BTS Mécanique et Automatismes Industriels

Statistiques inférentielles

Lycée Louis Armand, Poitiers, Année scolaire 2005/2006

Statistiques inférentielles

1. Introduction – vocabulaire

Pour étudier une population statistique, on a recours à deux méthodes :

- la *méthode exhaustive* (ou *recensement*) : on examine chacun des éléments de la population. En général, cette méthode est jugée trop longue.
- la *méthode des sondages* : on n'examine qu'une partie de la population pour essayer d'en déduire des informations sur la totalité de la population. Cette méthode comprend deux parties :
  - l'*échantillonage* qui permet de passer de la population totale à une partie seulement de cette population (l'*échantillon*).
  - l'*estimation* qui permet d'induire, à partir des résultats observés sur l'échantillon, des informations sur la population totale.

Nous ne nous préoccuperons pas ici des problèmes concernant l'échantillonage. Notre propos sera seulement d'examiner deux méthodes différentes d'estimation.

2. Principe de la théorie

On considère une population P d'effectif N. On suppose que, pour le caractère observé, la moyenne de P est m alors que son écart-type est σ. Ce sont ces deux valeurs que nous voudrions retrouver à partir des échantillons.

Supposons donc maintenant que nous disposons de k échantillons de P, chacun d'entre eux étant d'effectif n. On note $E_1, E_2, \dots, E_k$ ces k échantillons, de moyennes respectives $\overline{x}_1, \overline{x}_2, \dots, \overline{x}_k$, et d'écart-type respectifs $\sigma_1, \sigma_2, \dots \sigma_k$.

L'ensemble $\overline{X} = \{\overline{x}_1, \overline{x}_2, \dots, \overline{x}_k\}$ est une série statistique d'effectif k, série que l'on appelle *distribution des moyennes*. La théorie montre alors que

$$\boxed{E\left(\overline{X}\right) = m}, \qquad \text{et} \qquad \boxed{\sigma_{\overline{X}} = \frac{\sigma}{\sqrt{n}}}$$

De plus, pour $n > 30$, $\boxed{\text{la variable aléatoire } \overline{X} \text{ suit approximativement une loi normale } \mathcal{N}\left(m, \frac{\sigma}{\sqrt{n}}\right)}$. Autrement dit, $\boxed{\text{la variable aléatoire } \frac{\overline{X} - m}{\sigma/\sqrt{n}} \text{ suit approximativement une loi normale } \mathcal{N}(0, 1)}$.

3. Estimation ponctuelle

• Connaissant la moyenne $\overline{x}$ et l'écart-type σ' d'un échantillon de taille n, il s'agit d'estimer la moyenne m, l'écart-type σ et la variance σ^2 de la population totale.

Pour la moyenne, l'estimation ponctuelle est la méthode naïve qui consiste à confondre la mesure sur l'échantillon avec la moyenne de la population totale. On dira que $\boxed{\overline{x} \text{ est une } \textit{estimation ponctuelle de la moyenne } m}$.

Pour la variance et l'écart type, on admettra que : $\boxed{\frac{n}{n-1}\sigma'^2 \text{ est une } \textit{estimation ponctuelle de la variance } \sigma^2}$ et donc : $\boxed{\sqrt{\frac{n}{n-1}}\sigma' \text{ est une } \textit{estimation ponctuelle de l'écart-type } \sigma}$

• Dans le cas où, pour une population complète, c'est une fréquence p que l'on cherche à estimer à partir de la fréquence f observée sur un échantillon, on brosède comme pour une moyenne. Plus précisément, l'estimation ponctuelle est la méthode qui consiste à confondre la mesure sur l'échantillon avec la mesure de la population totale. Si f est la fréquence, sur l'échantillon, du caractère observé, on dira que $\boxed{f \text{ est une } \textit{estimation ponctuelle de la fréquence } p}$.

4. Estimation d'une moyenne par intervalle de confiance

On considère une population P d'effectif N. On suppose que, pour le caractère observé, la moyenne, inconnue, de P est m alors que son écart-type, connu, est σ. Situation résumée dans le diagramme ci-dessous :

m : inconnu
σ : connu

Échantillon $\boxed{n, \overline{x}}$

Population

On prélève au hasard, et avec remise, une succession d'échantillons de même effectif n dont on calcule les moyennes respectives : $\overline{x}_1$ pour le premier, $\overline{x}_2$ pour le deuxième, et ainsi de suite.

Notons maintenant $\overline{X}$ la variable aléatoire qui associe à un échantillon E_i sa moyenne x_i. La variable $\overline{X}$ prend donc successivement les valeurs $\overline{x}_1, \overline{x}_2, \ldots$

Pour finir, on suppose également que les conditions sont réunies pour pouvoir utiliser une conséquence du théorème de la limite centrée et faire l'approximation que X suit la loi normale $\mathcal{N}(m, \sigma/\sqrt{n})$. Autrement dit que la variable aléatoire $T = \dfrac{\sqrt{n}}{\sigma}\left(\overline{X} - m\right)$ suit la loi normale $\mathcal{N}(0, 1)$. On aura alors, pour tout $t \geqslant 0$,

$$\boxed{P(-t \leqslant T \leqslant t) = 2\Pi(t) - 1}$$

4.1 - Calcul sur un exemple : intervalle de confiance à 95%

Par exemple, si on veut obtenir un intervalle ayant 95% de chances de contenir la moyenne m de la population P, on procède de la manière suivante :

• On a $2\Pi(t) - 1 = 0,95 \iff \Pi(t) = 0,975$. Avec la table donnée dans le formulaire, on voit que cette valeur est obtenue pour $t = 1,96$. On a donc

$$\begin{aligned}
& P\left(-1,96 \leqslant \frac{\sqrt{n}}{\sigma}(\overline{X} - m) \leqslant 1,96\right) = 0,95 \\
\iff \quad & P\left(-1,96\frac{\sigma}{\sqrt{n}} \leqslant (\overline{X} - m) \leqslant 1,96\frac{\sigma}{\sqrt{n}}\right) = 0,95 \\
\iff \quad & P\left(m - 1,96\frac{\sigma}{\sqrt{n}} \leqslant \overline{X} \leqslant m + 1,96\frac{\sigma}{\sqrt{n}}\right) = 0,95
\end{aligned}$$

Autrement dit, **avant de prélever un échantillon** de taille n dans la population, il y a 95% de chances pour que cet échantillon ait une moyenne entre

$$m - 1,96\frac{\sigma}{\sqrt{n}} \qquad \text{et} \qquad m + 1,96\frac{\sigma}{\sqrt{n}}$$

• Comme m est inconnu, on se sert des résultats précédents pour encadrer m :

$$\begin{aligned}
& P\left(-\overline{X} - 1,96\frac{\sigma}{\sqrt{n}} \leqslant -m \leqslant -\overline{X} + 1,96\frac{\sigma}{\sqrt{n}}\right) = 0,95 \\
\iff \quad & P\left(\overline{X} + 1,96\frac{\sigma}{\sqrt{n}} \geqslant m \geqslant \overline{X} - 1,96\frac{\sigma}{\sqrt{n}}\right) = 0,95 \\
\iff \quad & P\left(\overline{X} - 1,96\frac{\sigma}{\sqrt{n}} \leqslant m \leqslant \overline{X} + 1,96\frac{\sigma}{\sqrt{n}}\right) = 0,95
\end{aligned}$$

Ainsi, **avant de prélever un échantillon** de taille n dans la population, il y a 95% de chances pour la moyenne $\overline{x}$ de cet échantillon vérifie

$$\boxed{\overline{x} - 1,96\frac{\sigma}{\sqrt{n}} \leqslant m \leqslant \overline{x} + 1,96\frac{\sigma}{\sqrt{n}}}$$

En revanche, **après** le prélèvement, il n'y a plus de probabilité à envisgager : il est vrai ou faux que la moyenne m se situe dans l'intervalle envisagé $\left[\overline{x} - 1,96\dfrac{\sigma}{\sqrt{n}}, \overline{x} + 1,96\dfrac{\sigma}{\sqrt{n}}\right]$. Cet intervalle est appelé *intervalle de confiance de la moyenne de la population avec le coefficient de confiance* 95% (ou *avec le risque* 5%).

4.2 - Cas général

On fonctionne exactement sur le même principe : un coefficient de confiance choisi à l'avance permet de définir un nombre positif t tel que $P(-t \leqslant T \leqslant t) = 2\Pi(t) - 1$ soit égal à ce coefficient de confiance.

