Exercices de biostatistique

Rappel: pour visualiser la formule associée aux résultats obtenus, il vous suffit d'aller cliquer sur la case concernée(uniquement dans excel et non avec "Adobe Acrobat") !!

Intervalles de confiance

Exercice 1

Calculez l'intervalle de confiance de la prévalence d'une maladie pour laquelle 200 cas ont été diagnostiqués dans 1000 individus pris au hasard ? Répétez si on a 1 cas sur 5.

Grand echantillon

p estimé = 200/1000 = 0.2

P{pe-1,96*sqrt(pe*qe/n) < p < pe+1,96*sqrt(pe*qe/n) } = 0.95

pi	0,17520774
ps	0,22479226

Petit echantillon

Chercher p tel que la probabilité d'avoir 1 cas ou moins = 0.025 => ps

B(0)+B(1) = 0.025	=>	(1-p)^4*(1+4p) = 0.025	
	=>	ps	0,71642019
		expr	0,02499926

Chercher p tel que la probabilité d'avoir 1 cas ou plus = 0,025 => pi

B(0) = 0.975	=>	(1-p)^5 = 0.975	
	=>	(1-p) = 0,975^0.2	
	=>	p = 1 - (0.975^0.2) =	0,00505076

Exercice 2

Calculez l'intervalle de confiance de la moyenne d'une distribution de Poisson si k observé est 1. Comment procéderiez vous si k=2 ?

Chercher µ tel que P(k=0) + P(k=1) = 0.025 => µs

exp(-µs)*(1+µs)=0.025	=>	µs	5,57168473
	=>	expr	0,02499912

Chercher µ tel que P(k=1) + P(k=2) + … = 0.025 <=> P(k=0) = 0.975

exp(-µl) = 0,975	=>	µl	0,02531781

Pour k=2, meme raisonnement:

P(k=0 | µs) + P(k=1 | µs) + P(k=2 | µs) = 0.025

P(k=0 | µl) + P(k=1 | µl) = 0.975

Exercice 3

On sait que, chez la vache, au moment du part, la calcémie tombe à la valeur moyenne de 7.6 mgr % (déviation standard: 1.4), ce qui donne parfois lieu à des accidents hypocalcémiques (fièvre de lait). Dans le but de prévenir ces accidents, on a préconisé l'emploi d'un produit, le tachystérol, qui mobilise le calcium osseux; celui-ci, administré à 100 vaches, une semaine avant le vêlage, a fait remonter la calcémie (moyenne) à 8.3 mgr %. Ce produit est-il vraiment efficace pour relever la calcémie ? Dans quelles limites pouvait-on attendre la calcémie si le traitement n'est pas efficace (c'est à dire, sous l'hypothèse nulle) ?

On peut reformuler le problème en se demandant: quelle est la probabilité d'obtenir une moyenne sur 100 vaches de 8.3 mgr % si ces vaches proviennent d'une population de moyenne 7.6 mgr % et de déviation standard 1.4 mgr % ?
Le raisonnement est bien entendu que si cette probabilité est anormalement petite, on aura tendance à considérer que l'échantillon ne provient pas de cette population (mais d'une population dans laquelle la moyenne est > 7.6 mgr %), ce qui signifie que le tachystérol a un effet.

On peut employer un test classique de z:

$$z = (X_m - \mu_m) / \sigma_m = (X_m - \mu) / (\sigma / \text{racine}(100))$$

où l'indice m indique que les paramètres correspondants sont les paramètres de la distribution des moyennes.
On obtient:

$$z = (8.3 - 7.6) / (1.4 / 10) = 5$$

ce qui correspond à un probabilité de : $P(z > 5) =$ 2,871E-07

Clairement, l'amélioration est très significative, et on rejette l'hypothèse nulle.

Si le traitement n'est pas efficace, $\mu = 7.6$, et on s'attend à trouver 95 % des moyennes calculées sur des échantillons de taille 100 dans les limites:

$$P(\mu - 1.96 * \sigma / 10 < Xm < \mu + 1.96 * \sigma / 10) = 95\ \%$$

soit,

$$P(7{,}3256 < Xm < 7{,}8744) = 95\%$$

La valeur observée (8.3) tombe nettement en dehors de cet intervalle (de confiance):
l'hypothèse nulle, qui a conduit à cet intervalle, est donc rejetée.

