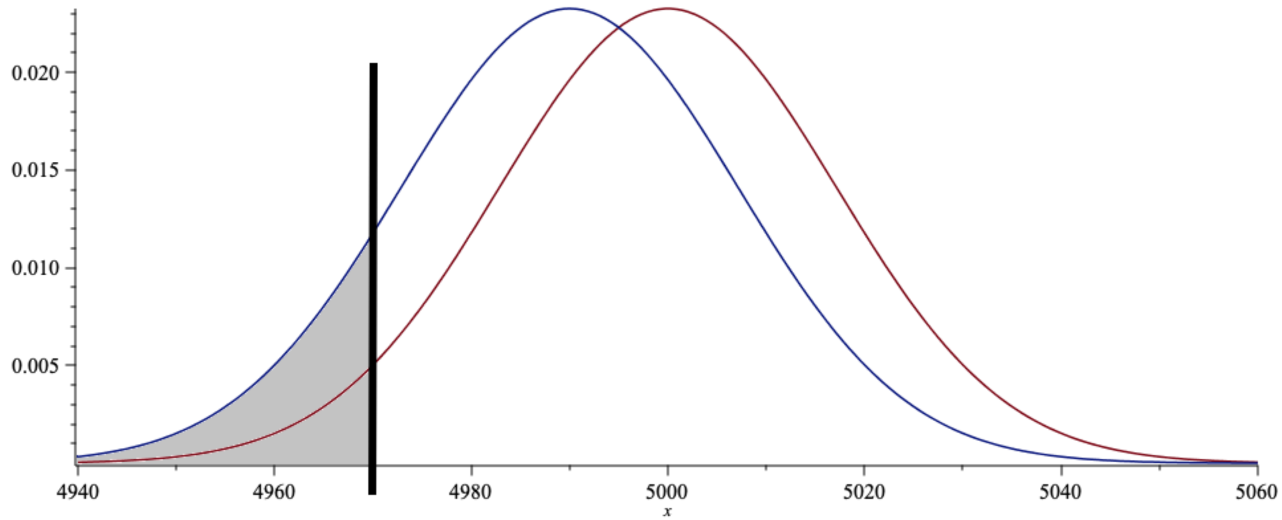
$$\begin{aligned}\beta = P(\overline{X} > 4970) = P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} > \frac{4970-4990}{120/7}\right) \approx P(Z > -1.17) \\ = 1 - \texttt{pnorm}(-1.17, 0, 1) \approx 0.879\,.\end{aligned}$$



$\beta = P(\text{ne pas rejeter } H_0 \mid H_0 \text{ n'est pas valide}) = P(\overline{X} > 4970 \mid \mu = 4990)$

ne pas rejeter $H_0 \Longrightarrow$ à droite de $\overline{X} = 4970$ (à l'ext. de la région critique)

H_0 n'est pas valide $\Longrightarrow$ sous la dist. **bleue** $(\mu = 4990)$



puissance $= P($rejeter $H_0 \mid H_0$n'est pas valide$) = P(\overline{X} < 4970) = 1 - \beta$

rejeter $H_0 \Longrightarrow$ à gauche de $\overline{X} = 4970$ (dans la région critique)

H_0 n'est pas valide $\Longrightarrow$ sous la dist. **bleue** $(\mu = 4990)$

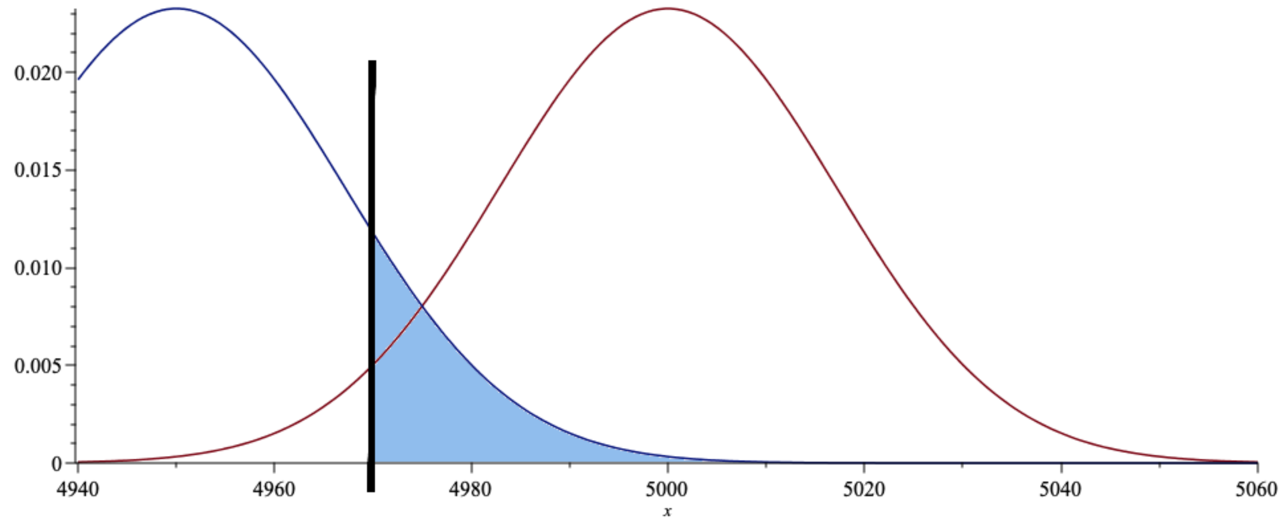
3. Évaluez la probabilité de commettre une erreur de type II si μ vaut en réalité 4950, disons (et non 5000, comme le prétend H_0).

Solution : par définition, nous avons

$$\begin{aligned}\beta &= P(\text{type II}) = P(\text{ne pas rejeter } H_0 \mid H_0 \text{ non valide}) \\ &= P(\overline{X} > 4970 \mid \mu = 4950).\end{aligned}$$

Ainsi, selon le TLC, nous avons

$$\begin{aligned}\beta &= P\left(\frac{\overline{X}-\mu}{\sigma/\sqrt{n}} > \frac{4970-4950}{120/7}\right) \approx P(Z > 1.17) \\ &= 1 - \texttt{pnorm}(1.17, 0, 1) \approx 0.121\,.\end{aligned}$$



$\beta = P(\text{ne pas rejeter } H_0 \mid H_0 \text{ n'est pas valide}) = P(\overline{X} > 4970)$

ne pas rejeter $H_0 \Longrightarrow$ à droite de $\overline{X} = 4970$ (à l'ext. de la région critique)

H_0 n'est pas valide $\Longrightarrow$ sous la dist. **bleue** ($\mu = 4950$)

6.4 – Les tests pour une moyenne quand la variance est connue

Supposons que $X_1, \ldots, X_n$ est un échantillon aléatoire provenant d'une population de moyenne μ et de variance σ^2, et que $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ désigne la moyenne de l'échantillon :

- si la population est normale, alors $\overline{X} \overset{\text{exacte}}{\sim} \mathcal{N}(\mu, \sigma^2/n)$;
- si la population est **non** normale, alors tant que n est **suffisamment élevé**, nous avons $\overline{X} \overset{\text{approx}}{\sim} \mathcal{N}(\mu, \sigma^2/n)$, selon le TLC.

Dans cette section, nous supposons que la variance de population σ^2 est connue, et que l'hypothèse concerne la moyenne de population μ (inconnue).

