

Université Abou Bakr Belkaid

1

Département d'Agronomie

Master 1 Production végétale S8 2019-2020

TESTS STATISTIQUES

Table des matières

1	Tests Statistiques	3
1.1	Test d'hypothèse avec un échantillon	3
1.1.1	Test d'hypothèse pour une proportion (fréquence) : . .	11

Chapitre 1

Tests Statistiques

1.1 Test d'hypothèse avec un échantillon

Introduction : Les deux activités majeures des statistiques inférentielles sont :

1. L'estimation des paramètres dans une population.
2. Les tests d'hypothèse.

Définition 1.1.1. *Une hypothèse est une affirmation ou un énoncé à propos d'une propriété d'une population.*

Définition 1.1.2. *Un test d'hypothèse est une procédure standard pour tester une affirmation (ou hypothèse).*

Définition 1.1.3. *Un test d'hypothèse (ou test de significativité) est une procédure standard pour tester un énoncé à propos d'une propriété d'une population.*

- Exemple 1.1.1.**
1. *Dans son expérience de croisement de pois à gousses jaunes avec des pois à gousses vertes Mendel affirme que sous certaines circonstances le pourcentage de pois à gousses jaunes est de 25% .*
 2. *Des chercheurs en médecine affirment que la température corporelle moyenne d'un adulte n'est pas égale à $37,0^{\circ}\text{C}$*

Bases des tests d'hypothèse :

Définition 1.1.4. *Les composantes formelles utilisées dans les tests d'hypothèse sont : hypothèse nulle, hypothèse alternative, statistique de test, région critique, niveau de significativité, valeurs critiques, p-value, erreur de première espèce et de deuxième espèce.*

Objectifs :

1. Pour une affirmation donnée, identifier l'hypothèse nulle, l'hypothèse alternative et les exprimer sous forme symbolique.
2. Etant donné une hypothèse et des données d'échantillon, calculer la valeur de la statistique de test.
3. Pour un niveau de significativité donné, identifier la valeur critique.
4. Etant donné la statistique de test, identifier la p-value.
5. Exprimer la conclusion du test d'hypothèse en termes simples et non techniques.
6. Quand on teste une affirmation donnée, identifier les erreurs de première et deuxième espèce.

Définition 1.1.5. *L'hypothèse nulle (notée H_0) est l'affirmation que la valeur d'un paramètre d'une population (comme une proportion, une moyenne ou un écart type) est égale à une certaine valeur supposée.*

Voici quelques hypothèses nulles typiques de celles considérées auparavant.

$$H_0 : p = 0,25; H_0 : \mu = 37,0; H_0 : \sigma = 15$$

Directement on teste l'hypothèse nulle au sens où on suppose qu'elle est vraie et où on arrive à une conclusion qui soit rejeter H_0 soit ne pas rejeter H_0).

Définition 1.1.6. *L'hypothèse alternative (notée H_1 ou H_a) est l'affirmation que le paramètre a une valeur différente de celle de l'hypothèse nulle. Pour les méthodes de notre cours, la forme symbolique de l'hypothèse alternative devra utiliser l'un des symboles $>$; $<$ ou \neq*

Voici neuf exemples :

1. *Proportion :*

$$H_1 : p > 0,5 \quad H_1 : p < 0,5 \quad H_1 : p \neq 0,5;$$

2. *Moyennes :*

$$H_1 : \mu > 37,0 \quad H_1 : \mu < 37,0 \quad H_1 : \mu \neq 37,0;$$

3. *Ecart types :*

$$H_1 : \sigma > 15 \quad H_1 : \sigma < 15 \quad H_1 : \sigma \neq 15.$$

Remarque 1.1.1. Bien que quelques livres utilisent les symboles \geq et \leq dans l'hypothèse nulle. Nous n'utiliserons que le signe égal comme la plupart des revues professionnelles

Exemple 1.1.2. Pour chaque affirmation, exprimez sous forme symbolique l'hypothèse nulle et l'hypothèse alternative correspondante.

1. La proportion de pois à gousses jaunes est égale à 0,25.
2. La taille moyenne d'un homme adulte est au plus de 183 cm.
3. L'écart type pour la taille des femmes adultes est supérieur à 6cm.

