L3 - Economie & Gestion

Statistique Inférentielle

Corrigé du Contrôle Continu N°1
Durée 1h30 - Année universitaire 2023 - 2024

Exercice 1 : **(Barème de notation : a) 2 pts b) 1.5 pt c) 1.5 pt d) 2.5 pts e) 2.5 pts f) 2.5 pts = 12.5 pts)**

a) Intervalle de confiance de la moyenne des températures en France-2022:

Données : températures de l'année 2022 en France :
$n_{22} = 20$; $\bar{x}_{22} = 14.51$; $s^2_{22} = 2.5^2$; $s^{*2}_{22} = \frac{n_{22}}{(n_{22}-1)} \times s^2_{22} = 6.579 = 2.565^2$.

Conditions : $n_{22} = 20$, échantillon de petite taille prélevé d'une population où la température est supposée suivre une loi normale de variance inconnue.

Statistique de test : $\frac{(\overline{X}_{n_{22}} - m_{22})}{\frac{s^*_{22}}{\sqrt{n_{22}}}} \rightarrow T_{n_{22}-1=19\,d.d.l.}$

Valeur critique de Student à $\nu = n_{22} - 1 = 19\,d.d.l.$: $t_{\alpha=2.5\%} = \pm 2.093$.

Marge d'erreur dans l'estimation de la production journalière moyenne m_{22} :

$$E = t_{2.5\%} \frac{s^*_{22}}{\sqrt{n_{22}}} = t_{2.5\%} \frac{s_{22}}{\sqrt{n_{22}-1}} = 2.093 \frac{2.5}{\sqrt{19}} = 1.20 \text{ degré Celsius.}$$

Estimation ponctuelle de la température moyenne en France-2022 : $\bar{x}_{22} = 14.51$ degrés Celsus.

Intervalle de confiance de niveau 95% de $m_{22} \in [13.31 \ ; \ 15.71]$

Conclusion : Puisque $15 \in [13.31 \ ; \ 15.71]$, on peut donc conclure avec un risque d'erreur $\alpha = 5\%$, que la température moyenne en France en 2022 n'est pas significativement différente de 15 degrés Celsus.

b) Taille d'échantillon requise pour estimer la température moyenne en France-2022 avec une précision au plus égale à $0.4°C$:

Risque d'erreur : $\alpha = 5\%$; $t_{\alpha=2.5\%}$ de la table de Student à $\nu = n^*_{22} - 1$ d.d.l.

On cherche n^*_{22}, la taille d'échantillon requise de telle sorte que la marge d'erreur :

$$E^* = t_{2.5\%} \frac{s^*_{22}}{\sqrt{n^*_{22}}} \leq 0.4 \ \Rightarrow E^* = u_{2.5\%} \frac{s^*_{22}}{\sqrt{n^*_{22}}} \leq 0.4$$

Approximation de la loi de Student $T_{\nu=n^*_{22}-1\,d.d.l.}$ par une loi normale N(0,1). la taille d'échantillon n^*_{22} sera grande (> 30), pour $\alpha = 5\%$, le fractile de la loi de Student $t_{2.5\%} \simeq u_{2.5\%} = 1.96$.

$$\Rightarrow \ n^*_{22} \geq (\frac{u_{\frac{\alpha}{2}} s^*_{22}}{E^*})^2 = (\frac{1.96 * 2.565}{0.4})^2 \ \Rightarrow \ n^*_{22} \geq 157.97 \ \Rightarrow \ n^*_{22} \geq 158 \text{ relevés de la température en 2022.}$$

c) Niveau de confiance de l'intervalle $[13.07°C \, ; \, 15.73°C]$ bilatéral symétrique de la température moyenne en France-2023 :

Données : températures de l'année 2023 en France : $n_{23} = 25$; $\bar{x}_{n_{23}} = 14.40$; $s^2_{n_{23}} = 3^2$; $s^{*2}_{n_{23}} = \frac{n_{23}}{(n_{23}-1)} \times s^2_{n_{23}} = 9.375 = 3.062^2$.

