4.3 ESTIMATION D'UNE PROPORTION

cours 24

<table>
<tr><th>$X \sim N(\mu, \sigma^2)$</th><th>$n \geq 30$</th><th>σ connue</th><th>Avec remise</th></tr>
<tr><td rowspan="2">oui</td><td>oui</td><td rowspan="3">oui</td><td rowspan="3">$\mu \in \left[\bar{x} - z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}, \bar{x} + z_{\frac{\alpha}{2}}\sqrt{\frac{\sigma^2}{n}}\right]$</td></tr>
<tr><td>non</td></tr>
<tr><td>non</td><td>oui</td></tr>
<tr><td>oui</td><td rowspan="2">oui</td><td rowspan="2">non</td><td rowspan="2">$\mu \in \left[\bar{x} - z_{\frac{\alpha}{2}}\sqrt{\frac{s^2}{n}}, \bar{x} + z_{\frac{\alpha}{2}}\sqrt{\frac{s^2}{n}}\right]$</td></tr>
<tr><td>non</td></tr>
<tr><td>oui</td><td>non</td><td>non</td><td>$\mu \in \left[\bar{x} - t_{\frac{\alpha}{2}}\sqrt{\frac{s^2}{n}}, \bar{x} + t_{\frac{\alpha}{2}}\sqrt{\frac{s^2}{n}}\right]$</td></tr>
</table>

sans remise

$$\mu \in \left[\bar{x} - z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}, \bar{x} + z_{\frac{\alpha}{2}} \sqrt{\frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)}\right]$$

$$\mu \in \left[\bar{x} - z_{\frac{\alpha}{2}} \sqrt{\frac{s^2}{n}\left(\frac{N-n}{N-1}\right)}, \bar{x} + z_{\frac{\alpha}{2}} \sqrt{\frac{s^2}{n}\left(\frac{N-n}{N-1}\right)}\right]$$

$$\mu \in \left[\bar{x} - t_{\frac{\alpha}{2}} \sqrt{\frac{s^2}{n}\left(\frac{N-n}{N-1}\right)}, \bar{x} + t_{\frac{\alpha}{2}} \sqrt{\frac{s^2}{n}\left(\frac{N-n}{N-1}\right)}\right]$$

Aujourd'hui, nous allons voir

- ✓ Distribution d'une proportion
- ✓ Estimation d'une proportion
- ✓ Intervalle de confiance sur une proportion
- ✓ Taille d'un échantillon

Pour calculer une moyenne d'une variable statistique,
il était sous-entendu qu'elle était quantitative.

On peut difficilement faire une moyenne d'une variable statistique qualitative.

On peut dans le cas d'une variable statistique ordinale attribuer un nombre à chacune des modalités et ensuite faire une moyenne.

Regardons que faire avec des variables statistiques qualitatives nominales.

Si X est une variable statistique qualitative nominale

ayant comme modalité

$\{\text{truc}, \text{much}, \text{machin chouette}\}$

On peut considérer la proportion de chacune de ces modalités dans la population.

$$\pi_{truc} = \frac{\text{nombre d'individus ayant truc}}{\text{population totale}}$$

$$p_{truc} = \frac{\text{nombre d'individus ayant truc dans l'échantillon}}{\text{taille de l'échantillon}}$$

Considérons une population de taille N.

Une proportion π de cette population, possède une certaine caractéristique

Si on pige au hasard avec remise n individus, la variable aléatoire

X : le nombre d'individus ayant la caractéristique

Alors

$$X \sim B(n, \pi)$$

$$E(X) = n\pi \qquad \mathrm{Var}(X) = n\pi(1-\pi)$$

$$X \sim B(n, \pi)$$

$$E(X) = n\pi \qquad \mathrm{Var}(X) = n\pi(1-\pi)$$

si $\quad n \geq 30 \qquad n\pi \geq 5 \qquad n(1-\pi) \geq 5$

alors on peut faire l'approximation

$X \sim N(n\pi, \sqrt{n\pi(1-\pi)})$ et donc $\dfrac{X - n\pi}{\sqrt{n\pi(1-\pi)}} \sim N(0,1)$

