IMBS3 - ISCID-CO, site de Dunkerque, 2015/2016

Mathématiques appliquées à la gestion - TESTS PARAMÉTRIQUES ET NON PARAMÉTRIQUES

Fiche de Mathématiques 2 - Notions générales sur les tests.

1 Notions générales

1.1 Hypothèses, erreurs et puissance

Une utilisation courante des statistiques est la notion de test. Un test est un mécanisme qui permet de trancher entre deux hypothèses au vu des résultats d'un échantillon. Dans les cas qui nous intéressent, ces hypothèses porteront sur des estimations (valeur d'un moment, égalité de variances, nature d'une loi de probabilité ...). Soient H_0 et H_1 ces deux hypothèses, dont une et une seule est vraie. La décision aboutira à choisir H_0 ou H_1. Il y a donc 4 cas possibles dont les probabilités sont résumées dans le tableau suivant :

	H_0 vraie	H_1 vraie
H_0 décidée	$1-\alpha$	β
H_1 décidée	α	$1-\beta$

α et β sont les **erreurs** (ou **risques d'erreurs) de première et deuxième espèces.**

Definition 1.1 *L'erreur de première espèce α est la probabilité de décider H_1 alors que H_0 est vraie,*

$$\boxed{\alpha = P(\textit{rejeter } H_0/H_0 \textit{ vraie}) = P(\textit{accepter } H_1/H_1 \textit{ fausse})}$$

Exercice 1 Si on cherche à tester l'hypothèse qu'une pièce de monnaie n'est pas "truquée", on adopte la règle de décision suivante, H_0 : "la pièce n'est pas truquée" est

- acceptée si $X \in [40, 60]$,
- rejetée si $X \notin [40, 60]$ (donc soit $X < 40$ ou $X > 60$),

avec X désignant le nombre de "faces" obtenues en lançant 100 fois la pièce.
Quel est dans ce cas le risque d'erreur de première espèce α?

Definition 1.2 *L'erreur de deuxième espèce β est la probabilité de décider H_0 alors que H_1 est vraie.*

$$\boxed{\beta = P(\textit{accepter } H_0/H_1 \textit{ vraie}) = P(\textit{rejeter } H_1/H_0 \textit{ fausse})}$$

Exercice 2 On reprend l'exemple précédent de la pièce de monnaie, et on suppose que la probabilité p d'obtenir "face" est de $0,6$ pour une pièce truquée. On adopte la même règle de décision que précédemment, H_0 : "la pièce n'est pas truquée" est

- acceptée si $X \in [40, 60]$,
- rejetée si $X \notin [40, 60]$,

avec X le nombre de "faces" obtenues en lançant 100 fois la pièce.
Quel est dans ce cas le risque d'erreur de seconde espèce β?

Ces deux erreurs sont antagonistes, plus α sera grand (resp. petit), plus β sera petit (resp. grand). Le fait d'imposer un α faible conduit à une règle de décision plus stricte qui aboutit le plus souvent à n'abandonner l'hypothèse H_0 que dans des cas rarissimes et donc à conserver cette hypothèse quelques fois à tort. Le compromis entre les valeurs de α et β est donc souhaitable bien que difficile à réaliser.

La figure FIG.1 présente, dans le cas d'une variable U (la statistique utilisée) suivant une loi normale centrée réduite, la probabilité $1-\alpha = P(|U| < V_S/H_0$ est vraie) où V_S désigne la valeur seuil déterminée par le risque α ainsi que la probabilité $\beta = P(|U| < V_S/H_1$ est vraie).