Marge d'erreur : $E = \frac{(15.73-13.07)}{2} = 1.33 = t_{\frac{\alpha}{2}} \frac{s^*_{n_{23}}}{\sqrt{n_{23}}} \Rightarrow t_{\frac{\alpha}{2}} = \frac{E\times\sqrt{n_{23}}}{s^*_{n_{23}}} = \frac{1.33\times 5}{3.062} = 2.1715$

cf. table Student à 24 d.d.l. , $P(\overline{X}_{n_{23}} \leq 2.1715) = F(2.172) = 0.98$

$\Rightarrow\ 1-\frac{\alpha}{2} = 98\% \ \Rightarrow\ \frac{\alpha}{2} = 2\% \Rightarrow\ \alpha = 4\% \ \Rightarrow\ 1-\alpha = 96\%.$

On attribue à l'intervalle $[13.07°C\ ;\ 15.73°C]$ de la température moyenne en France 2023, un niveau de confiance $1-\alpha = 96\%$.

d) Intervalle de confiance de la variance des températures en France-2023:

Année 2023 : $n_{23} = 25$; $\bar{x}_{n_{23}} = 14.40$; $s^2_{n_{23}} = 3^2$; $s^{*2}_{n_{23}} = \frac{n_{23}}{(n_{23}-1)} \times s^2_{n_{23}} = 9.375 = 3.062^2$.

Conditions d'application du test : échantillon de petite taille $n_{23} = 25$ prélevé d'une population où la température est supposée suivre une loi normale de variance inconnue.

Seuil de signification : $\alpha = 5\%$.

Niveau de confiance : $1-\alpha = 0.95 \ \Rightarrow\ \frac{\alpha}{2} = 0.025$ et $1-\frac{\alpha}{2} = 0.975$.

La statistique de test : $\frac{(n_{23}-1)S^{*2}_{n_{23}}}{\sigma^2_{n_{23}}} \ \rightarrow\ \chi^2_{(n_{23}-1=24\ d.d.l.)}$

Valeurs tabulées du khi-deux à 24 degrés de liberté (cf. table du khi-deux) :

$k_1 = \chi^2_{0.975\ ;\ 24} = 12.401$ et $k_2 = \chi^2_{0.025\ ;\ 24} = 39.364$

On en déduit alors l'intervalle de confiance suivant pour la variance $\sigma^2_{n_{23}}$:

$$\frac{(n_{23}-1)s^{*2}_{n_{23}}}{k_2} \leq \sigma^2_{n_{23}} \leq \frac{(n_{23}-1)s^{*2}_{n_{23}}}{k_1} \ \Rightarrow\ \frac{24\times 9.375}{39.364} \leq \sigma^2_{n_{23}} \leq \frac{24\times 9.375}{12.401} \ \Rightarrow\ 5.72 \leq \sigma^2_{n_{23}} \leq 18.14.$$

L'intervalle de confiance de niveau $1-\alpha = 95\%$ de la variance des températures en France en 2023 : $\sigma^2_{n_{23}} \in [5.72\ ;\ 18.14]$.

On en déduit l'intervalle de confiance de niveau $1-\alpha = 95\%$ de l'écart-type des températures en France en 2023 : $\sigma_{n_{23}} \in [2.39\ ;\ 4.26]$.

4.5 °C $\notin [2.39\ ;\ 4.26]$, on peut donc conclure que l'écart-type des températures en France en 2023 est significativement différent de 4.5 °C.

e) Étude de la variabilité des températures - Comparaison de variances :

Conditions d'application du test : grands échantillons $n_{22} = 20$ et $n_{23} = 25$ (échantillons indépendants provenant de deux populations normales de variances inconnues.

Les données :

Année 2022 : $n_{22} = 20$; $\bar{x}_{n_{22}} = 14.51$; $s^2_{n_{22}} = 2.5^2$; $s^{*2}_{n_{22}} = \frac{n_{22}}{(n_{22}-1)} \times s^2_{n_{22}} = 6.579 = 2.565^2$.

