24/09/2018

Numa BENAMER et Marie DÉNÈS

Estimation paramétrique

Travaux pratiques n°3 analyse de données

Question 1 1

Question 2 2

Question 3 3

Question 4 4

%\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*  
%\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\* TP 3 \*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*  
%\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

## Question 1

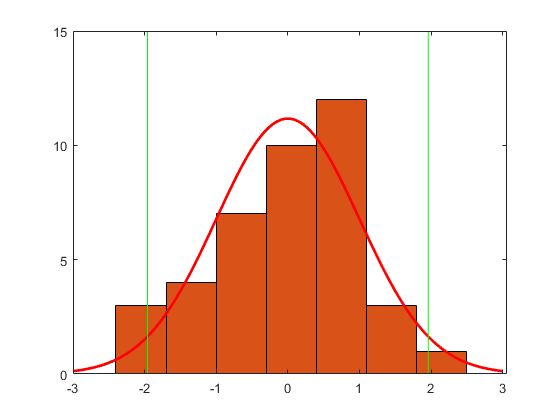
data1 = [0.82 0.87 0.77 0.96 0.75 0.83 0.87 0.81];  
%Nous avons ici un petit échantillon gaussien.  
  
mean\_data1 = mean(data1);  
std\_data1 = std(data1);  
  
a\_t = tinv((1+0.95)/2, length(data1) - 1);  
  
%Si l'échantillion data1 est généré par une loi gaussienne on peut estimer  
%qu'on a 95% de chance qu'une données produite le soit dans l'intervalle  
%suivant :  
interval1 = [mean\_data1 - a\_t .\* std\_data1/sqrt(length(data1)-1), mean\_data1 + a\_t .\* std\_data1/sqrt(length(data1)-1)]

interval1 =  
  
 0.7761 0.8939

## Question 2

data2 = [0.84 0.87 0.89 0.73 0.84 0.81 0.88 0.85 0.89 0.79 0.79 0.90 0.59 0.75 0.67 0.76 0.86 0.88 0.70 0.75 0.81 0.77 0.83 0.84 0.71 0.78 0.59 0.91 0.74 0.68 0.77 0.66 0.80 0.74 1.02 0.91 0.55 0.84 0.66 0.77];  
mean\_data2 = mean(data2);  
var\_data2 = var(data2);  
std\_data2 = std(data2);  
a\_t2 = norminv((1 + 0.95)/2, 0, 1);   
%ici on utilise norminv car nous avons un grand échantillon  
  
%Si l'echatillon est généré par une loi gaussienne on peut estimer qu'on a  
%95% de chance qu'une valeur produite soit dans l'interval suivant :  
  
interval2 = [mean\_data2 - a\_t2 .\* std\_data2/sqrt(length(data2)), mean\_data2 + a\_t2 .\* std\_data2/sqrt(length(data2))]  
  
figure(1)  
data\_centre\_reduit = (data2 - mean\_data2)./ std\_data2;  
histfit(data\_centre\_reduit)  
  
hold on  
plot([-a\_t2 a\_t2; -a\_t2 a\_t2], [0 0; 15 15], 'g')

interval2 =  
  
 0.7548 0.8162



## Question 3

n = 1000;  
sqrt2erfinv95 = norminv((1 + 0.95)/2, 0, 1);  
sqrt2erfinv99 = norminv((1 + 0.99)/2, 0, 1);  
  
%Dupond :  
p = 500/1000;  
%La pobabilité qu'un électeur vote pour dupond est de 0.5. Considerant un  
%tirage sur 1000 individus, nous pouvons affirmer avec un niveau de  
%certitude de 95% que le nombre de candidat qui voteront pour dupont est  
%compris dans l'intervalle suivant :  
  
%On multiplie par n pour ramener les probabilités en valeurs absolue ((plus  
%lisible).  
Vote\_pour\_dupond\_a\_95 = [p - sqrt2erfinv95 \* (p \* (1 - p) / n)^0.5, p + sqrt2erfinv95 \* (p \* (1 - p) / n)^0.5].\* n  
  
%Et à 99% :  
Vote\_pour\_dupond\_a\_99 = [p - sqrt2erfinv99 \* (p \* (1 - p) / n)^0.5, p + sqrt2erfinv99 \* (p \* (1 - p) / n)^0.5] .\* n  
  
%On constate que l'intervalle à 99% est plus étendu que celui à 95%. C'est  
%parfaitement logique : pour diminuer les chance de se trouver hors de  
%notre intervalle il convient de l'étendre.  
  
%De meme nous avons pour les autres candidats :  
  
%Durand :  
p = 250/1000;  
Vote\_pour\_Durand\_a\_95 = [p - sqrt2erfinv95 \* (p \* (1 - p) / n)^0.5, p + sqrt2erfinv95 \* (p \* (1 - p) / n)^0.5].\* n  
Vote\_pour\_Durand\_a\_99 = [p - sqrt2erfinv99 \* (p \* (1 - p) / n)^0.5, p + sqrt2erfinv99 \* (p \* (1 - p) / n)^0.5] .\* n  
  
%Duroc :  
p = 50/1000;  
Vote\_pour\_Duroc\_a\_95 = [p - sqrt2erfinv95 \* (p \* (1 - p) / n)^0.5, p + sqrt2erfinv95 \* (p \* (1 - p) / n)^0.5].\* n  
Vote\_pour\_Duroc\_a\_99 = [p - sqrt2erfinv99 \* (p \* (1 - p) / n)^0.5, p + sqrt2erfinv99 \* (p \* (1 - p) / n)^0.5] .\* n

|  |  |
| --- | --- |
| Vote\_pour\_dupond\_a\_95 =   469.0102 530.9898 | Vote\_pour\_dupond\_a\_99 =   459.2726 540.7274 |
| Vote\_pour\_Durand\_a\_95 =   223.1621 276.8379 | Vote\_pour\_Durand\_a\_99 =   214.7290 285.2710 |
| Vote\_pour\_Duroc\_a\_95 =   36.4919 63.5081 | Vote\_pour\_Duroc\_a\_99 =   32.2473 67.7527 |

## Question 4

% Pour Duval avec 17%  
  
p = 0.17;  
  
% Intervalle de confiance à 95%  
  
% Je garde la valeur de sqrt2erfinv(1-95) = 1.96  
% Avec une précision de 0.01 nous avons d'après le td n = 5420  
  
prec = 0.01;  
a\_t\_95 = tinv((1 + 0.95) ./ 2, 1000);  
  
  
%Nous avons donc un nombre minimum d'individu n de :  
  
n = (a\_t\_95 ./ prec .\* (p \* (1 - p))^0.5 )^2  
  
%Ce nombre varie légèrement de celui obtenue lors du td en classe. Cette  
%difference est due au calcul du erf inverse (1.9623 au lieu de 1.96 dans  
%la table).

n =  
  
 5433.4

[*Published with MATLAB® R2018a*](https://www.mathworks.com/products/matlab)