TS1

Contrôle du jeudi 23 mai 2013 (45 minutes)

Prénom : ………………………………………..………. **Nom :** ………………… **Note : …. / 20**

On prendra les intervalles de fluctuation non simplifiés et on arrondira au millième les bornes de tous les intervalles.

I. (5 points)

Des statistiques, réalisées sur un temps suffisamment long, amènent à penser qu'en période de compétition, la probabilité pour qu'un athlète pris au hasard soit déclaré positif au contrôle anti-dopage est $p = 0,02$.

Lors d'une compétition, 400 sportifs sont contrôlés.

1°) Déterminer l'intervalle de fluctuation asymptotique au seuil de 95 % (noté I) pour la fréquence des sportifs déclarés positifs.

2°) Le comité organisateur de la compétition décide que les résultats des épreuves seront annulés si, lors du contrôle anti-dopage, la fréquence des résultats positifs dépasse la borne supérieure de l'intervalle de fluctuation.

À partir de combien de résultats positifs au contrôle antidopage, la compétition sera-t-elle annulée ?

II. (5 points)

On a effectué un sondage auprès des 15-24 ans pour connaître leur préférence en matière de lecture. La question est formulée de la façon suivante : « Préférez-vous les romans ou les témoignages vécus ? ».

1°) On interroge un premier échantillon de 100 personnes. 48 préférèrent les romans et 52 les témoignages vécus.

a) Déterminer un intervalle de confiance au niveau 95 % pour la proportion des amateurs de romans dans la population 15-24 ans.

b) Cet intervalle permet-il de savoir lequel des deux types d'ouvrages a la préférence des jeunes ?

2°) En gardant la même fréquence observée qu'au 1°), quelle taille d'échantillon aurait permis de conclure ?

III. (5 points)

Partie A

Monsieur Z, chef du gouvernement d'un pays lointain, affirme au début de sa campagne que 52 % des électeurs lui font confiance. On réalise un sondage dans cette population en interrogeant 625 personnes au hasard.

1°) Préciser quel est l'intervalle de fluctuation, au seuil de 95 %, de la fréquence f d'opinions favorables sur un échantillon de taille 625 dans cette population. Calculer les bornes de cet intervalle.
2°) À l'issue de ce sondage, le taux d'opinions favorables est de 47 %. Doit-on émettre un doute sur le pourcentage de 52 % proclamé par Monsieur Z ?

Partie B

Les semaines passent et les élections approchent. La radio annonce : *« D'après les nouveaux sondages effectué auprès de 1000 personnes, la cote de Monsieur Z remonte de 49 à 52 %. Il a donc maintenant les faveurs d'une majorité de la population. »*

1°) a) Avec 49 % d'opinions favorables lors du premier sondage effectué auprès de 1000 personnes, estimer, au niveau de confiance 0,95, la proportion d'opinions favorables à Monsieur Z.
b) Avec 52 % d'opinions favorables lors du deuxième sondage effectué auprès de 1000 personnes, estimer, au niveau de confiance 0,95, la proportion d'opinions favorables à Monsieur Z.
2°) Représenter les deux intervalles précédents sur un même axe gradué et commenter l'annonce faite à la radio.

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

IV. (5 points)

Un maraîcher produit des tomates selon un mode d'agriculture biologique. Il a gardé des semences des deux années précédentes, 2010 et 2011, et il voudrait utiliser ces semences pour les planter en 2012.
Pour ce faire, il doit d'abord contrôler que les taux de germination des semences de 2011 et 2012 sont très proches.
Il décide alors de réserver deux parcelles distinctes pour tester dans l'une 200 graines de 2010, et dans l'autre 400 graines de 2011.
Dans la première parcelle, 180 germent, et dans la deuxième, 280.
On appelle p_0 le taux de germination des semences de 2010 et p_1 celui des semences de 2011.

1°) a) Déterminer un intervalle de confiance au niveau 0,95 pour p_0.
b) Même question pour p_1.
2°) Peut-on juger que la différence entre les deux taux de germination est significative au seuil de 5 % ?

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

……………………………………………………………………………………………

Corrigé du contrôle du 23-5-2013

I.

1°) **Déterminons l'intervalle de fluctuation asymptotique au seuil de 95 % (noté I) pour la fréquence des sportifs déclarés positifs.**

$n = 400 > 30$ $\quad p = 0,02$

$np = 8 > 5$ $\quad n(1 - p) = 392 > 5$

Donc les conditions sont réunies pour donner un intervalle de fluctuation asymptotique au seuil de 95 %.

L'intervalle de fluctuation asymptotique de la fréquence des sportifs déclarés positifs au seuil de 95 % est

$$\mathrm{I} = \left[0,02 - 1,96\sqrt{\frac{0,02(1-0,02)}{400}} \,;\, 0,02 + 1,96\sqrt{\frac{0,02(1-0,02)}{400}} \right].$$

Avec la calculatrice, on obtient :

$$0,02 - 1,96\sqrt{\frac{0,02(1-0,02)}{400}} = 0,00628...$$

$$0,02 + 1,96\sqrt{\frac{0,02(1-0,02)}{400}} = 0,03372...$$

En arrondissant les bornes au millième (comme le demande l'énoncé au début), on peut donner : $\mathrm{I} = [0,006\,;\,0,034]$.

