Probabilités et statistiques
Echantillonnage avec loi binomiale et loi normale

1 En classe de seconde : intervalle de fluctuation

Les données statistiques suivantes ont été relevées :
- en 2000, dans le village de Xicun, en Chine, il est né 20 enfants, parmi lesquels 16 garçons ;
- dans la réserve indienne d'Aamjiwnaag, située au Canada à proximité d'industries chimiques, il est né entre 1999 et 2003, 132 enfants dont 46 garçons.

Ces observations sont-elles le fruit du hasard ?

2 En classe de première : loi binomiale

Les services obstétriques d'une ville de province annoncent que sur 200 naissances, on a pu comptabiliser 116 garçons. On souhaite savoir à partir de quelles fréquences on pourra mettre en doute, au seuil de 5%, cette affirmation.

1. On fait l'hypothèse que l'on a bien 58% de chances d'accueillir un garçon. Par ailleurs, on suppose que la population de la ville est suffisamment grande pour considérer que les 200 naissances sont indépendantes les unes des autres. Montrer que dans ces conditions, la variable aléatoire X, correspondant au nombre de garçons, suit la loi binomiale de paramètres $n = 200$ et $p = 0,58$.
2. A l'aide d'un tableur, établir la loi de probabilité de X ainsi que la table des probabilités cumulées $P(X \leq k)$.
   (a) Déterminer les nombres a et b tels que :
   - a est le plus petit entier naturel tel que $P(X \leq a) > 0,025$;
   - b est le plus petit entier naturel tel que $P(X \leq b) \geq 0,975$.

   (b) Comparer l'intervalle de fluctuation à 95% $\left[\dfrac{a}{n} \, ; \, \dfrac{b}{n}\right]$ obtenu grâce à la loi binomiale avec l'intervalle $\left[p - \dfrac{1}{\sqrt{n}} \, ; \, p + \dfrac{1}{\sqrt{n}}\right]$. Essayer d'expliquer.
3. En réalité, on a comptabilisé, sur 200 naissances, 126 garçons. Peut-on accepter, au seuil de 5%, notre hypothèse de départ ?

utilisation sur tableur de LOI.BINOMIALE(k ;n ;p) et pour le cumul LOI.BINOMIALE(k ;n ;p ;VRAI) ou CRITERE.LOI.BINOMIALE(n ;p ;alpha)
simulation à l'aide d'un algorithme

Sur la calculatrice TI, menu $DISTR$ (touches $2ND$ puis VAR) :
- *binompdf* (n, p, k) renvoie la probabibilité $P(x = k)$ en suivant la loi binomiale de paramètres n et p ;
- *binomcdf* (n, p, k) renvoie la probabibilité $P(x \leq k)$ en suivant la loi binomiale de paramètres n et p.

3 En classe de terminale : loi normale

Les services obstétriques d'une ville de province annoncent que sur 200 naissances, on a pu comptabiliser 116 garçons. On souhaite savoir à partir de quelles fréquences on pourra mettre en doute, au seuil de 5%, cette affirmation.

1. On fait l'hypothèse que le nombre de naissances suit une loi normale de moyenne $\mu = 116$ et d'écart-type σ tel que $\sigma^2 = 48,72$.[1] Vérifier que la variable aléatoire $Z = \dfrac{X-\mu}{\sigma}$ suit la loi normale centrée réduite.
2. A l'aide d'un tableur, établir les probabilités $P(X \leq k)$ pour $k \in [0; 200]$.
   (a) Déterminer les nombres A et B tels que :
   - A est le plus petit entier naturel tel que $P(X \leq A) > 0,025$;
   - B est le plus petit entier naturel tel que $P(X \leq B) \geq 0,975$.

   (b) Comparer l'intervalle de fluctuation à 95% $\left[\dfrac{A}{n} ; \dfrac{B}{n}\right]$ ainsi obtenu avec celui obtenu avec la loi binomiale $\left[\dfrac{a}{n} ; \dfrac{b}{n}\right]$ et avec l'intervalle $\left[p - \dfrac{1}{\sqrt{n}} ; p + \dfrac{1}{\sqrt{n}}\right]$. Essayer d'expliquer.
3. En réalité, on a comptabilisé, sur 200 naissances, 126 garçons. Peut-on accepter, au seuil de 5%, notre hypothèse de départ ?
4. On souhaite calculer, grâce à la loi normale, l'intervalle de fluctuation obtenu en 2.(b).
   (a) Vérifier que, pout tout nombre k, $P(X \leq k) = P\left(Z \leq \dfrac{k-\mu}{\sigma}\right)$.