Par exemple, $2\Pi(t) - 1 = 0,99$ si et seulement si $\Pi(t) = 0,995$, ce qui correspond à $t = 2,58$ (d'après la table de la loi normale $\mathcal{N}(0, 1)$).

En reprenant tous les calculs ci-dessus, on obtient alors le résultat suivant :

L'intervalle
$$\left[\overline{x} - t\frac{\sigma}{\sqrt{n}}, \overline{x} + t\frac{\sigma}{\sqrt{n}}\right]$$
est l'intervalle de confiance de la moyenne m de la population avec le coefficient de confiance $2\Pi(t) - 1$, ayant pour centre la moyenne $\overline{x}$ de l'échantillon considéré.

Dans la pratique, on utilise souvent des coefficients de confiance de 95%, ce qui correspond à $t = 1,96$, ou à 99%, ce qui correspond à $t = 2,58$.

5. Estimation d'une fréquence par intervalle de confiance

On considère une population P d'effectif N. On suppose que, pour le caractère observé, la fréquence, inconnue, de P est p. On suppose également que l'on dispose d'un échantillon de atille n de cette population, échantillon sur lequel le caractère observé l'est avec la fréquence f. Situation résumée dans le diagramme ci-dessous :

p : inconnue Échantillon $\boxed{n, f}$

Population

On se place dans le cas où l'on peut considérer que la variable F qui, à tout échantillon aléatoire de taille n fixée, associe la fréquance de cet échantillon possédant la propriété fixée, suit la loi normale $\mathcal{N}\left(p, \sqrt{p(1-p)/n}\right)$.

Un raisonnement et des calculs analoguent à ceux du paragraphe précédent nous donnent alors le résultat suivant :

L'intervalle
$$\left[f - t\sqrt{\frac{f(1-f)}{n-1}}, f + t\sqrt{\frac{f(1-f)}{n-1}}\right]$$
est l'intervalle de confiance de la fréquence p de la population avec le coefficient de confiance $2\Pi(t) - 1$, ayant pour centre la fréquence f de l'échantillon considéré.

Dans la pratique, on utilise souvent des coefficients de confiance de 95%, ce qui correspond à $t = 1,96$, ou à 99%, ce qui correspond à $t = 2,58$.

6. Test de validité d'hypothèse : nature du problème

Depuis quelques décennies, on assiste à une « entrée en force » des méthodes statistiques dans le domaine règlementaire, lequel conduit à la **prise de décision** : on a ou on n'a pas le droit de ...

En particulier, l'augmentation des échanges commerciaux et des liens économiques entre les pays s'accompagne d'accords destinés à fixer les règles communes. Les statistiques inférentielles trouvent là un immense champ d'applications.

Cela se traduit par des règlementations définissant dans chaque cas particulier une procédure destinée à préciser sans ambiguïté :

- comment un ou plusieurs échantillons doîvent être prélevés dans la population étudiée;
- quelles mesures doîvent être effectuées sur ce ou ces échantillon(s);
- quelle décision doît être prise à propos de l'ensemble de la population.

Une telle procédure s'appelle, en statistique, un *test de validité d'hypothèse*.

De manière plus générale, il s'agit, à partir de l'étude d'un ou plusieurs échantillons, de prendre des décisions concernant l'ensemble de la population.

7. Exemple de test : comparaison de la moyenne d'une population à un nombre fixé

Une société s'approvisionne en pièces brutes qui, conformément aux conditions fixées par le fournisseur, doivent avoir une masse moyenne de 780 grammes.

Au moment où 500 pièces sont réceptionnées, on en prélève au hasard un échantillon de 36 pièces dont on mesure la masse.

On obtient les résultats suivants :

Masses des pièces (en grammes)	Nombre de pièces
$[745, 755[$	2
$[755, 765[$	6
$[765, 775[$	10
$[775, 785[$	11
$[785, 795[$	5
$[795, 805[$	2

La masse moyenne des pièces de l'échantillon est de 774, 7 g.

En supposant que l'écart-type des masses pour la population des 500 pièces est $\sigma = 12,5$ g, on obtient $[770,61 ; 778,79]$ comme intervalle de confiance à 95% de la moyenne inconnue m de cette population.

7.1 - Présentation du problème

Peut-on considérer que les 500 pièces de la population ont une masse moyenne de 780 g, comme le prévoient les conditions fixées par le fournisseur ? Autrement dit, doit-on ou non accepter la livraison de ces 500 pièces au vu du résultat obtenu sur l'échantillon ?

7.2 - Hypothèse nulle

On suppose que la moyenne de la population est 780. C'est *l'hypothèse nulle*, notée H_0 : $m = 780$.

Alors, la variable aléatoire $\overline{X}$ qui, à tout échantillon aléatoire non exhaustif de taille $n = 36$, associe la moyenne de cet échantillon suit approximativement la loi normale $\mathcal{N}(780; \sigma/\sqrt{n})$.

Cherchons h réel positif tel que

$$p\left(780 - h \leqslant \overline{X} \leqslant 780 + h\right) = 0,95.$$

Avec la méthode habituelle, on trouve $h = 4,08$, ce qui nous permet de conclure :

$$p\left(775,92 \leqslant \overline{X} \leqslant 784,08\right) = 0,95.$$

Ainsi, en supposant que $m = 780$, on sait, avant de prélever un échantillon aléatoire de taille 36, que l'on a 95% de chances que sa moyenne soit dans l'intervalle $[775,92 ; 784,08]$.

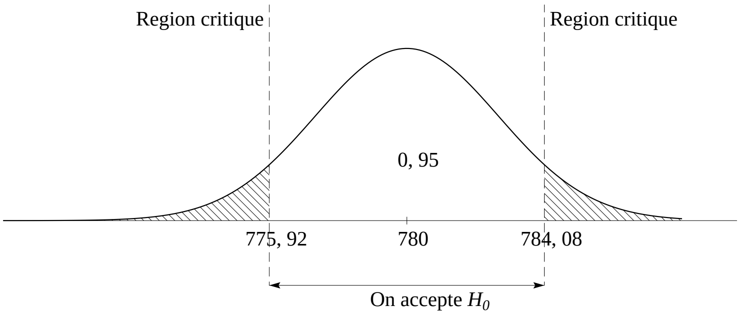
Autrement dit, si H_0 est vraie, il n'y a que 5% de chances de prélever un échantillon aléatoire de taille 36 dont la moyenne soit inférieure à 775, 92 ou supérieure à 784, 08.

7.3 - Règle de décision, région critique

On fixe alors la règle de décision suivante : on prélève avec remise un échantillon aléatoire non exhaustif de taille $n = 36$ et on calcule sa moyenne $\overline{x}$.

Si $\overline{x} \in [775,92 ; 784,08]$, on accepte H_0

Si $\overline{x} \notin [775,92 ; 784,08]$, on rejette H_0



Si H_0 est vraie, on prend donc le risque de se tromper dans 5% des cas en rejetant à tord H_0. On définit ainsi une *région critique au seuil* $\alpha = 5\%$.

Le seuil α est la probabilité de rejeter H_0 alors que H_0 est vraie. Il correspond à l'*erreur de première espèce*.

En général, on fixe *a priori* la valeur de α (ici égal à $0,05$).

Dans l'exemple qui nous occupe, on a $\overline{x} = 774,7$ pour l'échantillon considéré. On a $\overline{x} < 775,92$ et on rejette l'hypothèse H_0. Au seuil de 5%, on considère que les 500 pièces de la population n'ont pas une moyenne de 780 g et on refuse la livraison.

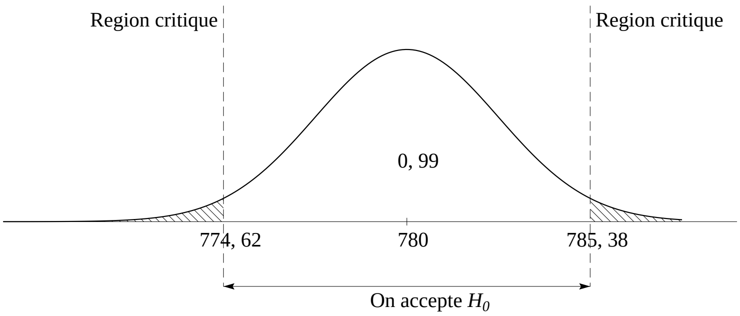
7.4 - Erreur de seconde espèce

On aurait pu choisir un seuil de 1% pour diminuer le risque de rejeter H_0 alors que H_0 est vraie.

