3- Test d'hypothèse

Dans tous les domaines, de l'expérimentation scientifique à la vie quotidienne, on est amené à prendre des décisions sur une activité risquée au vu de résultats d'expériences ou d'observation de phénomènes dans un contexte incertain. Par exemple :

- Santé : trancher sur la nocivité ou non des OGM ou des antennes de téléphonie mobile, décider s'il faut vacciner toute une population contre la grippe A.

- Essais thérapeutiques : décider si un nouveau traitement médical est meilleur qu'un ancien au vu du résultat de son expérimentation sur des malades.

- Finance : au vu du marché, décider si on doit ou pas se lancer dans une opération financière donnée.

- Justice : décider si l'accusé est innocent ou coupable à partir des informations acquises pendant le procès.

Dans chaque cas, le problème de décision consiste à trancher, au vu d'observations, entre une hypothèse appelée hypothèse nulle, notée H_0, et une autre hypothèse dite hypothèse alternative, notée H_1. En général, on suppose qu'une et une seule de ces deux hypothèses est vraie. Un test d'hypothèses est une procédure qui permet de choisir entre ces deux hypothèses [7, 14].

Pour atteindre une décision objective concernant une hypothèse particulière, nous devons suivre une procédure objective (méthodes publiques et répétables par d'autres chercheurs) permettant soit d'accepter soit de rejeter cette hypothèse. Cela consiste à formuler, en termes probabilistes, un jugement sur une hypothèse relative à une population, à partir des résultats observés sur un échantillon extrait au hasard de cette population [6, 15].

Cette procédure suit les étapes suivantes :

1- Etablir l'hypothèse nulle (H0) (considérer l'hypothèse alternative H1).

2- Choisir le test statistique approprié pour tester H0, 3- Spécifier un niveau de signification (alpha) et la taille de l'échantillon (N),

4- définir la région de rejet

5- Calculer la valeur du test statistique à l'aide des données de l'échantillon.

1. L'hypothèse nulle

C'est la première étape de la procédure. L'hypothèse nulle H_0 est une hypothèse de non différence (il n'y a pas de différence significative entre les échantillons A et B). Elle est formulée de façon à être rejetée. Dans le cas de son rejet, l'hypothèse alternative H_1 (il y a une différence significative entre les échantillons A et B) doit être acceptée. Cette dernière est la prédiction dérivée de la théorie à tester. Un test d'hypothèse constitue donc une sorte de démonstration par l'absurde en probabilité [15].

Supposons qu'une théorie scientifique nous conduise à prédire que deux régions en Algérie diffèrent par leur taux de chômage. Cette prédiction sera notre hypothèse de recherche. Pour tester cette hypothèse de recherche, nous la formulons en hypothèse alternative H_1. Cette dernière pose que la moyenne des jeunes sans emplois est différent dans les deux régions $m1 \neq m2$, alors que pour H_0 la moyenne est la même $m1=m2$. Si les données nous permettent de rejeter H_0, alors H_1 peut être acceptée, et cela supportera l'idée de la validité de l'hypothèse de recherche et de sa théorie sous-jacente.

La nature de l'hypothèse de recherche détermine comment H_1 doit être formulée :

Si elle pose que deux groupes différeront simplement par leur moyenne, alors H_1 est telle que $m1 \neq m2$. Les tests statistiques seront *bilatéraux.*

Au contraire, si la théorie prédit la direction de la différence, c'est-à-dire qu'un des groupes spécifiés aura une moyenne supérieure à celle de l'autre groupe, alors H_1 est telle que : soit $m1 > m2$, soit $m1 < m2$. Les tests applicables seront alors *unilatéraux*.

Il en résulte, que pour un même niveau de signification, la différence $x1 - x2$, doit être moins élevée pour être significative dans le cas unilatéral que dans le cas bilatéral. Les tables statistiques donnent les valeurs statistiques critiques dans les deux cas.

Les tables statistiques (ou les logiciels statistiques) fournissent les valeurs statistiques critiques dans les deux cas. Pour tous les tests, on définit donc une hypothèse nulle. Le calcul de probabilité p correspond à la probabilité que l'hypothèse nulle soit vraie (ou à la probabilité de se tromper en rejetant l'hypothèse nulle). Si $p>0{,}05$ (5%) ou $p>0{,}01$ (1%), on ne peut pas rejeter l'hypothèse nulle. On dit qu'on a une différence non significative entre les deux échantillons [16].

2. Choix du test statistique

On dispose actuellement de nombreux tests statistiques différents qui peuvent être utilisés pour arriver à une décision concernant une hypothèse. Le choix doit se faire sur des bases rationnelles. Un des critères de choix est la puissance du test utilisé. Mais d'autres critères sont importants pour déterminer l'adéquation d'un test lors de l'analyse de données particulières. Ces critères concernent :

- la façon dont l'échantillon a été réalisée,
- la nature de la population de laquelle a été tiré l'échantillon et
- la nature des mesures réalisées.

A chaque test statistique est associé un modèle et des contraintes de mesure. Ce test n'est alors valide que si les conditions imposées par le modèle et les contraintes de mesure sont respectées. Il est difficile de dire si les conditions d'un modèle sont remplies, et le plus souvent nous nous contentons d'admettre qu'elles le sont.

