

# **MODULE 10**

## **TESTS D'HYPOTHESES**

A decorative banner image featuring a calculator, a receipt with numbers like 852.97 and 423.99, a large dollar sign, and a blue bar chart.

**PAF-1010**

Analyse quantitative de problèmes de gestion

**Louis Houde**  
**Département de Mathématiques et d'informatique**  
**Université du Québec à Trois-Rivières**

# MODULE 10 Tests d'hypothèses

La notion de tests d'hypothèses est très importante en statistique, c'est le complément à l'estimation statistique : après avoir obtenu une approximation du paramètre d'intérêt, il est essentiel de le comparer à des standards ou de comparer plusieurs populations du point de vue de ce paramètre. Le principe est de diviser toutes les possibilités en deux catégories, l'hypothèse nulle notée  $H_0$  et l'alternative, notée  $H_1$ . C'est le test statistique qui permet, sur la base des observations, de décider en faveur de l'hypothèse la plus raisonnable :  $H_0$  ou  $H_1$ .

Le test statistique est une façon de faire une comparaison éclairée sur la base de l'observations d'un échantillon. Le tout étant basé sur la théorie des probabilités, il ne sera pas possible de savoir si la décision prise à la suite d'une test d'hypothèses est la bonne mais il est cependant possible de contrôler certaines des erreurs et de les maintenir à des taux acceptables dans le contexte. En cela, les tests d'hypothèses ne prouvent jamais rien mais ils permettent de comparer objectivement des possibilités. Dans ce contexte il est clair que la citation suivante "Il est prouvé statistiquement que...." que l'on retrouve dans beaucoup d'articles ou de communiqués non scientifiques est une aberration. On ne prouve pas mais on observe et c'est déjà bien.

## Objectifs et compétences

L'objectif de cette partie est de donner à l'étudiant les outils nécessaires pour appréhender les variations échantillonnales et les appliquer au problème du test d'hypothèses. L'étudiant devra identifier les paramètres d'intérêt et choisir la méthode appropriée.

L'étudiant sera en mesure de

- déterminer le paramètre et les hypothèses statistiques.
- déterminer le niveau du test en fonction de la puissance et des conséquences.
- effectuer les calculs pour les principaux tests.
- établir la conclusion en fonction des résultats du test d'hypothèses.

## 10.1 Hypothèses statistiques

Un test statistique découle souvent d'une question générale ou d'une hypothèse de travail. Prenons l'exemple d'une compagnie qui produit des pièces de précision à l'aide de machines automatisées. Le contexte et la spécificité des pièces impliquent un taux de rejet élevé qui peut être aussi grand que 10%. L'ingénieur responsable de la fabrication de pièces de précision doit

vérifier régulièrement si les machines sont bien calibrées. Le terme "bien calibrées" peut dans ce cas-ci s'interpréter comme étant un taux acceptable de rejet de 10%. Si c'est le cas, le test se fera en observant  $n$  pièces sortant de la machine et les hypothèses statistiques seront

$$\begin{cases} H_0 : \pi = 0.1 \\ H_1 : \pi > 0.1 \end{cases}$$

où  $\pi$  représente la probabilité qu'une pièce fabriquée par la machine soit rejetée. Si, à la suite du test c'est l'hypothèse  $H_0$  qui est retenue, la machine sera considérée comme bien calibrée tandis que si c'est l'hypothèse  $H_1$  qui est retenue alors la machine sera considérée comme mal calibrée.

Cet exemple illustre un caractère très important des tests d'hypothèses : on ne peut faire de test que sur des mesures numériques et celles-ci sont réductrices. La transcription de l'hypothèse générale en hypothèses sur un paramètre de la population,  $H_0$  et  $H_1$ , est une étape nécessaire pour utiliser les outils statistiques que sont les tests d'hypothèses. Cette transcription n'est pas unique et elle dépend de ce qui "intéresse" l'utilisateur.

Dans l'exemple sur les pièces de précision une première modélisation est de comparer le taux de rejet. Il est cependant possible de considérer les pièces produites et de décider que la machine est bien calibrée si la moyenne sur la déviance maximale entre la pièce et l'étalon est de moins de 0.1mm. On a alors un test sur le paramètre  $\mu$ .

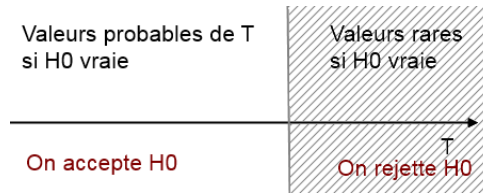
Si la pièce est un microprocesseur, il est possible de mesurer le "bruit" électronique que génère le microprocesseur à la sortie. Une machine peut alors être considérée comme mal calibrée si le bruit moyen des microprocesseurs qu'elle fabrique est de plus de 3 erreurs/microsec. Les hypothèses statistiques qui permettent de vérifier si la machine est bien calibrée sont  $H_0 : \mu = 3$  contre  $H_1 : \mu > 3$ , où  $\mu$  représente le bruit moyen des microprocesseurs.

Ces exemples de transcription des hypothèses de travail en hypothèses basées sur des paramètres statistiques démontrent que pour un même problème général il peut y avoir plusieurs interprétations possibles. Les comparaisons sont cependant basées sur les mêmes principes :

- deux hypothèses complémentaires concernant le paramètre
- acceptation d'une ou de l'autre à la suite de l'observation
- impossibilité de savoir quelle hypothèse est réellement vraie. Il est en effet rare de prendre une décision statistique et d'avoir la possibilité de connaître la "vérité" du moins à court terme.

Un test statistique est basé sur une variable aléatoire observable  $T$  et  $\theta$  un paramètre. On définit les valeurs pour lesquels la statistique  $T$  est probable sous  $H_0$  et les valeurs rares pour

cette statistique. C'est ce qui permet de choisir entre les deux hypothèses :



On doit connaître la loi de probabilité de la statistique  $T$  si  $H_0$  est vraie pour faire le test.

Dans l'exemple sur la calibration, la première modélisation est associée à une loi binomiale donnant 1 si la pièce est rejetée et 0 sinon tandis que le paramètre à "tester" est  $\pi$  la probabilité de "succès" de la loi binomiale. Dans la deuxième modélisation la v.a. est la déviance maximale par rapport à l'étalon et le paramètre est la moyenne de cette v.a. Les paramètres les plus souvent utilisés sont la moyenne, la probabilité de succès et la variance mais d'autres sont aussi possibles.

Un test d'hypothèses est une procédure qui consiste à vérifier à l'aide d'un échantillon si le paramètre est égal à une valeur donnée, disons  $\theta_0$ . L'idée est de poser comme hypothèse a priori l'égalité entre le paramètre  $\theta$  la valeur  $\theta_0$  puis de regarder des observations provenant de la population pour décider si cet a priori est réaliste. Si tel est le cas l'hypothèse a priori est considérée vraie et sinon elle est rejetée au profit de l'hypothèse complémentaire ou alternative.

Ainsi pour faire un test d'hypothèses, on considère une hypothèse nulle  $H_0$ , aussi appelée hypothèse principale, que l'on suppose vraie. L'hypothèse  $H_1$  appelée hypothèse rivale, contre-hypothèse ou hypothèse alternative, est l'hypothèse qui sera acceptée si on rejette  $H_0$ . L'hypothèse nulle fait **généralement** intervenir un signe d'égalité..

