STATISTIQUE: 4EME PARTIE
CORRIGES D'EXERCICES

Exercice 1

Supposons la série (xi; ni), représentative de la distribution d'une variable X dans une population, suivante: { (2 ; 1); (4 ; 2); (6 ; 3); (8 ; 2); (10 ; 1) }

1) déterminer l'espérance mathématique et la variance de la variable aléatoire dans la population.
2) établissez la liste de tous les échantillons possibles de taille 2, avec remise, et calculer pour chaque échantillon la moyenne.
Déterminer l'espérance mathématique et la variance de la distribution des moyennes des échantillons de taille 2 dans le cas d'un tirage avec remise.
3) établissez la liste de tous les échantillons possibles de taille 2, sans remise, et calculer pour chaque échantillon la moyenne.
Déterminer l'espérance mathématique et la variance de la distribution des moyennes des échantillons de taille 2 dans le cas d'un tirage avec remise.

1°) dans la population N = 9

xj	nj	xj*xj	xj²*nj
2	1	2	4
4	2	8	32
6	3	18	108
8	2	16	128
10	1	10	100
total	**9**	**54**	**372**

moyenne	$\mu = 6$
écart-type	$\sigma(x) = 2,309$
variance	$\sigma^2(x) = 5,333$

2°) distribution **des 81** moyennes d'échantillons **de taille n = 2**

avec remise	2	4	4	6	6	6	8	8	10
2	**2**	**3**	**3**	**4**	**4**	**4**	**5**	**5**	**6**
4	**3**	**4**	**4**	**5**	**5**	**5**	**6**	**6**	**7**
4	**3**	**4**	**4**	**5**	**5**	**5**	**6**	**6**	**7**
6	**4**	**5**	**5**	**6**	**6**	**6**	**7**	**7**	**8**
6	**4**	**5**	**5**	**6**	**6**	**6**	**7**	**7**	**8**
6	**4**	**5**	**5**	**6**	**6**	**6**	**7**	**7**	**8**
8	**5**	**6**	**6**	**7**	**7**	**7**	**8**	**8**	**9**
8	**5**	**6**	**6**	**7**	**7**	**7**	**8**	**8**	**9**
10	**6**	**7**	**7**	**8**	**8**	**8**	**9**	**9**	**10**

m2	nj	$m_2 * nj$	$m_2^2 * nj$
2	1	2	4
3	4	12	36
4	10	40	160
5	16	80	400
6	19	114	684
7	16	112	784
8	10	80	640
9	4	36	324
10	1	10	100
total	**81**	**486**	**3132**

moyenne	$\mu = 6$
écart-type	$\sigma(m) = 1{,}633$
variance	$\sigma^2(m) = 2{,}667$

On peut ainsi vérifier la validité des formules données en cours. $E(X) = E(m) = \mu$ et $\sigma^2(m) = \sigma^2(X)$ /

3°) distribution des **72 moyennes d'échantillons de taille n = 2**

sans remise	2	4	4	6	6	6	8	8	10
2		3	3	4	4	4	5	5	6
4	3		4	5	5	5	6	6	7
4	3	4		5	5	5	6	6	7
6	4	5	5		6	6	7	7	8
6	4	5	5	6		6	7	7	8
6	4	5	5	6	6		7	7	8
8	5	6	6	7	7	7		8	9
8	5	6	6	7	7	7	8		9
10	6	7	7	8	8	8	9	9	

m2	nj	$m_2 * nj$	$m_2^2 * nj$
3	4	12	36
4	8	32	128
5	16	80	400
6	16	96	576
7	16	112	784
8	8	64	512
9	4	36	324
total	**72**	**432**	**2760**

moyenne	$\mu = 6$
écart-type	$\sigma(m) = 1{,}528$
variance	$\sigma^2(m) = 2{,}333$

On peut ainsi vérifier la validité des formules données en cours. $E(X) = E(m) = \mu$ et
$\sigma^2(m) = (\sigma^2(X) / n)((N-n)/(N-1))$

Exercice 2

Supposons qu'un caractère soit présent dans une proportion de 80% dans une population de 10 unités statistiques. Après avoir établi la distribution des fréquences, déterminer :.
1) l'espérance mathématique et la variance de la distribution des fréquences des échantillons de taille 2 dans le cas d'un tirage avec remise.
2) l'espérance mathématique et la variance de la distribution des fréquences des échantillons de taille 2 dans le cas d'un tirage sans remise.

individu N°	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	9	10
caractère	1	1	1	1	1	1	1	1	0	0

(présent = 1; absent = 0)

dans la population: p = 0,8
q = 0,2

1°) distribution des 100 fréquences d'échantillons de taille n = 2

avec remise	1	2	3	4	5	6	7	8	9	10
1	1	1	1	1	1	1	1	1	0,5	0,5
2	1	1	1	1	1	1	1	1	0,5	0,5
3	1	1	1	1	1	1	1	1	0,5	0,5
4	1	1	1	1	1	1	1	1	0,5	0,5
5	1	1	1	1	1	1	1	1	0,5	0,5
6	1	1	1	1	1	1	1	1	0,5	0,5
7	1	1	1	1	1	1	1	1	0,5	0,5
8	1	1	1	1	1	1	1	1	0,5	0,5
9	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0	0
10	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0	0

f2	nj	$f_2 * nj$	$f_2^2 * nj$
0	4	0	0
0,5	32	16	8
1	64	64	64
total	**100**	**80**	**72**

moyenne	E(f) =	0,8
écart-type	σ(f) =	0,283
variance	σ²(f) =	0,080

On peut ainsi vérifier la validité des formules données en cours. $E(X) = E(f) = p$ et $\sigma^2(f) = (pq) / n$