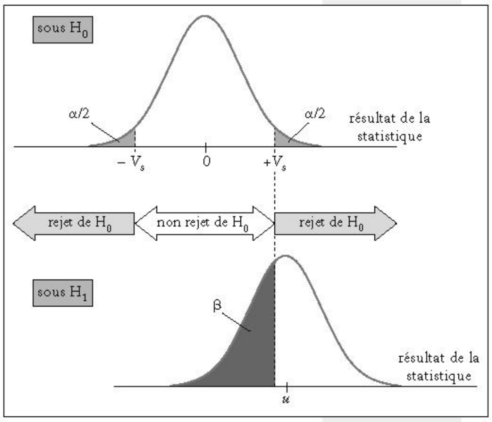

Figure 1: *Types d'erreurs possibles dans un test.*

Dans la pratique des tests statistiques, il est de règle de se fixer α comme donné (les valeurs les plus courantes sont 0.01, 0.05 ou 0.1) de préférence en fonction du risque de première espèce. En effet, H_0 joue le plus souvent un rôle prédominant par rapport à l'hypothèse H_1. Cela est la conséquence du fait que H_0 joue le rôle d'hypothèse de référence alors que H_1 est souvent limitée à l'hypothèse contraire. Par exemple, on peut avoir H_0 : "$m = m_0$" ce qui est relativement facile à tester et dans ce cas, H_1 est tout simplement "$m \neq m_0$".

Les tests ne sont pas faits pour démontrer H_0 mais pour rejeter H_0.

Dans certains cas, il n'est pas évident de formuler les hypothèses nulle et alternative. Il faut donc être très attentif à la formulation des hypothèses, afin d'être sûr qu'elles sont appropriées et que les conclusions du test d'hypothèses fournissent bien les informations souhaitées par le chercheur ou le responsable. On se placera dans les cas suivants :

- **Tester des hypothèses de recherche**:
  Dans des études comme celles-ci, les hypothèse nulle et alternative doivent être formulées de manière à ce que le rejet de H_0 conduise à la conclusion souhaitée.

  Exemple 1.1 On considère un modèle de voiture particulier qui consomme, en moyenne, un litre d'essence tous les 24 kilomètres. Un groupe de recherche a mis au point un nouveau moteur spécialement conçu pour augmenter le nombre de kilomètres effectués avec un litre d'essence. Pour évaluer les performances du nouveau moteur, plusieurs exemplaires du prototype ont été construits, installés sur des voitures et soumis à des tests de conduite par les groupes de recherche (le groupe de recherche cherche donc à prouver que le nouveau moteur augmente en moyenne le nombre de kilomètres effectués avec un litre d'essence). Par conséquent, les hypothèse nulle et alternative appropriées à cette étude sont :

  $$H_0 : \text{“}\mu \leq 24\text{”} \text{ et } H_1 : \text{“}\mu > 24\text{”}$$

  avec μ désignant la distance moyenne parcourue avec un litre d'essence par le nouveau moteur.

- **Tester la validité d'une assertion**:
  L'hypothèse nulle correspond généralement à l'hypothèse selon laquelle l'assertion est vraie.

  Exemple 1.2 On considère l'exemple d'un producteur de boisson non alcoolisée qui prétend que les bouteilles de deux litres contiennent en moyenne 2,028 litres de son produit, au minimum. Un échantillon de bouteilles de deux litres est sélectionné et leur capacité de contenance est mesurée pour tester l'affirmation du fabricant. Les hypothèses nulle et alternative sont formulées de la façon suivante :

  $$H_0 : \text{“}\mu \geq 2,028\text{”} \text{ et } H_1 : \text{“}\mu > 2,028\text{”}$$

  avec μ désignant la capacité de contenance moyenne mesurée.

- **Tester des hypothèses dans un contexte de prise de décision**:
  Dans les tests d'hypothèse de recherche ou de validité d'une assertion, des mesures sont prises lorsque H_0 est rejetée. Cependant, dans de nombreux cas, des mesures doivent être prises à la fois quand H_0 n'est pas rejetée et quand H_0 est rejetée. En général, ce type de situation apparaît lorsq'un responsable doit choisir entre deux actions, l'une associée à l'hypothèse nulle et l'autre associée à l'hypothèse alternative.