**Exemple 1.1** ★ On s'intéresse au taux de polluants dans la fabrication d'une composante électronique. Supposons que le taux acceptable est de 75ppm (parties par million), on s'intéresse au paramètre  $\mu$  qui donne le taux moyen du processus de fabrication, l'hypothèse nulle est  $H_0 : \mu = 75$  et l'alternative est  $H_1 : \mu > 75$

**Exemple 1.2** ★ On veut savoir si une pièce de monnaie est truquée. Les hypothèses sont

$$H_0 : \pi = 1/2$$

$$H_1 : \pi \neq 1/2$$

où  $\pi$  représente la probabilité d'obtenir un "face" au lancer de la pièce.

**Exemple 1.3** ★ Une usine de d'embouteillage doit mettre 10 onces de liquide dans chaque bouteille. Il n'est pas raisonnable de penser qu'il y aura exactement 10.000 onces dans chaque

bouteille mais on veut au moins qu'il y ait en moyenne 10 onces. Un contrôle de qualité consiste à vérifier si la quantité de liquide est bien 10 onces en moyenne. En pratique il faudra prendre un échantillon de bouteilles remplies puis de tester les hypothèses

$$H_0 : \mu = 10$$

$$H_1 : \mu \neq 10$$

où  $\mu$  représente la moyenne en onces de la v.a. représentant la quantité de liquide dans une bouteille au hasard.

Dans ce dernier exemple, une valeur moyenne plus petite que 10 lèse le consommateur tandis qu'une valeur plus grande que 10 lèse l'embouteilleur. Dans un cas comme dans l'autre la machine devrait être recalibrée. Dans l'exemple des composantes électroniques c'est le fait d'avoir trop de pollution qui est dommageable et ainsi on veut détecter uniquement cette alternative. On a donc une hypothèse alternative qui reflète cela :  $H_1 : \mu > 75$ . On dit dans le premier cas que l'hypothèse alternative est bilatérale et dans le deuxième cas elle est unilatérale.

- Soit  $\theta$  un paramètre et  $\theta_0$  une valeur fixée, les hypothèses statistiques sont **bilatérales** si elles sont de la forme

$$H_0 : \theta = \theta_0$$

$$H_1 : \theta \neq \theta_0$$

et elle sont **unilatérales** si elles sont de la forme

$$\begin{array}{ll} H_0 : \theta = \theta_0 & \text{ou} & H_0 : \theta = \theta_0 \\ H_1 : \theta < \theta_0 & & H_1 : \theta > \theta_0 \end{array}$$

Le test d'hypothèse doit conduire à une et une seule décision et cette décision est aléatoire.

**Définition 1.1** Un test statistique est une règle de décision qui permet de décider en faveur de  $H_0$  ou de  $H_1$  selon les observations.

Un embouteilleur veut vérifier si les bouteilles qu'il prépare contiennent bien 10 onces en moyenne. Il veut confronter les hypothèses

$$H_0 : \mu = 10$$

$$H_1 : \mu \neq 10$$

Il décide de prendre 10 bouteilles par heure et si 2 des 10 bouteilles contiennent moins de 9 onces ou plus que 11 onces alors l'hypothèse  $H_0$  sera rejetée et  $H_1$  acceptée. Cette règle de décision est un test statistique parce qu'il y a deux hypothèses complémentaires en jeu et une

règle de décision qui dépend des valeurs observées.

Cet exemple montre aussi une caractéristique importante des tests d'hypothèses : ils sont aléatoires et il y a une possibilité de se tromper. Dans un échantillon de 10 bouteilles il est possible, mais rare, de trouver 2 de ces bouteilles. Dans un tel cas l'hypothèse  $H_0$  sera rejetée et la machine sera déclarée non calibrée. Or elle peut être bien calibrée. C'est un type d'erreur qu'il faut accepter mais surtout contrôler.

La **statistique du test** est une quantité basée sur les valeurs de l'échantillon et sur laquelle le test est basé.

Dans le cas du test rejetant  $H_0$  s'il y a plus de 2 bouteilles sur 10 qui sont "extrêmes" la statistique est le nombre de bouteilles de moins de 9 onces ou de plus de 11 onces.

**Définition 1.2** La région critique d'un test est l'ensemble des valeurs de l'échantillon pour lesquels on rejette l'hypothèse nulle, on dit que c'est la région de rejet de l'hypothèse nulle associée au test.

Dans l'exemple des bouteilles la région critique est 2, 3, 4, ..., 10, soit le nombre de bouteilles trop ou pas assez remplies.

## 10.2 Niveau et puissance

Lorsqu'on confronte deux hypothèses statistiques à l'aide d'un test, on doit considérer qu'il y a la réalité (la valeur réelle du paramètre) et la décision qui est prise. Cela donne le tableau suivant :

		Décision	
		Rejeter $H_0$	Accepter $H_0$
Réalité	$H_0$ vraie	Erreur de 1 <sup>ère</sup> espèce	ok
	$H_1$ vraie	ok	Erreur de 2 <sup>ème</sup> espèce

L'erreur à contrôler est le fait de rejeter  $H_0$  étant donné que cette hypothèse est vraie. Cela correspond, dans le cas du test sur les bouteilles à dire que la machine n'est pas bien calibrée tandis qu'elle l'est dans la réalité. Il est possible de fixer la probabilité de cette erreur a priori à une valeur acceptable dans le contexte.



**Définition 2.1** Soit un test statistique. Le niveau du test noté  $\alpha$ , est la probabilité de faire une erreur de première espèce,

$$\alpha = \Pr(\text{rejeter } H_0 \mid H_0 \text{ vraie})$$

D'un autre côté il est important de pouvoir découvrir que l'hypothèse nulle n'est pas vraie si tel est le cas. Cette probabilité est la puissance du test. Puissance dans le sens de "capacité" à déceler un écart à  $H_0$ .

La **puissance** du test est donnée par 1 moins la probabilité de faire une erreur de deuxième espèce,

$$1 - \beta = \Pr(\text{rejeter } H_0 \mid H_1 \text{ vraie})$$

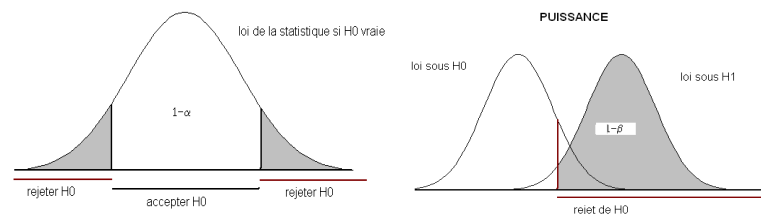
Dans le cas des bouteilles on a les éléments suivant :

- le niveau du test est la probabilité de dire que la machine est mal calibrée étant donné qu'elle est **effectivement bien calibrée**
- la puissance du test est la probabilité de réussir à **découvrir** que la machine est mal calibrée.

La théorie relative aux tests d'hypothèses est basée sur le fait que le niveau du test est un choix de l'utilisateur et généralement les valeurs utilisées sont 10%, 5% ou 1% selon l'importance de cette erreur dans le contexte particulier du problème.

Dans certains cas il faut aussi considérer le fait que la puissance et le niveau sont fortement liés puisque si le niveau augmente, la puissance augmente et si le niveau diminue, la puissance diminue pour une même taille d'échantillon. C'est donc qu'il faut prendre en considération la puissance dans le choix du niveau du test.

Le niveau et la puissance s'illustrent par les deux graphiques suivants :



Dans le graphique de gauche la zone grise indique la région de rejet et la probabilité  $\alpha$  associée. Dans le second on considère que la réalité est  $H_1$  et ainsi la zone ombrée indique  $1 - \beta$  soit la probabilité de rejeter  $H_0$  (zone de rejet) en considérant la distribution sous  $H_0$ .

