Eléments de statistiques

-

Rapport de la partie 2 du projet

Pazienza Laurie

-

s123514

3ème année de Bachelier Ingénieur Civil

Année académique 2014-2015

Questions

Les codes Matlab relatifs aux différentes questions se trouvent dans l'annexe.

1 Estimation

(a)

Le biais de l'estimateur m_x de la note finale moyenne de la population s'élève à 0.0803, la variance de ce même estimateur vaut 0.2862.

(b)

Le biais de l'estimateur $median_x$ de la note finale moyenne de la population s'élève à 0.1944, la variance de ce même estimateur vaut 0.2836.

(c)

Pour un échantillon de 50 étudiants, nous obtenons les résultats suivants :
— biais de $m_x = 0.0956$; variance de $m_x = 0.3026$
— biais de $median_x = 0.1111$; variance de $median_x = 0.3252$

Nous remarquons donc que la variance des deux différents estimateurs diminue lorsque la taille de l'échantillon augmente. Ceci semble logique car plus l'échantillon est de grande taille, plus il tend vers la population et donc offre des résultats proches de la réalité.

Le biais de l'estimateur de la médiane augmente lorsque la taille de l'échantillon croit alors que le biais de l'estimateur de la moyenne adopte le comportement inverse et se rapproche donc le la moyenne réelle. Nous pouvons donc en déduire que l'estimateur de la moyenne est meilleur que celui de la médiane.

(d)

i. Pour construire un intervalle de confiance selon la loi de student, nous utilisons la formule : $[m_x - t_{1-\frac{\alpha}{2}} \frac{s_{n-1}}{\sqrt{n}}, m_x + t_{1-\frac{\alpha}{2}} \frac{s_{n-1}}{\sqrt{n}}]$.

La moyenne de chaque échantillon est représentée par m_x, $t_{1-\frac{\alpha}{2}}$ vaut 2.093, il est calculé par la table de la loi de student pour $\alpha = 0.05$ et n-1 = 19 ddl. s_{n-1} est calculé par la formule suivante : $\sqrt{\frac{1}{n-1}\Sigma_{i=1}^{n}(x_i - m_x)^2}$

Les bornes moyennes de l'intervalle de confiance selon la loi de student sont : [11.9313, 14.1527].

En moyenne, il y a 94 intervalles de confiance qui contiennent la moyenne finale de la population.

ii. Selon la loi de Gauss, l'intervalle est construit par la formule : $[m_x - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}, m_x + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}]$. On calcule $u_{1-\frac{\alpha}{2}} = 1.96$ par la table de Gauss pour $\alpha = 0.05$, σ est l'écart-type de la population et n vaut, à nouveau, 20. La population étant considérée inconnue, nous approximerons σ par s_n qui sera calculé par la formule suivante : $s_n = \sqrt{\frac{n-1}{n}} s_{n-1}$

Les bornes moyennes de l'intervalle de confiance selon la loi de Gauss sont : [11.9476, 14.1364].

Il y a, en moyenne, 97 intervalles de confiance contenant la valeur de la population.

Dans les deux cas, les résultats sont proches des 95% fixés. De plus, la moyenne du nombre d'intervalles de confiance contenant la valeur de la population tend vers 95 lorsque le nombre d'échantillons testés croit. Nous en déduisons qu'il était correct de supposer que la variable parente utilisée était une variable Gaussienne.

2 Tests d'hypothèse

(a)

Vérifier l'hypothèse H_0 se résume à tester si $p \leq p_0 + u_{1-\alpha}\sqrt{\frac{p_0(1-p_0)}{n}}$. La proportion d'étudiants ayant obtenu une note inférieure ou égale à 10/20 est représentée par p, p_0 vaut 1/8 et représente la proportion fixée, $u_{1-\alpha}$ vaut 1.645 et s'obtient par la table de Gauss pour 95% car nous effectuons un test unilatéral, et enfin n est toujours égal à 20.

Après avoir fait tourner la fonction *Q2.m*, fournie dans l'annexe, plusieurs fois, nous en déduisons que l'Ulg rejette en moyenne 8 fois l'hypothèse H_0, ceci est acceptable mais ne témoigne pas de l'optimatilité du test au vu du α fixé à 5%. L'erreur vient en partie du fait que nous avons approximé une loi binomiale par une loi normale.