Explication : l'alternative à gauche

Supposons que nous souhaitions tester

$$H_0 : \mu = \mu_0 \quad \text{par rapport à} \quad H_1 : \mu < \mu_0.$$

Pour évaluer la preuve contre H_0, nous comparons $\overline{X}$ à μ_0 : sous H_0,

$$Z_0 = \frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \overset{\text{approx}}{\sim} \mathcal{N}(0,1).$$

La valeur observée Z_0 de la **statistique de test** Z est $z_0 = \frac{\overline{x}-\mu_0}{\sigma/n}$. Si $z_0 < 0$, nous avons des preuves que $\mu < \mu_0$. Cependant, nous ne rejetons H_0 en faveur de H_1 que si la preuve est **significative**.

Région critique: soit α le niveau de signification. Nous rejetons H_0 en faveur de H_1 uniquement si

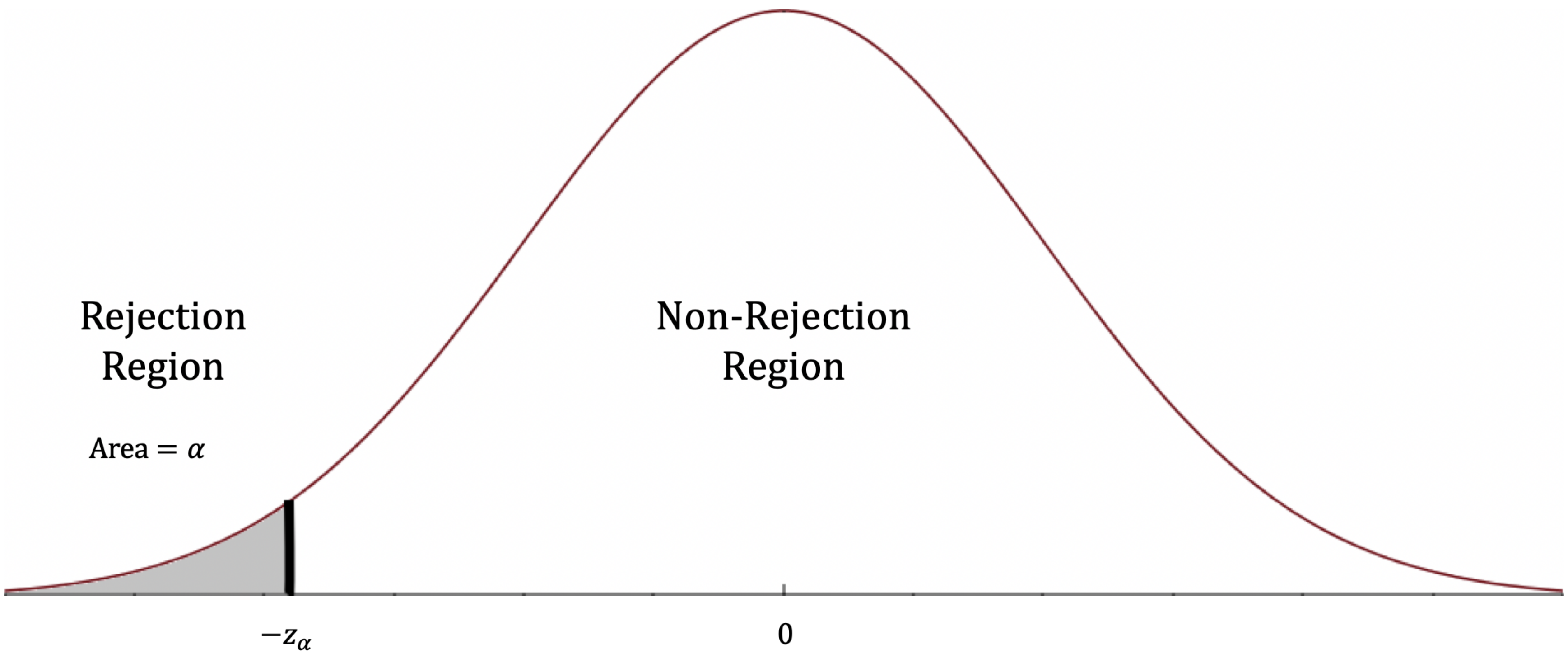
$$z_0 \leq -z_\alpha.$$

La **valeur**$-p$ correspondante du test est la probabilité d'observer des preuves aussi extrêmes (ou plus) que la preuve actuelle en faveur de H_1, en supposant que H_0 soit valide (c-à-d, simplement par hasard) ; plus extrême dans ce cas signifie plus à gauche :

$$\text{valeur} - p = P(Z \leq z_0) = \Phi(z_0),$$

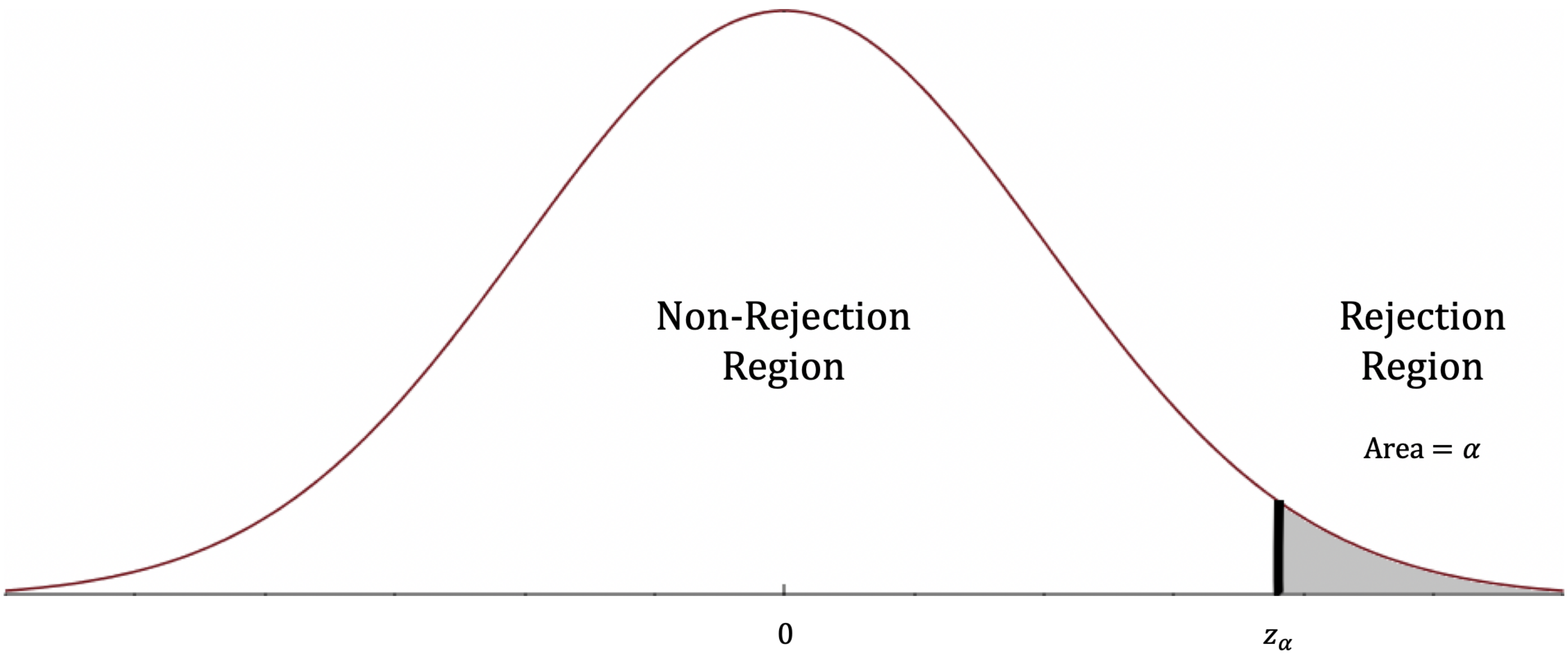
où z_0 est la valeur observée pour la statistique de test Z.

