STATISTIQUE INFÉRENTIELLE
DANS LES ÉPREUVES TERMINALES DE BTSA

Les épreuves nationales des BTSA, dont la plupart ne concernent qu'un nombre limité de candidats, vont néanmoins bien au-delà de leur mission première d'évaluer pour les enseignants que nous sommes. Nous pouvons les considérer comme des documents de référence pour l'organisation de nos modules soumis au contrôle continu en CCF.

Nous présentons ici des corrigés d'extraits d'une épreuve de mathématiques et d'épreuves professionnelles, la part de statistique inférentielle étant loin d'être négligeable dans ces dernières. En particulier, les référentiels des BTSA STA et ANABIOTEC comportent une partie *Statistique* assez importante dans leurs modules professionnels : M54 pour le premier et M53 pour le second. Leurs destins au niveau de l'évaluation sont assez croisés. En effet, feue l'épreuve ET2, épreuve scientifique du BTSA IAA, a fait place à l'épreuve E6 du BTSA STA alors que les étudiants de BTSA ANABIOTEC ont eu l'*agréable* surprise de voir apparaître une épreuve scientifique E7-1 en épreuve nationale alors qu'ils n'en avaient pas avant la réforme.

La première épreuve de référence que nous avons choisie est l'épreuve E4 2014. Commune à tous les BTSA, elle peut porter sur les objectifs 1 et 2 du module M41 et donc sur des notions de statistique inférentielle de l'objectif 2. Ensuite, nous proposons successivement un corrigé d'un extrait[1] (partie statistique) de l'épreuve E6 de BTSA STA 2014 et un corrigé d'un extrait (partie statistique) de l'épreuve E7-1 du BTSA ANABIOTEC 2014. Notons que nous aurions pu également proposer l'épreuve E5 du BTSA ANABIOTEC qui, en général, contient un exercice de statistique.

BTSA toutes options – Métropole Réunion - Épreuve E4 – Session 2014

Exercice 2 (5 points)

La verse est un accident de végétation atteignant les céréales, provoqué par la pluie ou le vent, et couchant les tiges au sol. Elle peut être due entre autres, à une fumure déséquilibrée (excès d'azote) provoquant une croissance exagérée des tiges.

Une étude statistique a été menée sur 200 parcelles de blé d'un même terroir sur lesquelles différents apports d'azote ont été réalisés, afin d'étudier un éventuel effet de cet apport sur l'apparition de la verse.

Les résultats sont consignés dans le tableau suivant :

Intensité de la verse / Apport d'azote	Parcelles non touchées par la verse	Parcelles peu touchées par la verse	Parcelles très touchées par la verse
Faible apport d'azote	30	13	7
Apport raisonné d'azote	49	37	14
Excès d'azote	17	18	15

Au seuil de risque 0,05, peut-on considérer que l'apport d'azote a une influence sur l'apparition de la verse ?

[1] Les sujets complets sont disponibles sur le site R2math :
http://r2math.enfa.fr/ressources-pedagogiques/py-math/py-math-bulletin-24-mai-2015/

Corrigé :

Réalisons un test du khi-deux d'indépendance.

- Les hypothèses de ce test sont :
  H_0 : "L'apport d'azote n'a pas eu d'influence sur l'apparition de la verse"
  H_1 : "L'apport d'azote a eu une influence sur l'apparition de la verse"

- Sous l'hypothèse d'indépendance H_0, on peut calculer les effectifs théoriques attendus.

Par exemple pour le premier : $\frac{30 + 13 + 7}{200} \times (30 \times 49 + 17) = \frac{50 \times 96}{200} = 24$. Ces effectifs $\widehat{n_{ij}}$ sont consignés dans le tableau suivant :

		Intensité de la verse		
		non	peu	très
Apport d'azote	faible	24	17	9
	raisonné	48	34	18
	excès	24	17	9

- Sous l'hypothèse d'indépendance H_0, la variable aléatoire $T = \sum_{(i;j)} \frac{\left(N_{ij} - \widehat{n_{ij}}\right)^2}{\widehat{n_{ij}}}$ est distribuée approximativement selon la loi de Khi-deux à $\nu = (3-1)(3-1) = 4$ degrés de liberté avec N_{ij} les variables aléatoires qui, à chaque échantillon de 200 parcelles, associent les effectifs de chacun des couples de modalités des variables (faible , non), (faible , peu)...