Solution 1. 1. L'affirmation donnée est exprimée par $p = 0,25$. Si $p = 0,25$ est fausse, alors $p \neq 0,25$ doit être vraie. On pose donc $H_1 : p \neq 0,25$ et $H_0 : p = 0,25$.

2. On exprime (la moyenne est au plus de 183cm) par la forme symbolique $\mu \leq 183$. Si la proposition $\mu \leq 183$ est fausse, alors la proposition $\mu > 183$ doit être vraie. On pose donc $H_1 : \mu > 183$ et $H_0 : \mu = 183$.

3. On exprime l'affirmation sous la forme symbolique par $\sigma > 6$. Si la proposition $\sigma > 6$ est fausse, la proposition $\sigma \leq 6$ doit être vraie. On pose donc $H_1 : \sigma > 6$ parce que'elle ne contient pas l'égalité et l'hypothèse nulle sera alors $H_0 : \sigma = 6$.

Définition 1.1.7. La statistique de test est une valeur calculée à partir des données d'échantillon. On utilisera les statistiques de test suivantes :

1. Statistique de test pour une proportion : $Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$.

2. Statistique de test pour une moyenne $Z = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{\sigma}{\sqrt{n}}}$.

ou $t = \frac{\bar{x} - \mu_{\bar{x}}}{\frac{s}{\sqrt{n}}}$.

3. Statistique de test pour un écart type $\chi^2 = \frac{(n-1).S^2}{\sigma^2}$

Exemple 1.1.3. Une expérience sur les pois aboutissait à 580 nouveaux pois, dont 26,2% à gousses jaunes.

Trouver la valeur de la statistique de test pour l'hypothèse de Mendel qui affirmait que la proportion de pois à gousses jaunes devrait être égale à 25%.

Solution 2. Compte tenu de l'expérience, il est raisonnable de supposer que l'échantillon est un échantillon aléatoire simple.

Les conditions d'une expérience binomiale sont satisfaites car il y a un nombre

fixe d'essais ($n = 580$) ; que les essais sont indépendants (la couleur d'une gousse n'affecte pas la probabilité de la couleur d'une autre gousse), qu'il n'y a que deux résultats (jaune, pas jaune) et que la probabilité de jaune reste constante. Au final, on utilise :

$n = 580$ et $p = 0,25$ pour voir que $np = 145 \geq 5$ et $nq = 445 \geq 5$, donc la loi normale peut être utilisée pour approximer la loi binomiale. Les conditions sont satisfaites et le test peut être effectué.

La statistique de test vaut :

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0,262 - 0,25}{\sqrt{\frac{0,25 \times 0,75}{580}}} = 0,67$$

Interprétation 1. Il semble que dans ce cas 26,2% n'est pas significativement différent de 25%.

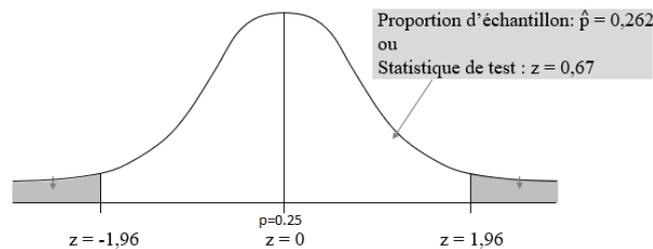


FIGURE 1.1 – Proportion de pois à gousses jaunes

Définition 1.1.8. Région critique :

La zone de rejet (ou région critique) est l'ensemble des valeurs de la statistique de test pour lesquelles on rejette l'hypothèse nulle H_0 .

Par exemple dans l'expérience génétique (voir la figure 3.1) Les zones du rejet de H_0 correspondent à $z < 1,96$ ou $z > 1,96$.

Puisque la statistique de test $z = 0,67$ appartient à l'intervalle $] -1,96; 1,96[$ donc l'hypothèse H_0 ne peut pas être rejetée.