Règle de décision: si la valeur$-p \leq \alpha$, alors nous **rejetons** H_0 **en faveur de** H_1. Si la valeur$-p > \alpha$, nous **ne rejetons pas** H_0.



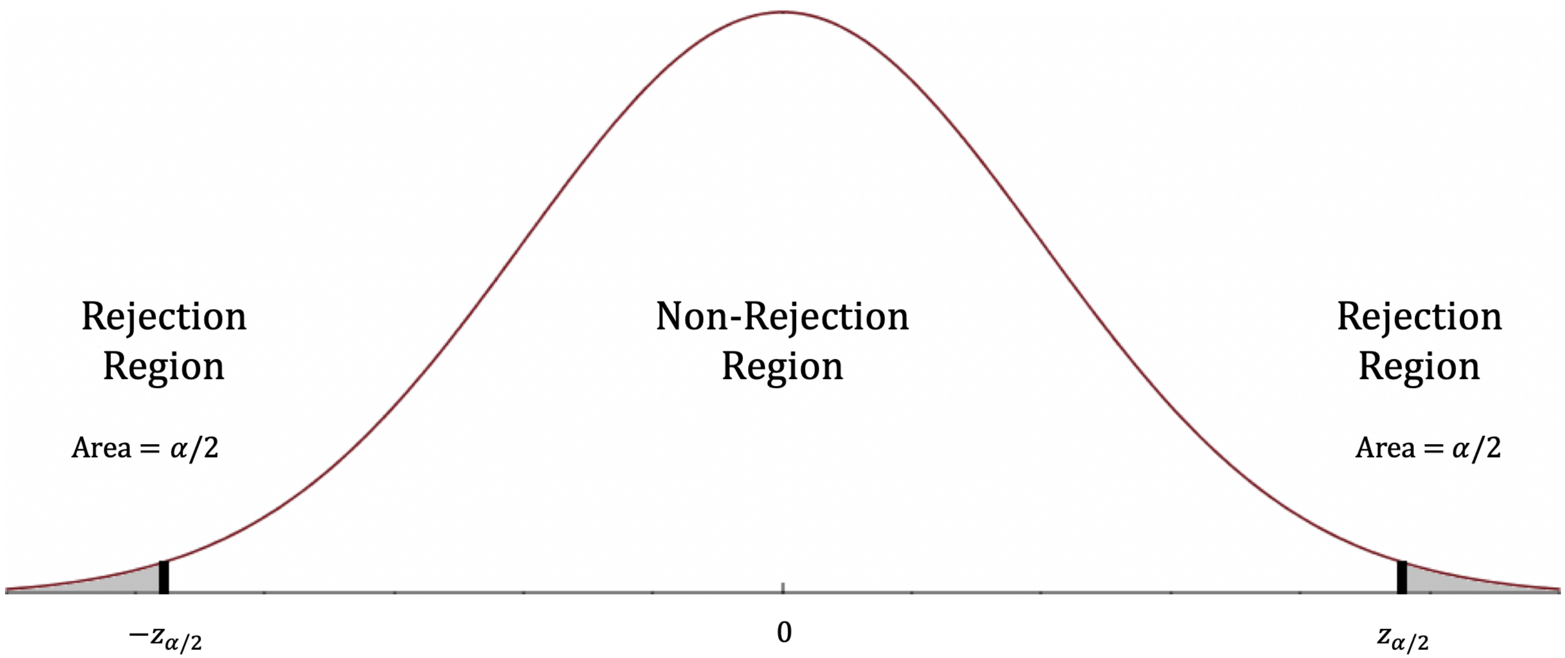
Test à gauche : $H_0 : \mu = \mu_0$ par rapport à $H_1 : \mu < \mu_0$.

Au niveau de signification α, si $z_0 = \frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}} \leq -z_\alpha$, nous rejetons H_0 en faveur de H_1.



Test à droite : $H_0 : \mu = \mu_0$ par rapport à $H_1 : \mu > \mu_0$.

Au niveau de signification α, si $z_0 = \frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}} \geq z_\alpha$, nous rejetons H_0 en faveur de H_1.



Test bilatéral : $H_0 : \mu = \mu_0$ par rapport à $H_1 : \mu \neq \mu_0$.

Au niveau de signification α, si $|z_0| = \left|\frac{\overline{x}-\mu_0}{\sigma/\sqrt{n}}\right| \geq z_{\alpha/2}$, nous rejetons H_0 en faveur de H_1.

Procédure

Étape 1 : définir l'hypothèse nulle $H_0 : \mu = \mu_0$

Étape 2 : choisir une hypothèse alternative H_1 (ce que nous essayons de démontrer en utilisant les données). Voici les alternatives :

- $H_1 : \mu < \mu_0$ (test unilatéral à gauche)
- $H_1 : \mu > \mu_0$ (test unilatéral à droite)
- $H_1 : \mu \neq \mu_0$ (test bilatéral)

Étape 3 : choisir un niveau de significance $\alpha = P($erreur de type I$)$, typiquement, $\alpha = 0.01$ ou 0.05.

Étape 4 : pour l'échantillon observé $\{x_1, \ldots, x_n\}$, calculer la valeur observée de la statistique de test $z_0 = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}}$.

Étape 5 : déterminer la région critique comme suit

Hypothèse alternative	Région critique
$H_1 : \mu > \mu_0$	$z_0 > z_\alpha$
$H_1 : \mu < \mu_0$	$z_0 < -z_\alpha$
$H_1 : \mu \neq \mu_0$	$\lvert z_0 \rvert > z_{\alpha/2}$

où z_α est la valeur critique satisfaisant $P(Z > z_\alpha) = \alpha$, avec $Z \sim \mathcal{N}(0,1)$:

α	z_α	$z_{\alpha/2}$
0.05	1.645	1.960
0.01	2.327	2.576

Étape 6 : calculer la valeur$-p$ associée comme suit

Hypothèse alternative	**Valeur**$-p$
$H_1 : \mu > \mu_0$	$P(Z > z_0)$
$H_1 : \mu < \mu_0$	$P(Z < z_0)$
$H_1 : \mu \neq \mu_0$	$2 \cdot \min\{P(Z > z_0), P(Z < z_0)\}$

où $Z \sim \mathcal{N}(0,1)$.

Règle de décision : si la valeur$-p \leq \alpha$, nous rejetons H_0 **en faveur de** H_1. Si la valeur$-p > \alpha$, nous **ne rejetons pas** H_0

C'est tout.

Voyons quelques exemples !

