U.F.R. SPSE Master 1
PMP STA 21 Méthodes statistiques pour l'analyse des données en psychologie

Fiche. Tests d'hypothèses statistiques

Préalable

Dans le cadre d'une recherche, on formule une hypothèse de recherche sur la population étudiée et on cherche à la valider à l'aide d'observations.

Test : méthode statistique permettant de confirmer ou d'infirmer l'hypothèse formulée sur la base d'observations faites sur un (ou plusieurs) échantillon(s) aléatoire(s) d'individus.

!! Au départ de l'étude, il faut définir précisément le problème à traiter :

population(s) et variable(s) étudiée(s) ; formuler l'hypothèse de recherche que l'on souhaite tester.

Sélectionner le test le plus approprié. Transcrire l'hypothèse de recherche dans le cadre du test que l'on va utiliser.

Les conditions d'application d'un test doivent être vérifiées : conditions sur les distributions des variables (par ex. loi normale), égalité des variances... Si les conditions du test ne sont pas vérifiées et que le résultat du test peut en être fortement affecté, considérer un test alternatif.

Mise en oeuvre d'un test

1. *Poser les deux hypothèses du test :*

   Un test oppose deux hypothèses : l'hypothèse nulle H_0 et l'hypothèse alternative H_1.

   L'hypothèse H_1 est en général celle que l'on cherche à valider (hypothèse de recherche transcrite dans le cadre du test). On l'oppose à l'hypothèse nulle H_0 qui est l'hypothèse de référence. On teste H_0 contre H_1.

   On admet provisoirement que H_0 est vraie. On rejette H_0 au profit de H_1 si, au vu des observations, la conviction que H_0 est fausse est « suffisamment » forte.

2. *Fixer le niveau α du test :*

   C'est le risque de 1re espèce, c-à-d la probabilité de rejeter à tort H_0, c-à-d de choisir H_1 alors qu'elle est fausse.

   On le fixe au départ, risque petit : $5\%; 1\%; 0,1\%$.

3. *Définir la statistique du test et sa loi sous H_0 :*

   La statistique du test est une variable aléatoire fonction des observations. Elle fournit un résumé des données qui permet de quantifier « l'écart à H_0 » observé sur l'échantillon. La décision du test est basée sur sa valeur observée.

   Pour effectuer le test, il est nécessaire de connaître sa loi sous H_0.

4. *Déterminer la région critique (de rejet) du test :*

   Pour préparer la décision, on partitionne le domaine de variation de la statistique en deux régions : l'intervalle d'acceptation (IA) et la région critique (RC).

   La région critique est l'ensemble des valeurs de la statistique qui amènent à rejeter H_0 et accepter H_1. Elle est associée au niveau α choisi. On rejette H_0 si la valeur observée de la statistique est dans la région critique.

   - IA : contient les valeurs de la statistique que l'on s'attend à observer quand H_0 est vraie ; valeurs conformes à H_0.
   - RC : contient les valeurs
     - qu'il est rare d'observer quand H_0 est vraie (rares et donc suspectes) : ce sont en général les valeurs les plus extrêmes ;
     - et que l'on s'attend à observer quand H_1 est vraie.
       Selon H_1 et le type de test, la RC est à une extrémité du domaine ou aux deux extrémités.

- les bornes (à calculer) de la RC sont déterminées par le niveau α choisi.
  Par ex. pour $\alpha = 5\%$, la RC contient les 5% de valeurs les plus extrêmes.
- Les bornes dépendent de la distribution de la statistique sous H_0.
  Cette distribution doit être connue et tabulée pour déterminer numériquement les bornes.

5. *Alternativement au calcul de la RC, calculer la p-valeur (ou niveau de signification)* α_{obs}
   Le calcul de la p-valeur rend inutile la détermination numérique de la région critique.

   Plus la valeur de la statistique est extrême dans le sens indiqué par H_1 et plus elle renforce la conviction que H_0 est fausse.

   p-valeur : probabilité quand H_0 est vraie d'observer une valeur de la statistique au moins aussi extrême, dans le sens indiqué par H_1, que la valeur effectivement observée.

   - On utilise la distribution sous H_0 de la statistique pour la calculer.
   - Plus la valeur observée est extrême et plus la p-valeur est petite.