2°) **Cherchons le nombre *x* de résultats positifs au contrôle antidopage de telle manière que la compétition soit annulée.**

La borne supérieure de I est environ égale à 0,034.

Pour les calculs (et c'est même obligatoire sur le plan de la rigueur mathématique), il faut travailler avec la valeur exacte (« grosse » expression).

La fréquence de résultats positifs au contrôle anti-dopage est égale à $\dfrac{x}{400}$.

La compétition est annulée si $\dfrac{x}{400} > 0,02 + 1,96\sqrt{\dfrac{0,02(1-0,02)}{400}}$

soit $x > 400\left(0,02 + 1,96\sqrt{\dfrac{0,02(1-0,02)}{400}}\right)$

soit $x > 13,488...$.

Comme *x* est un entier naturel, on en déduit que la compétition sera annulée à partir de 14 résultats positifs.

II.

1°) a) **Déterminons un intervalle de confiance au niveau 95 % pour la proportion des amateurs de romans dans la population 15-24 ans.**

La fréquence de lecteurs de romans est $f = \dfrac{48}{100} = 0,48$.

Un intervalle de confiance au niveau 95 % de la proportion de lecteurs de romans dans la population 15-24 ans est $\mathrm{I} = \left[0,48 - \dfrac{1}{\sqrt{100}}\,;\,0,48 + \dfrac{1}{\sqrt{100}}\right]$ soit $\mathrm{I} = [0,38\,;\,0,58]$ (il s'agit d'une « fourchette de sondage »).

On peut vérifier *a posteriori* que cet intervalle est valable car $\begin{cases} 100 \times 0,38 = 38 > 5 \\ 100 \times (1 - 0,58) = 42 > 5 \end{cases}$.

b) **Déterminons si l'intervalle de confiance déterminé à la question précédente permet de savoir lequel des deux types d'ouvrages a la préférence des jeunes.**

• 1ère manière de rédiger :

L'intervalle de confiance I déterminé à la question précédente ne permet pas de savoir quel type d'ouvrages a la préférence parmi les 15-24 car il contient des valeurs inférieures à 0,5 et supérieures à 0,5.

• 2e manière de rédiger :

$0,5 \in \mathrm{I}$ donc on ne peut pas savoir à l'aide de cet intervalle quel type d'ouvrages a la préférence des jeunes.

Idée exprimée par Dimitri Joly :

Il faut un intervalle plus large pour pouvoir conclure (cf. question suivante).

Erreur trouvée chez quelques élèves (confusion de méthode) :

On détermine l'intervalle de confiance J de la proportion de lecteurs de témoignages dans la population 15-24 ans.
On observe que I et J ne sont pas disjoints.
On conclut qu'il n'y a pas de différence significative.
Le raisonnement est complètement faux.

2°) **Déterminons la taille *n* d'un nouvel échantillon qui permettrait de conclure.**

En gardant la même fréquence observée qu'au 1°), un intervalle de confiance au niveau 0,95 de la proportion des amateurs de romans dans la population 15-24 ans pour un échantillon de taille *n* est $\mathrm{I} = \left[0,48 - \dfrac{1}{\sqrt{n}}\,;\,0,48 + \dfrac{1}{\sqrt{n}}\right]$.

La borne inférieure de I est toujours strictement inférieure à 0,5.
Donc on pourra conclure sur le type d'ouvrages qui a la préférence parmi les 15-24 ans lorsque la borne supérieure de I est strictement inférieure à 0,5 c'est-à-dire $0,48 + \dfrac{1}{\sqrt{n}} < 0,5$ (1).

$(1) \Leftrightarrow \frac{1}{\sqrt{n}} < 0,02$

$\Leftrightarrow \sqrt{n} > 50$

$\Leftrightarrow n > 2500$

Conclusion :

Pour pouvoir conclure, il suffit que la taille de l'échantillon soit strictement supérieure à 2500.

Un échantillon de 2501 jeunes aurait donc permis de conclure.

Autre rédaction possible en utilisant l'autre formule donnant un intervalle de confiance au niveau 0,95 :

On cherche n tel que la borne supérieure de l'intervalle de confiance soit strictement inférieure à 0,5 c'est-à-dire $f + 1,96\sqrt{\frac{f(1-f)}{n}} < 0,5 \quad (1)$.

$(1) \Leftrightarrow 0,48 + 1,96\sqrt{\frac{0,48 \times 0,52}{n}} < 0,5$

$\Leftrightarrow \sqrt{\frac{0,2496}{n}} < \frac{0,02}{1,96}$

$\Leftrightarrow \frac{0,2496}{n} < \left(\frac{0,02}{1,96}\right)^2$

$\Leftrightarrow n > \left(\frac{1,96}{0,02}\right)^2 \times 0,2496$

$\Leftrightarrow n > 2\,397,1584$

Un échantillon de 2 398 jeunes aurait donc permis de conclure.

