



**MIASS 241**  
**Mathématiques (appliquées aux sciences sociales) 4**  
**© El Hadj Touré, 2022**

**DIX EXERCICES RÉCAPITULATIFS**  
**Leçon 7. Introduction à l'inférence statistique**  
**(Solutionnaire)**

**1) Comparez, en termes de différence et de similitude, les paires de concepts ci-dessous :**

a) Une statistique & un paramètre

Une statistique est une caractéristique d'un échantillon : d'où le caractère aléatoire de sa valeur souvent connue. Par contre, un paramètre renvoie à une caractéristique d'une population : d'où le caractère déterministe de sa valeur souvent inconnue.

Au demeurant, la statistique et le paramètre sont deux valeurs calculées sur des données, celle-là étant un estimé de celui-ci.

**NB : Une statistique est à l'échantillon ce que le paramètre est à la population 😊**

b) Loi des grands nombres & théorème central limite

La loi des grands nombres postule que les caractéristiques de l'échantillon aléatoire se rapprochent d'autant plus de celles de la population que la taille de l'échantillon augmente. Parallèlement, le théorème central limite postule que la distribution d'échantillonnage issue d'échantillons aléatoires de grande taille suit un modèle normal prédictible, sa moyenne étant égale au paramètre de la population.

La loi des grands nombres et le théorème central limite sont deux lois statistiques sur lesquelles les sondes se basent pour connaître la marge d'erreur, et donc estimer le paramètre d'une population à partir d'une statistique de l'échantillon.

c) Biais & consistance

Le biais réfère à la situation où la valeur d'une statistique s'écarte de celle du paramètre qu'elle est censée estimer, au fur et à mesure que l'échantillon devient non aléatoire. Par contre, la consistance réfère à la faible variabilité de la distribution d'échantillonnage d'une statistique, laquelle se rapproche ainsi de la valeur du paramètre, au fur et à mesure que la taille de l'échantillon augmente.

L'absence de biais et la consistance permettent de rendre compte de l'efficacité de l'estimation du paramètre de la population à l'aide d'une statistique de l'échantillon.

**NB : Pour réduire le biais lié à un estimé, il faut que l'échantillon soit aléatoire. Pour**

augmenter la consistance d'un estimé, il faut augmenter la taille de l'échantillon ☺

d) Erreur-type & marge d'erreur

L'erreur-type renvoie à l'écart-type des statistiques de tous échantillons de taille  $n$  qu'il est possible de former dans une population  $N$ . Par contre, la marge d'erreur, pour un niveau de confiance donné (95% ou 99% en sciences sociales en général), réfère à l'écart entre une statistique d'un échantillon et le paramètre d'une population.

L'erreur-type est une forme de marge d'erreur, mais avec un niveau de confiance de 68%. Remarquez que la marge d'erreur est égale à  $1,96(z)$  fois l'erreur-type ( $E=1,96*\sigma_x$ ) à 95%. Eh bien, l'erreur-type est égale à  $1(z)$  fois l'erreur-type à 68%.

**NB : L'erreur-type s'apparente à une marge d'erreur. La marge d'erreur donne la même information que l'intervalle de confiance. L'intervalle de confiance, c'est la marge d'erreur en plus ou moins d'une statistique calculée ☺**

**2) Un sociologue étudie les inégalités de revenu (des parents) selon le programme d'études chez les étudiants d'une université. Il souhaite constituer un échantillon selon les 30 programmes que compte cette université. Pour ce faire, il compare trois modes d'échantillonnage : l'échantillonnage stratifié, l'échantillonnage par grappes et l'échantillonnage par quotas.**

a) Comment doit-il s'y prendre pour constituer un échantillon stratifié?

Pour constituer un échantillon stratifié, le sociologue doit recenser tous les 30 programmes comme étant des strates. *Au sein de chaque strate, il sélectionne aléatoirement le nombre d'étudiants voulu, sachant qu'il dispose d'une base de sondage fiable.* La sélection aléatoire peut se faire de façon simple ou systématique. Par exemple, le sociologue peut sélectionner au hasard 10 étudiants par programme à l'aide d'un tableur comme Excel ou d'un logiciel échantillonneur.

b) Comment doit-il s'y prendre pour constituer un échantillon par grappes?