6. *Règle de décision basée sur la p-valeur*
   Pour un test au niveau α donné :
   Si $\alpha_{obs} \leq \alpha$: on rejette H_0 et on accepte H_1 , au risque d'erreur α.
   Si $\alpha_{obs} > \alpha$: on conserve H_0 et on n'accepte pas H_1 , avec un risque d'erreur β inconnu.

Risques d'erreur et puissance d'un test

- *Risque d'erreur de 1^{re} espèce* α : risque de rejeter à tort H_0 (d'accepter H_1 quand H_1 est fausse). Il est fixé par l'utilisateur.
- *Risque d'erreur de 2^{e} espèce* β : risque de conserver à tort H_0 (de ne pas accepter H_1 quand H_1 est vraie). En général, il n'est pas calculable mais il diminue quand la taille de l'échantillon augmente.
- *Puissance du test* $= 1 - \beta$: probabilité d'accepter H_1 quand H_1 est vraie. La puissance d'un test augmente quand la taille de l'échantillon augmente.

Exemple. Test de Student sur la moyenne

Le problème traité

Dans une étude destinée à évaluer une thérapie cognitive de groupe s'adressant à des hommes ayant subi des agressions sexuelles durant l'enfance (ASE), les chercheurs ont constitué un échantillon de 24 hommes ayant vécu des ASE.
L'échelle TSC-40 a été utilisée pour mesurer le niveau de symptômes traumatiques de chaque patient.
Plus le score sur l'échelle est élevé et plus les symptômes traumatiques sont sévères.
Pour chaque patient, on a relevé le score au début de la thérapie (score T_1), le score à la fin de la thérapie (score T_2), puis on a calculé la différence entre les deux scores (variable $X = T_1 - T_2$).
Hypothèse de recherche : Les chercheurs veulent montrer qu'il y a une amélioration de l'état des patients en fin de thérapie. Pour cela, il veulent tester l'hypothèse d'une diminution du niveau de symptômes traumatiques entre le début et la fin de la thérapie.

Préalable

On étudie la population des hommes « ASE » qui suivent une thérapie cognitive de groupe.
L'évolution du niveau de symptômes traumatiques entre le début et la fin de la thérapie est mesurée à l'aide de la variable *quantitative* $X = T_1 - T_2$.
La variable X étant quantitative, on peut envisager le test standard de comparaison d'une moyenne à une valeur théorique. Le test porte sur la moyenne μ de la variable.
Conditions d'application du test :
L'échantillon étant de petite taille ($n = 24 < 30$), le test sur la moyenne est basé sur la loi de Student et n'est applicable que si la variable X se distribue selon la loi normale.
Les calculs faits ci-dessous ne sont donc valides que si la normalité de la variable est satisfaite.
Dans la suite, on admet donc que la variable X se distribue dans la population selon une loi normale $N(\mu\,;\sigma)$.
Les paramètres μ et σ sont *inconnus*.

Déroulement du test

- **Hypothèses du test** :
  $$\begin{cases} H_0 : \mu = 0 \\ H_1 : \mu > 0 \text{ unilatérale droite} \\ \text{On fixe le niveau du test : } \alpha = 1\,\%. \end{cases}$$
  En effet, sur un individu une diminution du niveau de symptômes traumatiques se traduit par une baisse du score TSC-40 et donc par une différence de scores $x = t_1 - t_2 > 0$.
  L'hypothèse H_1 correspond à une diminution des scores entre le début et la fin de la thérapie et donc à une amélioration de l'état des patients.
- **Observations :**
  Notons qu'au départ, on dispose de deux échantillons appariés de mesures : pour chaque homme de l'échantillon, on dispose d'une paire de mesures (score-début, score-fin).
  On est ramené à un échantillon unique de mesures après avoir calculé des $n = 24$ différences de scores x_i.
  Les différences de scores sont résumées par :
  la moyenne observée de l'échantillon $\bar{x} = 4,257$ et l'écart-type sans biais $s^* = 12,08$.
- **Statistique du test :**
  Pour le test, on utilise la statistique *moyenne empirique* $\overline{X}_n$. La décision du test est basée sur sa valeur $\bar{x}_{\text{obs}} = 4,257$ observée sur l'échantillon.
- **Loi de la statistique du test sous H_0 :**
  On a admis que la variable X se distribue dans la population selon une loi normale. Sous H_0, la moyenne de X est 0.
  On sait alors que sous H_0, la statistique $\overline{X}_n$ suit la loi normale $N(0\,;\dfrac{\sigma}{\sqrt{n}})$.