III.

Partie A

1°)

$p = 0,52$

$$\begin{cases} n = 625 > 30 \\ np = 325 > 5 \\ n(1-p) = 300 > 5 \end{cases}$$

• Avec la formule non simplifiée :

Un intervalle de fluctuation au seuil de 95 % de la fréquence f est $\mathrm{I} = \left[p - 1,96\sqrt{\frac{p(1-p)}{400}} \,;\, p + 1,96\sqrt{\frac{p(1-p)}{400}} \right]$ soit $\mathrm{I} = [0,481 \,;\, 0,559]$ (en arrondissant les bornes au millième).

• Avec la formule simplifiée :

Un intervalle de fluctuation au seuil de 95 % de la fréquence f d'opinions favorables pour Monsieur Z est $\mathrm{I} = \left[p - \frac{1}{\sqrt{625}} \,;\, p + \frac{1}{\sqrt{625}} \right]$ soit $\mathrm{I} = [0,480 \,;\, 0,560]$.

2°) **Cherchons si l'on doit émettre un doute sur le pourcentage de 52 % proclamé par Monsieur Z.**

La fréquence observée est $f = 0,47$.

On a : $f \notin \mathrm{I}$.

On peut donc rejeter l'hypothèse de Monsieur Z au risque d'erreur de 5 % (ou « au seuil de décision de 5 % »), c'est-à-dire que l'on peut émettre un doute sur le pourcentage de 52 % qu'il a donné.

Partie B

1°) a) **Avec 49 % d'opinions favorables lors du premier sondage effectué auprès de 1000 personnes, estimons, au niveau de confiance 0,95, la proportion d'opinions favorables à Monsieur Z.**

Un intervalle de confiance de la proportion d'opinions favorables à Monsieur Z au niveau de 95 % pour ce premier sondage est $\mathrm{I}_1 = \left[0,49 - \frac{1}{\sqrt{1000}} \,;\, 0,49 + \frac{1}{\sqrt{1000}} \right]$ soit $\mathrm{I}_1 = [0,458 \,;\, 0,522]$.

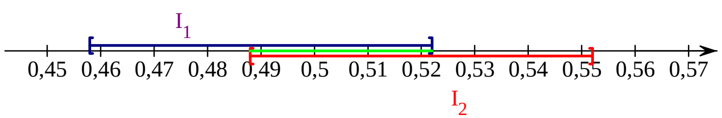
Il s'agit d'une fourchette de sondage.

On vérifie que les conditions d'application de la formule sont vérifiées.

b) **Avec 52 % d'opinions favorables lors du deuxième sondage effectué auprès de 1000 personnes, estimons, au niveau de confiance 0,95, la proportion d'opinions favorables à Monsieur Z.**

De même, un intervalle de confiance de la proportion d'opinions favorables à Monsieur Z au niveau de 95 % pour ce deuxième sondage est $\mathrm{I}_2 = [0,488 \,;\, 0,552]$.

2°) **Représentons les deux intervalles précédents sur un même axe gradué et commentons l'annonce faite à la radio.**



Monsieur Z n'a pas forcément les faveurs de la majorité car $0,5 \in \mathrm{I}_2$.

On ne parle pas de différence significative ; ce n'est pas du tout cette notion qui intervient ici.

IV.

1°) a) **Déterminons un intervalle de confiance au niveau 0,95 pour p_0.**

$$f_0 = \frac{180}{200} = 0,9$$

$$n_0 = 200$$

Un intervalle de confiance au niveau de 95 % de p_0 est $\mathrm{I}_0 = \left[0,9 - \frac{1}{\sqrt{200}} ; 0,9 + \frac{1}{\sqrt{200}}\right]$ donc $\mathrm{I}_0 = [0,829 ; 0,971]$ en arrondissant les bornes au millième.

On vérifie que les conditions d'application de la formule sont vérifiées.

b) **Déterminons un intervalle de confiance au niveau 0,95 pour p_1.**

$$f' = \frac{280}{400} = 0,7$$

$$n_0 = 400$$

Un intervalle de confiance au niveau de 95 % de p_1 est $\mathrm{I}_1 = [0,65 ; 0,75]$.
On vérifie que les conditions d'application de la formule sont vérifiées.

2°) **Cherchons si l'on peut juger que la différence entre les deux taux de germination est significative.**

Les intervalles I_0 et I_1 sont disjoints ($\mathrm{I}_0 \cap \mathrm{I}_1 = \varnothing$).
On peut donc considérer que la différence entre les taux de germination est significative (les graines de 2010 sont meilleures).

Il est à noter que d'après le cours, la méthode ne s'applique qu'au cas de deux échantillons de même taille, ce qui n'est pas le cas ici.

Attentes pour ce contrôle :

On attend que l'élève :

- fasse clairement la différence entre intervalle de fluctuation (fréquence) et intervalle de confiance (proportion) ;

- emploie à bon escient les termes : - de « fréquence » et de « proportion »

  - de « seuil » (risque) et de « niveau » (confiance)

L'énoncé guide en général pas mal sur les termes à employer.