**Exemple 2.1** ★ Une grande entreprise veut savoir si les hommes ont autant de chances que

les femmes pour l'obtention d'un poste. Les hypothèses à confronter sont : il y a autant d'hommes que de femmes engagés ( $H_0$ ) contre il y a une probabilité différente pour les hommes et les femmes d'être engagés ( $H_1$ ).

- Le niveau du test (ou le seuil de signification) est la probabilité de rejeter  $H_0$  étant donné  $H_0$  c'est-à-dire la probabilité de dire qu'il y a une probabilité différente (discrimination) étant donné qu'il n'y en a pas.
- La puissance est la probabilité de détecter qu'il y a une discrimination s'il y en a effectivement une.

**Exemple 2.2** ★ Une entreprise spécialisée dans la commercialisation de produits naturels veut promouvoir un produit stimulant les facultés mentales. Pour appuyer ses arguments de vente elle demande à un statisticien de faire un test pour vérifier si le produit est efficace sur la mémoire. L'expérience consiste à choisir 100 personnes au hasard et à quantifier leur mémoire à l'aide d'un test.

Chaque participant utilise le produit pendant 5 semaines puis refait le test de mémoire. À la fin de l'expérimentation on considère le nombre de personnes qui ont augmenté leur score au test de mémoire.

Les hypothèses statistiques permettant de vérifier l'efficacité du produit sont

$$H_0 : \pi = 0.5$$

$$H_1 : \pi > 0.5$$

où  $\pi$  est la probabilité qu'un individu augmente<sup>1</sup> son score au test de mémorisation après 5 semaines d'utilisation du produit.

Le niveau du test est défini par  $\alpha$  = probabilité de rejeter  $H_0$  étant donné que  $H_0$  est vraie. Cela veut dire "la probabilité de dire que le produit est efficace étant donné qu'il n'est pas efficace". Cette probabilité doit prendre une valeur de 10%, 5% ou 1%.

Pour déterminer une valeur raisonnable il faut considérer la conséquence de cette erreur ainsi que l'effet du niveau sur la puissance. Cette dernière est la probabilité de rejeter  $H_0$  étant donné que  $H_1$  est vraie, c'est-à-dire la probabilité de dire que le test n'est pas efficace étant donné qu'il est efficace.

Il y a deux points de vue à considérer

- Le promoteur du produit : celui-ci considère que le niveau doit être le plus grand possible puisque c'est une erreur relativement bénigne que de déclarer faussement le produit efficace considérant le manque de puissance. Il veut que si le produit est efficace il sera déclaré comme tel. Un niveau de 10% devrait donner une puissance suffisante.

<sup>1</sup> Cette hypothèse est commune pour des expériences dont la mesure est le fait d'avoir augmenter un indice après un traitement. Si la mesure est continue et que le traitement n'a pas d'effet, alors il y aura autant de chance d'augmenter que de diminuer l'indice donc  $\pi = 1/2$ . Par exemple, si on utilise une balance précise deux fois de suite, il y aura une certaine différence entre les poids pour les deux pesés. Il y a autant de chance que le poids soit plus petit lors de la deuxième pesée que plus grand.



- L'utilisateur du produit : l'utilisateur ne veut pas se faire refiler, à prix fort, un produit sans effet. Il veut donc un niveau le plus petit possible. Par contre il sera intéressé à acheter un produit pouvant effectivement augmenter sa mémoire. La puissance devrait donc être assez bonne. Un compromis est de prendre un niveau de 5%. L'erreur de première espèce est relativement petite et la puissance correcte.

**Remarque 2.1** Le niveau du test étant fixe, la taille de l'échantillon n'affecte pas cette quantité. Pour ce qui est de la puissance, plus la taille est grande plus la puissance est grande. Cette observation est très intuitive ; une expérience consistant à vérifier si une pièce de monnaie est équilibrée a peu de chance de détecter un biais après 3 tirages mais le test sera très puissant après 10000 expérimentations avec la même pièce.

**Exemple 2.3** ★ On veut savoir si une pièce de monnaie est truquée en faveur de "face". Les hypothèses sont

$$H_0 : \pi = 1/2$$

$$H_1 : \pi \neq 1/2$$

Le niveau du test est

$$\alpha = \Pr(\text{rejeter } H_0 \mid H_0)$$

soit la probabilité de décider que la pièce est non truquée étant donné qu'elle est bien équilibrée. La puissance est la probabilité de décider que la pièce est truquée étant donné qu'elle est effectivement truquée.

**Exemple 2.4** ★ Les machines vidéo-poker de loto Québec doivent redistribuer exactement 92% des mises comme gain. Le 92% correspond à un gain moyen chez le joueur. Sur 43570 mises de 25¢ répertoriées sur 10 machines dans le casino de Trois-Rivières on cherche à vérifier si la machine est correcte. Quelles sont les hypothèses ?

Solution : On veut confronter

$$H_0 : \mu = 0.92 * 0.25$$

$$H_1 : \mu \neq 0.92 * 0.25$$

soit «l'affirmation est bonne» contre «l'affirmation n'est pas valide».

### 10.3 Démarche pour faire un test statistique

Les tests statistiques sont tous exécutés de la même façon : déterminer ce qu'on veut faire, observer un échantillon et finalement prendre une décision. Il est possible de définir 6 étapes distinctes :

1. établir les hypothèses de travail
2. définir les hypothèses statistiques
3. déterminer le niveau  $\alpha$  et l'interpréter
4. déterminer le test statistique et faire les hypothèses supplémentaires en fonction des conditions d'application
5. observer les valeurs et faire les calculs
6. prendre une décision et conclure

Le passage de l'étape 1 à l'étape 2 est appelé la modélisation. Il est très important puisqu'il dicte la procédure statistique c'est-à-dire le test qui devra être exécuté. Il est aussi et surtout limitatif quant à la conclusion final de la procédure.

**Exemple 3.1** ★ Un ingénieur en charge du contrôle de la qualité doit faire un test pour vérifier la calibration de la machine remplissant les bouteilles à 10 onces. Il a comme hypothèses générales "la machine est bien calibrée" ou "la machine n'est pas bien calibrée". Si le paramètre choisi est la moyenne  $\mu$  alors les hypothèses seront

$$H_0 : \mu = 10$$

$$H_1 : \mu \neq 10$$

Ce test ne regardera pas les écarts entre la moyenne et les valeurs mais le total de liquide sur l'ensemble des bouteilles.

Dans l'exemple précédant on s'intéresse au paramètre  $\mu$  et on fait l'association entre la calibration et une moyenne de 10. Supposons que la machine donne beaucoup de bouteilles de moins de 8 onces et beaucoup de bouteilles de plus de 12 onces. Ce n'est certainement pas une bonne calibration mais avec la modélisation du problème basée uniquement sur le paramètre  $\mu$ , cela peut se produire et ce ne sera pas détecté.

Une façon de palier ce problème est de faire aussi un test sur le paramètre variance pour s'assurer que la variation du liquide est relativement petite. Des hypothèses du type

$$H_0 : \sigma^2 = 0.04$$

$$H_1 : \sigma^2 > 0.04$$

permettrait de s'assurer que la variance est raisonnable<sup>2</sup>.

