

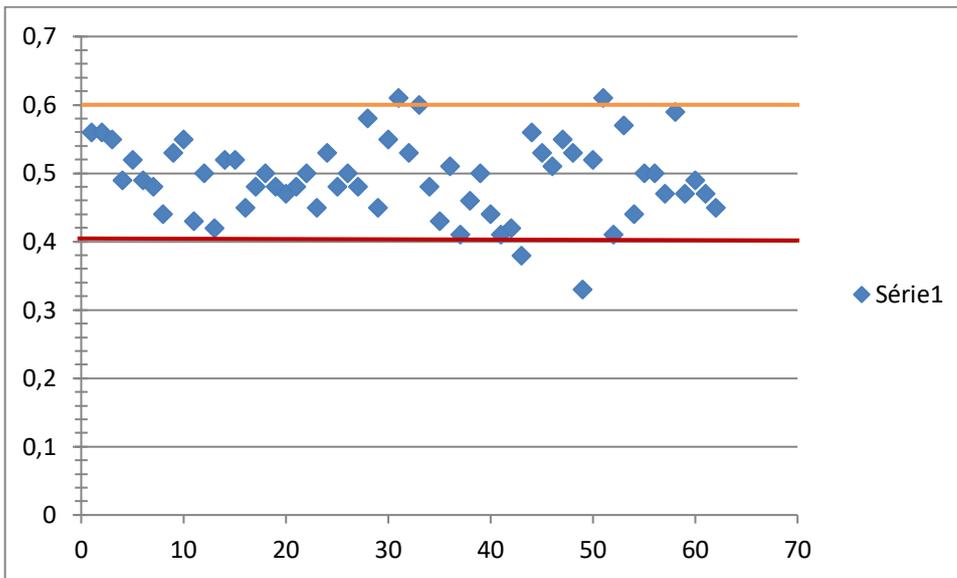
ECHANTILLONNAGE ET PRISE DE DECISION.

ECHANTILLONNAGE

Intervalle de fluctuation asymptotique, à environ 95%, d'une fréquence avec la loi normale.

On connaît la probabilité théorique et la taille de l'échantillon, on cherche dans quel intervalle se situent 95% des valeurs observées. C'est le problème typique du sondage au sortir des urnes. Penser également à la fréquence de l'apparition de pile dans le lancer d'une pièce de monnaie.

Voici la répartition de 70 échantillons de taille 100 qui représente la fréquence de piles dans le lancer d'une pièce.



En seconde, on trouve un intervalle sous la forme $\left[p - \frac{1}{\sqrt{n}}; p + \frac{1}{\sqrt{n}} \right]$. Avec $p=0,5$ et $n=100$ on a **[0,4 ; 0,6]**. Cela veut dire que dans un échantillon de 100 lancers, 95% des fréquences de piles sont dans [0,4 ; 0,6].

Utilisation de la loi binomiale.

Définition : On considère une population dont une proportion p des individus possède un caractère donné. On prélève dans cette population un échantillon de taille n .

Soit X la variable aléatoire associée au nombre d'individus possédant ce caractère. On considère que X suit une loi binomiale de paramètres n et p .

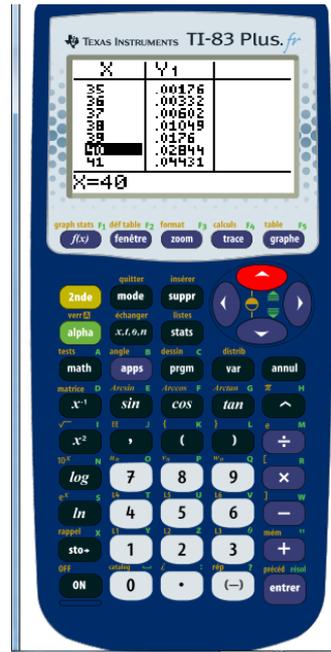
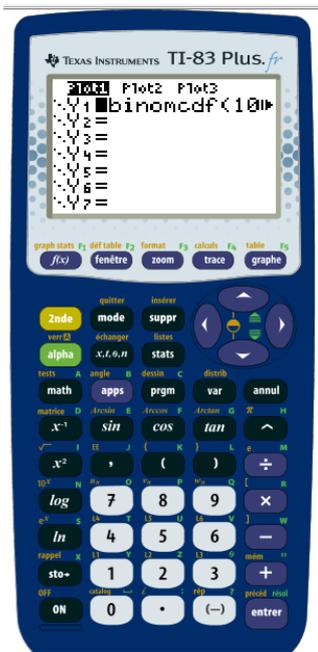
L'intervalle de fluctuation au seuil de 95 % associée à la variable aléatoire X est : $\left[\frac{a}{n}; \frac{b}{n} \right]$ où :

- a est le plus petit entier tel que : $P(X \leq a) > 0,025$,
- b est le plus petit entier tel que : $P(X \leq b) \geq 0,975$.