*Afin de former un échantillon par grappes, le sociologue doit choisir au hasard parmi les 30 programmes considérés comme des grappes.* Par exemple, il peut décider de retenir aléatoirement 2 programmes seulement et :

- soit sonder tous les étudiants appartenant à ces 2 programmes (il n'aurait pas ainsi besoin d'une base de sondage),
- soit sonder les étudiants choisis aléatoirement au sein de ces 2 programmes, sachant qu'il dispose d'une base de sondage fiable. On obtient ainsi un échantillon à plusieurs niveaux.

c) Comment doit-il s'y prendre pour constituer un échantillon par quotas?

*Pour constituer un échantillon par quotas, le sociologue doit reproduire de façon délibérée les caractéristiques de la population dans l'échantillon selon les mêmes quotas.* Par exemple, si le programme 1 compte 10% de l'effectif total d'étudiants de l'Université, il comptera 10% dans l'échantillon, ainsi de suite. Il ne lui restera plus qu'à sélectionner les étudiants, selon des modalités définies par lui-même, jusqu'à atteindre les quotas retenus pour chacun des 30 programmes. La sélection peut se faire de façon accidentelle, volontaire, typique ou par boule de neige.

**NB : Les catégories d'une même variable (programmes d'études) peuvent**

**parfois être considérées comme des strates ou des grappes ou des quotas. Dans ce cas, seule la procédure d'échantillonnage permet de les distinguer ☺**

- d) Quel est le mode d'échantillonnage le moins approprié s'il désire procéder à l'inférence statistique? Justifiez votre réponse.

Le mode d'échantillonnage le moins approprié en inférence statistique est l'échantillonnage par quotas, lequel produit un échantillon souvent non représentatif de la population. En fait, strictement parlant, l'échantillon par quotas est un **échantillon de convenance**, tous les étudiants n'ayant pas une même chance non nulle et indépendante d'être sélectionnés. C'est le contraire des modes d'échantillonnage stratifié et par grappes qui relèvent théoriquement de l'échantillonnage probabiliste ou aléatoire. Concernant l'échantillonnage stratifié, tous les étudiants ont la même chance de faire partie de façon indépendante de l'échantillon au sein de chacun des 30 programmes. Dans le cas de l'échantillonnage par grappes, chaque programme a une chance égale d'être sélectionné.

Habituellement, l'utilisation de l'inférence statistique n'est appropriée que lorsque l'échantillon est seulement aléatoire.

**NB : Étant donné qu'il est possible de calculer la probabilité qu'un individu soit sélectionné dans un échantillon aléatoire ( $n/N$ ), on peut estimer la probabilité de commettre une *erreur aléatoire d'échantillonnage* (ou marge d'erreur) en généralisant les résultats à la population dont est tiré l'échantillon ☺**

**3) Un sociologue de la santé étudie l'ampleur de la consommation concomitante de médicaments psychotropes dans une zone épidémiologique. Les psychotropes concernés sont : somnifères, antidépresseurs, régulateurs de l'humeur, antipsychotiques, anxiolytiques, stimulants.**

- a) Combien de combinaisons de deux psychotropes peut-il former dans cet ensemble de six psychotropes? Montrez les calculs effectués.

$$C^2_6 = \frac{6!}{2!(6-2)!} = \frac{6!}{2! \cdot 4!} \\ = \frac{6 \cdot 5 \cdot 4 \cdot 3 \cdot 2 \cdot 1}{2 \cdot 1 \cdot (4 \cdot 3 \cdot 2 \cdot 1)} = 15$$

Le sociologue peut former 15 combinaisons de deux psychotropes dans un ensemble formé de six psychotropes.

- b) Nommez les différentes combinaisons possibles.