Pour les calculs, c'est la loi centrée réduite qui doit être utilisée. Comme l'écart-type σ n'est pas connu, on centre et on réduit $\overline{X}_n$ par la formule

$$T_n = \frac{\overline{X}_n - 0}{\frac{S_n^*}{\sqrt{n}}}$$

Sous H_0, la statistique T_n suit la loi de Student à $n-1=23$ ddl.

- **Méthode 1. IA pour T_n, au risque $\alpha = 1\,\%$:**

  Orientation de la région critique dans le domaine de T_n :
  Sous H_0, on s'attend à observer une valeur de $\overline{X}_n$ proche de 0.
  Par contre, sous H_1, on s'attend à observer une valeur de $\overline{X}_n$ plus grande que 0.
  Ce sont les valeurs positives de $\overline{X}_n$ les plus grandes qui sont les plus significatives de H_1.
  On place donc la région critique RC à l'extrémité droite du domaine de $\bar{X}_n$.
  Le raisonnement est le même pour la statistique centrée réduite T_n.
  IA pour T_n :
  $IA =]-\infty\,; t_{1-\alpha}] =]-\infty\,; 2,5]$ où $t_{1-\alpha} = 2,5$ est le quantile d'ordre $1-\alpha = 0,99$ de la loi de Student à 23 ddl.
  Décision :
  On calcule d'abord la valeur observée de T_n :
  $t_{\text{obs}} = \dfrac{4,257-0}{\frac{12,08}{\sqrt{24}}} = 1,726.$
  Comme $t_{\text{obs}} \in IA$, on conserve H_0 et on ne valide pas H_1. Le risque d'erreur associé à cette décision est inconnu (risque β) .
  Conclusion. Le résultat du test est non significatif au niveau $\alpha = 1\,\%$. A ce niveau, on ne peut donc pas accepter l'hypothèse d'une diminution du niveau de symptômes traumatiques entre le début et la fin de la thérapie, c'est-à-dire conclure à une amélioration de l'état des patients en fin de thérapie.

- **Méthode 2. Décision basée sur le niveau de signification du test :**

  Pour un test unilatéral droit, on a :
  $\alpha_{\text{obs}} = P_{H_0}\left(\bar{X}_n \geq 4,257\right) = P_{H_0}\left(T_n \geq 1,726\right).$
  La valeur numérique de α_{obs} a été déterminée à l'aide du module `Calculateur de Probabilités` du logiciel `Statistica` :
  $\alpha_{\text{obs}} = 0,0489 = 4,89\,\%.$
  Comme $\alpha_{\text{obs}} = 4,89\% > \alpha = 1\%$, on ne peut pas rejeter H_0...

  Au niveau 1%, on ne peut conclure à une baisse significative du niveau des symptômes traumatiques. Cependant, si l'on fait le test au niveau 5%, le résultat devient significatif.

Remarque sur le test utilisé :
Il s'agit au départ d'un test *de comparaison de 2 moyennes sur 2 échantillons appariés tirés de la même population* : le principe consiste à travailler sur la variable différence $T1 - T2$ (ou $T2 - T1$) et à comparer sa moyenne à 0. On se ramène ainsi au test de comparaison d'une moyenne à une valeur théorique.

Tests alternatifs

La normalité de la variable X doit être testée à partir des données observées, avant de pouvoir effectuer le test sur la moyenne. Plusieurs tests de normalité existent (voir TD).
Si la normalité est rejetée, il faut envisager un test alternatif ne requérant pas la normalité de la variable X. Deux tests alternatifs (dits *non-paramétriques*) existent : le test de Wilcoxon et le test du signe pour 2 échantillons appariés. Chacun a toutefois ses propres conditions de validité , moins fortes que celle du test de Student : loi de X continue et symétrique pour le test de Wilcoxon et loi continue pour le test du signe.