Chapitre 5

Test d'hypothèses avec un échantillon

1) Introduction dans l'exemple 1 expérience de Mendel) du chapitre précédent, nous avons noté qu'il y avait 26.2% de pois de gousses jaunes alors que Mendel s'attendait à 25%, en statistique on utilise une approche qui nous permet de décider si le résultat obtenu diffère d'une quantité significative du résultat attendu ou prédit. On utilise pour cela une procédure appelé «test d'hypothèses ».

2) Définition en statistique, une hypothèse est une affirmation ou un énoncé à propos d'une propriété d'une population.

* un test statistique (ou test de significativité) est une procédure standard pour tester un énoncé à propos d'une propriété d'une population.

3)composantes d'un test d'hypothèse formel

***l'hypothèse nulle** (noté H_0) est l'affirmation que la valeur d'un paramètre d'une population est égale à une certaine valeur supposée (ex $H_0 : p=0.5$, $H_0 : \mu=37.0$, $H_0 : \sigma=15$)

On teste H_0 et on arrive à une conclusion qui soit rejette H_0, soit ne pas rejeter H_0.

***l'hypothèse alternative** (noté H_1) est l'affirmation que le paramètre a une valeur qui diffère de celle de H_0 , la forme

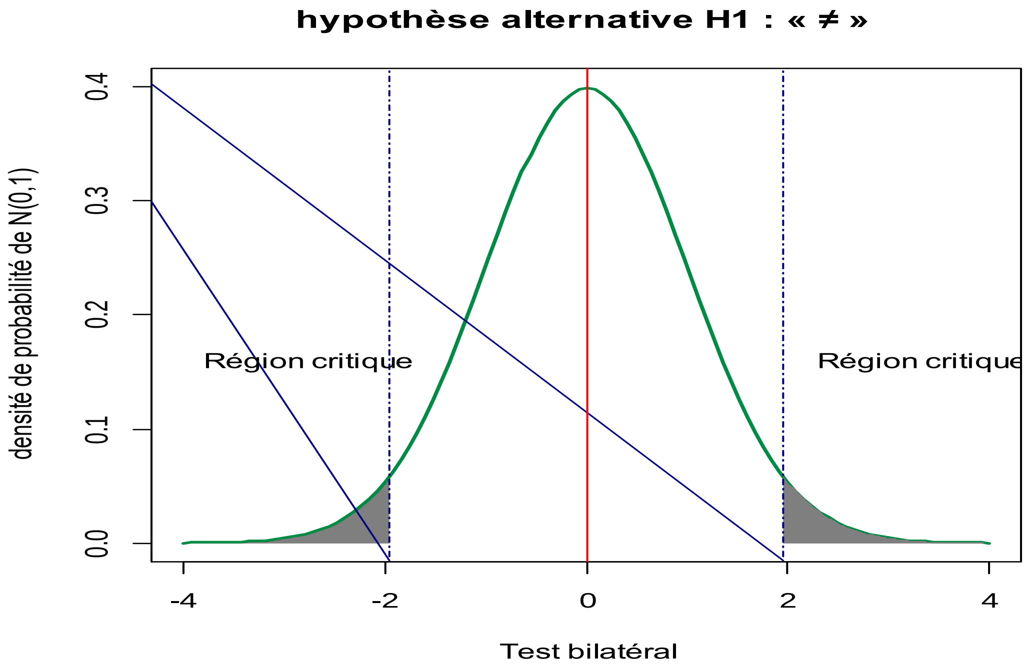
symbolique de H_0 devra utiliser l'une des symboles « > »,
« < », « $\neq$ » (ex $H_1 : p > 0.5$, $H_1 : \mu \neq 37.0$, $H_1 : \sigma < 15$

***Statistique de test** c'est une valeur calculée à partir des données de l'échantillon et elle est utilisée dans la prise de décision du rejet ou non de l'hypothèse nulle.

