CY-Tech	Statistiques	Mathématiques
2020 - 2021	TD 3 (Corrigé)	ING2-GI

Exercice 1. -

Une machine produit des billes de roulement de diamètre fixe. Si elle fonctionne normalement, il y a une proportion de 5% de billes défectueuses. Si elle est déréglée, la proportion de billes défectueuses passe à 10%.

Avant d'envoyer une commande à son client, le fabricant teste un lot de $n = 500$ billes. Soit X la variable aléatoire valant 1 si la bille est défectueuse et 0 sinon. On sait que X suit une loi de Bernouilli $\mathcal{B}(p)$ où $p = 0.05$ si la machine fonctionne correctement et $p = 0.1$ si la machine est déréglée.

1. Il décide de ne pas livrer son client si la machine est déréglée, c'est-à-dire si le nombre de billes défectueuses est supérieur ou égal à 50 ($= 500 \times 10\%$).
   (a) Exprimer les hypothèses (H_0) et (H_1) ainsi que les risques de première et de deuxième espèce.
   (b) Préciser la variable de décision et la valeur du seuil.
   (c) Calculer le risque de première espèce (α).
   (d) Calculer la puissance du test et en déduire si un client doit acheter ou non un lot ayant subi un tel test.
2. Devant les protestations du client, le fabricant re-définit son test en imposant un risque de première espèce de $\alpha = 1\%$.
   (a) Déterminer la région critique du test.
   (b) Calculer sa puissance.
   (c) Enoncer les règles de décision avec les erreurs associées.
   (d) Le client sera-t-il enclin à accepter ce nouveau test.

Correction 1. -

Fonctionnement normal : 5% de billes défectueuses
Fonctionnement anormal : 10% de billes défectueuses
Variable de décision
$F_n =$ proportion de billes défectueuses sur un lot de $n =$ 500 billes (fréquence empirique).
$n = 500$ est assez grand pour approcher la loi de F_n grâce au TCL.
On considère donc que F_n suit une $\mathcal{N}\left(\mu, \sigma^2\right)$ avec $\mu = p$ et $\sigma^2 = \dfrac{p(1-p)}{n}$.
$p = 0.05$ si la machine fonctionne correctement et $p = 0.1$ si elle est déréglée.

1. Test « naif » du fabricant.
   (a) Les hypothèses

$$\begin{cases} (H_0) & p = p_0 = 0.05 \quad \text{Machine bien réglée} \\ (H_1) & p = p_1 = 0.10 \quad \text{Machine déréglée} \end{cases}$$

   Les risques (erreurs) encourus sont
   Risque de 1ère espèce : Accepter (H_1) alors que (H_0) est vraie $\Longleftrightarrow$ ne pas livrer de la marchandise de bonne qualité
   Risque de 2ème espèce : Accepter (H_0) alors que (H_1) est vraie $\Longleftrightarrow$ livrer de la marchandise de mauvaise qualité (perdre client)

   (b) Variable de décision : F_n et seuil : $C = \dfrac{50}{500} = 0.1$

   avec les règles de décision :
   — Si $F_n < 0.1$, (< 50 billes défectueuses sur 500) on accepte l'hypothèse (H_0) , on livre la marchandise

— Si $F_n > 0.1$, (> 50 billes défectueuses sur 500) on valide (H_1), la marchandise n'est pas livrée.

(c) Risque de 1ère espèce :

On sait que $= P\,(\text{Accepter } (H_1)\ /(H_0) \text{ vraie}) = P\,(F_n \geq 0.1\ /(H_0) \text{ vraie})$.

Or sous (H_0) on a $\quad p = p_0 = 0.05 \quad$ et $\quad F_n \sim \mathcal{N}\left(p_0, \dfrac{p_0(1-p_0)}{n}\right)$

Donc : $$Z = \frac{F_n - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} \sim \mathcal{N}(0,1)$$

$$\alpha = P_{H_0}(F_n \geq 0.1) = P\left(\frac{F_n - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} \geq \frac{0.1 - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}\right) = P\left(Z \geq \frac{0.1 - 0.05}{\sqrt{0.05(0.95)}}\sqrt{500}\right)$$

$\alpha = P(Z \geq 5.13) = 1.4 \times 10^{-7}$

Le fabricant ne prend donc pratiquement aucun risque de ne pas livrer de la bonne marchandise.