1. Somnifères-antidépresseurs
2. Somnifères-régulateurs de l'humeur
3. Somnifères-antipsychotiques
4. Somnifères-anxiolytiques
5. Somnifères-stimulants
6. Antidépresseurs-régulateurs de l'humeur
7. Antidépresseurs-antipsychotiques
8. Antidépresseurs-anxiolytiques
9. Antidépresseurs-stimulants
10. Régulateurs-antipsychotiques
11. Régulateurs-anxiolytiques

- 12. Régulateurs-stimulants
- 13. Antipsychotiques-anxiolytiques
- 14. Antipsychotiques-stimulants
- 15. Anxiolytiques-stimulants

4) Pour chacun des deux sondages ci-dessous, peut-on inférer de façon fiable les résultats à toute la population en recourant à l'estimation par intervalle de confiance ? Justifiez votre réponse :

a) *Le premier sondage aléatoire porte sur 25 adolescents québécois rencontrés dans un Cegep. Parmi ces répondants, 30% déclarent posséder un iPhone. Les adolescents y utilisent Internet en moyenne 8 heures par semaine.*

Il n'est pas approprié d'utiliser l'estimation pour inférer les résultats à toute la population, car la taille de l'échantillon est tellement faible (25) que celui-ci n'est probablement pas représentatif de la population québécoise de laquelle il est tiré. La *loi des grands nombres*, qui postule que les caractéristiques de l'échantillon s'approchent d'autant plus de celles de la population que la taille de l'échantillon est importante, s'avérerait difficilement applicable à la situation en question.

b) *Un sondage CROP qui s'est terminé la veille du congédiement de Pierre Gauthier (Directeur général du Canadien de Montréal) indique que 61% des 800 Québécois ayant rempli le questionnaire souhaitaient son départ et celui de l'entraîneur Randy Cunneyworth. La collecte de données en ligne s'est déroulée du 28 au 29 mars 2012 par le biais d'un panel web auquel les répondants ont **volontairement** participé.*

Il n'est pas approprié d'utiliser l'estimation pour inférer les résultats à toute la population, car l'échantillonnage est de type volontaire. Certes, la taille de l'échantillon demeure importante. Mais l'échantillon de volontaires n'est pas complètement aléatoire puisque tous les individus de la population recensée n'ont pas une chance non nulle indépendante et égale d'y faire partie. Au contraire, le choix des unités statistiques demeure arbitraire. On peut supposer que les personnes qui se soucient, d'une façon ou d'une autre, de la problématique de l'étude tendent à participer au sondage web. Or ces « panélistes » ne se manifestent pas, comme par enchantement, par hasard au sens statistique du terme. La technique d'échantillonnage de volontaires ne fournit donc ni l'assurance d'une forte représentativité de l'échantillon ni la possibilité d'une estimation précise des biais en découlant inévitablement. Dans ce genre de situation, le *théorème central limite*, qui postule que la distribution des statistiques de tous les échantillons possibles tend à suivre un modèle normal prédictible, serait difficilement applicable.

NB : Si l'on infère à toute une population les résultats d'un **sondage non aléatoire**, où l'on ne peut estimer la probabilité qu'un répondant soit sélectionné plutôt qu'un autre, il faut être conscient des limites d'une telle inférence. Bref, lorsque vous utilisez l'inférence statistique sur la base d'un échantillon non aléatoire, vous commettez une **erreur biaisée d'échantillonnage** que l'on ne peut déterminer. Également, il existe des **erreurs non dues à l'échantillonnage** que l'on ne peut estimer : erreurs de mesure, de saisie, de calcul des données, non-réponses 😊

5) Les médias sociaux occupent de plus en plus de place dans la vie des jeunes. Selon un sondage effectué au hasard sur 2000 Québécois âgés de 18 à 34

**ans, ces derniers passent en moyenne 8 heures par semaine sur le site de médias sociaux tels Facebook, Instagram et Twitter. Ces résultats sont précis à  $\pm 56$  minutes, avec un risque d'erreur de 5%. Précisez les informations demandées ci-dessous:**

Variable étudiée : Temps passé sur le site des médias sociaux

Population étudiée : Les Québécois âgés de 18 à 34 ans (unité d'analyse)

Taille de l'échantillon : 2000 (cas)

Mode d'échantillonnage : Échantillonnage probabiliste ou aléatoire

Statistique mesurée : Moyenne

Valeur ponctuelle estimée du paramètre : 8 heures

Marge d'erreur : 56 minutes

Intervalle estimé du paramètre: 7 heures 4 mn et 8 heures 56 mn

Probabilité d'erreur : 5%

Niveau de confiance : 95%

Que signifie la valeur de la marge d'erreur? Elle signifie l'écart entre la moyenne de l'échantillon et la vraie moyenne de la population de laquelle est tiré cet échantillon. C'est l'erreur en plus ou en moins de la moyenne, mais pas les deux à la fois.

