

Tests statistiques élémentaires

Résumé

Il serait vain de chercher à présenter l'ensemble des tests statistiques, la littérature est très abondante sur le sujet. Cette vignette introduit les plus couramment calculés par les logiciels statistiques standards (SAS, Minitab, R...). Sont concernés les tests à un seul échantillon, de comparaison de deux échantillons, d'adéquation d'une loi et enfin non-paramétriques. Les différents tests sont décrits succinctement dans ce document en supposant que leur apprentissage est complété par des travaux pratiques mettant en œuvre un logiciel statistique. Dans ce contexte, les calculs et des exemples numériques ne sont pas systématiquement détaillés.

Retour au plan du cours.

1 Introduction

1.1 Objectif

Confronté à des phénomènes complexes et aléatoires, la prise de décision est difficile et les outils adaptés de la théorie des *tests* ont pour objet de guider les choix entre différentes alternatives. De façon générale, il s'agira de décider si des différences observées entre un modèle posé *a priori* et des observations sont *significatives* ou peuvent être considérées comme étant dues au simple effet du hasard consécutif aux aléas du tirage d'un échantillon. Réaliser un *test statistique* consiste à mettre en œuvre une procédure permettant :

- de confronter une hypothèse avec la réalité, ou plus exactement, avec ce que l'on perçoit de la réalité à travers les observations à disposition ;
- de prendre une décision à la suite de cette confrontation.

Si les problèmes traités par l'estimation (ponctuelle ou par intervalle de confiance) sont de type quantitatif, i.e. conduisent à un résultat numérique, ceux traités par les tests d'hypothèses sont d'ordre qualitatif, i.e. conduisent à une réponse du type rejet/acceptation de l'hypothèse statistique considérée.

1.2 Problématique

Le médicament testé est-il efficace ? Les pièces sortant d'une machine sont-elles conformes ? Un nouveau mode de culture bactérienne est-il plus efficace ? Quels sont les gènes significativement différentiellement exprimés dans un tissu pathologiques ?... Sont autant de questions auxquels des tests statistiques peuvent apporter des réponses sous 4 conditions :

1. la question est posée de sorte qu'il n'y ait que 2 réponses possibles : oui/non,
2. une expérimentation *planifiée* fournit des données relatives à cette question,
3. les données sont considérées comme la réalisation de variables aléatoires décrites par un *modèle statistique*,
4. la réponse à la question se traduit par l'acceptation ou le rejet d'une *hypothèse* (notée H_0) caractéristique du modèle précédent.

Dans ces conditions et avec une marge d'erreur plus ou moins bien contrôlée,

Accepter l'hypothèse, fait répondre *Non* à la question en considérant que les différences observées entre le modèle et la réalité des observations sont imputables au seul hasard.

Rejeter l'hypothèse fait *Oui* à la question : les différences sont jugées *significatives* car trop improbables ou invraisemblables.

1.3 Exemple

Des relevés effectués pendant de nombreuses années ont permis d'établir que le niveau naturel des pluies dans la Beauce X en millimètres par an suit une loi normale $\mathcal{N}(600, 100^2)$. Des entrepreneurs, surnommés "faiseurs de pluie", prétendaient pouvoir augmenter de 50mm le niveau moyen de pluie, ceci par insémination des nuages au moyen d'iodure d'argent. Leur procédé fut mis à l'essai entre 1951 et 1959 et on releva les hauteurs de pluies suivantes :

Année	1951	1952	1953	1954	1955	1956	1957	1958	1959
mm	510	614	780	512	501	534	603	788	650

Que pouvait-on en conclure ? Deux hypothèses s'affrontaient : ou bien l'insémination était sans effet, ou bien elle augmentait réellement le niveau moyen

de pluie de 50mm. Ces hypothèses pouvaient se formaliser comme suit, si μ désigne l'espérance mathématique de X , v.a.r. égale au niveau annuel de pluie.

$$\begin{cases} H_0 : \mu = 600mm \\ H_1 : \mu_1 = 650mm. \end{cases}$$

Les agriculteurs hésitaient à opter pour le procédé forcément onéreux des faiseurs de pluie. Ainsi, il fallait donc que l'expérience puisse les convaincre, c'est-à-dire que les faits observés contredisent nettement l'hypothèse H_0 , dite "hypothèse nulle" (H_1 s'appelle "hypothèse alternative"). Les agriculteurs n'étaient donc décidés à abandonner H_0 qu'en présence de faits expérimentaux traduisant une éventualité improbable compte-tenu de H_0 .

Par essence, l'objectif d'un test n'est pas de déterminer si H_0 est fondamentalement vraie ou non, mais plutôt de voir si H_0 est une hypothèse cohérente avec les données observées. On souhaite donc voir H_0 rejetée uniquement dans le cas où les observations la rendent invraisemblable. Ce qui amène généralement à la pratique suivante : on fixe une valeur $\alpha \in]0; 1[$, appelée risque de première espèce, et l'on impose au test d'être tel que si H_0 est vraie, la probabilité de rejeter à tort H_0 soit inférieure à α .

Dans notre exemple, les agriculteurs choisirent $\alpha = 0,05$, c'est-à-dire qu'ils assumaient le risque de se tromper dans 5 cas sur 100, en croyant les promesses des faiseurs de pluie, alors que ces faiseurs de pluie étaient en réalité des charlatans.

Comment décider ? Puisqu'il s'agit de "tester" la moyenne μ , il est naturel de s'intéresser à la moyenne empirique $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ que l'on sait convergente vers μ par la LGN. Pour ce test, \bar{X}_n est appelée la statistique de test. Alors, si H_0 est vraie, comme l'expérience porte sur $n = 9$ années, on a $\bar{X}_n \sim \mathcal{N}(600, 100^2/9)$.

Les données relevées indiquent que $\bar{x}_n = 610.2mm$ (c'est la moyenne obtenue sur toutes les années). L'hypothèse H_1 étant de la forme $\mu > 600$, on prendra la règle de décision suivante :

$$\begin{aligned} \text{Rejet de } H_0 &\iff \mathbb{P}_{H_0}(\text{Rejet de } H_0) < \alpha \\ &\iff \mathbb{P}_{H_0}(\bar{X}_n > \bar{x}_n) < \alpha, \end{aligned}$$

où le signe à l'intérieur de la probabilité précédente est le même que celui apparaissant dans H_1 . Il est alors facile de calculer cette dernière probabilité

car sous H_0 , la v.a.r. $Z = \frac{\bar{X}_n - 600}{100/3}$ suit une loi normale $\mathcal{N}(0, 1)$:

$$\begin{aligned} \mathbb{P}_{H_0}(\bar{X}_n > 610.2) &= \mathbb{P}\left(Z > \frac{610.2 - 600}{100/3}\right) \\ &= \mathbb{P}(Z > 0.306) \\ &= 1 - 0.6406 = 0.354. \end{aligned}$$

Comme cette probabilité, appelée *probabilité critique* ou *P-valeur*, est supérieure à $\alpha = 0.05$, la conclusion est donc que l'on accepte l'hypothèse H_0 , c'est-à-dire que les données ne sont pas incompatibles avec H_0 ; elles peuvent s'expliquer par le simple fait du hasard.