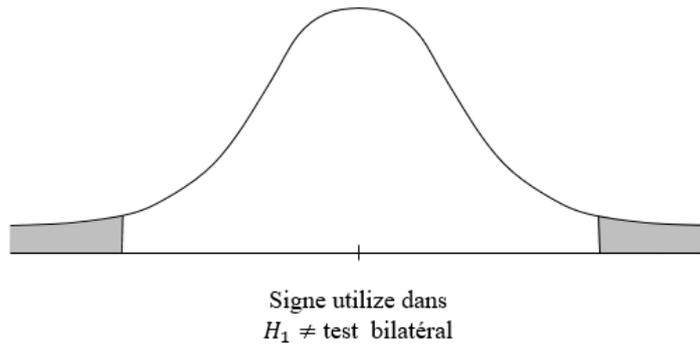
Définition 1.1.9. Niveau de significativité : Comme on avait défini le niveau de confiance pour un intervalle de confiance par la probabilité $1 - \alpha$, on dit alors que le niveau de significativité (noté α) est la probabilité que la statistique de test tombe dans la région critique quand l'hypothèse nulle est effectivement vraie.

Si la statistique de test tombe dans la région critique, on rejette l'hypothèse nulle est ainsi α est la probabilité de l'erreur de rejeter à tort l'hypothèse nulle alors qu'elle est vraie et α est appelée aussi erreur de première espèce.

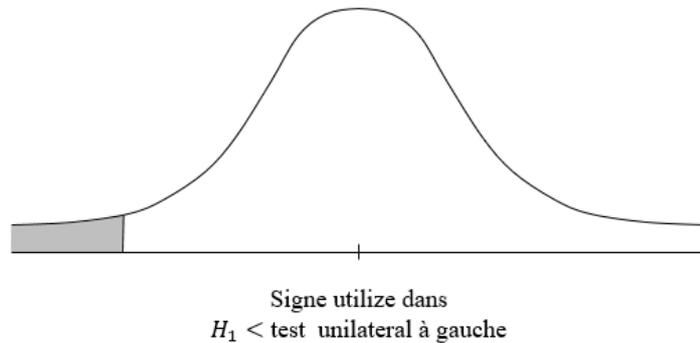
Les choix courants pour α sont 5% , 1% avec 0,05 le plus courant.

Définition 1.1.10. Valeur critique : C'est la valeur qui sépare la région où on rejette l'hypothèse nulle H_0 (région critique) des autres valeurs de la statistique de test qui ne mènent pas à ce rejet.

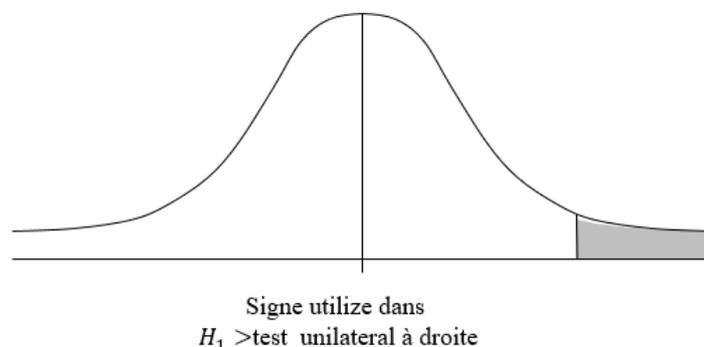
1. **Test bilatéral :** la région critique est dans les deux régions extrêmes (queues) sous la courbe.

FIGURE 1.2 – Signe dans $H_1; \neq$

2. **Test unilatéral à gauche :** la région critique est dans la région extrême gauche (queue) sous la courbe.

FIGURE 1.3 – Signe dans $H_1; <$

3. **Test unilatéral à droite :** la région critique est dans la région extrême droite (queue) sous la courbe.

FIGURE 1.4 – Signe dans $H_1; >$

Définition 1.1.11. *La p-value (ou valeur-p ou probabilité critique) : En supposant que l'hypothèse nulle est vraie, la probabilité d'obtenir une valeur de la statistique de test qui est au moins aussi extrême que celle qui représente les données d'échantillon.*

On n'accepte pas l'hypothèse nulle (rejet de H_0) si la p-value est très petite comme 0,05 ou moins.

Les p-values peuvent être trouvées comme l'indique la figure ci après (FIGURE 3.5- Méthode de la p-value).