On peut aussi définir une bonne calibration comme étant une probabilité faible de tomber sur

<sup>2</sup> Le paramètre  $\sigma^2 = 0.04$  est purement arbitraire pour illustrer le concept. Dans un problème réel il faudrait s'assurer que la valeur représente réellement la borne acceptable.

une bouteille qui n'en donne pas suffisamment au client soit une probabilité de moins de 10% d'obtenir une bouteille remplie sous les 9.9 onces. Les hypothèses seraient alors

$$H_0 : \pi = 0.1$$

$$H_1 : \pi < 0.1$$

où  $\pi$  est la probabilité d'obtenir une bouteille contenant moins de 9.9 onces.

Chacune de ces définitions est une façon de regarder la bonne calibration des machines qui font le remplissage.

**Exemple 3.2** ★ On veut comparer la valeur moyenne de vente des maisons en région rurale en Ontario et au Québec. Pour ce faire on prend un échantillon de ventes dans chacune des provinces et on veut confronter les hypothèses

$$H_0 : \mu_Q = \mu_O$$

$$H_1 : \mu_Q \neq \mu_O$$

On aura une comparaison de la valeur moyenne des maisons entre les deux provinces.

**Remarque 3.1** L'exemple précédant est parfaitement en accord avec la définition d'une hypothèse statistique puisqu'elle peut s'exprimer comme

$$H_0 : \mu_Q - \mu_O = 0$$

$$H_1 : \mu_Q - \mu_O \neq 0$$

c'est-à-dire l'égalité d'un paramètre ( $\mu_Q - \mu_O$ ) à une constante (0).

## 10.4 Hypothèses simples et composées

Dans la section précédente il est fait mention de la complémentarité des hypothèses statistique  $H_0$  et  $H_1$ . Or les hypothèses unilatérales ne sont pas complémentaires puisqu'on confronte les hypothèses suivantes :

$$H_0 : \pi = 1/2$$

$$H_1 : \pi > 1/2$$

Elles ne sont pas complémentaires puisque toutes les valeurs de  $\pi$  sous 1/2 ne sont pas représentées et il est certainement plausible que ce soit le cas dans plusieurs problèmes : un médicament naturel qui a supposément un effet d'augmenter la mémoire à court terme pourrait être dans la réalité une substance qui a l'effet contraire.

Lorsqu'on définit l'hypothèse nulle avec une égalité on est en présence d'une hypothèse sim-

ple<sup>3</sup> donc une seule possibilité. Cela est assez facile dans ce contexte de calculer le niveau et de le fixer puisqu'il n'y a qu'une possibilité. Si on pose une hypothèse du type  $H_0 : \pi \leq 1/2$  alors il y a une infinité de valeurs possibles pour le paramètre sous cette hypothèse. On dit que c'est une hypothèse composée c'est-à-dire qu'il y a plus qu'une valeur possible pour le paramètre. On aurait donc plusieurs niveaux, un pour chaque valeur du paramètre. Comment fixer une valeur si elle peut prendre une infinité de valeurs possibles.

Théoriquement cette multiplicité des niveaux se résout par la notion mathématique de "supremum". Or cette notion est plus théorique que pratique puisque généralement les tests qui sont définis en fonction de ce concept sont les mêmes que ceux qui supposent l'égalité du paramètre pour  $H_0$ . Comme les prérequis mathématiques pour travailler avec les hypothèses du type  $H_0 : \theta = \theta_0$  contre  $H_1 : \theta < \theta_0$  sont à la portée de tous il est donc préférable de s'en tenir à celles-ci.

Plusieurs auteurs donnent les tests pour les hypothèses bilatérales avec les signes  $\leq$  ou  $\geq$  sous  $H_0$  et ce sont les mêmes que ceux qui sont donnés dans la suite des notes. Ce n'est qu'une question de philosophie : nous préférons donner les formulations pour lesquelles les étudiants sont capables de faire le raisonnement complet même s'il y a manque de cohérence sur la complémentarité plutôt que d'avoir un manque de cohérence au niveau de la définition du concept de niveau. Il faut bien remarquer que dans les deux cas le test reste le même et les conclusions de l'analyse statistique restent aussi les mêmes donc c'est un choix.

La puissance est une hypothèse composée : il y a une infinité de possibilités dans tous les cas. Prenons le cas des hypothèses  $H_0 : \pi = 1/2$  contre  $H_1 : \pi > 1/2$  pour l'exemple d'une pièce de monnaie possiblement truquée en faveur du "face". C'est certain que si elle est truquée mais que la probabilité d'obtenir un "face" est de 0.5 il sera alors très difficile de faire la différence entre cela et une pièce équilibrée (puissance faible). Si, par contre, la probabilité d'un "face" est de 90% alors ce sera très facile de le démontrer par un échantillonnage de lancers (puissance très grande).

La puissance est donc une fonction qui dépend de l'alternative sous laquelle on l'analyse. On représente la puissance par un graphique qui donne pour les différentes alternatives les valeurs de la probabilité de rejeter  $H_0$ .

## 10.5 Tests pour un échantillon

On considère le contexte d'une population de référence dans laquelle on a mesuré une variable d'intérêt. C'est le contexte d'un échantillon et une variable. Les hypothèses statistiques sont

<sup>3</sup> Cette notion de hypothèse simple n'est pas exactement la même que celle définie dans les textes de statistique avancés mais elle est adaptée à la discussion.

de la forme

$$H_0 : \theta = \theta_0$$

pour  $\theta$  un paramètre simple contre une hypothèse bilatérale ou unilatérale concernant le même paramètre. Les tests concernant les différents paramètres sont étudiés dans les prochaines sections.

## Test pour une proportion

Le test pour une proportion se fait lorsque le paramètre est la probabilité d'un événement et que cette probabilité est supposée constante lorsqu'on répète l'expérience.

**Exemple 5.1** ★ On veut vérifier si la probabilité de tomber sur face est bien  $\frac{1}{2}$  lors du lancer d'une pièce de monnaie. Les hypothèses sont

$$\begin{aligned} H_0 &: \pi = \frac{1}{2} \\ H_1 &: \pi \neq \frac{1}{2} \end{aligned}$$

**Exemple 5.2** ★ On sait que le taux de masculinité à la naissance dans les sociétés occidentales est de 105 hommes pour 100 femmes. On veut vérifier si le taux de masculinité à la naissance au Québec est bien de 105 hommes pour 100 femmes :

$$\begin{aligned} H_0 &: \pi = 105/205 \\ H_1 &: \pi \neq 105/205 \end{aligned}$$

**Exemple 5.3** ★ On veut savoir si le taux de mauvaises facturations dans une entreprise est sous le seuil de 1%. Les hypothèses à confronter sont

$$\begin{aligned} H_0 &: \pi = 0.01 \\ H_1 &: \pi < 0.01 \end{aligned}$$

Il faut noter que dans cet exemple l'hypothèse  $H_0$  correspond à des entreprises qui ne respectent pas le seuil tandis que l'hypothèse alternative respecte le critère.

**Exemple 5.4** ★ Dans une usine de fabrication de pièces de précision on veut vérifier si la probabilité d'une pièce ne rencontrant pas les critères est de 10% ou moins. Les hypothèses sont

$$\begin{aligned} H_0 &: \pi = 0.1 \\ H_1 &: \pi > 0.1 \end{aligned}$$

Dans cet exemple  $H_0$  est l'hypothèse qui correspond aux pièces respectant les critères tandis que  $H_1$  correspond aux pièces qui ne les respectent pas.