Remarque : La probabilité critique est, dans ce cas simple, lue directement dans la table de la loi $\mathcal{N}(0, 1)$. C'est la quantité calculée (*P-valeur* ou *P-value*) et affichée par un logiciel pour tout test mais la décision peut être écrite d'une autre façon car la plupart des lois ne sont pas complètement tabulées. En effet, si le nombre k_α est la valeur seuil telle que

$$\mathbb{P}_{H_0}(\bar{X}_n > k_\alpha) = \alpha \iff \mathbb{P}_{H_0}\left(\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} > u_{1-\alpha}\right) = \alpha$$

où $u_{1-\alpha}$ est l' $(1 - \alpha)$ -quantile de la loi normale $\mathcal{N}(0, 1)$. Alors la règle de décision est la suivante :

$$\begin{aligned} \text{Rejet de } H_0 &\iff \frac{\bar{x}_n - \mu}{\sigma/\sqrt{n}} > u_{1-\alpha} \\ &\iff \bar{x}_n > k_\alpha = \mu + u_{1-\alpha}\sigma/\sqrt{n}. \end{aligned}$$

Avec $\alpha = 5\%$, la valeur $u_{0.95} = 1.64$ est lue dans la table idoine et $k_\alpha = 655$, qui définit la région critique, conduit à la même conclusion : pour le niveau $\alpha = 5\%$, on ne rejette pas l'hypothèse H_0 .

Une précision importante consiste à rechercher l'erreur de 2ème espèce (β) à savoir l'erreur commise en acceptant à tort H_0 . Supposons donc que les faiseurs de pluie ont raison. Alors \bar{X}_n suit une loi normale $\mathcal{N}(650, 100^2/9)$, c'est-à-dire que la moyenne théorique en millimètres n'est plus 600 mais 650, sous l'hypothèse H_1 . Ainsi, la probabilité de se tromper (accepter $H_0 : \mu = 600$ alors que $\mu = 650$) s'écrit en utilisant toujours $k_\alpha = 655$ comme dans le

calcul précédent mais en calculant la probabilité sous l'hypothèse H_1 :

$$\beta = \mathbb{P}_{H_1}(\bar{X}_n < 655) = \mathbb{P}\left(Z < \frac{655 - 650}{100/3}\right) = \mathbb{P}(Z < 0.15) = 0.56,$$

qui est un risque très élevée, au détriment des faiseurs de pluie, car l'échantillon est de trop petite taille pour espérer mettre en évidence ou rendre significative une différence de 50mm.

2 Généralités sur les tests

2.1 Rôle et forme de l'hypothèse nulle

Dans l'exemple et de façon générale, les hypothèses H_0 et H_1 ne jouent pas des rôles symétriques et la façon de rédiger H_0 doit respectée la démarche scientifique : H_0 est l'hypothèse communément admise sauf si une expérience *répétable* vient la *réfuter* ; elle est conservée si les résultats de l'expérience ne sont pas clairs. Par analogie avec la justice, H_0 est la *présomption d'innocence* et des éléments matériels sont requis pour prouver la culpabilité.

Le couple (H_0, H_1) (hypothèse nulle, hypothèse alternative) peut prendre plusieurs formes où généralement H_1 est la négation logique directe de H_0 . Deux cas sont classiques :

1. H_0 simple et $H_1 = H_0^c$, le test est dit *bilatéral*

$$H_0 = \{\theta = \theta_0\} \text{ et } H_1 = \{\theta \neq \theta_0\};$$

2. H_0 composite et $H_1 = H_0^c$, le test est dit *unilatéral*

$$H_0 = \{\theta \leq \theta_0\} \text{ et } H_1 = \{\theta > \theta_0\}.$$

2.2 Niveau et puissance

Dans un test d'hypothèse statistique, il y a deux manières de se tromper :

1. la possibilité de rejeter une hypothèse alors qu'elle est vraie. Le α ou risque de première espèce borne la probabilité de se tromper dans ce sens, c'est-à-dire la probabilité d'avoir un *faux-positif*.
2. La possibilité d'accepter une hypothèse alors qu'elle est fautive. La probabilité de se tromper dans ce sens, c'est-à-dire la probabilité d'avoir un *faux-négatif*, est le risque de deuxième espèce, et est notée β .

On appelle *puissance de test*, notée $\pi(\theta)$ la probabilité de rejeter H_0 alors qu'elle est fautive. La puissance de test mesure donc son aptitude à bien rejeter une hypothèse fautive : $\pi(\theta) = 1 - \beta(\theta)$.

L'enjeu en théorie des tests et plus généralement en statistique inférentielle est de construire des tests uniformément les plus puissants c'est-à-dire au moins aussi puissant que tout autre test de même niveau α quelle que soit l'hypothèse alternative.

Ainsi, dans notre exemple, la première façon de se tromper est de croire les faiseurs de pluie, alors qu'ils ne sont pour rien dans le résultat obtenu (rejeter H_0 alors qu'elle est vraie), tandis que la seconde manière de se tromper est de ne pas croire les faiseurs de pluie, alors que leur méthode est bonne et que seul le hasard (malencontreux pour eux), dû au faible nombre d'observations, a donné des résultats insuffisants pour convaincre les agriculteurs (accepter H_0 alors qu'elle est fautive).

Il faut bien noter le rôle particulier joué dans l'exemple par H_0 . Si la forme de la région de rejet est indiquée par la nature de H_1 (650 plus grand que 600), la valeur du seuil k ne dépend que des choix de H_0 et α , tandis que β est déterminé par la considération supplémentaire de H_1 . D'autre part, les deux types d'erreur possibles ne sont pas du tout de la même importance. En effet, rejeter l'hypothèse H_0 alors qu'elle est vraie (risque de première espèce, qui est maîtrisé) est beaucoup plus coûteux que de la conserver à tort (risque de deuxième espèce, non maîtrisé).

Exemple : Soit μ la moyenne du niveau de radioactivité en picocuries par litre. La valeur $\mu_0 = 5$ est considérée comme la valeur critique entre eau potable et non potable. On peut donc souhaiter tester H_0 : " $\mu \geq 5$ " contre H_1 : " $\mu < 5$ ". L'erreur de première espèce a alors pour conséquence de laisser boire de l'eau toxique, alors que l'erreur de deuxième espèce conduit seulement à jeter de l'eau potable...

2.3 Démarche d'un test

Après avoir clairement défini la *question* posée et le *modèle statistique* sous-jacent, une démarche de test suit généralement les étapes suivantes.

1. Choix de H_0 et de H_1 . Fixer α .
2. Détermination de la statistique de test.

3. Allure de la région de rejet en fonction de H_1 .
4. Calcul de la région de rejet en fonction de α et H_0 .
5. Calcul de la valeur observée de la statistique de test.
6. Conclusion : rejet ou acceptation de H_0 au risque α .
7. Si possible, calcul de la puissance du test : $1 - \beta$.

Construire un test conduit à la définition d'une règle de décision contrôlée par le risque α de première espèce. Le choix *a priori* de $\alpha = 10\%, 5\%, 1\%$ est le plus souvent très arbitraire avec un usage couramment répandu de 5% . Ce choix est reconsidérer en fonction des coûts des risques encourus. Prenons la pratique courante de l'industrie pharmaceutique réalisant des criblages systématiques de milliers de molécules à la recherche d'un principe actif. Connaissant le coût extrêmement élevé des essais pré-cliniques et cliniques avant d'aboutir à la mise sur le marché d'un médicament, seules quelques molécules, dont l'effet thérapeutique potentiel est solidement établi, peuvent être considérées. En conséquence, le niveau α des tests au moment du criblage est fixé à un seuil très bas $0, 1\%$ ou moins. L'hypothèse H_0 : absence d'effet de la molécule testée, est donc conservée sauf si l'expérience montre très nettement le contraire. D'autre part, tester au seuil $\alpha 10000$ molécules conduit nécessairement, par le simple fait du hasard, à rejet en moyenne $\alpha 10000$ tests et donc à considérer à tort ces molécules "fausses positives". Ceci pose une autre question : celle des tests multiples.