(d) Calcul de la puissance :

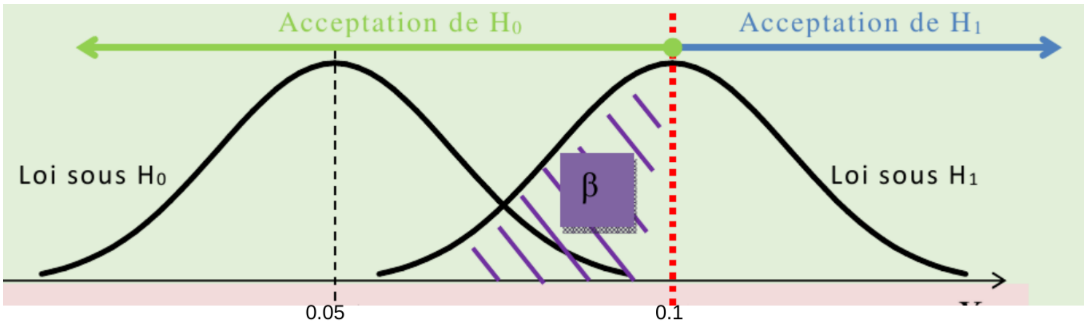
On sait que $\beta = P\,(\text{Accepter } (H_0)\ /(H_1) \text{ vraie}) = P_{H_1}(F_n < 0.1)$

Donc $\beta = P\left(\dfrac{F_n - p_1}{\sqrt{p_1(1-p_1)}}\sqrt{n} < \dfrac{0.1 - p_1}{\sqrt{p_1(1-p_1)}}\sqrt{n}\right) = P(Z < 0) = 0.5 = 50\%$!!!!??

Le fabricant a une chance sur deux de livrer de la marchandise de mauvaise qualité!!!

Cette règle de décision n'est pas du tout à l'avantage du client et risque de faire perdre sa clientèle au fabricant.

Puissance du test : $1 - \beta = 0.5 = 50\%$



2. Le fabricant a donc fixé $\alpha = 0.01$, c'est cette valeur qui va permettre de déterminer le seuil C et donc la région critique W.

(a) $\alpha = P_{H_0}(F_n \geq C) = P\left(\dfrac{F_n - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} \geq \dfrac{C - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}\right) = P\left(Z \geq \dfrac{C - 0.05}{\sqrt{0.05(0.95)}}\sqrt{500}\right)$

La table de $\mathcal{N}(0,1)$ donne la valeur $\quad t = z_{1-\alpha} = 2.33 \quad$ telle que $\quad 1 - \alpha = 0.99 = P(Z \leq t)$

Donc $\quad t = \dfrac{C - 0.05}{\sqrt{0.05(0.95)}}\sqrt{500} \quad$ et par suite $\quad C = 0.05 + 2.33\dfrac{\sqrt{0.05(0.95)}}{\sqrt{500}} = 0.073$

Le seuil est donc de $\quad C = 0.073 \quad$ et la région critique est : $\quad W = \{F_n > 0.073\}$

(b) $\beta = P_{H_1}(F_n < 0.073) = P\left(\dfrac{F_n - p_1}{\sqrt{p_1(1-p_1)}}\sqrt{n} < \dfrac{0.073 - p_1}{\sqrt{p_1(1-p_1)}}\sqrt{n}\right)$

$\beta = P\left(Z < \dfrac{0.073 - 0.1}{\sqrt{0.1(0.9)}}\sqrt{500}\right) \beta = P(Z < \ -2.01) = 1 - P(Z < 2.01) \simeq 0.022$

Le partage des risques semble moins choquant : $\alpha = 1\%$ de risque de ne pas livrer de la bonne marchandise et $\beta = 2.2\%$ de risque de livrer de la marchandise de mauvaise qualité.

