Annexe Chapitre 2 : RAPPELS

A. *Tests paramétriques de comparaisons de deux variables quantitatives*

1 Test de comparaison de deux moyennes pour deux échantillons indépendants Premier cas : petits échantillons, lois normales, même variance *Test de Student*

1.1 Contexte

Il s'agit d'un test portant sur deux échantillons indépendants issus de deux populations $\mathcal{P}_1$ et $\mathcal{P}_2$, et sur deux variables X et Y représentant le même caractère quantitatif continu, distribuées suivant une loi normale.

- $\mathcal{P}_1$ et $\mathcal{P}_2$ sont deux populations,
- X et Y sont deux variables quantitatives indépendantes, issues respectivement de $\mathcal{P}_1$ et $\mathcal{P}_2$, telles que

$$X \sim N(\mu_X, \sigma) \quad \text{et} \quad Y \sim N(\mu_Y, \sigma),$$

  avec μ_X, μ_Y et σ inconnus.
- il n'existe pas de conditions sur la taille des échantillons, mais en pratique on applique ce test pour des échantillons de petites tailles (n_1 ou $n_2 < 30$). Pour les grands échantillons (n_1 et $n_2 \geq 30$) on utilise plutôt son approximation normale (cf section **A**.2).

Exemple 1

Dans une étude s'intéressant au rôle de la valeur affective d'un texte dans la récupération du souvenir chez les personnes âgées, on a recueilli l'âge (en années) et le score au test de "Wescher Mémoire" d'un groupe de 10 personnes présentant un déficit mnésique et d'un groupe de 11 personnes ne présentant pas ce déficit.

sujet déficitaire	1	2	3	4	5	6	7	8	9	10
âge	80	91	82	87	82	85	84	85	88	87
score de Wescher	66	59	84	68	80	75	72	82	78	76

sujet non déficitaire	1	2	3	4	5	6	7	8	9	10	11
âge	80	81	82	84	85	85	86	89	91	92	86
score de Wescher	113	94	87	98	103	110	97	119	88	91	100

1. On cherche dans un premier temps à vérifier si les deux populations sont homogènes vis à vis de l'âge.
2. Dans un second temps on souhaite mettre en évidence, au risque $\alpha = 1\%$ le fait que les personnes présentant un déficit mnésique ont un score diminué.

Exemple 1.1

Contexte

$\mathcal{P}_1$: personnes âgées présentant un déficit mnésique

$\mathcal{P}_2$: personnes âgées ne présentant pas de déficit mnésique

X = `âge`, quantitative de moyenne μ_X et d'écart-type σ dans $\mathcal{P}_1$
Y = `âge`, quantitative de moyenne μ_Y et d'écart-type σ dans $\mathcal{P}_2$
Les deux variables X et Y représentent le même caractère quantitatif continu, et sont supposées suivre des lois normales dans $\mathcal{P}_1$ et $\mathcal{P}_2$ de même variance (σ^2).

1.2 Hypothèses de test et risque

L'hypothèse nulle H_0 correspond au fait que les moyennes des variables X et Y sont identiques $H_0 : \mu_X = \mu_Y$
alors que sous l'alternative H_1 on suppose, selon l'hypothèse de recherche envisagée
soit que les variables X et Y ont deux moyennes différentes $H_1 : \mu_X \neq \mu_Y$ test bilatéral,
soit que la moyenne de X est plus élevée que celle de Y $H_1 : \mu_X > \mu_Y$
soit que la moyenne de X est plus faible que celle de Y $H_1 : \mu_X < \mu_Y$ tests unilatéraux.

Le risque d'erreur de $1^{\text{ère}}$ espèce α, probabilité de rejeter H_0 à tort, est fixé à l'avance, et en général on choisit $\alpha = 5\%$.

Exemple 1.1
Hypothèses de test et risque
On veut tester l'hypothèse selon laquelle les âges des personnes déficitaires sont semblables à ceux des personnes non déficitaires, que l'on traduit puisque les variables sont normales, par le fait que les moyennes d'âge des deux populations sont égales, qui correspond à l'hypothèse nulle. Sous l'alternative les moyennes d'âge des deux populations sont différentes (pas d'orientation).
Il s'agit de tester $H_0 : \mu_X = \mu_Y$ contre $H_1 : \mu_X \neq \mu_Y$ alternative bilatérale, le risque n'étant pas précisé on prendra $\alpha = 5\%$.

1.3 Observations

On dispose de 2 échantillons tirés au hasard de manière indépendante dans les 2 populations.
On note :

$\mathcal{E}_1$ l'échantillon de taille n_1 issu de $\mathcal{P}_1$ et $(x_1, \ldots, x_{n_1})$ les mesures de $\mathcal{E}_1$,
$\mathcal{E}_2$ l'échantillon de taille n_2 issu de $\mathcal{P}_2$ et $(y_1, \ldots, y_{n_2})$ les mesures de $\mathcal{E}_2$

Les résultats sont résumés pour chaque échantillon par
les valeurs observées des moyenne $\overline{X}$, écart-type S_X et écart-type sans biais S_X^* empiriques sur l'échantillon $\mathcal{E}_1$ notées :

moyenne $\overline{x}$, écart-type s_X et écart-type sans biais (corrigé) s_X^*

les valeurs observées des moyenne $\overline{Y}$, écart-type S_Y et écart-type sans biais S_Y^* empiriques sur l'échantillon $\mathcal{E}_2$ notées :

moyenne $\overline{y}$, écart-type s_Y et écart-type sans biais (corrigé) s_Y^* .

Exemple 1.1
Observations
échantillon $\mathcal{E}_1$ issu de $\mathcal{P}_1$ de taille $n_1 = 10$
on observe : $\overline{x} = \frac{1}{10}\sum x_i = \frac{851}{10} = 85,10$

$$s_X = \sqrt{\frac{1}{10}\sum x_i^2 - \overline{x}^2} = \sqrt{\frac{72\,517}{10} - 85,10^2} = 3,11 \qquad s_X^* = \sqrt{\frac{10}{9}} \times 3,11 = 3,28$$

échantillon $\mathcal{E}_2$ issu de $\mathcal{P}_2$ de taille $n_2 = 11$
on observe : $\overline{y} = \frac{1}{11}\sum y_i = \frac{941}{11} = 85,55$

$$s_Y = \sqrt{\frac{1}{11}\sum y_i^2 - \overline{y}^2} = \sqrt{\frac{80\,649}{11} - 85,55^2} = 3,70 \qquad s_Y^* = \sqrt{\frac{11}{10}} \times 3,70 = 3,88$$

On estime à 85,10 ans l'âge moyen des personnes présentant un déficit mnésique et à 85,55 ans l'âge moyen des personnes ne présentant pas ce déficit.
On estime à 3,28 ans l'écart-type de l'âge des personnes présentant un déficit mnésique et à 3,88 ans l'écart-type de l'âge des personnes ne présentant pas ce déficit.

1.4 Statistique de test et loi sous H_0

- **Statistique de test de Student**

  L'écart entre les deux moyennes est quantifié par la différence des moyennes empiriques $\overline{X} - \overline{Y}$, qui sous l'hypothèse nulle H_0 d'égalité des deux moyennes μ_X et μ_Y doit être proche de 0. La statistique de test prend également en compte la variabilité de cette différence qui dépend de l'estimation de la variance σ^2 supposée commune à X et Y.

  Sous les conditions que X et Y suivent des lois normales de même variance σ^2,
  la statistique de test de Student

$$T = \frac{\overline{X} - \overline{Y}}{S^* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad \text{avec } S^* = \sqrt{\frac{n_1 \left(S_X\right)^2 + n_2 \left(S_Y\right)^2}{n_1 + n_2 - 2}}$$

  suit sous H_0 une loi de Student à $(n_1 + n_2 - 2)$ dl, notée $T_{n_1+n_2-2}$
  où S^{*2} estime la variance commune σ^2.