Que signifie l'expression « risque d'erreur de 5% »? Elle signifie que si l'on répétait 100 fois le sondage ou l'expérience de sélection de l'échantillon, on pourrait s'attendre à ce qu'il ait seulement 5 fois de chance que la vraie moyenne de la population ne soit pas comprise dans l'intervalle défini par l'estimation. Autrement dit, sur 100 échantillons, seuls 5 auraient des moyennes s'écartant de l'estimation.

**6) Un sondage aléatoire effectué par CROP en mars 2012 sur 1000 Québécois révèle que 51 % des répondants demeurent en faveur de la hausse des frais de scolarité de 1625 \$ sur cinq ans. Ces résultats sont précis à  $\pm 3,09\%$ , 19 fois sur 20. Précisez les informations demandées ci-dessous :**

Variable étudiée : Opinion concernant la hausse des frais de scolarité

Population étudiée : Les Québécois (unité d'analyse)

Taille de l'échantillon : 1000 (cas)

Mode d'échantillonnage : Échantillonnage probabiliste ou aléatoire

Statistique mesurée : Pourcentage (ou proportion)

Valeur ponctuelle estimée du paramètre : 51%

Marge d'erreur : 3,09%

Intervalle estimé du paramètre: 47,91% et 54,09%

Probabilité d'erreur : 1 fois sur 20 (ou 5%)

Niveau de confiance : 19 fois sur 20 (ou 95%)

Que signifie la valeur de la marge d'erreur?: Elle signifie l'écart entre la proportion de l'échantillon et la vraie proportion de la population de laquelle est tiré cet échantillon. C'est l'erreur en plus ou en moins du pourcentage, mais pas les deux à la fois.

Que signifie l'expression « 19 fois sur 20 »?: Si l'on répétait 20 fois le sondage, on pourrait s'attendre à ce qu'il y ait 19 fois de chance que la vraie proportion de la population soit comprise dans l'intervalle défini par l'estimation. Autrement dit, sur 100 échantillons, 95 auraient des proportions comprises dans l'estimation.

**7) Un sondage supposément aléatoire portant sur 2246 étudiants donne un âge moyen de 22,56 ans, avec plus ou moins 4,77 (écart-type). Vous voulez estimer la moyenne de la population. Pour cela, répondez aux questions ci-dessous.**

a) Calculez l'erreur-type.

L'erreur-type vaut 0,10066.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{4,77}{\sqrt{2246}} = 0,1006$$

b) Interprétez la valeur de l'erreur-type ainsi calculée

L'erreur-type signifie que l'écart-type de toutes les moyennes d'échantillons possibles de taille  $n=2246$  dans la population est égal à 0,1006, concernant l'âge des étudiants. Autrement dit, si l'on répétait 100 fois le sondage ou l'expérience de sélection de l'échantillon, 68% des moyennes d'échantillons possibles se situeraient à plus ou moins 0,1006 de la moyenne 22,56 ans.

**NB : Si l'écart-type concerne l'écart typique des scores d'une distribution d'un échantillon (scores-X ou âges individuels), l'erreur-type réfère à l'écart typique des moyennes ou proportions d'une distribution d'échantillonnage (statistiques comme les âges moyens de tous les échantillons possibles X). L'erreur-type s'interprète comme l'écart-type en conséquence 😊**

c) Calculez la marge d'erreur avec un niveau de confiance à 95%.

La marge d'erreur vaut 0,197.

$$E = z * \sigma_{\bar{x}} = 1,96 * 0,1006 = 0,197$$

d) Calculez l'intervalle de confiance à 95%.