Il est également très important de considérer la nature des données (observations) que l'on va tester. D'elle dépend la nature des opérations possibles et donc des statistiques utilisables dans chaque situation. Les observations peuvent être soit quantitatives soit qualitatives.

- Les données quantitatives comprennent les dénombrements (ou comptages) et les mesures (ou mensurations).
- Les données qualitatives peuvent être réalisées dans deux échelles de mesure : échelle de rangement et l'échelle nominale. Ces données ne sont pas manipulables par l'arithmétique.

3. Niveau de signification et la taille de l'échantillon

L'ensemble des valeurs observées pour lesquelles l'hypothèse nulle est admissible forme la région d'acceptation ou de non-rejet et les autres valeurs constituent la région de rejet ou domaine de rejet ou région critique. Mais le hasard de l'échantillonnage peut fausser les conclusions. Quatre situations doivent être envisagées :

- l'acceptation de l'hypothèse nulle alors qu'elle est vraie,

- le rejet de l'hypothèse nulle alors qu'elle est vraie,

- l'acceptation de l'hypothèse nulle alors qu'elle est fausse,

- le rejet de l'hypothèse nulle alors qu'elle est fausse.

Dans le premier et le dernier cas, la conclusion obtenue est correcte, mais non dans les deux cas intermédiaires. L'erreur qui consiste à rejeter une hypothèse vraie est appelée erreur de première espèce et celle commise en acceptant une hypothèse fausse est l'erreur de seconde espèce. Idéalement, alpha et bêta devraient être déterminés par l'expérimentateur préalablement à la recherche, ce qui détermine la taille de l'échantillon (N). Une diminution du risque alpha, augmente le risque bêta pour tout échantillon donné. La probabilité de commettre l'erreur de seconde espèce décroît lorsque la taille de l'échantillon augmente [17].

Pratiquement, on se donne une limite supérieure du risque de première espèce, le plus souvent 5% (significatif), 1% (très significatif) ou 1 pour mille (hautement significatif). Cette limite constitue aussi le niveau de signification du test et permet de définir la condition de rejet de l'hypothèse nulle. Le plus souvent, les logiciels de statistique donnent le niveau de signification réel. On rejette alors l'hypothèse nulle au niveau de signification nominal choisi (par exemple 0,05) si (et seulement si) le niveau de signification réel est inférieur ou égal au niveau de signification nominal ($p = 0,003 < 0,05$, rejet d'h_0). Cette attitude est dite conservatrice [18].

Le risque de première espèce étant donné, on peut s'efforcer de calculer le risque de deuxième espèce, grâce à la notion de puissance de test (P = 1-bêta). Mais ce problème possède rarement une solution simple et l'on perd souvent de vue l'existence même de ce risque. Cependant, la puissance d'un test dépend de la nature du test choisi, du niveau de signification du test, de la taille de l'échantillon, de la vraie valeur du paramètre ou mesure testée. En particulier, elle est liée à la nature de l'hypothèse alternative H_1. Comme nous l'avons déja dit, un test unilatéral est plus puissant qu'un test bilatéral. Aussi, souvent on se contente de préciser l'importance du risque de première espèce, sans se soucier de l'existence d'une seconde possibilité d'erreur [5, 19].

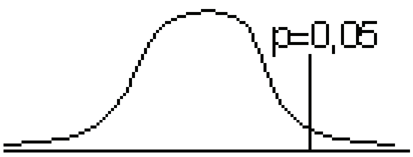
4. Région de rejet

Cette région est constituée par le sous-ensemble des valeurs de la distribution d'échantillonnage qui sont si extrêmes que lorsque H_0 est vrai, la probabilité que l'échantillon observé ait une valeur parmi celles-ci est très faible (la probabilité est alpha).

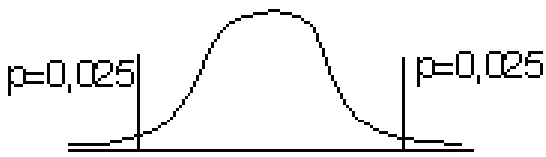
La position de cette région de rejet est affectée par la nature de H1, mais non pas sa taille :

Dans un test unilatéral, la région de rejet est entièrement située à une des extrémités de la distribution d'échantillonnage, alors que dans un test bilatéral, cette région est située aux deux extrémités de la distribution.

La taille de cette région de rejet est définie par alpha. Si alpha est = 0,05 (5%), la taille de la région de rejet correspond à 5% de l'espace inclus dans la courbe de la distribution d'échantillonnage. Cela signifie que dans d'une distribution suivant une loi normale, il n'y a que 5 chances sur 100 pour que l'écart entre la variable et sa valeur moyenne dépasse 2 fois l'écart-type [20].



Région de rejet d'un test unilatéral



Région de rejet d'un test bilatéral

5. **La décision**

Si le test statistique donne une valeur comprise dans la région de rejet, nous rejetons H0 (on adopte alors H1). Quand la probabilité associée à une valeur du test statistique est inférieure ou égale à la valeur alpha préalablement déterminée, nous concluons que H0 est faux. En effet, en rejetant l'hypothèse nulle au niveau 0,05, par exemple, nous avons 5 chances sur 100 seulement d'aboutir à une telle conclusion par le simple fait du hasard. Cette valeur est dite significative [6, 21].