Décisions et conclusions :

La procédure standard pour tester une hypothèse oblige de toujours tester l'hypothèse nulle et donc notre conclusion initiale est toujours l'une des deux conclusions suivantes :

1. Rejet de l'hypothèse nulle .
2. Echec du rejet de l'hypothèse nulle.

Critère de décision :

La décision de rejet ou d'échec de rejet pour l'hypothèse nulle est généralement faite par les trois méthodes suivantes à savoir :

1. **Méthode traditionnelle** : Si la statistique de test tombe dans la région critique, on rejette l'hypothèse nulle et si la statistique de test ne tombe pas dans la région critique il y a échec de rejet de l'hypothèse nulle.
2. **Méthode de la p-value** : On rejette l'hypothèse nulle si la p-value est $\leq \alpha$ (où α est le niveau de significativité, par exemple 0,05).

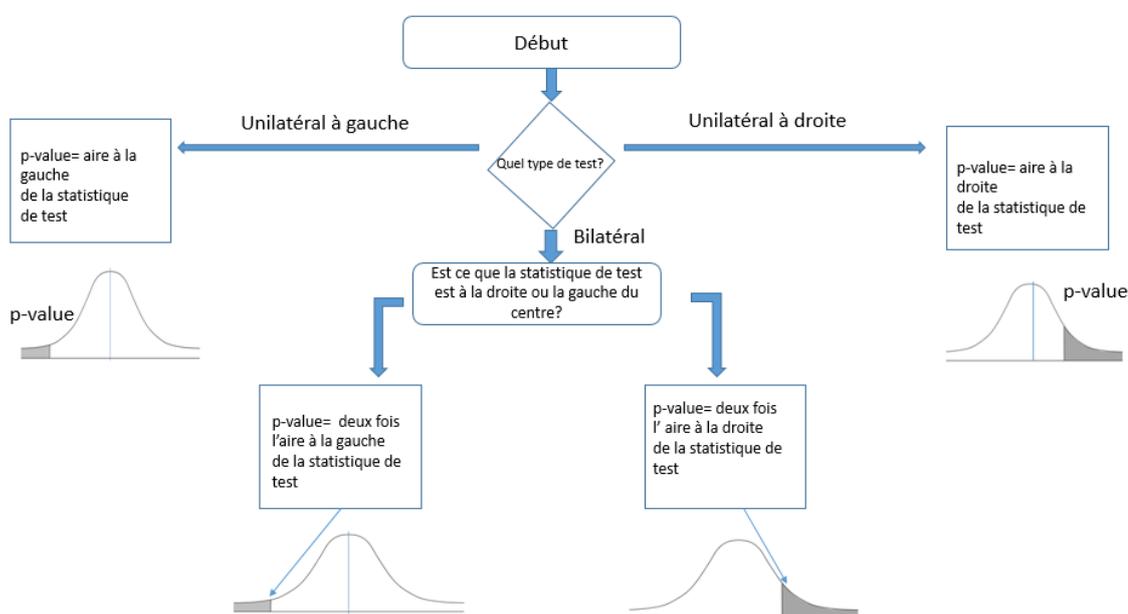


FIGURE 1.5 – Méthode de la p-value

3. **Autre option** : Au lieu d'utiliser un niveau de significativité comme $\alpha = 0,05$ identifiez simplement la p-value et à nous de décider.
4. **Intervalle de confiance** : Comme les valeurs probables d'un paramètre appartiennent à son intervalle de confiance donc on rejette l'affirmation que la valeur du paramètre de la population n'appartienne pas à cet intervalle de confiance.

Remarque 1.1.2. Dans ce cours, nous utilisons le critère de décision qui met en jeu la comparaison d'un niveau de significativité et de la p-value.

Exemple 1.1.4. Déterminez d'abord si les conditions fournies correspondent à un test bilatéral, unilatéral à gauche ou unilatéral à droite, puis trouvez la p-value et, enfin, donnez votre conclusion à propos de l'hypothèse nulle.