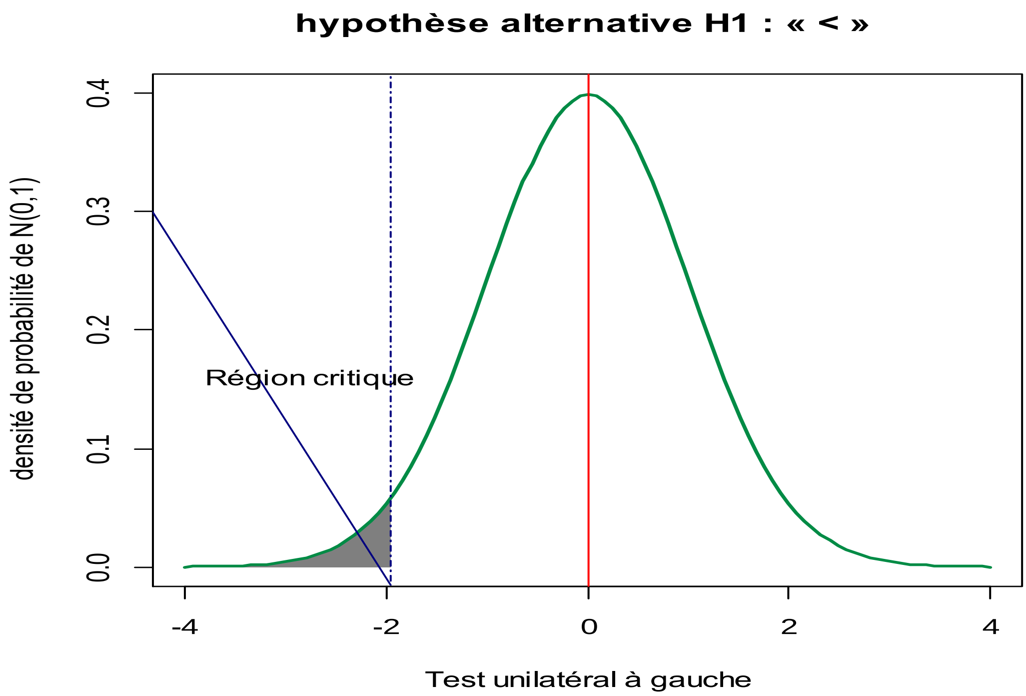
* **Région critique, niveau de significativité, valeur critique.**

- ❖ La région critique (ou zone de rejet) est l'ensemble des valeurs de la statistique de test qui nous font rejeter H_0.
- ❖ Le niveau de significativité (noté α) est la probabilité que la statistique de test tombe dans la région critique les choix courants pour α sont 0.05, 0.01, 0.1.
- ❖ La valeur critique est une valeur qui sépare la région critique (où on rejette H_0 des autres valeurs de la statistique de test. Elle dépend de la nature de H_0, de la distribution d'échantillonnage à appliquer et du niveau de significativité α.

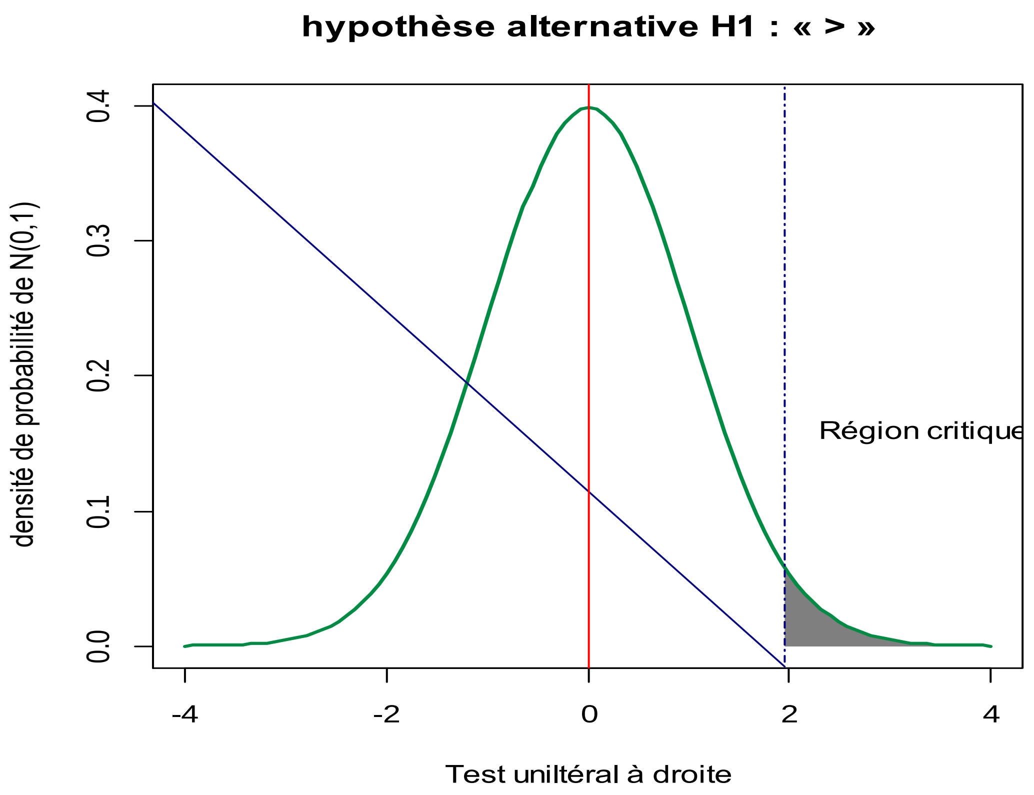
❖ Test bilatéral (H_1 : « ≠ ») la région critique est dans les deux régions extrêmes sous la courbe