(b)

Dans 42 cas, en moyenne, un article sera publié dans la gazette locale. Puisque nous nous intéressons maintenant aux 6 autres organismes, il est logique que ce résultat soit plus élevé qu'α. De plus, il semble également normal que ce résultat ne soit pas exactement égal à 6 fois le résultat obtenu au point précédent. En effet, un article est publié si au moins un des organismes rejette l'hypothèse. Il n'y aura donc qu'un seul article si les 6 organismes rejettent tous cette hypothèse.

L'événement X : "au moins un organisme rejette l'hypothèse et un article est publié" est le complémentaire de l'événement A : "aucun organisme ne rejette l'hypothèse et aucun article ne sera publié", nous pouvons donc calculer la probabilité de X par la formule suivante :

$$P(A) = (1-\alpha)^6 = 0.95^6 = 0.7351 \tag{1}$$

$$P(X) = 1 - P(A) = 0.2649 \tag{2}$$

Cette probabilité ne correspond pas aux nombres d'articles publiés obtenus en pratique, nous pouvons donc en conclure que le test n'est pas optimal dans ce cas. En effet, nous supposons que la proportion d'étudiants ayant obtenu une note inférieure ou égale à 10/20 suit une loi normale, cependant cette hypothèse n'est correcte que si $min(np_0, n(1-p_0)) \geq 5$ cette condition n'est pas vérifiée car 2.5 n'est pas supérieur ou égal à 5. La proportion étudiée ne suit donc pas une loi normale, ceci explique pourquoi ce test n'offre pas les résultats attendus.

(c)

L'avantage des instituts de sondage vient du fait qu'ils possèdent 6 échantillons différents, leur donner le même à tous réduirait cet avantage qu'ils ont sur les autorités de l'Ulg. Cependant il semblerait intéressant que ces organismes travaillent ensemble, afin d'alors avoir un échantillon plus grand et qui fournirait donc des résultats plus proches de ceux de la population entière.

Annexe

Q1A.m

```
1
2  function [matrice_echantillon_notefinale moyenne_notefinale moyenne_totale ...
       note_finale]=Q1A
3
4  resultats=xlsread('Proba1ereSession20132014.xls'); %rÃ©cupÃ©ration des ...
       donnÃ©es Excel
5  matrice_echantillon=zeros(20,100); %Matrice 20 lignes/100 colonnes pour ...
       stocker les Ã©chantillons
6
7  for i=1:100
8
9      echantillon = randsample(120,20,true); %CrÃ©ation d'un Ã©chantillon de 20 ...
           Ã©tudiants
10     matrice_echantillon(:,i)=echantillon; %Remplissage de la matrice ...
           contenant les 100 Ã©chantillons, une colonne = un Ã©chantillon
11
12 end
13
14
15
16 %Calcul des notes finales de la population :
17
18 note_finale=zeros(120,1);
19
20 for j=1:120
21
22     note_finale(j)=mean(resultats(j,:));
23
24 end
25
26 %Calcul des notes finales de chaque Ã©tudiant de chaque Ã©chantillon :
27
28 matrice_echantillon_notefinale=zeros(20,100);%Matrice qui contiendra les ...
       notes finales de chaque Ã©tudiant de chaque Ã©chantillon
29 moyenne_notefinale=zeros(100,1); %Vecteur de la nouvelle variable
30
31 for k=1:100
32
33     for l=1:20
34
35         matrice_echantillon_notefinale(l,k)=note_finale(matrice_echantillon(l,k));
36
37     end
38
39     moyenne_notefinale(k)=mean(matrice_echantillon_notefinale(:,k));
40
41 end
42
43 %Calcul du biais :
44
45 moyenne_totale = mean(note_finale);
46 biais = mean(moyenne_notefinale) - moyenne_totale
47
48 %Calcul de la variance :
49
50 variance = var(moyenne_notefinale,1)
51
52
53 end
```