L'intervalle de confiance à 95% vaut 22,36 et 22,76.

$$IC = \bar{x} \pm z * \sigma_{\bar{x}} = 22,56 \pm 0,197 = 22,56 - 0,197 \text{ à } 22,56 + 0,197 = 22,36 \text{ à } 22,76$$

e) Interprétez les valeurs de l'intervalle de confiance.

On est sûr à 95% que l'âge moyen chez la population étudiante se situe entre 22,36 et 22,76 ans, c'est-à-dire respectivement la plus petite moyenne échantillonnale et la plus grande moyenne échantillonnale que le hasard de l'échantillonnage peut donner. Autrement dit, les étudiants ont en moyenne 22,56 ans, avec une marge d'erreur de  $\pm 0,197$ , 19 fois sur 20 (ou 95 fois sur 100).

**8) Un sondage Web aléatoire a été mené par SOM du 11 au 17 août 2009 auprès**

de 1297 internautes québécois qui travaillent en entreprise ou dans une organisation. Parmi les internautes qui ont accès à Internet au travail, le taux d'utilisation du Web à des fins personnelles est de 66 %.

a) Estimez par intervalle de confiance le pourcentage de la population en indiquant la plus petite proportion échantillonnale et la plus grande proportion échantillonnale que le hasard peut donner, tout en optant pour une estimation précise dans 99 cas échantillonnés pour 100 (99%).

Calculons d'abord l'erreur-type :

$$\sigma_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{66 * 34}{1297}} = 1,31$$

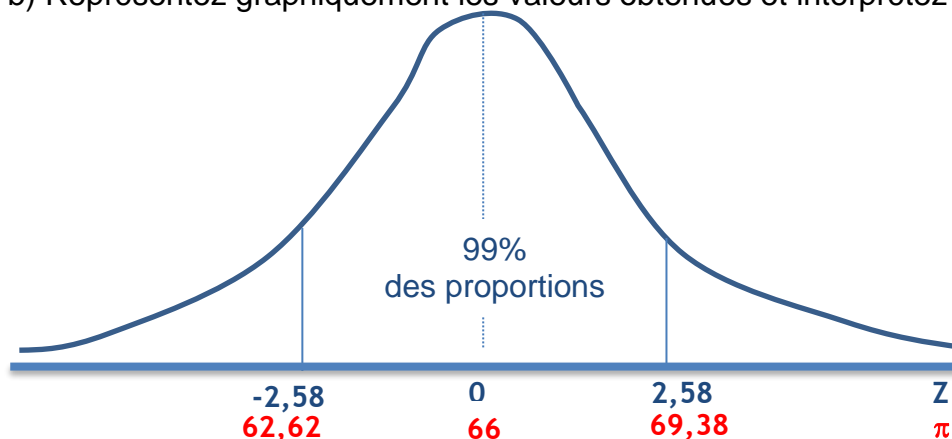
Calculons ensuite la marge d'erreur à 99%

$$E = z * \sigma_p = 2,58(1,31) = 3,38$$

Calculons enfin l'intervalle de confiance à 99%

$$IC = p \pm E = 66 \pm 3,38 = 62,62\% \text{ à } 69,38\%$$

b) Représentez graphiquement les valeurs obtenues et interprétez-les.



On est sûr à 99% que le taux d'utilisation du Web à des fins personnelles dans la population québécoise travaillant en entreprise ou dans une organisation se situe entre 62,6% et 69,4%, c'est-à-dire respectivement la plus petite proportion échantillonnale et la plus grande proportion échantillonnale que le hasard peut donner. Autrement dit, les travailleurs internautes québécois sont 66% à utiliser le Web à des fins personnelles, avec une marge d'erreur de  $\pm 3,38\%$ , 99 fois sur 100.

**9) Le changement climatique constitue un phénomène préoccupant et émergent. C'est ce que laisse croire une étude effectuée, entre 2010 et 2011, par Léger Marketing auprès de Nord-Américains âgés de 18 ans et plus. Parmi les 1214 Canadiens sélectionnés aléatoirement (2011), 80% croient au réchauffement de la terre. Par contre, parmi les 805 Américains sélectionnés aléatoirement (2010), 68% seulement croient à ce phénomène.**

a) Estimez par intervalle de confiance (99%) le pourcentage de la population canadienne qui croit au réchauffement de la terre.