  La valeur observée de T

$$t_{obs} = \frac{\overline{x} - \overline{y}}{s^* \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \qquad \text{avec } s^* = \sqrt{\frac{n_1\, s_X^2 + n_2\, s_Y^2}{n_1 + n_2 - 2}}$$

- **Région critique et niveau de signification du test**

  Test unilatéral

  $H_1 : \mu_X > \mu_Y$ — RC à droite pour T — $\alpha_{obs} = P_{H_0}\left[T \geq t_{obs}\right] = 1 - P_{H_0}\left[T \leq t_{obs}\right]$
  $H_1 : \mu_X < \mu_Y$ — RC à gauche pour T — $\alpha_{obs} = P_{H_0}\left[T \leq t_{obs}\right]$

  Test bilatéral

  $H_1 : \mu_X \neq \mu_Y$ — RC aux 2 extrémités de T — $\alpha_{obs} = 2 \times P_{H_0}\left[T \geq |t_{obs}|\right]$

  Chaque probabilité est calculée à partir de la loi $T_{n_1+n_2-2}$

Exemple 1.1

Statistique de test de Student et loi sous H_0

Puisque X et Y sont supposées normales de même variance, la statistique de test de Student T suit sous H_0 une loi de Student à $n_1 + n_2 - 2 = 10 + 11 - 2 = 19$ dl.

L'estimation de l'écart-type commun $s^* = \sqrt{\frac{10 \times 3,11^2 + 11 \times 3,70^2}{10+11-2}} = 3,61$

et la valeur observée de T : $t_{obs} = \frac{85,10 - 85,55}{3,61 \sqrt{\frac{1}{10} + \frac{1}{11}}} = -0,282402$

Niveau de signification

Sous $H_1 : \mu_X \neq \mu_Y$ alternative bilatérale, RC se situe aux 2 extrémités du domaine de variation de T, donc

$$\begin{aligned} \alpha_{obs} &= 2 \times P_{H_0}\left[T \geq |t_{obs}|\right] = 2 \times P_{H_0}\left[T \geq |-0,282401|\right] = 2 \times P_{H_0}\left[T \geq 0,282401\right] \\ &= 2 \times 0,390344 = 0,780689 \quad \text{(calculée à partir de la loi } T_{19}\text{)} \end{aligned}$$

Décision et conclusion

Pour un niveau $\alpha = 5\%$, $\alpha_{obs} = 78,1\% > \alpha$ donc on conserve H_0 (on rejette H_1) au seuil $\alpha = 5\%$ et au risque de 2^{de} espèce β (inconnu).

On ne peut donc pas conclure à l'existence d'une différence entre les âges des personnes présentant un déficit mnésique et ceux des personnes n'en présentant pas, au seuil $\alpha = 5\%$ et au risque de 2^{de} espèce β.

Exemple 1.2

Contexte

$\mathcal{P}_1$: personnes âgées présentant un déficit mnésique

$\mathcal{P}_2$: personnes âgées ne présentant pas de déficit mnésique

$X =$ `score de Wescher`, quantitative de moyenne μ_X et d'écart-type σ dans $\mathcal{P}_1$

$Y =$ `score de Wescher`, quantitative de moyenne μ_Y et d'écart-type σ dans $\mathcal{P}_2$

Les deux variables X et Y représentent le même caractère quantitatif continu, et sont supposées suivre des lois normales dans $\mathcal{P}_1$ et $\mathcal{P}_2$ de même variance (σ^2).

Hypothèses de test et risque

On veut tester l'hypothèse selon laquelle les scores de Wescher des personnes déficitaires sont plus faibles que ceux des personnes non déficitaires, que l'on traduit par le fait que le score moyen des personnes déficitaires est inférieur à celui des personnes non déficitaires, hypothèse alternative unilatérale.
Il s'agit de tester $H_0 : \mu_X = \mu_Y$ contre $H_1 : \mu_X < \mu_Y$ alternative unilatérale, au risque $\alpha = 1\%$.

Observations

échantillon $\mathcal{E}_1$ issu de $\mathcal{P}_1$ de taille $n_1 = 10$

on observe : $\overline{x} = \frac{1}{10}\sum x_i = \frac{740}{10} = 74$

$$s_X = \sqrt{\frac{1}{10}\sum x_i^2 - \overline{x}^2} = \sqrt{\frac{55\,310}{10} - 74^2} = 7,42 \qquad s_X^* = \sqrt{\frac{10}{9}} \times 7,42 = 7,82$$

échantillon $\mathcal{E}_2$ issu de $\mathcal{P}_2$ de taille $n_2 = 11$

on observe : $\overline{y} = \frac{1}{11}\sum y_i = \frac{1100}{11} = 100$

$$s_Y = \sqrt{\frac{1}{11}\sum y_i^2 - \overline{y}^2} = \sqrt{\frac{111\,082}{11} - 100^2} = 9,92 \qquad s_Y^* = \sqrt{\frac{11}{10}} \times 9,92 = 10,40$$

On estime à 74 le score de Wescher moyen des personnes présentant un déficit mnésique et à 100 le score moyen des personnes ne présentant pas ce déficit.
On estime à 7,82 l'écart-type du score des personnes présentant un déficit mnésique et à 10,4 l'écart-type du score des personnes ne présentant pas ce déficit.

Statistique de test de Student et loi sous H_0

Puisque X et Y sont supposées normales de même variance, la statistique de test de Student T suit sous H_0 une loi de Student à $n_1 + n_2 - 2 = 10 + 11 - 2 = 19$ dl.

L'estimation de l'écart-type commun $s^* = \sqrt{\frac{10\times 7,42^2 + 11\times 9,92^2}{10+11-2}} = 9,27$

et la valeur observée de T : $t_{obs} = \frac{74-100}{9,27\sqrt{\frac{1}{10}+\frac{1}{11}}} = -6,420617$

Niveau de signification

Sous $H_1 : \mu_X < \mu_Y$ alternative unilatérale, la différence des scores moyens $\overline{X} - \overline{Y}$ est négative ainsi que la valeur de T : RC se situe à gauche du domaine de variation de T, donc

$$\begin{aligned} \alpha_{obs} &= P_{H_0}[T \leq t_{obs}] = P_{H_0}[T \leq -6,420617] = 1 - P_{H_0}[T \leq 6,420617] \\ &= 0,00000186 \quad \text{(calculée à partir de la loi } T_{19}\text{)} \end{aligned}$$

Décision et conclusion

$\alpha_{obs} < \alpha = 1\%$, donc on rejette H_0 et on valide H_1 au risque $\alpha = 1\%$.

On peut donc conclure que les personnes présentant un déficit mnésique ont un score de Wescher moyen inférieur à celui des personnes n'en présentant pas, au risque $\alpha = 1\%$ et au niveau de signification $\alpha_{obs} < 10^{-5}$.

2 Test de comparaison de deux moyennes pour deux échantillons indépendants
Second cas : grands échantillons et lois quelconques

2.1 Contexte

Le contexte est identique à celui du test précédent, mais pour échantillons de grandes tailles (n_1 et $n_2 \geq 30$) ce qui se substitue aux conditions de normalité des variables et d'égalité des variances.
Il s'agit d'un test portant sur deux échantillons indépendants issus de deux populations $\mathcal{P}_1$ et $\mathcal{P}_2$, et sur deux variables X et Y représentant le même caractère quantitatif continu, de lois quelconques, de moyennes μ_X et μ_Y, d'écarts-types σ_X et σ_Y inconnus.

Exemple 2 (Tiré de L. Chanquoy, "Statistiques Appliquées à la Psychologie", pp. 100-101)
Dans une expérience sur les images mentales, 135 étudiants sont répartis en deux groupes (l'un de 60 et l'autre de 75) : ils doivent apprendre une liste de 40 mots concrets et les rappeler une heure après l'apprentissage. Le premier groupe utilise la méthode dite de "l'image construite", le second groupe la méthode de "l'image donnée". Les résultats (nombre de mots rappelés) sont donnés dans le tableau ci-dessous :

	effectif	moyenne	variance sans biais
méthode de l'image construite	$n_1 = 60$	$\overline{x} = 22$	$s_X^{*2} = 15,4$
méthode de l'image donnée	$n_2 = 75$	$\overline{y} = 16$	$s_Y^{*2} = 11,7$

On veut tester l'hypothèse selon laquelle la méthode de "l'image construite" est plus performante que celle de "l'image donnée", au risque $\alpha = 1\%$.