  Exemple 1.3 Sur la base d'un échantillon de pièces, issues d'une livraison qui vient d'être reçue, un inspecteur de contrôle de la qualité doit décider d'accepter ou de refuser la livraison, en fonction des critères de qualité établis. Supposons que les critères de qualité d'une pièce particulière correspondent à une longueur moyenne de deux centimètres. Si la moyenne est supérieure ou inférieure à deux centimètres, les pièces poseront problème dans le processus d'assemblage. Dans ce cas les hypothèses nulle et alternative sont formulées de la façon suivante :

  $$H_0 : \text{``}\mu = 2\text{''} \text{ et } H_1 : \text{``}\mu \neq 2\text{''}.$$

  avec μ désignant la mesure moyenne en centimètres des pièces de la livraison.

$\boxed{\textbf{Exercice 3}}$ Le responsable de la direction automobile d'une marque connue étudie un nouveau système de bonus destiné à accroître le volume des ventes. Actuellement, le volume moyen des ventes est de 14 automobiles par mois. Le responsable veut mener une étude pour évaluer l'effet du nouveau système de bonus sur le volume des ventes. Pour collecter des données sur les ventes avec le nouveau système de bonus, un échantillon du personnel des ventes a pu vendre des automobiles en utilisant le nouveau système de bonus pendant un mois.

1. Déterminer les hypothèses nulle et alternative les plus appropriées pour cette recherche.
2. Commenter le résultat obtenu lorsqu'on ne peut pas rejeter H_0.
3. Commenter le résultat obtenu lorsqu'on peut rejeter H_0.

Definition 1.3

- *Dans le cas où la valeur θ du paramètre inconnu θ est θ_0 sous H_0 et θ_1 sous H_1, θ se réduit à une seule valeur, l'hypothèse est dite alors simple.*
- *Au contraire, si H_0 : "$\theta = \theta_0$" et H_1 : "$\theta \neq \theta_0$" ou "$\theta > \theta_0$" ou "$\theta < \theta_0$" alors l'hypothèse est dite multiple.*

Definition 1.4 *L'aptitude d'un test à rejeter H_0 alors qu'elle est fausse constitue la puissance du test. On appelle puissance d'un test la quantité*

$$\boxed{1 - \beta = P(\textit{rejeter } H_0/H_0 \textit{ est fausse}) = P(\textit{accepter } H_1/H_1 \textit{ est vraie})}$$

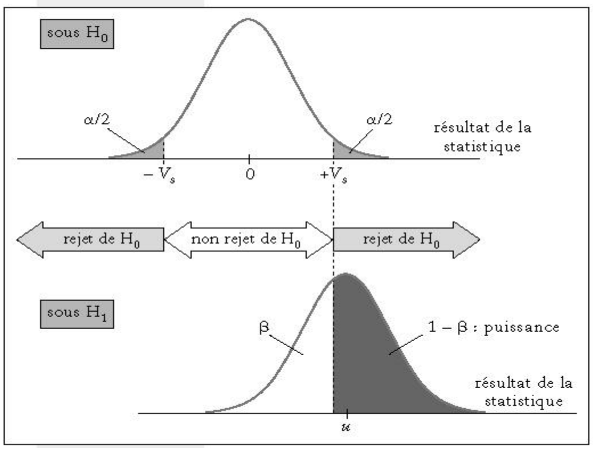

Figure 2: *Puissance d'un test*

Exercice 4 On se place dans le contexte des exercices 1.1 et 1.1, quelle est la puissance du test lorsque les probabilités d'obtenir "face" sont respectivement $0.3, 0.4, 0.5, 0.6, 0.7, 0.8$ pour une pièce truquée? Que constate t-on?

Exprimée simplement, la puissance d'un test d'hypothèses est la probabilité de discerner une différence lorsqu'elle existe. On peut alors facilement concevoir qu'un chercheur souhaite entreprendre son étude dans les meilleures conditions, c'est-à-dire avec une puissance à priori maximale. Cependant, le calcul de la puissance d'un test d'hypothèses est relativement complexe car il fait intervenir différents paramètres statistiques. On présente ci-dessous les variations de la puissance en fonction de certains de ces paramètres statistiques.