2°) distribution des **90 fréquences d'échantillons de taille n = 2**

sans remise	1	2	3	4	5	6	7	8	9	10
1		1	1	1	1	1	1	1	0,5	0,5
2	1		1	1	1	1	1	1	0,5	0,5
3	1	1		1	1	1	1	1	0,5	0,5
4	1	1	1		1	1	1	1	0,5	0,5
5	1	1	1	1		1	1	1	0,5	0,5
6	1	1	1	1	1		1	1	0,5	0,5
7	1	1	1	1	1	1		1	0,5	0,5
8	1	1	1	1	1	1	1		0,5	0,5
9	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5		0
10	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0,5	0	

f2	nj	$f_2 \cdot nj$	$f_2^2 * nj$
0	2	0	0
0,5	32	16	8
1	56	56	56
total	**90**	**72**	**64**

moyenne	E(f) =	0,8
écart-type	σ(f) =	0,267
variance	σ²(f) =	0,071

On peut ainsi vérifier la validité des formules données en cours. E(X) = E(f) =p et σ² (f) =[(p q) / n]*[(N-n) / (N - 1)]

Exercice 3

Soit une population de 340 unités statistiques sur laquelle la variable étudiée suit une loi normale de moyenne 250 et de variance 256.
Entre quelles limites doit se trouver la moyenne d'un échantillon de 40 unités pour être représentatif de la population pour un niveau de confiance de 0,94?

X: non définie dans l'exercice, variable mesurée sur 1 unité statistique
E(X) = 250 et σ²(X) = 256 donc σ(X) = 16
L(X) = N (250;16)   N = 340

Variable aléatoire: m_{40}: moyenne de 40 unités
Loi de la variable aléatoire: loi normale car n >30
Paramètres de la loi: **E(m) = E (X) = 250 = μ**
Taux de sondage = 100 * (40 / 340) = 11.8% (> 10%) donc sans remise:
σ²(m) = (16² / 40) *((340 – 40) / (340 – 1)) = 5.664 et **σ(m) = 2.38**
Donc: $L(m_{40})$ = N (250; 2.38) et on peut donc appliquer:

$$\mathbf{P\,[\,\mu - t_{(1-\alpha/2)}\,\sigma_m < m < \mu + t_{(1-\alpha/2)}\,\sigma_m\,] = (1 - \alpha)}$$

$t_{0.97}$ = 1.8808
P [250 - 1.8808*2.38 < m_{40} < 250 + 1.8808*2.38] = 0.94
P [245.5 < m_{40} < 254.5] = 0.94

Si on considère tous les échantillons exhaustifs de taille 40, issus d'une population centrée sur 250, avec un écart type de 16 alors 94% des moyennes des échantillons doivent se trouver entre 245.5 et 254.5.

Exercice 4

Supposons que sur une période de 4 ans on a étudié les rendements en blé de 370 parcelles. On a conclue que le rendement d'une parcelle suivait une loi normale de moyenne 590 et d'écart type 49.
1) Entre quelles limites doit se trouver la moyenne d'un échantillon de 25 parcelles pour être représentatif de la population pour un niveau de confiance de 0,90?
2) Quelle est la probabilité qu'un échantillon de 35 parcelles ait une moyenne comprise entre 580 et 610?
3) Quelle est la probabilité qu'un échantillon de 40 parcelles ait une moyenne comprise entre 570 et 600?

X: rendement d'une parcelle
$E(X) = 590$ et $\sigma(X) = 49$ N = 370
L(X) : N (590 ;49)

1°) Variable aléatoire: m_{25}: rendement moyen de 25 parcelles
Loi de la variable aléatoire: loi normale car : L(X) = N avec σx connu
Paramètres de la loi: $E(m) = E(X) = 590 = \mu$
Taux de sondage = 100 * (25 / 370) < 10% donc avec remise:
$\sigma^2(m) = 49^2 / 25 = 96.04$ et $\sigma(m) = 9.8$
Donc: $L(m_{25}) = N(590; 9.8)$ et on peut donc appliquer:

$$\mathbf{P\,[\,\mu - t_{(1-\alpha/2)}\,\sigma_m < m < \mu + t_{(1-\alpha/2)}\,\sigma_m\,] = (1-\alpha)}$$

$t_{0.95} = 1.6449$
$P\,[590 - 1.6449*9.8 < m_{40} < 590 + 1.6449*9.8] = 0.90$
$P\,[573.88 < m_{25} < 606.12] = 0.90$

Si on considère tous les échantillons de taille 25, issus d'une population centrée sur 590, avec un écart type de 49 alors 90% des moyennes des échantillons doivent se trouver entre 573.88 et 606.12 .

2°) Variable aléatoire: m_{35}: rendement moyen de 35 parcelles
Loi de la variable aléatoire: loi normale car n >30
Paramètres de la loi: $E(m) = E(X) = 590 = \mu$
Taux de sondage = 100 * (35 / 370) = 9.5% (< 10%) donc avec remise:
$\sigma^2(m) = 49^2 / 35 = 68.6$ et $\sigma(m) = 8.283$
Donc: $L(m_{35}) = N(590; 8.283)$ et on peut donc appliquer la variable centrée réduite et utiliser la table statistique 1.

