

Diplôme : BTSA Famille 1

Analyses agricoles, biologiques et biotechnologiques
Sciences et technologies des aliments
Productions animales
Aquaculture
Gestion forestière
Productions horticoles
Technologies végétales
Viticulture-Cœnologie
Développement de l'agriculture des régions chaudes

Module M41 : Traitement de données

Objectif général du module :

Choisir, dans une situation donnée, un modèle mathématique adapté au traitement de données.

Il s'agit d'un module fondé sur trois objectifs. Les deux premiers sont communs à tous les BTSA et constituent la base d'une culture commune statistique à tous les étudiants titulaires du diplôme.

Le troisième objectif, spécifique à l'option, tient compte des besoins professionnels relatifs à l'acquisition d'outils statistiques de base. Il s'agit de présenter aux étudiants les bases de la statistique inductive et de leur faire saisir les difficultés d'interprétation en considérant des modèles probabilistes. Ces modèles probabilistes sont abordés uniquement en vue de leur utilisation en statistique. Des situations suffisamment concrètes et issues du domaine professionnel permettent de donner du sens à cette approche. Les développements théoriques sont réduits et toujours présentés dans un cadre simple afin de donner du sens aux notions développées. Enfin des situations pluridisciplinaires valorisent et permettent d'appréhender les contenus.

Indications de contenus, commentaires,
recommandations pédagogiques

Objectif 1 : Utiliser les notions de statistique en vue d'une modélisation a priori.

-réalisation d'une modélisation simple : construire un ajustement affine avec un ou des changements de variable,

Dans le cadre de travaux dirigés, les connaissances acquises les années antérieures sur les séries statistiques à une variable (paramètres de position et de dispersion) sont à consolider.
Des situations issues de la vie économique, des sciences et techniques sont exploitées pour des études d'ajustement. On distingue variable explicative et variable expliquée. On veille à attirer l'attention des étudiants sur l'étude des résidus (on vérifie que leur moyenne est nulle).
Le coefficient de détermination peut être introduit comme quotient de la variance expliquée par la variance totale.
Le coefficient de corrélation linéaire est utilisé avec précaution. A partir de quelques exemples judicieux, le risque de mauvaise interprétation de ce coefficient et sa sensibilité aux valeurs extrêmes sont mis en évidence.
La calculatrice et le tableur sont largement utilisés.
Les fonctions logarithmes, exponentielles, puissances sont utilisées dans les changements de variable.

-détermination de la loi de probabilité d'une variable aléatoire discrète : calculer et interpréter les paramètres de cette variable,

Les notions de variable aléatoire, espérance mathématique, variance, sont présentées uniquement à partir d'exemples simples. Des tirages avec ou sans remise peuvent servir de support à des situations aléatoires.

-étude d'un couple de variables aléatoires discrètes : déterminer les lois marginales à partir d'une loi conjointe et reconnaître une situation de dépendance ou d'indépendance.

Les couples de variables aléatoires sont présentés uniquement à partir d'exemples simples.
Les étudiants doivent connaître :

- la somme de deux variables aléatoires,
- l'espérance mathématique de $aX+b$ et de $X+Y$.
- la variance de $aX+b$ et de $X+Y$ dans le cas où X et Y sont indépendantes.

-identification de situations où interviennent des lois usuelles discrètes et justification leur utilisation.

Sont étudiées la loi de Bernoulli et la loi binomiale ainsi que l'espérance mathématique et variance des variables sous-jacentes.

Une variable aléatoire de loi binomiale peut être présentée comme somme de variables de Bernoulli de même loi. D'autres lois en liaison avec le domaine professionnel peuvent être introduites.

-utilisation de variables aléatoires continues et en particulier de la fonction de répartition pour calculer des probabilités.

La notion de densité de probabilité est présentée en considérant des fonctions simples continues sur un intervalle.

La fonction de répartition F d'une variable aléatoire X est définie sur \mathbb{R} par $F(t) = P(X \leq t)$.

Pour une densité f définie sur $[a, b]$:

- pour $t < a$, $F(t) = 0$
- pour $a \leq t \leq b$, $F(t) = \int_a^t f(u) du$
- pour $t > b$, $F(t) = 1$.

Sont abordées l'espérance mathématique, la variance, l'écart-type d'une variable aléatoire continue ainsi que l'interprétation de ses paramètres.

Une loi uniforme est un bon support pour

- introduire la notion de densité,
- déterminer la fonction de répartition,
- calculer l'espérance mathématique et la variance.

La somme de deux variables aléatoires de même loi uniforme peut être simulée sur calculatrice ou tableur pour conjecturer sa densité.

Dans l'étude de variables aléatoires continues, on veille à lier calcul de probabilité, calcul intégral et calcul d'aire.

On peut être amené à utiliser les notations d'intégrales impropres mais aucune connaissance sur cette notion n'est exigible.

La notion de distribution normale est abordée à partir d'histogrammes.

Les étudiants doivent connaître la loi normale centrée réduite, la loi de $aX+b$ et de $X+Y$ lorsque X et Y sont des variables aléatoires de lois normales indépendantes.

Ils doivent être familiarisés avec la lecture directe et inverse de table.

