Estimation par intervalle de confiance

1/Estimation par intervalle de confiance d' une proportion :

Soit A le caractère observé dans la population ; X la (V.A) qui observe le caractère A

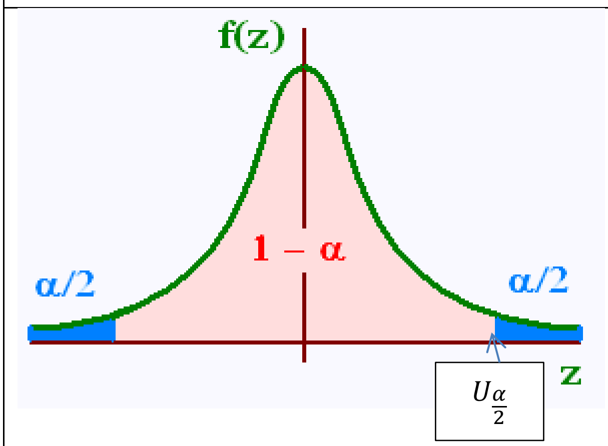
Pour un individu deux issus sont possibles

succès(avoir la caractère A)	$P(A) = p$	X suit une loi de Bernoulli
échec (ne pas avoir la caractère A)	$P(\bar{A}) = 1 - p$	

Pour n≥ 2 , X suit une loi de binomiale B(n,p)

Théorème

Si $n > 30, np > 5 \; et \; n(1-p) > $ 5 alors loi de binomiale B(n,p) peut être approximée par une loi normale $N(np \,; \sqrt{np(1-p)})$

intervalle de confiance d'une proportion	
condition d'application	IC
$n > 30, np > 5 \; et \; n(1-p) > 5$ échantillon aléatoire simple	$IC p_{pop} = \left[p - U_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} ; p + U_{\frac{\alpha}{2}} \sqrt{\frac{p(1-p)}{n}} \right]$
	La table de la loi normale N(0,1)

niveau de confiance	niveaude risque ∝	$F(U_{\frac{\alpha}{2}}) = 1 - \frac{\propto}{2}$	$U_{\frac{\alpha}{2}}$
90%	10%	0.950	1.645
95%	5%	0.975	1.96
98%	2%	0.990	2.325
99%	1%	0.995	2.575

Exemple Gregor Mendel a mené ses fameuses expériences génétiques avec des pois, un des échantillons des croisements a été obtenu en croisant **des pois à gousses vertes** et **des pois à gousses jaunes**. Cette lignée comportait **580 pois**. Parmi ces pois, **428 avait des gousses vertes** et **152 gousses jaunes**.

A partir de sa théorie des gènes, Mendel s'attendait à ce que **25%** des pois aient **des gousses jaunes**. Le pourcentage **de gousses jaunes est de 26.2%.**

a. Trouver la marge d'erreur qui correspond à un intervalle de confiance à 95%

b. Trouver l'intervalle de confiance à 95% de p de la population

c. A partir de ces résultats, que pouvons- nous conclure sur la théorie de Mendel qui déclare que le pourcentage de pois à gousse jaune devrait être de 25%?

Solution

On a: **n$\hat{p}$=152 $\geq$ 5, n$\hat{q}$=428$\geq$ 5**

a. **E=1.96** $\sqrt{\frac{0.262*0.738}{580}}$ **= 0.036**

b. **IC$_p$=[0.226, 0.298]** cet intervalle est décrit comme suit :

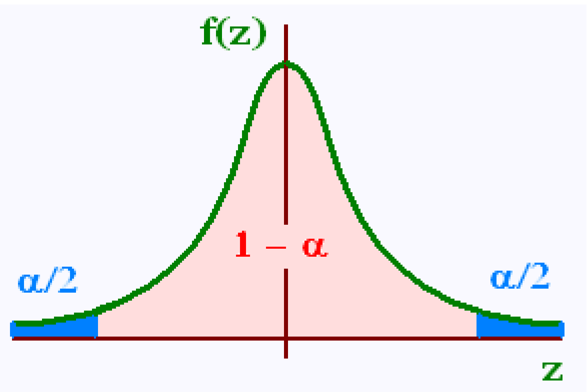
Le pourcentage de pois à gousse jaune est estimé à 26.2% avec une marge d'erreur de plus ou moins 3.6%.

c. A partir de ces résultats nous sommes sûr à 95% que les limites 22.6% et 29.8% contiennent le vrai pourcentage de pois à gousses jaunes.