$P\,[\,580 < m_{35} < 610\,] = (1-\alpha)$
$P\,[\,(580 - 590) / 8.283 < T < (610 - 590) / 8.283] = (1-\alpha)$
$P\,[\,-1.21 < T < +2.42\,] = 0.879$

La probabilité pour qu'un échantillon de 35 parcelles ait une moyenne comprise entre 580 et 610 est égale à 0.879 (87.9% des moyennes de 35 parcelles se trouvent dans cet intervalle).

3°) Variable aléatoire: m_{40}: rendement moyen de 40 parcelles
Loi de la variable aléatoire: loi normale car n >30
Paramètres de la loi: $E(m) = E(X) = 590 = \mu$
Taux de sondage = 100 * (40 / 370) = 10.8% (> 10%) donc sans remise:

$\sigma^2(m) = 49^2 / 40 *((370 – 40) / (370 – 1)) = 53.681$ et $\sigma(m) = 7.327$
Donc: $L(m_{40}) = N (590; 7.327)$, et on peut donc appliquer la variable centrée réduite et utiliser la table statistique 1.

$P [570 < m_{40} < 600] = (1 - \alpha)$
$P [(570 – 590) / 7.327 < T < (600 – 590) / 7.327] = (1 - \alpha)$
$P [-,2.73 < T < +1.37] = 0.912$

La probabilité pour qu'un échantillon de 40 parcelles ait une moyenne comprise entre 570 et 600 est égale à 0.912 (91.2% des moyennes de 40 parcelles se trouvent dans cet intervalle).

Exercice 5

Une entreprise fabrique des transistors utilisés dans un récepteur de haute qualité. Un contrôle régulier est effectué à l'aide d'un testeur électronique permettant de détecter d'une façon automatique les transistors défectueux.
1) Si le processus de fabrication produit, selon les normes, en moyenne une proportion de transistors défectueux de 3% entre quelles limites doit se trouver la fréquence de pièces défectueuses sur un échantillon de 200 pièces contrôlées pour un niveau de confiance de 0,90 pour que le processus de fabrication soit aux normes.
2) Afin d'améliorer la qualité de la fabrication on utilise sur la fabrication de 500 transistors un nouveau processus avec lequel on pense n'avoir que 2% de pièces défectueuses. Entre quelles limites doit se trouver la fréquence des pièces défectueuses sur un échantillon de 300 transistors contrôlés pour un niveau de confiance de 0,90?

X: aspect d'une pièce, X = 1 si défectueuse; X = 0 si non défectueuse
$L(X) = B (n;p)$

1°) Variable aléatoire: f_{200}: fréquence d'apparition du caractère dans un échantillon de 200 unités
$p = 0.03$ $q = 0.97$ $n = 200$ $1 – \alpha = 0.9$ $t_{0.95} = 1.645$
$L(f_{200}) = N (E(f); \sigma(f))$ car n p et n q >5
Paramètres de la loi : $\mathbf{E(f) = p = 0.03}$
N>>n donc tirage avec remise et $\sigma^2(f) = (p\ q) / n = 0.000146$ donc $\mathbf{\sigma(f) = 0.012}$
Donc $L(f_{200}) = N (0.03; 0.012)$ et on peut ainsi appliquer:

$P [p - t_{(1 - \alpha / 2)}\ \sigma_f < f < p + t_{(1 - \alpha / 2)}\ \sigma_f] = (1 - \alpha)$
$P [0.0102 < f_{200} < 0.0498] = 0.90$

Pour qu'un procédé de fabrication soit à la norme demandée (en moyenne 3% de pièces défectueuses) il faut une fréquence de pièces défectueuses comprise entre 1.02% et 4.98% dans l'échantillon contrôlé (90% des fréquences de pièces défectueuses, observées dans un échantillon de 200 pièces, sont comprises entre ces valeurs).

2°) Variable aléatoire: f_{300}: fréquence d'apparition du caractère dans un échantillon de 300 unités
$N = 500$ $p = 0.02$ $q = 0.98$ $n = 300$ $1 – \alpha = 0.9$ $t_{0.95} = 1.645$
$L(f_{300}) = N (E(f); \sigma(f))$ car n p et n q >5

Paramètres de la loi: **E(f) = p = 0.02**
Taux de sondage = (300 / 500)*100 = 60% (> 10%) donc tirage sans remise et:
$\sigma^2(f) = [(p\ q) / n]*[(N - n) / (N - 1)] = 2.62\ 10^{-5}$ donc **$\sigma(f)$ = 0.00512**
Donc $L(f_{300}) = N\ (0.02;\ 0.00512)$ et on peut ainsi appliquer:

$P\ [\ p - t_{(1 - \alpha / 2)}\ \sigma_f < f < p + t_{(1 - \alpha / 2)}\ \sigma_f\ ] = (1 - \alpha)$
$P\ [\ 0.0116 < f_{300} < 0.0284\ ] = 0.90$

Pour que le procédé de fabrication soit améliorer (centré en moyenne à 2% de pièces défectueuses) il faut une fréquence de pièces défectueuses comprise entre 1.16% et 2.84% dans l'échantillon contrôlé (90% des fréquences de pièces défectueuses, observées dans un échantillon de 300 pièces, seraient comprises entre ces valeurs).

Exercice 6

Si m est la moyenne d'un échantillon aléatoire de taille n tiré avec remise d'une population normale de moyenne μ et de variance 100, déterminer n tel que :
1) $P\ (\ \mu - 10 \le m \le \mu + 10\ ) = 0{,}9544$
2) $P\ (\ \mu - 5 \le m \le \mu + 5\ ) = 0{,}9544$
3) $P\ (\ \mu - 2 \le m \le \mu + 2\ ) = 0{,}9544$
4) $P\ (\ \mu - 2 \le m \le \mu + 2\ ) = 0{,}6826$
5) $P\ (\ \mu - 2 \le m \le \mu + 2\ ) = 0{,}8664$
On supposera le tirage avec remise.
Quelles conclusions pouvez-vous en tirer ?

