TUTORAT UE4 2011-2012 – Biostatistiques
CORRECTION Séance n°3 – Semaine du 17/10/2011

Lois de probabilités continues, estimations – Sabatier
Tests statistiques – Molinari

Séance préparée par Emmanuelle FAROUZ et Léo BARDOT

QCM n°1 : **B**

a) Faux : C'est l'inverse ! Il faut d'abord calculer les paramètres de l'échantillon avant de pouvoir en induire les estimations des paramètres de la population
b) Vrai
c) Faux : Toute la phrase est correcte sauf que le terme correcte est a priori et non a postériori
d) Faux : Attention aux parenthèses ! En effet E(T) doit être égal à θ , sinon le reste est vrai.
e) Faux : **Important :** M. Dujols nous a signifié qu'il fallait bien distinguer la variance de l'échantillon à la variance estimée de la population.

Ainsi, la **variance estimée dans la population** se calcule ainsi : $\hat{S_n}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \mu)^2$

Et la **variance observée de l'échantillon** est donc : $\hat{s_n}^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \mu)^2$

QCM n°2 : **B**

a) Faux : Au risque α l'intervalle à bien **1-α pourcent** de chance de contenir la valeur calculée. Si ce risque ce réparti symétriquement. Il aura donc α /2 % de chaque côté de l'intervalle.
b) Vrai
c) Faux : Justement ! Il dépend de la taille de l'échantillon
d) Faux : Par exemple pour l'intervalle de confiance sur l'espérance d'une v.a.r. suivant 'une loi normale, si n<30 on utilisera une table de Student.
e) Faux : Voir d).

QCM n°3 : **A, C, D**

a) **Vrai**
b) Faux : Calculons d'abord la variance observée dans l'échantillon : $s^2 = |\frac{1}{n}\sum_{i=1}^{n=10}(x_i - \bar{x})^2|$

$s^2 = |\frac{1}{10}\sum_{i=1}^{n=10}(x_i - 6{,}7)^2|$ = 2,61, **à noter** que la formule $\frac{1}{n}\sum_{i=1}^{n=10}x_i^2 - \bar{x}^2$ est bien plus pratique

Ainsi l'écart type $s = \sqrt{2{,}61}$ = 1,62. Or il s'agit de <u>l'écart type observé dans l'échantillon</u> (!).

Afin d'estimer ce paramètre dans la population il suffit de faire : S² = $\frac{n}{n-1}.s^2 = \frac{10}{9}.2{,}61$ = 2,9 afin d'obtenir la variance estimée dans la population. Il ne reste plus qu'à prendre la racine et on trouve $S = \sqrt{2{,}9}$ = **1,70**

c) **Vrai :** En effet étant donné que n<30 il faut supposer que le budget de chaque étudiant suit une loi normale.

d) **Vrai :** Nous cherchons à présent un intervalle de confiance de μ . On a un effectif n<30 donc on va lire le fractile dans la table de Student, au risque α = 5% et au degré de liberté n-1 = 9.

On obtient ainsi $t_{n-1,\alpha/2}$ = 2,262

Il suffit de remplacer : $[\overline{x}-t_{n-1,\alpha/2}\dfrac{S}{\sqrt{n}};\overline{x}+t_{n-1,\alpha/2}\dfrac{S}{\sqrt{n}}]$ avec $\overline{x} = 6{,}7$, $S = 1{,}7$ et n = 10

On obtient alors l'intervalle de confiance **[5,48 ; 7,92]**

e) Faux : ici il faut lire $t_{n-1,\alpha/2}$ avec un d.d.l.= 9 et un risque α = 10%, on lit dans la table : 1,833

On remplace et on trouve **[5,71 ; 7,69]**

QCM n°4 : B, E

a) Faux : Inutile de faire cette approximation : en effet il s'agit d'une proportion (loi binomiale donc !) avec un grand effectif dans l'échantillon (n>30)

b) **Vrai** : Il faut bien noter ici qu'il s'agit d'une proportion, on utilisera donc la formule de l'intervalle de confiance allant avec l'encadrement d'une binomiale.