Contexte
$\mathcal{P}_1$: étudiant utilisant la méthode de "l'image construite"
$\mathcal{P}_2$: étudiant utilisant la méthode de "l'image donnée"
X = `nombre de mots rappelés`, quantitative de moyenne μ_X et d'écart-type σ_X dans $\mathcal{P}_1$
Y = `nombre de mots rappelés`, quantitative de moyenne μ_Y et d'écart-type σ_Y dans $\mathcal{P}_2$
Les deux variables X et Y représentent le même caractère quantitatif continu.

Hypothèses de test et risque
Il s'agit de tester l'hypothèse selon laquelle le nombre de mots moyen rappelés avec la méthode de "l'image construite" est supérieur au nombre de mots moyen rappelés avec la méthode de "l'image donnée", c'est à dire : $\mu_X > \mu_Y$, alternative unilatérale. Cela revient à mettre en place le test de comparaison de deux moyennes sur deux échantillons indépendants

$$\begin{cases} H_0 : & \mu_1 = \mu_2 \\ H_1 : & \mu_1 > \mu_2 \quad \text{unilatéral droit au risque } \alpha = 1\% \end{cases}$$

Observations
échantillon $\mathcal{E}_1$ issu de $\mathcal{P}_1$ de taille $n_1 = 60$
on observe : $\overline{x} = 22 \quad s_X^{*2} = 15,4 \quad s_X^* = \sqrt{15,4} = 3,9$

échantillon $\mathcal{E}_2$ issu de $\mathcal{P}_2$ de taille $n_2 = 75$
on observe : $\overline{y} = 16 \quad s_Y^{*2} = 11,7 \quad s_Y^* = \sqrt{11,7} = 3,4$

On estime à 22 le nombre moyen de mots rappelés par les étudiants utilisant la méthode de "l'image construite" et à 16 le nombre de mots moyen rappelés avec la méthode de "l'image donnée".
On estime à 3,9 l'écart-type du nombre de mots rappelés par les étudiants utilisant la méthode de "l'image construite" et à 3,4 l'écart-type du nombre de mots rappelés avec la méthode de "l'image donnée".

2.2 Statistique de test et loi sous H_0

Sous la condition n_1 et $n_2 \geq 30$ et sous l'hypothèse nulle d'égalité des deux moyennes $H_0 : \mu_X = \mu_Y$ la statistique de test

$$Z = \frac{\overline{X} - \overline{Y}}{\sqrt{\frac{\left(S_X^*\right)^2}{n_1} + \frac{\left(S_Y^*\right)^2}{n_2}}} \overset{\text{approx.}}{\sim} \mathcal{N}(0,1)$$

- **Valeur observée de la statistique de test**
  La valeur observée de la statistique de test Z : $z_{obs} = \frac{\overline{x}-\overline{y}}{\sqrt{\frac{s_X^{*2}}{n_1}+\frac{s_Y^{*2}}{n_2}}}$

- **Région critique et niveau de signification du test**

  Test unilatéral

$H_1 : \mu_X > \mu_Y$	RC à droite pour Z	$\alpha_{obs} = P_{H_0}\left[Z \geq z_{obs}\right] = 1 - P_{H_0}\left[Z \leq z_{obs}\right]$
$H_1 : \mu_X < \mu_Y$	RC à gauche pour Z	$\alpha_{obs} = P_{H_0}\left[Z \leq z_{obs}\right]$

  Test bilatéral

$H_1 : \mu_X \neq \mu_Y$	RC aux 2 extrémités de Z	$\alpha_{obs} = 2 \times P_{H_0}\left[Z \geq \lvert z_{obs}\rvert\right]$

  Chaque probabilité est calculée à partir de la loi $\mathcal{N}(0,1)$.

Exemple 2

Statistique de test et loi sous H_0

Puisque $n_1 = 60 \geq 30$ et $n_2 = 75 \geq 30$ la statistique de test Z suit approximativement sous H_0 une loi normale centrée réduite $\mathcal{N}(0,1)$.

La valeur observée de Z : $z_{obs} = \frac{22-16}{\sqrt{\frac{15,4}{60}+\frac{11,7}{75}}} = 9,34$.

Niveau de signification

Sous l'alternative unilatérale $H_1 : \mu_X > \mu_Y$ on s'attend à observer une valeur de $\overline{X}$ plus grande que celle de $\overline{Y}$ donc une valeur de $\overline{X} - \overline{Y}$ largement positive, de même pour Z ; RC est à l'extrémité droite du domaine de variation de Z.

Il suffit alors de calculer la p-valeur $\alpha_{obs} = P_{H_0}\left[Z \geq 9,34\right] = 1 - P_{H_0}\left[Z \leq 9,34\right] \simeq 0,00000....$

Décision et conclusion

Comme $\alpha_{obs} < \alpha = 1\%$, on rejette H_0 avec un risque de 1% et on accepte l'hypothèse H_1.

On peut conclure que le nombre moyen de mots rappelés par les étudiants utilisant la méthode de "l'image construite" est supérieur au nombre de mots moyen rappelés avec la méthode de "l'image donnée" donc que la méthode de "l'image construite" est plus performante que celle de "l'image donnée", avec un risque de 1%.

Remarque

STATISTICA utilise systématiquement la statistique T (avec une loi Student), même si $n_1, n_2 \geq 30$.
On peut demander le calcul de la statistique Z en option (variances séparées) mais la loi utilisée par *STATISTICA* pour calculer la p-valeur n'est pas la loi normale mais une loi de Student dont des dl sont calculés à partir des variances observées.

3 Tests de comparaison de deux moyennes pour deux échantillons appariés

3.1 Contexte

Il s'agit d'un test portant sur deux échantillons appariés (ou pairés, ou appareillés) de deux variables X et Y représentant le même caractère quantitatif continu,

soit issus d'une même population $\mathcal{P}$, et dans ce cas les mesures de X et de Y sont faites sur les mêmes individus,

soit issus d'une population $\mathcal{P}$ composée de paires d'individus (2 individus) aussi semblables que possible :
- personnes de la même famille (paires de jumeaux, (père, fils), ...)
- ou bien, on définit des "variables d'appariement" (sexe, âge, durée ou gravité de la maladie, ...).

Pour comparer les moyennes de X et Y, on cherche à contrôler les facteurs "connus" et "inconnus" qui jouent un rôle dans la différence entre les variables X et Y, autres que le facteur étudié, qui pourraient être des "facteurs de confusion".

Exemples types d'utilisation de ce test :

∗ comparaison de type avant-après (un traitement, une thérapie, un régime, ...)

Pour chaque sujet, X représente le résultat avant traitement et Y le résultat après traitement.

Pour détecter un changement dû à ce traitement :

en hypothèse nulle, on suppose qu'il n'y a pas de changement $H_0 : \mu_X = \mu_Y$

en alternative, selon l'hypothèse de recherche envisagée

il y a une amélioration **alternative unilatérale** $H_1 : \mu_X > \mu_Y$ ou $H_1 : \mu_X < \mu_Y$

ou une modification (sans préciser l'orientation) **alternative bilatérale** $H_1 : \mu_X \neq \mu_Y$

La "paire" est constituée par un seul individu et l'appariement est idéal : le sujet est pris comme son "propre témoin" (ou "contrôle"). Ceci permet d'augmenter la puissance en diminuant la variabilité des résultats : en effet, la variabilité des résultats d'un même individu ("intra-sujet") est plus faible que celle d'individus différents ("inter-sujets").

∗ comparaison de l'efficacité de deux traitements

X désigne le résultat avec le traitement A et Y le résultat avec le traitement B.

Pour comparer l'efficacité des traitements :

en hypothèse nulle, on suppose que les deux traitements ont la même efficacité $H_0 : \mu_X = \mu_Y$

en alternative, selon l'hypothèse de recherche envisagée

un traitement est plus performant que l'autre, **alternative unilatérale**

$H_1 : \mu_X < \mu_Y$ ou $H_1 : \mu_X > \mu_Y$

ou ils ont simplement des efficacités différentes (sans orientation) **alternative bilatérale**

$H_1 : \mu_X \neq \mu_Y$

Si le sujet est son propre témoin, chaque sujet reçoit le traitement A et le traitement B dans un ordre qui peut-être aléatoire, ou prédéterminé si l'on veut étudier un "effet ordre" du traitement.

