Université de Picardie Jules Verne **Année 2014-2015**
IUT Techniques de commercialisation

TD de statistiques. Corrigé de l'évaluation 1.

Exercice 1. Une étude a montré que dans une promotion d'étudiants en première année à l'université, 3% démissionnent au bout d'un mois.

1) Sur une promotion de 160 étudiants TC1 quelle est la probabilité qu'exactement 4% démissionnent au bout d'un mois ?

2) Sur une promotion de 160 étudiants TC1 quelle est la probabilité qu'au moins 2% démissionnent au bout d'un mois ?

corrigé La population étudiée est celle des étudiants de première année à l'université, sur un grand nombre d'années. On suppose que son effectif est très grand $N = +\infty$.

La variable aléatoire étudiée est une variable qualitative. Etant donné un étudiant choisi aléatoirement, on se demande s'il a démissionné au bout d'un mois ou pas.

L'échantillon aléatoire est une promotion, de taille $n = 160$. C'est un grand échantillon $n \geq 30$.

La proportion théorique π d'étudiants ayant démissionné, sur l'ensemble de la population, vaut $\pi = 3\%$.

On note p la proportion (aléatoire) d'étudiants d'un échantillon de taille 160 ayant démissionné au bout d'un mois.

On sait que la variable aléatoire $Z = \frac{P-E(P)}{\sigma(P)}$ suit une loi normale $\mathcal{N}(0,1)$, avec $E(P) = \pi = 3\% = 0,03$ et $\sigma(P) = \sqrt{\pi(1-\pi)} \simeq 0,0135$.

1) On a alors

$$P(P = 4\%) = P\left(\frac{P - E(P)}{\sigma(P)} = \frac{0,04 - E(P)}{\sigma(P)}\right) = P\left(\frac{P - 0,03}{0,0135} = \frac{0,04 - 0,03}{0,0135}\right) = P(Z = 0,7407) = 0$$

car $Z = \frac{P-0,03}{0,0135}$ est une variable qui suit une loi normale (voir p.9 pour tout nombre z, on a $P(Z = z) = 0$).

La probabilité que la proportion d'étudiants qui démissionnent au bout d'un mois dans la promo soit exactement de 4% est nulle.

2) On a aussi

$$P(P > 2\%) = P\left(Z \geq \frac{0,02 - 0,03}{0,0135}\right) = P(Z \geq -0,7407) = P(Z \leq 0,7407) \simeq 0,7704\,.$$

Il y a donc une probabilité de 77% que au moins 2% des étudiants (en d'autres termes *plus de* 2%) démissionnent au bout d'un mois.

Exercice 2. Après la correction d'une épreuve d'examen nationale, on constate que les étudiants ont eu en moyenne 12 points avec une dispersion de plus ou moins 3 points.

1) Dans un groupe de TD de 36 étudiants, quelle est la probabilité que la moyenne soit au plus de 11 (c'est-à-dire *inférieure à* 11, ou *moins de* 11) ?

2) Dans un groupe de TD de 36 étudiants, quelle est la probabilité que la moyenne des notes soit comprise entre 12 et 14 ?

Corrigé On suppose que le nombre N d'étudiants est très grand. L'échantillon est un groupe de TD de $n = 36$ étudiants. C'est un grand échantillon. $n \geq 30$. La variable aléatoire étudiée est la note X d'un étudiant.

On note $\bar{X}$ la moyenne des notes sur un échantillon aléatoire de taille n.

La variable $Z = \frac{\bar{X}-E(\bar{X})}{\sigma(\bar{X}}$ suit une loi normale centrée réduite $\mathcal{N}(0,1)$, avec $E(\bar{X}) = \mu = 12$ et $\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}} = \frac{3}{\sqrt{36}} = 0,5$.

1) On a

$$P(\bar{X} \leq 11) = P\left(\frac{\bar{X} - E(\bar{X})}{\sigma(\bar{X})} \leq \frac{11 - 12}{0,5}\right) = P(Z \leq -2) = P(Z \geq 2) = 1 - P(Z \leq 2)\,.$$

On consulte la table pour avoir $P(Z \leq 2) = 0,9772$ d'où $P(\bar{X} \leq 11) = 1 - 0,9772 = 0,228 = 2,28\%$.

La probabilité que la moyenne des notes dans un groupe de TD de 36 étudiants soit inférieure à 11 vaut $2,28\%$.

2) De même, on calcule

$$\begin{aligned} P(12 \leq \bar{X} \leq 14) &= P\left(\frac{12-12}{0,5} \leq \frac{\bar{X}-E(\bar{X})}{\sigma(\bar{X})}) \leq \frac{14-12}{0,5}\right) \\ &= P(0 \leq Z \leq 4) = P(Z \leq 4) - P(Z \leq 0) = 0,9999 - 0,5 = 49,99\%\,. \end{aligned}$$

Ainsi, la probabilité que la moyenne des notes dans un groupe de TD de 36 étudiants soit comprise entre 12 et 14 vaut environ 50%.

Exercice 3. Une enqête d'un service marketing indique que sur 50 courriers publicitaires envoyés à des clients potentiels, 23 entraînent une visite en magasin de ces clients. Construire un intervalle de confiance à 95% pour la proportion de destinataires du courrier se rendant en magasin.