(c) Les règles de décision sont les suivantes :

— Si $F_n < 0.073$, (≤ 36 billes défectueuses sur 500) on accepte l'hypothèse (H_0) , on livre la marchandise

— Si $F_n > 0.073$, (≥ 37 billes défectueuses sur 500) on valide (H_1) , la marchandise n'est pas livrée.

(d) Ce deuxième test semble plus acceptable pour le client, mais il pourrait souihaiter avoir exactement le même niveau de risque que le fournisseur.

Exercice 2. -

Les habitants d'une région aéroportuaire se plaignent que le bruit des avions dépasse la limite autorisée de 80 décibels en moyenne imposée par la législation. On admet que l'intensité du bruit causé par les avions est une variable aléatoire X de loi gaussienne d'espérance μ et de variance $\sigma^2 = 64$.
On mesure un échantillon journalier de $n = 16$ variables aléatoires indépendantes $(X_1, ..., X_n)$ de l'intensité du bruit, et on effectue le test statistique suivant :

$$\begin{cases} (H_0) & \mu = \mu_0 = 80 \text{ décibels} \\ (H_1) & \mu = \mu_1 = 85 \text{ décibels} \end{cases}$$

1. Expliciter les risques de première et deuxième espèces. De quel point de vue est fait ce test ? Celui des habitants ou celui des responsables de l'aéroport ?
2. Quelle statistique (variable de décision) faut-il choisir et quelle est sa loi ?
3. Déterminer graphiquement l'allure de la région critique et représenter sur le graphique les erreurs de première et deuxième espèces.
4. Calculer le seuil de la région critique pour un risque $\alpha = 5\%$.
5. Calculer la puissance du test.
6. Enoncer les règles de décision avec les probabilités d'erreur.
7. La moyenne calculée sur l'échantillon est $\overline{x} = 83$ décibels. Les habitants ont-ils raison de se plaindre ? Le test d'hypothèses ainsi établi leur est-il favorable ou défavorable ?
8. Combien faudrait-il faire de relevés journaliers, pour que le risque de deuxième espèce soit de 5% ?
9. Quelle serait alors le seuil de décision ?

Correction 2. -

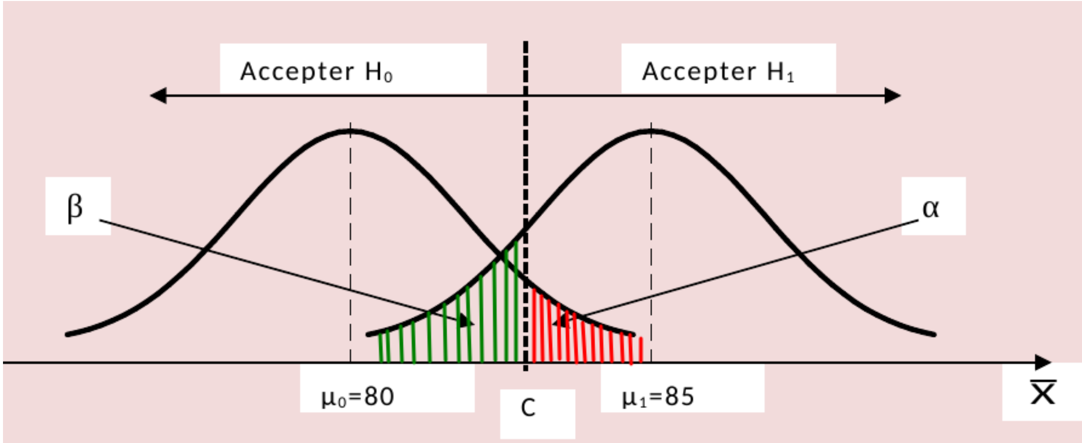
1. Les hypothèses à tester sont :

$$\begin{cases} (H_0) & \mu = \mu_0 = 80 \text{ dB} \\ (H_1) & \mu = \mu_1 = 85 \text{ dB} \end{cases}$$

   Risque de 1ère espèce : Accepter (H_1) alors que (H_0) est vraie $\Longleftrightarrow$ n« décider que l'aéroport n'est pas aux normes alors qu'il respecte la législation »

   Risque de 2ème espèce : Accepter (H_0) alors que (H_1) est vraie $\Longleftrightarrow$ « décider que l'aéroport est aux normes alors que le bruit dépasse le seuil autorisé »

   Le test est donc fait du point de vue des responsables de l'aéroport.
2. La moyenne empirique $\overline{X}$ est l'estimateur usuel de μ. L'échantillon est gaussien et $\sigma^2 = 64$ est connu. Donc $$\overline{X} \sim \mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$$

3. 