Indications calculatrices pour établir la table de la variable aléatoire B(100,0.5)

Sur TI :

Il faut taper Binomcdf(100,0.5,x) dans la panneau f(x) de saisie des fonctions.

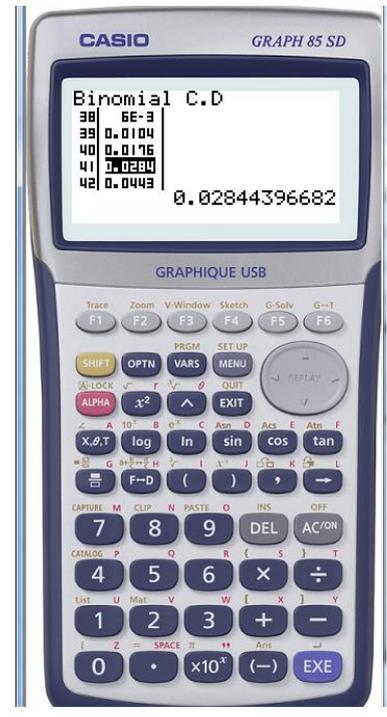
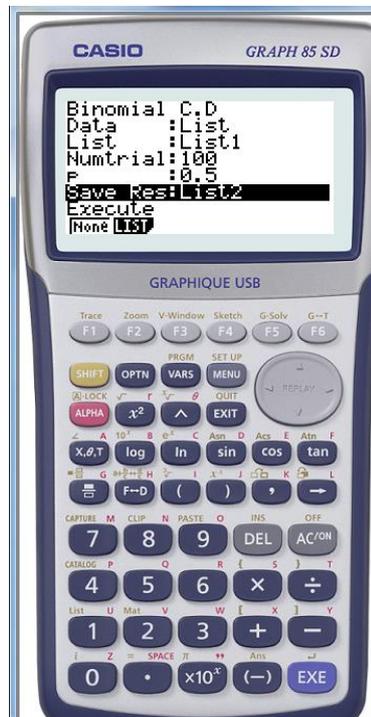
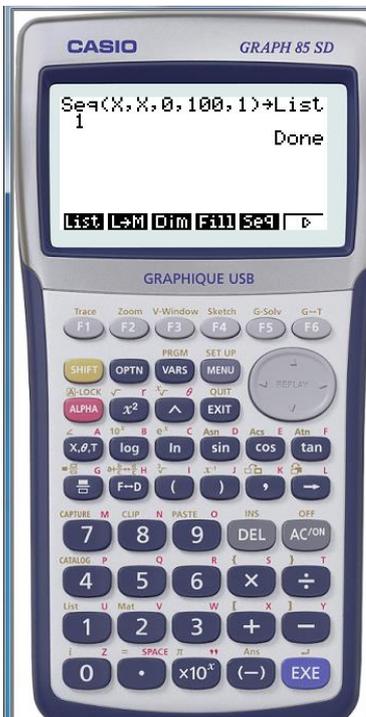


Sur Casio

Il faut d'abord installer une liste de nombres avec la commande $\text{Seq}(X,X,0,100,1) \rightarrow \text{List1}$

Seq se trouve dans CATALOG

List 1 se trouve dans OPTN F1 F1



Attention : il y a un décalage dans la numérotation avec CASIO, pour 41 il faut lire 40.

On trouve $a=40$ et $b=60$. L'intervalle de fluctuation au seuil de 95% est donc $[0,4 ; 0,6]$

1) Utilisation de la loi normale.