Il est facile de comprendre que le questionnement de chaque supporter quant à son pronostic constitue une épreuve, le questionnement des 80 sera donc une suite d'épreuves suivant une loi de Bernoulli : d'où l'emploi de la binomiale avec 2 issues possibles et opposées (victoire de Montpellier, défaite de Montpellier).

On rappelle la formule de l'encadrement : $[p_o - c_{\alpha/2}\sqrt{\dfrac{p_o(1-p_o)}{n}}; p_o + c_{\alpha/2}\sqrt{\dfrac{p_o(1-p_o)}{n}}]$

Avec p_o la proportion estimée dans l'échantillon

$c_{\alpha/2}$ = **1,96** à connaitre par cœur pour α =5%

Calculons d'abord p_o = 47/80 = 0,5875

Ainsi cela donne : $[0{,}5875-1{,}96\sqrt{\dfrac{0{,}5875_o(1-0{,}5875)}{80}};0{,}5875+1{,}96\sqrt{\dfrac{0{,}5875_o(1-0{,}5875)}{80}}]$

L'intervalle de confiance de la proportion de supporter prévoyant une victoire de Montpellier est donc bien : **[0,4796 ; 0,6953]**

c) Faux : voir b).

d) Faux : voir e).

e) **Vrai** : Tout pareil que pour un intervalle à 5% sauf que ici, avec α =10% on a : $c_{\alpha/2}$ = 1,645

Ainsi : $[0{,}5875-1{,}645\sqrt{\dfrac{0{,}5875.(1-0{,}5875)}{80}};0{,}5875+1{,}645\sqrt{\dfrac{0{,}5875.(1-0{,}5875)}{80}}]$

Nous donne bien l'intervalle de confiance **[0,4969 ; 0,6780]**

QCM n°5 : A, C, E

a) **Vrai** : Il suffit de faire $\dfrac{\sum_{i=1}^{15} x_i}{15}$ = 375/15 = 25ans.

b) Faux : Afin de calculer la variance lorsque l'on possède ce type d'information il vaut mieux utiliser la deuxième formule précédemment énoncée : s²= $\frac{\sum_{i=1}^{15} x_i^2}{15} - \bar{x}^2 = \frac{10635}{15} - 25^2 = 84$ **ans²**

Attentions aux unités !

c) **Vrai :** Estimons la variance dans la population : S² = $\frac{n}{n-1}.s^2$ = $\frac{15}{14}.84$ = 90

Donc l'écart-type estimé de la population est égal à S= $\sqrt{90}$ = **9,49 ans**

d) Faux : voir e)

e) **Vrai** : La formule correspondant à l'intervalle de confiance estimée de la variance dans la population est : $[(n-1)\frac{S^2}{b};(n-1)\frac{S^2}{a}]$ avec : S² = variance estimée de la population

a tel que $P(\chi_{14,10\%}{}^2 \leq a) = 0{,}10$

b tel que $P(\chi_{14,10\%}{}^2 \leq b) = 0{,}90$

On trouve ainsi a = 7,790 et b = 21,064

L'intervalle de confiance devient donc $[(15-1)\frac{90}{21{,}064};(15-1)\frac{90}{7{,}790}]$ soit **[59,82 ; 161,75].**

Ainsi l'intervalle de confiance de l'écart-type cette fois, est $[\sqrt{59{,}82};\sqrt{161{,}75}]$

C'est-à-dire **[7,73 ; 12,72].**

QCM n°6 : A, C, D

a) Vrai

b) Faux :

m=∑xi / n = 700 / 100 = 7

S²=(1/(n-1))∑(xi-m)²=(1/(n-1))(∑xi² - nm²)

=1/99 (5790 – 100×7²)=8,99

Donc S=2,99

Rmq: ici n est grand donc S²=s²

c) **Vrai** : cf correction au dessus

d) Vrai

e) Faux

QCM n°7 : D

P=P0 +/- c $\sqrt{\frac{P0(1-P0)}{n}}$

Précision=c $\sqrt{\frac{P0(1-P0)}{n}}$

0,02=1,96 $\sqrt{\frac{0{,}5\times 0{,}5}{n}}$

$\sqrt{n}$=49

D'où n=2401

Pourquoi prend-on un pq à 0,25 ?