Année 2023 : $n_{23} = 25$; $\bar{x}_{n_{23}} = 14.40$; $s^2_{n_{23}} = 3^2$; $s^{*2}_{n_{23}} = \frac{n_{23}}{(n_{23}-1)} \times s^2_{n_{23}} = 9.375 = 3.062^2$.

Solution 1 - Intervalle de confiance du rapport de deux variances :

Conditions d'application du test : petits échantillons indépendants provenant de deux populations normales de variances inconnues $\sigma^2_{n_{22}}$ et $\sigma^2_{n_{23}}$.

La statistique de test : $\frac{\sigma^2_{n_{23}} s^{*2}_{n_{22}}}{\sigma^2_{n_{22}} s^{*2}_{n_{23}}} \ \rightarrow\ F_{(n_{22}-1=19\ ;\ n_{23}-1=24)}$

Les valeurs critiques (cf. table de Fisher) :

$f_2 = f_{(2.5\%\ ,\ 19\ ,\ 24)} = 2.345$; $f_1 = f_{(97.5\%\ ,\ 19\ ,\ 24)} = \frac{1}{f_{(2.5\%\ ,\ 24\ ,\ 19)}} = \frac{1}{2.452} = 0.408.$

$$f_1 < \frac{\sigma^2_{n_{23}} s^{*2}_{n_{22}}}{\sigma^2_{n_{22}} s^{*2}_{n_{23}}} < f_2 \;\Leftrightarrow\; f_1 \frac{s^{*2}_{n_{23}}}{s^{*2}_{n_{22}}} < \frac{\sigma^2_{n_{23}}}{\sigma^2_{n_{22}}} < f_2 \frac{s^{*2}_{n_{23}}}{s^{*2}_{n_{22}}} \;\Leftrightarrow\; 0.581 < \frac{\sigma^2_{n_{23}}}{\sigma^2_{n_{22}}} < 3.342$$

$I.C._{95\%}$: Intervalle de confiance de niveau 95% de $\frac{\sigma^2_{n_{23}}}{\sigma^2_{n_{22}}} \in [0.581\ ;\ 3.342]$.

Conclusion : $1 \in I.C._{95\%}$, on peut donc conclure avec un risque d'erreur $\alpha = 5\%$, qu'il n'y a pas de différence significative entre les variances des températures. On peut les supposer égales $\sigma^2_{n_{22}} = \sigma^2_{n_{23}}$.

Solution 2 - Test bilatéral symétrique du rapport de deux variances :

Les données :

Année 2022 : $n_{22} = 20$; $\bar{x}_{n_{22}} = 14.51$; $s^2_{n_{22}} = 2.5^2$; $s^{*2}_{n_{22}} = \frac{n_{22}}{(n_{22}-1)} \times s^2_{n_{22}} = 6.579 = 2.565^2$.

Année 2023 : $n_{23} = 25$; $\bar{x}_{n_{23}} = 14.40$; $s^2_{n_{23}} = 3^2$; $s^{*2}_{n_{23}} = \frac{n_{23}}{(n_{23}-1)} \times s^2_{n_{23}} = 9.375 = 3.062^2$.

Hypothèses statistiques :

$$\begin{cases} H_0 : \frac{\sigma^2_{n_{22}}}{\sigma^2_{n_{23}}} = 1 \;\Leftrightarrow\; \sigma^2_{n_{23}} = \sigma^2_{n_{22}} \quad \text{égalité des variances} \\ H_1 : \frac{\sigma^2_{n_{22}}}{\sigma^2_{n_{23}}} \neq 1 \;\Leftrightarrow\; \sigma^2_{n_{23}} \neq \sigma^2_{n_{22}}. \end{cases}$$