La région critique est donc de la forme : $W = \{\overline{X} > C\}$

4. sous l'hypothèses (H_0), $\mu = 80$.

$\alpha = P_{H_0}(\overline{X} \geq C) = P_{H\,0}\left(\frac{\overline{X}-\mu}{\sigma}\sqrt{n} \geq \frac{C-\mu}{\sigma}\sqrt{n}\right) = P\left(Z \geq \frac{C-80}{8}\sqrt{16}\right)$

Or, d'après la table de la loi normalr, $\alpha = 0.05 = P(Z \geq 1.645)$

Donc : $\frac{C-80}{8}\sqrt{16} = 1.645 \quad \Longrightarrow \quad C = 80 + 1.645\frac{8}{\sqrt{16}}80 + 1.645 \times 2 = 83.3$

Le seuil est donc $C = 83.3$ décibels.

5. sous l'hypothèses (H_1), $\mu = 85$.

$\beta = P_{H\,1}(\overline{X} \leq C) = P_{H\,0}\left(\frac{\overline{X}-\mu}{\sigma}\sqrt{n} \leq \frac{C-\mu}{\sigma}\sqrt{n}\right) = P\left(Z \leq \frac{83.3-85}{8}\sqrt{16}\right)$

$\beta = P(Z \leq -0.85) = 1 - P(Z \leq -0.85) = 1 - 0.80 = 0.2$

La puissance de ce test est donc $1 - \beta = 0.8 \ = 80\%$.

6. Les règles de décision sont :

• Si $\overline{x} > 83.3$ décibels alors on accepte (H_1), i.e on considère que l'aéroport n'est pas aux normes, avec 5% de chance de se tromper.

• Si $\overline{x} < 83.3$ alors on garde (H_0), i.e on considère que l'aéroport est aux normes, avec 20% chance de se tromper.

7. La moyenne calculée sur l'échantillon est $\overline{x} = 83 < 83.3$. Donc les habitants n'ont pas de raison de se plaindre. Cependant, le risque de deuxième espèce étant très élevé, ils sont en droit de remettre en cause le test.

8. Supposons que la taille de l'échantillon ne soit pas fixée. Alors le calcul fait pour un risque de 1ère espèce de $\alpha = 5\%$ donne l'équation :

$$C = 80 + 1.645\frac{8}{\sqrt{n}}$$

Et le calcul du risque de deuxième espèce donne :

$\beta = 0.05 = P\left(Z \leq \frac{C-85}{8}\sqrt{n}\right)$ et donc : $\frac{C-85}{8}\sqrt{n} = -1.645,$ soit :

$$C = 85 - 1.645\frac{8}{\sqrt{n}}$$

La soustraction des deux formules pour C donne : $\sqrt{n} = \frac{2 \times 1.645 \times 8}{85-80} = 5.264$

$n = (5.264)^2 = 27.7$

Il faudrait donc tester un échantillon de taille $n = 28$.

9. Le seuil est alors égal à : $C = 80 + 1.645\frac{8}{\sqrt{28}} = 82.5$

Le résulat, milieu du segment [80 ; 85] est logique car les deux gaussiennes sont symétriques l'une par rapport à 80 et l-'autre par rapport à 85.