Exemples :

1. Des composantes sont fabriquées de sorte que la résistance suit une loi normale de moyenne $\mu = 40$ et d'écart-type $\sigma = 1.2$. Le procédé de fabrication a été modifié, et une augmentation de la résistance moyenne est revendiquée (l'écart-type reste le même). Un échantillon aléatoire de $n = 12$ composantes produits à l'aide du procédé modifié présente les résistances suivantes :

   ```
   42.5, 39.8, 40.3, 43.1, 39.6, 41.0,
   39.9, 42.1, 40.7, 41.6, 42.1, 40.8
   ```

   Les données fournissent-elles une preuve solide que la résistance moyenne dépasse maintenant 40 unités ? Utilisez $\alpha = 0.05$.

Solution : nous suivons la procédure décrite plus tôt, en testant $H_0 : \mu = 40$ par rapport à $H_1 : \mu > 40$.

La valeur observée de la moyenne de l'échantillon est $\overline{x} = 41.125$. Ainsi,

$$\text{valeur}-p = P(\overline{X} \geq \overline{x}) = P(\overline{X} \geq 41.125)$$
$$= P\left(\frac{\overline{X} - \mu_0}{\sigma/\sqrt{n}} \geq \frac{41.125 - \mu_0}{\sigma/\sqrt{n}}\right) = P(Z \geq 3.25) \approx 0.006.$$

Comme la valeur$-p$ est **inférieure** à α, nous **rejetons** H_0 **en faveur de** H_1.

Une autre façon de voir les choses est que si le modèle '$\mu = 40$' est valide, alors il est **très peu probable** que nous observions l'événement $\{\overline{X} \geq 41.125\}$ entièrement par hasard, et donc le processus de fabrication a **probablement un effet** dans la direction revendiquée.

2. Une balance fonctionne correctement si les mesures diffèrent du poids réel par un terme d'erreur aléatoire suivant une loi normale avec ua plus $\sigma = 0.007$ grammes. On soupçonne que la balance ajoute systématiquement aux poids. Pour tester cette hypothèse, on effectue $n = 10$ mesures sur un poids étalon de 1.0g, donnant un ensemble de mesures dont la moyenne est de 1.0038g. La balance ajoute-t-elle aux poids des mesures ? Utilisez $\alpha = 0.05,\ 0.01$.

   Solution : soit μ le poids qui serait enregistré en l'absence d'erreurs aléatoires. Nous testons pour $H_0 : \mu = 1.0$ par rapport à $H_1 : \mu > 1.0$.

   La statistique de test observée est $z_0 = \frac{1.0038-1.0}{0.007/\sqrt{10}} \approx 1.7167$. Puisque

$$z_{0.05} = 1.645 < z_0 = 1.7167 \leq z_{0.01} = 2.327,$$

   nous **rejetons** H_0 lorsque $\alpha = 0.05$, mais **non** lorsque $\alpha = 0.01$.

3. À l'exemple précédent, supposons que nous nous intéressons au bon fonctionnement de la balance, ce qui signifie que les enquêteurs pensent qu'il pourrait y avoir une erreur systématique de lecture, mais qu'ils ne sont pas sûrs de la direction dans laquelle cette erreur se produirait. Les données de l'échantillon fournissent-elles la preuve que la balance est systématiquement biaisée ? Utilisez $\alpha = 0.05, 0.01$.

   Solution : soit μ comme à l'exemple précédent. Nous testons pour $H_0 : \mu = 1.0$ par rapport à $H_1 : \mu \neq 1.0$.

   La statistique de test est toujours $z_0 = 1.7167$.

   Puisque $|z_0| \leq z_{\alpha/2}$ et pour $\alpha = 0.05$, et pour $\alpha = 0.01$, nous **ne rejetons pas** H_0 lorsque $\alpha = 0.05$ ou $\alpha = 0.01$.

   L'interprétation de la statistique de test dépend du type d'hypothèse alternative choisi (et donc, du contexte).

4. En général, les notes d'une classe suivent une loi normale de moyenne 60 et de variance 100. Neuf élèves sont choisis au hasard ; leur note moyenne est de 55. Ce sous-groupe est-il 'inférieur à la moyenne' ?

   Solution : soit μ la moyenne rélle d'un sous-groupe de 9 élèves. Nous testons $H_0 : \mu = 60$ par rapport à $H_1 : \mu < 60$. La statistique de test observée est

$$z_0 = \frac{55 - 60}{10/\sqrt{9}} = -1.5.$$

   La valeur$-p$ correspondante est

$$P(\overline{X} \leq \overline{x}) = P(Z \leq -1.5) = 0.07.$$

   L'évidence n'est ainsi **pas suffisament forte pour rejeter** l'affirmation que le sous-groupe est 'moyen', que nous utilisions $\alpha = 0.05$ ou $\alpha = 0.01$.

5. Nous considérons la même configuration que dans l'exemple précédent, mais cette fois la taille de l'échantillon est de $n = 100$, et non de 9. Qu'en est-il alors ?

Solution : soit μ la moyenne rélle d'un sous-groupe de 100 élèves. Le test ne change pas, mais la statistique de test observée est maintenant

$$z_0 = \frac{55 - 60}{10/\sqrt{100}} = -5$$

et la valeur$-p$ correspondante est

$$P(\overline{X} \leq \overline{x}) = P(Z \leq -5) \approx 0.00.$$

Nous **rejetons** H_0 **en faveur de** H_1, que nous utilisions $\alpha = 0.05$ ou $\alpha = 0.01$. La taille de l'échantillon joue un rôle !

Les tests et les intervalles de confiance

Il est de plus en plus courant pour les analystes d'éviter complètement le calcul de la valeur$-p$, en faveur d'une approche se basant sur les **intervalles de confiance**.

Pour un α donné, nous rejetons $H_0 : \mu = \mu_0$ en faveur de $H_1 : \mu \neq \mu_0$ si μ_0 **ne se retrouve pas** dans l'I.C. de μ à environ $100(1-\alpha)\%$.

Exemple : Un fabricant affirme qu'un type particulier de moteur utilise 20 gallons de carburant pour fonctionner pendant une heure. D'après des études antérieures, on sait que la quantité horaire de carburant utilisée suit une loi normale de variance $\sigma^2 = 25$. Un échantillon de taille $n = 9$ est prélevé et on obtient la moyenne empirique $\overline{X} = 23$. Acceptons-nous l'affirmation du fabricant ? Utilisez $\alpha = 0.05$.

Solution : nous testons $H_0 : \mu = 20$ vs. $H_1 : \mu \neq 20$. La statistique de test observée est

$$z_0 = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{23 - 20}{5/\sqrt{9}} = 1.8.$$

Pour un test bilatéral avec $\alpha = 0.05$, la valeur critique est $z_{0.025} = 1.96$. Comme $|z_0| \leq z_{0.025}$, z_0 **ne se retrouve pas** dans la région critique, nous **ne rejetons pas** H_0.

L'avantage de l'approche des intervalles de confiance est qu'elle permet de tester plusieurs affirmations simultanément. Puisque nous connaissons la variance de la population sous-jacente, on obtient un I.C. de μ à environ $100(1-\alpha)\%$ à l'aide de

$$\overline{X} \pm z_{\alpha/2}\sigma/\sqrt{n} = 23 \pm 1.96 \cdot 5/\sqrt{9} = (19.73, 26.26).$$

Nous ne rejetons pas les affirmations que $\mu = 20$, $\mu = 19.8$, $\mu = 26.2$, etc.