On a

$$p(774,62 \leqslant \overline{X} \leqslant 785,38) = 0,99.$$

Au seuil de 1%, on accepte H_0 puisque $\overline{x}$ appartient à l'intervalle considéré, et on accepte alors la livraison des 500 pièces.



Mais, en acceptant H_0 au seuil de 1%, on court un second risque : celui d'accepter H_0 alors que H_0 est fausse : c'est l'*erreur de seconde espèce*, dont la probabilité est notée β.

En général, lorque la taille n de l'échantillon est fixée, on a α qui diminue lorsque β augmente, et réciproquement. Le seule façon de diminuer en même temps α et β est d'augmenter n, ce qui n'est pas toujours possible.

En fait, la plupart du temps, les erreurs des deux types n'ont pas la même importance, et on essaie de limiter la plus grave.

7.5 - Hypothèse alternative

Il faut définir plus précisément le cas où H_0 est fausse.

Dans ce qui précède, on a choisi implicitement $m \neq 780$ comme *hypothèse alternative* H_1. Le test est alors *bilatéral*, car la région critique est située des deux côtés de la région où on accepte H_0.

Si on décide, par exemple, de prendre $m < 780$ comme hypothèse alternative H_1, le test est alors *unilatéral* et la région critique est située entièrement d'un côté de la région où on accepte H_0.

7.6 - Résumé

En général, les questions faisant intervenir un test de validité d'hypothèse peuvent être résolues en adoptant le plan suivant :

1. **Construction du test**
   a) Choix de l'hypothèse nulle H_0 et de l'hypothèse alternative H_1.
   b) Détermination de la région critique à un seuil α donné.
   c) Énoncé de la règle de décision : si un paramètre du ou des échantillon(s) est dans la région critique, on rejette H_0, sinon on l'accepte.
2. **Utilisation du test**
   a) Calcul du paramètre de l'échantillon mentionné dans la règle de décision,
   b) Application de la règle de décision.

8. Exemple de test : comparaison de la moyenne de deux populations

8.1 - Présentation du problème

Un second fournisseur B livre 800 pièces du même modèle. On prélève au hasard et avec remise un échantillon de 50 pièces dont on mesure la masse. On obtient les résultats suivants :

Masses des pièces (en grammes)	Nombre de pièces
$[745, 755[$	6
$[755, 765[$	12
$[765, 775[$	16
$[775, 785[$	11
$[785, 795[$	4
$[795, 805[$	1

La masse moyenne des pièces de l'échantillon est de $779,6$ alors que l'échantillon de 36 pièces provenant du premier fournisseur a pour moyenne $774,7$ g.

La différence de $4,9$ entre ces moyennes provient-elle d'une différence entre les productions des deux fournisseurs ou du choix des échantillons ?

Autrement dit, comment construire et utiliser un test permettant de décider, à partir des échantillons ci-dessus, s'il y a une différence significative, au seuil de 5%, entre les moyennes des masses des pièces livrées par les deux fournisseurs ?

8.2 - Un peu de théorie

Nous sommes en présence de deux échantillons extraits de deux populations correspondant aux deux fournisseurs A et B. On peut schématiser la situation de la façon suivante :

Population A	Échantillon	Population B	Échantillon
m_A : inconnue $\sigma_A = 12,5$	$n_A = 36$ $\overline{x}_A = 774,7$ $\sigma'_A = 12,36$	m_B : inconnue $\sigma_B = 12,1$	$n_B = 50$ $\overline{x}_B = 779,6$ $\sigma'_B = 11,99$

Population A *Population B*

Soit $\overline{X}_A$ (resp $\overline{X}_B$) la variable aléatoire qui, à tout échantillon de taille $n_A = 36$ (resp $n_B = 50$) prélevbé aléatoirement et avec remise dans la population A (resp B), associe la moyenne des masses de pièces de l'échantillon.

On se place dans le cas où $\overline{X}_A$ suit approximativement la loi normale $\mathcal{N}(m_A, \sigma_A/\sqrt{n_A})$ et $\overline{X}_A$ suit approximativement la loi normale $\mathcal{N}(m_B, \sigma_B/\sqrt{n_B})$.

Par définition, la variable aléatoire $D = \overline{X}_B - \overline{X}_B$ associe à tout échantillon de taille 36 ainsi prélevé dans la population A et à tout échantillon ainsi prélevé dans lma population B la différence des moyennes de l'échantillon B et de l'échantillon A.

On suppose que les variables $\overline{X}_A$ et $\overline{X}_B$ sont **indépendantes**.

Alors $D = \overline{X}_B - \overline{X}_A$ suit une loi normale et

$$E(D) = E\left(\overline{X}_B - \overline{X}_A\right) = E\left(\overline{X}_B\right) - E\left(\overline{X}_A\right) = m_B - m_A$$
$$V(D) = V\left(\overline{X}_B - \overline{X}_A\right) = V\left(\overline{X}_B\right) + V\left(\overline{X}_A\right) = \frac{\sigma_B^2}{n_B} + \frac{\sigma_A^2}{n_A}$$

L'écart-type de D est donc $\sqrt{\frac{\sigma_B^2}{n_B} + \frac{\sigma_A^2}{n_A}} \approx 2,7$, et D suit une loi normale $\mathcal{N}(m_B - m_A\,;2,7)$.

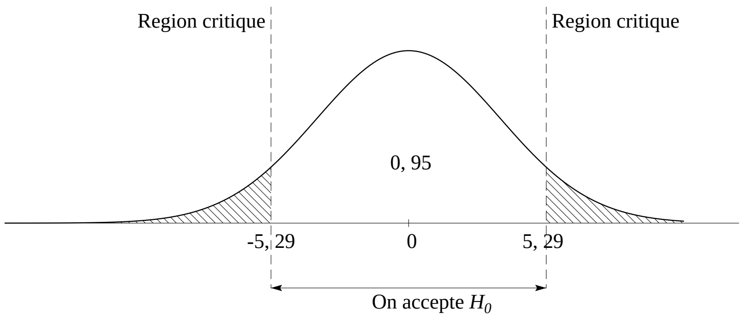
8.3 - Construction du test

- *Choix de* H_0 : $m_A = m_B$.

  Choix de H_1 : $m_A \neq m_B$.

  Nous allons tester la validité de l'hypothèse : « *la moyenne des masses des pièces sur l'ensemble de chaque livraison est la même pour les fournisseurs A et B* ».
- *détermination de la région critique au seuil de* 5%

  Sous l'hypothèse H_0, D suit la loi normale $\mathcal{N}(0\,;2,7)$, donc $D/2,7$ suit la loi normale centrée réduite $\mathcal{N}(0,1)$. En particulier, on a $p(-t \leqslant D/2,7 \leqslant t) = 0,95$ lorsque $t = 1,96$, et donc $p(-5,29 \leqslant D \leqslant 5,29) = 0,95$.



- *Énoncé de la règle de décision*

  On prélève avec remise un échantillon aléatoire de taille $n_A = 30$ de la population A et on calcule sa moyenne $\overline{x}_A$; on fait de même pour la population B avec $n_B = 50$. On pose $d = \overline{x}_B - \overline{x}_A$.

  si $d \in [-5,29\,;5,29]$ on accepte H_0.

  si $d \notin [-5,29\,;5,29]$ on rejette H_0 et on accepte H_1.

8.4 - Utilisation du test

- *Calcul de d*

  On a $d = \overline{x}_B - \overline{x}_A = 779,6 - 774,7 = 4,9$
- *Application de la règle de décision*

  Comme $4,9 \in [-5,29\,;5,29]$, on accepte H_0 : au seuil de 5%, il n'existe pas de différence significative entre les moyennes des masses des pièces livrées par les deux fournisseurs.