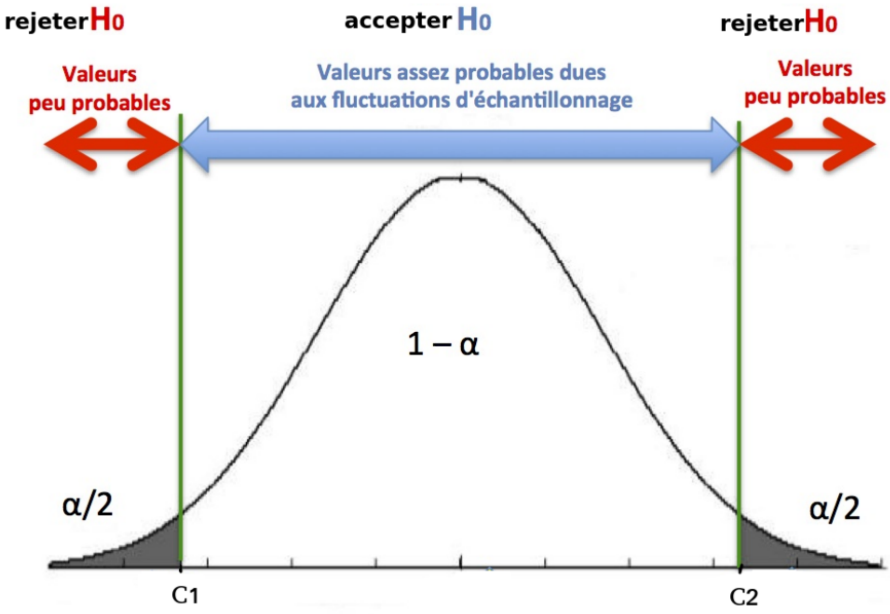
Exercice 3. -

Un fabricant de conserves de petits pois produit des boîtes où l'étiquette annonce un poids net égoutté de 560 gr. Il souhaite construire un test pour s'assurer, d'une part qu'il n'aura pas d'ennui à l'issu d'un contrôle éventuel, et d'autre part, que le poids moyen des boîtes n'est pas excédentaire. Pour ce faire, il compte prélever un lot de $n = 25$ boites et relever le poids moyen ainsi que l'écart-type.

1. Déterminer les hypothèses et expliciter les risques de première et deuxième espèces.
2. Quelle est la statistique (variable de décision) ? Préciser sa loi.
3. Déterminer graphiquement l'allure de la région critique.
4. Calculer les seuils de la région critique sachant que le risque de première espèce est de 10%.
5. Peut-on calculer la puissance du test ?
6. Il prélève un lot de 25 boites et il pèse un poids moyen de $\overline{x} = 556$ gr avec un écart-type empirique $s^* = 10$ gr . Quelle décision doit-il prendre ?

Correction 3. -

1. Les hypothèses à tester sont :
$$\begin{cases} (H_0) \quad \mu = \mu_0 = 560 \text{ gr} \\ (H_1) \quad \mu = \mu_1 \neq 560 \text{ gr} \end{cases}$$
   — Risque de 1ère espèce : Accepter (H_1) alors que (H_0) est vraie $\Longleftrightarrow$ « refuser de mettre en vente des boites qui sont pourtant conformes »
   — Risque de 2ème espèce : Accepter (H_0) alors que (H_1) est vraie $\Longleftrightarrow$ « décider de vendre des boites qui ne sont pas conformes »
2. La variable de décision est $\overline{X}$, estimateur usuel de μ.
   La taille de l'échntillon, $n = 25$, n'est pas suffisante pour utiliser le TCL. Nous devons donc supposer $X \sim \mathcal{N}(\mu, \sigma^2)$ avec μ et σ^2 inconnus.
   Par conséquent : $T = \dfrac{\overline{X} - \mu}{S^*}\sqrt{n} \sim T_{n-1}$ loi de Student à $n - 1 = 24$ d.d.l.
3. L'hypothèse alternative est $(H_1) : \mu = \mu_1 \neq 560$ gr. Il s'agit donc d'un test bilatéral.
   (H_1) sera logiquement validée pour de petites valeurs aussi bien que pour de grandes valeurs $\overline{X}$.
   La région critique sera donc de la forme : $W = \left\{\overline{X} \leq C_1 \text{ ou } \overline{X} \geq C_2\right\}$