L'appariement permet de contrôler les facteurs connus (ou inconnus) pour influencer l'évolution de la maladie, autres que le traitement (âge, intensité, durée, ...).
Il permet de limiter le "biais de confusion" qui consisterait à conclure à l'existence d'une différence d'efficacité entre les traitements alors qu'elle n'est due qu'à un (ou des) facteur(s) de confusion.

On dispose donc de deux échantillons appariés (en paires d'individus "jumeaux") de taille n.

Principe pour des échantillons appariés :

on travaille sur la variable "différence" notée $D = X - Y$ (ou $Y - X$).

3.2 Hypothèses de test et risque

Les hypothèses se traduisent sur la moyenne de la variable D notée μ en remarquant que $\mu = \mu_X - \mu_Y$ (ou $\mu_Y - \mu_X$).

L'hypothèse nulle	$H_0 : \mu_X = \mu_Y$	s'écrit	$H_0 : \mu = 0$	
l'alternative bilatérale	$H_1 : \mu_X \neq \mu_Y$	s'écrit	$H_1 : \mu \neq 0$	
ou les alternatives unilatérales	$H_1 : \mu_X > \mu_Y$	devient	$H_1 : \mu > 0$	unilatérale droite,
et	$H_1 : \mu_X < \mu_Y$		$H_1 : \mu < 0$	unilatérale gauche.

On est donc ramené à un test de comparaison d'une moyenne (de la différence) à la valeur théorique (de référence) 0.

3.3 Observations

On dispose de 2 échantillons appariés de même taille n. On note :
$\mathcal{E}_1$ l'échantillon de taille n et $(x_1, \ldots, x_n)$ les mesures de X,
$\mathcal{E}_2$ l'échantillon de taille n et $(y_1, \ldots, y_n)$ les mesures de Y.

On calcule la variable différence pour chaque paire $d_i = x_i - y_i$ (ou $y_i - x_i$) :
on dispose d'un échantillon de taille n de D dont les mesures sont notées $(d_1, \ldots, d_n)$.
Les différences sont résumés par
les valeurs observées des moyenne $\overline{D}$, écart-type S_D et écart-type sans biais S^*_D empiriques de D notées : moyenne $\overline{d}$, écart-type s_D et écart-type sans biais (corrigé) s^*_D

3.4 Statistiques de test et lois sous H_0

Les statistiques utilisées sont celles du test de comparaison d'une moyenne à une valeur théorique μ_0, pour $\mu_0 = 0$.

1. ***petites tailles d'échantillon*** ($n < 30$)
   Sous la condition que la variable "différence" D suit une loi normale, la statistique de test de Student $T = \frac{\overline{D}}{\frac{S^*_D}{\sqrt{n}}}$ suit sous H_0 une loi de Student à $(n-1)$ dl, notée T_{n-1}.
2. ***grandes tailles d'échantillon*** ($n \geq 30$)
   Sous la condition $n \geq 30$, la statistique de test $Z = \frac{\overline{D}}{\frac{S^*_D}{\sqrt{n}}}$ suit approximativement sous H_0 une loi normale $\mathcal{N}(0, 1)$.

Exemple 3

On étudie le comportement agressif d'enfants ayant des difficultés de comportement, avant et après la projection d'un film d'aventures. Pour cela, on a noté le nombre de comportements agressifs pendant la demi-journée précédant la projection du film et pendant la demi-journée suivant la projection du film de 26 enfants considérés comme ayant des difficultés de comportement.

enfant	1	2	3	4	5	6	7	8	9	10	11	12	13
avant projection	46	31	65	61	47	32	58	14	7	48	43	33	14
après projection	71	79	27	39	75	28	32	36	61	83	80	28	72

Peut-on dire, au risque $\alpha = 5\%$, que le film a une influence négative sur le comportement agressif des enfants ?

Contexte

$\mathcal{P}$: enfants ayant des difficultés de comportement

$X =$ `nombre de comportements agressifs avant la projection du film`, quantitative de moyenne μ_X dans $\mathcal{P}$

$Y =$ `nombre de comportements agressifs après la projection du film`, quantitative de moyenne μ_Y dans $\mathcal{P}$

Les deux variables X et Y représentent le même caractère quantitatif continu.

Hypothèses de test et risque

On veut tester l'hypothèse selon laquelle les nombres de comportements agressifs sont plus faibles avant la projection du film qu'après, que l'on traduit par le fait que le nombre moyen de comportements agressifs avant la projection est inférieur à celui après la projection, hypothèse alternative unilatérale. Il s'agit de tester $H_0 : \mu_X = \mu_Y$ contre $H_1 : \mu_X < \mu_Y$ alternative unilatérale, au risque $\alpha = 5\%$, test de comparaison de deux moyennes sur deux échantillons appariés.
On traduit les hypothèses sur la moyenne μ de la variable différence D
$H_0 : \mu = 0$ contre $H_1 : \mu < 0$ alternative unilatérale gauche, au risque $\alpha = 5\%$.

Observations

échantillon $\mathcal{E}_1$ issu de $\mathcal{P}$ de taille $n = 13$: $\overline{x} = \frac{1}{13}\sum x_i = \frac{499}{13} = 38,38$
échantillon $\mathcal{E}_2$ issu de $\mathcal{P}$ de taille $n = 13$: $\overline{y} = \frac{1}{13}\sum y_i = \frac{711}{13} = 54,69$

les deux échantillons étant appariés, on définit la variable différence $D = X - Y$

enfant i	1	2	3	4	5	6	7	8	9	10	11	12	13
avant projection x_i	46	31	65	61	47	32	58	14	7	48	43	33	14
après projection y_i	71	79	27	39	75	28	32	36	61	83	80	28	72
différence $d_i = x_i - y_i$	-25	-48	38	22	-28	4	26	-22	-54	-35	-37	5	-58

on observe : $\overline{d} = \frac{1}{13}\sum d_i = \frac{-212}{13} = -16,31 \; (= \overline{x} - \overline{y})$

$$s_D^{*2} = \frac{1}{12}\left(\sum d_i^2 - 13\overline{d}^2\right) = \frac{1}{12}\left(15\,716 - 13 \times (-16,31)^2\right) = 1\,021,56$$

$$s_D^* = \sqrt{1\,021,56} = 31,96$$

On estime à $38,38$ le nombre moyen de comportements agressifs des enfants avant la projection du film et à $54,69$ le nombre moyen de comportements agressifs des enfants après la projection.
On estime à $-16,31$ la moyenne des différences du nombre de comportements agressifs des enfants avant et après la projection du film et à $31,96$ l'écart-type de cette différence.

Statistique de test de Student et loi sous H_0

Sous la condition que la variable D suit une distribution normale, la statistique de test de Student T suit sous H_0 une loi de Student à $n - 1 = 13 - 1 = 12$ dl.
La valeur observée de T : $t_{obs} = \frac{-16,31}{\frac{31,96}{\sqrt{13}}} = -1,840$

Niveau de signification

Sous $H_1 : \mu < 0$ alternative unilatérale gauche, la différence $\overline{D}$ est négative ainsi que la valeur de T : RC se situe à gauche du domaine de variation de T, donc

$$\begin{aligned} \alpha_{obs} &= P_{H_0}[T \leq t_{obs}] = P_{H_0}[T \leq -1,84] = 1 - P_{H_0}[T \leq 1,84] \\ &= 0,04534 \quad \text{(calculée à partir de la loi } T_{12}\text{)} \end{aligned}$$

Décision et conclusion

$\alpha_{obs} < \alpha = 5\%$, donc on rejette H_0 et on valide H_1 au risque $\alpha = 5\%$.
On peut donc conclure que la projection du film a une influence négative sur les comportements agressifs des enfants ayant des difficultés de comportement, au risque $\alpha = 5\%$ et au niveau de signification $\alpha_{obs} \simeq 4,6\%$.

4 Vérification des conditions d'application des tests

4.1 Graphiques et indices

Les conclusions tirées de l'examen d'un graphique ont nécessairement un caractère subjectif. Seuls des tests peuvent infirmer l'hypothèse de normalité d'une distribution qui n'est connue que sur un échantillon, ou confirmer l'égalité des variances de deux distributions observées sur deux échantillons. En revanche, ces tests sont incapables de mettre en évidence certaines caractéristiques essentielles des échantillons, que l'on pourra déceler avec un graphique.