6.5 – Les tests pour une moyenne quand la variance est inconnue

Si les données sont normales et que σ est inconnu, nous pouvons l'estimer à l'aide de l'**écart-type de l'échantillon**

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

Comme nous l'avons vu pour les I.C., la statistique de test

$$T = \frac{\overline{X} - \mu}{S/\sqrt{n}} \sim t(n-1)$$

suit une loi T **de Student avec** $n-1$ **degrés de liberté**.

Nous pouvons suivre les mêmes étapes que pour le test à variance connue, avec des régions critiques et des valeurs$-p$ modifiées :

Hypothèse alternative	**Région critique**
$H_1 : \mu > \mu_0$	$t_0 > t_\alpha(n-1)$
$H_1 : \mu < \mu_0$	$t_0 < -t_\alpha(n-1)$
$H_1 : \mu \neq \mu_0$	$\|t_0\| > t_{\alpha/2}(n-1)$

où $t_0 = \frac{\overline{x}-\mu_0}{S/\sqrt{n}}$, $t_\alpha(n-1)$ est la valeur$-t$ satisfaisant $P(T > t_\alpha(n-1)) = \alpha$, et T suit une loi de Student avec $n-1$ degrés de liberté.

Hypothèse alternative	**Valeur**$-p$
$H_1 : \mu > \mu_0$	$P(T > t_0)$
$H_1 : \mu < \mu_0$	$P(T < t_0)$
$H_1 : \mu \neq \mu_0$	$2 \cdot \min\{P(T > t_0), P(T < t_0)\}$

Exemple : considérons les observations suivantes, issues d'une population normale dont la moyenne et la variance sont inconnues

```
18.0, 17.4, 15.5, 16.8, 19.0, 17.8, 17.4, 15.8,
17.9, 16.3, 16.9, 18.6, 17.7, 16.4, 18.2, 18.7.
```

Testez $H_0 : \mu = 16.6$ vs. $H_1 : \mu > 16.6$, en utilisant $\alpha = 0.05$.

Solution : la taille de l'échantillon est $n = 16$, et la moyenne et l'écart-type de l'échantillon sont $\overline{X} = 17.4$ et $S = 1.078$, respectivement.

La statistique de test observée est

$$t_0 = \frac{\overline{x} - \mu_0}{S/\sqrt{n}} = \frac{17.4 - 16.6}{1.078/\sqrt{16}} \approx 2.968,$$

de sorte que la valeur$-p$ est

$$\text{valeur}-p = P(\overline{X} \geq \overline{x}) = P(\overline{X} \geq 17.4) = P(T > 2.968),$$

où T suit une loi de Student avec $\nu = n - 1 = 15$ degrés de liberté. Selon les tables t, on reqarque que, pour $\nu = 15$,

$$P\left(T(15) \geq 2.947\right) \approx 0.005 \quad \text{et} \quad P\left(T(15) \geq 3.286\right) \approx 0.0025\,.$$

Par conséquent, la valeur $p-$ correspondante se situe quelque part entre ces points extrêmes, c'est-à-dire dans l'intervalle $(0.0025, 0.005)$.

En particulier, la valeur$-p \leq 0.05$, ce qui supporte fortement l'affirmation **contre** $H_0: \mu = 16.6$.

6.6 – Les tests pour une proportion

Le principe est à peu près le même, comme nous pouvons le voir dans l'exemple suivant.

Exemple : on a demandé à un groupe de 100 catholiques américains adultes prélevés au hasard s'ils étaient favorable à ce que les femmes puissent devenir prêtres : 60 d'entre eux ont répondu "Oui". La preuve est-elle assez forte pour conclure que plus de la moitié des catholiques américains sont favorables à ce que les femmes puissent être prêtres ?

Solution : soit X le nombre de réponses "Oui". Nous supposons que $X \sim \mathcal{B}(100, p)$, où p est la proportion réelle de catholiques américains favorables à ce que les femmes deviennent prêtres.

Testons $H_0 : p = 0.5$ vs. $H_1 : p > 0.5$. Sous H_0, $X \sim \mathcal{B}(100, 0.5)$.

The $p-$value that corresponds to the observed sample is

$$\begin{aligned} P(X \geq 60) &= 1 - P(X < 60) = 1 - P(X \leq 59) \\ &\approx 1 - P\left(\frac{X + 0.5 - np}{\sqrt{np(1-p)}} \leq \frac{59 + 0.5 - 50}{\sqrt{25}}\right) \\ &\approx 1 - P(Z \leq 1.9) = 0.0287, \end{aligned}$$

où les $+\ 0.5$ proviennent de la correction normale de la loi binomiale (cf. chapitre 3).

Ainsi, nous **rejetons** H_0 lorsque $\alpha = 0.05$, mais nous **ne rejetons pas** H_0 lorsque $\alpha = 0.01$.

6.7 – Les tests pour deux échantillons appariés

Soit $X_{1,1},\ldots,X_{1,n}$ un échantillon aléatoire provenant d'une population normale de moyenne inconnue μ_1 et de variance inconnue σ^2 ; soit $X_{2,1},\ldots,X_{2,n}$ un échantillon aléatoire provenant d'une population normale de moyenne inconnue μ_2 et de variance inconnue σ^2, les deux populations étant **dépendantes** l'une de l'autre (i.e., les échantillons proviennent de la même population, ou soint des mesures sur les mêmes unités, etc.).

Nous souhaitons tester $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 \neq \mu_2$. Pour ce faire, nous calculons les différences $D_i = X_{1,i} - X_{2,i}$ et considérons le test t (car nous ne connaissons pas la variance). La statistique de test est ainsi

$$T_0 = \frac{\overline{D}}{S_D/\sqrt{n}} \sim t(n-1), \text{ où } \overline{D} = \frac{1}{n}\sum_{i=1}^n D_i \text{ et } S_D^2 = \frac{1}{n-1}\sum_{i=1}^n (D_i - \overline{D})^2.$$

Exemple : les connaissances de $n = 10$ ingénieurs sur les concepts statistiques de base ont été mesurées sur une échelle de 0 à 100, avant et après un court cours sur le contrôle statistique de la qualité. Les résultats sont les suivants :

Individu	1	2	3	4	5	6	7	8	9	10
Avant $X_{1,i}$	43	82	77	39	51	66	55	61	79	43
Après $X_{2,i}$	51	84	74	48	53	61	59	75	82	48

Soient μ_1 et μ_2 le score moyen avant et après le cours, respectivement, les scores suivant des lois normales. Testez $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 < \mu_2$.

