7. Principe d'un test statistique

Objectifs :

- Aborder les tests statistiques en examinant leur principe
- Comprendre la logique commune des tests statistiques
- Connaître et savoir appliquer les différentes étapes d'un test statistique

---

A) Principe d'un test statistique

Nous allons voir comment, en considérant les notions vues dans le cadre des probabilités ainsi que le principe de l'échantillonnage, on peut prendre des décisions objectives par rapport à la "véracité" de certaines prédictions. Le travail du biologiste, et plus généralement celui du scientifique, consiste à apporter des explications cohérentes aux observations qu'il peut faire dans la nature ou en laboratoire. Face à un(e) phénomène/observation biologique donné(e), la démarche consiste à émettre une explication potentielle (ce n'est alors qu'une hypothèse). Cette hypothèse est ensuite mise à l'épreuve en confrontant des **prédictions** qui sont émises à partir de cette hypothèse avec de **nouvelles observations**. L'outil qui permet ensuite de dire objectivement si les nouvelles observations sont conformes aux prédictions est un **test statistique**.

On utilisera 2 exemples : le premier non biologique est très intuitif, et le second moins intuitif est biologique.

Exemple 1: On propose à un étudiant de jouer son argent de poche du mois avec un jeu de Pile/Face (pièce de monnaie). L'enseignant propose d'utiliser une pièce à lui. L'étudiant a un doute sur l'honnêteté de l'enseignant, plus particulièrement sur la qualité de sa pièce. Est-elle truquée ou non?
Question : comment feraient les étudiants pour le vérifier avant de se décider à jouer leur argent de poche?

Exemple 2: Vous êtes chargé par un parc naturel de réintroduire la truite dans un cours d'eau duquel elle a disparu il y a plusieurs années, en raison d'une détérioration de son habitat. La rivière a depuis été remise en état. Plusieurs piscicultures vous proposent de vous vendre des truitelles pour effectuer cette réintroduction. Vous consultez un spécialiste de cette espèce et il vous apprend que la taille des individus relâchés est un facteur déterminant dans le succès de la réintroduction. Plus les truitelles relâchées sont grosses, plus elles donneront de descendants au moment de la reproduction. Un pisciculteur vous affirme qu'il peut vous fournir des truitelles de 12 mois ayant un poids moyen de 250gr. Comme il est celui qui propose les poissons les plus gros, vous êtes évidemment très intéressé par sa proposition. Très intéressé, mais aussi méfiant…

Vous pouvez à ce stade vous trouver dans 1 des 4 situations suivantes :

- ce pisciculteur est honnête et vous achetez vos truitelles chez lui.
- ce pisciculteur est malhonnête, ses truitelles sont en réalité plus petites, et ses affirmations avaient pour seul but d'emporter le marché! Si vous faites affaire avec lui, vous vous ferez avoir…
- ce pisciculteur est très prudent, il tend à sous-estimer la taille de ses truitelles pour ne pas être accusé d'être malhonnête. Elles sont en réalité plus grosses. Tant mieux pour vous!
- Vous avez en fait rencontré le fils du pisciculteur qui débute dans le métier, et il vous a avancé le chiffre de 250 gr sans trop savoir de quoi il en est en réalité. Ses truitelles ne font pas 250 gr mais vous ne savez pas si elles sont plus grosses ou plus petites.

Votre problème est maintenant de décider quelle position adopter, en étant le plus objectif possible. Comment faire?

Principe de la démarche

Si vous souhaitez vérifier par vous-mêmes que les truitelles de 12 mois de ce pisciculteur ont bien un poids moyen de 250gr, le plus simple est d'aller directement chez ce pisciculteur et d'échantillonner quelques truitelles de 12 mois (par exemple une vingtaine). Vous les pesez toutes et vous calculez ensuite le poids moyen de ces individus. Plusieurs possibilités :