2.4 Probabilité critique ou P-valeur

L'usage ancien des tables statistiques donnant les quantiles des différentes lois usuelles n'a plus lieu d'être avec la pratique d'un logiciel statistique. En effet, ceux-ci fournissent directement la probabilité critique ou *P-valeur* (en anglais *P-value*) associée à un test donné. Plutôt que de chercher dans une table les α -quantiles des lois associées, il suffit donc de comparer la probabilité critique fournit avec le seuil ou niveau de test prédéterminé.

La *probabilité critique* est la probabilité pour que la statistique de test T dépasse, sous l'hypothèse H_0 , la valeur seuil. Plus cette probabilité est proche de 0, plus forte est la contradiction entre H_0 et le résultat observé avec l'échantillon.

Dans le cas d'un test bilatéral, la région de rejet est de la forme :

$$\mathcal{R} = \{|T| > l\}.$$

La probabilité critique ou *P-valeur* est la quantité :

$$P_c(t) = \mathbb{P}_{H_0}\{|T| \geq l\}.$$

Le cas du test unilatéral se déduit facilement en supprimant les signes de valeur absolue.

Une fois le niveau α fixé, et le test mis en œuvre par un logiciel qui fournit la *P-valeur*, la décision est immédiate : $\{P_c(t) < \alpha\} \Leftrightarrow \{t \in \mathcal{R}\}$.

D'autre part l'édition de *P-valeurs* par un logiciel est souvent accompagnée par des étoiles, de 1 à 3, pour une lecture rapide de résultats multiples. Ces étoiles indiquent le *niveau* plus ou moins élevé de significativité d'un test : 1 étoile $P_c \leq 0,05$, 2 étoiles si $P_c \leq 0,01$, 3 si $P_c \leq 0,001$.

2.5 Choix du test

Le choix du test est guidé par la question posée et la structure des données issues de l'expérience. La littérature statistique est très abondante sur ce sujet, seuls sont référencés ici les tests élémentaires les plus couramment utilisés et généralement proposés par les logiciels statistiques ; en voici un guide sommaire.

Tests paramétriques les observations sont supposées suivre un modèle gaussien ou l'échantillon est de suffisamment grande taille pour accepter la normalité asymptotique par le théorème de la limite centrale.

Un échantillon

- Comparaison de la moyenne de l'échantillon à une valeur théorique lorsque la variance est supposée connue (Gauss)
- Comparaison de la moyenne de l'échantillon à une valeur théorique lorsque la variance est inconnue et estimée (Student)
- Comparer une proportion à une valeur théorique

Deux échantillons indépendants

- Comparaison de deux moyennes (variances égales ou échantillon suffisamment grand, Student)

- Comparaison de deux variances (Fisher)
- Comparaison de deux proportions

Deux échantillons appariés Le même échantillon est observé à deux instants différents ou dans deux conditions différentes (Student apparié)

Plusieurs échantillons ANOVA à un facteur

Tests d'adéquation

- Comparaison de deux distributions (chi-deux)
- Normalité d'une distribution (Kolmogorov, Shapiro Wilks)

Tests non-paramétrique Dans le cas : petit échantillon et distribution non gaussienne (cf. tests d'adéquation)

Deux échantillons indépendants Comparaison de deux médianes (Mann-Whitney)

Deux échantillons appariés Test de Wilcoxon sur les différences

Plusieurs échantillons Kruskal-Wallis

Liaisons entre variables D'autres questions, qui conduisent à tester l'influence d'une variable quantitative (régression) ou qualitative (ANOVA et comparaison de plusieurs moyennes) sur une variable quantitative, considèrent d'autres modèles et sont traitées dans les chapitres [Régression](#) et [Analyse de variance](#).

3 Tests paramétriques sur un échantillon

Dans cette section, la question posée revient à comparer la caractéristique d'un échantillon (moyenne, écart-type, proportion) à une valeur de référence connue *a priori*. Dans le cas dit *paramétrique*, l'échantillon est supposé issu d'une distribution gaussienne ou binomiale. Dans le cas contraire et si le nombre n d'observations est suffisamment grand, on se ramène à une loi normale et donc au cas paramétrique grâce au théorème de la limite centrale. La loi de \bar{X}_n est par exemple approchée par $\mathcal{N}\left(\mu, \frac{\sigma^2}{n}\right)$. Les procédures de test sont alors les mêmes. Dans le cas le plus défavorable petit échantillon non gaussien, ce sont les procédures de tests dits non paramétriques qu'il faut mettre en œuvre.

3.1 Test sur la moyenne d'une loi gaussienne

Moyenne μ inconnue et variance σ^2 connue

On suppose que l'on observe les réalisations d'un n -échantillon (X_1, \dots, X_n) issu d'une loi $\mathcal{N}(\mu, \sigma^2)$ avec μ inconnue et σ^2 connue. On se donne donc l'hypothèse nulle $H_0 : \mu = \mu_0$, μ_0 étant une valeur donnée par l'énoncé ou provenant de connaissances théoriques. Ensuite, il existe trois types d'hypothèse H_0 correspondant à leur alternative spécifique H_1 :

- H_0 composite : $\mu \leq \mu_0$ contre $H_1 : \mu > \mu_0$, test unilatéral à droite ;
- H_0 composite : $\mu \geq \mu_0$ contre $H_1 : \mu < \mu_0$, test unilatéral à gauche ;
- H_0 simple : $\mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$, test bilatéral.

Test unilatéral à droite $H_1 : \mu > \mu_0$ (cas de l'exemple introductif).

- Statistique de test : la moyenne empirique $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$.
- Région de rejet :

$$\text{Rejet de } H_0 \iff \mathbb{P}_{H_0}(\bar{X}_n > \bar{x}_n) < \alpha,$$

la région de rejet étant donnée par la forme de H_1 .

- Calcul de \bar{x}_n , la valeur observée de \bar{X}_n .
- Détermination de $\mathbb{P}_{H_0}(\bar{X}_n > \bar{x}_n)$. Sous H_0 , on a que $\bar{X}_n \sim \mathcal{N}(\mu_0, \frac{\sigma^2}{n})$. On définit alors $Z = \frac{\bar{X}_n - \mu_0}{\sigma/\sqrt{n}}$, qui suit $\mathcal{N}(0, 1)$. Ainsi on a :

$$\mathbb{P}_{H_0}(\bar{X}_n > \bar{x}_n) = \mathbb{P}\left(Z > \frac{\bar{x}_n - \mu_0}{\sigma/\sqrt{n}}\right).$$

- $\mathbb{P}_{H_0}(\bar{X}_n > \bar{x}_n)$ est la *probabilité critique* calculée par le logiciel et à comparer avec α .

Test unilatéral à gauche $H_1 : \mu < \mu_0$. Même méthode que précédemment, sauf que l'on remplace l'événement $\{\bar{X}_n > \bar{x}_n\}$ par $\{\bar{X}_n < \bar{x}_n\}$.