ME abs = $t_{1-\alpha/2} * \sigma x / \sqrt{n}$ donc $\boxed{n = t^2_{1-\alpha/2} * \sigma^2 x / ME^2_{abs}}$

Avec $\sigma x = 10$ on obtient:

question	**1 - α**	**1 – α / 2**	**$t_{1-\alpha/2}$**	**α**	**ME abs**	**n**
1	0,9544	0,9772	2	0,0456	10	4
2	0,9544	0,9772	2	0,0456	5	16
3	0,9544	0,9772	2	0,0456	2	100
4	0,6826	0,8413	1	0,3174	2	25
5	0,8664	0,9332	1,5	0,1336	2	56,25 = 57

On remarque que:

- pour diviser une marge d'erreur absolue par un facteur 2 (passer de 10 à 5) il faut multiplier l'effectif de l'échantillon par un facteur 4 (2^2) – pour un risque d'erreur α donné (question 1 et 2 à comparer).
- pour diviser une marge d'erreur absolue par un facteur 5 (passer de 10 à 2) il faut multiplier l'effectif de l'échantillon par un facteur 25 (5^2) – pour un risque d'erreur α donné (questions 1 et 3 à comparer).
- pour une marge d'erreur absolue donnée on peut réduire la taille de l'échantillon n si au augmente le risque d'erreur α commis sur la détermination des limites de l'intervalle (questions 3 ; 4 ; 5 à comparer).

Exercice 7

Le directeur des ressources humaines d'une entreprise de 200 employés veut estimer la dextérité manuelle de sa nouvelle main-d'œuvre. Des résultats aux tests de dextérité sur une ancienne main d'œuvre étaient distribués selon une loi normale de variance 25.
1) Les tests sur un échantillon de 40 employés sélectionnés au hasard donnent une moyenne égale à 71. Donner les limites entre lesquelles se trouve la dextérité moyenne de tous les employés avec 1% de risque d'erreur. Calculer les marges d'erreur.
2) Si les tests avaient eu lieu avec un échantillon de 20 employés avec une moyenne de 71, entre quelles limites se trouverait la dextérité moyenne de tous les employés ? Calculer les marges d'erreur.

X: mesure de la dextérité manuelle d'un individu
$E(X) = \mu$ et $\sigma^2(X) = 25$ $\quad$ $L(X) = N(\mu; 5)$ $\quad$ $N = 200$

1°) Variable aléatoire: m_{40}: moyenne des résultats de 40 employés
Loi de la variable aléatoire: loi normale car $L(X) = N$ et $\sigma^2(x)$ connue
Paramètres de la loi: $E(m) = E(X) = \mu$ avec $\hat{\mu} = 71$
Taux de sondage = (40 / 200)*100 = 20% (> 10%) donc tirage sans remise
$\sigma^2(m) = 25 / 40 *((200 - 40) / (200 - 1)) = 0.502$ et $\sigma(m) = 0.709$
Donc: $L(m_{40}) = N(71; 0.709)$ et on peut donc appliquer:

$$\mathbf{P[\,m - t_{(1-\alpha/2)}\,\sigma_m < \mu < m + t_{(1-\alpha/2)}\,\sigma_m\,] = (1 - \alpha)}$$

$t_{0.995} = 2.576$
$P[71 - 2.576*0.709 < \mu < 71 + 2.576*0.709] = 0.99$
$P[69.174 < \mu < 72.826] = 0.99$

L'intervalle [69.17 ; 72.83] à 1% de risque de ne pas contenir la vraie valeur μ.
L'amplitude de l'intervalle = 72.826 – 69.174 = 3.652
La marge d'erreur absolue = 3.652 / 2 = 1.826
La marge d'erreur relative = 100* 1.826 / 71 = 2.6%

2°) Variable aléatoire: m_{20}: moyenne des résultats de 20 employés
Loi de la variable aléatoire: loi normale car $L(X) = N$ et $\sigma^2(x)$ connue
Paramètres de la loi: $E(m) = E(X) = \mu$ avec $\hat{\mu} = 71$
Taux de sondage = (20 / 200)*100 = 10% ($\leq$ 10%) donc tirage avec remise
$\sigma^2(m) = 25 / 20 = 1.25$ et $\sigma(m) = 1.118$
Donc: $L(m_{20}) = N(71; 1.118)$ et on peut donc appliquer:

$$\mathbf{P[\,m - t_{(1-\alpha/2)}\,\sigma_m < \mu < m + t_{(1-\alpha/2)}\,\sigma_m\,] = (1 - \alpha)}$$

$t_{0.995} = 2.576$
$P[71 - 2.576*1.118 < \mu < 71 + 2.576*1.118] = 0.99$
$P[68.12 < \mu < 73.88] = 0.99$

L'intervalle [68.12 ; 73.88] à 1% de risque de ne pas contenir la vraie valeur μ.
L'amplitude de l'intervalle = 73.88 – 68.12 = 5.76
La marge d'erreur absolue = 5.76 / 2 = 2.88
La marge d'erreur relative = 100* 2.88 / 71 = 4.1%

Exercice 8

Des études sur 500 sportifs minimes ont montré que le temps de réponse à un stimulus a une dispersion de 15μs.
1) Si on admet que la dispersion ne varie pas, et que pour un échantillon de 20 sportifs minimes tirés au hasard parmi 500, on a un résultat moyen de 71 μs. Estimer par intervalle de confiance le temps de réponse moyen de la catégorie minime avec un niveau de confiance de 0,95. Calculer les marges d'erreur.
2) Mêmes questions si sur un échantillon de taille 40 on observe un temps moyen de 71 μs.
3) Mêmes questions si sur un échantillon de taille 60 on observe un temps moyen de 71 μs.

