

TD3 : Test d'hypothèses



Version web

Ex 1

Selon le choix de l'hypothèse

trhY

★ ★

Dans une grande entreprise, le salaire moyen annuel des hommes ayant entre 3 et 5 ans d'ancienneté est de 28 000 euros. Pour préparer une négociation sur la parité des salaires hommes femmes, on fait un sondage sur 10 femmes qui donne les résultats suivants (en milliers d'euros) :

24 27 31 19 26 27 22 15 33 21

A l'ouverture de la réunion, Monsieur A, délégué du personnel annonce : « Le sondage montre qu'il n'y a pas lieu de penser que le salaire des femmes est différent de celui des hommes ». Il est aussitôt interrompu par Madame B, Directrice des Ressources Humaines, qui dit « Pas du tout, le sondage prouve que le salaire des femmes est inférieur à celui des hommes ». Qu'en pensez-vous ?

Pour argumenter les réponses, on pourra se poser les questions suivantes pour Monsieur A et Madame B.

1. Que doit-on supposer sur la distribution des salaires des femmes pour pouvoir faire des tests statistiques ?



On doit supposer que la distribution dans la population suit une loi normale.

2. Si on note $\mu_0 = 28$ et $H_0: \mu = \mu_0$ l'hypothèse nulle de Monsieur A / Madame B, quelle est l'hypothèse alternative pour chacun ?



Monsieur A fait son test sur le jeu d'hypothèses : $\begin{cases} \mu = 28 \\ \mu \neq 28 \end{cases}$, partant du principe que la moyenne de salaire des femmes peut être aussi bien supérieur qu'inférieur à la moyenne.

Quant à Mme B, elle fait son test sur les hypothèses : $\begin{cases} \mu = 28 \\ \mu < 28 \end{cases}$, partant du principe qu'il est improbable que les femmes aient une moyenne salariale supérieure.

3. Déterminer la variable de décision et sa loi sous H_0 .



On pose $Z = \frac{\bar{X}-28}{\frac{s}{\sqrt{10}}}$, cette variable suit une loi $St(9)$ d'après le cours.

4. Déterminer la région critique pour un risque de première espèce $\alpha = 5\%$.



Par lecture de table, on détermine la région critique du test bilatéral de M. A : $W_A =] - \infty; -2.26[\cup] 2.26; +\infty[$.

Pour Mme B, le test unilatéral à gauche donne la région critique $W_B =] - \infty; -1.83[$.

5. Calculer une réalisation de la variable de décision.



Après calculs sur cet échantillon, on obtient $\bar{x}_{obs} = 24.5$ et $s_{obs} \approx 5.46$.

6. Peut-on dire que Monsieur A ou Madame B se trompe ?



Ex 2

Contrôle qualité

aUn2

★ ★

Désirant juger le travail d'un ouvrier ajusteur, un chef d'atelier prélève un échantillon de 50 pièces métalliques dans sa production. On note X l'épaisseur de ses pièces. L'objectif est d'avoir $\mathbb{E}(X) = 5 \text{ mm}$. Les résultats des mesures sur cet échantillon sont portés dans le tableau suivant :

n_i	5	15	20	10
x_i en mm	4.8	4.9	5.0	5.1

1. Cette vérification permet-elle de conclure que le résultat est conforme aux exigences, au seuil de confiance de 99% ?



$$\begin{aligned}\bar{x} &= \frac{1}{50} \sum_{i=1}^4 n_i x_i \\ &= \frac{1}{50} (5 \times 4.8 + 15 \times 4.9 + 20 \times 5.0 + 10 \times 5.1) \\ &= 4.98\end{aligned}$$

donc une estimation sans biais de l'épaisseur moyenne des pièces est $\bar{x} = 4.98 \text{ mm}$.
De plus, la variance observée dans cet échantillon est :

$$\begin{aligned}\sigma^2 &= \frac{1}{50} \sum_{i=1}^4 n_i (x_i - \bar{x})^2 \\ &= \frac{1}{50} (5 \times (4.8 - 4.98)^2 + 15 \times (4.9 - 4.98)^2 + 20 \times (5.0 - 4.98)^2 + 10 \times (5.1 - 4.98)^2) \\ &= 0.081\end{aligned}$$

donc une estimation sans biais de la variance de l'épaisseur des pièces est $s^2 = \frac{50}{49} \sigma^2 = 0.083$.
On réalise le test d'hypothèse suivant :

$$H_0 : \mathbb{E}(X) = 5$$

$$H_1 : \mathbb{E}(X) \neq 5$$

avec un risque de première espèce de 1%.

La variable de décision est $Z = \frac{\bar{x} - 5}{\frac{s}{\sqrt{n}}} \sim \mathcal{N}(0, 1)$.

On fait un test bilatéral, donc on rejette H_0 si $|Z| > z_{\frac{\alpha}{2}} = 2.58$ par lecture de la table de la loi normale. Or la valeur observée est $Z_{obs} = \frac{4.98 - 5}{\frac{\sqrt{0.083}}{\sqrt{50}}} = -2.33$. Donc on ne rejette pas H_0 . On peut donc conclure que le résultat est conforme aux exigences, au seuil de confiance de 99%.

2. Quel risque de première espèce devrait-on prendre pour que la prise de décision soit différente ?



Pour que la prise de décision soit différente, il faudrait que la valeur critique soit 2.33, ce qui correspond, par lecture de table, à un risque de première espèce de 1.98%.

Ex 3

Test du comportement des clients

r5Yy

Soit X le nombre d'incidents de paiements pour un crédit à la consommation observés sur la durée du prêt. On suppose que X suit une loi de Poisson de paramètre λ positif. On dispose d'un échantillon de n clients appartenant à la banque A.

1. Donner un estimateur de λ centré et convergent.
2. On désire tester l'hypothèse nulle selon laquelle les clients de la banque A sont faiblement risqués sous la forme suivante : $\begin{cases} \lambda = 1 \\ \lambda > 1 \end{cases}$.

Pour un échantillon de 100 clients de la banque A, nous observons les résultats suivants :

Incidents :	0	1	2	3	4	5	6	Total
Effectifs :	38	31	16	10	2	2	1	100

Que peut-on conclure pour un risque de première espèce de 5% ? Sur un même graphique, superposer une représentation graphique de la loi de Poisson de paramètre $\lambda = 1$ et l'histogramme des effectifs observés.

Ex 4

Contrôle qualité

3GmQ

Une entreprise fabrique des pièces en sous-traitance. Au sein d'une démarche qualité, toutes les machines ont été systématiquement révisées et on a défini une nouvelle organisation dans l'atelier : les tâches de contrôle sont réparties à chaque étape du processus de fabrication et le taux de pièces défectueuses est tombé à 1%.

Quelques mois plus tard, une opération de contrôle est effectuée pour vérifier si la norme de 1% (hypothèse H_0) de pièces défectueuses reste valable. Sur les 5 000 pièces contrôlées 100 s'avèrent défectueuses, soit 2% (hypothèse H_1).

Mme de Mainard, chef d'entreprise, décide que si l'hypothèse nulle est vérifiée, elle ne modifiera plus son processus de production (décision D_0) et au contraire, si c'est l'hypothèse alternative, elle entreprendra une action de sensibilisation des salariés de cet atelier au problème de la qualité (décision D_1).

Pour choisir entre ces deux hypothèses, elle tire un échantillon de 1 500 pièces.

1. Si la chef d'entreprise se fixe un risque de 1% d'entreprendre une action de sensibilisation des salariés à tort, quel sera le taux critique de pièces défectueuses qui fera prendre une décision ?



On réalise les premières étapes d'un test de conformité d'une proportion :

(a) Hypothèses : $\begin{cases} H_0 : p = 0.01 \\ H_1 : p > 0.01 \end{cases}$

(b) Variable de décision : $Z = \frac{F - 0.01}{\sqrt{\frac{0.01 \times 0.99}{1500}}} \sim \mathcal{N}(0, 1)$

(c) Zone de rejet : $W = [2.326; +\infty[$ pour une erreur de première espèce $\alpha = 1\%$

(d) proportion critique : on cherche p_C tel que $\frac{F - 0.01}{\sqrt{\frac{0.01 \times 0.99}{1500}}} = 2.326$ et on trouve $p_C = 0.016 = 1,6\%$.

Au delà de 1,6% de pièces défectueuses observées, on rejette l'hypothèse H_0 avec un risque de première espèce $\alpha = 1\%$.

2. Si dans l'échantillon prélevé, le nombre de pièces défectueuses est 18, quelle sera la décision de la chef d'entreprise ?