Test bilatéral $H_1 : \mu \neq \mu_0$. Même méthode que précédemment, sauf que l'on remplace l'événement $\{\bar{X}_n > \bar{x}_n\}$ par $\{|\bar{X}_n - \mu_0| > |\bar{x}_n - \mu_0|\}$.

Remarque : Il est facile de voir que les valeurs critiques d'un test (niveau α) de comparaison à une valeur de référence μ_0 sont les mêmes que les bornes de l'intervalle de confiance de probabilité $(1 - \alpha)$. Il est donc équivalent de tester l'hypothèse $H_0 : \mu = \mu_0$ que de s'assurer que la valeur de référence est à l'intérieur (acceptation) ou à l'extérieur (rejet) de l'intervalle de confiance estimant μ_0 .

Moyenne μ inconnue et variance σ^2 inconnue

C'est le test de Student. On suppose que l'on observe les réalisations d'un n -échantillon (X_1, \dots, X_n) issu d'une loi $\mathcal{N}(\mu, \sigma^2)$ avec μ inconnue et σ^2 inconnue.

Test bilatéral On teste $H_0 : \mu = \mu_0$ contre $H_1 : \mu \neq \mu_0$. Comme σ^2 est inconnue, on l'estime par la variance empirique S_n^2 . Par conséquent, la règle de décision est la suivante :

$$\text{Rejet de } H_0 \iff \mathbb{P}_{H_0} \left(\frac{|\bar{X}_n - \mu_0|}{S_n/\sqrt{n}} > \frac{|\bar{x}_n - \mu_0|}{s_n/\sqrt{n}} \right) < \alpha,$$

où s_n est l'estimation ponctuelle de l'écart-type empirique S_n associé à la réalisation du n -échantillon (X_1, \dots, X_n) . Sous H_0 , on sait que $T_n = \frac{\bar{X}_n - \mu_0}{S_n/\sqrt{n}}$ suit la loi de Student à $(n-1)$ degrés de liberté. Ainsi, la valeur de la probabilité critique est calculée à partir de la loi de Student, et l'on compare finalement avec le risque de première espèce α .

Les tests unilatéraux se déduisent de la même manière.

3.2 Test sur l'écart type d'une loi gaussienne

Un ensemble de tests (bilatéraux, unilatéral) sont construits de manière analogue à ceux sur la moyenne. La statistique de test vise à comparer S_n^2 et un paramètre σ_0^2 par le rapport $(n-1) \frac{S_n^2}{\sigma_0^2}$. Cette variable suit cette soit une loi du chi-deux à $(n-1)$ degrés de liberté. Le test le plus utilisé, par exemple en contrôle de qualité, est celui unilatéral. On veut par exemple détecter la dérive d'un procédé, tester si la variance d'un paramètre de la production a significativement augmenté même si la moyenne est restée proche de la valeur nominale.

on veut tester l'hypothèse composite $H_0 : \sigma \leq \sigma_0$ contre l'alternative $H_1 : \sigma > \sigma_0$. La probabilité critique est

$$\mathbb{P}_{H_0} \left((n-1) \frac{S_n^2}{\sigma_0^2} > (n-1) \frac{s_n^2}{\sigma_0^2} \right).$$

Le test de niveau α consiste à rejeter H_0 si :

$$\frac{(n-1)s_n^2}{\sigma_0^2} > \chi_{n-1;1-\alpha}^2 \iff s_n^2 > \frac{\sigma_0^2 \chi_{n-1;1-\alpha}^2}{n-1},$$

où $\chi_{n-1;1-\alpha}^2$ est l' $(1-\alpha)$ -quantile d'une loi du chi-deux à $(n-1)$ degrés de liberté.

Remarque : les procédures industrielles de maîtrise statistique des procédés (MSP) font appel à des échantillonnages séquentielles (cartes de contrôle) donc des tests spécifiques plus complexes.

3.3 Test sur une proportion

Test bilatéral On teste $H_0 : \pi = \pi_0$ contre l'hypothèse alternative $H_1 : \pi \neq \pi_0$ en supposant que la variable X suit une loi binomiale $\mathcal{B}(n, \pi)$.

Il est possible de définir un test exact mais comme la loi binomiale est associée à une variable aléatoire discrète, la fonction de répartition est étagée et les quantiles associés ne sont définis que sur des valeurs entières ; on ne peut donc pas obtenir un test de niveau exactement α pour toute valeur. Il est souvent plus simple de considérer une approximation gaussienne de la loi binomiale par le théorème de la limite centrale sous la condition que les nombres $n\pi_0$ et $n(1-\pi_0)$ soient suffisamment grands, en pratique, plus grands que 10.

La statistique de test :

$$\frac{|X - n\pi_0|}{\sqrt{n\pi_0(1-\pi_0)}}$$

suit une loi gaussienne centrée réduite et sa réalisation est comparée avec le $1 - \alpha/2$ -quantile de cette loi.

Test unilatéral On teste $H_0 = \pi \leq \pi_0$ contre l'hypothèse alternative $H_1 = \pi > \pi_0$ en supposant que la variable X suit une loi binomiale $\mathcal{B}(n, \pi)$. La statistique du test est la même mais comparée cette fois avec $le 1 - \alpha$ -quantile de la loi gaussienne pour un test unilatéral.

4 Comparaison de deux échantillons

L'objectif dans cette section est de comparer des caractéristiques de deux populations à l'aide d'échantillons extraits de chacune de celle-ci.

4.1 Deux échantillons indépendants

Comparaison de deux moyennes

Cette section considère deux variables X et Y gaussiennes de moyennes (μ_1, μ_2) et de variances (σ_1^2, σ_2^2) dont sont tirés des échantillons d'effectifs (n_1, n_2) ; \bar{X} (resp. \bar{Y}) désigne toujours l'estimateur de l'espérance de la variable X (resp. Y) et S_1^2 (resp. S_2^2) celui de la variance. Notons également :

$$S^2 = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2}.$$

σ_1 et σ_2 connus Le test de l'hypothèse $H_0 : \mu_1 = \mu_2$ est basé sur la loi des différences des moyennes. Sachant que \bar{X} et \bar{Y} sont des variables aléatoires gaussiennes de moyennes respectivement (μ_1, μ_2) et de variances $(\sigma_1^2/n_1, \sigma_2^2/n_2)$, la loi de la variable $(\bar{X} - \bar{Y})$ est gaussienne d'espérance nulle et de variance $(\sigma_1^2/n_1 + \sigma_2^2/n_2)$.

Tester $\mu_1 = \mu_2$ revient donc à tester la moyenne $\mu_1 - \mu_2 = 0$ d'une variable aléatoire $(\bar{X} - \bar{Y})$ gaussienne et de variance connue comme dans un test à un échantillon.

$\sigma_1 = \sigma_2$ inconnues Cette hypothèse d'égalité des variances peut être vérifiée avec le test ci-dessous. Dans ce cas, la variable aléatoire $(\bar{X} - \bar{Y})$ est gaussienne d'espérance nulle et de variance $\sigma^2(1/n_1 + 1/n_2)$. La variance doit être estimée par S^2 et la statistique de test est le quotient d'une variable gaussienne par une variable du chi-deux ; c'est une variable de Student à $(n_1 + n_2 - 2)$ degrés de liberté :

$$\frac{(\bar{X} - \bar{Y})}{S\sqrt{(1/n_1 + 1/n_2)}} \sim \mathcal{T}_{n_1 + n_2 - 2}.$$

Tester $\mu_1 = \mu_2$ revient donc à tester $\mu_1 - \mu_2 = 0$ d'une variable aléatoire $\bar{X} - \bar{Y}$ gaussienne et de variance inconnue comme dans un test à un échantillon.