X: temps de réponse à un stimulus d'un individu
$E(X) = \mu$ et $\sigma(X) = 15$ L(X) = inconnue N = 500

1°) Variable aléatoire: m_{20}: moyenne des résultats de 20 individus
Loi de la variable aléatoire: inconnue
Paramètres de la loi: $E(m) = \mu$ avec $\hat{\mu} = 71$
Loi de la variable aléatoire: inconnue car loi de X inconnue et n <30 Donc: on ne peut donc pas appliquer: $P [m - t_{(1-\alpha/2)}\, \sigma_m < \mu < m + t_{(1-\alpha/2)}\, \sigma_m] = (1-\alpha)$
On ne peut pas répondre à la question, il faudrait un échantillon de taille > 30 par exemple.

2°)Variable aléatoire: m_{40}: moyenne des résultats de 40 individus
Loi de la variable aléatoire: loi normale car n >30 et $\sigma^2(x)$ connue
Paramètres de la loi: $E(m) = \mu$ avec $\hat{\mu} = 71$
Taux de sondage = (40 / 500)*100 = 8% (< 10%) donc tirage avec remise
$\sigma^2(m) = 225 / 40 = 5.625$ et $\sigma(m) = 2.372$
Donc: $L(m_{40}) = N(\mu; 2.372)$ et on peut donc appliquer:
$\mathbf{P [m - t_{(1-\alpha/2)}\, \sigma_{(m)} < \mu < m + t_{(1-\alpha/2)}\, \sigma_{(m)}] = (1-\alpha)}$
m = 71 $(1-\alpha) = 0.95$ $t_{0.975} = 1.96$
$P [66.35 < \mu < 75.65] = 0.95$
L'intervalle [66.35 ; 75.65] à 5% de risque de ne pas contenir la vraie valeur μ.
L'amplitude de l'intervalle = 75.65 – 66.35 = 9.3
La marge d'erreur absolue = 9.3 / 2 = 4.65
La marge d'erreur relative = 100* 4.65 / 71 = 6.55%

3°)Variable aléatoire: m_{40}: moyenne des résultats de 60 individus
Loi de la variable aléatoire: loi normale car n >30 et $\sigma^2(x)$ connue
Paramètres de la loi: $E(m) = \mu$ avec $\hat{\mu} = 71$
Taux de sondage = (60 / 500)*100 = 12% (> 10%) donc tirage sans remise
$\sigma^2(m) = 225 / 60 *((500 - 60) / (500 - 1)) = 3.307$ et $\sigma(m) = 1.818$
Donc: $L(m_{60}) = N(\mu; 1.818)$ et on peut donc appliquer:
$\mathbf{P [m - t_{(1-\alpha/2)}\, \sigma_{(m)} < \mu < m + t_{(1-\alpha/2)}\, \sigma_{(m)}] = (1-\alpha)}$
m = 71 $(1-\alpha) = 0.95$ $t_{0.975} = 1.96$
$P [67.437 < \mu < 74.563] = 0.95$
L'intervalle [67.44 ; 74.56] à 5% de risque de ne pas contenir la vraie valeur μ.
L'amplitude de l'intervalle = 74.563 – 67.437 = 7.126
La marge d'erreur absolue = 7.126 / 2 = 3.563
La marge d'erreur relative = 100* 3.563 / 71 = 5%

Exercice 9

Le distributeur d'un appareil estime que celui-ci doit être vendu 50$. Afin de vérifier si ce prix est respecté dans ses 300 magasins, il vérifie le prix moyen proposé par un échantillon de n magasins. On choisira un niveau de confiance de 0.95.
1) Sur 20 magasins le prix moyen est de 55$ avec un écart type égal à 5$. Quel doit être la conclusion du distributeur?
2) N'étant pas satisfait du résultat il complète son étude avec 20 autres magasins tirés au hasard. Sur les 40 magasins le prix moyen est de 56$ avec un écart type égal à 6$.. Quel doit être la conclusion du distributeur?

X: prix de vente d'un appareil dans un magasin
$E(X) = 50 = \mu$ et $\sigma(X)$ inconnu L(X) inconnue N = 300

1°) Variable aléatoire: m_{20}: moyenne des prix de 20 appareils
Loi de la variable aléatoire: inconnue car L(X) inconnue et n <30
Donc: $L(m_{20})$ inconnue et on ne peut donc pas appliquer:

$$\mathbf{P\,[\,m - t_{(1-\alpha/2)}\,\sigma_{(m)} < \mu < m + t_{(1-\alpha/2)}\,\sigma_{(m)}\,] = (1 - \alpha)}$$

On ne peut pas répondre à la question, il faudrait un échantillon de taille > 30 par exemple.

2°) Variable aléatoire: m_{40}: moyenne des prix de 40 appareils
Pour traiter cette question il y a 3 méthodes possibles:
méthode A: l'échantillonnage, on utilise la moyenne connue de la population et on va déduire dans quel intervalle doit se trouver la moyenne de l'échantillon. On vérifiera si notre moyenne d'échantillon appartient à l'intervalle de prédiction.
méthode B: l'estimation de μ, on suppose connus les paramètres de l'échantillon et on calculera entre quelles valeurs doit se trouver la moyenne de la population. On vérifiera si notre moyenne μ se trouve dans cet intervalle.
méthode C: test statistique de comparaison d'un nombre à une norme, variance de la population inconnue (partie 5 du cours) pour le risque d'erreur donné.