On affine un résultat connu qui est que dans le cadre d'une variable aléatoire X suivant une loi normale de paramètres $(\mu; \delta)$, on a $P(X \in [\mu - 2\delta; \mu + 2\delta]) \approx 0,95$

Dans le cadre d'une approximation d'une loi binomiale par une loi normale, on pose $\delta = \sqrt{n \times p \times (1 - p)}$ et $\mu = n \times p$. On affine la précision en remplaçant 2 par 1,96

Cas général

Dans le cas général le même raisonnement conduit à :

$$P(Y \in [\mu - 1,96\sigma, \mu + 1,96\sigma]) \approx 0,95$$

c'est-à-dire, en divisant par l'effectif n de l'échantillon pour passer des effectifs aux fréquences :

$$P\left(\frac{Y}{n} \in \left[\frac{\mu - 1,96\sigma}{n}, \frac{\mu + 1,96\sigma}{n}\right]\right) \approx 0,95.$$

$$\text{Or } \mu = np \text{ et } \sigma = \sqrt{np(1-p)}.$$

$$\text{Donc } \frac{\mu - 1,96\sigma}{n} = \frac{np - 1,96\sqrt{n}\sqrt{p(1-p)}}{n},$$

$$\frac{\mu - 1,96\sigma}{n} = p - 1,96\sqrt{\frac{p(1-p)}{n}}$$

$$\text{et de même } \frac{\mu + 1,96\sigma}{n} = p + 1,96\sqrt{\frac{p(1-p)}{n}}.$$

On peut alors définir un intervalle de fluctuation asymptotique.

DÉFINITION

L'intervalle de fluctuation asymptotique à environ 95 % d'une fréquence obtenue sur un échantillon de taille n est

$$\left[p - 1,96\sqrt{\frac{p(1-p)}{n}}, p + 1,96\sqrt{\frac{p(1-p)}{n}} \right]$$

où p est la proportion connue dans la population.

D'une manière générale, ces deux intervalles (binomiale et normale) sont proches dès que $n \geq 30$, $np \geq 5$ et $np(1-p) \geq 5$

Dans notre cas de lancer de pièce de monnaie avec $n=100$ et $p=0,5$ on trouve comme intervalle de fluctuation à environ 95% : **[0,402 ; 0,598]**

Prise de décision à l'aide d'un intervalle de fluctuation lié à la loi asymptotique.

RÈGLE DE DÉCISION

Au seuil de 5 %, si la fréquence observée f appartient à l'intervalle de fluctuation asymptotique $\left[p - 1,96\sqrt{\frac{p(1-p)}{n}}, p + 1,96\sqrt{\frac{p(1-p)}{n}} \right]$ on accepte l'hypothèse selon laquelle la proportion est p dans la population ; sinon, on rejette l'hypothèse selon laquelle cette proportion vaut p .

Attention, erreur dans le texte : 55% au lieu de 52%

Sondage avant une élection

ÉNONCÉ

Un mois avant l'élection où il se présente, le candidat A affirme que 52 % des électeurs veulent voter pour lui. Son adversaire B commande immédiatement un sondage portant modestement sur 100 personnes prises au hasard parmi les électeurs.

Le résultat de ce sondage donne 47 de votes favorables au candidat A.

Peut-on considérer au seuil de 5 % que l'affirmation du candidat A est vraie ?

SOLUTION

On fait l'hypothèse : l'affirmation est vraie, c'est-à-dire la proportion des électeurs qui veulent voter pour lui est $p = 0,55$.

L'intervalle de fluctuation asymptotique à 95 % de la fréquence des votes favorables au candidat A sur un échantillon de taille $n = 100$ est :

$$\left[p - 1,96\sqrt{\frac{p(1-p)}{n}}, p + 1,96\sqrt{\frac{p(1-p)}{n}} \right] = [0,4525 ; 0,6475].$$

La fréquence observée des votes favorables au candidat A dans l'échantillon est $f = 0,47$.

Elle appartient à l'intervalle de fluctuation asymptotique.

On accepte donc l'affirmation du candidat A : 55 % des électeurs veulent voter pour lui.

Extension : faire le même exercice avec la loi binomiale.

ESTIMATION

Intervalle de confiance d'une proportion au niveau de confiance de 95%.