- Notons t_{obs} la valeur observée de T sur l'échantillon.
  Le test étant unilatéral à droite, cette valeur doit être comparée à la valeur critique $\chi_{1-\alpha}$. La règle de décision est la suivante :
  - si $t_{obs} \geq \chi_{1-\alpha}$ alors on rejette H_0 ;
  - si $t_{obs} < \chi_{1-\alpha}$ alors on n'est pas en mesure de rejeter H_0.

- Valeur observée de T : $t_{obs} = \sum_{(i;j)} \frac{\left(n_{ij} - \widehat{n_{ij}}\right)^2}{\widehat{n_{ij}}}$ où les n_{ij} sont les valeurs observées des N_{ij} sur l'échantillon de 200 parcelles étudié.

$$t_{obs} = \frac{(30-24)^2}{24} + \frac{(13-17)^2}{17} + \frac{(7-9)^2}{9} + \frac{(49-48)^2}{48} + \frac{(37-34)^2}{34} + \frac{(14-18)^2}{18} + \frac{(17-24)^2}{24} + \frac{(18-17)^2}{17} + \frac{(15-9)^2}{9} \approx 10{,}06.$$

- Le seuil est $\alpha = 0{,}05$, la valeur critique associée est $\chi_{0{,}95} = 9{,}49$ (lecture dans la table).

- Comme $10{,}06 \geq 9{,}49$ alors $t_{obs} \geq \chi_{0{,}95}$: on rejette l'hypothèse d'indépendance H_0.

- **Conclusion :** Au seuil de risque de 0,05, nous pouvons considérer que l'apport d'azote a eu de l'influence sur l'apparition de la verse.

Énoncé (extrait)

La PME CROQ'LEG, installée à Carpentras, produit des légumes de quatrième gamme pour la grande distribution. Dans son atelier réfrigéré à 6°C, elle est en train de mettre en place la fabrication d'un nouveau produit : mélange de légumes conditionnés en sachets de 1 kg destiné à la préparation de soupe.

L'entreprise utilise exclusivement des matières premières d'origine française.
[...]

Pour respecter la réglementation (en particulier le décret 78-166), le cahier des charges interne à l'entreprise précise que la masse réelle d'un produit préemballé dont la masse nominale est de 1 000 g doit être comprise entre 985 g et 1 015 g.

Une étude précédente a montré que pour l'ensemble de la production, les masses exprimées en grammes sont distribuées selon une loi normale.

1.3 Afin d'éviter la mise en vente de produits hors tolérance, on admet que l'amplitude 30 de l'intervalle [985 ; 1 015] permet d'estimer l'écart type $\hat{\sigma}$ de la production en utilisant la relation suivante : $8.\hat{\sigma} = 30$.

En déduire la valeur de $\hat{\sigma}$.

1.4 On souhaite vérifier que la variance de la production n'est pas supérieure à $3{,}75^2$ (notée σ_0^2). Pour cela, on prélève de manière aléatoire un échantillon de 51 sachets de produits finis. On note n la taille de cet échantillon. L'écart type de ces 51 masses est 4,5. On le note s.

1.4.1 Mettre en œuvre un test statistique permettant de répondre à la question suivante :
Peut-on considérer que la variance de la production est supérieure à $3{,}75^2$?
(On prendra un seuil de risque de première espèce de 0,05.)

1.4.2 Conclure sur la conformité de la production.

On rappelle que, sous certaines conditions, la variable aléatoire K définie par $K = \dfrac{n\,S^2}{\sigma_0^2}$ est distribuée selon la loi du χ^2 à $n - 1$ degrés de liberté.

La table du χ^2 à k degrés de liberté est donnée au **document 6**.

Note de la rédaction :

La question 1.3 peut soulever des commentaires, à commencer par la capacité évaluée dans cette question. De plus, une confusion fâcheuse au niveau des écarts types est glissée dans la phrase : "On admet que l'amplitude 30 de l'intervalle [985 ; 1 015] permet d'estimer l'écart type $\hat{\sigma}$ de la production en utilisant la relation suivante : $8.\hat{\sigma} = 30$".
Il eût été préférable de dire : "On admet que l'amplitude 30 de l'intervalle [985 ; 1 015] permet d'estimer l'écart type σ de la production par la valeur, notée $\hat{\sigma}$, devant vérifier la relation suivante : $8.\hat{\sigma} = 30$".