❖ Test unilatéral à gauche : (H_1 : « < ») la région critique est dans la région extrême à gauche sous la courbe.



❖ Test unilatéral à droite : (H_1 : « > ») la région critique est dans la région extrême à droite sous la courbe.



***Décisions et conclusions** la décision est l'une des deux affirmations :

1. Rejet de H_0.

2. Echec du rejet de H_0.

Selon la méthode traditionnelle on décide :

« le rejet de H_0 »Si la statistique de test tombe dans la région critique.

« Echec du rejet de H_0 » si la statistique de test ne tombe pas dans la région critique.

4)Test d'hypothèse pour une moyenne

***cas de variance connue**

Conditions d'application du test

1. L'échantillon est aléatoire simple
2. La variance de la population est connue
3. Une des deux conditions est satisfaite : la population est normalement distribuée ou n>30.

Statistique de test

$$z=\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}}$$

Valeurs critiques utiliser une table de loi normale.

Exemple un échantillon de 106 températures humaines dont la moyenne est 36.78°C. Supposer que l'échantillon est aléatoire simple, que l'écart type σ est connu et vaut 0.34. utilisez un niveau de significativité de 0.05 pour tester la croyance commune que la température d'un adulte en bonne santé est 37.0°C

Solution les conditions requises sont satisfaites (cf exemple 1 du chapitre 3)

$H_0 : \mu = 37.0$

$H_1 : \mu \neq 37.0$

$$z=\frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{36.78-37}{\frac{0.34}{\sqrt{106}}} =-6.66$$

il s'agit d'un test bilatéral, on doit comparer la statistique z aux valeurs critiques de la loi normale ; α=0.05

F($t_{\alpha/2}$)=1- $\alpha/2$ = 0.975 $\xrightarrow{la\ table\ N(0,1)}$ $t_{\alpha/2} = 1.96$

z= -6.66 <-$t_{\alpha/2}$ =-1.96 d'où le rejet de H_0

conclusion : on conclut qu'il y a suffisamment de preuves pour dire que la moyenne des températures diffère de 37.0°C.

***cas de variance inconnue** l'objectif principal de cette section est de développer la capacité de tester des affirmations à propos de µ quand σ est inconnu

Conditions d'application du test

1. L'échantillon est aléatoire simple
2. La variance de la population est inconnue
3. Une des deux conditions est satisfaite : la population est normalement distribuée ou n>30.