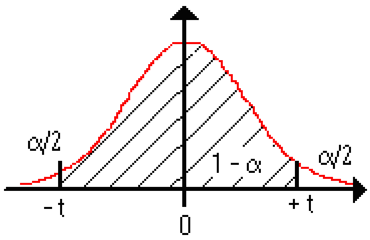
Le vrai pourcentage peut être vraisemblablement n'importe quelle valeur entre ces deux limites. Comme cet intervalle contient la valeur 25%, la valeur de Mendel ne peut pas être considéré comme fausse.

2/Estimation par intervalle de confiance d'une moyenne

Soit $X\ suit\ N(\mu,\sigma^2)$

<table>
<tr><th>cas de variance connue</th><th>cas de variance inconnue</th></tr>
<tr>
<td>$X\ suit\ N(\mu,\sigma^2)$ ou N>30
$\mu \in \left[m - U_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}; m - U_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}\right]$
m : moyenne estimée
Uα/2 est lue sur la table normale
.

</td>
<td>$X\ suit\ N(\mu,\sigma^2)$ ou N>30 ou N$\leq$ 30
échantillon aléatoire simple
$\mu \in \left[m - t_{\frac{\alpha}{2};v}\frac{S_{est}}{\sqrt{n}}; m + t_{\frac{\alpha}{2};v}\frac{S_{est}}{\sqrt{n}}\right]$
m : moyenne estimée.
σ : écart-type de la population or ici il est inconnu on va le remplacer par l' écart-type estiméS_{est}
$t_{\frac{\alpha}{2},v}$: une valeur lue sur la table Student
un d.d.l (v=n-1)
$\mu \in \left[m - t_{\frac{\alpha}{2};v}\frac{S_{est}}{\sqrt{n}}; m + t_{\frac{\alpha}{2};v}\frac{S_{est}}{\sqrt{n}}\right]$</td>
</tr>
</table>

niveau de confiance	niveaude risque $\propto$	$F(U_{\frac{\alpha}{2}}) = 1 - \frac{\propto}{2}$	$U_{\frac{\alpha}{2}}$
90%	10%	0.950	1.645
95%	5%	0.975	1.96
98%	2%	0.990	2.325
99%	1%	0.995	2.575



comme l'intervalle de confiance et aussi bilatérale donc on lit directement de la table

$$\propto_{risque} = \propto_{table}$$

pour $\alpha = 0{,}05$ et v=24

$\downarrow v/\alpha \rightarrow$	0.9	0.8	0.7	0.6	….	0.05	.02	0.01
1	0.1584	0.3249	0.5095	0.7265	1	12.706	31.821	63.656
…	…			…				
24	0.127	0.2562	0.39	0.5314	0.6848	2.0639	2.4922	0.127

Exemple 1 « Température du corps humain ».

Pour un échantillon des températures corporelles, on a n=106 et $\bar{x}$=36.78°C.

Supposer que l'échantillon est un échantillon aléatoire simple et que σ est connue et vaut $0.34^{\circ}C$.

➢ Donner l'intervalle de confiance à 95% de la température corporelle.

Il faut d'abord vérifier les conditions requises pour estimer µ

On a la normalité de la population et n>30.

n>30 est suffisante même si on n'a pas la normalité.

$$E = t_{\frac{\alpha}{2}} * \frac{\sigma}{\sqrt{n}} = 1.96 \frac{0.34}{\sqrt{106}} = 0.06$$

IC_{μ} =[36.72, 36.84]

Exemple 2 Considérons l'exemple de la température corporelle mais avec l'écart type estimé s=0.34.