- **La variation de la puissance en fonction de α.**
  Supposons que le même test d'hypothèse soit réalisé, d'une part au risque α_1 et, d'autre part, au risque α_2, avec $\alpha_2 < \alpha_1$. Considérons V_{s1} la valeur seuil déterminée avec le risque α_1 et V_{s2} la valeur seuil déterminée avec le risque α_2 et $V_{s2} > V_{s1}$. La diminution du risque α augmente la valeur seuil, entraînant de ce fait une diminution de la puissance du test, la puissance et α varient donc de la même manière.
  On illustre cette variation en considérant la loi normale centrée réduite. $\alpha_1 = 5\%$ donne $V_{s1} = 1,96$ et $\alpha_2 = 1\%$ donne $V_{s2} = 2,58$ (voir FIG. 3).

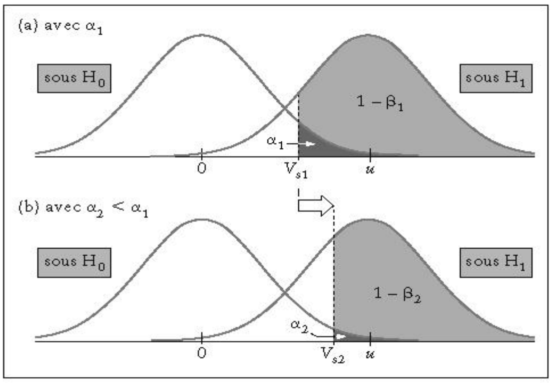

Figure 3: *Variation de la puissance en fonction de α.*

- **La variation de la puissance en fonction de la taille de l'échantillon.**
  La précision d'une estimation augmente avec la taille de l'échantillon. Graphiquement, cela se traduit par un resserrement de la courbe de distribution autour de la valeur estimée. Ainsi, sous H_0, la courbe de la distribution théorique d'un échantillon de taille $n_2 > n_1$ se resserre autour de son paramètre central et, pour conserver un risque α, la valeur seuil du test diminue. La courbe de distribution sous H_1 se resserrant également du fait de l'augmentation de n, il s'ensuit une augmentation de la puissance, la puissance et n varient de la même manière (voir FIG. 4).
- **La variation de la puissance en fonction de l'écart des paramètres sous H_0 et sous H_1.**
  Plus un test est puissant et plus petit peut être l'écart entre le paramètre sous H_0 et celui sous H_1, la puissance et l'écart entre H_0 et H_1 varient de la même manière (voir FIG. 5).

1.2 Test bilatéral et unilatéral

Avant d'appliquer tout test statistique, il s'agit de bien définir le problème posé. En effet, selon les hypothèses formulées, on applique soit un test bilatéral, soit un test unilatéral. La nature de H_0 détermine la façon de formuler H_1 et par conséquent la nature unilatérale ou bilatérale du test.

Definition 1.5 *Un test bilatéral s'applique quand on cherche une différence entre deux estimations, ou entre une estimation et une valeur donnée sans se préoccuper du signe ou du sens de la différence. Dans ce cas, la zone de rejet (cf. section suivante) de l'hypothèse principale se fait de part et d'autre de la distribution de référence.*

Exemple 1.4 Si H_0 consiste à dire que la population estudiantine avec la fréquence de fumeurs p est représentative de la population avec la fréquence de fumeurs p_0, on pose alors :