Q1B.m

```
1
2  function Q1B
3
4  %Cette fonction est la mÃªme que la fonction Q1A mais pour le calcul de la
5  %mÃ©diane
6
7  resultats=xlsread('Proba1ereSession20132014.xls'); %rÃ©cupÃ©ration des ...
       donnÃ©es Excel
8  matrice_echantillon=zeros(20,100); %Matrice 20 lignes/100 colonnes pour ...
       stocker les Ã©chantillons
9
10 for i=1:100
11
12     echantillon = randsample(120,20,true); %CrÃ©ation d'un Ã©chantillon de 20 ...
           Ã©tudiants
13     matrice_echantillon(:,i)=echantillon; %Remplissage de la matrice ...
           contenant les 100 Ã©chantillons, une colonne = un Ã©chantillon
14
15 end
16
17
18 %Calcul des notes finales de la population:
19
20 note_finale=zeros(120,1);
21
22 for j=1:120
23
24     note_finale(j)=mean(resultats(j,:));
25
26 end
27
28 %Calcul des notes finales de chaque Ã©tudiant de chaque Ã©chantillon :
29
30 matrice_echantillon_notefinale=zeros(20,100);%Matrice qui contiendra les ...
       notes finales de chaque Ã©tudiant de chaque Ã©chantillon
31 mediane_notefinale=zeros(100,1); %Vecteur de la nouvelle variable
32
33 for k=1:100
34
35     for l=1:20
36
37         matrice_echantillon_notefinale(l,k)=note_finale(matrice_echantillon(l,k));
38
39     end
40
41     mediane_notefinale(k)=median(matrice_echantillon_notefinale(:,k));
42
43 end
44
45 %Calcul du biais :
46
47 mediane_totale = median(note_finale);
48 biais = median(mediane_notefinale) - mediane_totale
49
50 %Calcul de la variance :
51
52 variance = var(mediane_notefinale,1)
53
54
55 end
```

Q1C.m

```
1
2  function Q1C
3
4  %MÃªme fonction que les deux prÃ©cÃ©dentes mais cette fois pour un Ã©chantillon
5  %de taille 50
6
7  resultats=xlsread('Probalere Session20132014.xls'); %rÃ©cupÃ©ration des ...
       donnÃ©es Excel
8  matrice_echantillon=zeros(20,50); %Matrice 20 lignes/50 colonnes pour stocker ...
       les Ã©chantillons
9
10 for i=1:50
11
12     echantillon = randsample(120,20,true); %CrÃ©ation d'un Ã©chantillon de 20 ...
           Ã©tudiants
13     matrice_echantillon(:,i)=echantillon; %Remplissage de la matrice ...
           contenant les 50 Ã©chantillons, une colonne = un Ã©chantillon
14
15 end
16
17 %Calcul des notes finales :
18
19 note_finale=zeros(120,1);
20
21 for j=1:120
22
23     note_finale(j)=mean(resultats(j,:));
24
25 end
26
27 %Calcul des notes finales de chaque Ã©tudiant de chaque Ã©chantillon :
28
29 matrice_echantillon_notefinale=zeros(20,50);%Matrice qui contiendra les notes ...
       finales de chaque Ã©tudiant de chaque Ã©chantillon
30 moyenne_notefinale=zeros(50,1);
31 mediane_notefinale=zeros(50,1);
32
33
34 for k=1:50
35
36     for l=1:20
37
38         matrice_echantillon_notefinale(l,k)=note_finale(matrice_echantillon(l,k));
39
40     end
41
42     moyenne_notefinale(k)=mean(matrice_echantillon_notefinale(:,k));
43     mediane_notefinale(k)=median(matrice_echantillon_notefinale(:,k));
44
45 end
46
47 %Calcul du biais :
48
49 moyenne_totale = mean(note_finale);
50 biaisA = mean(moyenne_notefinale) - moyenne_totale
51 mediane_totale = median(note_finale);
52 biaisB = median(mediane_notefinale) - mediane_totale
53
54 %Calcul de la variance :
55
56 varianceA = var(moyenne_notefinale,1)
57 varianceB = var(mediane_notefinale,1)
```

```
58
59
60 end
```