Calculons d'abord l'erreur-type :

$$\sigma_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{80 * 20}{1214}} = 1,15$$

Calculons ensuite la marge d'erreur à 99%

$$E = z * \sigma_p = 2,58(1,15) = 2,96$$

Calculons enfin l'intervalle de confiance à 99%

$$IC = p \pm E = 80 \pm 2,96 = 77,04\% \text{ à } 82,96\%$$

Selon les résultats du sondage, on peut estimer qu'il y a 99% de chance que le pourcentage de Canadiens qui croient au réchauffement de la terre dans la population tout entière se situe entre 77% et 83%, c'est-à-dire respectivement la plus petite proportion échantillonnale et la plus grande proportion échantillonnale que le hasard peut donner. Autrement dit, les Canadiens sont 80% à croire au changement climatique, avec une marge d'erreur de  $\pm 2,96\%$ , 99 fois sur 100.

- b) Estimez par intervalle de confiance (99%) le pourcentage de la population américaine qui croit au réchauffement de la terre.

Calculons d'abord l'erreur-type

$$\sigma_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{68 * 32}{805}} = 1,64$$

Calculons ensuite la marge d'erreur à 99%

$$E = z * \sigma_p = 2,58(1,64) = 4,23$$

Calculons enfin l'intervalle de confiance à 99%

$$IC = p \pm E = 68 \pm 4,23 = 63,77\% \text{ à } 72,23\%$$

Selon les résultats du sondage, on peut estimer qu'il y a 99% de chance que le pourcentage d'Américains qui croient au réchauffement de la terre dans la population tout entière se situe entre 63,8% et 72,2%, c'est-à-dire respectivement la plus petite proportion échantillonnale et la plus grande proportion échantillonnale que le hasard peut donner. Autrement dit, les Américains sont 68% à croire au changement climatique, avec une marge d'erreur de  $\pm 4,23$ , 99 fois sur 100.

- c) Construisez un tableau présentant les intervalles de confiance pour les sondages canadien et américain (% , marge d'erreur et nombre de cas)

Croyance au réchauffement climatique selon le pays

Croyance au réchauffement	Pays	
	Canada	États-Unis
Pourcentage	80	68
Marge d'erreur	$\pm 2,96$	$\pm 4,23$
Nombre de cas (n)	1214	805

Source : Léger-Marketing, 2010-2011



- d) Pourquoi la marge d'erreur est-elle plus importante dans le sondage américain que dans le sondage canadien concernant la croyance au changement climatique?

D'une part, parce que la taille de l'échantillon américain est plus petite que celle de l'échantillon canadien. Plus la taille de l'échantillon augmente, plus la marge d'erreur tend à diminuer.

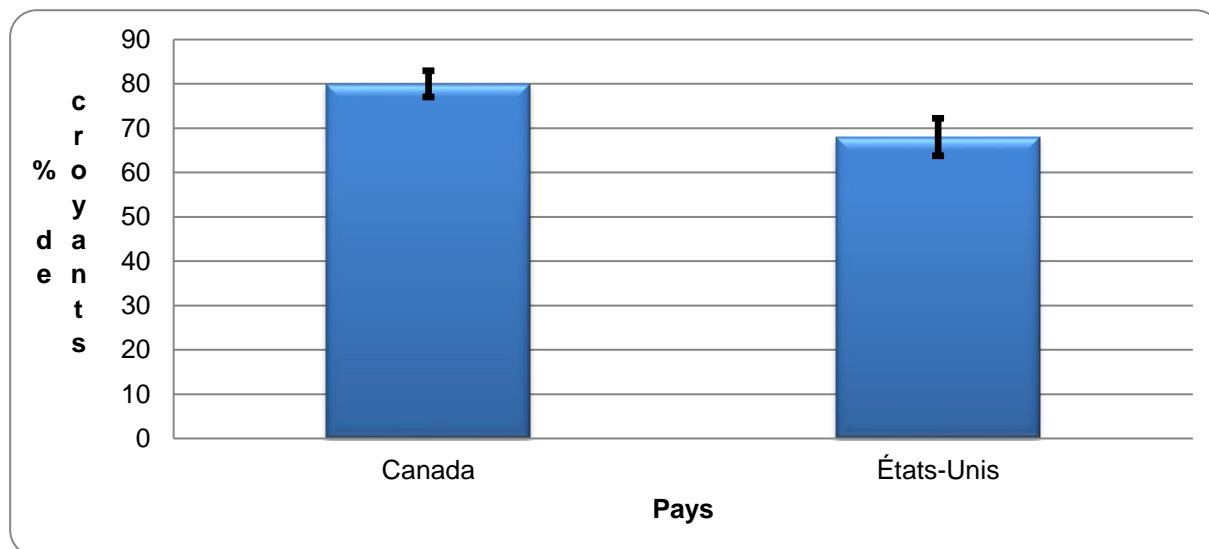
D'autre part, parce que le pourcentage de l'échantillon américain est plus proche de 50%, qui donne une marge d'erreur maximale. Plus le pourcentage à estimer s'éloigne de 50% (0,50) en plus ou moins, plus la marge d'erreur tend à diminuer.