$\sigma_1 \neq \sigma_2$ inconnues Lorsque les effectifs des deux échantillons sont élevés (supérieurs à 20), le test précédent peut être utilisé.

Comparaison de σ^1 et σ^2 Il est utile de tester l'égalité de deux variances $H_0 : \sigma^1 = \sigma^2$ préalablement au test d'égalité des moyennes. La statistique de test est S_1^2/S_2^2 qui ne doit être ni trop petite ni trop grande afin d'accepter H_0 . Ce rapport de deux variables suivant des lois du chi-deux est une variable aléatoire suivant une loi de Fisher à $(n_1 - 1)$ et $(n_2 - 1)$ degrés de liberté.

En pratique, la plus grande quantité, disons S_1^2 , est placée au numérateur et le test de niveau α conduit à rejeter H_0 si la réalisation de la statistique S_1^2/S_2^2 est plus grande que le quantile $\mathcal{F}_{(n_1-1), (n_2-2); \alpha/2}$ de la loi de Fisher.

Comparaisons de deux proportions

Les variables X et Y ne sont plus supposées gaussiennes mais binomiales de paramètres (n_1, π_1) et (n_2, π_2) . Le test qui suit n'est valide qu'approximativement pour des échantillons de tailles n_1 et n_2 suffisamment grandes car il est basée sur des approximations gaussiennes des lois binomiales.

L'hypothèse H_0 est $\pi_1 = \pi_2$ et on note :

$$P = \frac{X + Y}{n_1 + n_2}; \quad S_d^2 = P(1 - P)(1/n_1 + 1/n_2).$$

Sous H_0 , $\pi_1 = \pi_2 = \pi$ et, moyennant une approximation gaussienne, la variable $(X/n_1 - Y/n_2)$ suit une loi gaussienne de moyenne nulle et de variance $\pi(1 - \pi)(1/n_1 + 1/n_2)$; π est estimée par P et le test approché de niveau α conduit à rejeter H_0 si la réalisation de la statistique de test :

$$\frac{|X/n_1 - Y/n_2|}{\sqrt{P(1 - P)(1/n_1 + 1/n_2)}}$$

dépasse le quantile $u_{1-\alpha/2}$ de la loi gaussienne centrée réduite.

4.2 Deux échantillons appariés

Deux variables sont observées sur le même échantillon ou encore ce même échantillon est soumis à deux mesures successives de la même variable : des

copies sont soumises à une double correction, deux tests biologiques sont comparés sur le même ensemble de patients... Les mesures sont alors dites appariées ou en couple par opposition à des échantillons distincts et donc indépendants.

La variable D désigne la différence $X - Y$ de variables gaussiennes de moyennes μ_1, μ_2 et de même variance σ^2 ; D est donc supposée suivre une loi gaussienne de moyenne $\mu_1 - \mu_2$ et de variance σ^2 ; \bar{D} désigne l'estimateur de la moyenne de D et S_D celui de sa variance. Le test de $H_0 : \mu_1 = \mu_2$ contre $H_1 : \mu_1 \neq \mu_2$ revient à comparer la moyenne de D à la valeur nulle. Comme pour un test à un échantillon gaussien de variance inconnue, la statistique de test $\sqrt{n} \frac{\bar{D}}{S_D}$ suit une loi de Student à $(n - 1)$ degrés de liberté. L'hypothèse H_0 est rejetée si la réalisation de la statistique de test

$$\sqrt{n} \frac{|\bar{D}|}{S_D}$$

dépasse le quantile $t_{n-1; 1-\alpha/2}$ de la loi de Student.

Remarque Deux tests différents sont ainsi proposés pour comparer deux moyennes celui basé sur deux échantillons indépendants et celui basé sur des échantillons appariés. La différence est importante : la variance de la statistique de test intègre (S dans le cas indépendant), ou non (S_D dans le cas apparié), la variabilité des échantillons. Comme le test apparié est épargné par cette source de variabilité supplémentaire, il est plus *puissant* que le test sur échantillons indépendants ; ce test est à favoriser lorsque cela est expérimentalement possible.

5 Comparaison de plusieurs échantillons

La comparaison des moyennes de plusieurs échantillons se présente sous une forme générale en considérant les m échantillons comme les m classes d'unités statistiques prenant les modalités d'une variable qualitative X . Comparer les moyennes d'une variable quantitative Y revient alors à s'intéresser à l'influence possible de la variable qualitative X ou facteur sur la variable quantitative Y . Cette présentation est en fait un cas particulier de *modèle linéaire* couramment appelée **ANOVA** pour méthode d'*AN*alyse de la *VA*riance. La présentation générale est beaucoup plus élégante mais ne permet pas d'aller à l'essentiel, à savoir la présentation élémentaire et opérationnelle du test sans

doute le plus utilisé en biostatistique comme dans bien d'autres domaines. Les compléments ANOVA à 2 facteurs et plus, comparaisons multiples sont développés par ailleurs.

Attention aux subtilités sémantiques, ce test compare bien des moyennes en testant $H_0 : \mu_1 = \dots = \mu_k$ contre H_1 : au moins deux moyennes sont différentes. Néanmoins la statistique de test est basée sur la comparaison des deux variances inter et intra-classes d'où le nom de ce modèle.

Deux hypothèses importantes sont à vérifiées préalablement à l'analyse :

1. la variable Y suit une loi gaussienne,
2. les variances des sous-groupes (variances conditionnelles) sont identiques (homoscédasticité).

La première hypothèse est moins contraignante car l'ANOVA est robuste si les tailles des échantillons sont suffisantes. En revanche celle d'homoscédasticité est sensible ; elle peut être vérifiée par différents tests : Bartlett dans le cas gaussien strict, Levene sinon. Ils ne sont pas décrits ici mais présents dans la plupart des logiciels. Dans le cas de violation de ces hypothèses préalables, une approche non-paramétrique est préférable (Kruskal-Wallis).

5.1 Notations

Une variable aléatoire supposée gaussienne Y est observée sur m échantillons de tailles n_k . C'est aussi la partition de l'échantillon en m classes issue de la variable qualitative X ; \bar{y} désigne la moyenne générale, \bar{y}_k , la moyenne dans chaque classe ou pour chaque échantillon.

L'ANOVA est basée sur la décomposition de la variance :

$$\begin{aligned} \text{SST} &= \text{SSB} + \text{SSW} \\ \sum_{i=1}^n (y_i - \bar{y})^2 &= \sum_{k=1}^m n_k (\bar{y}_k - \bar{y})^2 + \sum_{k=1}^m \sum_{i=1}^{n_k} (y_{i,k} - \bar{y}_k)^2 \end{aligned}$$

ou SST traduit la variabilité totale, c'est la somme des carrés totaux. SSB traduit la variabilité des moyennes des groupes ou moyennes conditionnelles par rapport à la moyenne générale. C'est la variabilité inter-classe (*between*) ou expliquée par le modèle. SSW traduit la variabilité résiduelle à l'intérieur des groupes ou variabilité intra-classe (*within*).

5.2 Statistique de test

Trois autres quantités sont définies, les carrés moyens $MSB=SSB/(m-1)$ et $MSW=SSW/(n-m)$ ainsi que la statistique de test $F=MSB/MSW$. Plus cette quantité F est grande, plus la variance inter-classe l'emporte sur celle intra-classe et donc, au regard de la dispersion globale, la dispersion des moyennes des classes est d'autant plus significative.