On a $F_{obs} = \frac{18}{1500} = 0.012$: la décision prise est donc D_0 (on ne rejette pas H_0).

3. Calculer alors le risque de l'acheteur, c'est-à-dire ne pas modifier le processus de production alors qu'on le devrait. Comment s'appelle ce risque ?



On cherche la probabilité de prendre la décision D_0 par erreur, c'est-à-dire si H_1 est vraie. Sous l'hypothèse $p = 0.02$, on a la variable $Z_2 = \frac{F - 0.02}{\sqrt{\frac{0.02 \times 0.98}{1500}}}$ qui suit une loi $\mathcal{N}(0, 1)$.

Ex 5

Contrôle qualité : test de proportion

6YqE

Une machine de production fabrique des pièces en grande série. Lorsque la machine est bien réglée, la proportion de pièces défectueuses est $p_0 = 0.05$. Lorsqu'elle est dérégulée, cette proportion passe à $p_1 = 0.10$.

Pour vérifier l'état de la machine, on prélève un échantillon aléatoire de $n = 100$ pièces dans la production, et on observe $k = 8$ pièces défectueuses.

On note $\hat{p} = \frac{k}{n}$ la fréquence observée de pièces défectueuses et on souhaite tester :

$$H_0 : p = p_0 = 0.05 \quad (\text{la machine est bien réglée})$$

$$H_1 : p = p_1 = 0.10 \quad (\text{la machine est dérégulée})$$

1. Justifier que, sous H_0 , la statistique de test

$$T_n = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$$

suit approximativement une loi $\mathcal{N}(0, 1)$ pour n assez grand.



Soit X_i la variable indicatrice valant 1 si la i -ème pièce est défectueuse, 0 sinon. On a $X_i \sim \mathcal{B}(p)$ et $\hat{p} = \frac{1}{n} \sum_{i=1}^n X_i$.

D'après le théorème central limite :

$$\frac{\hat{p} - \mathbb{E}[\hat{p}]}{\sqrt{\text{Var}(\hat{p})}} = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(0, 1)$$

Sous H_0 , on a $p = p_0$, donc :

$$T_n = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \stackrel{H_0}{\sim} \mathcal{N}(0, 1) \text{ approximativement}$$

Les conditions d'application sont vérifiées car :

- $np_0 = 100 \times 0.05 = 5 \geq 5$
- $n(1 - p_0) = 100 \times 0.95 = 95 \geq 5$

2. On note π la valeur critique : on rejette H_0 si $\hat{p} > \pi$. En utilisant la définition du risque de première espèce, exprimer α en fonction de π , p_0 et n .



Le risque de première espèce est la probabilité de rejeter H_0 alors qu'elle est vraie :

$$\begin{aligned} \alpha &= \mathbb{P}_{H_0}(\hat{p} > \pi) \\ &= \mathbb{P}_{H_0} \left(\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > \frac{\pi - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \right) \\ &= \mathbb{P} \left(Z > \frac{\pi - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \right) \quad \text{où } Z \sim \mathcal{N}(0, 1) \\ &= 1 - \Phi \left(\frac{\pi - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \right) \end{aligned}$$

Avec $p_0 = 0.05$ et $n = 100$:

$$\alpha = 1 - \Phi \left(\frac{\pi - 0.05}{\sqrt{\frac{0.05 \times 0.95}{100}}} \right) = 1 - \Phi \left(\frac{\pi - 0.05}{0.0218} \right)$$

3. Pour un seuil $\alpha = 0.05$, déterminer la valeur critique π et en déduire la région d'acceptation de l'hypothèse H_0 en termes de fréquence observée.



On cherche π tel que $\alpha = 0.05$.

D'après la question précédente :

$$1 - \Phi \left(\frac{\pi - 0.05}{0.0218} \right) = 0.05$$

Donc :

$$\Phi \left(\frac{\pi - 0.05}{0.0218} \right) = 0.95$$

Or $\Phi(1.645) = 0.95$, donc :

$$\frac{\pi - 0.05}{0.0218} = 1.645$$

D'où :

$$\pi = 0.05 + 1.645 \times 0.0218 = 0.05 + 0.0359 = 0.0859$$

Région d'acceptation de H_0 : $\hat{p} \leq 0.0859$, soit environ 8.6%.

Région de rejet de H_0 : $\hat{p} > 0.0859$.

4. Au vu des données ($\hat{p} = 0.08$), peut-on conclure que la machine est bien réglée au seuil $\alpha = 5\%$? Calculer également la p -valeur associée à ce test.



On observe $\hat{p} = \frac{8}{100} = 0.08$.

Décision : Puisque $\hat{p} = 0.08 < 0.0859 = \pi$, on se situe dans la région d'acceptation. **On ne rejette pas H_0 au seuil $\alpha = 5\%$.**

Conclusion : Les données ne permettent pas de conclure que la machine est dérégulée. Au niveau de confiance de 95%, on peut considérer que la machine est bien réglée.

Calcul de la p -valeur :

La p -valeur est la probabilité d'observer une valeur au moins aussi extrême que celle observée sous H_0 :

$$\begin{aligned} p\text{-valeur} &= \mathbb{P}_{H_0}(\hat{p} \geq 0.08) \\ &= \mathbb{P}\left(Z \geq \frac{0.08 - 0.05}{0.0218}\right) \\ &= \mathbb{P}(Z \geq 1.376) \\ &= 1 - \Phi(1.376) \\ &\approx 0.084 \end{aligned}$$

La p -valeur est d'environ 8.4%, ce qui est supérieur à $\alpha = 5\%$, confirmant la non-rejet de H_0 .

5. Calculer le risque de seconde espèce β , c'est-à-dire la probabilité d'accepter H_0 alors que la machine est effectivement dérégulée.



Le risque de seconde espèce est :

$$\begin{aligned} \beta &= \mathbb{P}_{H_1}(\hat{p} \leq \pi) \\ &= \mathbb{P}_{H_1}(\hat{p} \leq 0.0859) \end{aligned}$$

Sous H_1 , on a $p = p_1 = 0.10$, donc :

$$\frac{\hat{p} - p_1}{\sqrt{\frac{p_1(1-p_1)}{n}}} \sim \mathcal{N}(0, 1)$$

L'écart-type sous H_1 est :

$$\sigma_1 = \sqrt{\frac{0.10 \times 0.90}{100}} = \sqrt{0.0009} = 0.03$$

Donc :

$$\begin{aligned} \beta &= \mathbb{P}_{H_1}\left(\frac{\hat{p} - 0.10}{0.03} \leq \frac{0.0859 - 0.10}{0.03}\right) \\ &= \mathbb{P}\left(Z \leq \frac{-0.0141}{0.03}\right) \\ &= \mathbb{P}(Z \leq -0.47) \\ &= \Phi(-0.47) \\ &= 1 - \Phi(0.47) \\ &\approx 0.319 \end{aligned}$$

Le risque de seconde espèce est d'environ **32%**.

Interprétation : Si la machine est effectivement dérégulée (avec $p = 0.10$), il y a environ 32% de chances de ne pas le détecter avec ce plan de contrôle.

La **puissance du test** est $1 - \beta \approx 0.68$, soit 68%.

6. On souhaite désormais dimensionner le plan de contrôle. Quelle taille d'échantillon n permettrait de rejeter H_0 avec une fréquence observée de $\hat{p} = 0.08$, tout en conservant un risque de première espèce $\alpha = 5\%$?

Interpréter ce résultat : pourquoi faut-il augmenter la taille de l'échantillon ?



On cherche n tel que $\hat{p} = 0.08$ soit dans la région critique pour $\alpha = 5\%$.

La condition de rejet est :

$$\frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} > z_{1-\alpha}$$

avec $z_{0.95} = 1.645$.

On veut que $\hat{p} = 0.08$ soit juste à la limite, donc :

$$\frac{0.08 - 0.05}{\sqrt{\frac{0.05 \times 0.95}{n}}} = 1.645$$

D'où :

$$\frac{0.03}{\sqrt{\frac{0.0475}{n}}} = 1.645$$

$$0.03 = 1.645 \times \sqrt{\frac{0.0475}{n}}$$

$$\frac{0.03}{1.645} = \sqrt{\frac{0.0475}{n}}$$

$$\left(\frac{0.03}{1.645}\right)^2 = \frac{0.0475}{n}$$

$$n = \frac{0.0475}{\left(\frac{0.03}{1.645}\right)^2} = \frac{0.0475 \times (1.645)^2}{(0.03)^2}$$

$$n = \frac{0.0475 \times 2.706}{0.0009} = \frac{0.1285}{0.0009} \approx 143$$

Il faudrait un échantillon d'au moins **143 pièces**.