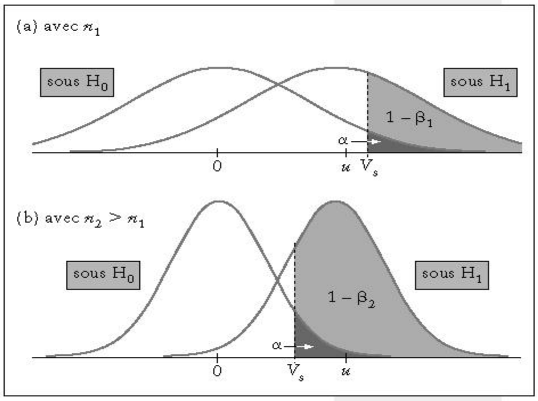

Figure 4: *Variation de la puissance en fonction de la taille de l'échantillon*

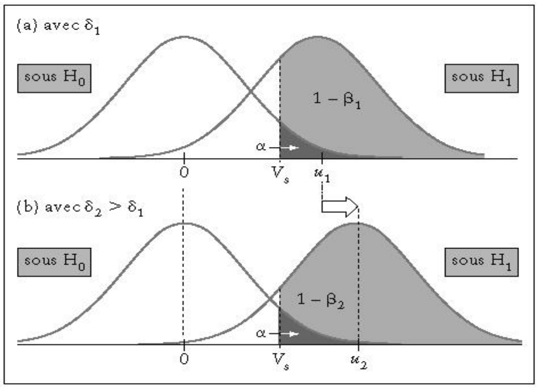

Figure 5: *Variation de la puissance en fonction de l'écart entre les paramètres sous H_0 et sous H_1.*

$$H_0 : \text{“}p = p_0\text{”} \;\text{ et } H_1 : \text{“}p \neq p_0\text{”}.$$

Definition 1.6 *Un test unilatéral s'applique quand on cherche à savoir si une estimation est supérieure (ou inférieure) à une autre ou à une valeur donnée. La zone de rejet de l'hypothèse principale est située d'un seul côté de la distribution de probabilité de référence. Certains tests comme l'analyse de la variance ou le test du χ^2 sont pratiquement toujours unilatéraux.*

Exemple 1.5 Si on fait l'hypothèse que la fréquence de fumeurs dans la population estudiantine p est supérieure à la fréquence de fumeurs dans la population p_0, on pose alors

$$H_0 : \text{“}p = p_0\text{”} \;\text{ et } H_1 : \text{“}p > p_0\text{”}.$$

Le raisonnement inverse peut être formulé avec l'hypothèse suivante : H_0 : “$p = p_0$” et H_1 : “$p < p_0$”.

1.3 Région d'acceptation et région critique

Definition 1.7

- *On appelle région critique, et on note W, l'ensemble des valeurs de la variable de décision qui conduisent à écarter H_0 au profit de H_1. On peut relier W à α par $p(W/H_0) = \alpha$.*
- *On appelle région d'acceptation, et on note $\overline{W}$, la région complémentaire de la région critique. On a alors $p(\overline{W}/H_0) = 1 - \alpha$.*

On a également des relations avec l'erreur de deuxième espèce : $p(W/H_1) = 1 - \beta$ et $p(\overline{W}/H_1) = \beta$

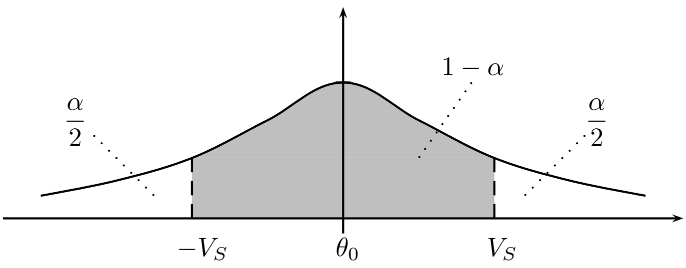

Figure 6: *Test bilatéral*

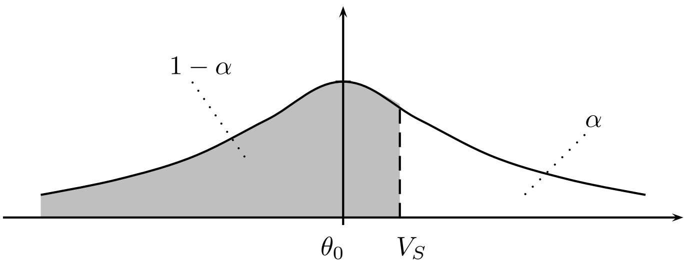

Figure 7: *Test unilatéral*

La zone ou région d'acceptation correspond à l'intervalle dans lequel les différences observées entre les réalisations et la théorie sont attribuables aux fluctuations d'échantillonnage. La région critique ou zone de rejet correspond donc aux intervalles dans lesquels les différences sont trop grandes pour être le fruit du hasard d'échantillonnage.