Les représentations graphiques de quelques densités de variables aléatoires de lois normales judicieusement choisies permettent de mettre en évidence l'importance de l'écart-type et de l'espérance mathématique.

Est abordé l'approximation d'une loi binomiale par une loi normale : Le traitement par tableur d'exemples illustre

efficacement le bien fondé de l'approximation.

Il convient de mettre en évidence la raison d'être de la correction de continuité lors de l'approximation d'une loi binomiale par une loi normale.

Objectif 2 : Utiliser les notions de statistique et de probabilités en vue d'une estimation et d'une prise de décision

-distribution d'échantillonnage : savoir prélever un échantillon de façon aléatoire simple, déterminer les lois des variables aléatoires d'échantillonnage des moyennes et des proportions

L'échantillonnage aléatoire simple correspond à des tirages successifs équiprobables et indépendants les uns des autres.

La distribution d'échantillonnage sera ainsi définie :

- on imagine que l'on prélève, dans une population, tous les échantillons de taille n .
- selon la nature du caractère étudié, à chaque échantillon sont associés une moyenne \bar{x} et une variance s^2 , ou la proportion f d'une modalité.

La distribution de l'ensemble des \bar{x} (respectivement des s^2 , des f) est la distribution d'échantillonnage des moyennes. (respectivement des variances, des proportions).

On définit les variables aléatoires \bar{X} , S^2 , F dont on précise les espérances mathématiques. Ces paramètres peuvent être conjecturés lors d'un TP.

Lorsque la loi de la variable aléatoire sous-jacente X est normale, la loi de \bar{X} est elle-même une loi normale.

Dans les autres cas lorsqu'il s'agit de grands échantillons, les lois de \bar{X} et de F sont approchées par des lois normales. Ce résultat peut faire l'objet d'un TP.

-estimation : déterminer une estimation ponctuelle ou par intervalle de confiance en liaison avec les variables d'échantillonnage

Les espérances mathématiques précédentes permettent de définir des estimations ponctuelles

Les intervalles de confiance sont introduits à partir d'exemples en adoptant la démarche suivante :

- choisir une variable d'échantillonnage dont on explicite la loi de probabilité
- construire un intervalle aléatoire
- établir l'intervalle de confiance.

On considère essentiellement des intervalles de confiance à 0,95 (ou 0,99) symétriques en probabilité.

Le niveau de confiance 0,95 est la probabilité que la méthode fournisse un intervalle contenant le paramètre estimé.

Ce résultat peut faire l'objet d'un TP.

On sensibilise les étudiants à la double influence de la taille d'échantillon et de la variabilité du phénomène étudié sur l'amplitude de l'intervalle de confiance.

La loi de Student est introduite dans le cadre de l'estimation par intervalle de confiance de la moyenne d'un caractère distribué selon une loi normale de variance inconnue.

-statistique inférentielle bidimensionnelle : mise en œuvre d'un test d'indépendance.

Le test retenu est le test du χ^2 d'indépendance de deux caractères qualitatifs.

On organise des données sous forme de tableau de contingence.

On s'assure que les effectifs théoriques sont supérieurs à 5 et on exprime l'hypothèse H_0 en terme d'indépendance de deux caractères qualitatifs.

Objectif 3 : Acquérir des outils mathématiques de base nécessaires à l'interprétation de résultats expérimentaux

-tests d'hypothèse : mise en œuvre des différents tests statistiques classiques et interprétation des résultats obtenus :

Les tests sont introduits à partir d'exemples en adoptant la démarche suivante :

- choisir une hypothèse H_0 et une hypothèse H_1 déduites du problème posé (l'hypothèse H_1 déterminant le caractère bilatéral ou unilatéral du test)
- choisir une variable de décision dont on explicite la loi de probabilité
- déterminer la région critique pour un risque α donné
- énoncer la règle de décision
- conclure.

Les résultats obtenus par logiciel sont à expliciter en liaison avec cette démarche.

Les deux types d'erreur, de première et de deuxième espèce sont identifiés.

Document d'accompagnement - Inspection de l'Enseignement Agricole

3

Diplôme : BTS Analyses agricoles, biologiques et biotechnologiques / Sciences et technologies des aliments / Productions animales / Aquaculture / Gestion forestière / Productions horticoles / Technologies végétales / Viticulture-Œnologie / Développement de l'agriculture des régions chaudes.

Module : M41 : Traitement de données

Date : 6 juillet 2009

-conformité d'une proportion, d'une moyenne, comparaison de proportions, de moyennes, de variances.

Le test de conformité d'une proportion est traité uniquement dans le cas des grands échantillons.

On rappelle que la variable aléatoire nF est distribuée suivant la loi binomiale de paramètres n et p .

Si n est grand, la loi de nF peut être approchée par une loi normale.

Le cas des petits échantillons peut être évoqué en considérant une loi binomiale.

En complément du cours, on utilise un logiciel de traitement de données pour conduire et interpréter des résultats obtenus dans le cas de situations concrètes étudiées en activités pluridisciplinaires.

La comparaison de deux moyennes est uniquement traitée dans le cas d'échantillons indépendants.

La comparaison des variances à l'aide d'un test de FISCHER est mise en œuvre dans des études d'homogénéité de lots