Interprétation :

- Avec $n = 100$, une fréquence de 8% n'est pas statistiquement significative car elle pourrait résulter de fluctuations d'échantillonnage.
- En augmentant la taille de l'échantillon à $n = 143$, on réduit la variabilité d'échantillonnage, ce qui permet de détecter des écarts plus faibles.
- Plus l'échantillon est grand, plus on peut détecter de petites différences entre p_0 et p_1 .
- Compromis à trouver entre : précision du test vs. coût du contrôle.

7. **[Bonus]** Représenter graphiquement sur un même schéma :

- la densité de probabilité de $\hat{p}$ sous H_0
- la densité de probabilité de $\hat{p}$ sous H_1
- la région critique
- les risques α et β

Ce graphique permet de visualiser la *puissance* du test.



Voici les éléments à représenter :

Distributions :

- Sous H_0 : $\hat{p} \sim \mathcal{N}(0.05, \sqrt{0.0475/100})$ (courbe centrée en 0.05)
- Sous H_1 : $\hat{p} \sim \mathcal{N}(0.10, \sqrt{0.09/100})$ (courbe centrée en 0.10)

Région critique : Zone $\hat{p} > 0.0859$ (à droite de la valeur critique)

Visualisation des risques :

- α (risque de 1ère espèce) : aire sous la courbe de H_0 à droite de $\pi = 0.0859$ (zone hachurée en rouge) $\approx 5\%$
- β (risque de 2ème espèce) : aire sous la courbe de H_1 à gauche de $\pi = 0.0859$ (zone hachurée en bleu) $\approx 32\%$
- Puissance = $1 - \beta$: aire sous la courbe de H_1 à droite de $\pi \approx 68\%$

Code pour un tableur ou Python :

On peut générer ce graphique avec les valeurs de $\hat{p}$ allant de 0 à 0.15, et calculer les densités correspondantes.

Ex 6

Contrôle qualité

aW1E

Un pharmacien de banlieue désire être approvisionné par un grossiste chaque soir avant 18h00. Le grossiste doit donc estimer le temps de trajet moyen de son coursier entre le dépôt et la pharmacie. Pour cela, il relève le temps de parcours (en minutes) pendant 13 trajets sur un itinéraire fixé :

20 30 45 39 16 25 23 19 40 21 30 19 40

On supposera que ce temps est distribué selon une loi normale.

1. Donner une estimation ponctuelle de la moyenne du temps de parcours μ et de son écart-type s en rappelant les estimateurs utilisés et leurs propriétés.
2. Le pharmacien affirme qu'il faut en moyenne 25 minutes pour faire le trajet, alors que le grossiste, quant à lui, pense qu'il faut en moyenne 30 minutes.
 - (a) On cherche à tester $H_0: \mu = 25$ contre $H_1: \mu = 30$. Préciser la variable de décision et sa loi sous H_0 .
 - (b) Déterminer la région critique pour un seuil $\alpha = 5\%$.
 - (c) Calculer le risque de seconde espèce, c'est-à-dire la probabilité de faire une erreur sachant que H_1 est vraie.
 - (d) Avec un risque de première espèce $\alpha = 5\%$, peut-on accepter H_0 à partir de ces données ?

Ex 7

Expérience scientifique

7aqI

En 1908, en France, Binet publie une étude concernant la mesure de l'intelligence des enfants. Son échelle d'intelligence se base sur un certain nombre d'épreuves classées dans un ordre croissant de difficulté. On peut donc classer l'enfant dans une des 5 catégories suivantes : retardé de 2 ans, retardé d'1 an, régulier, avancé de 1 an et avancé de 2 ans. Voici les données, en fréquences relatives, que Binet obtient sur un échantillon de 192 enfants :

-2	-1	0	+1	+2
0,06	0,23	0,48	0,22	0,01

Des chercheurs, un américain (Goddard) et un allemand (Bobertag) font passer le test de Binet à un échantillon de 1547 enfants américains (Goddard) et 228 enfants allemands (Bobertag). Ils obtiennent les effectifs suivants :

	-2	-1	0	+1	+2
Goddard	294	309	557	325	62
Bobertag	6	40	119	57	6

Avec un niveau de confiance de 95%, peut-on dire que les données chez les enfants américains et allemands sont cohérentes avec la distribution obtenue par Binet chez les enfants français ?



Les effectifs théoriques sous hypothèse de distribution de Binet :

	-2	-1	0	+1	+2
Goddard	92.82	355.81	742.56	340.34	15.47
Bobertag	13.68	52.44	109.44	50.16	2.28

On fait un test du χ^2 avec une variable de décision qui suit une loi du $\chi^2(4)$ la valeur critique est 9,49 pour une erreur de première espèce $\alpha = 5\%$: la valeur observée pour Goddard est 629,21, celle pour Bobertag est 15.09, on rejette l'hypothèse H_0 pour les deux test.

Ex 8

Test d'adéquation à une loi

DkOf

★ ★

Une enquête a été réalisée auprès de 100 demandeurs d'emploi sur le nombre d'entretiens d'embauche obtenus ces deux derniers mois. Elle donne les résultats suivants :

Nombre d'entretiens :	0	1	2	3	4	5	6 et plus
Effectifs observés :	15	30	27	13	9	5	1

Peut-on dire que le nombre d'entretiens d'embauche est distribué selon une loi de Poisson ? On utilisera un seuil de significativité $\alpha = 5\%$.



Préambule : Rappels sur le test d'adéquation du χ^2

Le test d'adéquation du χ^2 permet de vérifier si un échantillon suit une loi théorique donnée.

Principe : On compare les effectifs observés o_i avec les effectifs théoriques e_i calculés à partir de la loi supposée. Si les écarts sont trop importants, on rejette l'hypothèse.

Conditions d'application :

- Les observations sont indépendantes
- Tous les effectifs théoriques $e_i \geq 5$ (sinon regrouper les classes)
- La taille de l'échantillon est suffisante ($n \geq 30$)

ÉTAPE 1 : FORMULATION DES HYPOTHÈSES

Soit X la variable aléatoire représentant le nombre d'entretiens d'embauche obtenus par un demandeur d'emploi.

- **Hypothèse nulle** H_0 : X suit une loi de Poisson de paramètre λ
- **Hypothèse alternative** H_1 : X ne suit pas une loi de Poisson

Il s'agit d'un test bilatéral.

ÉTAPE 2 : ESTIMATION DU PARAMÈTRE λ

Sous H_0 , si $X \sim \mathcal{P}(\lambda)$, alors $\mathbb{E}(X) = \lambda$.

On estime λ par la moyenne empirique des observations :

$$\hat{\lambda} = \bar{x} = \frac{1}{n} \sum_i x_i \cdot o_i$$

Calcul détaillé :

$$\begin{aligned} \hat{\lambda} &= \frac{1}{100} (0 \times 15 + 1 \times 30 + 2 \times 27 + 3 \times 13 + 4 \times 9 + 5 \times 5 + 6 \times 1) \\ &= \frac{1}{100} (0 + 30 + 54 + 39 + 36 + 25 + 6) \\ &= \frac{190}{100} \\ &= 1.90 \end{aligned}$$

On estime donc $\lambda \approx 1.90$.

ÉTAPE 3 : CALCUL DES EFFECTIFS THÉORIQUES

Pour une loi de Poisson $\mathcal{P}(1.90)$, la probabilité d'obtenir k entretiens est :

$$P(X = k) = \frac{e^{-1.90} \times (1.90)^k}{k!}$$

Les effectifs théoriques sont : $e_k = n \times P(X = k) = 100 \times P(X = k)$

Ex 9

Contrôle de qualité

XwAO

★ ★ ★

Pour simuler le tirage aléatoire uniforme avec un tableur, on peut utiliser la commande

`=ALEA.ENTRE.BORNES(1;8)`

qui génère un nombre entier aléatoire entre 1 et 8.

À l'aide d'un test du χ^2 et une simulation de 500 tirages de dé, peut-on affirmer que cette commande simule correctement un tirage uniforme? On pourra illustrer le résultat à l'aide d'un histogramme et calculer la p -valeur de ce test.

Ex 10

Test d'indépendance

5FLs

On veut étudier la liaison entre les caractères : «être fumeur» (plus de 20 cigarettes par jour, pendant 10 ans) et «avoir un cancer de la gorge», sur une population de 1000 personnes, dont 500 sont atteintes d'un cancer de la gorge. Voici les résultats observés :

Observé	cancer	non cancer	TOTAL
fumeur	342	258	600
non fumeur	158	242	400
TOTAL	500	500	1000

Faire un test d'indépendance pour établir la liaison entre ces caractères.