Dans les 3 cas la conclusion sera la même quand au respect ou non des prix!!

Méthode A: utilisation de la théorie de l'échantillonnage
Variable aléatoire: m_{40}: moyenne des prix de 40 appareils
Loi de la variable aléatoire: loi normale car n >30
$L(m_{40}) = N\,(E(m)\,;\,\sigma(m))$
Paramètres de la loi: $E(m) = \mu = 50$
Taux de sondage = (40 / 300)*100 = 13.3% (> 10%) donc tirage sans remise

$\sigma^2(m) = 6^2 / 39 *((300 - 40) / (300 - 1)) = 0.8027$ et $\sigma(m) = 0.896$
Donc: $L(m_{40}) = N\,(50;\,0.896)$ et on peut donc appliquer:

$$\mathbf{P\,[\,\mu - t_{(1-\alpha/2)}\,\sigma_{(m)} < m < \mu + t_{(1-\alpha/2)}\,\sigma_{(m)}\,] = (1 - \alpha)}$$

$(1 - \alpha) = 0.95$ $t_{0.975} = 1.96$
P [48.24 < m < 51.76] = 0.95

Si le prix moyen dans 40 magasins a été de 56$, ce prix ne se trouve pas dans l'intervalle calculé, donc les prix ne sont pas respectés mais on a 5% de chance de se tromper en disant cela.

Méthode B: détermination de l'intervalle d'estimation de μ
Variable aléatoire: m_{40}: moyenne des prix de 40 appareils
Loi de la variable aléatoire: loi normale car n >30
$L(m_{40}) = N(E(m)\,;\,\sigma(m))$
Paramètres de la loi: $E(m) = \mu$ avec $\hat{\mu} = 56$
Taux de sondage = (40 / 300)*100 = 13.3% (> 10%) donc tirage sans remise

$\sigma^2(m) = 6^2 / 39 *((300 - 40) / (300 - 1)) = 0.8027$ et $\sigma(m) = 0.896$
Donc: $L(m_{40}) = N(\mu; 0.896)$ et on peut donc appliquer:

$$\mathbf{P\left[m - t_{(1-\alpha/2)}\,\sigma_{(m)} < \mu < m + t_{(1-\alpha/2)}\,\sigma_{(m)} \right] = (1-\alpha)}$$

$(1 - \alpha) = 0.95$ $t_{0.975} = 1.96$
$P[54.24 < \mu < 57.76] = 0.95$

Le prix de vente souhaité de 50$ ne se trouve pas dans l'intervalle calculé, donc les prix ne sont pas respectés mais l'intervalle d'estimation a 5% de chance de ne pas contenir la vraie valeur μ.

Méthode C: test de comparaison d'une moyenne à un nombre
Variable aléatoire: m_{40}: moyenne des prix de 40 appareils
Loi de la variable aléatoire: loi normale car n >30
$L(m_{40}) = N(E(m); \sigma(m))$
Paramètres de la loi: $E(m) = \mu = 50$
Taux de sondage = (40 / 300)*100 = 13.3% (> 10%) donc tirage sans remise

$\sigma^2(m) = 6^2 / 39 *((300 - 40) / (300 - 1)) = 0.8027$ et $\sigma(m) = 0.896$
Donc: $L(m_{40}) = N(50; 0.896)$

H0 : $\mu = 50$ (les prix sont respectés)
H1 : $\mu \neq 50$ (les prix ne sont pas respectés)
Test bilatéral

Critère statistique calculé: $t_{\text{calculé}} = (56 - 50) / 0.896 = 6.696$
$t_{\text{théorique}} = t_{0.975} = 1.96$
| t calculé | > t théorique, on rejette H0 avec un risque de le faire à tort inférieur à 5%. On accepte H1, on a pu mettre en évidence qu'il existait une différence entre les prix pratiqués entre ces 40 magasins et le prix préconisé, avec en réalité un risque inférieur à 0.1% de chance de se tromper (la valeur 6.70 n'est plus dans la table 1).

Exercice 10

On souhaite estimer le salaire mensuel moyen de 2 niveaux de cadres d'entreprises.
L'étude porte sur des échantillons et on a obtenu :

Cadres	effectif	Moyenne ($)	Ecart type ($)
Niveau 1	26	3521	338
Niveau 2	57	2756	245

Estimer, par intervalle de confiance à l'unité près, pour un niveau de confiance de 0,98 le salaire mensuel moyen pour chaque niveau. Préciser dans chaque cas les marges d'erreur absolue et relative. On admettra que le salaire d'un cadre de niveau j suit une loi normale.