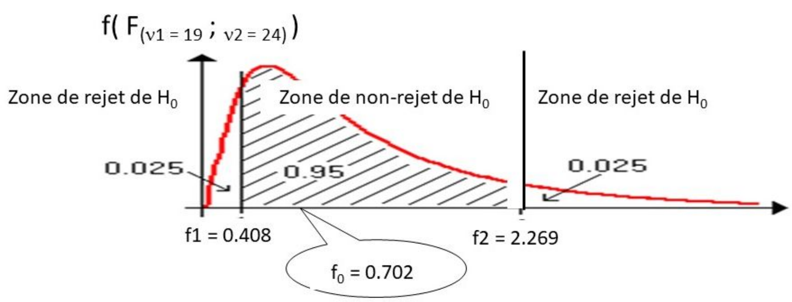
La statistique de test : $\frac{\sigma^2_{n_{23}} s^{*2}_{n_{22}}}{\sigma^2_{22} s^{*2}_{n_{23}}} \rightarrow F_{(n_{22}-1=19\ ;\ n_{23}-1=24)}$

Seuil de signification : $\alpha = 5\%$. Les valeurs critiques (cf. table de Fisher) :

$f_2 = f_{(2.5\%\ ,\ 19\ ,\ 24)} = 2.269$; $f_1 = f_{(97.5\%\ ,\ 19\ ,\ 24)} = \frac{1}{f_{(2.5\%\ ,\ 24\ ,\ 19)}} = \frac{1}{2.452} = 0.408$.

Calcul de la valeur de la statistique de test sous l'hypothèse nulle d'égalité des variances $H_0 : \sigma^2_{n_{22}} = \sigma^2_{n_{23}}$:

$$f_0 = \frac{s^{*2}_{n_{22}}}{s^{*2}_{n_{23}}} = \frac{6.579}{9.375} = 0.702$$



Règle de décision et conclusion : vu que $f_1 = 0.441 \leq f_0 = 0.702 \leq f_2 = 2.269$ alors non-rejet de l'hypothèse nulle H_0. On ne peut rejeter l'hypothèse nulle H_0. On peut donc conclure avec un seuil de signification $\alpha = 5\%$, qu'il n'y a pas de différence significative entre les variances ; on peut alors les supposer comme égales $\sigma^2_{n_{22}} \approx \sigma^2_{n_{23}}$.

f) Comparaison de deux moyennes :

Conditions d'application du test : petits échantillons $n_{22} = 20$ et $n_{23} = 25$ (échantillons indépendants provenant de deux populations normales de variances inconnues, testées égales à la question e) : $\sigma^2_{22} \approx \sigma^2_{23}$.

Les données :

Année 2022 : $n_{22} = 20$; $\bar{x}_{n_{22}} = 14.51$; $s^2_{n_{22}} = 2.5^2$; $s^{*2}_{n_{22}} = \frac{n_{22}}{(n_{22}-1)} \times s^2_{n_{22}} = 6.579 = 2.565^2$.

Année 2023 : $n_{23} = 25$; $\bar{x}_{n_{23}} = 14.40$; $s^2_{n_{23}} = 3^2$; $s^{*2}_{n_{23}} = \frac{n_{23}}{(n_{23}-1)} \times s^2_{n_{23}} = 9.375 = 3.062^2$.

Solution 1 - Intervalle de confiance de la différence de deux moyennes :

La statistique de test de la différence de deux moyennes variances égales :

$$\frac{(\overline{X}_{n_{22}}-\overline{X}_{n_{23}})-(m_{22}-m_{23})}{s^*\sqrt{\frac{1}{n_{22}}+\frac{1}{n_{23}}}} \rightarrow T_{n_{22}+n_{23}-2=43\,d.d.l.}$$

Estimation de la variance commune : $s^{*2} = \frac{n_{22}s_{22}^2+n_{23}s_{23}^2}{n_{22}+n_{23}-2} = \frac{20\times 2.5^2+25\times 3^2}{43} = 8.139 \Rightarrow s^* = 2.853.$

Les valeurs critiques (cf. table Student) avec $\alpha = 5\%$: $t_{2.5\%} = \pm 2.017$.