9. Un dernier exemple

9.1 - Définition du problème

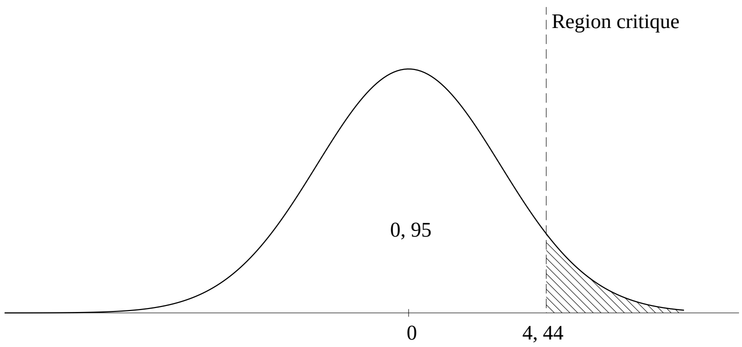
Toujours avec les donnéees précédentes, comment construire et utiliser un test permettant de décider, à partir des mêmes échantillons, si la moyenne des masses des pièces livrées par le fournisseur B est significativement supérieure, au seuil de 5%, à celle du fournisseur A ?

9.2 - Construction du test unilatéral

- *Choix de* H_0 : $m_A = m_B$.

  Choix de H_1 : $m_A > m_B$.
- *détermination de la région critique au seuil de* 5%

  Sous l'hypothèse H_0, D suit la loi normale $\mathcal{N}(0\,;2,7)$, donc $D/2,7$ suit la loi normale centrée réduite $\mathcal{N}(0,1)$. En particulier, on a $p(D/2,7 \leqslant t) = 0,95$ lorsque $t = 1,645$ (par lecture du formulaire), et donc $p(D < 4,44) = 0,95$.



- *Énoncé de la règle de décision*

  On prélève avec remise un échantillon aléatoire de taille $n_A = 30$ de la population A et on calcule sa moyenne $\overline{x}_A$; on fait de même pour la population B avec $n_B = 50$. On pose $d = \overline{x}_B - \overline{x}_A$.

  si $d \leqslant 4,44$ on accepte H_0.

  si $d > 4,44$ on rejette H_0 et on accepte H_1.

9.3 - Utilisation du test unilatéral

- *Calcul de d*

  On a $d = \overline{x}_B - \overline{x}_A = 779,6 - 774,7 = 4,9$
- *Application de la règle de décision*

  Comme $4,9 > 4,44$, on rejette H_0 et on accepte H_1.

Au seuil de 5%, la moyenne des masses des pièces livrées par le fournisseur B est significativement supérieure à celle du fournisseur A.

Exercices

Exercice 1 : Introduction aux intervalles de confiance

1. Soit X une variable aléatoire suivant la loi normale centrée réduite $\mathcal{N}(0, 1)$.
   a) Déterminer le réel a tel que $p(-a \leqslant X \leqslant a) = 0,99$.
   b) Déterminer le réel a tel que $p(-a \leqslant X \leqslant a) = 0,95$.
   c) Déterminer le réel a tel que $p(-a \leqslant X \leqslant a) = 0,90$.
   d) Déterminer le réel a tel que $p(-a \leqslant X \leqslant a) = 0,70$.
2. Soit Y une variable aléatoire suivant la loi normale centrée réduite $\mathcal{N}(20, 2)$.
   a) Déterminer le réel a tel que $p(20 - a \leqslant Y \leqslant 20 + a) = 0,99$.
   b) Déterminer le réel a tel que $p(20 - a \leqslant Y \leqslant 20 + a) = 0,95$.
   c) Déterminer le réel a tel que $p(20 - a \leqslant Y \leqslant 20 + a) = 0,90$.
   d) Déterminer le réel a tel que $p(20 - a \leqslant Y \leqslant 20 + a) = 0,70$.

Exercice 2 : La machine à bouchons, bts mai**,** *1996*

Une machine fabrique plusieurs milliers de bouchons cylindriques par jour. On admet que la variable aléatoire X qui, à chaque bouchon, associe son diamètre exprimé en millimètres, suit la loi normale de moyenne $m = 22$ mm et d'écart-type $\sigma = 0,025$ mm.

Les bouchons sont acceptables si leur diamètre appartient à l'intervalle $[21,95; 22,05]$.

Les trois questions de cet exercice peuvent être traitées de manière indépendante.

1. Quelle est la probabilité qu'un bouchon pris au hasard dans la production soit acceptable ?
2. Dans cette question, on considère que la probabilité qu'un bouchon soit défectueux est $q = 0,05$.

   On prélève au hasard un échantillon de 80 bouchons (ce prélèvement est assimilé à un tirage de 80 bouchons avec remise). On nomme Y la variable aléatoire mesurant le nombre de bouchons défectueux d'un tel échantillon.

   a) Quelle est la loi suivie par la variable aléatoire Y ? Déterminer l'espérance mathématique de la variable Y.
   b) On approche Y par une variable aléatoire Y_1 qui suit une loi de Poisson $\mathcal{P}(\lambda)$. Quelle est la valeur du paramètre λ ?

   Calculer la probabilité que l'échantillon prélevé contienne exactement 10 bouchons défectueux.
3. En vue du contrôle de réglage de la machine, on prélève régulièrement dans la production des échantillons de 100 bouchons.

   On appelle $\overline{X}$ la variable aléatoire qui, à chaque échantillon de 100 bouchons, associe le diamètre moyen des bouchons de cet échantillon.

   Lorsque la machine est bien réglée, $\overline{X}$ suit la loi normale de paramètres m et $\sigma' = \sigma/10$ (on rappelle que $m = 22$ et $\sigma = 0,025$).

   a) Déterminer le réel a tel que $P(22 - a \leqslant \overline{X} \leqslant 22 + a) = 0,95$.
   b) Sur un échantillon de 100 bouchons, on a les résultats suivants (les mesures des diamètres étant réparties en classe d'amplitude 0,02 mm) :

Classes de diamètres	effectif correspondant
$[21,93; 21,95[$	3
$[21,95; 21,97[$	7
$[21,97; 21,99[$	27
$[21,99; 22,01[$	30
$[22,01; 22,03[$	24
$[22,03; 22,05[$	7
$[22,05; 22,07[$	2

   En supposant que tous les bouchons d'une classe ont pour diamètre la valeur centrale de cette classe, donner la moyenne et l'écart-type de cette série (aucune justification demandée; résultats arrondis à l'ordre 10^{-4}).

En utilisant la question précédente, peut-on accepter au seuil de risque 5%, l'hypothèse selon laquelle la machine est bien réglée ?

Exercice 3 : Des glaces ! bts mai, *1995*

Une fabrique de desserts glacés dispose d'une chaîne automatisée pour remplir et emballer des cônes de glace.

– Partie A –

Chaque cône est rempli avec de la glace à la vanille. On désigne par X la variable aléatoire qui, à chaque cône, associe la masse (exprimée en grammes) de glace qu'il contient. On suppose que X suit la loi normale de paramètres $m = 100$ et σ.

1. Dans cette question, $\sigma = 2\sqrt{2}$.

   On choisit au hasard un cône rempli de glace. Calculer, à 10^{-2} près, la probabilité que la masse qu'il contient soit comprise entre 95 g et 105 g.
2. Un cône est considéré comme « bon » lorsque la masse de glace qu'il contient appartient à l'intervalle [95 ; 105]. Déterminer la valeur du paramètre σ telle que la probabilité de l'événement « *le cône est bon* » soit égale à 0, 95 (on donnera le résultat avec deux décimales).

– Partie B –

Les cônes de glace sont emballés individuellement puis conditionnés en lots de 2 000 pour la vente en gros.

On considère que la probabilité qu'un cône présente un défaut quelconque avant son conditionnement en gros est égale à 0, 000 5.

On nomme Z la variable aléatoire qui, à chaque lot de 2 000 cônes prélevés au hasard dans la production, associe le nombre de cônes défectueux présents dans le lot.

1. Quelle est la loi suivie par Z ?
2. On admet que la loi de Z peut être approchée par une loi de Poisson.
   a) Déterminer le paramètre de cette loi.
   b) Si un client reçoit un lot contenant au moins 5 cônes défectueux, l'entreprise procède alors à un échange de ce lot.

   Calculer la probabilité qu'un lot soit inchangé.

Exercice 4 : Des tiges métalliques. . ., *bts mai, session 2004*

Les trois parties de cet exercice sont indépendantes

Une entreprise fabrique, en grande quantité, des tiges métalliques cylindriques pour l'industrie. Leur longueur et leur diamètre sont exprimés en millimètres.

Dans cet exercice, les résultats approchés sont à arrondir à 10^{-2}

A - Loi normale.

Une tige de ce type est considérée comme conforme pour la longueur lorsque celle-ci appartient à l'intervalle [99, 45; 100, 55].

On note X la variable aléatoire qui, à chaque tige prélevée au hasard dans la production, associe sa longueur.