Histogramme

Boîte à moustache (*"Box plot"*)

Ce graphique synthétise la distribution de la variable. Classiquement la boîte est délimitée par les 1$^{\text{er}}$ et 3$^{\text{ème}}$ quartiles de la distribution (indice de dispersion ou de variabilité), la médiane (indicateur de valeur centrale ou de position) est représentée à l'intérieur de la boîte, et les moustaches figurent les minimum et maximum (extrêmes, représentent l'étendue).

La position de la médiane par rapport aux limites de la boîte, ainsi que la position de la boîte par rapport aux extrêmes donne des indications sur la forme de la distributions, notamment sur la symétrie.

D'autres graphiques proposent de choisir comme limites de la boîte :

'moyenne $\pm$ écart-type' ou 'moyenne $\pm$ erreur-type' (erreur-type = écart-type/$\sqrt{n}$)

et comme extrémités (moustaches) :

'moyenne $\pm 1,96\times$écart-type' ou 'moyenne $\pm 1,96\times$ erreur-type'

ce qui ne renseigne pas sur la symétrie de la distribution mais peut aider à dépister visuellement, en comparant les boîtes à moustaches de deux (ou plusieurs) groupes, des variances inégales ou des différences entre moyennes.

Droite de Henry (*"Q-Q plot"*)

On compare les quantiles empiriques (en abscisse) aux quantiles théoriques du même ordre (en ordonnée) de la loi normale centré réduite.
Si la variable est gaussienne, les points sont alignés le long d'une droite appelée "droite de Henry" (on peut en déduire approximativement la moyenne et l'écart-type de la loi normale).

Ce diagramme (appelé aussi "diagramme Quantile-Quantile", ou "Q-Q Plot") est un graphique permettant de comparer visuellement un échantillon et une distribution théorique de référence (le plus souvent normale), dans le but de décider s'il est vraisemblable que cette distribution théorique ait généré l'échantillon, et d'analyser les raisons qui peuvent éventuellement faire rejeter cette hypothèse.

Ce diagramme peut donc être considéré comme un "test de normalité visuel".

Cœfficient d'asymétrie (*"skewness"*)

Le cœfficient d'asymétrie est défini par : $\dfrac{\mu_3}{\sigma^3}$

où μ_3 est le moment centré d'ordre 3 de X et σ l'écart-type de X.

Il vaut 0 pour une loi symétrique (en particulier normale),
il est > 0 pour une distribution concentrée à gauche et étalée à droite (dans ce cas moy $>$ med $>$ mode) et < 0 dans le cas contraire (pour une distribution concentrée à droite et étalée à gauche ; moy $<$ med $<$ mode).

Cœfficient d'aplatissement (*"kurtosis"*)

Le cœfficient d'aplatissement est défini par : $\dfrac{\mu_4}{\sigma^4} - 3$

où μ_4 est le moment centré d'ordre 4 de X et σ l'écart-type de X.

Il vaut 0 pour une loi normale, il est > 0 pour une distribution moins aplatie (plus pointue) que la loi normale et < 0 dans le cas contraire (pour une distribution plus aplatie).

4.2 Tests de normalité

Il s'agit de tests d' "adéquation", c'est à dire dont l'objectif est d'établir la plausibilité de l'hypothèse selon laquelle l'échantillon a été prélevé dans une population ayant une distribution donnée. Un autre test d'adéquation est le "test du khi-deux d'adéquation".

Test de Kolmogorov-Smirnov

Le test de Kolmogorov est non-paramétrique : il ne place aucune contrainte sur la distribution de référence, et ne demande pas qu'elle soit connue sous forme analytique (bien que ce soit pourtant le cas le plus courant).

Etant donnés un échantillon d'une variable quantitative X de fonction de répartition F, et une fonction de répartition théorique (de référence) F_0, le test de Kolmogorov teste l'hypothèse H_0 selon laquelle les fonctions de répartition F et F_0 sont égales.

Pour cela, il calcule sur l'échantillon de taille n une quantité D_n, appelée "statistique de Kolmogorov", dont la distribution est connue lorsque H_0 est vraie.

D_n est l'écart maximum entre les fonctions de répartition théorique F_0 et empirique (observée) F_n : $D_n = \sup_x |F_n(x) - F_0(x)|$.
Si F_0 est continue, sous H_0 : "la distribution de X est la même que la distribution théorique"
$\sqrt{n}D_n$ suit approximativement quand n est grand, la distribution de Kolmogorov (table des valeurs critiques).

Une valeur élevée de D_n indique que la distribution de l'échantillon s'éloigne sensiblement de la distribution de référence F_0, et qu'il est donc peu probable que H_0 soit correcte : donc si l'écart entre F_n et F_0 est trop grand on rejette H_0.

Pour tester la normalité de X : F_0 est la fonction de répartition théorique d'une loi normale de moyenne et d'écart-type donnés.

Remarque

Dans *STATISTICA* on ne peut pas spécifier la moyenne et l'écart-type de la distribution théorique.

Test de Lilliefors

C'est l'adaptation du test de Kolmogorov-Smirnov au cas où la moyenne et l'écart-type de la distribution normale sous H_0 ne sont pas spécifiés.

La procédure du test commence donc par estimer la moyenne et la variance sur les observations, puis calcule l'écart maximum entre la fonction de répartition empirique et la fonction de répartition théorique dont la moyenne et l'écart-type ont été estimés sur les observations.

Si l'écart est trop grand (tables des valeurs critiques de Lilliefors) on rejette H_0.

Ce test est assez peu puissant : un grand nombre d'observations est nécessaire pour rejeter l'hypothèse de normalité.

Test de Shapiro-Wilk

Contrairement aux deux tests précédents, il ne teste que la normalité c'est à dire H_0 : la distribution de X est une loi normale.
La statistique de test W cœfficient de détermination (carré du cœfficient de corrélation) entre les observations ordonnées et des cœfficients centrés réduits basés sur les valeurs d'ordre attendues sous la loi normale centré réduite, W est donc compris entre 0 et 1.
Si W est suffisamment proche de 1 on conserve l'hypothèse de normalité H_0, en revanche
si W est trop faible on rejette H_0 (table des valeurs critiques).
Ce test est plus puissant que les tests précédents Kolmogorov-Smirnov et Lilliefors.

Exemple 1.1

Contexte

$\mathcal{P}_1$: personnes âgées présentant un déficit mnésique
$\mathcal{P}_2$: personnes âgées ne présentant pas de déficit mnésique
X = `âge`, quantitative de moyenne μ_X et d'écart-type σ dans $\mathcal{P}_1$
Y = `âge`, quantitative de moyenne μ_Y et d'écart-type σ dans $\mathcal{P}_2$

Pour comparer les moyennes des deux variables X et Y on a supposé qu'elles suivaient des lois normales dans $\mathcal{P}_1$ et $\mathcal{P}_2$. On vérifie ici la plausibilité de ces suppositions.

Graphiques et indices

On représente l'histogramme de l'ge pour chaque échantillon (variables X et Y, avec des échelles identiques), ainsi que la boîte à moustaches : les étendues sont similaires pour les deux variables et on ne repère pas d'asymétrie évidente.

groupe	taille	asymétrie	aplatissement
déficitaire	10	$0,196255$	$-0,251361$
non déficitaire	11	$0,333284$	$-0,696849$

Les coefficients d'asymétrie sont légèrement positifs dans les deux groupes, indiquant une légère concentration à gauche et ceux d'aplatissement sont négatifs, signe de distributions un peu plus aplaties que celle d'une loi normale.

Tests de normalité

Hypothèses de test et risque

Pour vérifier la normalité de X on teste :

H_0 : X suit une loi normale, de moyenne et d'écart-type inconnus dans $\mathcal{P}_1$ contre
H_1 : X ne suit pas une loi normale dans $\mathcal{P}_1$, alternative bilatérale, au risque $\alpha = 5\%$ (risque classique à utiliser quand il n'est pas spécifié).

Statistiques de test et niveaux de signification

statistique de test		valeur observée	$p - valeur$
Kolmogorov-Smirnov (D max)	D	$D_{obs} = 0,1276$	p Lilliefors $\alpha_{obs} < 1$
Shapiro-Wilk (SW-W)	W	$W_{obs} = 0,976$	$\alpha_{obs} = 0,94$

Décision et conclusion

Pour chacun des deux tests, $\alpha_{obs} < 5\%$ on conserve donc H_0 au seuil $\alpha = 5\%$ (et au risque β inconnu).