4. Les deux seuils sont symétriques par rapport à la valeur de $\mu = E(X) = E(\overline{X}) = 560$ sous (H_0).
$C_1 = \mu - \ell$ et $C_2 = \mu + \ell$. Avec $\alpha = 10\%$, on va avoir :
$0.9 = 1 - \alpha = P\ (\mu - \ell \leq \overline{X} \leq \mu + \ell) \ = \ P(-\ell \leq \overline{X} - \mu \leq \ell\ )$
$$0.9 = P\ \left(-\ell\frac{\sqrt{n}}{S^*} \leq \frac{\overline{X} - \mu}{S^*}\sqrt{n}\ \leq \ell\frac{\sqrt{n}}{S^*}\ \right) = P\ \left(-\ell\frac{\sqrt{n}}{S^*} \leq T\ \leq \ell\frac{\sqrt{n}}{S^*}\ \right)$$
La table de la loi de Student à 24 d.d.l donne une valeur $k = 1.71$ qui vérifie :
$P\ (k \leq T\ \leq k\ \ ) = 0.9$, on en déduit :
$$\ell\frac{\sqrt{n}}{s^*} = 1.71 \qquad \text{et donc} \qquad \ell = 1.71\frac{s^*}{\sqrt{n}} = 1.71\frac{10}{\sqrt{25}} = 3.42$$
Finalement : $C_1 = 560 - 3.42 = 556.58$
et $C_2 = 560 + 3.42 = 563.42$
5. On ne peut pas calucler la puissance du test car son calcul se fait sous l'hypothèse (H_1) et nous n'avons pas de valeur précise de μ sous (H_1).
6. $\overline{x} = 556 < C_1$ donc on doit valider (H_1) et revoir la chaîne de production.

Exercice 4. -

Sur un échantillon de 900 naissances, on constate qu'il y a 470 garçons. Un généticien décide d'utiliser ces données pour tester si la proportion de garçons est significativement plus importante que la proportion de filles dans cette population.
Quelle sera sa conclusion avec un risque de 5% ?

Correction 4. -

$$\begin{cases} (H_0) & p = p_0 = 0.5 \\ (H_1) & p = p_1 > 0.5 \end{cases}$$

(a) Variable de décision
La fréquence empirique F_n est un estimateur de p et avec un échantillon de taille $n = 900$, on peut considérer grâce au TCL que F_n suit une loi normale $F_n\ \sim \mathcal{N}\left(p, \frac{p(1-p)}{n}\right)$ où $p = 0.5$ sous l'hypothèse (H_0) et p inconnu sous l'hypothèse (H_1) .
(b) Allure de la région critique
La région critique= W est la région d'acceptation de (H_1) d'où $W = \{F_n \geq C\}$.
(c) Calcul du seuil
$$\alpha = P_{H_0}(F_n \geq C) = P\left(\frac{F_n - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n} \geq \frac{C - p_0}{\sqrt{p_0(1-p_0)}}\sqrt{n}\right) = P\left(Z \geq \frac{C - 0.5}{\sqrt{0.5(0.5)}}\sqrt{900}\right)$$
La table de $\mathcal{N}(0,1)$ donne la valeur 1.645 telle que $1 - \alpha = 0.95 = P(Z \leq t)$
Donc $t = \frac{C - 0.5}{\sqrt{0.5(0.5)}}\sqrt{900}$ et par suite $C = 0.5 + 1.645\frac{\sqrt{0.5(0.5)}}{\sqrt{900}} = 0.527$
Le seuil est donc de $C = 0.527$ et la région critique est : $W = \{F_n > 0.527\}$
(d) Règles de décision
- Si $f_n > 0.527$ alors on accepte (H_1), i.e on considère qu'il y a plus de garçons que de filles avec 5% de risque de se tromper.
- Si $f_n < 0.527$ alors on garde (H_0), i.e on considère qu'il y a autant de filles que de garçons (sans connaitre le risque se tromper).

L'échantillon considéré indique $f_n = \frac{470}{900} = 0.522 < C$ donc le généticien conclut qu'il y autant de garçons que de filles.

Exercice 5. -

Un fabricant produit des piles dont la durée de vie suit une loi normale d'espérance 80 h et d'écart-type 6.44 h.