On suppose que X suit une loi normale de moyenne 100 et d'écart-type 0, 25.

1. Calculer la probabilité qu'une tige prélevée au hasard dans la production soit conforme pour la longueur.
2. Déterminer le nombre réel h positif tel que :

$$p(100 - h \leqslant X \leqslant 100 + h) = 0,95.$$

   Interpréter le résultat à l'aide d'une phrase.

B - Loi binomiale et loi de Poisson.

Dans un lot de ce type de tiges, 3% des tiges ne sont pas conformes pour la longueur. On prélève au hasard 50 tiges de ce lot pour vérification de la longueur. Le lot est suffisamment important pour que l'on puisse assimiler ce prélèvement à un tirage avec remise de 50 tiges.

On considère la variable aléatoire Y qui, à tout prélèvement de 50 tiges, associe le nombre de tiges non conformes pour la longueur.

1. Justifier que la variable aléatoire Y suit une loi binomiale dont on déterminera les paramètres.

2. Calculer la probabilité que, dans un tel prélèvement, au plus deux tiges ne soient pas conformes pour la longueur.

3. On considère que la loi suivie par Y peut-être approchée par une loi de Poisson. Déterminer le paramètre λ de cette loi de Poisson.

4. On désigne par Z une variable aléatoire suivant la loi de Poisson de paramètre λ où λ a la valeur obtenue au **4.**

Calculer $p(Z = 2)$ et $p(Z \leqslant 2)$.

C - Intervalle de confiance.

Dans cette question, on s'intéresse au diamètre des tiges, exprimé en millimètres.

On prélève au hasard et avec remise un échantillon de 50 tiges dans la production d'un journée.

Soit $\overline{D}$ la variable aléatoire qui, à tout échantillon de 50 tiges prélevées au hasard et avec remise dans la production d'un journée, associe la moyenne des diamètres des tiges de cet échantillon.

On suppose que $\overline{D}$ suit une loi normale de moyenne inconnue μ et d'écart-type $\sigma/\sqrt{50}$ avec $\sigma = 0,19$.

Pour l'échantillon prélevé, la moyenne obtenue, arrondie à 10^{-2}, est $\overline{x} = 9,99$.

1. À partir des informations portant sur cet échantillon, donner une estimation ponctuelle de la moyenne μ des tiges produites dans cette journée.

2. Déterminer un intervalle de confiance centré sur $\overline{x}$ de la moyenne μ des diamètres des tiges produites pendant la journée considérée, avec le coefficient de confiance de 95%.

3. On considère l'affirmation suivante : « *la moyenne μ est obligatoirement dans l'intervalle de confiance obtenue à la question* **2.** ».

Est-elle vraie ? (on ne demande pas de justification.)

Exercice 5 : Des pièces en série. . . , *bts mai, session 1997*

Une entreprise fabrique en série des pièces dont le diamètre, mesuré en millimètres, définit une variable aléatoire D.

On admet que cette variable aléatoire D suit la loi normale de moyenne m et d'écart type σ.

1. *Estimation de m et σ :*

a) Un échantillon de 100 pièces est prélevé au hasard dans la production. Les mesures des diamètres des pièces de cet échantillon son regroupées dans le tableau suivant :

Mesures des diamètres (en mm)	$[4,0;4,2[$	$[4,2;4,4[$	$[4,4;4,6[$	$[4,6;4,8[$	$[4,8;5,0[$
effectif	6	24	41	25	4

En faisant l'hypothèse que, pour chaque classe, les valeurs mesurées sont égales à celle du centre de la classe, calculer, à 10^{-2} près, la moyenne d et l'écart type s de cet échantillon.

En déduire l'estimation ponctuelle de σ fournie par cet échantillon.

b) On appelle $\overline{D}$ la variable aléatoire qui, à chaque échantillon de 100 pièces, associe la moyenne des diamètres des pièces de l'échantillon.

On rappelle que $\overline{D}$ suit la loi normale de moyenne m et d'écart type $\sigma/10$.

Déterminer un intervalle de confiance de la moyenne m de D au seuil de confiance de 95%.

2. Dans cette question, on admet que la production comporte 5 % de pièces inutilisables.

a) L'entreprise conditionne ses pièces par boîtes de 25.

On tire une boîte au hasard (on assimilera cette épreuve à un tirage successif avec remise de 25 pièces dans la production).

On désigne par K le nombre de pièces inutilisables dans cette boîte.

Quelle est la loi suivie par la variable aléatoire K ?

Calculer, à 10^{-3} près, la probabilité que cette boîte contienne au plus une pièce inutilisable.

b) Un client qui a besoin de 185 pièces commande 8 boîtes de pièces (on assimile cette épreuve à un tirage successif et avec remise de 200 pièces dans la production).

On désigne par L le nombre de pièces inutilisables dans cette commande. On admet que L suit la loi de Poisson de paramètre 10.

Quelle est la probabilité que le client dispose d'un échantillon suffisant de pièces utilisables dans sa commande ?

Exercice 6 : Production de rondelles. . . *bts mai, session 2005*

Les quatre parties de cet exercice peuvent être traitées de façon indépendante.

Une usine fabrique, en grande quantité, des rondelles d'acier pour la construction. Leur diamètre est exprimé en millimètres.

Dans cet exercice, sauf indication contraire, les résultats approchés sont à arrondir à 10^{-2}**.**

– Partie A – loi normale –

Une rondelle de ce modèle est conforme pour le diamètre lorsque celui-ci appartient à l'intervalle $[89,6;90,4]$.

1. On note X_1 la variable aléatoire qui, à chaque rondelle prélevée au hasard dans la production, associe son diamètre; On supppose que X_1 suit la loi normale de moyenne 90 et d'écart type $\sigma = 0,17$.

Calculer la probabilité qu'une rondelle prélevée au hasard dans la production soit conforme.

2. L'entreprise désire améliorer la qualité de la production des rondelles : il est envisagé de modifier le réglage des machines produisant les rondelles.

On note D la variable aléatoire qui, à chaque rondelle prélevée dans la production future, associera son diamètre. On suppose que la variable aléatoire D suit une loi normale de moyenne 90 et d'écart type σ_1.

Déterminer σ_1 pour que la probabilité qu'une rondelle prélevée au hasard dans la production future soit conforme pour le diamètre soit égale à $0,99$.

– Partie B – Loi binomiale –

On note E l'événement : *une rondelle prélevée au hasard dans un stock important a un diamètre défectueux.*

On supppose que $p(E) = 0,02$.

On prélève au hasard quatre rondelles dans le stock pour vérification de leur diamètre. Le stock est assez important pour que l'on puisse assimiler ce prélèvement à un tirage avec remise de quatre rondelles.

On considère la variable aléatoire Y_1 qui à tout prélèvement de quatre rondelles associe le nombre de rondelles de ce prélèvement ayant un diamètre défectueux.

1. Justifier que la variable aléatoire Y_1 suit une loi binomiale dont on déterminera les paramètres.

2. Calculer la probabilité que, dans un tel prélèvement, aucune rondelle n'ait un diamètre défectueux. Arrondir à 10^{-3}.

3. Calculer la probabilité que, dans un tel prélèvement, au plus une rondelle ait un diamètre défectueux. Arrondir à 10^{-3}.

– Partie C – Approximation d'une loi binomiale par une loi normale –

Les rondelles sont commercialisées par lot de 1 000.

On prélève au hasard un lot de 1 000 dans un dépôt de l'usine. On assimile ce prélèvement à un tirage avec remise de 1 000 rondelles.

On considère la variable aléatoire Y_2 qui, à tout prélèvement de 1 000 rondelles, associe le nombre de rondelles non conformes parmi ces 1 000 rondelles.

On admet que la variable aléatoire Y_2 suit la loi binomiale de paramètres $n = 1\,000$ et $p = 0,02$.

On décide d'approcher la loi de la variable aléatoire Y_2 par la loi normale de moyenne 20 et d'écart type $4,43$.

On note Z une variable aléatoire suivant la loi normale de moyenne 20 et d'écart type $4,43$.

1. Justifier les paramètres de cette loi normale.

2. Calculer la probabilité qu'il y ait au plus 15 rondelles non conformes dans le lot de 1 000 rondelles, c'est à dire calculer $p(Z \leqslant 15,5)$.

– Partie D – Test d'hypothèse –

On se propose de construire un test d'hypothèse pour contrôler la moyenne μ de l'ensemble des diamètres, en millimètres, de rondelles constituant une livraison à effectuer.