Les conditions requises pour l'estimation de µ sont satisfaites.

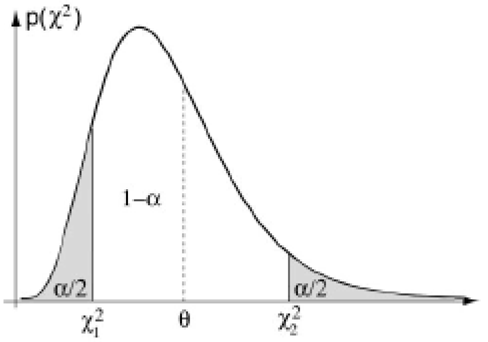
(n étant >30 , on n'a pas besoin de vérifier la normalité) pour α=0.05, n-1=105 on a $t_{\alpha/2,n-1}$=1.984

$$E = t_{\frac{\alpha}{2}} * \frac{s}{\sqrt{n}} = 1.984\frac{0.34}{\sqrt{106}} = 0.06$$

IC_{μ}= [36.72, 36.84]

Interprétation à partir des résultats de l'échantillon on peut être sûr à 95% que l'intervalle [36.72; 36.84] contient effectivement la vraie valeur de la moyenne de la population

3/ Estimation par intervalle de confiance d'une variance

intervalle de confiance d'une variance σ^2	
conditions d'application	IC
* l'échantillon est aléatoire simple * la population doit avoir une distribution normale (même si n>30).	$IC_{\sigma^2} = \left[\frac{(n-1)S_{est}^2}{\chi_1^2} ; \frac{(n-1)S_{est}^2}{\chi_2^2}\right]$ **un d.d.l (v=n-1)**
	$F(\chi_1^2) = 1 - \frac{\propto}{2}; \ F(\chi_2^2) = \frac{\propto}{2}$ si N est assez grand et ne se trouve pas dans la table du χ^2 ; on fait une approximation par laloi normale théorème si X suit une loi du χ^2 $si\ le\ d.d.l > 30\ alors\ la\ (V.A)U = \sqrt{2X^2} - \sqrt{2v-1}$ suit une loi normale N(0 ,1)

Exercice dans un échantillon de taille n=20 on a pu établir une estimation de la variance d'une (V.A) X

$$S_{est}^2 = 25.3$$

Donner un IC de la variance au risque 5%.Refaire l'exercice pour N=45

Exemple : « Températures corporelles »

On liste 106 températures corporelles prises par des chercheurs. Supposer que c'est un échantillon simple et utiliser les caractéristiques suivantes pour construire un **intervalle de confiance à 95%** de **l'écart type** pour les températures corporelles de l'ensemble de la population.

a. D'après la représentation graphique (histogramme) de l'échantillon, **Il n'y a pas de valeurs extrêmes** et les **données semblent suivre une loi normale**.

$\bar{x}$ **= 36.78°C,** $s = 0.34$ **°C,** $n = 106$

Solution

1. La condition de normalité est satisfaite.
2. **n=106, le ddl=106-1=105 .** Si ce ddl n'est pas dans la table, nous prendrons la valeur la plus proche. Par exemple **ddl=100.**
3. Pour un niveau de confiance 95%, on se réfère aux valeurs 0.975 et 0.025 comme en- têtes de colonnes, On a pour **ddl=105** :$\chi^2{}_1 =$ **78,54**, $\chi^2{}_2$=**135,2**
4. Pour **ddl=100** : $\chi^2{}_1 =$ 74.22, $\chi^2{}_2$=129.561
5. $IC_{\sigma^2} = \left[\frac{(n-1)s^2}{\chi^2{}_2}, \frac{(n-1)s^2}{\chi^2{}_1}\right]$
6. $IC_{\sigma^2} = \left[\frac{(106-1)*(0.34)^2}{135.2}, \frac{(106-1)*(0.34)^2}{78.54}\right]$
7. $\boldsymbol{IC_{\sigma^2}}$ =**[0.0898, 0.155]**
8. Si on prend la racine carré **IC $_{\sigma}$=[0.3, 0.39]**