Tout logiciel présente les résultats dans un tableau dit d'analyse de la variance à un facteur :

Source de variation	d.d.l.	Somme des carrés	Variance	F
Modèle (inter)	$m-1$	SSB	$MSB=SSB/(m-1)$	MSB/MSW
Erreur (intra)	$n-m$	SSW	$MSW=SSW/(n-m)$	
Total	$n-1$	SST		

Sous les hypothèses de normalité de Y , d'homoscédasticité (égalité des variances conditionnelles) et sous l'hypothèse $H_0 : \mu_1 = \dots = \mu_m$, la statistique de test est un rapport de variables suivant des lois du chi-deux, c'est donc une variable de Fisher à $(m-1)$ et $(n-m)$ degrés de liberté. Tout logiciel fournit la probabilité critique (P -valeur) que cette variable de Fisher dépasse la réalisation de la statistique de test.

Si H_0 est rejetée (P -valeur inférieure au seuil), une comparaison des moyennes par paires permet de juger de celles qui sont effectivement significativement différentes. Il s'agit alors de comparaisons multiples nécessitant des ajustements des seuils (Scheffé, Tukey...) avec l'affichage en parallèle des intervalles de confiance corrigés des estimations des moyennes.

Enfin, l'analyse de variance à deux facteurs est nettement plus complexe et sort du cadre de cet exposé.

6 Tests d'ajustement

Dans le paragraphe précédent, nous avons construit des tests portant sur un paramètre réel. Nous souhaitons désormais, à partir de l'observation d'un échantillon de taille n issu d'une loi de fonction de répartition F , tester si $F = F_0$ où F_0 est une fonction de répartition que l'on se donne. Il s'agit donc de tester que la loi des X_i est une loi donnée. Par exemple

- $\mathcal{N}(0, 1)$,
- $\mathcal{E}(5)$,
- si les X_i ne peuvent prendre que 3 valeurs, $\{1, 2, 3\}$, il s'agira, par exemple de tester que $P(X_i = i) = 1/3$ pour tout $i \in \{1, 2, 3\}$.

Lorsque les X_i sont des v.a.r. discrètes, on utilise le test du chi-deux, alors que dans le cas où F_0 est une fonction de répartition continue, on utilise plutôt le test de Kolmogorov.

6.1 Test du chi-deux

6.1.1 Problématique

Le test du chi-deux utilise des propriétés de la loi multinomiale. Il permet de juger si une hypothèse concernant la loi de probabilité d'une variable aléatoire est compatible avec la réalisation d'un échantillon. Deux cas sont à distinguer :

- la fonction de répartition F est entièrement spécifiée, et ses paramètres sont connus.
- on connaît seulement la forme de la loi de distribution, et ses paramètres sont estimés à partir d'un échantillon.

Soit (X_1, \dots, X_n) un n -échantillon d'une v.a.r. X prenant ses valeurs dans l'ensemble $\{x_1, \dots, x_k\}$. On se donne un vecteur de \mathbb{R}^k , $p = (p_1, \dots, p_k)$, tel que $p_1 + \dots + p_k = 1$. On souhaite tester l'hypothèse

$$H_0 : \text{pour tout } i \text{ de } 1 \text{ à } k, \mathbb{P}(X = x_i) = p_i$$

contre l'hypothèse

$$H_1 : \text{il existe } i \text{ de } 1 \text{ à } k \text{ tel que } \mathbb{P}(X = x_i) \neq p_i.$$

Test du χ^2 pour une loi uniforme discrète

Lors d'un lancer de dé, on cherche à déterminer si le dé est biaisé ou non au risque de 1%. L'hypothèse nulle est que le dé est non biaisé, c'est-à-dire que la loi de probabilité sous-jacente est uniforme :

x_i	1	2	3	4	5	6
$\mathbb{P}(X = x_i)$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$	$\frac{1}{6}$

Dans cet exemple, il y a $k = 6$ classes différentes et pour tout i variant de 1 à 6, $p_i = \frac{1}{6}$. Parmi les n observations à notre disposition, on note N_i le nombre

de variables parmi X_1, \dots, X_n qui prennent la valeur x_i , c'est-à-dire

$$N_i = \sum_{j=1}^n 1_{\{X_j=x_i\}}.$$

Statistique de test

La v.a.r. N_i correspond à l'effectif observé de la valeur x_i , tandis que np_i correspond à l'effectif théorique espéré de la valeur x_i (effectif calculé à partir de la distribution théorique F). Sous l'hypothèse H_0 , la fréquence observée N_i/n est "proche" de la fréquence théorique p_i pour tout i . En effet, par la LGN, elle converge en probabilité vers l'espérance de la v.a.r. $1_{\{X_1=x_i\}}$, qui n'est autre que p_i .

La statistique de test que l'on considère est :

$$D^2 = \sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i}.$$

Elle exprime une *distance* dite du chi-deux entre la distribution empirique et celle théorique sous H_0 . On admet que cette v.a.r. D^2 suit une loi du chi-deux à ν degrés de liberté, quelle que soit la loi considérée, lorsque le nombre n d'observations tend vers l'infini. Le nombre ν de degrés de liberté est égal à :

- $(k - 1)$ si la distribution théorique est entièrement déterminée, aucun paramètre n'ayant été estimé.
- $(k - 1 - r)$ si r paramètres ont dû être estimés à partir des observations, pour définir complètement la distribution.

Règle de décision :

$$\text{Rejet de } H_0 \iff \mathbb{P}_{H_0} \left(\sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i} > \sum_{i=1}^k \frac{(N_{i,obs} - np_i)^2}{np_i} \right) < \alpha.$$

Sinon, on garde l'hypothèse H_0 et on considère que la distribution théorique spécifiée est plausible, c'est-à-dire que $F = F_0$.

Exemple : On note les résultats sur 100 lancés et l'on obtient le tableau suivant :

Face	1	2	3	4	5	6
$N_{i,obs}$: Effectifs	7	18	26	15	18	16
np_i	16.67	16.67	16.67	16.67	16.67	16.67
$\frac{(N_{i,obs} - np_i)^2}{np_i}$	5.61	0.11	5.23	0.17	0.11	0.03

où les effectifs théoriques sont $np_i = 100 \times 1/6 = 16.67$ si le dé est équilibré (c'est-à-dire sous H_0). D'où la valeur de

$$D_{obs}^2 = \sum_{i=1}^k \frac{(N_{i,obs} - np_i)^2}{np_i} = 11.24.$$

Comme on a 6 classes, on utilise donc la loi du χ^2 à 5 degrés de liberté. À l'aide des tables, on trouve la valeur de la probabilité cherchée :

$$\mathbb{P}_{H_0}(D^2 > D_{obs}^2) = 0.0468.$$

Cette probabilité étant supérieure à $\alpha = 0.01$, alors l'hypothèse H_0 "le dé est équilibré" est acceptée au risque de $\alpha = 0.01$.

Test du chi-deux de contingence

Ce test est une application particulière du test d'adéquation du chi-deux car il est utilisé pour caractériser la liaison entre deux variables qualitatives. La question posée concerne l'indépendance en deux variables à partir d'observations consignées dans une table de contingence. Il s'agit donc de comparer les effectifs observés en croisant les deux variables à ceux qui seraient obtenus sous l'hypothèse H_0 : les variables sont indépendantes.