On note X_2 la variable aléatoire qui, à chaque rondelle prélevée au hasard dans la livraison, associe son diamètre.

La variable aléatoire X_2 suit la loi normale de moyenne inconnue μ et d'écart type $\sigma = 0,17$.

On désigne par $\overline{X_2}$ la variable aléatoire qui, à chaque échantillon aléatoire de 100 rondelles prélevé dans la livraison, associe la moyenne des diamètres de ces rondelles (la livraison est assez importatne pour que l'on puisse assimiler ces prélèvements à des tirages avec remise).

L'hypothèse nulle est H_0 : $\mu = 90$. Dans ce cas la livraison est dite conforme pour le diamètre.

L'hypothèse alternative est H_1 : $\mu \neq 90$.

Le seuil de signification du test est fixé à 0, 05.

1. Énoncer la règle de décision permettant d'utiliser ce test en admettant, sous l'hypothèse nulle H_0, le résultat suivant qui n'a pas à être démontré :

$$p(89,967 \leqslant \overline{X_2} \leqslant 90,033) = 0,95.$$

2. On prélève un échantillon de 100 rondelles dans la livraison et on observe que, pour cet échantillon, la moyenne des diamètres est $\overline{x} = 90,02$.

   Peut-on, au seuil de 5%, conclure que la livraison est conforme pour le diamètre ?

Exercice 7 : Variation du coefficient de confiance

Dans tout ce qui suit, les résultats seront donnés à 10^{-2} près.

Les ampoules électriques d'un certain modèle ont une durée de vie exprimée en heures, dont la distribution est normale d'écart-type $\sigma = 200$ heures. Avec cette hypothèse, on se propose d'estimer la moyenne m de la durée de vie des ampoules de la production à partir d'un échantillon de 36 ampoules dont la moyenne des durées de vie est égale à 3 000 heures.

On assimile cet échantillon à un échantillon prélevé au hasard et avec remise parmi la production.

1. Déterminer une estimation de m par un intervalle de confiance avec le coefficient de confiance 95%.
2. Même question avec le coefficient de confiance 90% puis avec le coefficient de confiance 99%.
3. Qu'observe-t-on sur les intervalles de confiance de la moyenne m de la population obtenus à partir d'un **même** échantillon lorsque le coefficient de confiance varie ?
4. Peut-on situer exactement la position de la moyenne m par rapport à l'intervalle obtenu à la question **1.** ?

Exercice 8 : On roule pour vous

Une entreprise utilise des camions pour transporter sa production. Elle dispose de 100 camions.

Elle repère sur un échantillon de 30 jours choisis au hasard, le nombre de camions en panne. Voici les résultats :

$$5, 6, 4, 6, 6, 8, 3, 5, 5, 5, 4, 3, 6, 5, 6, 4, 7, 6, 6, 5, 4, 3, 6, 5, 4, 5, 4, 5, 5, 4.$$

1. Calculer la moyenne $\overline{x}$ et l'écart-type σ du nombre de camions en panne chaque jour pour l'échantillon étudié.
2. À partir des résultats obtenus pour cet écantillon, proposer une estimation ponctuelle de la moyenne μ et de l'écart-type s du nombre de camions en panne chaque jour pour la population correspondant aux jours ouvrables de l'année.
3. On suppose que la variable aléatoire $\overline{X}$ qui, à tout échantillon de taille 30 prélevé au hasard et avec remise, associe la moyenne du nombre de camions en panne chaque jour, suit la loi normale $\mathcal{N}(\mu, s/\sqrt{n})$. On prend pour valeur l'estimation ponctuelle obtenue au **2.**.

   Déterminer un intervalle de confiance de la moyenne μ de la population avec le coefficient de confiance 95%.

Exercice 9 : Recherche de l'effectif d'un échantillon

Sur une portion de route où la vitesse des véhicules est limitée à 90 km/h, on effectue un contrôle des vitesses avecun instrument de mesure de grande précision.

On mesure la vitesse (en km/h) d'un véhicule sur vingt et on obtient les résultats suivants pour un échantillon de 100 véhicules que l'on assimile à un échantillon obtenu par prélèvement aléatoire avec remise :

Vitesse (en km/h)	Effectif
[75, 80 [	5
[80, 85 [	10
[85, 90 [	20
[90, 95 [	36
[95, 100 [	15
[100, 105 [	8
[105, 110 [	6

1. **1.** En supposant que les valeurs observées sont celles du centre de la classe, calculer, à 10^{-2} près, la moyenne $\overline{x}$ et l'écart-type σ des vitesses pour cet échantillon.
2. **2.** À partir des résultats obtenus pour cet échantillon, proposer une estimation ponctuelle de la moyenne μ et de l'écart-type s des vitesses des 2 000 véhicules de la population observée.
3. **3.** On suppose que la variable aléatoire $\overline{X}$ qui, à tout échantillon de taille $n = 100$ obtenu comme précédemment, associe la moyenne des vitesses de l'échantillon, suit la loi normale $\mathcal{N}(\mu, \sigma/\sqrt{n})$. On prend pour valeur de s l'estimation ponctuelle obtenue au **2.**.

   Déterminer un intervalle de confiance de la vitesse moyenne μ de la population avec le coefficient de confiance 99%.
4. **4.** Quelle doît être la taille minimale n de l'échantillon pour connaître, avec le coefficient de confiance 95% la vitesse moyenne de la population à $0,5\ \mathrm{km}/\mathrm{h}$ près ?

Exercice 10 : Longueur des pièces (test bilatéral)

Dans un atelier, une machine fabrique des pièces en grande série ; on s'intéresse à leur longueur mesurée en cm. On admet que la variable aléatoire X qui, à chaque pièce tirée au hasard dansla production associe sa longueur, suit une loi normale de moyenne m et d'écart-type $\sigma = 0,14$.

Afin de contrôler le fait que la moyenne m des longueurs des pièces produites est 150, on se propose de construire un test d'hypothèse.

On prélève des échantillons aléatoires de 49 pièces (chaque échantillon étant obtenu par tirage avec remise).

À chaque échantillon ainsi défini, on associe la moyenne $\overline{x}$ des longueurs des 49 pièces ; on définit ainsi une variable aléatoire $\overline{X}$.

L'hypothèse nulle est H_0 : $m = 150$; l'hypothèse alternative est H_1 : $m \neq 150$.

Le suil de signification du test est fixé à 0, 05.

1. **1.** Quelle est, sous l'hypothèse nulle H_0, la loi de la variable aléatoire $\overline{X}$?
2. **2.** Déterminer le nombre réel positif h tel que $p(150 - h \leqslant \overline{X} \leqslant 150 + h) = 0,95$.
3. **3.** Énoncer la règle de décision permettant d'utiliser ce test.
4. **4.** La moyenne observée sur un échantillon de 49 pièces est $\overline{x} = 149,9$. Que peut-on en conclure au seuil de signification 5% quant à la qualité des pièces produites ?

Exercice 11 : Il était un petit navire

Les tôles constituant les ponts d'un paquebot subissent des déformations lors des opérations d'assemblage par soudure. Les tôles doivent être redressées ; cette opération nécessite de nombreuses heures de travail.

1. **1.** Lors d'une construction, on relève les durées nécessaires au redressage d'un échantillon représentatif de 50 tôles. Les résultats obtenus sont résumés dans le tableau suivant :

x_i	[0, 10[	[10, 20[	[20, 30[	[30, 40[	[40, 50[	[50, 60[	[60, 70[
n_i	1	4	8	20	12	3	2

   où x_i : durée en heures et n_i : nombre de tôles.

   Calculer la moyenne $\overline{x}$ et l'écart-type s de ce échantillon.
2. **2.** À l'occasion d'une nouvelle construction, on cherche une estimation de la durée moyenne du redressage d'une tôle prise dans une population de 1 700 tôles.

   Soit X la variable aléatoire qui, à chaque tôle à redresser prise au hasard, associe la durée du redressage. On admet que X suit la loi normale $\mathcal{N}(m, \sigma)$, où m est la durée moyenne (inconnue) du redressage par tôle pour la population et σ l'écart-type de la variable aléatoire. On prend pour valeur de σ l'estimation ponctuelle obtenue à partir de l'échantillon étudié au **1**.

   a) Déterminer un intervalle de confiance contenant, avec 95% de certitude, la moyenne m, en supposant que la variable aléatoire qui, à tout échantillon de 50 tôles prélevées au hasard, associe la durée moyenne du redressage pour une tôle de cet échantillon, suit la loi normale $\mathcal{N}(m, \sigma/\sqrt{n})$ où $n = 50$.

   b) Quelle devrait être la taille de l'échantillon à étudier pour que l'amplitude de l'intervalle de confiance de la moyenne soit 3 heures avec une probabilité de 0, 95 ?