On conserve l'hypothèse que l'âge suit une loi normale dans la population des personnes âgées présentant un déficit mnésique, au seuil $\alpha = 5\%$.

On conserve également l'hypothèse que l'âge suit une loi normale dans la population des personnes âgées ne présentant pas de déficit mnésique, au seuil $\alpha = 5\%$. En effet, pour la variable Y : $D_{obs} = 0,1807$ p Lilliefors < 1 et $W_{obs} = 0,9512$ $\alpha_{obs} = 0,6592$.

4.3 Tests d'égalité (homogénéité) des variances

L'une des conditions pour les tests de comparaison de deux moyennes (T de Student) (ou plus, ANOVA) est l'égalité des variances (homogénéité) dans chaque groupe (variances intra-groupe). Les tests suivants permettent de tester cette homogénéité, à condition que les variables soient régies par des lois normales.
Cependant la plupart de ces tests sont peu robustes aux écarts à la normalité. Dans la plupart des cas, si l'on suspecte une hétérogénéité des variances, il est plus avisé d'utiliser le test de Student pour variances séparées, ou une transformation de variable, ou de pratiquer un test non paramétrique.

Test de Fisher : test du rapport des variances (*"ratio F variances"*)

Ce test teste l'égalité de deux variances en calculant le rapport des estimations des deux variances, pour deux échantillons indépendants de tailles respectives n_1 et n_2.

Sous l'hypothèse nulle d'égalité des deux variances H_0 : $\sigma_1^2 = \sigma_2^2$ la statistique de test $F = \dfrac{S_{n_1}^{2*}}{S_{n_2}^{2*}}$ suit approximativement une loi de Fisher $F(n_1 - 1, n_2 - 1)$.

Ce test n'est pas robuste : un écart même minime à la normalité fausse les résultats. Il faut absolument s'assurer du caractère gaussien des variables.

Remarque

STATISTICA calcule systématiquement le rapport de la plus grande variance à la plus petite et donne une $p - valeur$ bilatérale ($2 \times P_{H_0}[F > f_{obs}]$ double de la $p - valeur$ de la loi $F(n_1 - 1, n_2 - 1)$).

Test de Levene

Il teste l'hypothèse nulle d'homogénéité des variances de k variables $X_1, X_2, \ldots, X_k$

H_0 : égalité des k variances $\sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$

contre l'hypothèse alternative d'hétérogénéité des variances des k variables

H_1 : il existe (au moins) deux variances différentes $\sigma_i^2 \neq \sigma_j^2$

Ce test est basé sur le principe que plus la variance d'une variable est grande, plus les écarts absolus à la moyenne de la variable sont grands.

Il consiste à comparer les moyennes (ANOVA) des écarts (absolus) U des valeurs à la moyenne de leur groupe, avec pour l'observation i du groupe j : $u_{ij} = |x_{ij} - \overline{x}_j|$.

La statistique de test : $W = \dfrac{(n-k)\sum_{j=1}^{k} n_j(\overline{U}_j - \overline{U})^2}{(k-1)\sum_{j=1}^{k}\sum_{i=1}^{n_j}(U_{ij} - \overline{U}_j)^2}$

suit approximativement une loi de Fisher $F(k-1, n-k)$ sous l'hypothèse nulle d'égalité des k variances.

Ce test est plus robuste et moins sensible aux écarts à la normalité que les tests de Bartlett, de Hartley ou de Cochran.

Test de Brown-Forsythe

Sur le même principe que le test de Levene, ce test consiste à comparer les moyennes (ANOVA) des écarts (absolus) U des valeurs à la médiane de leur groupe, avec pour l'observation i du groupe j : $u_{ij} = |x_{ij} - \widetilde{x}_j|$ où $\widetilde{x}_j$ est la médiane de X pour le groupe j.

Ce test est une généralisation du test de Levene, plus robuste que celui-ci, particulièrement dans les cas de distributions non symétriques.

Cependant, la robustesse de ce test et de celui de Levene ne sont pas établies en présence de tailles d'échantillons différentes.

Exemple 1.1

Contexte

$\mathcal{P}_1$: personnes âgées présentant un déficit mnésique
$\mathcal{P}_2$: personnes âgées ne présentant pas de déficit mnésique
$X =$ `âge`, quantitative de moyenne μ_X et d'écart-type σ_X dans $\mathcal{P}_1$
$Y =$ `âge`, quantitative de moyenne μ_Y et d'écart-type σ_Y dans $\mathcal{P}_2$

Pour comparer les moyennes des deux variables X et Y on a supposé qu'elles avaient la même variance. On vérifie ici la plausibilité de cette supposition.

Graphiques

On représente l'histogramme de l'ge pour chaque échantillon (variables X et Y, avec des échelles identiques), ainsi que la boîte à moustaches : les étendues sont similaires pour les deux variables mais la boîte est légèrement plus grande pour le groupe "non déficitaire" suggérant une variabilité un peu plus grande. Il faut faire le test affirmer ou infirmer pour la significativité de cette différence observée.

Tests de normalité

Hypothèses de test et risque

Pour vérifier l'égalité des variances de X et de Y on teste :
H_0 : les variances de X et de Y sont égales $\sigma_X^2 = \sigma_Y^2$ contre
H_1 : les variances de X et de Y sont différentes $\sigma_X^2 \neq \sigma_Y^2$ alternative bilatérale, au risque $\alpha = 5\%$ (risque classique à utiliser quand il n'est pas spécifié).

Statistiques de test et niveaux de signification

statistique de test		valeur observée	dl	$p-valeur$
Fisher (Ratio F)	F	$F_{obs} = 1,399944$	$(1,19)$	$\alpha_{obs} = 0,624016$
Levene	W	$W_{obs} = 0,218055$	$(1,19)$	$\alpha_{obs} = 0,645836$
Brown-Forsythe (Brn-Fors)	BF	$BF_{obs} = 0,174573$	$(1,19)$	$\alpha_{obs} = 0,680763$

Décision et conclusion

Pour chacun des trois tests, $\alpha_{obs} > 5\%$ on conserve donc H_0 au seuil $\alpha = 5\%$ (et au risque β inconnu).

On conserve l'hypothèse que la variance de l'âge est la même dans la population des personnes âgées présentant un déficit mnésique et dans celle des personnes âgées ne présentant pas de déficit mnésique, au seuil $\alpha = 5\%$.

B. *Tests de comparaisons de deux variables qualitatives*

1 Test du khi-deux d'homogénéité

1.1 Contexte

Test de comparaison de deux variables qualitatives sur deux échantillons indépendants

Il s'agit d'un test portant sur deux échantillons indépendants issus de deux populations $\mathcal{P}_1$ et $\mathcal{P}_2$, et sur deux variables X_1 et X_2 représentant le même caractère qualitatif avec un nombre fini $l \geq 2$ de modalités.
On note p_{i1} pour $i = 1, \ldots, l$ les l proportions des modalités de X_1 dans $\mathcal{P}_1$
et p_{i2} pour $i = 1, \ldots, l$ celles de X_2 dans $\mathcal{P}_2$

NB : le test d'homogénéité s'applique aussi au cas où le nombre de populations c est ≥ 3.

Exemple 1

On souhaite comparer les comportements à risque d'adolescents victimes de mauvais traitements pendant l'enfance à ceux d'adolescents n'en ayant pas subi. On a relevé comme comportement à risque le fait d'avoir des idées suicidaires, en trois modalités : jamais, parfois ou souvent, pour 85 adolescents considérés comme victimes de mauvais traitements et pour 104 adolescents non maltraités.

Peut-on, au risque $\alpha = 5\%$, accepter l'hypothèse d'une répartition différente des idées suicidaires chez les adolescents maltraités pendant l'enfance de celle des adolescents non maltraités ?

Contexte

$\mathcal{P}_1$: Adolescents maltraités pendant l'enfance

$\mathcal{P}_2$: Adolescents non maltraités

X_1 : `idées suicidaires` des adolescents maltraités, dans $\mathcal{P}_1$
qualitative à trois modalités ($l = 3$) : `jamais, parfois, souvent`.