Solution : les différences $D_i = X_{1,i} - X_{2,i}$ sont :

Individu	1	2	3	4	5	6	7	8	9	10
Avant X_{1i}	43	82	77	39	51	66	55	61	79	43
Après X_{2i}	51	84	74	48	53	61	59	75	82	48
Différence D_i	-8	-2	3	-9	-2	5	-4	-14	-3	-5

La moyenne et la variance observées sont, respectivement, $\overline{d} = -3.9$ et $s_D^2 = 31.21$. La statistique de test est

$$T_0 = \frac{\overline{D} - 0}{S_D/\sqrt{n}} \sim t(n-1)\ ; \text{la valeur observée est } t_0 = \frac{-3.9}{\sqrt{31.21/10}} \approx -2.21.$$

Nous calculons

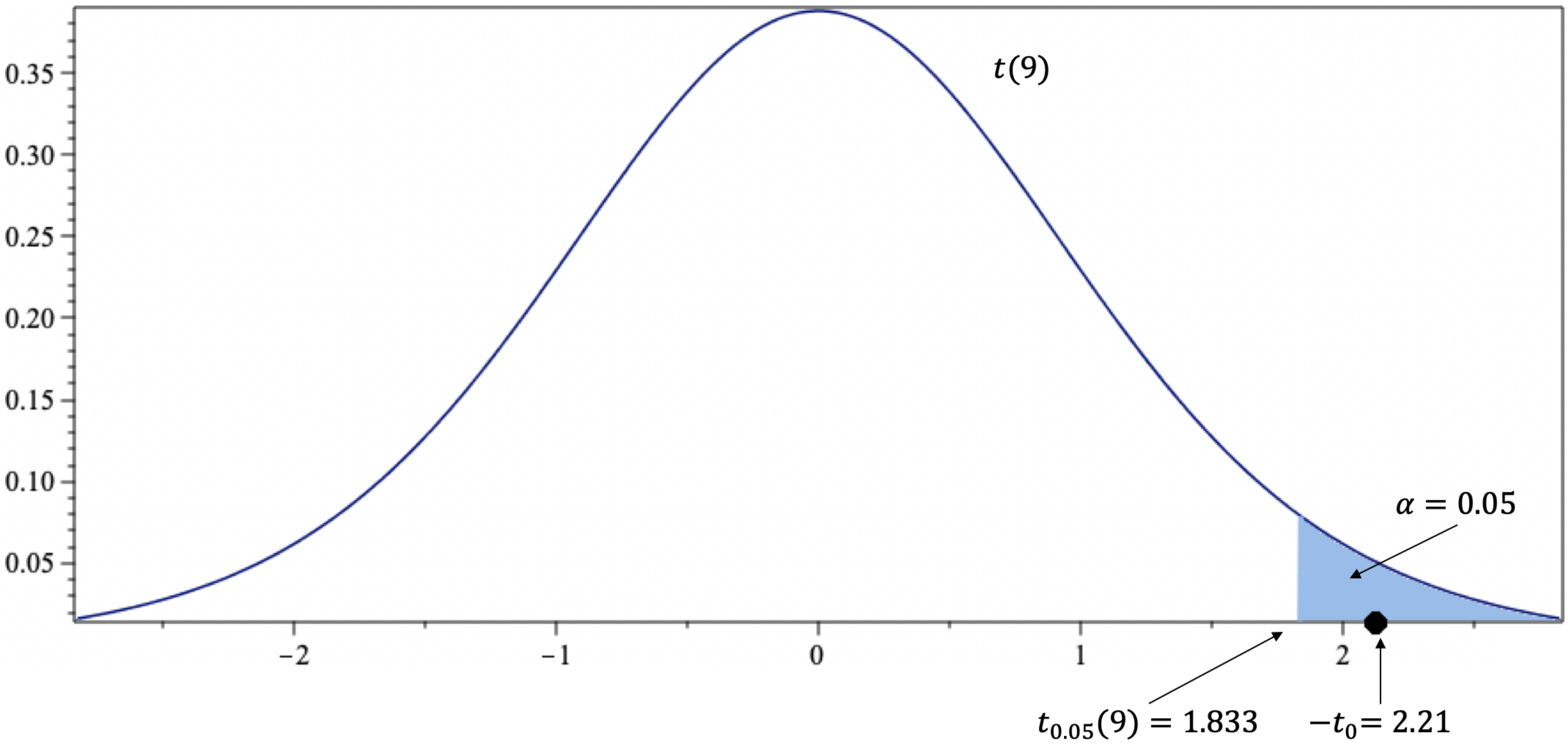
$$P(\overline{D} \leq \overline{d}) = P(\overline{D} \leq -3.9) = P(T(9) \leq -2.21) = P(T(9) > 2.21).$$

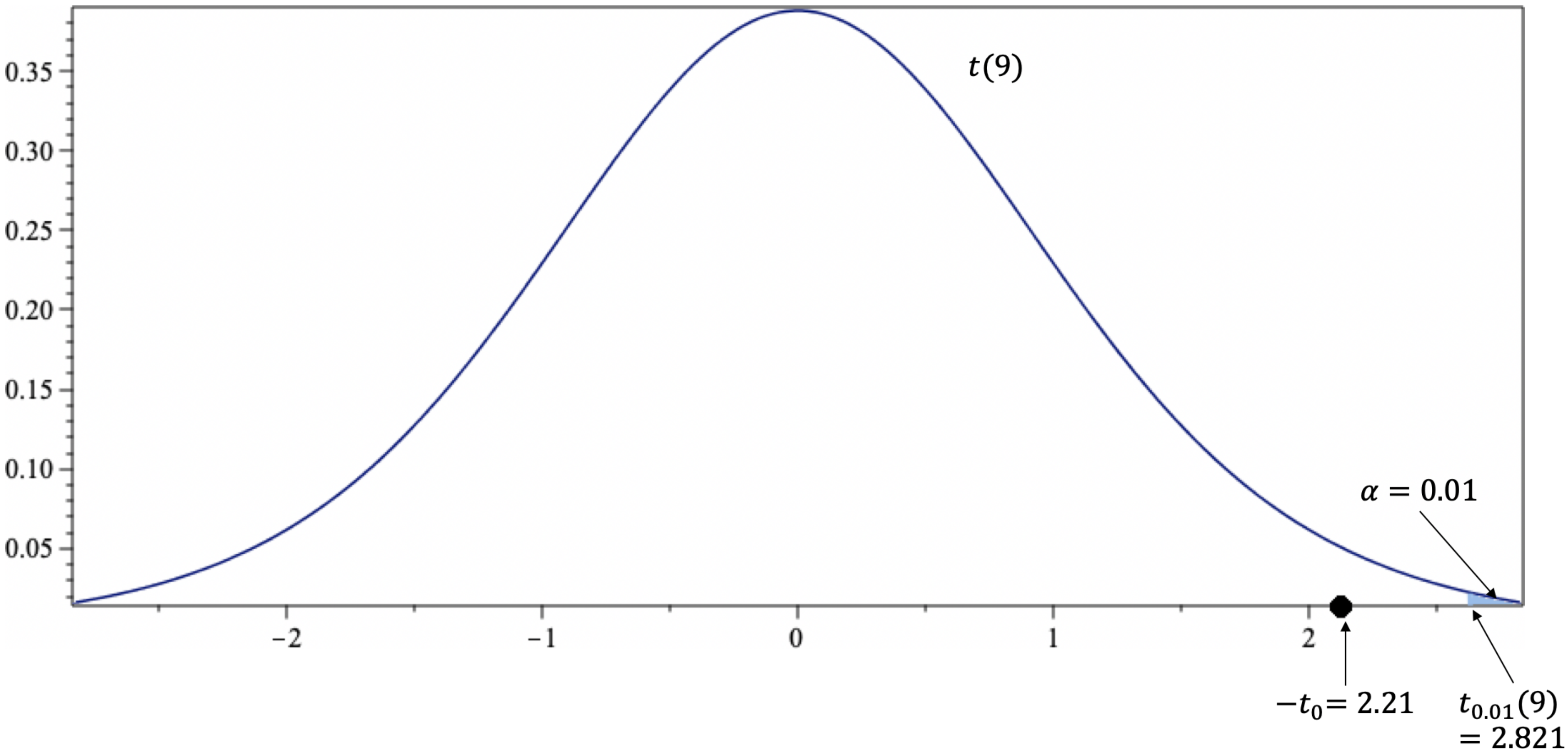
Mais

$$t_{0.05}(9) = 1.833 < t_0 = 2.21 < t_{0.01}(9) = 2.821,$$

ainsi nous **rejetons** H_0 lorsque $\alpha = 0.05$, mais **pas** H_0 lorsque $\alpha = 0.01$.