1. On utilise le niveau de significativité de $\alpha = 0,05$ pour tester que $p > 0,25$. et les données d'échantillon donnent comme résultat une statistique de test $z = 2,34$.
2. On utilise le niveau de significativité de $\alpha = 0,05$ pour tester que $p \neq 0,25$. et les données d'échantillon donnent comme résultat une statistique de test $z = 0,67$ Utilisez un

Solution 3. 1. Avec l'affirmation $p > 0,25$ le test est unilatéral à droite. On peut trouver la p-value à l'aide de la figure (3.5). Parce que le test est

unilatéral à droite, la figure (3.5) montre que la p -value est l'aire à la droite de la statistique de test $z = 2,34$. Une table de la loi normale montre que l'aire à droite de $2,34$ est $1-0,9904=0,0096$. La p -value $0,0096$ est inférieur ou égale au niveau de significativité $\alpha = 0,05$ donc on rejette l'hypothèse. La p -value $0,0096$ est relativement petite, ce qui indique que le résultat d'échantillon ne survient pas facilement par hasard.

2. Avec l'affirmation $p \neq 0,25$ le test est bilatéral. On peut trouver la p -value à l'aide de la figure (3.5). Parce que le test est bilatéral et parce que la statistique de test

$z = 0,67$ est à la droite du centre, la p -value est deux fois l'aire à la droite de $z = 0,67$. Une table de la loi normale montre que l'aire à droite de $z = 0,67$ est $1-0,7486=0,2514$ donc la p -value est $2 \times 0,2514 = 0,5028$. La p -value $0,5028$ est plus grande que le niveau de significativité $\alpha = 0,05$ donc il y a échec du rejet de l'hypothèse nulle. La grande p -value $0,5028$ montre que le résultat d'échantillon peut facilement survenir par hasard.

1.1.1 Test d'hypothèse pour une proportion (fréquence) :

Tester des hypothèses à propos d'une proportion p :

1. Les observations d'échantillon proviennent d'un échantillon aléatoire simple. Il ne faut jamais oublier l'importance des méthodes d'échantillonnage.
2. Les conditions pour une loi binomiale sont satisfaites :
Il y a un nombre fixé d'essais indépendants avec une même probabilité et chaque essai n'a que deux résultats, nommés "succès" et "échec".
3. Les conditions $np \geq 5$ et $nq \geq 5$ sont toutes les deux satisfaites ; ainsi la loi binomiale peut être approximée par une loi normale avec $\mu = np$ et $\sigma = \sqrt{npq}$.

Notation 1.

- $n =$ taille d'échantillon ou nombre d'essais.
- $\hat{p} = \frac{x}{n}$ (proportion d'échantillon)
- $p =$ proportion de la population (utilisée dans l'hypothèse nulle).
- $q = 1 - p$

Définition 1.1.12. La statistique de test pour tester une affirmation à propos d'une proportion est :

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}}$$

Valeurs critiques : Utiliser la table de la loi normale.

P -value : suivant le genre du test.

Exemple 1.1.5. Une expérience sur les pois aboutissait à 580 nouveaux pois, dont 26,2% à gosses jaunes.

Mendel affirmait que la proportion de pois à gousse jaune devrait être égale à 25%.

1. Résumez les données de l'échantillon et de l'affirmation.
2. Illustrez le test d'hypothèse à l'aide de la méthode des valeurs critiques.
3. Illustrez le test d'hypothèse à l'aide de la méthode de la P -value

Solution 4.

- Compte tenu de l'expérience, il est raisonnable de supposer que l'échantillon est un échantillon aléatoire simple.

- Les conditions d'une expérience binomiale sont satisfaites car il y a un nombre fixe d'essais ($n = 580$); que les essais sont indépendantes (la couleur d'une gousse n'affecte pas la probabilité de la couleur d'une autre gousse), qu'il n'y a que deux résultats (jaune, pas jaune) et que la probabilité de jaune reste constante. Au final, on utilise :

$n = 580$ et $p = 0,25$ pour voir que $np = 145 \geq 5$ et $nq = 445 \geq 5$, donc la loi normale peut être utilisée pour approximer la loi binomiale. Les conditions sont satisfaites et le test peut être effectué.