- le poids moyen dans votre échantillon est exactement de 250 gr : vous décidez de faire confiance à ce pisciculteur, il vous a dit vrai.
- le poids moyen dans votre échantillon est légèrement plus faible (par exemple 247,5gr) ou légèrement plus élevé (par exemple 252,4gr) que le poids annoncé (250 gr). Vous considérez que la différence entre ce que vous observez (le poids moyen dans votre échantillon) et ce que vous attendiez (le poids moyen annoncé par le pisciculteur) est petite et qu'elle est simplement due aux fluctuations d'échantillonnage. Le hasard de l'échantillonnage a fait que votre échantillon comporte des individus dont le poids moyen est légèrement différent du poids moyen de l'ensemble des truitelles de 12 mois de cette pisciculture. Notez que toutes les truitelles de 12 mois n'ont pas exactement le même poids, certaines font à peine 230 gr alors que d'autres dépassent les 280 gr. Il est alors tout à fait possible d'avoir tiré un échantillon identique au vôtre dans une telle population statistique. Vous décidez donc de faire confiance au pisciculteur.
- le poids moyen dans votre échantillon (par exemple 225 gr) est très différent de la valeur annoncée. Même s'il existe des truitelles de 12 mois pesant moins de 230 gr, elles sont relativement peu nombreuses. Il est alors peu probable qu'en échantillonnant dans cette pisciculture on obtienne un échantillon dont le poids moyen soit aussi petit que 225 gr. Vous avez donc des doutes sur l'affirmation du pisciculteur et vous décidez de ne pas lui faire confiance (vous pensez qu'il vous ment).

Ce type de raisonnement correspond exactement à la logique des tests statistiques. Ce que nous venons de faire est un test statistique (non formalisé).

B) Les étapes d'un test statistique

En reprenant l'exemple développé précédemment, nous allons examiner les différentes étapes d'un test statistique en les formalisant plus précisément.

1. Poser clairement le problème (la question)

Devez-vous faire confiance au pisciculteur? Est-il trop prudent? Est-il un escroc? Est-il incompétent?

2. Les hypothèses (H0 et H1)

Un test statistique consiste toujours à opposer 2 hypothèses. A l'issue du test, l'une doit être invalidée au profit de l'autre.
La première est **l'hypothèse nulle**, notée **H0**. L'hypothèse nulle correspond toujours à l'absence de différence entre ce qu'on observe et ce qu'on attend par rapport à l'affirmation ou au modèle proposé(e).
L'hypothèse alternative (parfois appelée hypothèse explicative) est la seconde hypothèse et elle est notée **H1**. Elle correspond à la différence (pas forcément quantifiée) que l'on attend entre ce qu'on observe et ce qui est prédit par l'affirmation ou le modèle.

Exemple
A. Imaginons dans un premier temps que l'on soupçonne ce pisciculteur d'essayer de vous arnaquer.
Si il dit vrai, vous savez à peu près à quelle valeur de poids moyen vous attendre dans votre échantillon. Si il dit faux, vous ne savez pas à quelle valeur vous attendre ; vous vous attendez seulement à une valeur inférieure à celle qu'il annonce. On écrira les hypothèses de la manière suivante.
H0 : le poids moyen des truitelles de 12 mois de ce pisciculteur n'est pas inférieur à 250gr.
H1 : le poids moyen des truitelles de 12 mois de ce pisciculteur est inférieur à 250gr.
Ou encore
H0 : $\mu \geq 250$ gr
H1 : $\mu < 250$ gr

B. Si vous le soupçonner d'être très prudent
H0 : $\mu \leq 250$ gr
H1 : $\mu > 250$ gr

C. Si vous le soupçonner d'être incompétent
H0 : $\mu = 250$ gr
H1 : $\mu \neq 250$ gr

Les 2 hypothèses doivent être mutuellement exclusives et doivent collectivement inclure toutes les possibilités. De cette manière, il y en a forcément une qui est vraie et une qui est fausse.

Les 2 premières situations correspondent à un test dit **unilatéral** et la seconde à un test **bilatéral**. Nous y reviendrons plus tard.

3. Choisir un seuil de signification

Pour décider de faire confiance ou non au pisciculteur, vous examinez la différence entre ce que vous observez (poids moyen de votre échantillon) et ce que vous attendez si l'affirmation est vraie (en fait, si H0 est vraie). Si ce que vous observez est proche de ce que vous attendez, vous considérez que H0 est vraie (le pisciculteur dit vrai). Si ce que vous observez est loin de ce que vous attendez, vous rejeter H0 (le pisciculteur ment, volontairement ou non). Tout le problème est de déterminer ce qui est proche ou éloigné de ce vous attendez si H0 est vraie. En fait, on ne s'intéresse pas directement à la valeur de la différence entre ce qui est observé et attendu, mais à la probabilité d'observer une telle différence si H0 est vraie. Est-ce probable d'observer un tel écart entre le poids moyen de l'échantillon et 250gr si cette valeur annoncée par le pisciculteur est la bonne (H0 vraie)? Cet écart n'est-il pas "douteux" par rapport à ce qui est annoncé?
La limite entre ce que l'on considère comme probable ou improbable si H0 est vraie correspond à ce qu'on appelle le **seuil de signification** et on le note **α**. C'est une limite choisie **arbitrairement**. Le seuil le plus utilisé est α=5%. Autrement dit, si on a moins de 5% de chances d'observer un poids moyen dans l'échantillon aussi différent de la valeur attendue, on considèrera que H0 doit être rejetée.