Marge d'erreur : $E = t_{\frac{\alpha}{2}} s^* \sqrt{\frac{1}{n_{22}}+\frac{1}{n_{23}}} = 2.017 * 2.853\sqrt{\frac{1}{20}+\frac{1}{25}} = 1.726$

Ecart observé des températures moyennes : $(\bar{x}_{n_{22}} - \bar{x}_{n_{23}}) = 14.51 - 14.40 = 0.11$ °C

$I.C._{95\%}$: Intervalle de confiance de niveau 95% de la différence des températures moyennes :

$$(m_{22} - m_{23}) \in [-1.616\ ;\ 1.836].$$

Conclusion : $0 \in I.C._{95\%}$, on peut donc conclure avec un risque d'erreur $\alpha = 5\%$, qu'il n'y a pas de différence significative entre les températures moyennes de 2022 et 2023 en France. On peut les supposer comme égales $m_{22} \simeq m_{23}$.

Solution 2 - Test bilatéral symétrique de la différence de deux moyennes :

Hypothèses statistiques :

$$\begin{cases} H_0 : m_{22} - m_{23} = 0 \\ H_1 : m_{22} - m_{23} \neq 0 \quad \text{Les températures moyennes en France en 2022 et 2023 sont différentes} \end{cases}$$

Seuil de signification : $\alpha = 5\%$.

La statistique de test de la différence de deux moyennes variances égales :

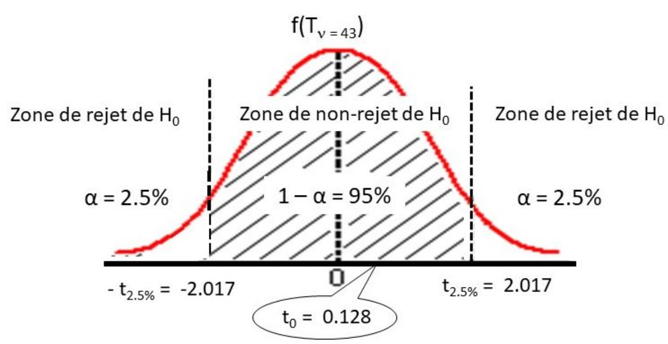
$$\frac{(\overline{X}_{n_{22}}-\overline{X}_{n_{23}})-(m_{22}-m_{23})}{s^*\sqrt{\frac{1}{n_{22}}+\frac{1}{n_{23}}}} \rightarrow T_{n_{22}+n_{23}-2=43\,d.d.l.}$$

Estimation de la variance commune : $s^{*2} = \frac{n_{22}s_{22}^2+n_{23}s_{23}^2}{n_{22}+n_{23}-2} = \frac{202.5^2+25*3^2}{43} = 8.139 \Rightarrow s^* = 2.853.$

Les valeurs critiques (cf. table Student à 48 d.d.l.) : avec $\alpha = 5\%$: $t_{2.5\%} = \pm 2.017$.

Calcul de la valeur de la statistique de test sous l'hypothèse nulle d'égalité des moyennes $H_0 : m_{22} = m_{23}$:

$$t_0 = \frac{(\bar{x}_{n_{22}}-\bar{x}_{n_{23}})}{s^*\sqrt{\frac{1}{n_{22}}+\frac{1}{n_{23}}}} = \frac{0.11}{2.853\sqrt{\frac{1}{20}+\frac{1}{25}}} = 0.128$$



Règle de décision et conclusion : vu que $t_0 = 0.128 \in [-2.017, 2.017]$ alors non-rejet de l'hypothèse nulle H_0. On peut donc conclure au seuil de signification $\alpha = 5\%$, qu'il n'y a pas de différence significative entre les températures moyennes de 2022 et 2023 en France.

************** °°° **************

Exercice 2 : **(Barème de notation : a) 2.5 pts b) 1.5 pt c) 1.5 pt d) 2 pts = 7.5 pts)**

EnR 2021	France	Allemagne
Nombre de relevés	40	40
Part des EnR	19.30%	41.00%

a) Intervalle de confiance de niveau $1-\alpha = 95\%$ de la part des EnR en France-2021 :

Conditions d'application : échantillon aléatoire $n_F = 40$ de grande taille.