r	$t_{0.40}(r)$	$t_{0.25}(r)$	$t_{0.10}(r)$	$t_{0.05}(r)$	$t_{0.025}(r)$	$t_{0.01}(r)$	$t_{0.005}(r)$
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169


0.35
0.30
0.25
0.20
0.15
0.10
0.05
−2
−1
0
1
2
$t(9)$
$\alpha = 0.05$
$t_{0.05}(9) = 1.833$
$-t_0 = 2.21$


t(9)
α = 0.01
0.35
0.30
0.25
0.20
0.15
0.10
0.05
−2
−1
0
1
2
−t0= 2.21
t0.01(9)
= 2.821

6.8 – Les tests pour deux échantillons non appariés

Soit $X_{1,1},\ldots,X_{1,n}$ un échantillon aléatoire provenant d'une population normale de moyenne μ_1 et variance σ_1^2 inconnues ; soit $Y_{2,1},\ldots,Y_{2,m}$ un échantillon aléatoire provenant d'une population normale de moyenne μ_2 et de variance σ_2^2 inconnues, les deux populations étant **indépendantes**.

Nous voulons tester

$$H_0 : \mu_1 = \mu_2 \text{ par rapport à } H_1 : \mu_1 \neq \mu_2.$$

Soient $\overline{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$, $\overline{Y} = \frac{1}{m}\sum_{i=1}^{m} Y_i$. Les valeurs observées sont à nouveau désignées par des lettres minuscules : $\overline{x}$, $\overline{y}$.

1er cas: les variances σ_1^2 et σ_2^2 sont connues

Hypothèse alternative	**Région critique**
$H_1 : \mu_1 > \mu_2$	$z_0 > z_\alpha$
$H_1 : \mu_1 < \mu_2$	$z_0 < -z_\alpha$
$H_1 : \mu_1 \neq \mu_2$	$\lvert z_0 \rvert > z_{\alpha/2}$

où $z_0 = \frac{\overline{x}-\overline{y}}{\sqrt{\sigma_1^2/n+\sigma_2^2/m}}$, z_α est tel que $P(Z > z_\alpha) = \alpha$, et $Z \sim \mathcal{N}(0,1)$.

Hypothèse alternative	**Valeur**$-p$
$H_1 : \mu_1 > \mu_2$	$P(Z > z_0)$
$H_1 : \mu_1 < \mu_2$	$P(Z < z_0)$
$H_1 : \mu_1 \neq \mu_2$	$2 \cdot \min\{P(Z > z_0), P(Z < z_0)\}$

Exemple : un échantillon de $n = 100$ Albertains donne un revenu moyen de $\overline{X} = 33,000\$$. Un échantillon de $m = 80$ Ontariens donne un revenu moyen de $\overline{Y} = 32,000\$$. D'après des études antérieures, on sait que les écarts-types du revenu de la population sont, respectivement, $\sigma_1 = 5000\$$ en Alberta et $\sigma_2 = 2000\$$ en Ontario. Les Albertains gagnent-ils plus que les Ontariens, en moyenne ?

Solution : nous testons $H_0 : \mu_1 = \mu_2$ par rapport à $H_1 : \mu_1 > \mu_2$ (ce n'est pas la seule possibilité). La différence observée est $\overline{X} - \overline{Y} = 1000$; la statistique de test observée est

$$z_0 = \frac{\overline{X} - \overline{Y}}{\sqrt{\sigma_1^2/n + \sigma_2^2/m}} = \frac{1000}{\sqrt{5000^2/100 + 2000^2/80}} = 1.82;$$

la valeur$-p$ est $P\left(\overline{X} - \overline{Y} > 1000\right) = P(Z > 1.82) = 0.035$, et nous **rejetons** H_0 lorsque $\alpha = 0.05$, mais **pas** lorsque $\alpha = 0.01$.

2e cas: les variances σ_1^2 et σ_2^2 sont inconnues, avec de petites tailles d'échantillon

Hypothèse alternative	Région critique
$H_1 : \mu_1 > \mu_2$	$t_0 > t_\alpha(n+m-2)$
$H_1 : \mu_1 < \mu_2$	$t_0 < -t_\alpha(n+m-2)$
$H_1 : \mu_1 \neq \mu_2$	$\lvert t_0 \rvert > t_{\alpha/2}(n+m-2)$

où $t_0 = \frac{\overline{X}-\overline{Y}}{\sqrt{S_p^2/n+S_p^2/m}}$, $S_p^2 = \frac{(n-1)S_1^2+(m-1)S_2^2}{n+m-2}$, $t_\alpha(n+m-2)$ est tel que $P(T > t_\alpha(n+m-2)) = \alpha$, et $T \sim t(n+m-2)$.

Hypothèse alternative	Valeur$-p$
$H_1 : \mu_1 > \mu_2$	$P(T > t_0)$
$H_1 : \mu_1 < \mu_2$	$P(T < t_0)$
$H_1 : \mu_1 \neq \mu_2$	$2 \cdot \min\{P(T > t_0), P(T < t_0)\}$

Exemple : une chercheuse veut tester si, en moyenne, un nouvel engrais donne de plus grandes plantes. Les plantes ont été divisées en deux groupes : un groupe témoin traité avec l'ancien engrais et un groupe d'étude traité avec le nouvel engrais. Les données suivantes sont obtenues :

Taille	Moyenne	Variance
$n = 8$	$\overline{X} = 43.14$	$S_1^2 = 71.65$
$m = 8$	$\overline{Y} = 47.79$	$S_2^2 = 52.66$

Testez $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 < \mu_2$.

Solution : la différence moyenne observée est de $\overline{d} = -4.65$; la variance groupée observée est

$$S_p^2 = \frac{(n-1)S_1^2 + (m-1)S_2^2}{n+m-2} = \frac{7(71.65) + 7(52.66)}{8+8-2} = 62.155 = 7.88^2.$$

La statistique de test observée est

$$t_0 = \frac{\overline{X} - \overline{Y}}{\sqrt{S_p^2/n + S_p^2/m}} = \frac{-4.65}{7.88\sqrt{1/8 + 1/8}} = -1.18;$$

la valeur$-p$ correspondante est

$$P\left(\overline{X} - \overline{Y} < -4.65\right) = P(T(14) < -1.18) = P(T(14) > 1.18) \in (0.1, 0.25)$$

(selon la table), nous **ne rejetons** H_0 **ni** lorsque $\alpha = 0.05$, **ni** lorsque $\alpha = 0.01$.