4. Choisir un test statistique

Cela consiste à choisir une variable aléatoire dite **auxiliaire**, dont on détermine ensuite la fonction de répartition. La valeur de cette variable sera ensuite calculée à partir des données.

Exemple
Pour décider de rejeter ou d'accepter H0, on cherche à obtenir la probabilité d'observer une valeur de poids moyen au moins aussi différente (de ce qu'on attend, ici 250gr) que celle obtenue à partir des données. En d'autre terme, on cherche la **distribution d'échantillonnage** de la moyenne d'un échantillon (ou sa fonction de répartition).

Distribution d'échantillonnage de la moyenne d'un échantillon

A partir de la distribution de la taille des truitelles dans la pisciculture (= population), on peut déterminer la probabilité de tirer au hasard une truitelle plus grande ou plus petite qu'une certaine valeur, en utilisant la fonction de répartition. Le problème est ici de déterminer la probabilité d'obtenir une moyenne plus grande ou plus petite qu'une certaine valeur.
Quand on prélève un échantillon, on ne sait pas à l'avance quelle valeur de moyenne on va obtenir. La moyenne d'un échantillon est donc aussi une variable aléatoire dont on va chercher à détermine la fonction de densité.
Quand on prélève un échantillon dans une population qui a une distribution normale $N(\mu, \sigma)$ pour la variable étudiée, la moyenne d'un échantillon suit aussi une distribution normale $N(\mu, \sigma/\sqrt{n})$.
Quand on prélève un échantillon dans une population qui a une distribution non normale, la moyenne d'un échantillon s'approche d'une distribution normale si *n* est suffisamment grand.

Donc $Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}}$ suit une distribution normale centrée réduite.

Pour tester si μ=250, nous utiliserons donc la variable auxiliaire Z telle que définie dans l'encadré précédent. Sa fonction de densité est la distribution normale centrée réduite.

Rq : d'une manière générale, en statistiques, on essaie souvent de rapprocher la distribution de la variable qui nous intéresse d'une distribution théorique connue.

5. Calculer la valeur de la variable auxiliaire à partir des données

Imaginons qu'à partir des 25 truitelles échantillonnées, vous obteniez un poids moyen de 248 grammes. Imaginons que le poids suive une distribution normale. D'après H0, la moyenne est μ=250gr. Imaginons que l'écart-type de la taille soit de 5 grammes (connu ou estimé à partir de l'échantillon).

$$Z_c = \frac{248 - 250}{5 / \sqrt{20}} = -1.79$$

6. Déterminer la probabilité d'observer une valeur aussi extrême que celle calculée à partir des données

$$X \sim N(250,5) \quad \text{d'où} \quad \overline{X} \sim N(250, 5/\sqrt{20}) \quad \text{donc} \quad Z = \frac{\overline{X} - 250}{5 / \sqrt{20}} \sim N(0,1)$$

a) Dans la situation où on soupçonne que le pisciculteur essaie de nous arnaquer, on cherche la probabilité d'avoir un poids moyen au moins aussi petit que 248gr (valeur observée) si H0 est vraie (le pisciculteur est honnête).

P(poids moyen < 248) = P(Z < -1,79) = 1 – F(1,79) = 0,0367
Si on prélève un échantillon de 20 truitelles dans une population dont le poids des individus est distribué normalement, avec pour moyenne 250 gr et écart-type 5 gr, il y a une probabilité de 0,0367 d'obtenir un poids moyen inférieur ou égal à 248 gr.

b) On peut aussi déterminer la valeur seuil (ou valeur critique) du test. C'est la valeur pour laquelle il y a une probabilité de 5% d'obtenir une valeur égale ou plus extrême si H0 est vraie.

Si α=5%, la valeur critique pour Z est -1,645 (que l'on compare à la valeur de Z calculée sur les données).

7. Décision statistique et conclusion biologique

Il y a donc 2 manières de décider de rejeter ou d'accepter H0 :
a) On compare la probabilité associée à la valeur calculée au seuil de signification (seuil α) choisi a priori

Si cette probabilité est inférieure à α, on rejette H0 (cette probabilité indique qu'on a très peu de chances d'observer un résultat tel que celui obtenu à partir de l'échantillon si H0 est vraie ; on remet donc en cause H0)
Si cette probabilité est supérieure à α, on accepte H0.

b) On compare la valeur observée de la variable auxiliaire à la valeur seuil de cette variable. Selon que la valeur observée est supérieure ou inférieure à la valeur seuil, on rejette ou on accepte H0.