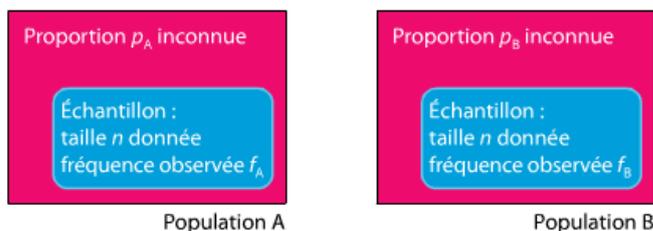
On connaît la taille de l'échantillon et une fréquence observée. On cherche dans quel intervalle va se situer la probabilité réelle. C'est le cas de la recherche d'une probabilité dans une population à partir d'échantillons.

DÉFINITION

L'intervalle $\left[f - 1,96\sqrt{\frac{f(1-f)}{n}}, f + 1,96\sqrt{\frac{f(1-f)}{n}} \right]$, où f est une fréquence obtenue sur un échantillon de taille n , est un **intervalle de confiance, au niveau de confiance 95 %**, pour la proportion correspondante p dans la population.

Comparaison de deux proportions à l'aide d'intervalles de confiance.

Cas général



On note I_A l'intervalle de confiance $\left[f_A - 1,96 \sqrt{\frac{f_A(1-f_A)}{n}}, f_A + 1,96 \sqrt{\frac{f_A(1-f_A)}{n}} \right]$, au niveau de confiance 95 % pour la proportion p_A dans la population A et on définit de même I_B pour la population B.

À RETENIR

La différence entre les deux fréquences observées f_A et f_B est considérée comme significative quand les intervalles de confiance à 95 % I_A et I_B sont disjoints.

On juge alors que les deux proportions p_A et p_B sont différentes (avec un petit risque d'erreur).

Dans le cas contraire, on juge que les deux proportions p_A et p_B sont égales (avec un petit risque d'erreur).

Campagne publicitaire

ÉNONCÉ

Une importante société de vente au détail de matériel informatique veut juger de l'impact d'une campagne publicitaire menée dans les médias pour une tablette numérique particulière.

Dans un échantillon, considéré comme prélevé au hasard et avec remise, de 200 ventes de tablettes effectuées avant la campagne publicitaire, on observe que 44 ventes concernent cette tablette.

Dans un nouvel échantillon de même taille et prélevé dans les mêmes conditions après la campagne publicitaire, on observe que 66 ventes concernent cette tablette.

- 1° Déterminer un intervalle de confiance, avec un niveau de confiance de 95 %, pour la proportion des ventes de cette tablette avant la campagne publicitaire.
- 2° Même question après la campagne publicitaire.
- 3° La différence entre les deux fréquences observées avant et après la campagne publicitaire est-elle significative ?
- 4° Le responsable de la campagne publicitaire affirme : « Les deux échantillons montrent que les ventes de cette tablette ont augmenté de 50 % ; donc la campagne publicitaire a été très efficace ». Qu'en pensez-vous ?

SOLUTION

1° L'intervalle de confiance, avec un niveau de confiance 95 %, pour la proportion p_A des ventes de cette tablette avant la campagne publicitaire est :

$$I_A = \left[f_A - 1,96 \sqrt{\frac{f_A(1-f_A)}{n}}, f_A + 1,96 \sqrt{\frac{f_A(1-f_A)}{n}} \right] \text{ où } f_A = \frac{44}{200} \text{ et } n = 200.$$

$$I_A = [0,163 ; 0,277].$$

2° De même $f_B = \frac{66}{200}$, $n = 200$ et $I_B = [0,265 ; 0,395]$.

3° Les deux intervalles $I_A = [0,163 ; 0,277]$ et $I_B = [0,265 ; 0,395]$ ne sont pas disjoints car $I_A \cap I_B = [0,265 ; 0,277]$.

Au seuil de 5 %, la différence entre les deux fréquences observées f_A et f_B avant et après la campagne publicitaire n'est donc pas significative.

$$4° \frac{f_B - f_A}{f_A} = \frac{0,33 - 0,22}{0,22} = 0,5 = 50 \text{ \%}.$$

Le responsable de la campagne a raison d'affirmer que **pour les deux échantillons** les ventes de cette tablette ont augmenté de 50 %.

Cependant, comme la différence entre f_A et f_B n'est pas significative, on juge que les deux proportions p_A et p_B sont égales (avec un petit risque d'erreur).

Le responsable de la campagne publicitaire ne peut donc pas affirmer, **sur la seule base de ces deux échantillons**, que la campagne publicitaire a été très efficace.