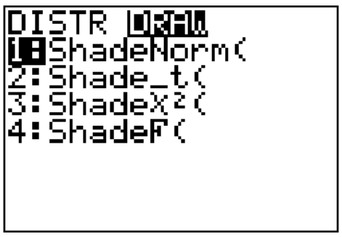
   (b) Vérifier que $P(-1,96 \leq Z \leq 1,96) = 0,95$.

   (c) En déduire que $P(-1,96\sigma + \mu \leq X \leq 1,96\sigma + \mu) = 0,95$.

   (d) Calculer $-1,96\sigma + \mu$ et $1,96\sigma + \mu$. Comparer ces valeurs avec A et B.
5. Quels résultats obtiendrait-on pour accepter, au seuil de 1% cette fois-ci, les valeurs annoncées par ces mêmes services obstétriques ?

Utilisation sur tableur de $LOI.NORMALE(x; \mu; \sigma; VRAI)$ qui donne $\displaystyle\int_{-\infty}^{x} \frac{e^{-(t-\mu)^2}}{\sqrt{2\pi\sigma^2}}\,dt$. Attention à bien préciser la valeur cumulative sur VRAI. Sur FAUX, elle renvoie la valeur de la fonction de densité en x.

Utilisation sur tableur de $LOI.NORMALE.INVERSE(valeurproba; \mu; \sigma)$ qui donne la valeur de x pour laquelle $\displaystyle\int_{-\infty}^{x} \frac{e^{-(t-\mu)^2}}{\sqrt{2\pi\sigma^2}}\,dt = valeurproba$.
simulation à l'aide d'un algorithme



Sur la calculatrice TI, menu $DISTR$ (touches $2ND$ puis $VARS$) :
- *normalpdf*(a, μ, σ) renvoie la valeur de la densité de probabilité en a de la loi $\mathcal{N}(\mu, \sigma)$;
- *normalcdf*(a, b, μ, σ) renvoie la probabilité $P(a \leq X \leq b)$ en suivant la loi $\mathcal{N}(\mu, \sigma)$ (prendre $a = -1E99$ pour obtenir $P(x \leq b)$) ;
- $InvNorm(p, \mu, \sigma)$ renvoie la valeur k pour laquelle $P(X \leq k) = p$ en suivant la loi $\mathcal{N}(\mu, \sigma)$;
- $ShadeNorm(a, b, \mu, \sigma)$ dessine la surface correspondant à $P(a \leq X \leq b)$ en suivant la loi $\mathcal{N}(\mu, \sigma)$.

1. La loi binomiale avait pour moyenne $n \times p = 200 \times 0,58 = 116$ et pour écart-type $\sqrt{np(1-p)} = \sqrt{200 \times 0,58 \times 0,42} = 6,9799135$.

4 En classe de terminale : loi binomiale et intervalle de fluctuation

Les services obstétriques d'une ville de province annoncent que sur 200 naissances, on a pu comptabiliser 116 garçons. On souhaite savoir à partir de quelles fréquences on pourra mettre en doute, au seuil de 5%, cette affirmation.

1. On fait l'hypothèse que l'on a bien 58% de chances d'accueillir un garçon. Par ailleurs, on suppose que la population de la ville est suffisamment grande pour considérer que les 200 naissances sont indépendantes les unes des autres.
   Montrer que dans ces conditions, la variable aléatoire X, correspondant au nombre de garçons, suit la loi binomiale de paramètres $n = 200$ et $p = 0,58$.
2. Préciser la moyenne μ et l'écart-type σ de X.

   On considère que la population est suffisammment importante pour que les résultats de cette loi binomiale soient "proches" de ceux de la loi normale de paramètres μ et σ^2.
3. Vérifier que, dans ces conditions, la variable aléatoire $Z = \dfrac{X - np}{\sqrt{np(1-p)}}$ suit la loi normale centrée réduite.
4. Vérifier que $P(-1,96 \leq Z \leq 1,96) = 0,95$.
5. En déduire que $P\left(-1,96\sqrt{np(1-p)} + np \leq X \leq 1,96\sqrt{np(1-p)} + np\right) = 0,95$.
6. Si l'on désigne par F_n la variable aléatoire donnant la fréquence $\dfrac{X}{n}$, vérifier alors que $F_n \in \left[p - 1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + 1,96\dfrac{\sqrt{p(1-p)}}{\sqrt{n}}\right]$ avec une probabilité de 95%.
7. En étudiant les variations de la fonction $t \to 1,96\sqrt{t(1-t)}$ sur l'intervalle $[0;1]$, montrer que :
$$\left[p - 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\ ;\ p + 1,96\frac{\sqrt{p(1-p)}}{\sqrt{n}}\right] \subset \left[\ p - \frac{1}{\sqrt{n}}\ ;\ p + \frac{1}{\sqrt{n}}\ \right].$$
8. Conclusions ?