➔ On veut avoir le nombre de sujets nécessaires pour déterminer la précision demandée. Or on ne sait pas de combien p est. On prend le pire des cas = celui où on a un pq le plus élevé possible, comme ça on aura un n qui est un NSN même dans le pire des cas. Si au final on avait un p à 0,4 ce n'est pas grave (il fallait un moindre NSN mais on en a pris plus et c'est tant mieux !). Par calcul, on peut faire p=0,1 et q=0,9 et on retrouve pq=0,09, p=0,2/0,3/0,4. Le p pour lequel pq est le plus grand est p=0,5 (pq=0,25).

QCM n°8 : **D, E**

a) Faux : Les deux hypothèses sont définies à priori, c'est à dire, au tout début du test.
b) Faux : Elles peuvent aussi porter sur la nature d'une ou plusieurs distributions.
c) Faux : C'est le risque de deuxième espèce. Tout le reste est vrai.
d) Vrai
e) Vrai

QCM n°9 : **A, B, C, D**

a) Vrai
b) Vrai
c) Vrai
d) Vrai
e) Faux : Elle est indépendante des risques de première et de deuxième espèces.

QCM n°10 : **A, C, D, E**

a) Vrai
Remarque : pour information le risque α est généralement de 5%, mais étant défini a priori, il peut prendre d'autres valeurs.
b) Faux : a priori ! C'est la p-value que l'on trouve a posteriori.
c) Vrai.
d) Vrai.
e) Vrai

QCM n°11 : **B, C, D, E**

a) Faux
On fera un test bilatéral (par exemple Student) pour montrer une différence, alors que l'on fera un test unilatéral (par exemple Fischer) pour montrer une supériorité/infériorité entre 2 groupes ce que l'on recherche ici.
b) Vrai
c) Vrai
En effet, on a H_0 : l'effectif des patients est le même que l'effectif des patientes, autrement dit que le pourcentage de patients (ou de patientes) est de 50%.
Et H_1 : l'effectif des patientes est supérieur à l'effectif des patients (cf a)
d) **Vrai**
e) Vrai (mais item annulé ^^)

QCM n°12 : **A, E**

a) Vrai :
On a $z=\frac{m-m1}{\sqrt{\frac{s^2+s1^2}{n}}} = \frac{(4{,}75-4{,}2)}{\sqrt{\frac{(1+3)}{100}}}$ = 2,75
Or, 2,75 > 1,96 donc on peut rejeter H_0.
b) Faux : cf item a
c) Faux.
d) Faux : il y a eu TAS de 2 échantillons.
e) Vrai.

QCM n°13 : **A, E**

a) Vrai : Ici z=0 car $m_A=m_B$
b) Faux
c) Faux
d) Faux

e) Vrai : Aucun calcul nécessaire ici, car $z=\frac{mA-mB}{\sqrt{\frac{sA^2+sB^2}{n}}}$ or $m_A=m_B$ d'où $m_A-m_B=0$

QCM n°14 : **E**

a) Faux

b) Faux

On a p_0 qui est toujours estimé et p qui est le paramètre vrai ou théorique. Or, on pose les hypothèses toujours sur les paramètres VRAIS !

c) Faux

d) Faux

e) Vrai

On a donc H_0 : p=1/6 et H_1 : p>1/6 (test unilatéral)