<u>Le test est légèrement différent si les hypothèses testées sont différentes.</u>

B. H0 : $\mu \leq 250$ gr
H1 : $\mu > 250$ gr

Si on cherche à tester l'hypothèse que le pisciculteur est trop prudent, on cherche la probabilité d'avoir un poids moyen au moins aussi grand que 248gr. (si H1 est vraie, on attend une valeur de moyenne très grande)
H0 : $\mu \leq 250$ gr
H1 : $\mu > 250$ gr

C. H0 : $\mu = 250$ gr
H1 : $\mu \neq 250$ gr

Si on cherche à tester l'hypothèse que le pisciculteur est incompétent, on cherche la probabilité d'avoir une moyenne au moins aussi différente que 248gr, mais en valeur absolue.

Les 2 premiers cas (A et B) nécessitent un test unilatéral (zone de rejet de H0 d'un seul côté de la distribution). La dernière situation (C) correspond à un test bilatéral.

<u>C) Notion d'erreur</u>

Les hypothèses H0 et H1 sont mutuellement exclusives et décrivent collectivement toutes les situations possibles. Donc l'une est forcément vraie alors que l'autre est forcément fausse. Laquelle est vraie? Nous ne le savons pas et nous ne le saurons probablement jamais... En revanche, il nous fait prendre une décision et considérer l'un comme vraie et l'autre comme fausse.

On peut donc se retrouver dans une des 4 situations suivantes :

- H0 est vraie et on l'accepte : on prend la bonne décision
- H0 est fausse et la rejette : on prend la bonne décision
- H0 est vraie et on la rejette : on commet ce qu'on appelle une **erreur de type I** (ou **erreur de première espèce**)
- H0 est fausse et on l'accepte : on commet une **erreur de type II** (ou **erreur de seconde espèce**)

Revenons à notre exemple. On soupçonne que le pisciculteur est malhonnête. On cherche donc la probabilité d'observer un poids moyen de l'échantillon au moins aussi petit que 248gr si H0 est vraie, c'est-à-dire si notre échantillon provient bien d'une population dans laquelle le poids moyen est de 250gr (comme l'affirme la pisciculteur). On a vu précédemment que cette probabilité est de 0,0367. Comme cette probabilité est très faible (la limite entre ce que nous considérons comme probable et improbable si H0 est vraie a été fixée arbitrairement à 5% (p=0.05)), on décide de rejeter H0. On considère donc que le pisciculteur ne dit pas la vérité (H0 fausse) et qu'il essaie de nous arnaquer.
Cependant, obtenir un poids moyen pour l'échantillon au moins aussi petit que 248gr n'est pas une chose <u>**impossible**</u> si on échantillonne dans une population dont le poids moyen est de 250gr. C'est juste <u>**improbable**</u> (ce qui est très différent!). Donc il se pourrait tout à fait que notre échantillon provienne d'une population ayant un poids moyen de 250gr, c'est-à-dire que H0 soit vraie. Nous avons simplement très peu "confiance" dans l'affirmation "le poids moyen de la population est de 250gr" (H0). Nous rejetons cette hypothèse et nous sommes donc contraints d'accepter H1 (le poids moyen de la population est inférieur à 250gr). En prenant la décision de rejeter H0, on prend donc le risque de commettre une erreur de première espèce (rejeter H0 alors qu'elle est vraie).
Quelle est la probabilité de commettre une telle erreur?
Réponse : p = 0.0367 (Cf cours)
La probabilité de commettre une erreur de type I est donc égale au seuil de signification choisi arbitrairement (α=5% dans la plupart des cas).

L'erreur de type II correspond au fait d'accepter H0 alors qu'elle est fausse. La probabilité de commettre ce type d'erreur est notée β, mais contrairement à α, elle est difficilement quantifiable.

On pourrait a priori penser qu'il suffit de choisir une valeur de α la plus petite possible pour limiter le risque de commettre une erreur de type I. Le problème est qu'une diminution du risque d'erreur de première espèce (α) s'accompagne presque toujours d'une augmentation du risque d'erreur de seconde espèce (β).

RQ : le résultat d'un test (la décision) n'est aucunement une preuve de quoi que ce soit. La décision prise l'est toujours avec un certain risque d'erreur (parfois quantifié).