$\widehat{p}_F = 19.30\%$: estimation ponctuelle de la part des EnR en France en 2021.

La statistique de test : $\frac{\widehat{p}_F - p_F}{\sqrt{\frac{p_F \times q_F}{n_F}}} \rightarrow N(0\,,\,1)$

Risque d'erreur : $\alpha = 5\%$; $u_{\frac{\alpha}{2}} = 1.96$ table N(0 , 1).

Marge d'erreur : $E = u_{\frac{\alpha}{2}}\sqrt{\frac{\widehat{p}_F \times \widehat{q}_F}{n_F}} = 1.96\sqrt{\frac{0.1930\times 0.8070}{40}} = 12.23\%$

$I.C._{95\%}$ de la part des EnR en France-2021 : $p_F \in [7.07\%\ ;\ 31.53\%]$.

La part des EnR de l'UE-27 en 2021 : $p_{UE} = 21.80\% \in [7.07\%\ ;\ 31.53\%]$. On peut donc conclure avec un risque d'erreur $\alpha = 5\%$, que la part des EnR en France en 2021 n'est pas significativement différente de celle de l'UE-27 en 2021.

b) Taille d'échantillon requise pour estimer la part des EnR en France-2021 avec une précision au plus égale à 8% :

Risque d'erreur : $\alpha = 5\% \Rightarrow u_{\frac{\alpha}{2}} = 1,96$ table N(0 , 1)

On cherche la taille d'échantillon n^* telle que la marge d'erreur :

$$E^* = u_{\frac{\alpha}{2}}\sqrt{\frac{\widehat{p}_F\widehat{q}_F}{n^*}} \leq 0.08 \Rightarrow n* \geq (\frac{u_{\frac{\alpha}{2}}}{0.08})^2\widehat{p}_F\widehat{q}_F$$

$$\Rightarrow n^* \geq (\frac{1.96}{0.08})^2 0.1930 \times 0.8070 \Rightarrow n^* \geq 93.49 \Rightarrow n^* \geq 94 \text{ relevés.}$$

c) Niveau de confiance de l'intervalle $[22.91\%\ ;\ 59.09\%]$: de la part des EnR en Allemagne-2021

Marge d'erreur : $E = u_{\frac{\alpha}{2}}\sqrt{\frac{\widehat{p}_A \times \widehat{q}_A}{n_A}} = \frac{(0.5909-0.2291)}{2} = 18.09\%$

$$\Rightarrow u_{\frac{\alpha}{2}} = E \times \sqrt{\frac{n_A}{\widehat{p}_A \times \widehat{q}_A}} = 0.1809 \times \sqrt{\frac{40}{0.41 \times 0.59}} = 2.3262$$

cf. table N(0 , 1), $P(U \leq 2.3262) = \Phi(2.33) = 0.99$

$$\Rightarrow 1 - \frac{\alpha}{2} = 99\% \Rightarrow \frac{\alpha}{2} = 1\% \Rightarrow \alpha = 2\% \Rightarrow 1-\alpha = 98\%.$$

On attribue à l'intervalle $[22.91\%\ ;\ 59.09\%]$ de la part des EnR en Allemagne en 2021, un niveau de confiance $1-\alpha = 98\%$.

d) Comparaison de 2 proportions :

Solution 1 - Intervalle de confiance de la différence de deux proportions :

Conditions d'application du test : échantillons indépendants provenant de deux populations normales - Grands échantillons $n_F = n_A = 40$.