3. Construire un test permettant d'accepter ou de refuser, au seuil de 5%, l'hypothèse selon laquelle la durée nécessaire au redressage d'une tôle est en moyenne de 40 heures.

Utiliser ce test avec l'échantillon de l'énoncé.

4. Même question avec le seuil de signification de 1%.

Exercice 12 : Statistiques inférentielles : un problème de synthèse. *Bts Maintenance industrielle, 1996*

Les parties I et II peuvent être traitées indépendamment l'une de l'autre.

Un groupe industriel possède deux filiales MAT et MATIC qui produisent des petits moteurs destinés au montage de jouets.

– Partie I –

La variable aléatoire X qui, à chaque moteur tiré au hasard dans la production, associe sa durée de vie moyenne exprimée en heures, suit la loi normale de moyenne 400 et d'écart type 40.

1. Un moteur est déclaré non commercialisable si sa durée de vie est inférieure à 318 heures. Calculer, à 10^{-4} près la probabilité p qu'un moteur prélevé au hasard dans la production ne soit pas commercialisable.

2. On admet que $p = 0,02$. Soit Y la variable aléatoire qui, à tout lot de 50 moteurs, associe le nombre de moteurs non commercialisables. La production est assez importante pour que l'on puisse assimiler le prélèvement de 50 moteurs à un prélèvement aléatoire avec remise.

a) Quelle est la loi suivie par Y ? Justifier la réponse et donner ses paramètres.

b) Calculer à 10^{-3} près la probabilité de l'événement : « *il y a au plus trois moteurs non commercialisables* ».

– Partie II –

La filiale MAT prélève un échantillon de taille 100 sur la production d'un jour et mesure la durée de vie, en heures, des moteurs. Les résultats obtenus sont les suivants :

durée de vie	$[300,340[$	$[340,380[$	$[380,420[$	$[420,460[$	$[460,500[$
Effectifs	7	21	48	16	8

1. En faisant l'hypothèse que les valeurs mesurées sont celles du centre de classe, calculer, à 10^{-2} près, la moyenne m_1 et l'écart type σ_1 de cette série statistique.

La filiale MATIC, dans des conditions similaires, contrôle un échantillon de taille 100 et obtient pour résultats $m_2 = 406,8$ et $\sigma_2 = 40,5$.

2. On désigne par $\overline{X_1}$ la variable aléatoire qui, à chaque échantillon de 100 moteurs prélevés au hasard par la filiale MAT, associe sa moyenne, et par $\overline{X_2}$ la variable aléatoire qui, à chaque échantillon de 100 moteurs prélevés au hasard par la filiale MATIC, associe sa moyenne.

Tous les échantillons considérés sont assimilés à des échantillons prélevés avec remise.

On suppose que les variables aléatoires $\overline{X_1}$, $\overline{X_2}$, $D = \overline{X_1} - \overline{X_2}$ suivent des lois normales de moyennes respectives M_1, M_2, $M_1 - M_2$ inconnues, et on estime l'écart type de D par

$$\sigma_D = \sqrt{\frac{\sigma_1^2 + \sigma_2^2}{100}}.$$

(On prend comme valeur approchée à 10^{-1} près de σ_1 la valeur 39, 4.)

On décide de construire un test bilatéral permettant de savoir s'il existe une différence significative au seuil de 5% entre les durées de vie des moteurs fabriqués par les filiales MAT et MATIC.

On choisit pour hypothèse $H_0 : M_1 = M_2$, et pour hypothèse alternative $H_1 : M_1 \neq M_2$.

a) Sous l'hypothèse H_0, D suit la loi normale $\mathcal{N}(0, \sigma_D)$. Déterminer l'intervalle $[-h, h]$ tel que $P(-h \leqslant D \leqslant h) = 0,95$.

b) Énoncer la règle de décision du test.

c) Utiliser ce test avec les deux échantillons de l'énoncé et conclure.

Exercice 13 : Une usine de plaquettes, bts mai, *1998*

Dans tout l'exercice, on arrondira les résultats à 10^{-2} près.

Une entreprise fabrique des plaquettes dont la longueur et la largeur sont mesurées en mm.

– Partie A –

Sur un échantillon de 100 plaquettes, on a mesuré la longueur de chaque plaquette et obtenu le tableau suivant :

Longueur	$[35, 37[$	$[37, 39[$	$[39, 41[$	$[41, 43[$	$[43, 45[$
effectif	3	25	50	20	2

1. On veut calculer une valeur approchée de la moyenne m et de l'écart type s de l'échantillon. Pour cela, on fait comme si toutes les observations d'une classe étaient situées au centre de la classe. Calculer m et s. Compte tenu de l'erreur de méthode induite par l'approximation précédente, les résultats seront donnés à 10^{-1} près.

2. On suppose que la variable aléatoire L qui à chaque plaquette associe sa longueur suit une loi normale de moyenne μ et d'écart type $1,6$.

a) Donner une estimation ponctuelle de μ.

b) Déterminer un intervalle de confiance à 95% de μ centré sur la valeur obtenue précédemment.

– Partie B –

On suppose dans cette partie que L suit une loi normale de moyenne 40 et d'écart type $1,6$ et que la largeur ℓ suit une loi normale de moyenne 25 et d'écart type $1,2$.

1. On tire une plaquette au hasard dans la production.

a) Quelle est la probabilité d'obtenir une longueur comprise entre 37 et 43 mm ?

b) Quelle est la probabilité d'obtenir une largeur comprise entre 22 et 28 mm ?

2. Une plaquette est acceptée si sa longueur est comprise entre 37 et 43 mm et sa largeur est comprise entre 22 et 28 mm.

En admettant que L et ℓ sont des variables aléatoires indépendantes, quelle est la probabilité d'obtenir une plaquette qui soit acceptée ?

– Partie C –

La probabilité d'obtenir une plaquette qui soit rejetée est égale à $0,07$.

On appelle X la variable aléatoire qui à un lot de 100 plaquettes extraites de la fabrication associe le nombre de plaquettes rejetées contenues dans ce lot.

1. Quelle est la loi de probabilité suivie par X ? Préciser ses paramètres et son espérance mathématique.

2. En admettant que la loi de X peut être approchée par une loi de Poisson, préciser son paramètre.

Quelle est alors la probabilité d'obtenir strictement moins de 10 plaquettes rejetées dans un lot de 100 plaquettes ?

Exercice 14 : Des paquets de farine, *bts MAI, 1993*

Une machine est chargée de conditionner des paquets de farine. La masse M d'un paquet est une variable aléatoire qui suit une loi normale d'écart-type constant $\sigma = 30$, et dont la moyenne m peut-être modifiée. Un paquet est refusé si sa masse est inférieure à 995 grammes.

1. On suppose que la moyenne m est égale à 1 000.

a) Quelle est la probabilité pour qu'un paquet soit refusé ?

b) On suppose que la probabilité pour qu'un paquet soit refusé est $p = 0,07$. On dispose de 100 paquets. Soit X la variable aléatoire dénombrant le nombre de paquets à rejeter.

Quelle est la loi de X ? Calculer $p(X = 3)$.

On assimile la loi de X à une loi de Poisson. Indiquer le paramètre de cette loi de Poisson et déterminer la valeur indiquée pour $p(X \leqslant 5)$.

2. Afin de diminiuer le nombre de paquets refusés, on décide de modifier le réglage de la machine.

a) Quelle doit être la valeur de m pour que la probabilité d'accepter un paquet soit égale à $0,99$?

b) La machine est réglée de telle sorte que $m = 1\,025$. Afin de vérifier ce réglage, on prélève un échantillon de 20 paquets et on déterminer la masse moyenne $\overline{x}$.

Déterminer l'intervalle centré en m contenant $\overline{x}$ avec une probabilité de $0,95$.

Exercice 15 : Assurances d'une flotte de véhicules

Les quatres questions de cet exercice sont indépendantes.

Dans un groupe d'assurances on s'intéresse aux sinistres suceptibles de survenir, une année donnée, aux véhicules de la flotte d'une importante entreprise de maintenance de chauffage collectif.