Mise en oeuvre du test :

- On définit un risque : 5%. Pour étudier la dépendance de ces caractères faisons l'hypothèse H_0 : «les deux caractères sont indépendants » et voyons ce qui se passerait sous cette hypothèse. Notons les événements :

— C : «avoir un cancer dans la population observée»

— F : «être fumeur dans la population observée»

Si les événements F et C sont indépendants, alors : $P(F \cap C) = P(F) \cdot P(C)$ et de même pour les trois autres possibilités : $P(\bar{C} \cap F)$, $P(\bar{C} \cap \bar{F})$, $P(C \cap \bar{F})$, quantités que l'on peut donc calculer sous H_0 :

$P(F) = \frac{600}{1000}$, $P(C) = \frac{500}{1000}$, $P(F) \cdot P(C) = \frac{3}{10}$, alors l'effectif théorique correspondant à la catégorie «fumeur et cancéreux» est de 300.

- On en déduit le tableau théorique sous H_0 :

Théorique	cancer	non cancer	marge
fumeur	300	300	600
non fumeur	200	200	400
marge	500	500	1000

- On calcule alors la valeur de $s = \sum_{i=1}^{i=4} \frac{(O_i - T_i)^2}{T_i}$: on obtient : $s = 34.73$. On a précisé le risque de %, mais pour $\alpha = 0,001$, on lit dans la table du khi-deux à un degré de liberté : $P[\chi^2 \geq 10.83] = 0.001$ et le χ^2 calculé est 34.73!
- On décide de rejeter H_0 . Ainsi, en rejetant l'hypothèse de l'indépendance des caractères «être fumeur» et «avoir un cancer de la gorge», on a moins d'une chance sur 1000 de se tromper, puisque moins d'un tableau possible sur mille conduit à un calcul de χ^2 plus grand que 10.83; beaucoup moins sans doute, conduiraient à un calcul de χ^2 plus grand que 34.73.

Ex 11

Comparaison d'homogénéité des évaluations

qAu0

★ ★ ★

Deux départements Mesures Physiques d'IUT utilisent le même référentiel de compétences pour évaluer leurs étudiants en statistiques. On souhaite comparer l'homogénéité de leur système de notation.

On suppose que les notes dans chaque IUT suivent une loi normale. Pour effectuer cette comparaison, on prélève :

- un échantillon de $n_1 = 10$ notes d'étudiants de l'IUT 1, qui donne un écart-type empirique $s_1 = 5$ points
- un échantillon de $n_2 = 25$ notes d'étudiants de l'IUT 2, qui donne un écart-type empirique $s_2 = 3$ points

Contexte : Une différence significative d'écart-type pourrait indiquer :

- des critères d'évaluation plus ou moins stricts
- une hétérogénéité différente dans les groupes d'étudiants
- des méthodes pédagogiques ayant des impacts différents

1. Avant de réaliser le test, calculer le coefficient de variation pour chaque IUT si on vous donne les moyennes suivantes : $\bar{x}_1 = 12.5$ et $\bar{x}_2 = 13.2$.

Rappel : le coefficient de variation est $CV = \frac{s}{\bar{x}} \times 100\%$. Que peut-on en conclure intuitivement ?



Calcul des coefficients de variation :

Pour l'IUT 1 :

$$CV_1 = \frac{s_1}{\bar{x}_1} \times 100 = \frac{5}{12.5} \times 100 = 40\%$$

Pour l'IUT 2 :

$$CV_2 = \frac{s_2}{\bar{x}_2} \times 100 = \frac{3}{13.2} \times 100 \approx 22.7\%$$

Interprétation intuitive :

- L'IUT 1 présente une dispersion relative de 40% autour de sa moyenne, contre 22.7% pour l'IUT 2
- L'IUT 2 semble avoir une notation plus homogène (écarts plus faibles par rapport à la moyenne)
- L'IUT 1 présente une plus grande variabilité des notes, ce qui pourrait indiquer soit une plus grande hétérogénéité des étudiants, soit des critères de notation plus différenciants

Toutefois, cette observation descriptive ne suffit pas : il faut un test statistique pour déterminer si cette différence est significative ou due au hasard de l'échantillonnage.

2. On souhaite tester si les deux IUT ont la même dispersion de notes.

- Formuler les hypothèses H_0 et H_1 de ce test
- Quelle statistique de test doit-on utiliser ? Préciser sa loi sous H_0
- Ce test nécessite-t-il des conditions particulières ?

**Hypothèses du test :**

On teste l'égalité des variances (ou des écarts-types, ce qui est équivalent) :

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad \text{ou de manière équivalente} \quad H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2 \quad \text{ou} \quad H_1 : \frac{\sigma_1^2}{\sigma_2^2} \neq 1$$

Il s'agit d'un test bilatéral car on ne privilégie aucune direction a priori.

Statistique de test (test de Fisher) :

On utilise le rapport des variances empiriques :

$$F = \frac{s_1^2}{s_2^2}$$

Par convention, on place la plus grande variance au numérateur pour avoir $F \geq 1$.

Sous H_0 (si $\sigma_1^2 = \sigma_2^2$), cette statistique suit une loi de Fisher :

$$F \sim \mathcal{F}(n_1 - 1, n_2 - 1) = \mathcal{F}(9, 24)$$

Conditions d'application :

- (a) Les deux échantillons sont indépendants
- (b) Les notes dans chaque IUT suivent une loi normale (hypothèse de l'énoncé)
- (c) Les échantillons sont aléatoires simples

Attention : Le test de Fisher est sensible à la non-normalité. Si les distributions ne sont pas normales, ce test peut donner des résultats incorrects.

3. Au seuil de signification $\alpha = 5\%$, peut-on considérer que les enseignants des deux IUT ont la même manière de noter les étudiants ?


Calcul de la statistique de test :

On calcule le rapport des variances (la plus grande au numérateur) :

$$F_{\text{obs}} = \frac{s_1^2}{s_2^2} = \frac{5^2}{3^2} = \frac{25}{9} \approx 2.778$$

Détermination de la région critique :

Pour un test bilatéral au seuil $\alpha = 5\%$, on rejette H_0 si :

$$F_{\text{obs}} > F_{1-\alpha/2; n_1-1, n_2-1} = F_{0.975; 9, 24}$$

ou si

$$F_{\text{obs}} < F_{\alpha/2; n_1-1, n_2-1} = F_{0.025; 9, 24}$$

En pratique, comme on a placé la plus grande variance au numérateur ($F > 1$), on compare uniquement avec la valeur critique supérieure.

D'après les tables de Fisher (ou avec un tableur) :

$$F_{0.975; 9, 24} \approx 2.90$$

Pour la borne inférieure (si nécessaire) :

$$F_{0.025; 9, 24} = \frac{1}{F_{0.975; 24, 9}} \approx \frac{1}{3.61} \approx 0.277$$

Décision :

$F_{\text{obs}} = 2.778 < 2.90$: on se situe dans la région d'acceptation.

On ne rejette pas H_0 au seuil $\alpha = 5\%$.

Calcul de la p -valeur :

Pour un test bilatéral, avec $F = 2.778$:

$$p\text{-valeur} = 2 \times \mathbb{P}(F_{9, 24} > 2.778) = 2 \times (1 - \mathbb{P}(F_{9, 24} \leq 2.778))$$

Avec un tableur : $p\text{-valeur} \approx 0.053$ (légèrement supérieure à 5%).

Conclusion :

Au niveau de confiance de 95%, on ne peut pas conclure à une différence significative entre les écarts-types des deux IUT.

Les données ne permettent pas d'affirmer que les enseignants ont des manières de noter différentes en termes d'homogénéité.

Remarque importante : La p -valeur étant très proche de 5% (environ 5.3%), le résultat est « limite ». Avec un échantillon plus grand, on pourrait peut-être détecter une différence significative. C'est ce qu'on appelle un résultat « marginalement significatif ».

4. Construire un intervalle de confiance à 95% pour le rapport $\frac{\sigma_1^2}{\sigma_2^2}$. Que peut-on en déduire concernant la comparaison des variances ?