Considérons une variable dichotomique pour laquelle une des valeurs est classée "succès". La probabilité de succès,  $\pi$  est le paramètre d'intérêt et les hypothèses à confronter sont

$$H_0 : \pi = \pi_0$$

$$H_1 : \pi \neq \pi_0$$

pour  $0 < \pi_0 < 1$  une valeur fixée.

Le test approximatif de niveau  $\alpha$  est de rejeter  $H_0$  si

$$\frac{|p - \pi_0|}{\sqrt{\pi_0(1 - \pi_0)}} \sqrt{n} \geq z_{\alpha/2}$$

où  $z_{\alpha/2}$  est le point critique de niveau  $\alpha/2$  pour une loi  $N(0, 1)$ .

Conditions d'application Le test est valide si

- a. le caractère mesuré sur chaque unité statistique est dichotomique
- b. la probabilité est identique à chaque élément de l'échantillon
- c. la taille d'échantillon,  $n$ , est assez grande : soit  $n \geq 20$  si  $\pi_0$  est près de 0.5,  $n \geq 30$  si  $\pi_0$  est près de 15% (85%) et  $n \geq 50$  si  $\pi_0$  près de 5% (95%).

**Remarque 5.1** Les test unilatéraux correspondant sont de rejeter  $H_0$  si

$$\begin{aligned} \frac{p - \pi_0}{\sqrt{\pi_0(1 - \pi_0)}} \sqrt{n} \geq z_\alpha & \quad \text{pour} \quad \begin{cases} H_0 : \pi = \pi_0 \\ H_1 : \pi > \pi_0 \end{cases} \\ \frac{p - \pi_0}{\sqrt{\pi_0(1 - \pi_0)}} \sqrt{n} \leq -z_\alpha & \quad \text{pour} \quad \begin{cases} H_0 : \pi = \pi_0 \\ H_1 : \pi < \pi_0 \end{cases} \end{aligned}$$

**Exemple 5.5** ★ Une économiste étudie le risque de faillite personnelle après 2 ans sur les prêts hypothécaires à risque dans les états du sud ouest des États-Unis. Ce taux était de 10% en 1995 et l'objectif est de vérifier s'il est différent pour 2005. Un échantillon de 300 prêts à risque donne 44 faillites personnelles après deux ans. Au niveau  $\alpha = 5\%$ , peut-on dire que la probabilité de faillite est différente de celle observée en 1995 ?

1. On cherche à vérifier si le taux de faillite aujourd'hui est différent de ce qu'il était en 1995 pour les prêts hypothécaires à risque.



2. Les hypothèses statistiques sont

$$H_0 : \pi = 0.1$$

$$H_1 : \pi \neq 0.1$$

où  $\pi$  est la probabilité de faillite personnelle.

3. Le niveau du test est  $\alpha = 5\%$ , soit la probabilité de dire que le taux est plus élevé étant donné qu'il est inchangé.

4. Le test de niveau  $\alpha$  est de rejeter  $H_0$  si

$$\frac{|p - \pi_0|}{\sqrt{\pi_0(1 - \pi_0)}} \sqrt{n} \geq z_{\alpha/2}$$

c'est-à-dire si

$$\frac{|p - 0.1|}{\sqrt{0.1(1 - 0.1)}} \sqrt{300} \geq 1.96$$

Les conditions d'application sont respectées puisqu'on est en présence d'un échantillon aléatoire et que la variable est dichotomique. De plus, la taille est assez grande,  $n = 300$ .

5. On observe

$$p = \frac{44}{300} = 0.14667$$

et

$$\frac{|p - 0.1|}{\sqrt{0.1(1 - 0.1)}} \sqrt{300} = 2.6946$$

6. Cette valeur est nettement plus grande que le point critique de 1,96 donc on rejette  $H_0$  et on peut conclure que le taux de 2005 est significativement différent du taux de 10% répertorié en 2005 et cela avec un niveau de  $\alpha = 5\%$ .

**Remarque 5.2** Le test est approximatif dans le sens que le point critique a été évalué en utilisant le théorème central limite pour une variable dichotomique. Si les tailles d'échantillon ne respectent pas les valeurs données dans les conditions d'application il en résulterait un point critique plus ou moins précis. C'est dire qu'il faudrait faire attention aux conclusions pour toutes les valeurs observées de la statistique qui seraient près du point critique.

**Exemple 5.6** ★ Un chercheur croit que le taux de fraudeurs pour un programme de subvention est en augmentation depuis quelques années. Il y a 3 ans le taux était de 3% dans la population. Pour vérifier son hypothèse, il choisit aléatoirement 400 personnes parmi celles qui ont bénéficié du programme de subvention et une vérification est faite. Il y a exactement 19 personnes qui ont fraudées. Son hypothèse est-elle vérifiée ?

1. Pour déterminer si l'hypothèse est vérifiée, il faut faire un test statistique sur un paramètre de la population. Dans le cas présent la v.a.  $X$  prend la valeur 1 s'il y a fraude et 0 sinon. On a  $X \sim \text{Bin}(1, \pi)$ , où  $\pi$  est le taux de fraudeur.

2. On veut confronter

$$H_0 : \pi = 0.03$$

$$H_1 : \pi > 0.03$$

3. Le niveau du test est défini par  $\alpha = \Pr(\text{rejeter } H_0 \mid H_0)$ , c'est-à-dire la probabilité de dire que le taux est en augmentation étant donné qu'il est fixé à 3%. Si on ne sait pas si l'erreur est grave, on fixe  $\alpha = 5\%$ .

4. Le test d'hypothèse est le suivant : rejeter  $H_0$  si

$$\frac{p - 0.03}{\sqrt{0.03(1 - 0.03)}} \sqrt{n} > z_\alpha$$

5. Or on observe  $p = 19/400 = .0475$ ,  $n = 400$  et  $z_\alpha = z_{0.05} = 1.645$ . La statistique du test donne

$$\frac{19/400 - 0.03}{\sqrt{0.03(1 - 0.03)}} \sqrt{400} = 2.0517$$

6. Puisque la valeur de la statistique du test est dans la zone de rejet, on rejette  $H_0$  et on conclue avec un niveau de 5% que le taux est en augmentation.

**Exemple 5.7** ★ Dans une usine d'électronique, on fabrique des composantes pour construire un circuit. On a calibré les machines pour que les pertes ne dépassent pas les 10%. Pour vérifier si la calibration est adéquate, 300 composantes sont vérifiées et sur le lot il y en a 45 qui sont défectueuses. Peut-on dire avec un niveau de 5% que la calibration est correcte ?

On est en présence d'un échantillon de taille 300 de v.a.  $\text{Bin}(1, \pi)$ , où  $\pi$  est la probabilité qu'une composante soit défectueuse. On veut confronter les hypothèses (2)<sup>4</sup>

$$\begin{cases} H_0 : \pi = 0.1 \\ H_1 : \pi > 0.1 \end{cases}$$

avec un niveau  $\alpha = 5\%$ , soit la probabilité de dire que la machine n'est pas calibrée étant donné qu'elle l'est effectivement (3). La taille de l'échantillon est très grande et ainsi un test approximatif est de rejeter  $H_0$  si (4)

$$\frac{p - 0.1}{\sqrt{0.1(0.9)}} \sqrt{300} > z_{0.05} = 1.645$$

On observe  $p = 0.15$  ce qui donne comme statistique du test (5) :

$$\frac{p - 0.1}{\sqrt{0.1(0.9)}} \sqrt{300} = 2.8868 > 1.645$$

d'où le rejet de  $H_0$ . On peut donc conclure qu'il y a un problème de calibration au niveau 5% (6).