Considérons maintenant la variable aléatoire

$\bar{P}$: la proportion du nombre de succès de notre échantillon

$$X \sim N(n\pi, \sqrt{n\pi(1-\pi)})$$

$$\frac{X - n\pi}{\sqrt{n\pi(1-\pi)}} \sim N(0,1)$$

$\bar{P}$: la proportion du nombre de succès de notre échantillon

$$\bar{P} = \frac{X}{n}$$

$$E(\bar{P}) = E\left(\frac{X}{n}\right) = \frac{1}{n}E\left(X\right) = \frac{1}{n}n\pi = \pi$$

$$\mathrm{Var}\left(\bar{P}\right) = \mathrm{Var}\left(\frac{X}{n}\right) = \frac{1}{n^2}\mathrm{Var}\left(X\right) = \frac{1}{n^2}n\pi(1-\pi) = \frac{\pi(1-\pi)}{n}$$

$$\frac{\bar{P} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \sim N(0,1)$$

$$\frac{\bar{P}-\pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \sim N(0,1)$$

et si l'échantillon est fait sans remise, la variance est multipliée par le facteur

$$\frac{N-n}{N-1}$$

$$\frac{\bar{P}-\pi}{\sigma_{\bar{P}}} \sim N(0,1)$$

avec remise

$$\sigma_{\bar{P}} = \sqrt{\frac{\pi(1-\pi)}{n}}$$

sans remise

$$\sigma_{\bar{P}} = \sqrt{\frac{\pi(1-\pi)}{n}\left(\frac{N-n}{N-1}\right)}$$

Remarque

Puisque nous faisons une approximation d’une loi binomiale avec une loi normale, il ne faut pas oublier de faire la correction de continuité.

Faites les exercices suivants

#4.6, 4.7, 4.8, 4.9

$$\text{Puisque} \quad E(\bar{P}) = \pi$$

On peut donc conclure que $\bar{P}$ est un estimateur sans biais de π

$$Z = \frac{\bar{P} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \sim N(0,1)$$

$$1 - \alpha = P(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}})$$

$$= P\left(-z_{\frac{\alpha}{2}} \leq \frac{\bar{P} - \pi}{\sqrt{\frac{\pi(1-\pi)}{n}}} \leq z_{\frac{\alpha}{2}}\right)$$

$$= P\left(-z_{\frac{\alpha}{2}}\sqrt{\frac{\pi(1-\pi)}{n}} \leq \bar{P} - \pi \leq z_{\frac{\alpha}{2}}\sqrt{\frac{\pi(1-\pi)}{n}}\right)$$

$$1-\alpha = P(-z_{\frac{\alpha}{2}} \leq Z \leq z_{\frac{\alpha}{2}})$$

$$= P\left(-z_{\frac{\alpha}{2}}\sqrt{\frac{\pi(1-\pi)}{n}} \leq \bar{P} - \pi \leq z_{\frac{\alpha}{2}}\sqrt{\frac{\pi(1-\pi)}{n}}\right)$$

$$= P\left(\pi - z_{\frac{\alpha}{2}}\sqrt{\frac{\pi(1-\pi)}{n}} \leq \bar{P} \leq \pi + z_{\frac{\alpha}{2}}\sqrt{\frac{\pi(1-\pi)}{n}}\right)$$

$$\pi \in \left[\bar{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\pi(1-\pi)}{n}}, \bar{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\pi(1-\pi)}{n}}\right]$$