	Test unilatéral H_0 : "$\theta = \theta_0$"		Test bilatéral H_0 : "$\theta = \theta_0$"
Hypothèse alternative	H_1 : "$\theta > \theta_0$"	H_1 : "$\theta < \theta_0$"	H_1 : "$\theta \neq \theta_0$"
Valeur de $U = \theta - \theta_0$ sous H_1	$U > 0$	$U < 0$	$\|U\| \neq 0$
Niveau de signification α	$P(U > V_S) = \alpha$	$P(U < V_S) = \alpha$	$P(\|U\| > V_S) = \alpha$

La construction d'un test est la détermination à priori de la région critique sans connaître le résultat de l'expérience. On peut donc résumer cette démarche de la manière suivante:

- Choix de H_0 et H_1
- Détermination de la variable de décision (choix du test)
- Allure de la région critique en fonction de H_1
- Calcul de la région critique en fonction de α
- Calcul éventuel de la puissance du test $1 - \beta$
- Calcul expérimental de la variable de décision
- Conclusion du test: rejet ou acceptation de H_0

1.4 Choix d'un test

Plusieurs tests de conception très différente sont souvent disponibles pour soumettre à une épreuve de vérité une hypothèse principale. Dans un tel cas, le test qui fournit l'erreur β la plus petite, pour une même valeur de α, est par définition le plus puissant (celui ayant la plus grande valeur de la puissance de test $1 - \beta$). En effet, il peut détecter les plus petites différences entre les populations sans pour autant augmenter l'erreur de première espèce.

La majorité des tests statistiques repose sur le respect d'un certain nombre de conditions. Selon le degré de respect de ces conditions d'application, la validité des résultats se trouve plus ou moins affectée et elle l'est d'autant plus que le test est moins robuste. Ainsi, la robustesse d'un test équivaut à sa tolérance vis-à-vis du respect des conditions.

Si le statisticien dispose de plusieurs tests pour vérifier une hypothèse, il choisira bien sûr le plus puissant et le plus robuste.

Les tests peu puissants augmentent la probabilité de commettre une erreur de deuxième espèce. Or, cette erreur peut s'avérer particulièrement grave. En effet, en médecine par exemple, une analyse qui classerait comme malade un individu bien portant peut avoir des conséquences aussi graves qu'une analyse qui classerait comme bien portants des individus malades (erreur de première espèce). Dans de tels cas, il y a intérêt à tracer la courbe de puissance du test, aussi appelée courbe caractéristique d'efficacité, qui indique la probabilité de prendre une bonne décision si H_1 est vraie. La puissance est mesurée par la valeur de $1-\beta$ pour un α donné.

1.5 Influence de l'échantillonnage

Pour comparer les moyennes, les variances ou les autres paramètres estimés de deux échantillons, il faut prendre en considération la technique conduisant à la constitution des deux échantillons. Si la sélection des éléments est aléatoire, et si le choix des éléments du premier échantillon n'a aucune influence sur le choix des éléments du second, les deux échantillons sont alors appelés indépendants.

Si l'on prélève aléatoirement des paires d'éléments, et non les éléments eux-mêmes, on constitue deux échantillons appariés. Dans ce cas, le premier élément de chaque paire appartient au premier échantillon et le deuxième est affecté au second. Parfois, la paire d'éléments peut se rapporter au même individu sur lequel on mesure la même variable à deux occasions différentes, par deux moyens différents par exemple.

La technique de l'échantillonnage apparié présente l'avantage d'éliminer un maximum de sources de variations non reliées au facteur que l'on étudie. En règle générale, plus les critères d'appariement des données sont nombreux, plus grand sera cet avantage.