Q1D.m

```
1
2  function Q1D
3
4  resultats=xlsread('Proba1ereSession20132014.xls'); %rÃ©cupÃ©ration des ...
       donnÃ©es Excel
5
6  [matrice_echantillon_notefinale moyenne_notefinale moyenne_totale ...
       note_finale]=Q1A;%rÃ©cupÃ©ration des donnÃ©es calculÃ©es Ã  la question 1A
7
8  borne_min1=zeros(100,1);%Initialisation des vecteurs contenant les bornes de ...
       l'intervalle de confiance pour la loi de student
9  borne_max1=zeros(100,1);
10 borne_min2=zeros(100,1);%Initialisation des vecteurs contenant les bornes de ...
       l'intervalle de confiance pour la loi de Gauss
11 borne_max2=zeros(100,1);
12 somme=zeros(100,1);
13 sn_1=zeros(100,1);
14 sn=zeros(100,1);
15
16 %Calcul de s_n-1 :
17
18 for j=1:100
19
20     for k=1:20
21
22         somme(j)= somme(j) + (matrice_echantillon_notefinale(k,j)- ...
               moyenne_notefinale(j))^2;
23
24     end
25
26     sn_1(j) = sqrt(somme(j)/19);
27     sn(j) = sqrt(19/20)*sn_1(j);
28
29 end
30
31 %Calcul des bornes des intervalles de confiance :
32
33 t = 2.093;
34 u = 1.96;
35 compteur1=0; %Compteur pour la loi de student
36 compteur2=0; %Compteur pour la loi de Gauss
37
38 for i=1:100
39
40     borne_min1(i) = moyenne_notefinale(i)-(t*sn_1(i)/sqrt(20));
41     borne_max1(i) = moyenne_notefinale(i)+(t*sn_1(i)/sqrt(20));
42     borne_min2(i) = moyenne_notefinale(i)-(u*sn(i)/sqrt(20));
43     borne_max2(i) = moyenne_notefinale(i)+(u*sn(i)/sqrt(20));
44
45     if moyenne_totale≥borne_min1(i) && moyenne_totale≤borne_max1(i)
46         compteur1=compteur1+1;
47     end
48
49     if moyenne_totale≥borne_min2(i) && moyenne_totale≤borne_max2(i)
50         compteur2=compteur2+1;
51     end
```

```
52 end
53 borne_min1bis=mean(borne_min1)
54 borne_max1bis=mean(borne_max1)
55 borne_min2bis=mean(borne_min2)
56 borne_max2bis=mean(borne_max2)
57 compteur1
58 compteur2
59
60
61 end
```

Q2.m

```
1
2
3  function Q2
4
5  resultats=xlsread('Proba1ereSession20132014.xls'); %rÃ©cupÃ©ration des ...
       donnÃ©es Excel
6  matrice_echantillon=zeros(20,7); %Matrice 20 lignes/7 colonnes pour stocker ...
       les Ã©chantillons
7
8  %Calcul des notes finales :
9
10 note_finale=zeros(120,1);
11
12 for j=1:120
13
14     note_finale(j)=mean(resultats(j,:));
15
16 end
17
18 %100 tirages des 7 Ã©chantillons de 20 Ã©tudiants :
19
20 proportion=1/8;
21 compteur=zeros(7,1);
22 ecart_type = sqrt((proportion*(1-proportion))/20);
23 proportion_moins10=zeros(7,1);
24 z=1.645;
25 borne_sup=proportion+(z*ecart_type);
26 nbre_articles=0;
27
28 for m=1:100
29
30 for i=1:7
31
32     echantillon = randsample(120,20,true); %CrÃ©ation d'un Ã©chantillon de 20 ...
           Ã©tudiants
33     matrice_echantillon(:,i)=echantillon; %Remplissage de la matrice ...
           contenant les 7 Ã©chantillons, une colonne = un Ã©chantillon
34
35 end
36
37
38 %Calcul des notes finales de chaque Ã©tudiant de chaque Ã©chantillon :
39
40 matrice_echantillon_notefinale=zeros(20,7);%Matrice qui contiendra les notes ...
       finales de chaque Ã©tudiant de chaque Ã©chantillon
41 nombre_moins10=zeros(7,1);
42 x=0;
43
44 for k=1:7
```

```
45
46      for l=1:20
47
48          matrice_echantillon_notefinale(l,k)=note_finale(matrice_echantillon(l,k));
49
50          if matrice_echantillon_notefinale(l,k)≤10
51              nombre_moins10(k)= nombre_moins10(k) + 1;
52
53          end
54
55
56      end
57      proportion_moins10(k)=nombre_moins10(k)/20;
58
59      if proportion_moins10(k)>borne_sup
60          %HypothÃ¨se rejetÃ©e
61          compteur(k)=compteur(k)+1;
62
63      end
64
65      if proportion_moins10(k)>borne_sup && k≠1
66          %Un organisme rejette l'hypothÃ¨se
67          x=1;
68      end
69
70  end
71
72  nbre_articles=nbre_articles+x;
73
74  end
75  disp('Les autoritÃ©s de l Ulg ont rejetÃ© l hypothÃ¨se dans le nombre de cas ...
        suivant : ');
76  compteur(1)
77  disp('Un article sera publiÃ© le nombre de fois suivant :');
78  nbre_articles
79
80  end
```