Corrigé :

1.3. $\hat{\sigma} = \dfrac{30}{8} = 3{,}75$ g.

1.4.1. Le test statistique à mettre en place est un test de conformité de la variance à une variance théorique. Le seuil de risque choisi est $\alpha = 0{,}05$.

- Les hypothèses de ce test sont :
  H_0 : "$\sigma^2 = 3{,}75^2$"
  H_1 : "$\sigma^2 > 3{,}75^2$"

  Le test est donc unilatéral à droite.

- Sous l'hypothèse H_0, la variable aléatoire $K = \dfrac{n\,S^2}{\sigma_0^2} = \dfrac{51\,S^2}{3{,}75^2}$ est distribuée selon la loi du Khi-deux à $n - 1 = 51 - 1 = 50$ degrés de liberté avec S la variable aléatoire qui, à chaque échantillon de 51 sachets, associe l'écart type des masses.

- Notons k_{obs} la valeur observée de K sur l'échantillon.
  Le test étant unilatéral à droite, cette valeur doit être comparée à la valeur critique $\chi_{1-\alpha}$. La règle de décision est la suivante :
  - si $k_{obs} \geq \chi_{1-\alpha}$ alors on rejette H_0 ;
  - si $k_{obs} < \chi_{1-\alpha}$ alors on n'est pas en mesure de rejeter H_0.

- Valeur observée de K : $k_{obs} = \dfrac{n\,s^2}{\sigma_0^2} = \dfrac{51 \times 4{,}5^2}{3{,}75^2} = 73{,}44$.

- Le seuil est $\alpha = 0{,}05$, la valeur critique associée est $\chi_{0{,}95} = 67{,}50$ (lecture dans la table).

- Comme $73{,}44 \geq 67{,}50$ alors $k_{obs} \geq \chi_{1-\alpha}$: on rejette l'hypothèse H_0.

- **Conclusion :** Au seuil de risque de 0,05, nous pouvons considérer que la variance de la production est significativement supérieure à $3{,}75^2$.

1.4.2. De la conclusion précédente, on en déduit que la production n’est pas conforme.

BTSA option ANABIOTEC – France Métropolitaine - Épreuve E7-1 – Session 2014

Note de la rédaction :

- *la question **3** de la partie 2 porte sur le test de conformité d'une variance à une variance théorique ;*
- *la question **4.1** de la partie 3 vérifie si l'étudiant sait utiliser les résultats d'une analyse de la variance à un facteur ;*
- *la question **4.2** de la partie 3 porte sur le test de comparaison de moyennes dans le cas d'échantillons appariés.*

Partie 2 : Dosage des nitrites dans le magret séché tranché par un laboratoire prestataire de service

[...]

3. Reproductibilité de la méthode de Griess

Le laboratoire participe à une étude inter-laboratoire. Un même échantillon est envoyé à 10 laboratoires différents ; les résultats obtenus par chaque laboratoire par la méthode de Griess, sont consignés dans le tableau ci-dessous :

Laboratoires	1	2	3	4	5	6	7	8	9	10
Teneur en $NaNO_2$ ($mg.kg^{-1}$)	80	81	81	84	84	90	87	79	82	87

On souhaite tester la reproductibilité de la méthode. La reproductibilité n'est pas satisfaisante lorsque les résultats sont trop fluctuants entre les laboratoires ; on admet que ceci se produit lorsque la variance de la teneur en nitrites est supérieure à 12.

3.1. Calculer, à 10^{-4} près, la variance s^2 des 10 résultats. Les calculs intermédiaires ne sont pas demandés.

3.2. On cherche à savoir si la variance de la teneur en nitrites est satisfaisante, en effectuant un test de conformité de la variance. Le seuil de risque α est fixée à 0,05.

3.2.1. Formuler l'hypothèse nulle H_0 et l'hypothèse alternative H_1.

3.2.2. On rappelle que sous H_0, avec les notations usuelles, la variable aléatoire $\dfrac{n\,S^2}{\sigma^2}$ est distribuée selon la loi du Khi2 à $n-1$ degrés de liberté.

Calculer, à 10^{-2} près, la valeur observée notée χ_0^2.

Déterminer la valeur critique $\chi^2_{1-\alpha\,;\,n-1}$ et rédiger la conclusion. On pourra utiliser [*la table du*] **document 6**.