Construction de l'intervalle de confiance :

Un intervalle de confiance à $1 - \alpha = 95\%$ pour le rapport des variances est donné par :

$$IC_{95\%} \left(\frac{\sigma_1^2}{\sigma_2^2} \right) = \left[\frac{s_1^2/s_2^2}{F_{1-\alpha/2; n_1-1, n_2-1}} ; \frac{s_1^2/s_2^2}{F_{\alpha/2; n_1-1, n_2-1}} \right]$$

$$= \left[\frac{F_{\text{obs}}}{F_{0.975; 9, 24}} ; \frac{F_{\text{obs}}}{F_{0.025; 9, 24}} \right]$$

Avec les valeurs numériques :

- $F_{\text{obs}} = 2.778$
- $F_{0.975; 9, 24} = 2.90$
- $F_{0.025; 9, 24} = 0.277$

$$IC_{95\%} \left(\frac{\sigma_1^2}{\sigma_2^2} \right) = \left[\frac{2.778}{2.90} ; \frac{2.778}{0.277} \right]$$

$$= [0.958 ; 10.03]$$

Interprétation :

Avec 95% de confiance, le rapport des variances $\frac{\sigma_1^2}{\sigma_2^2}$ se situe entre 0.958 et 10.03.

- Cet intervalle contient la valeur 1 (test de $H_0 : \frac{\sigma_1^2}{\sigma_2^2} = 1$)
- Cela confirme qu'on ne peut pas rejeter l'hypothèse d'égalité des variances
- L'intervalle est très large, ce qui reflète l'incertitude due aux petits échantillons
- La variance de l'IUT 1 pourrait être jusqu'à 10 fois plus grande que celle de l'IUT 2, ou légèrement plus petite (facteur 0.96)

Conclusion pratique :

La grande largeur de l'intervalle suggère qu'il faudrait des échantillons plus grands pour conclure avec précision sur la comparaison des variances.

5. Pour aller plus loin :

- Que se passerait-il si on augmentait la taille des échantillons (par exemple $n_1 = 30$ et $n_2 = 50$) tout en conservant les mêmes écarts-types ?
- Quelle serait la taille d'échantillon minimale nécessaire pour l'IUT 1 (en gardant $n_2 = 25$ et les mêmes écarts-types) pour pouvoir rejeter H_0 au seuil de 5% ?
- Dans le contexte pédagogique, quelles pourraient être les causes d'une différence d'écart-type si elle était avérée ?



1. Impact de l'augmentation des échantillons :

Si $n_1 = 30$ et $n_2 = 50$ avec les mêmes écarts-types :

- La statistique F reste la même : $F = 2.778$
- Mais la distribution de référence change : $\mathcal{F}(29, 49)$
- La valeur critique diminue : $F_{0.975;29,49} \approx 1.87$ (au lieu de 2.90)
- Avec $F_{\text{obs}} = 2.778 > 1.87$, on rejeterait H_0
- La p -valeur diminuerait significativement (environ 0.003)

Conclusion : Des échantillons plus grands permettent de détecter des différences plus faibles avec plus de confiance.

2. Taille minimale pour n_1 :

On cherche n_1 tel que $F_{\text{obs}} = 2.778 > F_{0.975;n_1-1,24}$ au seuil $\alpha = 5\%$.

Par tâtonnements (ou résolution numérique) :

- Pour $n_1 = 11$: $F_{0.975;10,24} \approx 2.74 \Rightarrow$ rejet
- Pour $n_1 = 10$: $F_{0.975;9,24} \approx 2.90 \Rightarrow$ non-rejet

Il faudrait au moins $n_1 = 11$ observations pour l'IUT 1.

3. Causes possibles d'une différence d'écart-type :

Si les écarts-types étaient significativement différents, cela pourrait s'expliquer par :

Côté étudiants :

- Hétérogénéité différente des niveaux (recrutement, prérequis)
- Motivation et implication variables
- Taille des groupes différente (petits groupes vs grands amphithéâtres)

Côté enseignement :

- Critères d'évaluation plus ou moins stricts/différenciants
- Barèmes différents (notation sur 20 vs notation plus « resserrée »)
- Type d'épreuves (QCM vs exercices ouverts)
- Pédagogie différenciée vs enseignement uniforme

Côté organisationnel :

- Nombre d'évaluations différent (moyennes de plusieurs notes vs note unique)
- Contrôle continu vs examen terminal
- Politique de rattrapage/seconde chance

Une dispersion plus forte n'est ni bonne ni mauvaise en soi : elle doit être cohérente avec les objectifs pédagogiques.

Ex 12

Comparaison de consommations énergétiques

rhUy

★ ★ ★

Un fournisseur d'énergie souhaite comparer la consommation de chauffage durant les mois d'hiver (décembre à février) entre deux régions géographiques A et B afin d'optimiser ses approvisionnements et son réseau de distribution.

On suppose que la consommation mensuelle de chauffage (en kWh) dans chaque région suit une loi normale. Pour effectuer cette comparaison, on prélève un échantillon aléatoire de ménages dans chaque région et on mesure leur consommation moyenne sur la période.

Résultats observés :

Région A : $n_1 = 200$ ménages, moyenne $\bar{x}_1 = 600$ kWh, variance empirique $s_1^2 = 20\,000$ kWh² (donc $s_1 \approx 141.4$ kWh)

Région B : $n_2 = 100$ ménages, moyenne $\bar{x}_2 = 500$ kWh, variance empirique $s_2^2 = 160\,000$ kWh² (donc $s_2 = 400$ kWh)

Question générale : L'écart de consommation entre les deux régions est-il statistiquement significatif au seuil de 5% ?

1. Avant tout test, analyser les données de manière descriptive :
 - Calculer la différence observée entre les moyennes
 - Calculer les coefficients de variation pour chaque région

- Que peut-on dire intuitivement sur la dispersion dans chaque région ?



Différence des moyennes :

$$\bar{x}_1 - \bar{x}_2 = 600 - 500 = 100 \text{ kWh}$$

La région A consomme en moyenne 100 kWh de plus que la région B, soit 20% de plus ($100/500 = 0.20$).

Écart-types :

— Région A : $s_1 = \sqrt{20000} \approx 141.4 \text{ kWh}$

— Région B : $s_2 = \sqrt{160000} = 400 \text{ kWh}$

Coefficients de variation :

$$CV_A = \frac{s_1}{\bar{x}_1} \times 100 = \frac{141.4}{600} \times 100 \approx 23.6\%$$

$$CV_B = \frac{s_2}{\bar{x}_2} \times 100 = \frac{400}{500} \times 100 = 80\%$$

Analyse descriptive :

- La région A présente une consommation plus élevée mais plus homogène ($CV = 23.6\%$)
- La région B a une consommation plus faible en moyenne mais avec une très forte hétérogénéité ($CV = 80\%$)
- Cette forte dispersion en région B pourrait s'expliquer par :
 - une diversité importante des types d'habitation (surface, isolation)
 - des conditions climatiques locales variables (altitude, exposition)
 - des comportements de chauffage très différents entre ménages
- Les variances semblent très différentes : s_2^2 est 8 fois plus grande que s_1^2

2. Avant de comparer les moyennes, doit-on vérifier l'égalité des variances ? Réaliser un test de Fisher au seuil de 5% pour comparer σ_1^2 et σ_2^2 .



Importance de cette étape :

Pour comparer deux moyennes de lois normales, on utilise :

- Un test de Student avec variances égales (pooled variance) si $\sigma_1^2 = \sigma_2^2$
- Un test de Welch avec variances inégales si $\sigma_1^2 \neq \sigma_2^2$

Il est donc important de tester d'abord l'égalité des variances.

Test de Fisher :

Hypothèses :

$$H_0 : \sigma_1^2 = \sigma_2^2$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2$$

Statistique de test (plus grande variance au numérateur) :

$$F = \frac{s_2^2}{s_1^2} = \frac{160000}{20000} = 8$$

Sous $H_0 : F \sim \mathcal{F}(n_2 - 1, n_1 - 1) = \mathcal{F}(99, 199)$

Valeur critique :

Pour un test bilatéral à $\alpha = 5\%$:

$$F_{0.975;99,199} \approx 1.39$$

Décision :

$F_{\text{obs}} = 8 > 1.39$: on rejette H_0 .

La p -valeur est extrêmement faible (quasi nulle).

Conclusion :

Les variances des deux régions sont significativement différentes au seuil de 5%. Il faudra donc utiliser le test de Welch (et non le test de Student classique) pour comparer les moyennes.

3. Formuler les hypothèses pour tester si la consommation moyenne est différente entre les deux régions. S'agit-il d'un test bilatéral ou unilatéral ? Justifier.