Corrigé On suppose que la population est grande $N = +\infty$. On considère un échantillon de clients de taille $n = 50$. La proportion p de clients de l'échantillon qui vont en magasin après avoir reçu un courrier vaut $p = 23/50 = 0,46$. On estime la proportion π théorique de clients qui vont en magasin après avoir reçu un courrier à l'aide d'un intervalle de confiance.

Au niveau de confiance 95%, cet intervalle de confiance se calcule ainsi :

$$IC_{95\%}(\pi) = \left[p - Z_{0,975}\sqrt{\frac{p(1-p)}{n}}\,; p - Z_{0,975}\sqrt{\frac{p(1-p)}{n}}\right],$$

On a $p = 0,46$, $Z_{0,975} = 1,96$, et $n = 50$ d'où

$$IC_{95\%}(\pi) = [0,46 - 0,1381; 0,46 + 0,1381] = [0,3219; 0,5981] \simeq [32\%; 60\%]\,.$$

Il y a donc 95% de chance que la proportion π de clients qui vont se déplacer au magasin après avoir reçu un courrier appartienne à l'intervalle $[32\%; 60\%]$.

Exercice 4. Une entreprise de menuiserie fabrique des pièces de bois pour la marque de meubles Mobelia. Une nouvelle chaine de production de planches circulaires pour des dessus de tables rondes vient d'être mise en service. Le responsable de la chaine de production veut vérifier que la chaine est bien calibrée pour répondre aux préconisations de la marque Mobelia.

1) Le responsable de la production commence par prélever un échantillon de 36 planches, sur lequel il mesure un diamètre moyen de 104cm, avec un écart-type des diamètres de l'échantillon de 2cm. Quelles estimations rapides du diamètre et de l'écart-type de l'ensemble des planches produites peut-il faire ?

2) Il s'interroge sur la validité de son estimation. Il souhaite alors donner un intervalle de confiance au niveau de confiance de 90% pour le diamètre moyen des planches. Quel sera cet intervalle de confiance ? On arrondira les bornes de l'intervalle à deux décimales après la virgule.

3.) Il trouve les estimations ci-dessus trop imprécises, et veut prélever un nouvel échantillon pour améliorer la précision de son estimation. Il souhaite estimer le diamètre moyen des planches produites avec d'une part un niveau de confiance plus élevé, et d'autre part une précision bien plus grande. Il voudrait un niveau de confiance de 95% et une précision de $0,1$cm. Quelle taille d'échantillon lui faudra-t-il considérer ?

Corrigé La population étudiée est celle des planches fabriquées, de taille supposée très grande $N = +\infty$. L'{echantillon est un grand échantillon $n = 36$.

La variable aléatoire X étudiée est le diamètre d'une planche. Sur l'échantillon, la moyenne des diamètres vaut $m = 104cm$, et l'écart-type $s = 2cm$.

1) L'estimation la plus rapide est l'estimation ponctuelle. On estime la valeur théorique μ du diamètre moyen des planches par l'estimation $\hat{\mu} = m = 104cm$.

On estime la valeur théorique σ de l'écart-type du diamètre des planches par l'estimation $\hat{\sigma} = s.\sqrt{\frac{n}{n-1}} \simeq 2,02cm$.

2) On veut trouver un intervalle de confiance au niveau de confiance de 90% pour la moyenne. On a $1 - \alpha = 90\% = 0,9$ donc $\alpha = 0,1$ et $\alpha/2 = 0,05$, donc $1 - \frac{\alpha}{2} = 0,95$. On cherche $Z_{1-\frac{\alpha}{2}} = Z_{0,95}$; c'est la valeur telle que si Z est une variable aléatoire suivant une loi normale centrée réduite $\mathcal{N}(0,1)$, alors $P(Z \leq Z_{0,95}) = 0,95$. On cherche cette valeur dans la table. Elle n'y est pas exactement, mais on peut écrire $Z_{0,95} \simeq 1,65$ (ou $1,645$).

Ensuite, comme l'échantillon est grand et l'écart-type théorique σ est inconnu, on a

$$IC_{90\%}(\mu) = [m - Z_{0,95}.\frac{s}{\sqrt{n-1}}; m + Z_{0,95}.\frac{s}{\sqrt{n-1}}]$$

. Avec $m = 104$, $Z_{0,95} = 1,645$, $s = 2$ et $n - 1 = 48$, on trouve

$$IC_{90\%}(\mu) = [104 - 0,4749; 104 + 0,4749] = [103,5251; 104,4749] \simeq [103,52; 104,48]\,.$$

Il y a donc 90% de chances pour que le diamètre moyen des planches soit compris entre $103,52cm$ et $104,48cm$.

3) On veut trouver la taille minimale n d'échantillon qu'il faut choisir pour avoir un intervalle de confiance $IC_{95\%}$ qui approche μ à la précision $c = 1cm$. Pour cela, on suppose $n \geq 30$, et il suffit ensuite de choisir n de sorte que

$$n \geq \left(\frac{Z_{0,975}.\sigma}{c}\right)^2 = (\frac{1,96 \times 2,02}{1})^2 \simeq 15,67.$$

Il faut donc prendre $n \geq 16$. Comme de toute façon l'échantillon doit être grand pour faire ce calcul, on prendra donc $n = 30$.