Statistique de test

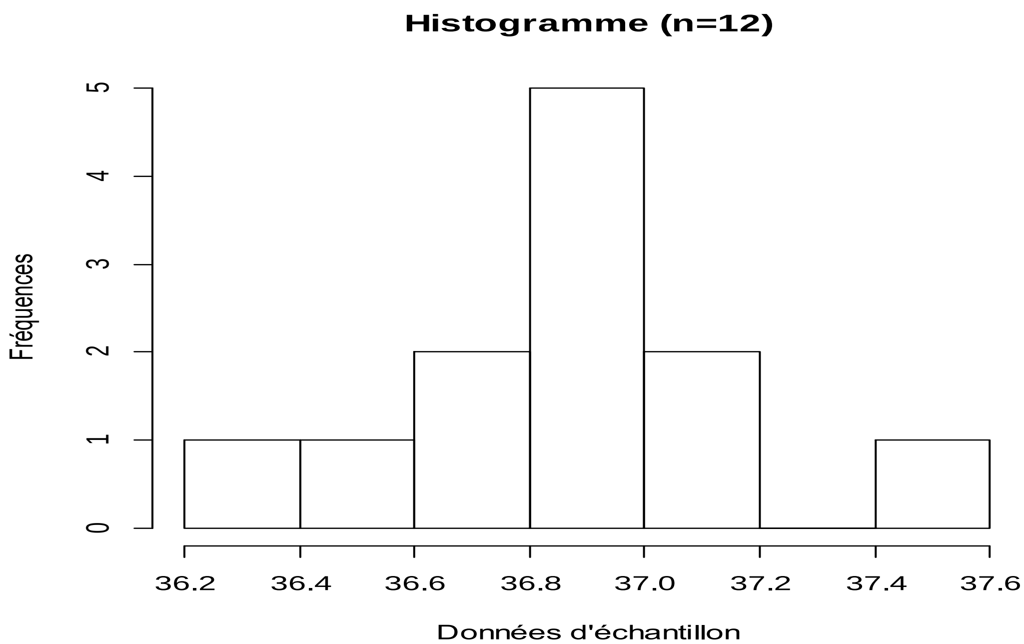
$$t=\frac{\bar{x}-\mu}{\frac{s}{\sqrt{n}}}$$

Valeurs critiques utiliser une table de la loi t de student à n-1 degrés de liberté, ddl=n-1.

Exemple on a les données de température de 12 personnes en bonne santé, on a obtenu les températures listés ci-dessous. Utiliser un niveau de significativité de 0.05 pour tester l'hypothèse que la moyenne de ces températures est issue d'une population dont la moyenne est inférieure à 37.0°C.

36.67	36.39	37.00	37.11	36.67	36.94
37.00	37.44	36.89	37.06	37.00	36.44

Solution il faut d'abord vérifier que les conditions requises sont satisfaites. On a un échantillon aléatoire simple, reste à vérifier la condition de normalité puisque n=12<30. L'histogramme suivant fourni par le logiciel R montre que les données suivent une distribution pas très éloignée de la loi normale.



On effectue le test

On a $\bar{x}$ =36.88°C

$H_0 : \mu = 37.0$

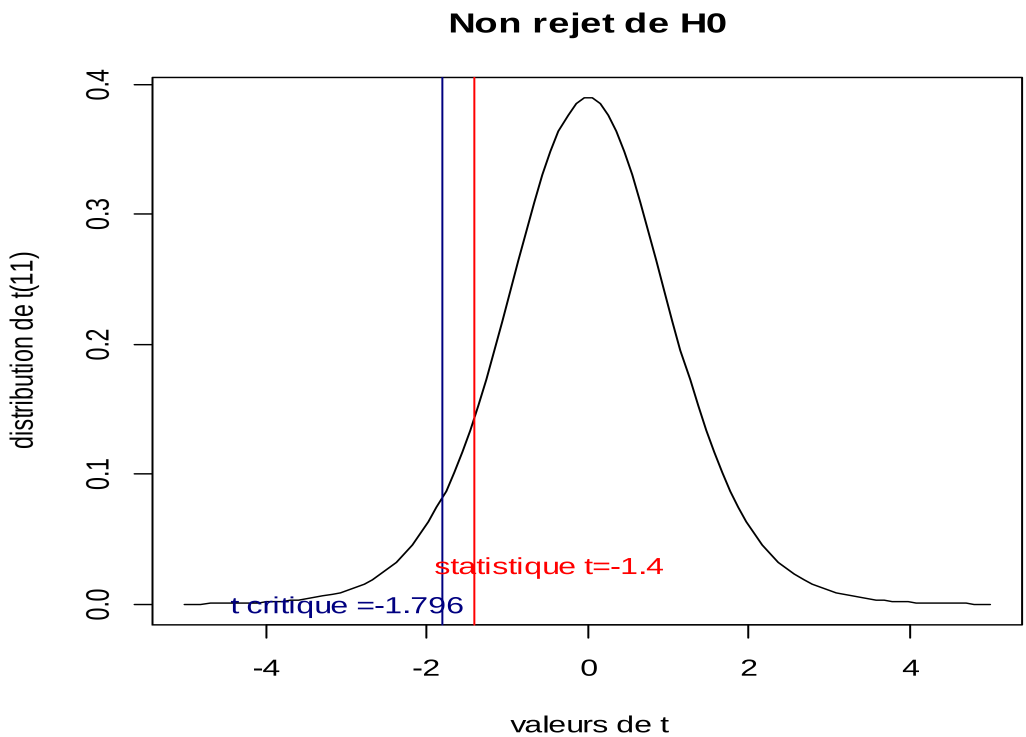
$H_1 : \mu < 37.0$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{36.88 - 37}{\frac{0.297}{\sqrt{12}}} = -1.4$$

la valeur critique $-t_{\alpha,11}$ = -1.796 (à partir de la table de la loi de student)

t > $-t_{\alpha,11}$ = -1.796, comme la valeur de t ne tombe pas dans région critique, on ne peut pas rejeter H_0.