**Hypothèses :**

On souhaite tester s'il existe une différence de consommation moyenne :

$$H_0 : \mu_1 = \mu_2 \quad \text{ou de manière équivalente} \quad H_0 : \mu_1 - \mu_2 = 0$$

$$H_1 : \mu_1 \neq \mu_2 \quad \text{ou} \quad H_1 : \mu_1 - \mu_2 \neq 0$$

Type de test : bilatéral**Justification :**

- La question demande si l'écart est significatif, sans préciser de direction
- On ne cherche pas à montrer qu'une région consomme plus que l'autre, mais simplement à détecter une différence
- L'objectif du fournisseur est de détecter toute différence (dans un sens ou dans l'autre) pour optimiser sa logistique
- Un test bilatéral est plus conservateur et plus approprié dans un contexte exploratoire

Remarque : Si l'objectif était de montrer que la région A consomme *plus* que la région B (par exemple pour justifier des investissements supplémentaires en région A), on utiliserait un test unilatéral avec $H_1 : \mu_1 > \mu_2$.

4. Les variances étant inégales, on utilise le test de Welch. La statistique de test est :

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

qui suit approximativement une loi de Student à ν degrés de liberté, où :

$$\nu = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{(s_1^2/n_1)^2}{n_1-1} + \frac{(s_2^2/n_2)^2}{n_2-1}}$$

Calculer la statistique de test observée et le nombre de degrés de liberté.



Calcul des variances estimées des moyennes :

$$\frac{s_1^2}{n_1} = \frac{20000}{200} = 100$$

$$\frac{s_2^2}{n_2} = \frac{160000}{100} = 1600$$

$$\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} = 100 + 1600 = 1700$$

Statistique de test :

$$t_{\text{obs}} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{600 - 500}{\sqrt{1700}} = \frac{100}{\sqrt{1700}} = \frac{100}{41.23} \approx 2.426$$

Calcul des degrés de liberté (formule de Welch-Satterthwaite) :

Numérateur :

$$\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2} \right)^2 = (1700)^2 = 2\,890\,000$$

Dénominateur :

$$\begin{aligned} \frac{(s_1^2/n_1)^2}{n_1 - 1} + \frac{(s_2^2/n_2)^2}{n_2 - 1} &= \frac{100^2}{199} + \frac{1600^2}{99} \\ &= \frac{10000}{199} + \frac{2560000}{99} = 50.25 + 25858.59 = 25908.84 \end{aligned}$$

Donc :

$$\nu = \frac{2890000}{25908.84} \approx 111.57$$

On arrondit à $\nu = 111$ ou $\nu = 112$ degrés de liberté.

Remarque : Les degrés de liberté sont plus proches de $n_2 - 1 = 99$ que de $n_1 + n_2 - 2 = 298$, car la variance de la région B domine le calcul.

5. Au seuil de signification $\alpha = 5\%$, l'écart de consommation entre les deux régions est-il significatif? Calculer également la p -valeur et interpréter le résultat dans le contexte.

**Région critique :**

Pour un test bilatéral au seuil $\alpha = 5\%$ avec $\nu \approx 111$ ddl, on rejette H_0 si $|t_{\text{obs}}| > t_{0.975;111}$.
D'après les tables de Student (ou un tableur) :

$$t_{0.975;111} \approx 1.982$$

Décision :

$|t_{\text{obs}}| = 2.426 > 1.982$: **on rejette H_0 au seuil $\alpha = 5\%$.**

Calcul de la p -valeur :

Pour un test bilatéral :

$$p\text{-valeur} = 2 \times \mathbb{P}(T_{111} > 2.426) = 2 \times (1 - \mathbb{P}(T_{111} \leq 2.426))$$

Avec un tableur : $p\text{-valeur} \approx 0.0167$ soit environ 1.67%.

Conclusion statistique :

Au niveau de confiance de 95%, on peut conclure que la consommation moyenne de chauffage est significativement différente entre les deux régions.

Interprétation dans le contexte :

- La région A consomme en moyenne 100 kWh de plus par ménage et par mois que la région B, soit une surconsommation de 20%
- Cette différence est statistiquement significative et ne peut pas être attribuée au hasard de l'échantillonnage
- La p -valeur de 1.67% indique un résultat assez robuste (bien inférieur au seuil de 5%)
- Cette différence peut s'expliquer par :
 - Des conditions climatiques plus rigoureuses en région A (températures plus basses, exposition au vent)
 - Des caractéristiques d'habitation différentes (surface moyenne, qualité d'isolation, type de chauffage)
 - Des profils socio-économiques différents (présence au domicile, température de confort recherchée)
- **Implications opérationnelles pour le fournisseur :**
 - Adapter les stocks et approvisionnements selon les régions
 - Prévoir des capacités de distribution plus importantes en région A
 - Cibler différemment les campagnes de maîtrise de l'énergie
 - Ajuster les prévisions de demande régionales

6. Construire un intervalle de confiance à 95% pour la différence des moyennes $\mu_1 - \mu_2$.
Que peut-on en déduire ?


Construction de l'intervalle de confiance :

L'intervalle de confiance à $1 - \alpha = 95\%$ pour $\mu_1 - \mu_2$ est :

$$IC_{95\%}(\mu_1 - \mu_2) = (\bar{x}_1 - \bar{x}_2) \pm t_{1-\alpha/2; \nu} \times \sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}$$

Avec les valeurs numériques :

- $\bar{x}_1 - \bar{x}_2 = 100$ kWh
- $t_{0.975; 111} \approx 1.982$
- $\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}} = \sqrt{1700} \approx 41.23$ kWh

Marge d'erreur :

$$ME = 1.982 \times 41.23 \approx 81.72 \text{ kWh}$$

Donc :

$$IC_{95\%}(\mu_1 - \mu_2) = [100 - 81.72; 100 + 81.72] = [18.28; 181.72] \text{ kWh}$$

Interprétation :

Avec 95% de confiance, la surconsommation moyenne de la région A par rapport à la région B se situe entre 18.3 et 181.7 kWh par ménage et par mois.

Conséquences :

- L'intervalle ne contient pas 0, ce qui confirme le rejet de H_0 (cohérence test/IC)
- La borne inférieure (18.3 kWh) représente environ 3.7% de la consommation moyenne de la région B
- La borne supérieure (181.7 kWh) représente environ 36% de la consommation de la région B
- L'intervalle est assez large, reflétant l'importante variabilité en région B ($s_2 = 400$ kWh)
- Dans le pire cas (borne supérieure), la région A consomme presque 200 kWh de plus, ce qui justifie des investissements différenciés

Quantification de l'incertitude :

La largeur de l'intervalle (163.4 kWh) est importante, principalement à cause de :

- La forte variance en région B ($s_2^2 = 160000$)
- La taille d'échantillon relativement modeste en région B ($n_2 = 100$)

Pour réduire cette incertitude, il faudrait augmenter la taille de l'échantillon en région B.

7. Synthèse :

une analyste propose d'utiliser le test de Student classique (avec hypothèse de variances égales) au lieu du test de Welch.

- Expliquer pourquoi ce serait inapproprié dans ce cas
- Calculer néanmoins la statistique de test qu'on obtiendrait avec le test de Student classique (variance poolée)
- Comparer les deux approches : quelle différence observe-t-on dans la conclusion ?
- Quel enseignement peut-on tirer sur l'importance de vérifier les hypothèses d'un test ?



1. Pourquoi le test de Student classique serait inapproprié :

Le test de Student avec variances égales suppose $\sigma_1^2 = \sigma_2^2$. Or, nous avons montré (question 2) que cette hypothèse est fortement rejetée :

- $F = 8$ avec p -valeur < 0.0001
- La variance de B est 8 fois celle de A
- Le CV de B (80%) est 3.4 fois celui de A (23.6%)

Utiliser le test classique violerait une hypothèse fondamentale et pourrait conduire à des conclusions erronées.

