Estimation des paramètres à partir du pourcentage, et de la moyenne

1ère année Médecine 2016/2017

1- Introduction

Faire une estimation, c'est tenter de définir les paramètres d'une population à partir des paramètres observés sur un échantillon.

Lorsqu'on observe un paramètre sur un échantillon, on sait:

Que la valeur observée a fort peu de chance d'être exactement la valeur inconnue de la population

- Que cette valeur est néanmoins assez proche de la valeur inconnue si notre échantillon est représentatif
- Qu'en répétant l'échantillonnage, on trouverait d'autres valeurs, toutes assez proches les unes des autres.

2- Estimation d'une moyenne inconnue

- M est la moyenne de la variable quantitative sur un échantillon
- Problème: μ est l'estimation de la moyenne ce cette variable dans la population, donc

$$\mu = ?$$

a- Ecart type de la moyenne:

Il faut savoir que la M est elle-même une variable aléatoire. Elle varie selon les échantillons, elle suit une loi normale et on peut lui calculer son écart type, écart type de la moyenne Sm, à ne pas confondre avec l'écart type des valeurs de l'échantillon.

$$Sm = S/\sqrt{N}$$

- S : écart type des valeurs de l'échantillon
- N : taille de l'échantillon

b- Intervalle de confiance (IC)

- Le but: tenter d'estimer **μ**
- Pour cela il faut estimer un intervalle dans lequel la moyenne a la plus grande probabilité de se trouver
- On démontre que à 95% de chance cette moyenne **μ** se trouve dans un intervalle compris entre

M-1,96. Sm et **M+ 1,96. Sm**

On appelle cet intervalle, intervalle de confiance (IC) à 95% de la moyenne μ

L'IC à 95% s'exprime:

$$\mu = M \pm 1.96 . Sm$$

Avec

µ : la moyenne inconnue de la population

M : la moyenne calculée sur l'échantillon

Sm : l'écart type de la moyenne

A 95% de chance la µ se trouve dans cet intervalle, mais on peut dire en complément qu'à 5% µ peut être à l'extérieur de cet intervalle.

c- Exemple

- Dans une enquête sur la durée du sommeil des enfants de 2 à 3 ans effectuée sur un échantillon de 540 enfants d'un département français, on a trouvé une moyenne de ce temps par nuit de 11.7 heures. L'écart type est de 1.3 heures.
- **On veut reconnaitre la moyenne générale du temps de sommeil chez tous les enfants du département**
- > l'écart type de la moyenne $Sm = 1.3/\sqrt{540} = 0.056$ heures.
- > l'IC à 95% est de 11.7+/- (1,96. 0,056)= 11.7+/-0,11 heures
- > la moyenne du temps du sommeil est donc comprise entre 11,6 et 11,8 heures

3- Estimation d'un pourcentage inconnu

Lorsqu'on a observé un pourcentage sur un échantillon p, le problème est d'estimer le véritable pourcentage **P** inconnu de la population d'où est extrait l'échantillon.

a- Ecart type du pourcentage

Il faut savoir que p est elle-même une variable aléatoire. Elle varie selon les échantillons, elle suit une loi normale

et on peut lui calculer son écart type , écart type du pourcentage Sp.

$$Sp = \sqrt{[p(1-p)/N]}$$

- Sp : écart type du pourcentage
- p : le pourcentage calculé sur l'échantillon
- N : taille de l'échantillon

b- Intervalle de confiance (IC)

- Le but: tenter d'estimer **P**
- Pour cela il faut estimer un intervalle dans lequel la moyenne a la plus grande probabilité de se trouver
- On démontre que à 95% de chance ce pourcentage **P** se trouve dans un intervalle compris entre

p-1,96.Sp et **p+1,96.Sp**

On appelle cet intervalle, intervalle de confiance (IC) à 95% du pourcentage P

L'IC à 95% s'exprime:

$$P= p+/-1,96.Sp$$

Avec

Sp : écart type du pourcentage

p : le pourcentage calculé sur l'échantillon

A 95% de chance le P se trouve dans cet intervalle, mais on peut dire en complément que à 5% P peut être à l'extérieur de cet intervalle.

c- <u>Exemple</u>

- Dans une enquête sur la durée du sommeil des enfants de 2 à 3 ans effectuée sur un échantillon de 540 enfants d'un département français , on a trouvé 86 enfants présentant des troubles de sommeil. On veut connaitre la proportion des troubles du sommeil chez tous les enfants du département.