Suite à des réclamations, il veut vérifier si la qualité a baissé ou non. Dans le cas où la moyenne de la durée de vie tombe à 75 h, il sera obligé de baisser le prix de vente.
Construire un test lui permettant de prendre une décision sachant qu'il veut que le risque de déclasser des piles de bonne qualité soit de 2,5% et le risque de vendre des piles non conformes soit de 5%.

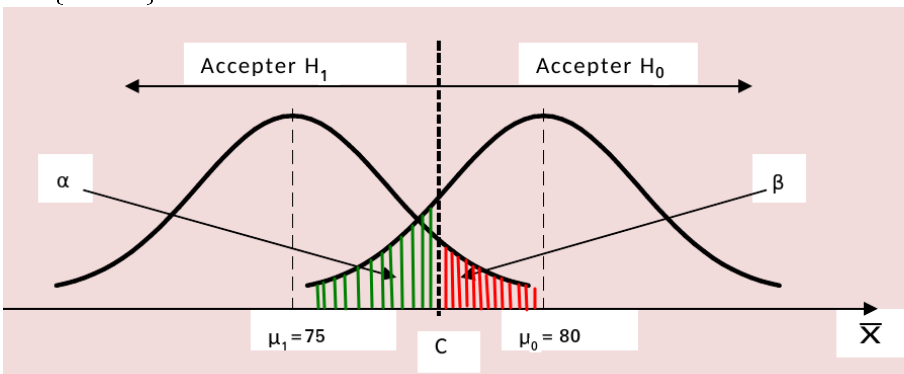
Correction 5. -
X =durée de vie d'une pile. $E(X) = \mu$ et $Var(X) = \sigma^2 = (6.44)^2$
Test sur μ à construire avec $\alpha = 2.5\%$ et $\beta = 5\%$.
(a) Les hypothèses à tester sont :

$$\begin{cases} (H_0) & \mu = \mu_0 = 80 \text{ h} \\ (H_1) & \mu = \mu_1 = 75 \text{ h} \end{cases}$$

Test à construire avec $\alpha = 2.5\%$ et $\beta = 5\%$.
(b) La variable de décision est $\overline{X}$, estimateur usuel de μ.

$X \sim \mathcal{N}\left(\mu, \sigma^2\right)$ par conséquent : $Z = \dfrac{\overline{X} - \mu}{\sigma}\sqrt{n} \sim \mathcal{N}(0,1)$

(c) Région critique :
$W = \{\overline{X} < C\}$



(d) Calcul de seuils :

$$\alpha = P_{H_0}(\overline{X} \leqslant C) = P_{H0}\left(\frac{\overline{X} - \mu}{\sigma}\sqrt{n} \leqslant \frac{C - \mu}{\sigma}\sqrt{n}\right) = P\left(Z \leqslant \frac{C - 80}{6.44}\sqrt{n}\right)$$

$$\alpha = 0.025 \implies \frac{C - 80}{6.44}\sqrt{n} = -1.96$$

$$\beta = P_{H1}(\overline{X} > C) = P_{H0}\left(\frac{\overline{X} - \mu}{\sigma}\sqrt{n} > \frac{C - \mu}{\sigma}\sqrt{n}\right) = P\left(Z > \frac{C - 75}{6.44}\sqrt{n}\right)$$

$$\beta = 0.05 \implies \frac{C - 75}{6.44}\sqrt{n} = 1.645$$

On a donc à résoudre le système : $\begin{cases} C = & 80 - 1.96\frac{6.44}{\sqrt{n}} \\ C = & 75 + 1.645\frac{6.44}{\sqrt{n}} \end{cases}$

La soustraction donne : $5 = \dfrac{6.44}{\sqrt{n}}(1.645 + 1.96)$

D'où : $\sqrt{n} = \dfrac{6.44}{5}3.605 = 4.64 \implies n = 21.56$ arrondi à 22

Ensuite : $C = 75 + 1.645\dfrac{6.44}{\sqrt{22}} = 77.26$

(e) Règles de décision
- Si $\overline{x} < 77,26$ alors on accepte (H_1), i.e on doit baisser le prix des ampoules, avec 2.5% de risque de se tromper.
- Si $\overline{x} > 77,26$ alors on garde (H_0), i.e on vend les piles au prix habituel, avec 5% de risque de se tromper.