Les notations reprennent celles de la [description statistique](#) de deux variables qualitatives. Deux caractères ou variables qualitatives X (r modalités notées $x_1, \dots, x_\ell, \dots, x_r$) et Y (c modalités notées $y_1, \dots, y_h, \dots, y_c$) sont observées sur un même échantillon de taille n . La *table de contingence* associée a pour élément générique l'effectif conjoint $n_{\ell h}$; $n_{\ell+} = \sum_{h=1}^c n_{\ell h}$ et $n_{+h} = \sum_{\ell=1}^r n_{\ell h}$ sont les effectifs marginaux. La division des lignes et colonnes par ces effectifs marginaux fournit des distributions empiriques constituées de fréquences conditionnelles.

Ce sont les profils lignes : $\frac{n_{\ell h}}{n_{\ell+}}$ et les profils colonnes : $\frac{n_{\ell h}}{n_{+h}}$.

Pour le modèle probabiliste, les observations proviennent d'un échantillon bidimensionnelle. L'hypothèse à tester est que les deux marginales de cette loi sont indépendantes. Dans ce cas, les profils lignes seront tous très proches de la distribution empirique de Y et les profils colonnes de celle de X :

$$f_{h|\ell} = \frac{n_{\ell h}}{n_{\ell+}} \approx f_{+h} = \frac{n_{+h}}{n} \quad \text{et} \quad f_{\ell|h} = \frac{n_{\ell h}}{n_{+h}} \approx f_{\ell+} = \frac{n_{\ell+}}{n}$$

C'est équivalent à dire que les fréquences conjointes sont proches des produits des fréquences marginales :

$$f_{h\ell} = \frac{n_{\ell h}}{n} \approx f_{+h} f_{\ell+} = \frac{n_{+h}}{n} \frac{n_{\ell+}}{n}.$$

La statistique de test est la distance du chi-deux entre ces deux quantités, elles s'écrivent :

$$D_{\chi^2} = \sum_{\ell=1}^r \sum_{h=1}^c \frac{(f_{\ell h} - f_{\ell+} f_{+h})^2}{f_{\ell+} f_{+h}} = \sum_{\ell=1}^r \sum_{h=1}^c \frac{n_{\ell h}^2}{n_{\ell+} n_{+h}} - 1.$$

Pour n assez grand, nD_{χ^2} est approchée par une loi du χ^2 à $(r-1)(c-1)$ degrés de liberté. Il s'agit donc d'un test asymptotique qui nécessite, en pratique, des effectifs théoriques $\frac{n_{+h} n_{\ell+}}{n} > 5$.

6.2 Test de Kolmogorov

Le test de Kolmogorov est employé dans le cas où l'on veut tester que la loi des X_i a pour fonction de répartition théorique F_{th} , lorsque F_{th} est continue (loi normale, loi exponentielle).

On considère un n -échantillon (X_1, \dots, X_n) , de fonction de répartition F . On veut tester l'hypothèse $H_0 : F = F_{th}$ contre l'hypothèse $H_1 : F \neq F_{th}$. Pour cela, on introduit la fonction de répartition empirique pour le n -échantillon (X_1, \dots, X_n) :

$$F_n(t) = \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{\{X_i \leq t\}} = \begin{cases} 0 & \text{si } t < x_{(1)} \\ \frac{i}{n} & \text{si } x_{(i-1)} \leq t < x_{(i)} \\ 1 & \text{si } t > x_n \end{cases}$$

les $x_{(i)}$ désignant la valeur prise par X dans l'échantillon au rang i des valeurs rangées par ordre croissant (c'est en fait une fonction en escalier, dont la

taille de la i -ème marche fait $1/n$ en hauteur et $x_{(i+1)} - x_{(i)}$ en largeur). Cette fonction estime (sans biais) la fonction de répartition F des variables aléatoires X_i . En effet, par la LGN, on a pour tout $t \in \mathbb{R}$ la convergence en probabilité de $F_n(t)$ vers $F(t)$.

La statistique de test est la v.a.r. D_n définie par :

$$D_n = \sup_{t \in \mathbb{R}} |F_n(t) - F_{th}(t)|.$$

C'est le maximum des écarts entre les deux fonctions de répartition F_n et F_{th} .

Exemple : Soit X la v.a.r. dont on recherche l'ajustement à la loi exponentielle avec $\lambda = 4.10^{-4}$. On a procédé à un échantillonnage de taille $n = 10$ qui a donné le résultat suivant :

i	1	2	3	4	5
x_i	124	512	1004	1129	2007
i	6	7	8	9	10
x_i	2248	2619	3492	4750	5497

La figure 1 représente les fonctions de répartition empirique F_n et théorique F_{th} .

On calcule la valeur observée de la statistique de test : $D_{n,obs} = 0,152$. La règle de décision est alors la suivante :

$$\text{Rejet de } H_0 \iff \mathbb{P}_{H_0} \left(\sup_{t \in \mathbb{R}} |F_n(t) - F_{th}(t)| > D_{n,obs} \right) < \alpha.$$

Sinon, on garde l'hypothèse H_0 et on considère que la distribution théorique spécifiée est plausible, c'est-à-dire que $F = F_{th}$.

Suite de l'exemple : Pour $n = 10$, on lit dans les tables du test de Kolmogorov, la valeur du seuil k_α tel que

$$\mathbb{P}_{H_0} \left(\sup_{t \in \mathbb{R}} |F_n(t) - F_{th}(t)| > k_\alpha \right) = \alpha.$$

On trouve pour $\alpha = 0,1$, la valeur seuil $k_\alpha = 0,36866$. Or si $D_{n,obs} > k_\alpha$ alors on a :

$$\mathbb{P}_{H_0} \left(\sup_{t \in \mathbb{R}} |F_n(t) - F_{th}(t)| > D_{n,obs} \right) \leq \mathbb{P}_{H_0} \left(\sup_{t \in \mathbb{R}} |F_n(t) - F_{th}(t)| > k_\alpha \right) = \alpha.$$

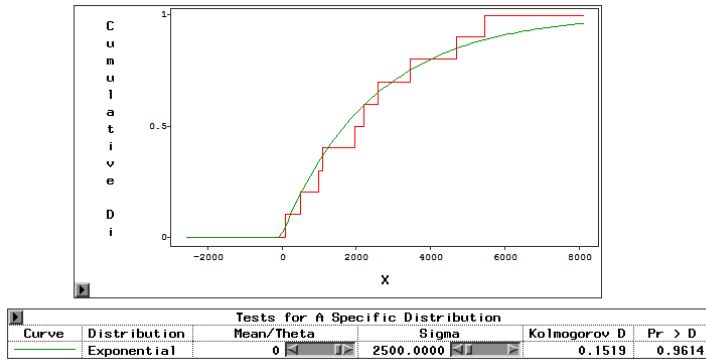


FIGURE 1 – Fonction de répartition empirique et théorique (loi exponentielle de paramètre $\lambda = 10^{-4}$) de la variable X ; test d'ajustement de Kolmogorov

La règle de décision se réécrit alors de la manière suivante :

$$\text{Rejet de } H_0 \iff D_{n,obs} > k_\alpha.$$

Dans cet exemple $D_{n,obs} = 0,152 < k_\alpha$ et au risque $\alpha = 0.1$, on donc accepte H_0 .