<sup>4</sup> Les numéros entre parenthèses correspondent aux étapes de réalisation du test. Il est parfois plus élégant de décrire le test statistique dans une suite de paragraphes sans identifier nécessairement les numéros d'étapes. Il faut cependant s'assurer que tous les éléments sont là.

## Test pour une moyenne

Le test sur une moyenne permet de comparer une moyenne à une norme déjà établie. La notion de moyenne étant relativement bien intégrée dans notre compréhension du monde, ce test est très fréquent. Il faut distinguer deux contextes ; lorsque le caractère d'intérêt est de distribution normale (test exact) et lorsque le caractère n'est pas distribué normalement mais que la taille de l'échantillon est relativement grande (test approximatif).

### Population normale

Le terme population normale veut dire que la variable mesurée est distribuée selon une loi normale ou du moins selon une distribution qui est très semblable.

Les hypothèses à confronter sont données par

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

Le test de niveau  $\alpha$  est de rejeter  $H_0$  si

$$\frac{|\bar{X} - \mu_0|}{S} \sqrt{n} \geq t_{n-1; \alpha/2}$$

où  $S^2$  est la variance échantillonnale et  $t_{n-1; \alpha/2}$  est le point critique de niveau  $\alpha/2$  pour une loi de Student de paramètre  $n - 1$ .

Conditions d'application Pour que le test soit valide il faut

- un échantillon aléatoire de  $n$  valeurs
- les mesures sont de loi normale

**Remarque 5.3** Les tests unilatéraux correspondants sont de rejeter  $H_0$  si

$$\begin{aligned} \frac{(\bar{X} - \mu_0)}{S} \sqrt{n} \geq t_{n-1; \alpha} & \quad \text{pour} \quad \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} \\ \frac{(\bar{X} - \mu_0)}{S} \sqrt{n} \leq -t_{n-1; \alpha} & \quad \text{pour} \quad \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \end{aligned}$$

**Exemple 5.8** ★ Les plantations d'eucalyptus est une base de l'économie régionale de plusieurs pays d'Afrique. Un chercheur en agronomie veut vérifier si un nouveau clone d'eucalyptus (le no 51n78) est assez productif pour que son exploitation soit rentable. L'exploitation est rentable si le rendement moyen est de plus de  $2\text{m}^3$  de bois en moyenne par arbre après 10 ans.

Pour vérifier si le clone proposé est rentable, 57 arbres ont été plantés et les résultats après 10 ans sont les suivants :  $\bar{x} = 2.2\text{m}^3$  et  $s = 1\text{m}^3$ . Des données biologiques montrent que le

rendement d'un arbre en  $m^3$  est une mesure dont la distribution est très près d'une loi normale.

On veut vérifier l'hypothèse selon laquelle le clone d'eucalyptus no 51n78 est rentable ce qui se traduit par les hypothèses statistiques (2)

$$H_0 : \mu = 2$$

$$H_1 : \mu > 2$$

où  $\mu$  est l'espérance de la v.a. qui donne la production de bois en  $m^3$  du clone d'eucalyptus no 51n78. Le niveau du test statistique est défini par  $\alpha = \Pr(\text{rejeter } H_0 \mid H_0)$ , soit de dire que le clone est rentable étant donné qu'il ne l'est pas. Cette erreur est grave pour l'exploitant et ainsi il prendra un niveau de 1% (3).

Étant donné que la distribution de la population est normale alors le test est de rejeter  $H_0$  si (4)

$$\frac{\bar{X} - \mu_0}{S} \sqrt{n} = \frac{\bar{X} - 2}{S} \sqrt{57} > t_{n-1;\alpha} = 2.6665$$

On observe (5)

$$\frac{\bar{x} - 2}{s} \sqrt{57} = \frac{2.2 - 2}{1} \sqrt{57} = 1.51$$

et on accepte  $H_0$  au niveau 1% (6). On doit donc conclure que le clone 51n78 n'est pas rentable.

### Population quelconque

Le paramètre à tester est la moyenne de la population mais il n'y a pas d'indication sur la forme de la distribution de la variable aléatoire en jeu. La taille de l'échantillon est cependant assez grande (au moins  $n \geq 30$ ) pour justifier l'utilisation du TCL pour calculer les points critiques.

Les hypothèses à confronter sont données par

$$H_0 : \mu = \mu_0$$

$$H_1 : \mu \neq \mu_0$$

et le test de niveau  $\alpha$  est de rejeter  $H_0$  si

$$\frac{|\bar{X} - \mu_0|}{S} \sqrt{n} \geq z_{\alpha/2}$$

où  $S$  est la variance échantillonnale et  $z_{\alpha/2}$  est le point critique d'une loi  $N(0, 1)$ .

Conditions d'application Pour que le test soit valide il faut

- un échantillon aléatoire de  $n$  valeurs
- une taille d'échantillon assez grande  $n \geq 30$

**Remarque 5.4** Les tests unilatéraux correspondants sont de rejeter  $H_0$  si

$$\begin{aligned} \frac{(\bar{X} - \mu_0)}{S} \sqrt{n} \geq z_\alpha & \quad \text{pour} \quad \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu > \mu_0 \end{cases} \\ \frac{(\bar{X} - \mu_0)}{S} \sqrt{n} \leq -z_\alpha & \quad \text{pour} \quad \begin{cases} H_0 : \mu = \mu_0 \\ H_1 : \mu < \mu_0 \end{cases} \end{aligned}$$

**Exemple 5.9** ★ Une entreprise utilise un test de personnalité pour détecter les éléments les plus prometteurs parmi les candidats. Or ce test de personnalité n'a pas été modifié depuis 25 ans et le directeur des ressources humaines a quelques doutes sur sa validité dans le monde actuel. Pour dissiper ses craintes il demande à un psychologue industriel d'investiguer la question.

Les données disponibles lors de la mise en place du test donne un score moyen de 70 dans la population générale et c'est ce score moyen qui permet de détecter les éléments prometteurs. Le psychologue tire un échantillon de 40 personnes et observe une moyenne échantillonnale de 82 avec un écart type de 20. A-t-on des raisons de croire que le test n'est plus fiable ?

1. L'objectif est de comparer la population actuelle à la population d'il y a 25 ans sur la base du score au test de personnalité.

2. Les hypothèses statistiques sont

$$H_0 : \mu = 70$$

$$H_1 : \mu \neq 70$$

3. Le niveau est la probabilité de dire que la population actuelle a une moyenne différente de celle d'il y a 25 ans étant donné qu'elle est semblable. C'est la probabilité de dire que le test n'est plus fiable étant donné qu'il est encore fiable. Les coûts reliés à cette erreur ne sont pas connus donc le niveau  $\alpha = 5\%$  est employé.

4. Puisque la taille de l'échantillon est assez grande et qu'il s'agit d'un échantillon, le test est de rejeter  $H_0$  si

$$\frac{|\bar{X} - 70|}{S} \sqrt{40} \geq z_{\alpha/2} = 1.96$$

5. Les observations donnent

$$\frac{|\bar{X} - 70|}{S} \sqrt{40} = 3.7947$$

6. On rejette  $H_0$  au niveau 5% et on peut dire que la population actuelle est significativement différente de la population d'il y a 25 ans du point de vue de ce test.

## Test pour une variance

Le paramètre d'intérêt est la variance et l'objectif est de comparer ce paramètre à une norme connue.