X: salaire d'une personne
$E(X) = \mu$ et $\sigma^2(X)$ inconnue $L(X) = N(\mu\ ; \sigma(X))$

Attention!!
Dans cet exercice 3521$ représente la moyenne des salaires des 26 cadres de niveau 1 et 338$ représente l'écart type s(x) de ces 26 salaires donc $s(x) = (SCE_x / 26)^{0.5.}$ Même remarque pour les cadres niveau 2.
Remarque: on peut noter $\sigma^2(m)$ ou $\hat{\sigma}^2(m)$, l'important est surtout de faire la différence entre $\sigma^2(x)$ (variance connue dans la population) et $\hat{\sigma}^2(x)$ (variance dans la population estimée = SCEx / (n-1) à partir des n données de l'échantillon)

Cadres de niveau 1: Variable aléatoire: m_{26}: salaire moyen de 26 cadres du niveau 1
Loi de la variable aléatoire :Student (25) car loi X normale, $\sigma^2(X)$ inconnue et n < 30
Donc: $L(m_{26})$ = St (25) et on peut donc appliquer:
$$\mathbf{P\,[\,m - t_{(1-\alpha/2)}(\nu)\,\sigma_m < \mu < m + t_{(1-\alpha/2)}(\nu)\,\sigma_m\,] = (1-\alpha)}$$
N >> n donc avec remise:
$\sigma^2(m) = \hat{\sigma}^2(x) / n$ ou encore $\sigma^2(m) = s^2_x / (n-1)$ et donc $\sigma^2(m) = 338^2/25 = 4569.76$
$\sigma(m) = 67.6$
$(1-\alpha) = 0.98$ $t_{0.99}(25) = 2.485$
$P\,[3353 < \mu < 3689] = 0.98$

L'intervalle [3353; 3689] a 2% de risque de ne pas contenir le vrai salaire moyen des cadres de niveau 1.

On peut calculer les marges d'erreur pour les cadres de niveau 1 :
L'amplitude de l'intervalle = 3689 - 3353 = 336
La marge d'erreur absolue = 336 / 2 = 168
La marge d'erreur relative = 100* 168 / 3521 = 4.8%

Cadres de niveau 2: Variable aléatoire: m_{57}: salaire moyen de 57 cadres du niveau 2
Loi de la variable aléatoire: loi normale car n >30
Paramètres de la loi: $E(m) = E(X) = \mu$ avec $\hat{\mu} = 2756$
N >> n donc avec remise:

$\sigma^2(m) = \hat{\sigma}^2(x) / n$ ou encore $\sigma^2(m) = s^2_x / (n-1)$ et donc $\sigma^2(m) = 245^2/56 = 1071.875$
$\sigma(m) = 32.74$
Donc: $L(m_{57}) = N(\mu; 32.74)$ et on peut appliquer:
$$\mathbf{P}\,[\,\mathbf{m} - \mathbf{t}_{(1-\alpha/2)}\,\sigma_m < \mu < \mathbf{m} + \mathbf{t}_{(1-\alpha/2)}\,\sigma_m\,] = (1-\alpha)$$
$(1 – \alpha) = 0.98$ $\quad t_{0.99} = 2.3263$
P [2680 < μ < 2832] = 0.98

L'intervalle [2680; 2832] a 2% de risque de ne pas contenir le vrai salaire moyen des cadres de niveau 2.

On ne peut calculer les marges d'erreur pour les cadres de niveau 2 :
L'amplitude de l'intervalle = 2832 - 2680 = 152
La marge d'erreur absolue = 152 / 2 = 76
La marge d'erreur relative = 100* 76 / 2756 = 2.8%

Exercice 11

Des essais en laboratoire sur 20 lampes témoins utilisées sur des panneaux de contrôle électroniques conduisent aux durées de vie en heures ci-dessous :

450	410	412	380	407
455	375	390	355	364
410	413	345	430	390
330	440	381	451	415

Supposons que la durée de vie soit distribuée normalement, estimer par intervalle de confiance la durée de vie moyenne de toute la production au niveau 0,98.

X: durée de vie d'une lampe
E(X) = μ et $\sigma^2(X)$ inconnue
L(X) = N (E(x); σ(X)) on suppose que l'on travaille dans un domaine où X suit une loi normale.
Variable aléatoire: m_{20}: durée de vie moyenne de 20 lampes
Loi de la variable aléatoire: loi de Student (ddl = 19) car L(X) = N (E(x); σ(X)) $\sigma^2(X)$ inconnue et n<30
Paramètres de la loi: E(m) = E (X) = μ $\quad$ avec $\hat{\mu} = m = 400.15$
N >> n donc avec remise:
$\sigma^2(m) = \hat{\sigma}^2(x) / n = 35.72^2 / 20 = 63.8$
Ou encore $\sigma^2(m) = s^2_x / (n-1) = 34.82^2 / 19 = 63.8$ $\quad$ et donc **σ(m) = 7.99**
Ou encore $\sigma^2(m)$ = SCE / n (n-1) = 24244.55 / (20*19) = 63.8

Remarque : on peut noter $\sigma^2(m)$ ou $\hat{\sigma}^2(m)$, l'important est surtout de faire la différence entre $\sigma^2(x)$ (variance connue dans la population) et $\hat{\sigma}^2(x)$ (variance dans la population estimée = SCEx / (n-1) à partir des n données de l'échantillon)

Donc: L (m_{20}) = S(19) (μ; 7.99) et on peut donc appliquer:
$$\mathbf{P}\,[\,\mathbf{m} - \mathbf{t}(\nu)_{(1-\alpha/2)}\,\sigma_m < \mu < \mathbf{m} + \mathbf{t}(\nu)_{(1-\alpha/2)}\,\sigma_m\,] = (1-\alpha)$$
$(1 – \alpha) = 0.98$ $\quad t_{0.99}(19) = 2.539$
P [379.86 < μ < 420.44] = 0.98

L'intervalle [379.86;420.44] a 2% de risque de ne pas contenir la vraie durée de vie moyenne des lampes fabriquées.

Exercice 12

Sur un échantillon aléatoire de 18 prélèvements indépendants dans une population binomiale de taille infinie on observe 9 unités statistiques présentant une caractéristique étudiée. Déterminer une estimation par intervalle de confiance à 99% de la proportion d'unités présentant la caractéristique dans la population.