r	$t_{0.40}(r)$	$t_{0.25}(r)$	$t_{0.10}(r)$	$t_{0.05}(r)$	$t_{0.025}(r)$	$t_{0.01}(r)$	$t_{0.005}(r)$
1	0.325	1.000	3.078	6.314	12.706	31.821	63.657
2	0.289	0.816	1.886	2.920	4.303	6.965	9.925
3	0.277	0.765	1.638	2.353	3.182	4.541	5.841
4	0.271	0.741	1.533	2.132	2.776	3.747	4.604
5	0.267	0.727	1.476	2.015	2.571	3.365	4.032
6	0.265	0.718	1.440	1.943	2.447	3.143	3.707
7	0.263	0.711	1.415	1.895	2.365	2.998	3.499
8	0.262	0.706	1.397	1.860	2.306	2.896	3.355
9	0.261	0.703	1.383	1.833	2.262	2.821	3.250
10	0.260	0.700	1.372	1.812	2.228	2.764	3.169
11	0.260	0.697	1.363	1.796	2.201	2.718	3.106
12	0.259	0.695	1.356	1.782	2.179	2.681	3.055
13	0.259	0.694	1.350	1.771	2.160	2.650	3.012
14	0.258	0.692	1.345	1.761	2.145	2.624	2.997
15	0.258	0.691	1.341	1.753	2.131	2.602	2.947
16	0.258	0.690	1.337	1.746	2.120	2.583	2.921
17	0.257	0.689	1.333	1.740	2.110	2.567	2.898
18	0.257	0.688	1.330	1.734	2.101	2.552	2.878
19	0.257	0.688	1.328	1.729	2.093	2.539	2.861
20	0.257	0.687	1.325	1.725	2.086	2.528	2.845

3e cas: les variances σ_1^2 et σ_2^2 sont inconnues, avec de grandes tailles d'échantillon

Hypothèse alternative	**Région critique**
$H_1 : \mu_1 > \mu_2$	$z_0 > z_\alpha$
$H_1 : \mu_1 < \mu_2$	$z_0 < -z_\alpha$
$H_1 : \mu_1 \neq \mu_2$	$\lvert z_0 \rvert > z_{\alpha/2}$

où $z_0 = \frac{\overline{X}-\overline{Y}}{\sqrt{S_1^2/n+S_2^2/m}}$, z_α est tel que $P(Z > z_\alpha) = \alpha$, et $Z \sim \mathcal{N}(0,1)$.

Hypothèse alternative	**Valeur**$-p$
$H_1 : \mu_1 > \mu_2$	$P(Z > z_0)$
$H_1 : \mu_1 < \mu_2$	$P(Z < z_0)$
$H_1 : \mu_1 \neq \mu_2$	$2 \cdot \min\{P(Z > z_0), P(Z < z_0)\}$

Exemple : on répéte l'exemple précédent, mais avec des tailles d'échantillons plus élevés: $n = m = 100$. Testez $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 < \mu_2$.

Solution : la différence observée est (toujours) de $\overline{d} = -4.65$. La statistique de test observée devient cependant

$$z_0 = \frac{\overline{X} - \overline{Y}}{\sqrt{S_1^2/n + S_2^2/m}} = \frac{-4.65}{\sqrt{71.65^2/100 + 52.66^2/100}} = -4.17;$$

la valeur$-p$ correspondante est

$$P\left(\overline{X} - \overline{Y} < -4.65\right) = P(Z < -4.17) \approx 0.0000;$$

nous **rejetons** H_0 **et** lorsque $\alpha = 0.05$, **et** lorsque $\alpha = 0.01$.

6.9 – La différence de deux proportions

Comme toujours, nous pouvons transférer ces tests aux proportions, en utilisant l'approximation normale de la loi binomiale.

Exemple : considérez les proportions de papillons recapturés dans des populations claires (p_1) et foncées (p_2). Parmi les $n_1 = 137$ papillons de couleur claire, $y_1 = 18$ ont été recapturés ; parmi les $n_2 = 493$ papillons de nuit, $y_2 = 131$ ont été recapturés. Y a-t-il une différence significative entre la proportion de papillons recapturés dans les deux populations ?

Solution : nous testons $H_0 : p_1 = p_2$ vs. $H_1 : p_1 \neq p_2$. Les proportions observées sont

$$\hat{p}_1 = \frac{y_1}{n_1} = 0.131, \ \hat{p}_2 = \frac{y_2}{n_2} = 0.266, \quad \text{et} \quad \hat{p}_1 - \hat{p}_2 = -0.135.$$

La valeur$-p$ correspondante est $2 \cdot P(\hat{p}_1 - \hat{p}_2 \leq -0.135)$:

$$2 \cdot P\left(\frac{\hat{p}_1 - \hat{p}_2 - 0}{\sqrt{\hat{p}(1-\hat{p})}\sqrt{1/n_1 + 1/n_2}} \leq \frac{-0.135 - 0}{\sqrt{0.2365(1-0.2365)}\sqrt{1/137 + 1/493}}\right).$$

Nous avons ainsi

$$2 \cdot P(\hat{p}_1 - \hat{p}_2 \leq -0.135) \approx 2P(Z < -3.29) \approx 0.0000,$$

et nous **rejetons** H_0 lorsque $\alpha = 0.05$ ou $\alpha = 0.01$.

Note : la quantité $\hat{p}$ est la **proportion combinée**

$$\hat{p} = \frac{n_1}{n_1 + n_2}\hat{p}_1 + \frac{n_2}{n_1 + n_2}\hat{p}_2.$$

6.10 – Les tests d'hypothèses avec R

- `t.test(x,mu=5)` est un test de $H_0 : \mu = 5$ vs. $H_1 : \mu \neq 5$ lorsque σ est inconnu (test t)

- `t.test(x,mu=5,alternative="greater")` est un test de $H_0 : \mu = 5$ vs. $H_1 : \mu > 5$ lorsque σ est inconnu (test t)

- `t.test(x,mu=5,alternative="less")` est un test de $H_0 : \mu = 5$ vs. $H_1 : \mu < 5$ lorsque σ est inconnu (test t)

- `t.test(x,y,var.equal=TRUE)` est un test de $H_0 : \mu_1 = \mu_2$ par rapport à $H_1 : \mu_1 \neq \mu_2$ lorsque nous avons deux échantillons indépendants, avec des variances égales mais inconnues

- `t.test(x,y,var.equal=TRUE,alternative="greater")` est un test de $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 > \mu_2$ lorsque nous avons deux échantillons indépendants, avec des variances égales mais inconnues

- `t.test(x,y,var.equal=TRUE,alternative="less")` est un test de $H_0 : \mu_1 = \mu_2$ vs. $H_1 : \mu_1 < \mu_2$ lorsque nous avons deux échantillons indépendants, avec des variances égales mais inconnues

Exemple :

```
> x=c(4,5,4,6,4,4,5)
> t.test(x,mu=5)

One Sample t-test
data:  x
t = -1.4412, df = 6, p-value = 0.1996
alternative hypothesis: true mean is not equal to 5
95 percent confidence interval:
 3.843764 5.299093
sample estimates:
mean of x
 4.571429
```

Ici, nous ne parvenons pas à rejeter l'hypothèse nulle selon laquelle la vraie moyenne est de 5.

Exemple :

```
> x=c(1,2,1,4,3,2,4,3,2)
> t.test(x,mu=5)

One Sample t-test

data:  x
t = -6.7823, df = 8, p-value = 0.0001403
alternative hypothesis: true mean is not equal to 5
95 percent confidence interval:
 1.575551 3.313338
sample estimates:
mean of x
 2.444444
```

Qu'en est-il ici ?