Les hypothèses à confronter sont

$$H_0 : \sigma^2 = \sigma_0^2$$

$$H_1 : \sigma^2 \neq \sigma_0^2$$

et le test est de rejeter  $H_0$  si

$$\frac{(n-1)S^2}{\sigma_0^2} \geq \chi_{n-1;\alpha/2}^2 \text{ ou si } \frac{(n-1)S^2}{\sigma_0^2} \leq \chi_{n-1;1-\alpha/2}^2$$

où  $\chi_{n-1;\alpha/2}^2$  est le point critique de niveau  $\alpha/2$  d'une loi Khi-deux de paramètre  $n-1$  et  $\chi_{n-1;1-\alpha/2}^2$  est le point critique de niveau  $1-\alpha/2$  d'une loi Khi-deux de paramètre  $n-1$ .

Conditions d'application Pour que le test soit valide il faut

- un échantillon de  $n$  valeurs
- que la mesure soit issue d'une population normale.

**Remarque 5.5** Les tests unilatéraux correspondants sont de rejeter  $H_0$  si

$$\begin{aligned} \frac{(n-1)S^2}{\sigma_0^2} \geq \chi_{n-1;\alpha}^2 & \text{ pour } \begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 > \sigma_0^2 \end{cases} \\ \frac{(n-1)S^2}{\sigma_0^2} \leq \chi_{n-1;1-\alpha}^2 & \text{ pour } \begin{cases} H_0 : \sigma^2 = \sigma_0^2 \\ H_1 : \sigma^2 < \sigma_0^2 \end{cases} \end{aligned}$$

**Exemple 5.10** ★ Un embouteilleur de bières annonce 12 onces par bouteille. Une vérification des quantités de liquide produit et du nombre de bouteilles indique que la moyenne est effectivement de 12 onces. Sachant les amateurs de ce nectar particulièrement pointilleux sur la quantité de liquide dans chaque bouteille il doit aussi s'assurer qu'il n'y a pas trop de variation entre les bouteilles.

Pour cela, il faut que la variance soit d'au plus 0.1. Considérant le procédé d'embouteillage, on peut supposer que la quantité de liquide dans les bouteilles suit une loi normale.

Un échantillon de 24 bouteilles sont tirées au hasard de la production et on observe une variance échantillonnale de  $s^2 = 0.14$ . Peut-on conclure que la variance est supérieure à la norme à un niveau de 5% ?

- On cherche à vérifier si les bouteilles n'ont pas trop de variations sur le niveau de liquide.



2. Les hypothèses statistiques sont

$$H_0 : \sigma^2 = 0.1$$

$$H_1 : \sigma^2 > 0.1$$

3. Le niveau est la probabilité de rejeter  $H_0$  étant donné  $H_0$ , c'est-à-dire la probabilité de dire que la variation est trop grande étant donné qu'elle est "raisonnable". On utilise  $\alpha = 5\%$ .

4. Le test est de rejeter  $H_0$  si

$$\frac{(n-1) S^2}{\sigma_0^2} \geq \chi_{n-1;\alpha}^2 = 35.172$$

5. On observe  $S^2 = 0.14$  et

$$\frac{(n-1) S^2}{\sigma_0^2} = \frac{23(0.14)}{0.1} = 32.2$$

6. Puisque la valeur de la statistique est plus petite que le point critique, il faut accepter  $H_0$  au niveau 5% et on peut dire que la différence n'est pas significative donc que la variation n'est pas trop grande.

## Test pour échantillons appariés

Il arrive qu'il y a deux mesures similaires sur chaque unité statistique pour comparer une évolution dans le temps ou sur des éléments ayant un rapport. On dit que ce sont des données appariées<sup>5</sup>. On est alors dans le contexte d'un seul échantillon mais avec deux mesures, souvent les mesures sont du type "avant traitement" et "après traitement".

Considérons l'exemple d'une entreprise qui veut vérifier si un programme de formation des vendeurs a eu un effet bénéfique. Les ventes mensuelles corrigées pour le mois de l'année ont été enregistrées pour 23 vendeurs. Par la suite ces vendeurs ont bénéficiés d'une formation. Un mois après la formation, les ventes, toujours corrigées pour le mois de l'année, sont enregistrées pour ces mêmes 23 vendeurs. On veut vérifier si la formation a eu un effet bénéfique sur les ventes. Dans ce problème, il n'y a qu'un échantillon, soit les 23 vendeurs mais il y a 2 mesures par individu, les ventes avant la formation et les ventes après la formation. La variable d'intérêt est la différence des ventes entre "après la formation" et "avant la formation". Les hypothèses à confronter sont

$$H_0 : \mu_d = 0$$

$$H_1 : \mu_d > 0$$

<sup>5</sup> Plusieurs documents font référence à des données "pairees". Cette formulation est un anglicisme provenant de l'expression anglaise "paired sample".

où  $\mu_d$  représente la moyenne de la différence entre "après" moins "avant". C'est donc un problème de comparaison de la moyenne à une norme qui égale dans ce cas 0.

**Exemple 5.11 ★★** Une chercheuse en biologie veut déterminer l'effet de la congélation à  $-80^\circ\text{C}$  sur le nombre de bactéries dans une culture. L'hypothèse étant que la congélation à une telle température devrait détruire une certaine proportion des bactéries. Pour faire l'expérimentation, la chercheuse observe 30 cultures pour déterminer le nombre de bactéries présentes dans la culture puis chaque culture est réfrigérée à  $-80^\circ\text{C}$  pendant 1 semaine pour ensuite être décongelée. Le nombre de bactéries est alors comptabilisé dans chaque culture décongelée. Si on observe les données standardisées ( $\times 10^2$ ) suivantes

Culture	1	2	3	4	5	6	7	8	9	10
Avant	57	45	70	62	47	55	50	61	45	71
Après	55	50	60	59	50	49	43	55	49	66

Culture	11	12	13	14	15	16	17	18	19	20
Avant	45	66	69	54	47	53	58	67	59	47
Après	40	59	70	49	53	45	50	63	55	50

Culture	21	22	23	24	25	26	27	28	29	30
Avant	57	64	73	48	55	59	62	61	51	54
Après	53	56	63	45	59	51	56	62	48	51

peut-on déterminer si la congélation a détruit des bactéries ?

**Solution :** On est en présence de deux échantillons appariés. On s'intéresse à la différence du nombre de bactéries dans la culture entre avant et après. Ainsi les valeurs observées pour cette différence sont

$$2, -5, 10, -3, -3, \dots, 3, 3$$

Ce qui donne une moyenne de 3.4 et un écart type de 4.81.

1. On cherche à savoir si la congélation a un effet sur la destruction des cellules
2. Les hypothèses statistiques sont

$$H_0 : \mu_d = 0$$

$$H_1 : \mu_d > 0$$

3. Le niveau du test est  $\alpha = \Pr(\text{rejeter } H_0 \mid H_0)$ , soit la probabilité de dire que la congélation a une influence sur la destruction des bactéries étant donné que cela n'a pas d'influence. On ne connaît pas les conséquences et ainsi le niveau est fixé à 5%.
4. On a un échantillon de taille relativement grande ( $n = 30$ ) et ainsi le test d'hypothèses approximatif de niveau  $\alpha$  est de rejeter  $H_0$  si

$$\frac{\bar{X} - 0}{S} \sqrt{n} \geq z_{\alpha/2}$$

5. Or on observe

$$\frac{\bar{x}}{s} \sqrt{n} = \frac{3.4}{4.81} \sqrt{30} = 3.8716$$

et  $z_{0.05} = 1.645$ .