- e) Peut-on affirmer que la croyance au changement climatique est plus forte au Canada qu'aux USA ? Pour savoir s'il y a une différence significative entre les deux pourcentages, comparez les deux intervalles de confiance à l'aide des barres d'erreur. Si elles sont disjointes, la différence est significative.

Les barres d'erreur montrent que les intervalles de confiance pour le cas canadien (83,88% à 86,12%) et pour le cas américain (63,77% à 72,23%) ne se chevauchent pas. On peut donc affirmer que la croyance au changement climatique est significativement plus forte au Canada qu'aux USA. La différence entre les pourcentages est tellement importante qu'elle ne peut être due au hasard de l'échantillonnage. Elle semble significative.

**NB : Pour conclure à une différence significative entre deux groupes eu égard à un phénomène quantitatif, il faut que les intervalles de confiance (barres d'erreur) soient disjoints ☺**

Croyance concernant le réchauffement climatique en % au Canada et aux États-Unis



- f) Quelle est la marge d'erreur maximale du sondage canadien avec un niveau de confiance de 99%?

D'une part, calculons l'erreur-type pour une proportion de 50%

$$\sigma_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{50 * 50}{1214}} = 1,44$$

D'autre part, calculons la marge d'erreur du sondage à 99%

$$E = z * \sigma_p = 2,58(1,44) = 3,7$$

La marge d'erreur maximale du sondage canadien est de 3,7, 99 fois sur 100.

- g) Quelle est la marge d'erreur maximale du sondage américain avec un niveau de confiance de 99%?

D'une part, calculons l'erreur-type pour une proportion de 50%

$$\sigma_p = \sqrt{\frac{pq}{n}} = \sqrt{\frac{50 * 50}{805}} = 1,76$$

D'autre part, calculons la marge d'erreur du sondage à 99%

$$E = z * \sigma_p = 2,58(1,76) = 4,55$$

La marge d'erreur maximale du sondage américain est de 4,55, 99 fois sur 100.

- e) Comparez les deux marges d'erreur maximales et justifiez.

Le sondage américain présente la marge d'erreur la plus élevée en raison de l'importance moins grande de la taille de l'échantillon.

### 10) Pour démarrer un nouveau programme de criminologie dans une université, un jury recrute des candidats par le biais d'une entrevue de sélection.

- a) Quels sont les deux types d'erreur que le jury peut commettre en prenant une décision eu égard aux candidats?

D'une part, **rejeter un candidat alors qu'il est bon** : les conditions de l'entrevue ou les caractéristiques du candidat peuvent le faire échouer alors qu'il est bon.

D'autre part, **accepter un candidat alors qu'il est mauvais** : les conditions de l'entrevue et les caractéristiques du candidat peuvent l'aider à réussir alors qu'il n'est pas bon.