Dans cet exercice, sauf mention du contraire, les résultats sont à arrondir à 10^{-3} près.

1. *Étude du nombre de sinistres par véhicule*

Soit X la variable aléatoire qui, à tout véhicule tiré au hasard dans un des parcs de la flotte, associe le nombre de sinistres survenant pendant l'année considérée.

On admet que X suit la loi de Poisson de paramètre $0,28$.

a) Calculer la probabilité de l'événement A : « *un véhicule tiré au hasard dans le parc n'a aucun sinistre pendant l'année considérée* ».

b) Calculer la probabilité de l'événement B : « *un véhicule tiré au hasard dans le parc a, au plus, deux sinistres pendant l'année considérée* ».

2. *Étude du nombre de sinistres dans une équipe de 15 conducteurs*

On note E l'événement : « *un conducteur tiré au hasard dans l'ensemble des conducteurs de l'entreprise n'a pas de sinistre pendant l'année considérée* ».

On suppose que la probabilité de l'événement E est $0,6$.

On tire au hasard 15 conducteurs dans l'effectif des conducteurs de l'entreprise. Cet effectif est assez important pour que l'on puisse assimiler ce prélèvement à untirage avec remise de 15 conducteurs.

On coinsidère la variable aléatoire Y qui, à tout prélèvement de 15 conducteurs, associe le nombre de conducteurs n'ayant pas de sinistre pendant l'année considérée.

a) Justifier que la variable aléatoire Y suit une loi binomiale et déterminer ses paramètres.

b) Calculer la probabilité que, dans un tel prélèvement, 10 conducteurs n'aient pas de sinistre pendant l'année considérée.

3. *Étude du coût des sinistres*

Dans ce qui suit, on s'intéresse au coût d'une certaine catégorie de sinistres survenus dans l'entreprise pendant l'année considérée.

On considère la variable aléatoire C qui, à chaque sinistre tiré au hasard parmi les sinistres de cette catégorie, associe son coût en euros.

On suppose que C suit la loi normale de moyenne 1 200 et d'écart type 200.

Calculer la probabilité qu'un sinistre tiré au hasard parmi les sinistres de ce type coûte entre 1 000 euros et 1 500 euros.

4. La question ci-dessous doit être traitée par les candidats de toutes les spécialités de BTS du groupement B, à l'exception du BTS Maintenance et exploitation des matériels aéronautiques.

On considère un échantillon de 100 véhicules prélevés au hasard dans le parc de véhicules mis en service depuis 6 mois. Ce parc contient suffisamment de véhicules pour qu'on puisse assimiler ce tirage à un tirage avec remise.

On constate que 91 véhicules de cet échantillon n'ont pas eu de sinistre.

a) Donner une estimation ponctuelle du pourcentage p de véhicules de ce parc qui n'ont pas eu de sinistre 6 mois après leur mise en service.

b) Soit F la variable aléatoire qui à tout échantillon de 100 véhicules prélevés au hasard et avec remise dans ce parc, associe le pourcentage de véhicules qui n'ont pas eu de sinistre 6 mois après leur mise en service.

On suppose que F suit la loi normale

$$\mathcal{N}\left(p, \sqrt{\frac{p(1-p)}{100}}\right)$$

où p est le pourcentage inconnu de véhicules du parc qui n'ont pas eu de sinistre 6 mois après leur mise en service.

Déterminer un intervalle de confiance du pourcentage p avec le coefficient de confiance 95%.

c) On considère l'affirmation suivante : « *le pourcentage p est obligatoirement dans l'intervalle de confiance obtenu à la question b*) ». Est-elle vraie ? (On ne demande pas de justification.)

4. La question ci-dessous doit être traitée uniquement par les candidats au BTS Maintenance et exploitation des matériels aéronautiques.

Pour un parc de véhicules, on a relevé le nombre de sinistres par véhicule pendant la première année de mise en service.

Pour les véhicules ayant eu, au plus, quatre sinistres, on a obtenu :

Nombre de sinistres : x_i	0	1	2	3	4
Effectif : n_i	1 345	508	228	78	35

a) Compléter, après l'avoir reproduit, le tableau suivant :

Nombre de sinistres : x_i	0	1	2	3	4
$y_i = \ln n_i$					

b) Déterminer, à l'aide d'une calculatrice, une équation de la droite de régression de y en x sous la forme

$$y = ax + b$$

où a et b sont à arrondir à 10^{-2}.

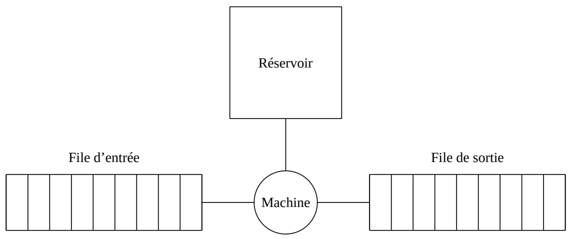
c) À l'aide de l'équation précédente, estimer le nombre de véhicules ayant eu six sinistres pendant leur première année de mise en circulation.

Exercice 16 : Une chaîne d'embouteillage, *bts mai, mai 2003*

Les quatres questions de cet exercice sont indépendantes

Dans une usine du secteur de l'agroalimentaire, une machine à embouteiller est alimentée par un réservoir d'eau et par une file d'approvisionnement en bouteilles vides, selon le schéma ci-contre.

L'exercice consiste en une étude statistique du bon fonctionnement de ce système.



1. *Défaut d'approvisionnement*

On considère qu'il y a défaut d'approvisionnement :

– soit lorsque la file d'entrée des bouteilles est vide,
– soit lorsque le réservoir est vide.

On tire au hasard un jour ouvrable dans une année. On note A l'événement : « *la file d'attente est vide au moins une fois dans la journée* » et B l'événement : « *le réservoir est vide au moins une fois dans la journée* ».

On suppose que les événements A et B sont indépendants et une étude statistique a montré que

$$p(A) = 0,04 \qquad \text{et} \qquad p(B) = 0,02.$$

Calculer la probabilité des événements suivants :

a) $E_1 = A \cap B$.

b) E_2 : « *la machine a connu au moins un défaut d'approvisionnement dans la journée* ».

2. *Pannes de la machine sur une durée de 100 jours*

On note X la variable aléatoire qui à toute période de 100 jours consécutifs, tirée au hasard dans les jours ouvrables d'une année, associe le nombre de pannes de la machine. Une étude, menée par le constructeur sur un grand nombre de machines de ce type, permet d'admettre que X suit la loi de Poisson de paramètre $\lambda = 0,5$.

Déterminer, à l'aide de la table du formulaire :

a) $p(X \leqslant 2)$;

b) la probabilité de l'événement « *la machine a au plus quatre pannes pendant la période de 100 jours consécutifs* »;

c) le plus petit entier n tel que : $p(X \leqslant n) \geqslant 0,99$.

Dans tout ce qui suit, les volumes sont exprimés en litres et tous les résultats approchés sont à arrondir à $\mathbf{10^{-3}}$.

3. *Qualité de l'embouteillage à la sortie*

On désigne par Y la variable aléatoire qui, à toute bouteille prise au hasard dans la production d'une heure, associe le volume d'eau qu'elle contient. On admet que, lorsque la machine est bien réglée, Y suit la loi normale de moyenne $1,5$ et d'écart-type $0,01$.

Une bouteille d'eau est conforme aux normes de l'entreprise lorsqu'elle contient entre $1,47$ et $1,53$ litre d'eau.

Calculer la probabilité qu'une bouteille satisfasse à la norme.

4. *Fiabilité d'une machine à embouteiller*

On s'intéresse à une machine à embouteiller prélevée au hasard dans le parc des machines sur le point d'être livrées par le constructeur.

On désigne par T la variable aléatoire qui, à toute machine prélevée au hasard dans le parc, asssocie sa durée de vie avant une défaillance.

On note $p(T > t)$ la probabilité qu'une machine prélevée au hasard dans le parc n'ait pas de défaillance avant l'instant t, exprimé en jours.

On suppose que $p(T > t) = e^{-0,005t}$.

a) Calculer la probabilité qu'une machine prélevée au hasard dans le parc fonctionne plus de 200 jours sans panne.

b) Déterminer t pour que la probabilité qu'une machine prélevée au hasard dans le parc fonctionne plus de t jours, soit égale à $0,8$. Arrondir à l'entier par défaut.