- Méthode des valeurs critiques :

1. Hypothèse de test :

$$H_0 : p = 0,25$$

$H_1 : p \neq 0,25$ On prend comme niveau de significativité

$$\alpha = 5\%$$

Calculons la statistique de test :

$$Z = \frac{\hat{p} - p}{\sqrt{\frac{pq}{n}}} = \frac{0,262 - 0,25}{\sqrt{\frac{0,25 \times 0,75}{580}}} = 0,67$$

C'est un test bilatéral. Et parce que la statistique de test 0,67 ne tombe pas dans la région critique, on ne peut pas rejeter H_0

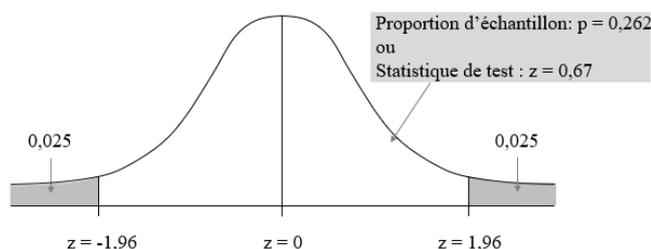


FIGURE 1.6 – Méthode de la p-value

Conclusion : 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 gousses jaunes.

2. Utilisons la P-value :

$$H_0 : p = 0,25$$

$$H_1 : p \neq 0,25$$

— La statistique de test est $Z = 0,67$

Il s'agit d'un test bilatéral, la P-value est le double de l'aire à droite de $Z = 0,67$. À l'aide d'une table de la loi normale il y a une aire de 0,7486 pour l'aire cumulée à gauche de la statistique de test. L'aire à droite de $Z = 0,67$ est donc $1 - 0,7486 = 0,2514$ qui doit être doublée, soit une P-value de 0,5028.

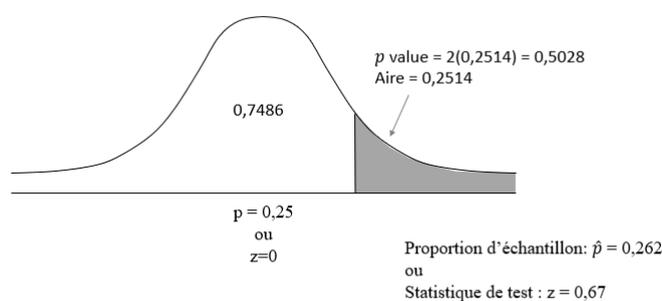


FIGURE 1.7 – Méthode de la p-value

Conclusion : 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 gousses jaunes

EXERCICES

Exercice 1.1.1. Dans une des fameuses expériences d'hybridation de Mendel, 8023 pois ont été obtenus et 24,94% ont des fleurs vertes. Les autres ont des fleurs blanches. Considérez un test d'hypothèse qui utilise un niveau de significativité 0,05 pour tester que les pois à fleurs vertes surviennent avec un taux de 25%.

1. Quelle est la statistique de test ?
2. Quelles sont les valeurs critiques ?
3. Quelle est la P -value ?
4. Quelle est la conclusion ?
5. Est-ce qu'un test d'hypothèse peut être utilisé pour "prouver" que le taux de pois à fleurs vertes est de 25% comme affirmé ?

Exercice 1.1.2. Supposez qu'une méthode de choix de sexe est testée avec des couples qui veulent avoir des filles.

Identifiez l'hypothèse nulle, l'hypothèse alternative, la statistique de test, la P -value ou les valeurs critiques, la conclusion sur l'hypothèse nulle et la conclusion finale qui décide de l'affirmation originale. On utilisera la méthode de la P -value.

1. Parmi 100 bébés nés de 100 couples qui utilisent une méthode de choix de sexe pour augmenter les chances d'avoir une fille, 65 sont des filles. Utilisez un niveau de significativité 0,01 pour tester l'affirmation que pour les couples qui utilisent cette méthode, la proportion des filles est plus grande que 0,5.
2. Parmi 50 bébés nés de 50 couples qui utilisent une méthode de choix de sexe, 15 sont des garçons. Utilisez un niveau de significativité 0,05 pour tester l'affirmation que pour les couples qui utilisent cette méthode, la proportion de garçon est plus petite que 0,5.

Bibliographie

- [1] MF.Triola, Bio statistique pour les sciences de la vie et de la santé
- [2] MF.Triola, Exercices de statistiques et probabilités avec rappel de cours en 12 fiches ,(MAURICE LETHIELLEUX).
- [3] MF.Triola, Flash : Probabilités et statistique DEUG et ST BTS biologiques et agricoles.
- [4] MF.Triola, Statistiques et probabilités en 30 fiches ,(DANIEL FREDON),(DUNOD).