24/09/2018

Estimation paramétrique

Travaux pratiques n°3 analyse de données

Numa BENAMER et Marie DÉNÈS

```
%*************************************************************************
%******************************* TP 3 ************************************
%*************************************************************************
```

Question 1

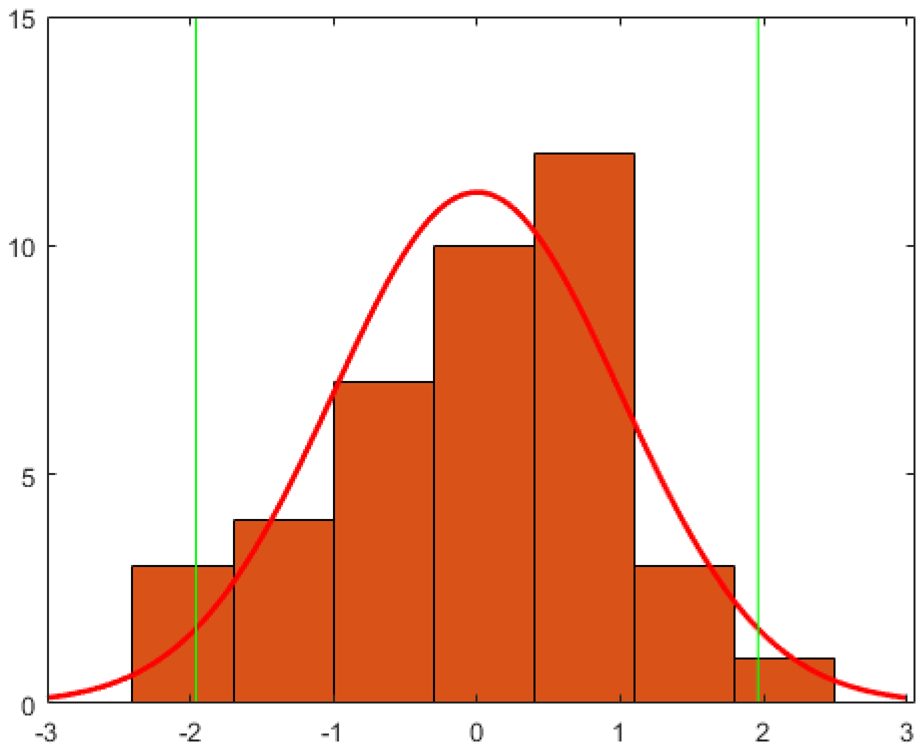
```
data1 = [0.82 0.87 0.77 0.96 0.75 0.83 0.87 0.81];
%Nous avons ici un petit échantillon gaussien.

mean_data1 = mean(data1);
std_data1 = std(data1);

a_t = tinv((1+0.95)/2, length(data1) - 1);

%Si l'échantillon 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