2 Exercices

$\boxed{\textbf{Exercice 5}}$ Après une campagne électorale, l'institut de sondage affirme que Mr X est passé de 39% à 41% d'intention de vote en sa faveur. Mr X vous demande alors de vérifier cette affirmation en choisissant l'une des deux hypothèses :

$$H_0 : \text{“}p = 39\%\text{”} \text{ et } H_1 : \text{“}p = 41\%\text{”}.$$

1. En considérant un 100-échantillon, quel est le risque de première espèce α correspondant à ce test?
2. En posant $\alpha = 5\%$, quelle est la région d'acceptation de l'hypothèse H_0?
3. Calculer le risque de seconde espèce β sachant que $\alpha = 5\%$.
4. Le 100-échantillon indique que 40 personnes ont l'intention de voter pour Mr X. Que répondre à Mr X?
5. Que doit être l'effectif n (supposé élevé) de l'échantillon pour qu'une fréquence f_n d'intention de vote pour le candidat de $0,4$ permette de ne pas donner raison à l'institut de sondage?
6. On considère maintenant le test suivant :

$$H_0 : \text{“}p = 0,39\text{”} \text{ et } H_1 : \text{“}p = 0,42\text{”}.$$

   Cacluler la taille de l'échantillon n (supposée toujours élevée) pour avoir un risque β inférieur à $0,1$.

$\boxed{\textbf{Exercice 6}}$ La proportion de grévistes lors des dernières manifestations était de 53% lorsque les leaders syndicaux réussissaient à convaincre les populations concernées, et elle était de 47% si au contraire, les gens n'étaient pas convaincus.
Après une intervention télévisée, un leader syndical désire savoir si son intervention était réussie ou non. Il interroge pour ce fait 100 personnes et recueille 51 avis favorables à la grève.

1. Quel est le risque de première espèce α?
2. En posant $\alpha = 5\%$, déterminer la région d'acceptation de l'hypothèse nulle H_0.
3. Le leader syndical a t-il réussi à convaincre les gens au risque $\alpha = 5\%$?
4. Quelle est dans ces conditions l'erreur de seconde espèce β?

5. Calculer alors la puissance du test η
6. Quel doit être l'effectif de l'échantillon à interroger pour qu'une fréquence de 51% de grévistes permette de conclure que le leader syndical a convaincu les gens?

Exercice 7 Dans un centre de renseignements téléphoniques, une étude statistique a montré que l'attente (en secondes) avant que la communication soit amorcée suit une loi normale de moyenne 18 et d'écart-type 7,2.
Après une réorganisation du service une étude est effectuée pour contrôler s'il y a eu une amélioration du temps d'attente

Attente	$[2,6[$	$[6,10[$	$[10,12[$	$[12,14[$	$[14,18[$	$[18,26[$	$[26,34[$
Effectif	8	14	9	11	30	16	12

1. Construire un test permettant de décider si le temps d'attente a diminué depuis la réorganisation du service (seuil de risque de 2%).
2. Calculer la probabilité critique liée à ce test.
3. Conclure à l'aide du sondage effectué.

Exercice 8 Soit la variable X distribuée suivant une loi uniforme sur $[0,c]$. On veut tester H_0 : "$c=1$" vs H_1 : "$c>1$". On recueille une seule observation et on rejette H_0 si $x_1 > 0,9$.

1. Quel est le risque α de ce test?
2. Quelle est la puissance du test si en réalité $c = 1,5$?

Exercice 9 On s'intéresse à la durée de vie des pneus de marque A (en dizaine de milliers de kilomètres). On désire tester H_0 : "$\mu = 50$" vs H_1 : "$\mu < 50$" au niveau 5%. On suppose $X \rightsquigarrow \mathcal{N}(\mu, 25)$.

1. Quelle est la puissance du test si $n = 9$ et $\mu = 45$?
2. Quelle valeur de n assure une puissance supérieure à $0,9$?