**NB : Les statisticiens nomment ces deux types d'erreur : erreur de type I ou de première espèce (rejeter un bon candidat) et erreur de type II ou de deuxième espèce (accepter un mauvais candidat). Aussi, rejeter l'hypothèse nulle alors qu'elle est vraie (bonne) relève de l'erreur de type I; accepter l'hypothèse nulle alors qu'elle est fausse (mauvaise) relève de l'erreur de type II 😊**

- b) Peut-il éviter de commettre ces types d'erreur? Justifiez.

Il est difficile pour le jury d'éviter de commettre les deux types d'erreur en raison de la

situation d'**information imparfaite** dans laquelle il se trouve. Au fait, il est difficile en quelques minutes de recueillir toutes les informations nécessaires à une prise de décision éclairée, d'autant plus que les membres du jury n'ont pas forcément une attitude neutre à l'égard des candidats.

c) Peut-il minimiser ou réduire ces types d'erreur? Si oui, comment?

Il est toujours possible pour le jury de minorer les deux types d'erreur en cherchant à obtenir **plus d'informations** sur les candidats soumis au processus de sélection. À titre d'exemple, les membres du jury peuvent soumettre les candidats à un test écrit de statistiques sociales pour jauger leur habileté à décrire, expliquer et prévoir des comportements criminels à l'aide de techniques statistiques appropriées.

d) Si l'on minimise l'un des types d'erreur, comment se comporte l'autre type?

En tentant de minimiser l'un des types d'erreur, on court le risque de commettre l'autre type d'erreur. Par exemple, si le programme de criminologie est **moins exigeant**, le jury aura tendance à tout faire pour ne pas commettre l'erreur de type I, c.-à-d. **rejeter de bons candidats**. Mais, en même temps, on court le risque d'**accepter de mauvais étudiants**, soit de commettre l'erreur de type II.

De même, si le programme de criminologie est **très exigeant | contingenté**, le jury aura tendance à tout faire pour ne pas commettre l'erreur de type II, c.-à-d. **accepter de mauvais étudiants**. Mais, par la même occasion, on court le risque de **rejeter de bons candidats**, soit de commettre l'erreur de type I.

**NB : Chaque fois qu'on est en présence d'une situation de décision où l'on doit choisir entre deux possibilités, on commet presque toujours deux types d'erreur et si on fait tout pour ne pas commettre l'une des erreurs, on augmente les chances de commettre quand même l'autre. Ces deux types d'erreur font partie intégrante de la vie quotidienne : le juge qui décide de la culpabilité ou non d'un prévenu, le professeur qui décide d'accommoder ou non un étudiant dont l'ordinateur s'est supposément planté, le joueur de soccer qui décide de placer ou de tirer en force un penalty, etc. On ne peut, en aucune façon, éliminer complètement la possibilité de commettre un type d'erreur particulier. Par contre, on peut déterminer la probabilité de commettre par exemple une **erreur de type I**, c.-à-d. rejeter l'hypothèse nulle alors qu'elle s'avère vraie : c'est tout le sens des tests d'hypothèse ☺**

Par ailleurs, dans *La logique de la découverte scientifique* (1934), Popper considère qu'une hypothèse scientifique doit pouvoir être réfutée, falsifiable. Le principe de réfutabilité ou falsifiabilité est illustré par l'**exemple des cygnes**. L'hypothèse "Tous les cygnes sont blancs" ne peut être vérifiée dans la mesure où il est impossible de recenser l'intégralité de cette population. Il est, en revanche, facile de la réfuter, dès l'instant où l'on découvre un cygne noir (Popper : 1934). Pour vérifier une hypothèse sur des données d'échantillon, il faut démontrer son contraire de façon à la réfuter. Ainsi, les tests d'hypothèse reposent sur le principe épistémologique de la réfutabilité ou falsifiabilité de Popper. Ils renseignent sur la probabilité de commettre une erreur en rejetant l'hypothèse nulle  $H_0$  (type I). Si cette probabilité est faible (c.-à-d. inférieure à un seuil de signification  $\alpha = 0,05$  ou 5%), on rejette l'hypothèse nulle  $H_0$ . Par contre, si elle est élevée, on accepte plutôt l'hypothèse nulle  $H_0$  ☺