X_2 : `idées suicidaires` des adolescents non maltraités, dans $\mathcal{P}_2$
qualitative à 3 modalités ($l = 3$)

Les 2 variables représentent le même caractère qualitatif à 3 modalités : jamais , parfois, souvent.

1.2 Hypothèses de test et risque

L'hypothèse nulle H_0 correspond au fait que les variables X_1 et X_2 ont la même distribution, alors que l'alternative H_1 suppose que les variables X_1 et X_2 ont deux distributions différentes :

$$\begin{cases} H_0 : & X_1 \text{ et } X_2 \text{ ont la même distribution} \\ H_1 : & X_1 \text{ et } X_2 \text{ ont des distributions différentes} \end{cases} \quad \text{test bilatéral, au risque } \alpha \text{ (dans l'exemple } \alpha = 5\%)$$

Ces hypothèses se traduisent sur les proportions par :

$$\begin{cases} H_0 : & \text{pour chaque modalité } i \quad p_{i1} = p_{i2} \\ H_1 : & \text{il existe une modalité } i \text{ pour laquelle } p_{i1} \neq p_{i2} \end{cases}$$

Exemple 1

On veut tester l'hypothèse selon laquelle la répartition des idées suicidaires dans la population des adolescents maltraités est différente de celle des adolescents non maltraités.

H_0 : la répartition des idées suicidaires est la même dans les deux populations d'adolescents

H_1 : les répartitions des idées suicidaires sont différentes dans les deux populations d'adolescents

test bilatéral au risque $\alpha = 5\%$

1.3 Observations

On dispose de 2 échantillons tirés au hasard de manière indépendante dans les 2 populations.
On note :

$\mathcal{E}_1$ l'échantillon de taille n_1 issu de $\mathcal{P}_1$
$\mathcal{E}_2$ l'échantillon de taille n_2 issu de $\mathcal{P}_2$
n représente la taille totale des 2 échantillons : $n = n_1 + n_2$

Les résultats sont résumés dans un tableau de contingence où les effectifs observés sur l'échantillon $\mathcal{E}_1$ sont notés $n_{i1}, i = 1, \ldots, l$ et ceux observés sur l'échantillon $\mathcal{E}_2$ sont notés $n_{i2}, i = 1, \ldots, l$.
Les proportions de chaque modalité des variables X_1 et X_2 sont estimées par les fréquences observées sur l'échantillon correspondant :

pour $i = 1, \ldots, l$ $\quad p_{i1}$ est estimée par $f_{i1} = \frac{n_{i1}}{n_1}$ et p_{i2} est estimée par $f_{i2} = \frac{n_{i2}}{n_2}$

Exemple 1

On dispose de 2 échantillons indépendants :
$\mathcal{E}_1$ de taille $n_1 = 85$ issu de $\mathcal{P}_1$ et $\mathcal{E}_2$ de taille $n_2 = 104$ issu de $\mathcal{P}_2$.

Tableau de contingence des effectifs observés

`idées suicidaires`	adolescents maltraités	adolescents non maltraités	Total Ligne L_i
`jamais`	$n_{11} = 50$	$n_{12} = 92$	L_1
`parfois`	$n_{21} = 25$	$n_{22} = 10$	L_2
`souvent`	$n_{31} = 10$	$n_{32} = 2$	L_3
Total Colonne C_j	$n_1 = 85 = C_1$	$n_2 = 104 = C_2$	$n = 189$

Tableau de contingence des fréquences observées (Pourcentage des effectifs en Colonne)

`idées suicidaires`	adolescents maltraités	adolescents non maltraités	Total
`jamais`	$f_{11} = \frac{50}{85} = 0,5882$	$f_{12} = \frac{92}{104} = 0,8846$	$f_1 = \frac{50+92}{189} = 0,7513$
`parfois`	$f_{21} = \frac{25}{85} = 0,2941$	$f_{22} = \frac{10}{104} = 0,0962$	$f_2 = \frac{25+10}{189} = 0,1852$
`souvent`	$f_{31} = \frac{10}{85} = 0,1177$	$f_{32} = \frac{2}{104} = 0,0192$	$f_3 = \frac{10+2}{189} = 0,0635$
Total	1	1	$n = 189$

Ainsi, on estime à $11,77\%$ la proportion d'adolescents maltraités ayant `souvent` des idées suicidaires, alors qu'elle est estimée à $1,92\%$ pour les adolescents non maltraités.
En revanche la proportion d'adolescents n'ayant `jamais` d'idées suicidaires est estimée à $58,82\%$ pour les adolescents maltraités, et à $88,46\%$ pour les adolescents non maltraités.

1.4 Statistique de test et loi sous H_0

Sous l'hypothèse nulle H_0 les variables X_1 et X_2 ont la même distribution, les proportions de la modalité i sont donc égales dans les deux populations.
La proportion de la modalité i est estimée par : $f_i = \frac{n_{i1}+n_{i2}}{n} = \frac{L_i}{n}$ et l'effectif de la modalité i attendu sous H_0 est égal à :

$f_i \times n_1 = \frac{n_{i1}+n_{i2}}{n} \times n_1 = \frac{L_i C_1}{n}$ dans l'échantillon $\mathcal{E}_1$ et
$f_i \times n_2 = \frac{n_{i1}+n_{i2}}{n} \times n_2 = \frac{L_i C_2}{n}$ dans l'échantillon $\mathcal{E}_2$.

De manière générale l'effectif attendu sous H_0 (effectif théorique) pour la modalité i de la variable X_j (ligne i colonne j) est égal à :

$$e_{ij} = \frac{L_i C_j}{n} \quad (= f_i\, C_j)$$

Exemple 1

Sous H_0, on estime à $6,3\%$ la proportion d'adolescents ayant `souvent` des idées suicidaires, et à $75,13\%$ la proportion d'adolescents n'ayant `jamais` d'idées suicidaires.

Tableau des effectifs théoriques

`idées suicidaires`	adolescents maltraités	adolescents non maltraités	Total Ligne L_i
`jamais`	$e_{11} = \frac{142\times 85}{189} = 63,86$	$e_{12} = \frac{142\times 104}{189} = 78,14$	$L_1 = 142$
`parfois`	$e_{21} = \frac{35\times 85}{189} = 15,74$	$e_{22} = \frac{35\times 104}{189} = 19,26$	$L_2 = 35$
`souvent`	$e_{31} = \frac{12\times 85}{189} = 5,40$	$e_{32} = \frac{12\times 104}{189} = 6,60$	$L_3 = 12$
Total Colonne C_j	$C_1 = n_1 = 85$	$C_2 = n_2 = 104$	$n = 189$

- **Statistique de test du khi-deux de Pearson**

  La statistique du khi-deux de Pearson

$$Q^2 = \sum_{i,j} \frac{(N_{ij} - e_{ij})^2}{e_{ij}}$$

  permet de quantifier l'écart (la distance) entre effectifs empiriques N_{ij} et théoriques e_{ij} donc l'écart à l'hypothèse nulle.
  Sa valeur observée sur les deux échantillons $q^2_{obs} = \sum_{i,j} \frac{(n_{ij}-e_{ij})^2}{e_{ij}}$

- **Loi de la statistique sous H_0**

  Sous les conditions $n \geq 30$ et tous les effectifs théoriques $e_{ij} \geq 5$
  la statistique de test Q^2 suit approximativement sous H_0, une loi du khi-deux à $(l-1) \times (c-1) = (l-1)$ degrés de liberté (dl) notée χ^2_{l-1}.

- **Niveau de signification du test**

  Le niveau de signification α_{obs} ou *p-valeur* est la probabilité d'observer sous H_0 une valeur de Q^2 supérieure à q^2_{obs} :

$$\alpha_{obs} = P_{H_0}\left[Q^2 \geq q^2_{obs}\right] \quad \text{probabilité calculée à partir de la loi } \chi^2_{l-1}.$$

Exemple 1

Puisque $n = 189 \geq 30$ et tous les $e_{ij} \geq 5$ car $min(e_{ij}) = 5,40 \geq 5$ la statistique Q^2 suit approximativement sous H_0 une loi du khi-deux à $l - 1 = 2$ dl.
La valeur observée de Q^2 : $q_{obs} = 22,502$
Le niveau de signification : $\alpha_{obs} = P_{H_0}\left[Q^2 \geq 22,502\right] = 0,000051$ d'après la loi χ^2_2

1.5 Décision et conclusion

Règle de décision basée sur le niveau de signification α_{obs} :

- si $\alpha_{obs} > \alpha$ on conserve H_0 (on ne valide pas H_1) au seuil α et au risque de 2de espèce β inconnu
- si $\alpha_{obs} \leq \alpha$ on rejette H_0 en faveur de H_1 (on valide H_1) au risque α et au niveau de signification (*p-valeur*) α_{obs}

Exemple 1

Décision et conclusion

Pour $\alpha = 5\%$, on voit que $\alpha_{obs} = 0,0051\% < \alpha$ ce qui nous amène à rejeter H_0 en faveur de H_1 au risque $\alpha = 5\%$.