Exercice 4

On a mesuré l'épaisseur du lard dorsal chez 10 porcs.
On demande de calculer les limites de l'intervalle de confiance de μ, la vraie valeur de la moyenne de la population.
Peut-on conclure que l'épaisseur du lard dorsal dans cette lignée est différente de celle rencontée dans la population utilisée plus haut ?

L'intervalle de confiance de la vraie moyenne μ se calcule en partant de :

$$t = (Xm - \mu) / (s / \text{racine}(n))$$ où t a (n-1) degrés de liberté

Par conséquent,

$$\mu = Xm + t * s / \text{racine}(n)$$

On ne connaît pas la valeur de t qui résoud cette équation, mais on peut prendre une valeur (négative) de t telle que μ n'a alors que 2.5 % de chance d'être inférieure à:

$$\mu = Xm + t1 * s / \text{racine}(n)$$

et on peut prendre une valeur positive de t telle que μ n'a alors que 2.5 % de chances d'être supérieur à:

$\mu = Xm + t2 * s / racine(n)$

Il est clair, par symétrie de la distribution de t, que t1 = - t2, et par conséquent:

$P (Xm - t2 * s / racine(n) < \mu < Xm + t2 * s / racine(n)) = 95\%$

La valeur correspondante de t2 se trouve dans la table avec (n-1) degrés de liberté, qui fournit:

t2 = **2,26215889**

Par ailleurs, n vaut 10 et

s = **5,16935414**
Xm = **38,5**

L'intervalle de confiance correspondant est donc:

$P (34,802 < \mu < 42,198) = 95\%$

Si on considère l'échantillon constitué par les porcs XI à XX, on obtient une moyenne de:

Xm = **39,3**

Cette moyenne tombe dans l'intervalle de confiance: il n'y a pas de raison d'après ces résultats de douter que cet échantillon provient de la même population que celle dont est issu le premier échantillon.

Exercice 5

Si, pour une dilution donnée, on a obtenu expérimentalement une concentration en microorganismes de 1 par unité de volume, quelles sont les limites entre lesquelles on peut attendre la concentration à cette dilution dans la population (au seuil 95%) ?

La concentration suit une distribution de Poisson, et la moyenne expérimentale (calculée sur les données) est $\mu = 1$.
Il est bien entendu possible que la vraie moyenne soit en réalité plus grande ou plus petite que cette moyenne expérimentale, et que l'on ait obtenu $\mu = 1$ que par fluctuation statistique. On peut chercher les valeurs de moyennes $\mu 1$ et $\mu 2$ qui sont telles que la valeur observée (ou une valeur plus extrême de μ) n'a qu'une probabilité de 2.5 % de se produire. Au vu des données, l'intervalle [$\mu 1$; $\mu 2$] est donc constitué de valeurs de μ pour lesquelles la valeur observée a plus de 5% de chances de se produire: il s'agit de l'intervalle de confiance de μ.

1) Pour calculer $\mu 1$ (limite inférieure), on veut qu'obtenir 1 ou plus que 1 n'ait que 2.5% de chance de se produire, ou, de manière équivalente, que 0 µorganisme ait 97.5 % de chances d'arriver.

Employant la distribution de Poisson, on peut donc écrire:

$P(0) = 0.975 = exp(-\mu 1) * (\mu 1\text{^}0) / (0!) = exp(-\mu 1)$
=> $\mu 1 = -ln(0.975) =$ **0,02531781**

2) Pour calculer $\mu 2$ (limite supérieure), on veut qu'obtenir 1 ou moins que 1 n'ait que 2.5% de chance de se produire.

=> $P(0) + P(1) = 0.025 = exp(-\mu 2) * [(\mu 2\text{^}0) / 0! + (\mu 2\text{^}1) / 1!]$
=> $0.025 = exp(-\mu 2) * (1 + \mu 2)$

La résolution de cette équation fournit:

$\mu 2 =$ **5,57165919**

(on peut vérifier ce résultat en introduisant la valeur trouvée dans l'équation précédente).