3.2.3. Conclure quant à la reproductibilité inter-laboratoire de la méthode.

Corrigé :

3.1. Sans faire de calculs intermédiaires, il suffit de lire l'écart type s fourni par la calculatrice pour en déduire la variance.

Note de la rédaction :

Le résultat attendu pour s^2 doit être arrondi à 10^{-4} près, mais par ailleurs le calcul (non demandé) est $s^2 = \dfrac{\sum_{i=1}^{10} x_i^2}{10} - \bar{x} = \dfrac{69\,837}{10} - 83{,}5^2$ *ce qui donne exactement* $s^2 = 11{,}45$ *!*

Nous nous autorisons à penser que demander un arrondi de s^2 à 10^{-4} près n'était pas très pertinent. Retournons au corrigé…

La valeur affichée sur une calculatrice de s est 3,383 784 86, ce qui donne $s^2 \approx 11{,}450\ 0$.

3.2.1. La variance de la teneur en nitrites dans la population étant notée σ^2, l'hypothèse nulle est : H_0 : "$\sigma^2 = 12$" et l'hypothèse alternative est : H_1 : "$\sigma^2 > 12$".

(en effet l'objectif est de savoir si la variance est supérieure à 12, ou pas).

3.2.2. • La valeur observée de $\dfrac{n\,S^2}{\sigma^2}$ est : $\chi_0^2 = \dfrac{10 \times 11{,}450\ 0}{12}$ c'est-à-dire : $\chi_0^2 \approx 9{,}541\ 67$.

Soit, à 10^{-2} près, $\chi_0^2 = 9{,}54$.

- Au seuil $\alpha = 0{,}05$, la valeur critique est $\chi^2_{1-\alpha\,;\,n-1} = \chi^2_{0{,}95\,;\,9} = 16{,}92$.

3.2.3. $9{,}54 \leq 16{,}92$, donc $\chi_0^2 < \chi^2_{1-\alpha\,;\,n-1}$ et on n'est pas en mesure de rejeter H_0.

Ainsi la variance n'étant pas significativement supérieure à 12, nous pouvons conclure au seuil $\alpha = 0{,}05$ que la reproductibilité est satisfaisante.

Partie 3 : Recherche d'adultérations de viandes

[...]

4. Comparaison de méthodes de détermination du taux d'adultération

4.1. Le laboratoire veut comparer statistiquement trois méthodes de détermination du taux d'adultération des viandes.

- Test sur boîte méthode Mancini.
- Test sur lame méthode Mancini.
- Dosage par Elisa compétition.

Afin de déterminer un éventuel effet « méthode », le laboratoire effectue dans un premier temps 24 prélèvements d'un même pâté permettant la détermination du taux d'adultération par les trois méthodes. On obtient les résultats (en g/100g) consignés dans le tableau 1.

Tableau 1 :

Immuno diffusion Mancini	Elisa compétition	Mancini sur lame
5,64	6,30	3,71
4,99	6,07	5,69
6,07	6,09	5,45
6,03	5,43	4,96
5,90	5,66	5,84
4,00	6,49	4,63
5,18	5,99	3,76
4,37	5,68	5,00

Le technicien a réalisé un test d'analyse de la variance (ANOVA), dont le tableau récapitulatif est donné dans le tableau 2.

Tableau 2 :

	SCE	ddl	CM	f	f_c
Factoriel	4,817 1	2	2,408 5		
Résiduel	9,745 5	21	0,464 1		
Total	14,562 6	23			

Dans ce tableau 2 :

SCE signifie « somme des carrés des écarts »,
ddl signifie « degrés de liberté »,
CM signifie « carrés moyens »,
f est la valeur observée de la variable aléatoire F,
f_c est la valeur critique obtenue pour un seuil de risque de 0,05.

4.1.1. À partir du tableau 2, donner la valeur de f à 10-2 près, en explicitant le calcul.

4.1.2. En vous aidant [*des tables*] du **document 8**, donner un encadrement de f_c d'amplitude 0,1.

4.1.3. Conclure.

Le but de cette analyse de variance est de tester si les taux d'adultération moyens donnés par les trois méthodes sont différents (auquel cas la variance factorielle est supérieure à la variance résiduelle, c'est-à-dire si $CM_{factoriel} > CM_{résiduel}$). Le test utilisé est le test de comparaison de variances de Fisher, unilatéral à droite.