Remarque : Le test de Kolmogorov est préférable au test du chi-deux pour des variables continues car il présente un avantage considérable. En effet, la v.a.r. de décision D_n utilise l'échantillon tel qu'il se présente, tandis que le test du chi-deux appauvrit l'information en regroupant les données par classes et en assimilant les données d'une classe à une valeur centrale.

6.3 Droite de Henry

Une autre méthode couramment utilisée en pratique pour visualiser plutôt que tester si la loi d'une v.a.r. est normale est la droite de Henry. L'utilisation d'une échelle spéciale, appelée échelle galtonienne ou gaussio-arithmétique, permet de transformer les fonctions de répartition complexes de type gaussienne en droites. Pour ce faire, l'échelle des ordonnées est transformée par la fonction inverse de la fonction de répartition de la loi gaussienne.

Soit $X \sim \mathcal{N}(m, \sigma^2)$ et $Z \sim \mathcal{N}(0, 1)$. Désignons par F_X et F les fonctions

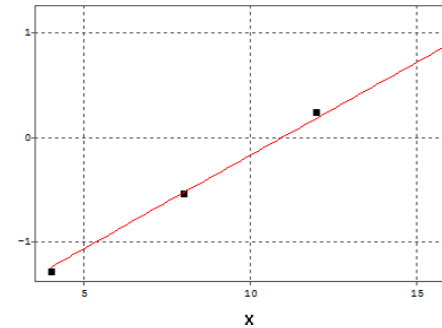


FIGURE 2 – Droite de Henry de la série agrégée des notes d'examen (y_i, x_i)

de répartition de X et Z , respectivement. Si l'on note x_i l'ensemble des valeurs prises par X , il est clair que l'on a $F_X(x_i) = F(y_i)$, où $y_i = (x_i - m)/\sigma$. Ainsi, une v.a.r. est normale si et seulement si les points de coordonnées (x_i, y_i) sont alignés sur la droite d'équation $y = (x - m)/\sigma$.

Exemple : Lors d'un examen noté sur 20, on obtient les résultats suivants : 10% des candidats ont obtenu moins de 4, 30% moins de 8, 60% moins de 12 et 80% moins de 16. On cherche à déterminer si la distribution des notes X est normale, et, si oui, ce que valent son espérance et son écart-type. On connaît donc 4 valeurs x_i , et, pour ces 4 valeurs, on connaît $F_X(x_i) = \mathbb{P}(X < x_i)$. Par la table de la loi normale, on détermine les y_i satisfaisant $F_X(x_i) = F(y_i)$. On obtient alors le tableau suivant, ainsi que le tracé des points de coordonnées (x_i, y_i) (figure 2) :

x_i	$F_X(x_i)$	y_i
4	0.1	-1,28
8	0.3	-0.525
12	0.6	0.255
16	0.8	0.84

Les points paraissent alignés. La droite coupe l'axe des abscisses au point $x = 11$ et le coefficient directeur est $(0.84 + 1.28)/12$, ce qui donnerait un écart-type de $12/2.12 = 5.7$. Cela laisse penser que la distribution est normale de paramètres $m = 11$ et d'écart-type $\sigma = 5,7$.

7 Tests non-paramétriques

Cette section décrit un ensemble de tests dits *non-paramétriques* car, basés sur des statistiques de rang, ils sont exploitables quelle que soit la loi des échantillons ; en anglais : *distribution free*. Lorsque les hypothèses d'un test paramétrique sont vérifiées, un test non-paramétrique est généralement moins puissant que son homologue, il peut conduire puis facilement à accepter à tort l'hypothèse H_0 . Leur usage ne se justifie donc pleinement que lorsque l'hypothèse sur la distribution du ou des échantillons (gaussiennes, binomiales) ne sont pas vérifiées et si le ou les échantillons sont de petites tailles. Ils sont décrits de façon succincte sans application numérique car tout logiciel fournit les probabilités critiques ou P -valeurs.

7.1 Test de Wilcoxon-Mann-Whitney pour deux échantillons indépendants

Ce test repose sur l'idée que deux séries de valeurs mélangées et ordonnées par valeurs croissantes, doivent conduire à un mélange homogène si l'hypothèse H_0 d'identité des distributions est vérifiée. Soit (x_1, \dots, x_n) et (y_1, \dots, y_m) deux échantillons, les deux suites étant fusionnées et réordonnées, R_1 (resp. R_2) désigne la somme des rangs des observations de l'échantillon 1 (resp. l'échantillon 2) dans cette suite en affectant éventuellement un rang moyen en cas d'égalité entre valeurs.

La statistique U est définie par la plus petite des deux valeurs :

$$U_1 = nm + \frac{n(n+1)}{2} - R_1 \quad \text{et} \quad U_2 = nm + \frac{m(m+1)}{2} - R_2.$$

La loi de U est tabulée (effectifs < 16) ou, lorsque n_1 et n_2 sont suffisamment grands, U suit une loi gaussienne de moyenne $(nm + 1)/2$ et de variance $nm(n + m + 1)/12$; le test se ramène au test de la moyenne d'un échantillon gaussien.

7.2 Test de Wilcoxon pour échantillons appariés

Comme pour des échantillons gaussiens, la variable considérée est celle $X - Y$ des différences. Le test dit des signes consiste à dénombrer le nombre K de différences positives. Sous l'hypothèse d'homogénéité, k suit une loi binomiale $\mathcal{B}(n, 0.5)$. Ce test est facile à mettre en œuvre mais s'avère moins

puissant que le test de Wilcoxon pour données appariées.

Soit (x_i, y_i) n paires d'observations de deux variables X et Y sur le même échantillon ; (d_1, \dots, d_n) désigne la suite des différences qui sont ordonnées par ordre croissant des valeurs absolues $|d_i|$; R^+ (resp. R^-) désigne la somme des rangs correspondants aux valeurs positives (resp. négatives) ; R désigne alors la plus petite des valeurs R^+ et R^- . En cas de différence nulle $d_i = 0$, l'effectif n est réduit d'une unité et en cas d'ex-æquo, le rang moyen est utilisé.

La loi de la statistique R est tabulée et pour des valeurs plus grandes R suit approximativement une loi gaussienne de moyenne $n(n + 1)/4$ et de variance $n(n + 1)(2n + 1)/24$.

7.3 Test de Kruskal-Wallis pour plusieurs échantillons

Le test de Kruskal-Wallis est la généralisation à m échantillons de celui de Wilcoxon-Mann-Whitney auquel il est équivalent pour $k = 2$. Chaque observation $x_{i,k}$ appartient à l'un k des m échantillons et est remplacée par son rang $r_{i,k}$ dans la suite ordonnée de toutes les valeurs ; $\bar{r} = (n + 1)/2$ et \bar{r}_k désignent respectivement la moyenne globale des rangs et la moyenne des rangs pour les observations de l'échantillon k . La statistique H de Kruskal-Wallis est définie de façon analogue à celle d'une variance inter-classe par :

$$H = \frac{12}{n(n+1)} \sum_{k=1}^m n_k (\bar{r}_k - \bar{r})^2 = \frac{12}{n(n+1)} \sum_{k=1}^m \frac{S_k^2}{n_k} - 3(n+1)$$

où S_k est la somme des rangs du k -ième échantillon, n_k son effectif.

La loi de H est tabulée pour de petits effectifs et, pour des effectifs plus importants ($n_k > 5$ pour tout k), la loi de H est approchée par une loi du chi-deux à $m - 1$ degrés de liberté.