2. Calcul avec le test de Student classique (à titre de comparaison) :

Variance poolée (estimateur de la variance commune supposée) :

$$\begin{aligned} s_p^2 &= \frac{(n_1 - 1)s_1^2 + (n_2 - 1)s_2^2}{n_1 + n_2 - 2} = \frac{199 \times 20000 + 99 \times 160000}{298} \\ &= \frac{3980000 + 15840000}{298} = \frac{19820000}{298} \approx 66510.07 \end{aligned}$$

Donc $s_p \approx 257.9$ kWh

Statistique de test :

$$\begin{aligned} t &= \frac{\bar{x}_1 - \bar{x}_2}{s_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} = \frac{100}{257.9 \times \sqrt{\frac{1}{200} + \frac{1}{100}}} \\ &= \frac{100}{257.9 \times \sqrt{0.005 + 0.01}} = \frac{100}{257.9 \times 0.1225} = \frac{100}{31.59} \approx 3.166 \end{aligned}$$

Degrés de liberté : $\nu = n_1 + n_2 - 2 = 298$

Valeur critique : $t_{0.975;298} \approx 1.968$

3. Comparaison des deux approches :

Méthode	Test de Welch	Student classique
Hypothèse sur variances	Inégales	Égales (fausse ici)
Statistique t	2.426	3.166
Degrés de liberté	111	298
Valeur critique (5%)	1.982	1.968
p -valeur	0.0167	0.0017
Conclusion	Rejet de H_0	Rejet de H_0

Observations :

- Les deux tests rejettent H_0 , mais avec des niveaux de significativité différents
- Le test de Student classique donne une statistique plus élevée (3.166 vs 2.426) et une p -valeur plus faible (0.17% vs 1.67%)
- Le test classique est ici **anti-conservateur** : il surestime la significativité (risque d'erreur de type I plus élevé que prévu)
- La variance poolée ($s_p^2 \approx 66510$) est intermédiaire entre $s_1^2 = 20000$ et $s_2^2 = 160000$, mais ne représente bien ni l'une ni l'autre

4. Enseignements sur l'importance de vérifier les hypothèses :

- **Les hypothèses ne sont pas des formalités** : elles conditionnent la validité des conclusions
- Dans ce cas, bien que les deux tests rejettent H_0 , le test classique donne une fausse impression de forte significativité
- Dans des cas limites (résultat proche du seuil), l'utilisation d'un test inapproprié pourrait inverser la conclusion
- **Principe de précaution** : en cas de doute sur l'égalité des variances, préférer le test de Welch (plus robuste)
- La vérification préalable (test de Fisher) n'était pas une perte de temps, elle était essentielle pour choisir le bon test
- **En pratique** : les logiciels modernes utilisent souvent le test de Welch par défaut, car il est plus robuste et ne pénalise que peu quand les variances sont effectivement égales

Citation pertinente :

« *All models are wrong, but some are useful* » (George Box). Les hypothèses de normalité et d'égalité de variance sont des modèles simplificateurs. Quand ils sont clairement violés (comme l'égalité des variances ici), il faut utiliser des méthodes plus robustes.

Ex 13

Test d'indépendance

DzLc

Dans le cadre d'une enquête sur le SIDA réalisée en Allemagne durant l'été 1990 (A. Hahn, W.H. Eirmbter et R. Jacob), on a interrogé 2089 personnes. Le questionnaire comportait notamment l'item suivant : Le sida est un péril omniprésent. Indiquez si vous êtes : d'accord, indécis, pas d'accord. Le croisement de la réponse du sujet avec son âge donne le tableau de contingence suivant :

Classe d'âge	d'accord	indécis	pas d'accord	Total
18 à < 30	43	116	365	524
30 à < 40	36	116	273	425
40 à < 50	32	95	217	344
50 à < 60	38	114	167	319
60 et plus	67	160	250	477
Total	216	601	1272	2089

1. Réaliser un test du χ^2 permettant de répondre à la question suivante : « Les réponses des sujets dépendent-elles de leur âge ? »



Le tableau des effectifs théoriques est donné par : Celui des contributions au χ^2 est donné par :

Classe d'âge	d'accord	indécis	pas d'accord
18 à < 30	2.31	8.01	6.61
30 à < 40	1.44	0.32	0.78
40 à < 50	0.36	0.16	0.27
50 à < 60	0.76	5.58	3.82
60 et plus	6.33	3.78	5.63

On obtient $\chi_{obs}^2 = 46$. Or pour un seuil de 1% et $ddl = (5 - 1)(3 - 1) = 8$, on a : $\chi_c^2 = 20.09$. La réponse du sujet est donc dépendante de son âge.

2. Comparer de même à l'aide d'un test du χ^2 les réponses des deux dernières classes d'âge, puis les réponses des moins de 30 ans à celles des 60 ans et plus.



En ne considérant que les deux dernières classes d'âge, on obtient le tableau d'effectifs théoriques suivant :

Classe d'âge	d'accord	indécis	pas d'accord
50 à < 60	42.1	109.8	167.1
60 et plus	62.9	164.2	249.9

On obtient alors $\chi_{obs}^2 = 0.92$. Or, pour $\alpha = 5\%$ et $ddl = 2$, $\chi_c^2 = 5.99$. Les réponses des sujets sont cette fois indépendantes de leur appartenance à l'une ou l'autre classe d'âge. Dans le dernier cas, $\chi_{obs}^2 = 31.62$, ce qui est significatif d'une dépendance.

Ex 14

Analyse de la précision d'un système d'arme

QeZr

★★★★

Dans le cadre de la qualification d'une section d'infanterie, on évalue la précision du tir au FAMAS sur cible à 200 mètres. Le référentiel de l'armée de terre fixe un écart-type maximal de $\sigma_0 = 8$ cm pour les impacts par rapport au centre de la cible (tous tireurs confondus) pour qu'une section soit considérée comme opérationnelle.

Un chef de section souhaite vérifier si sa section respecte ce standard. Lors d'une séance de tir,

25 soldats effectuent chacun un tir, et on mesure la distance de l'impact au centre de la cible (en cm).

Distances observées (en cm) :

5.2	12.3	8.7	6.5	15.1	9.2	7.4	11.8	4.9	13.6	9.8
4.1	11.2	7.8	13.5	6.9	10.5	8.3	14.2	5.7	7.2	10.1
8.9	12.5	6.3								

Statistiques calculées (pour vérification) : $\bar{x} = 9.27$ cm, $s^2 = 9.83$ cm², $s = 3.13$ cm

On suppose que la distance des impacts suit une loi normale.

- Avant de réaliser les tests statistiques, visualiser les données :
 - Représenter les distances par un histogramme ou un diagramme en boîte
 - Cette représentation est-elle cohérente avec l'hypothèse de normalité ?
 - Y a-t-il des valeurs aberrantes qui pourraient indiquer un problème (tireur mal formé, incident de tir, etc.) ?



Calculs des quartiles et statistiques descriptives :

En ordonnant les données : 4.1, 4.9, 5.2, 5.7, 6.3, 6.5, 6.9, 7.2, 7.4, 7.8, 8.3, 8.7, 8.9, 9.2, 9.8, 10.1, 10.5, 11.2, 11.8, 12.3, 12.5, 13.5, 13.6, 14.2, 15.1

- Effectif : $n = 25$
- Minimum : 4.1 cm
- Q1 (25%) : 6.9 cm (7^e valeur)
- Médiane (50%) : 8.9 cm (13^e valeur)
- Q3 (75%) : 11.8 cm (19^e valeur)
- Maximum : 15.1 cm
- Écart interquartile : $IQR = Q3 - Q1 = 11.8 - 6.9 = 4.9$ cm

Recherche de valeurs aberrantes :

Critère de Tukey : une valeur est aberrante si elle est en dehors de $[Q1 - 1.5 \times IQR; Q3 + 1.5 \times IQR]$

$$[6.9 - 1.5 \times 4.9; 11.8 + 1.5 \times 4.9] = [-0.45; 19.15]$$

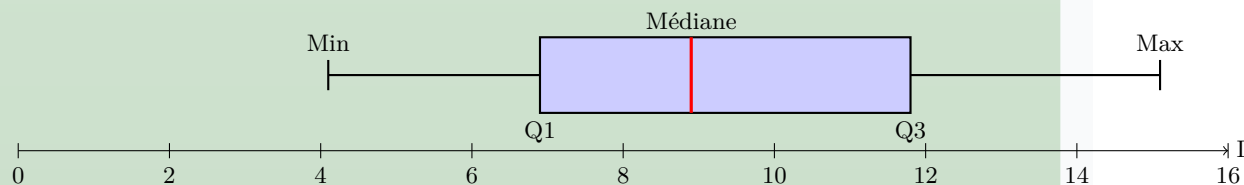
Toutes les valeurs sont dans cet intervalle : **pas de valeur aberrante.**

Cohérence avec la normalité :

- La distribution semble relativement symétrique autour de la médiane (8.9 cm)
- La médiane (8.9 cm) est très proche de la moyenne (9.27 cm), ce qui suggère une distribution symétrique
- L'écart-type $s = 3.13$ cm représente environ 34% de la moyenne, ce qui est raisonnable
- Pas de valeurs extrêmes selon le critère de Tukey
- Distribution unimodale concentrée autour de 9 cm
- L'écart-type est nettement inférieur au standard requis ($3.13 < 8$ cm)

L'hypothèse de normalité semble acceptable pour ces données.