4.1.1. $f = \frac{CM_{factoriel}}{CM_{résiduel}} = \frac{2,4085}{0,4641}$ soit, à 10^{-2} près, $f = 5,19$.

4.1.2 $f_c = f_{0,95;2;21}$ ainsi comme $f_{0,95;2;25} < f_{0,95;2;21} < f_{0,95;2;20}$ on a $3,39 < f_c < 3,49$.

4.1.3. On observe que $f > 3,49 > f_c$ ce qui valide, au seuil $\alpha = 0,05$, l'hypothèse que la variance factorielle est significativement supérieure à la variance résiduelle et donc que les taux d'adultération moyens donnés par les trois méthodes sont différents.

4.2. Avant d'envisager de mettre en place la méthode Mancini sur lame, le laboratoire veut comparer cette méthode à la méthode Elisa compétition pour des pâtés de foie à faible taux d'adultération. Sur dix pâtés différents, il effectue le dosage par les deux méthodes. On obtient les résultats suivants (en g/100g) :

N° pâté	1	2	3	4	5	6	7	8	9	10
Elisa compétition	2,8	2,7	3,2	3,4	2,2	1,0	2,0	3,0	1,4	1,6
Mancini sur lame	0,8	1,0	3,0	2,3	2,0	1,0	0,8	1,2	2,3	1,0

On note :

X_1 : la variable aléatoire prenant pour valeur le taux d'adultération, exprimé en pourcentage, obtenu par la méthode Elisa compétition sur un magret pris au hasard, est supposée de loi normale.

X_2 : la variable aléatoire prenant pour valeur le taux d'adultération, exprimé en pourcentage, obtenu par la méthode Mancini sur lame sur un magret pris au hasard, est supposée de loi normale.

D : la variable aléatoire égale à la différence $D = X_1 - X_2$. On admet que la loi de probabilité de la variable aléatoire D est la loi normale de moyenne μ_D et d'écart type σ_D.

$\overline{D}$: la variable aléatoire qui, à tout échantillon aléatoire simple de 10 différences, associe sa moyenne.

S_D : la variable aléatoire qui, à tout échantillon aléatoire simple de 10 différences, associe son écart type.

On cherche à savoir, au seuil de risque $\alpha = 0{,}05$, si les résultats obtenus par la méthode Elisa sont supérieurs à ceux obtenus par la méthode Mancini.

4.2.1. Expliquer pourquoi le test de comparaison de moyennes dans le cas d'échantillons appariés est approprié.

4.2.2. Formuler l'hypothèse nulle H_0 et l'hypothèse alternative H_1.

4.2.3. On admet que sous H_0 la loi de probabilité de la variable aléatoire $T = \dfrac{\overline{D}}{\dfrac{S_D}{\sqrt{n-1}}}$ est la loi de Student à $n-1$ degrés de liberté.

Calculer, à 10^{-2} près, la valeur observée t_0 définie par $t_0 = \dfrac{\overline{d}}{\dfrac{s_D}{\sqrt{n-1}}}$.

Déterminer la valeur critique $t_{1-\alpha\,;\,n-1}$ et rédiger la conclusion pour ce test statistique. On pourra utiliser le **document 9**.

4.2.1. Le test de comparaison de moyennes dans le cas d'échantillons appariés est approprié car :

- Les variables X_1 et X_2 sont distribuées normalement (ce qui entraîne que $D = X_1 - X_2$ l'est également), condition nécessaire pour l'utilisation du modèle de Student.
- Les échantillons sont appariés, c'est-à-dire que pour tout i prenant une valeur entière de 1 à 10, les valeurs de x_i et de y_i sont couplées rendant les deux échantillons simples de même taille non indépendants.

4.2.2. H_0 : « $\mu_D = 0$ » et H_1 : « $\mu_D > 0$ ».

4.2.3. $t_0 = \dfrac{0{,}79}{\dfrac{0{,}88}{\sqrt{9}}}$ soit $t_0 \approx 2{,}69$.

$t_{1-\alpha\,;\,n-1} = t_{0{,}95;9}$ soit $t_{1-\alpha\,;\,n-1} = 1{,}83$.

$t_0 > t_{1-\alpha\,;\,n-1}$ donc on rejette H_0, ce qui pousse à conclure, au seuil $\alpha = 0{,}05$, que les résultats obtenus avec la méthode Elisa sont supérieurs à ceux obtenus par la méthode Mancini.