Conclusion : il n'y a pas suffisamment de preuves pour dire que l'échantillon vient d'une population dont la moyenne est inférieur à 37°C. L moyenne μ peut être bien inférieur à 37 °C mais les 12 valeurs de l'échantillon ne fournissent pas assez d'éléments pour le confirmer.



5) Test d'hypothèse pour une proportion

Conditions requises pour ce test

1. Les observations d'échantillon proviennent d'un échantillon aléatoire simple
2. Les conditions pour une loi binomiales sont satisfaites (un nombre fixés d'essais indépendants avec une même probabilité et chaque essai n'a que deux résultats « succés » et « échec »)
3. np> 5, nq > 5 sont toutes les deux satisfaites ; ainsi la loi binomiale peut être approximée par une loi normale avec μ = np et σ = $\sqrt{npq}$

Notations

n= taille d'échantillon ou nombre d'essais

$\hat{p}=\frac{x}{n}$ (proportion d'échantillon)

p= proportion de la population utilisée dans H_0

Statistique de test

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

Valeurs critiques utiliser une table de loi normale.

Exemple Nous avons noté dans un exemple précédent, qu'une expérience sur des pois aboutissait à 580 nouveaux pois, dont 26.2% à gousse jaune. Mendel affirmait que la proportion de pois à gousse jaune devrait être égale à 25%.

On veut tester l'hypothèse H_0 : p = 0.25 au niveau de significativité α=0.05.

Solution on vérifie les hypothèses d'application du test ;

1. l'échantillon est aléatoire simple.
2. np=145 $\geq$ 5 et n(1-p)=445 $\geq$ 5, donc la loi normale peut être utilisée pour approximation de la loi binomiale. Les conditions sont satisfaites et le test peut être effectué.
   n=580, $\hat{p}$ = 0.262

H_0 : p=0.25

H_1 : p$\neq$0.25

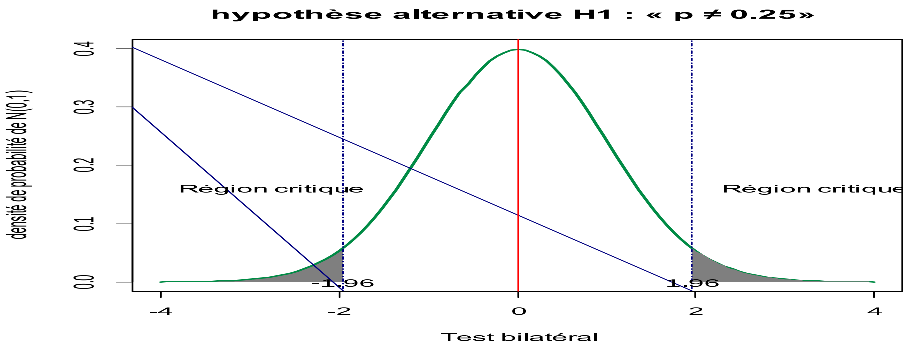
La statistique de test

$$z= \frac{\hat{p}-p}{\sqrt{\frac{p(1-p)}{n}}} = \frac{0.262-0.25}{\sqrt{\frac{0.25(1-0.25)}{580}}} = 0.67.$$

c'est un test bilatéral, à l'aide d'une table de loi normale, on trouve les valeurs critiques ;

$F(t_{\alpha/2})=1-0.05/2=0.975 \xRightarrow{table\ N(0,1)} t_{\alpha/2}=1.96,\ -t_{\alpha/2}=-1.96$

Z < 1.96, la statistique de test ne tombe pas dans la région critique d'où le non rejet deH_0.



Conclusion : on conclut qu'il n'y a pas suffisamment d'éléments pour garantir le rejet de l'affirmation de Mendel qu'il y a 25% de pois à gousse jaune.