L'intervalle cherché:

P (0,02532 < μ < 5,57166) = 95%

Exercice 6

On a mesuré (en mm) l'épaisseur du lard dorsal chez 10 porcs et on a obtenu les résultats suivants: 47, 38 , 39 , 32 , 34 , 37 , 31 , 43 , 41 , 43. On demande de calculer les limites de l'intervalle de confiance de la moyenne de population et sa vraie valeur.

	Données	
	47	2209
	38	1444
	39	1521
	32	1024
	34	1156
	37	1369
	31	961
	43	1849
	41	1681
	43	1849
Somme	385	15063
Somme²	148225	
Moyenne	38,5	
n	10	
S²	26,7222222	
S	5,16935414	

La formule à utiliser pour trouver les limites de l'intervalle de confiance est la suivante:

$$\mu = \bar{x} \pm t_{0.05}^{n-1} \frac{S}{\sqrt{n}}$$

où $\frac{S}{\sqrt{n}}$ est l'erreur standard de l'échantillon et est égale à : 1,63469331

Il s'agit d'un test bilatéral, il faut donc chercher la valeur de t correspondante dans la table
t 0,05, c'est à dire : 2,26215889

Limite inf.moyenne popul.:	34,802064
Limite sup.moyenne popul.:	42,197936

Exercice 7

D'une distribution normale dont on connaît la variance(=25), on prélève 64 observations au hasard et la moyenne de ces observations =11,1. Testez l'hypothèse que la vraie moyenne est égale à 10.

Ici la variance étant connue, il faut utiliser la statistique z et non t !

n:	64
sigma²:	25
sigma:	5
Moy.échant.:	11,1

Pour un intervalle de confiance de 95 %, on aura un z correspondant de: 1,95996108

Limite sup.:	12,3249757
Limite inf.:	9,87502432

La vraie moyenne pourrait être de 10 vu que cette valeur est bien comprise entre les limites de notre intervalle de confiance!

Exercice 8

Dans une région, on a pesé 68 veaux de 2 semaines et de ces pesées, on a extrait les valeurs suivantes: Moyenne= 73,33 et déviation standard= 6,39.
Montrez que si tous les veaux de cet âge avaient été pesés, la moyenne de ces mesures aurait 95 chances sur 100 de tomber dans les limites 71,77 et 74,89?

n:	68
Moy.échant.:	73,33
S:	6,39

La valeur de t correspondante pour un intervalle de confiance de 95 %: 1,9960089

Limite sup.:	74,8767099
Limite inf.:	71,7832901

Exercice 9

La distibution des numérations globulaires de porcelets âgés de 2 semaines peut être considérée comme normale avec une moyenne de population égale à 7000 et une déviation standard de population de 1200 (en 1000/mm³). Quelle est la probabilité d'une numération supérieure à 10600 ? 21 porcelets ont reçu un régime spécial, enrichi en fer. La numération globulaire moyenne a été de 9400 et la variance de 1960000, soit S=1400. Le régime a-t-il modifié la numération globulaire moyenne ? A-t-il modifié la variation des numérations ?

a) Moy.pop.:	7000
Dév.st.pop.:	1200
X	10600
Z	3
P(>10600)	0,00134997

b) n:	21
Moy.échant.:	9400
S²:	1960000
S:	1400
I.C.:	95%

H0: le régime n'a pas d'effet et donc la moyenne de population est égale à la moyenne de l'échantillon, les différences observées sont dues à des fluctuations statistiques.

Limite inf.:	8886,75152

Limite sup.: 9913,24848

L'hypothèse nulle est donc rejetée puisque 7000 n'est pas compris dans cet intervalle de confiance et donc le régime a bien un effet sur la numération globulaire.

Une autre façon de procéder aurait également pu être de faire comme suit:

Z= 9,16515139 Ce qui est beaucoup plus grand que 1,96 (valeur seuil pour 95%) et donc l'hypothèse nulle est rejetée!

c) H0: La variance de population est égale à la variance de l'échantillon.

$\frac{\Sigma x^2}{\sigma^2}$ est distribué comme un khi² avec 20 ddl

$\frac{S^2(n-1)}{\sigma^2}$ 27,2222222 Ce qui est inférieur à la valeur théorique pour un khi² ayant un seuil de 5% et 20 ddl(31,41) et donc H0 est acceptée.

Vérification: 31,4104204