Diagramme en boîte (boîte à moustaches) :



- Le chef de section souhaite tester si sa section respecte le standard de l'armée.
 - Formuler les hypothèses H_0 et H_1 de ce test
 - Quel type de test doit-on utiliser ? (test bilatéral ou unilatéral ?)
 - Justifier le choix de la statistique de test et donner sa loi sous H_0

**Formulation des hypothèses :**

Le standard fixe $\sigma_0 = 8$ cm comme écart-type maximal. On veut tester si la dispersion observée est conforme (ou meilleure) à ce standard.

$$H_0 : \sigma^2 \leq \sigma_0^2 = 64 \text{ cm}^2 \quad (\text{section opérationnelle})$$

$$H_1 : \sigma^2 > \sigma_0^2 = 64 \text{ cm}^2 \quad (\text{précision insuffisante})$$

Type de test : Test unilatéral à droite, car on cherche à détecter une dispersion *trop grande* (problème opérationnel).

Statistique de test :

Pour tester une variance avec un échantillon de loi normale, on utilise :

$$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$$

où $S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$ est la variance empirique.

Sous H_0 (en prenant l'égalité $\sigma^2 = \sigma_0^2$), cette statistique suit une loi du χ^2 à $(n-1)$ degrés de liberté :

$$\chi^2 \sim \chi_{n-1}^2 = \chi_{24}^2$$

Justification :

Si $X_1, \dots, X_n \stackrel{iid}{\sim} \mathcal{N}(\mu, \sigma^2)$, alors :

$$\frac{(n-1)S^2}{\sigma^2} \sim \chi_{n-1}^2$$

3. Au seuil de signification $\alpha = 5\%$, peut-on conclure que la section respecte le standard de précision de l'armée de terre ?

**Calcul de la statistique de test :**

Avec $n = 25$, $s^2 = 9.83 \text{ cm}^2$ et $\sigma_0^2 = 64 \text{ cm}^2$:

$$\chi_{\text{obs}}^2 = \frac{(n-1)s^2}{\sigma_0^2} = \frac{24 \times 9.83}{64} = \frac{235.92}{64} \approx 3.69$$

Région critique :

Pour un test unilatéral à droite avec $\alpha = 5\%$ et 24 degrés de liberté, on rejette H_0 si $\chi_{\text{obs}}^2 > \chi_{0.95;24}^2$.

En consultant la table du χ^2 : $\chi_{0.95;24}^2 = 36.415$

Décision :

$\chi_{\text{obs}}^2 = 3.69 < 36.415$: on se trouve dans la région d'acceptation.

On ne rejette pas H_0 au seuil $\alpha = 5\%$.

Conclusion opérationnelle :

Au niveau de confiance de 95%, les données ne permettent pas de rejeter l'hypothèse que la section respecte le standard de précision de l'armée de terre. La dispersion observée ($s \approx 3.13$ cm) est **significativement inférieure** au standard requis ($\sigma_0 = 8$ cm).

La section peut être considérée comme opérationnelle du point de vue de la précision du tir.

4. Calculer un intervalle de confiance à 95% pour l'écart-type σ de la population. Interpréter ce résultat dans le contexte militaire.


Intervalle de confiance pour σ^2 :

L'intervalle de confiance à $1 - \alpha = 95\%$ pour σ^2 est :

$$\left[\frac{(n-1)s^2}{\chi_{1-\alpha/2; n-1}^2}; \frac{(n-1)s^2}{\chi_{\alpha/2; n-1}^2} \right]$$

Avec $n = 25$, $s^2 = 9.83$, et les valeurs critiques du χ_{24}^2 :

- $\chi_{0.975; 24}^2 = 39.364$
- $\chi_{0.025; 24}^2 = 12.401$

$$IC_{95\%}(\sigma^2) = \left[\frac{24 \times 9.83}{39.364}; \frac{24 \times 9.83}{12.401} \right] = [5.99; 19.02] \text{ cm}^2$$

Intervalle de confiance pour σ :

$$IC_{95\%}(\sigma) = [\sqrt{5.99}; \sqrt{19.02}] = [2.45; 4.36] \text{ cm}$$

Interprétation militaire :

Avec 95% de confiance, l'écart-type réel de la section se situe entre 2.45 et 4.36 cm.

- Le standard de l'armée fixe $\sigma_0 = 8$ cm
- La borne supérieure de l'IC est 4.36 cm, **bien inférieure** à 8 cm
- Même dans le pire scénario (borne supérieure), la section reste largement en deçà du seuil maximal
- Cela confirme que la section a une précision de tir **excellente**
- Le chef de section peut valider sa section comme opérationnelle avec une grande confiance

5. Le commandant de compagnie examine les résultats et fait la remarque suivante :

« Votre section montre d'excellents résultats avec un écart-type de 3.1 cm, bien meilleur que le standard de 8 cm. Cependant, je constate que la moyenne des impacts est à 9.3 cm du centre. Ne faudrait-il pas aussi vérifier que le tir n'est pas systématiquement décentré ? »

- Le commandant a-t-il raison de s'interroger sur ce point ?
- Proposer et réaliser un test statistique pour vérifier si le tir de la section présente un biais systématique (hypothèse : le tir devrait être centré, donc $\mu = 0$)
- Que recommanderiez-vous au chef de section suite à cette analyse complète ?


Pertinence de la remarque :

Le commandant a tout à fait raison. Un système d'arme peut être :

- **Précis** (faible dispersion, σ petit) : c'est le cas ici
- **Juste** (pas de biais, impacts centrés sur la cible)

Un système précis mais biaisé reste problématique opérationnellement (tous les tirs manquent le but dans la même direction).

Test de la moyenne :

On teste :

$$H_0 : \mu = 0 \quad (\text{tir centré})$$

$$H_1 : \mu \neq 0 \quad (\text{tir biaisé})$$

Test bilatéral (le biais peut être dans n'importe quelle direction).

Statistique de test (loi de Student car σ inconnu) :

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}} = \frac{9.27 - 0}{\sqrt{9.83}/\sqrt{25}} = \frac{9.27}{3.135/5} = \frac{9.27}{0.627} = 14.78$$

Sous $H_0 : t \sim \mathcal{T}_{24}$ (Student à 24 ddl).

Pour $\alpha = 5\%$ bilatéral : $t_{0.975;24} = 2.064$

Décision :

$|t_{\text{obs}}| = 14.78 > 2.064$: **on rejette** H_0 .

La p -valeur est extrêmement faible (quasi nulle).

Conclusion :

Le tir de la section présente un **bias systématique très significatif** d'environ 9.27 cm par rapport au centre.

Recommandations opérationnelles :
(a) Diagnostic du biais :

- Vérifier le réglage des hausses et des organes de visée
- Contrôler l'alignement des armes (armurerie)
- Analyser si le biais est identique pour tous les tireurs (problème matériel) ou variable (problème de formation)

(b) Actions correctives :

- Faire réviser les armes par l'armurerie
- Refaire un tir de réglage après correction
- Vérifier les conditions de tir (vent latéral ? position de tir ?)

(c) Qualification :

- La section est **précise** (bonne dispersion)
- Mais elle n'est **pas juste** (biais significatif)
- Après correction du biais, elle devrait être pleinement opérationnelle

6. Un officier du bureau emploi propose de comparer cette section avec une autre section de la même compagnie qui a obtenu les résultats suivants (20 tireurs) : $\bar{y} = 8.1$ cm, $s_y^2 = 15.3$ cm².

Les deux sections ont-elles le même niveau de précision ? Réaliser un test de comparaison de variances au seuil de 5%.

**Test de comparaison de deux variances (test de Fisher) :**

On teste :

$$H_0 : \sigma_1^2 = \sigma_2^2 \quad (\text{même précision})$$

$$H_1 : \sigma_1^2 \neq \sigma_2^2 \quad (\text{précisions différentes})$$

Statistique de test :

$$F = \frac{s_y^2}{s^2} = \frac{15.3}{9.83} = 1.556$$

(On place la plus grande variance au numérateur pour avoir $F \geq 1$)

Sous $H_0 : F \sim \mathcal{F}(n_2 - 1, n_1 - 1) = \mathcal{F}(19, 24)$ (loi de Fisher).

Pour un test bilatéral à $\alpha = 5\%$, on compare F_{obs} avec les valeurs critiques correspondantes.

En pratique, pour un test bilatéral avec $F \geq 1$, on compare avec $F_{1-\alpha/2}$:

$$F_{0.975;19,24} \approx 2.35$$

Décision :

$F_{\text{obs}} = 1.556 < 2.35$: **on ne rejette pas H_0 .**

Conclusion :

Au seuil de 5%, on ne peut pas conclure à une différence significative de précision entre les deux sections. Les écarts observés peuvent être attribués à la variabilité d'échantillonnage.

Interprétation opérationnelle :

Les deux sections ont un niveau de précision comparable. Le commandant de compagnie peut considérer que la formation au tir est homogène entre les sections.