- La proportion d'enfants présentant des troubles du sommeil dans l'échantillon est de 86/540= 15.9%.

- L'écart type est = **Sp = √[p(1-p)/N]**

  = √[0,159(1- 0,159)/540] = 0,016

- L'IC à 95% est : 0,159+/-1,96. 0,016 = 0,159+/-0,031

- La proportion d'enfants présentant des troubles dans ce département est donc comprise entre 12,8% et 19,0%

4- Risque d'erreur consentie α

Le risque d'erreur α est un choix, mais ça n'empêche pas de choisir un autre risque bien évidement moindre, il faudrait alors remplacer la valeur de Zα (1.96), dans la table de Z(loi normale centrée réduite) par la valeur correspondant valeur du risque α, comme pour le risque α=2%, Zα= 2.33, pour α=1%, Zα=2,58 … .

Les formules pour calculer la moyenne ou le pourcentage se généralisent alors et deviennent:

Moyenne: μ= M+/- Zα.Sm

Pourcentage: P= p+/- Zα. Sp

Plus le risque d'erreur est petit, plus le risque de se tromper est moindre et plus l'IC s'élargit pour contenir plus de valeurs mais, plus on perd la précision sur l'estimation des paramètres.

5- Taille d'un échantillon

La précision d'une estimation dépend de:

1- Choix du risque d'erreur qui détermine $Z\alpha$

2- La taille N de l'échantillon qui intervient dans les formules déterminant l'écart type de la moyenne et du pourcentage. Dans ces deux formules, la taille est au dénominateur et donc:

- Plus la taille est grande
- Plus les écarts types ne sont petits
- Plus l'IC est resserré

Plus est grande la précision

Calcul de la taille de l'échantillon

a- A partir d'une moyenne

$$N= S^2(Z\alpha^2/i^2)$$

Où:

--pour α= 5%, $Z\alpha$= 1,96

--S^2 est la variance de la variable dans la population (elle n'est pas connue, donc on l'estime à partir d'études antérieures ou d'études pilotes.

--i est la précision désirée c'est la moitié de l'intervalle de confiance (plus elle est petite plus N est grand)

b- A partir d'un pourcentage

$$N=P(1-P) (Z\alpha^2/i^2)$$

Où:

--pour α= 5%, $Z\alpha$= 1,96

--P est le pourcentage de la variable dans la population(elle n'est pas connue, donc on l'estime à partir d'études antérieures ou d'études pilotes.

--i est la précision désirée c'est la moitié de l'intervalle de confiance (plus elle est petite plus N est grand)

Exemple

ON désire estimer la proportion des troubles du sommeil chez des enfants de 2 à 3 ans dans un département français. Des études effectuées dan d'autres régions ont montré que la proportion de ces troubles est d'environs 16%. On désire une précision de +/- 3% et on choisit un risque α de 5%

De combien sera notre échantillon N?

N= 0,16(1-0,16)($1,96^2$/$\mathbf{0,03^2}$)=**574**

Les commanditaires de cette étude jugent qu'un i de 2% sera meilleur

N=0,16(1-0,16)($1,96^2$/$\mathbf{0,02^2}$)=**1291**.
Pour gagner 1% de précision, la taille de l'échantillon double

6- Conclusion

Le calcul d'une moyenne ou d'un pourcentage observé sur un échantillon représentatif d'une population a pour but d'estimer la moyenne ou le pourcentage inconnu dans cette population.

On estime ces paramètres inconnus en calculant un intervalle de confiance, cet intervalle est composé de deux bornes entre lesquelles la valeur inconnue a la plus grande probabilité de se situer.

La le degré de confiance qu'on exige est déterminé dans la formule de calcul.

On se fixe en général un IC à 95%.

-Pour une moyenne M, l'IC à 95% = $M+/- 1.96(s/\sqrt{N})$

-Pour un pourcentage p, l'IC à 95% = $p+/-1,96 \sqrt{[p(1-p)/N]}$

- Cela signifie que le paramètre a 95% de chance d'être situé dans cette intervalle, cela signifie qu'il existe un risque d'erreur (α) de 5%.
- Si on désire une plus grande certitude d'encadrer la valeur du paramètre inconnu, on choisit un risque d'erreur α plus faible, l'IC sera plus large et la précision de de l'estimation sera moins bonne
- La largeur de l'IC dépend aussi de la taille de l'échantillon.
- Plus la taille est plus élevée, meilleure est la précision