6) test d'hypothèse pour une variance ou un écart type

Beaucoup d'organisations de productions et de service ont le même but : améliorer la qualité en réduisant la variation. Les ingénieurs de qualité veulent s'assurer qu'un produit a une moyenne acceptable, mais ils veulent aussi produire des articles de qualité à peu près constante pour qu'il y ait peu de défauts. L'objectif de cette section est de présenter une méthode pour tester des affirmations à propos de l'écart type ou de la variance d'une population.

Conditions d'application du test

1. l'échantillon est aléatoire simple
2. la population doit avoir une distribution normale.

Statistique de test

$$X^2=\frac{(n-1)s^2}{\sigma^2}$$

Valeurs critiques on utilise une table de khi deux avec ddl=n-1 pour le degrés de liberté.

Exemple la valeur du QI de adultes est normalement distribué avec une moyenne de 100 et un écart type de 15. Un échantillon aléatoire simple de 13 professeurs de statistiques fournit un écart type s=7.2, un psychologue affirme que les professeurs de statistiques ont un écart type égal à 15 qui est l'écart type de la population générale. Supposer que les scores des QI des professeurs de statistiques sont normalement distribués et utiliser un niveau de significativité de 0.05 pour tester l'affirmation que σ = 15. D'après ce résultat, que pouvez – vous conclure à propos de l'écart-type des sores de QI pour les professeurs de statistique?

Solution il faut d'abord vérifier que les conditions requises sont satisfaites. On suppose que l'échantillon est aléatoire simple. De plus, la population est normalement distribuée Donc les conditions sont vérifiées et le test peut être effectué.

H_0 : $\sigma = 15$

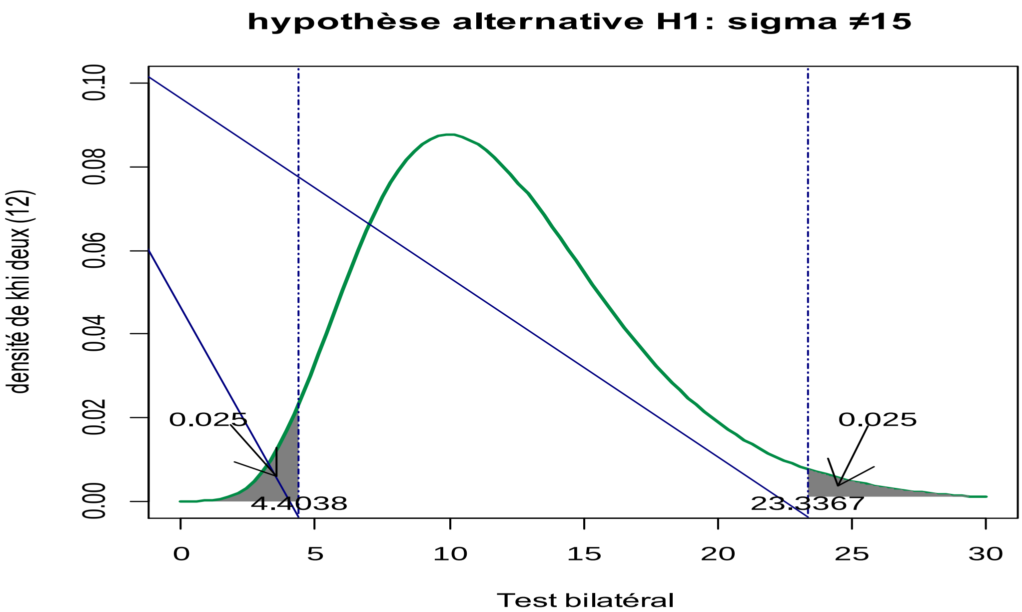
H_1 : $\sigma \neq 15$

Statistique de test

$$X^2 = \frac{(n-1)s^2}{\sigma^2} = \frac{(13-1)7.2^2}{15^2} = 2.765$$

Les valeurs critiques 4.404 et 23.337 dans la table de la loi de khi deux à la ligne 12 (ddl=13-1) pour les colonnes correspondant à 0.975 et 0.025.

Comme la statistique de test est dans la région critique on rejette H_0.



Conclusion on a suffisamment de preuves pour garantir le rejet de l'affirmation que l'écart type est égal à 15. Il aparraît que les professeurs de statistique ont des scores de QI avec un écart type différent de 15 qui est l'écart type pour la population générale.