La statistique de test : $\frac{(\widehat{P}_F - \widehat{P}_A)-(p_F - p_A)}{\sqrt{\frac{p_F q_F}{n_F} + \frac{p_A q_A}{n_A}}} \rightarrow N(0\,,\,1)$

France : $n_F = 40$; $\widehat{p}_F = 19.30\%$

Allemagne : $n_A = 40$; $\widehat{p}_A = 41\%$

Ecart observé : $\widehat{p}_F - \widehat{p}_A = 19.30\% - 41.00\% = -21.70\%$

Risque d'erreur : $\alpha = 5\% \Rightarrow u_{\alpha=2.5\%} = u_{2.5\%} = \pm 1.96$ cf. table N(0 , 1)

Marge d'erreur $E = u_{\alpha=2.5\%}\sqrt{\frac{\widehat{p}_F\widehat{q}_F}{n_F} + \frac{\widehat{p}_A\widehat{q}_A}{n_A}} = 1.96\sqrt{\frac{0.1930\times0.8070}{40} + \frac{0.41\times0.59}{40}} = 19.54\%$

$I.C._{95\%}$: de la différence des parts d'EnR : $(p_F - p_A) \in [-41.20\% \, ; \, -2.20\%]$.

Règle de décision et conclusion :

Conclusion : puisque $0 \notin I.C_{95\%} = [-41.20\% \, ; -2.20\%]$ on peut conclure avec un risque d'erreur $\alpha = 5\%$, que la différence observée $(\widehat{p}_F - \widehat{p}_A) = -21.70\%$ est significative, on peut donc conclure que les parts d'EnR en France et en Allemagne en 2021, sont significativement différentes.

Solution 2 - Test bilatéral symétrique de la différence de deux proportions :

Hypothèses statistiques :

$$\begin{cases} H_0 : p_F - p_A = 0 \\ H_1 : p_F - p_A \neq 0 \ \text{ les parts d'EnR en 2021 en France et en Allemagne sont différentes} \end{cases}$$

Conditions d'application du test : échantillons indépendants provenant de deux populations normales - Grands échantillons $n_F = n_A = 40$.

La statistique de test : $\frac{(\widehat{P}_F - \widehat{P}_A) - (p_F - p_A)}{\sqrt{\frac{p_F q_F}{n_F} + \frac{p_A q_A}{n_A}}} \rightarrow N(0\,,\,1)$

Echantillon France : $n_F = 40$; $\widehat{p}_F = 19.30\%$

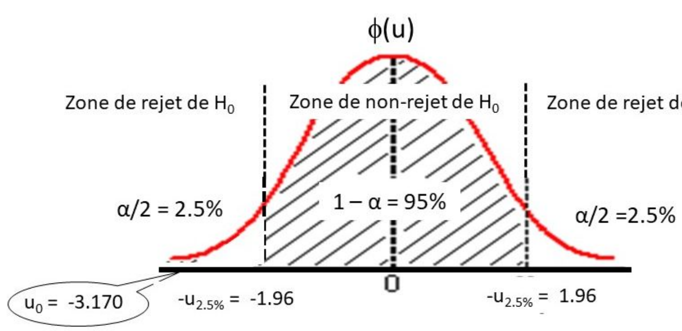
Echantillon Allemagne : $n_A = 40$; $\widehat{p}_A = 41\%$

Ecart observé : $\widehat{p}_F - \widehat{p}_A = 19.30\% - 41\% = -21.70\%$

Risque d'erreur : $\alpha = 5\% \Rightarrow u_{2.5\%} = \pm 1.96$ cf. table N(0 , 1)

Valeur de la statistique de test sous l'hypothèse nulle H_0 : $p_F = p_A$

$$u_0 = \frac{(0.1930 - 0.41)}{\sqrt{\frac{0.1930\times0.8070}{40} + \frac{0.41\times0.59}{40}}} = -3.170$$



Conclusion : $u_0 = -3.170 \notin [-u_{2.5\%} = -1.96 \, ; \, u_{2.5\%} = 1.96]$ n'appartient pas à la zone de Non-Rejet de H_0, on peut donc conclure avec un risque d'erreur $\alpha = 5\%$, que les parts des EnR en France et en Allemagne sont significativement différentes.

*************** °°° ***************