Mais là on a un petit problème, car on cherche à estimer π

mais on en a de besoin

$$\pi \in \left[\bar{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\pi(1-\pi)}{n}}, \bar{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\pi(1-\pi)}{n}}\right]$$

on n'a donc pas le choix d'estimer l'écart type aussi

$$\sigma_{\bar{P}} = \sqrt{\frac{\pi(1-\pi)}{n}} \qquad s_{\bar{P}} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n-1}}$$

$$\sigma_{\bar{P}} = \sqrt{\frac{\pi(1-\pi)}{n}\left(\frac{N-n}{N-1}\right)} \qquad s_{\bar{P}} = \sqrt{\frac{\bar{p}(1-\bar{p})}{n-1}\left(\frac{N-n}{N-1}\right)}$$

$$\pi \in \left[\bar{p} - z_{\frac{\alpha}{2}}\sqrt{\frac{\pi(1-\pi)}{n}}, \bar{p} + z_{\frac{\alpha}{2}}\sqrt{\frac{\pi(1-\pi)}{n}}\right]$$

on va donc plutôt utiliser

$$\pi \in \left[\bar{p} - z_{\frac{\alpha}{2}} \boxed{s_{\bar{P}}}, \bar{p} + z_{\frac{\alpha}{2}} \boxed{s_{\bar{P}}}\right]$$

Avec remise

$$s_{\bar{P}} = \boxed{\sqrt{\frac{\bar{p}(1-\bar{p})}{n-1}}}$$

Sans remise

$$s_{\bar{P}} = \boxed{\sqrt{\frac{\bar{p}(1-\bar{p})}{n-1}\left(\frac{N-n}{N-1}\right)}}$$

Faites les exercices suivants

#4.20 à 4,23

Lorsqu’on estime une moyenne ou une proportion, quelle est la taille d’échantillon avons-nous de besoin.

Naturellement, si possible, il faut au moins qu’on ait un échantillon de taille 30.

Mais jusqu’où faut-il aller?

50? 100? 250? 1000?

Et bien ça dépend de la marge d’erreur et du niveau de confiance avec lequel on est à l’aise.

Pour une moyenne la marge d'erreur est

$$EEM = c_{\frac{\alpha}{2}}\sqrt{\frac{b^2}{n}}$$

avec $c = z, t$ $b = \sigma, s$

en fonction du contexte.

$$EEM^2 = c^2_{\frac{\alpha}{2}}\frac{b^2}{n} \qquad n = \left(\frac{c_{\frac{\alpha}{2}}b}{EEM}\right)^2$$

Si on connait σ, il n'y a pas de problème

Mais si on n'a pas la taille de l'échantillon comment on trouve s?

Dans la pratique, on commence par prendre un échantillon de taille 30, on calcul le s avec cet échantillon pour ensuite déterminer la bonne taille de l'échantillon à prendre.

Pour une proportion la marge d'erreur est

$$EEM = z_{\frac{\alpha}{2}} \sqrt{\frac{\bar{p}(1-\bar{p})}{n-1}}$$

$$\boxed{EEM^2} = z^2_{\frac{\alpha}{2}} \frac{\bar{p}(1-\bar{p})}{\boxed{n-1}}$$

$$\boxed{n-1} = z^2_{\frac{\alpha}{2}} \frac{\bar{p}(1-\bar{p})}{\boxed{EEM^2}}$$

$$n = \frac{z^2_{\frac{\alpha}{2}} \bar{p}(1-\bar{p})}{EEM^2} + 1$$

Encore une fois, on a le problème qu'on ne connait pas $\bar{p}$.

On peut soit commencer avec un échantillon de 30.

Ou bien prendre le pire des scénarios; $\bar{p} = \dfrac{1}{2}$

Pour une proportion la marge d'erreur est

$$n = \frac{z^2_{\frac{\alpha}{2}} \bar{p}(1-\bar{p})}{EEM^2} + 1$$

Ou bien prendre le pire des scénario; $\bar{p} = \dfrac{1}{2}$

et dans ce cas

$$n = \frac{z^2_{\frac{\alpha}{2}}}{4EEM^2} + 1$$

Faites les exercices suivants

#4.18, 4.19, 4,24

Devoir: 4.8, 4.9 et 4.20 à 4.29