On peut donc conclure que la répartition des idées suicidaires chez les adolescents maltraités pendant l'enfance est significativement différente de celle des adolescents non maltraités, au risque $\alpha = 5\%$ et au niveau de signification $\alpha_{obs} = 0,0051\%$.

Remarque
STATISTICA fournit également la valeur observée de la statistique de test du khi-deux du maximum de vraisemblance : $MV = \sum_{ij} N_{ij}\ ln\left(\frac{N_{ij}}{e_{ij}}\right)$ qui, sous les conditions $n \geq 30$ et tous les $e_{ij} \geq 5$, suit approximativement une loi du khi-deux à $(l-1)(c-1)$ dl.
Cette statistique est équivalente à celle du khi-deux de Pearson.

Exemple 1

$MV_{obs} = 23,161$ et $\alpha_{obs} = 0,000037$: décision et conclusion restent identiques.

2 Cas particulier : test de comparaison de deux proportions sur deux échantillons indépendants

2.1 Contexte

Il s'agit d'un cas particulier du test d'homogénéité lorsque le caractère qualitatif étudié n'a que deux modalités $l = 2$, représentées par "oui" ou "non" (souvent codées 1 ou 0). On peut donc se contenter de comparer les proportions de "oui" p_1 et p_2 dans les deux populations $\mathcal{P}_1$ et $\mathcal{P}_2$ (ou de manière équivalente, les proportions de "non").

2.2 Hypothèses de test

L'hypothèse nulle correspond à l'égalité des deux proportions $H_0 : p_1 = p_2$
et l'alternative peut prendre, selon l'hypothèse de recherche envisagée, l'une des trois formes

bilatérale : les deux proportions sont différentes $\quad H_1 : p_1 \neq p_2$

ou unilatérales :

la proportion de "oui" dans $\mathcal{P}_1$ est plus élevée que celle de $\mathcal{P}_2$ $\quad H_1 : p_1 > p_2$

ou bien la proportion de "oui" dans $\mathcal{P}_1$ est plus faible que celle de $\mathcal{P}_2$ $\quad H_1 : p_1 < p_2$

2.3 Statistiques de test du khi-deux et lois sous H_0

- **Statistique de test du khi-deux de Pearson**
  Le test du khi-deux d'homogénéité s'applique, et sous les mêmes conditions, la statistique de test du khi-deux de Pearson Q^2 suit approximativement sous H_0, une loi du khi-deux à 1 dl.

- **Statistique de test du khi-deux de Yates**
  Lorsque les effectifs théoriques sont proches de 5 (compris entre 5 et 10), on utilise la statistique du khi-deux de Yates (correction de continuité appliquée à la statistique de Pearson) :

$$Q_Y^2 = \sum_{i,j} \frac{\left(|N_{ij} - e_{ij}| - 0,5\right)^2}{e_{ij}}$$

  Sous H_0, la statistique du khi-deux de Pearson suit approximativement une loi du khi-deux à 1 dl.

2.4 Niveau de signification du test

Quelle que soit la statistique de test utilisée (statistique de Pearson, de Yates ou du maximum de vraisemblance), le niveau de signification du test obtenu avec une loi du khi-deux correspond à celui d'une alternative bilatérale.
En présence d'un test unilatéral, le niveau de signification obtenu doit être divisé par 2.

$$\alpha_{obs} = \frac{1}{2} \quad P_{H_0}\left(Q^2 \geq q^2_{obs}\right)$$

2.5 Décision et conclusion

La décision se fait en comparant le niveau de signification α_{obs} au risque d'erreur de 1ère espèce α fixé à l'avance (en général $\alpha = 5\%$).
Pour conclure en présence d'un test unilatéral, il faut **vérifier au préalable** que les fréquences observées sont compatibles avec l'hypothèse alternative, c'est à dire que

$f_1 > f_2$ si $H_1 : p_1 > p_2$

et inversement $f_1 < f_2$ si $H_1 : p_1 < p_2$

Exemple 2

On souhaite savoir si les adolescents victimes de mauvais traitements pendant l'enfance ont tendance à fuguer plus que les autres. On a relevé 28 fugueurs parmi les 85 adolescents maltraités et 9 parmi les 104 adolescents non maltraités.

Peut-on, au risque $\alpha = 5\%$, accepter l'hypothèse émise ?

Contexte

$\mathcal{P}_1$: Adolescents maltraités pendant l'enfance $\quad \mathcal{P}_2$: Adolescents non maltraités
$X_1 =$ `fugues` dans $\mathcal{P}_1$ qualitative à deux modalités ($l = 2$) : `oui, non`
$p_1 =$ proportion de fugueurs parmi les adolescents maltraités, dans $\mathcal{P}_1$

$X_2 =$ `fugues` dans $\mathcal{P}_2$ qualitative à deux modalités ($l = 2$)
$p_2 =$ proportion de fugueurs parmi les adolescents non maltraités, dans $\mathcal{P}_2$
Les 2 variables représentent le même caractère qualitatif à 2 modalités.

Hypothèses de test et risque

On veut tester l'hypothèse selon laquelle la proportion d'adolescents fugueurs est plus grande chez les adolescents maltraités que chez les adolescents non maltraités, c'est à dire réaliser le test de $\left\{\begin{array}{ll} H_0: & p_1 = p_2 \\ H_1: & p_1 > p_2 \end{array}\right.$ unilatéral, au risque $\alpha = 5\%$

Observations

On dispose de deux échantillons indépendants :
$\mathcal{E}_1$ issu de $\mathcal{P}_1$ de taille $n_1 = 85$ sur lequel on observe un effectif de fugueurs $n_{11} = 28$ et une fréquence de fugueurs $f_1 = \frac{n_{11}}{n_1} = \frac{28}{85} = 0,3294$
$\mathcal{E}_2$ issu de $\mathcal{P}_2$ de taille $n_2 = 104$ sur lequel on observe un effectif de fugueurs $n_{12} = 9$ et une fréquence de fugueurs $f_2 = \frac{n_{12}}{n_2} = \frac{9}{104} = 0,0865$
La proportion d'adolescents fugueurs est estimée à $32,94\%$ parmi les adolescents maltraités et à $8,65\%$ parmi les adolescents non maltraités, ce qui permet de vérifier que les fréquences observées sont compatibles avec l'hypothèse alternative, ici $f_1 = 32,94\% > f_2 = 8,65\%$ compatible avec $H_1 : p_1 > p_2$.

Statistique de test du khi-deux de Pearson

La valeur observée de la statistique du khi-deux de Pearson Q^2 : $q^2_{obs} = 17,524$

Loi de la statistique sous H_0 et niveau de signification

Puisque $n = 85 + 104 = 189 \geq 30$ et que le plus petit effectif théorique $\frac{(28+9)\times 85}{189} = 16,64 \geq 5$ la statistique de test du khi-deux de Pearson suit approximativement une loi du khi-deux à 1 dl sous H_0.
Pour le test unilatéral : $2\,\alpha_{obs} = P_{H_0}\left[Q^2 \geq 17,524\right] = 0,000028$ donc $\alpha_{obs} = 0,000014$.

Décision et conclusion

Puisque $\alpha_{obs} = 0,0014\% < \alpha = 5\%$ on rejette H_0 en faveur de H_1 au risque $\alpha = 5\%$.

On peut conclure que la proportion de fugueurs parmi les adolescents victimes de mauvais traitements pendant l'enfance est significativement supérieure (puisque $f_1 = 32,94\% > f_2 = 8,65\%$) à celle des adolescents non maltraités, au risque $\alpha = 5\%$ et au niveau de signification $\alpha_{obs} = 0,0014\%$.