6. On doit donc rejeter  $H_0$  au niveau 5% et conclure que la congélation a un effet destructeur sur les bactéries.

**Exemple 5.12** ★ On pense que les automobilistes sont plus soucieux de l'usure des pneus d'hiver que des pneus d'été. Sur un échantillon de 15 automobilistes on remarque les usures (en %) pour les pneus d'été :

45, 48, 38, 46, 37, 39, 29, 46, 47, 43, 31, 43, 37, 27, 49

L'usure des pneus d'hiver a été mesurée chez les mêmes automobilistes :

41, 37, 44, 28, 28, 32, 40, 41, 31, 30, 30, 49, 35, 35, 37

Peut-on dire que les automobilistes ont des pneus d'hiver moins usés au niveau 10% en considérant que la mesure de l'usure est une v.a. approximativement normale ?

Solution : On est en présence de deux échantillons appariés. Soit  $X_d$  la v.a. qui donne la différence entre les l'usure des pneus d'hiver et des pneus d'été, on veut confronter les hypothèses

$$H_0 : \mu_d = 0$$

$$H_1 : \mu_d < 0$$

avec  $\alpha = 5\%$  Le test est de rejeter  $H_0$  si

$$\frac{\bar{X}_d}{S_d} \sqrt{n} \leq -t_{0.1;14} = -1.345$$

On observe :  $X_d = -4 \quad -11 \quad 6 \quad -18 \quad -9 \quad -7 \quad 11 \quad -5 \quad -16 \quad -13 \quad -1 \quad 6 \quad -2 \quad 8 \quad -12$   $s_D = 9.0543$  et  $\bar{x}_d = -4.4667$ . La statistique devient

$$\frac{\bar{x}_d}{s_d} \sqrt{n} = \frac{-4.4667}{9.0543} \sqrt{15} = -1.9106 \leq -1.345$$

On rejette  $H_0$  au niveau 10%

**Exemple 5.13** ★ On pense que la fatigue de la journée mène à plus d'erreurs lorsqu'on utilise un traitement de texte pour reproduire un texte. Pour vérifier cette hypothèse on choisit un échantillon de 12 personnes à qui on demande de recopier un texte au début de la matinée. Voici le nombre d'erreurs de frappe :

7, 15, 6, 7, 13, 15, 9, 15, 5, 8, 8, 11

À la fin de la journée les mêmes personnes doivent recopier un texte. Voici le nombre d'erreurs de frappe pour les 12 personnes dans le même ordre :

9, 7, 12, 10, 12, 9, 4, 6, 8, 7, 11, 7

En supposant que le nombre d'erreurs de frappe est une variable de loi normale, vérifier l'affirmation au niveau 5%.

Solution : Les hypothèses à confronter sont  $H_0 : \mu_d = 0$  VS  $H_1 : \mu_d < 0$  où  $X_d = X_{Matin} - X_{Soir}$ .

Le test est de rejeter  $H_0$  si

$$\frac{\bar{X}_d}{S_d} \sqrt{n} < -t_{0.05;11} = -1.7959.$$

On a  $\bar{x}_d = 1.4167$ ,  $s_d = 4.926$  et

$$\begin{aligned} \frac{\bar{x}_d}{s_d} \sqrt{n} &= 1.4167/4.926 \sqrt{12} \\ &= 0.99626 \not< -1.7959 \end{aligned}$$

donc on accepte  $H_0$

## Tableau des tests pour un échantillon

Voici un tableau récapitulatif des tests disponibles pour un échantillon :

Hypothèses	Population	Statistique du test :	Région de rejet
$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$N(\mu, \sigma^2)$ $\sigma^2$ inconnu	$T = \frac{ \bar{X} - \mu_0  \sqrt{n}}{S}$	$T \geq t_{n-1; \alpha/2}$
$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$N(\mu, \sigma^2)$ $\sigma^2$ inconnu	$T = \frac{(\bar{X} - \mu_0) \sqrt{n}}{S}$	$T \leq -t_{n-1; \alpha}$
$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$N(\mu, \sigma^2)$ $\sigma^2$ inconnu	$T = \frac{(\bar{X} - \mu_0) \sqrt{n}}{S}$	$T \geq t_{n-1; \alpha}$
$H_0 : \mu = \mu_0$ $H_1 : \mu \neq \mu_0$	$n \geq 30$ $\sigma^2$ inconnu	$T = \frac{ \bar{X} - \mu_0  \sqrt{n}}{S}$	$T \geq z_{\alpha/2}$
$H_0 : \mu = \mu_0$ $H_1 : \mu < \mu_0$	$n \geq 30$ $\sigma^2$ inconnu	$T = \frac{(\bar{X} - \mu_0) \sqrt{n}}{S}$	$T \leq -z_{\alpha}$
$H_0 : \mu = \mu_0$ $H_1 : \mu > \mu_0$	$n \geq 30$ $\sigma^2$ inconnu	$T = \frac{(\bar{X} - \mu_0) \sqrt{n}}{S}$	$T \geq z_{\alpha}$
$H_0 : \pi = \pi_0$ $H_1 : \pi \neq \pi_0$	$n \geq 20$	$T = \frac{ p - \pi_0 }{\sqrt{\pi_0(1-\pi_0)}} \sqrt{n}$	$T \geq z_{\alpha/2}$
$H_0 : \pi = \pi_0$ $H_1 : \pi < \pi_0$	$n \geq 20$	$T = \frac{(p - \pi_0)}{\sqrt{\pi_0(1-\pi_0)}} \sqrt{n}$	$T \leq -z_{\alpha}$
$H_0 : \pi = \pi_0$ $H_1 : \pi > \pi_0$	$n \geq 20$	$T = \frac{(p - \pi_0)}{\sqrt{\pi_0(1-\pi_0)}} \sqrt{n}$	$T \geq z_{\alpha}$
$H_0 : \sigma^2 = \sigma_0^2$ $H_1 : \sigma^2 \neq \sigma_0^2$	$N(\mu, \sigma^2)$	$T = \frac{(n-1)S^2}{\sigma_0^2}$	$T \geq \chi_{n-1; \alpha/2}^2$ $T \leq \chi_{n-1; 1-\alpha/2}^2$
$H_0 : \sigma^2 = \sigma_0^2$ $H_1 : \sigma^2 < \sigma_0^2$	$N(\mu, \sigma^2)$	$T = \frac{(n-1)S^2}{\sigma_0^2}$	$T \leq \chi_{n-1; 1-\alpha}^2$
$H_0 : \sigma^2 = \sigma_0^2$ $H_1 : \sigma^2 > \sigma_0^2$	$N(\mu, \sigma^2)$	$T = \frac{(n-1)S^2}{\sigma_0^2}$	$T \geq \chi_{n-1; \alpha}^2$

### 10.6 Résumé

Dans ce module on aborde la notion de tests d'hypothèses avec quelques exemples de tests se faisant facilement avec la calculatrice. Les principes de base sont

- Deux hypothèses complémentaires  $H_0$  et  $H_1$
- Niveau du test,  $\alpha$  fixé par l'utilisateur à 1%, 5% ou 10% a priori.

- La puissance du test,  $1 - \beta$  dépend de  $\alpha$  (même sens) et de la taille de l'échantillon
- On suppose a priori  $H_0$  vraie et si les observations contredisent cette hypothèse on opte pour  $H_1$
- Un test est toujours de la forme : "Rejeter  $H_0$  si..."

Pour faire un test il faut déterminer les hypothèses statistiques, les conditions d'application puis choisir le test dans la table des formules pour les tests à un échantillon.