Exercice 6. -

L'objectif de cet exercice est d'apprendre à utiliser un test d'hypothèses avec le logiciel R et à interpréter le résultat. Pour cela, il suffit de :

— Connaitre les hypothèses (H_0) et (H_1)
— Vérifier les conditions d'applications du test
— Savoir lire une p-valeur

On rappelle que la p-valeur indique le niveau α minimum du risque de première espèce à partir duquel l'hypothèse (H_0) est rejetée, connaissant la valeur prise par la variable de décision sur l'échantillon,

Si la p-valeur est inférieure à la valeur du seuil préalablement défini (traditionnellement 5 % ou 1 %), on rejette (H_0).

Nous utiliserons ici le test de Student.
S'il est utilisé pour comparer une moyenne μ à une valeur théorique : $(H_0) : \mu = \mu_0$, alors il s'écrit de la façon suivante, ***t.test(data ,mu=μ_0,...).***
Il faut préciser si le test est unilatéral avec l'argument ***alternative = "less"*** ou ***"greater"*** ou bilatéral avec l'argument ***alternative = "two.sided"***.
Le test n'est valable que pour les échantillons gaussiens. Si les échantillons ne sont pas gaussiens, il est peut être possible d'appliquer le test s'ils sont suffisamment grands pour que le théorème de la limite centrale s'applique.
Dans le fichier **Echantillons.zip**, vous trouverez des échantillons de valeurs entre 0 et 1.

1. Calculez la moyenne et l'écart-type des échantillons.
2. Pensez-vous que les conditions d'applications du test de Student soient vérifiées sur ces échantillons ?
3. Pour les échantillons 1, 2, 3 testez : $(H_0) : \mu = 0.5$ contre $(H_1) : \mu > 0.5$
   Comparez l'évolution de la p-valeur en fonction de la moyenne des échantillons.
4. Effectuez le même test pour l'échantillon 4. Que constatez-vous ? Quel test faut-il faire ?
5. Effectuez le même test pour l'échantillon 5. Comment expliquez-vous que la conclusion soit différente de celle de l'échantillon 1 ?

Correction 6. -

Commandes à utiliser avec R-studio :
Mydata=read.table("Echantillon1.txt",header=T)
mean(Mydata)
sd(Mydata)
boxplot(Mydata)
hist(Mydata)
t.test(Mydata ;mu=0.5 ;alternative="greater")

1. -

	Ech1	Ech2	Ech3	Ech4	Ech5
Moyenne	0.54	0.58	0.67	0.40	0.54
Ecart-type	0.32	0.30	0.28	0.31	0.28
P-valeur	0.1837	0.0263	1.17 E-5	0.99157	0.0016

2. L'histogramme ne permet pas de valider que les échantillons sont gaussiens.
   Cependant, ils sont grands et leur distribution est symétrique sans individus atypiques.
   Le test de Student est donc envisageable.
3. Si on prend un risque de $\alpha = 5\%$, alors (H_0) est acceptée pour Ech1 avec une moyenne estimée à 0.54. On peut donc considérer que $\mu = 0.5$
   En revanche, la moyenne de Ech2 est plus grande 0.58 et dans ce cas (H_1) est acceptée avec un P-valeur= $2,6\%$. Avec une moyenne encore plus grande pour Ech3, (H_1) est acceptée plus facilement avec une P-valeur à 10^{-3} %.

4. Avec une moyenne de 0.40, (H_1) ne peut pas être acceptée d'où une P-valeur très grande. En revanche si on teste $\mu < 0.5$, alors (H_1) est acceptée
5. La moyenne de Ech5 est la même que celle de Ech1 et pourtant ici (H_1) est acceptée.
   On ne peut donc pas considérer que $\mu = 0.5$. Cela est dû au fait que Ech5 est de plus grande taille que Ech1 et donc le test est plus « exigeant » avec lui. Car avec un plus grand échantillon on a une meilleure précision.