On a n prélèvements indépendants $n = 18$

$f = 9 / 18 = 0.5$ $(1 - \alpha) = 0.99$ $(1 - \alpha/2) = 0.995$

$P[p_1 < p < p_2] = (1 - \alpha)$

$n < 100$ et $0.1 < f < 0.9$, les fréquences ne suivent pas une loi Normale on applique donc:

$p_1 = \dfrac{k}{k + (n - k + 1) F_{1-\alpha/2}(\nu 1, \nu 2)}$ avec $\nu_1 = 2(n - k + 1)$ et $\nu_2 = 2k$

$p_1 = \dfrac{\mathbf{9}}{\mathbf{9 + 10 * 3.50}} = 0.2045$ avec $\nu_1 = 20$ et $\nu_2 = 18$

$p_2 = \dfrac{(k+1) F_{1-\alpha/2}(\nu 1, \nu 2)}{n - k + (k+1) F_{1-\alpha/2}(\nu 1, \nu 2)}$ avec $\nu_1 = 2(k + 1)$ et $\nu_2 = 2(n - k)$

$p_2 = \dfrac{\mathbf{10 * 3.50}}{\mathbf{9 + 10 * 3.50}} = 0.7955$ avec $\nu_1 = 20$ et $\nu_2 = 18$

$P[0.2045 < p < 0.7955] = 0.99$

L'intervalle [20.45%; 79.55%] a 1% de risque de ne pas contenir la vraie proportion d'individus possédant la caractéristique étudiée.

Exercice 13

Un sondage mené auprès de 1000 adultes pour connaître le niveau de satisfaction d'une émission de télévision a donné les résultats suivants :
satisfaits : 146 insatisfaits : 851 sans avis : 3
1) Entre quelles limites peut-on s'attendre de trouver la vraie proportion de personnes satisfaites au risque de se tromper 1 fois sur 20 ?
2) Même question pour les personnes sans avis.

Dans cet exercice on va estimer la proportion des personnes satisfaites (1°) et celle des sans avis (2°).
Dans les 2 cas N>>>n donc le tirage est supposé avec remise.

1°) $n = 1000$ $k = 146$ $f = 146 / 1000 = 0.146$

☛ $n \geq 100$ et $0.1 \leq f \leq 0.9$ donc on applique $L(f_{1000}) = N(E(f); \sigma(f))$

Paramètres de la loi : E(f) = p avec $\hat{p}$ = 0.146 σ (f) = $\sqrt{\frac{f(1-f)}{n}}$.= 0.0112

Remarque: on peut noter σ^2(f) ou $\hat{\sigma}^2$(f),
Donc L(f_{1000}) = N (p; 0.0112) et on applique:
P [f - $t_{1-\alpha/2}$ σ (f) < p < f + $t_{1-\alpha/2}$ σ (f)] = (1 - α)
α = 1/20 = 0.05 (1 - α) = 0.95 $t_{1-\alpha/2} = t_{0.975} = 1.96$
P [0.1240 < p < 0.1680]= 0.95

L'intervalle [12.4% ; 16.8%] a 5% de risque de ne pas contenir la vraie proportion de personnes satisfaites de l'émission télévisée.

2°) n = 1000 k = 3 f = 3 / 1000 = 0.003

☛ n ≥100 et f ≤ 0.1 donc on applique donc: P [p_1 < p < p_2] = (1 - α) avec :

$p_1 = \frac{1}{2n} X^2_{\alpha/2}(\nu)$ $\nu = 2k$

$p_2 = \frac{1}{2n} X^2_{1-\alpha/2}(\nu)$ $\nu = 2k + 2$

$p_1 = \frac{\mathbf{1}}{\mathbf{2*1000}}$ *1.24 = 0.00062 ν = 6 $X^2_{0.025}(6) = 1.24$

$p_2 = \frac{\mathbf{1}}{\mathbf{2*1000}}$ * 17.5 = 0.00875 ν = 8 $X^2_{0.975}(8) = 17.5$

P [0.00062 < p < 0.00875]= 0.95

L'intervalle [0.06%; 0.88%] a 5% de risque de ne pas contenir la vraie proportion de personnes sans avis sur l'émission télévisée.

Exercice 14

Déterminer un intervalle de confiance à 95% pour la variance et l'écart-type de la population de l'exercice 11.

Si on a n prélèvements indépendants et si on admet que la distribution de la variable suit une loi Normale dans la population.

$$P\left[\frac{\sum_{i=1}^{i=n}(xi-\bar{x})^2}{X^2_{1-\alpha/2}(\nu)} < \sigma^2(x) < \frac{\sum_{i=1}^{i=n}(xi-\bar{x})^2}{X^2_{\alpha/2}(\nu)}\right] = (1-\alpha) \qquad \nu = n-1$$

Pour l'exercice 11 où n = 20
SCEx = 24244.55
(1 – α) = 0.95 (1 – α / 2) = 0.975 ν = 19
$X^2_{0.025}(19) = 8.91$ $X^2_{0.025}(19) = 32.9$

P [736.9 < σ^2(x) < 2721] = 0.95
L'intervalle [736.9 ; 2721] a 5% de risque de ne pas contenir la vraie valeur de la variance.

On en déduit :
P [27.15 < $\sigma(x)$ < 52.16] = 0.95
L'intervalle [27.15; 52.16] a 5% de risque de ne pas contenir la vraie valeur de l'écart-type.

Exercices supplémentaires

Exercice 1s

On veut estimer le salaire annuel moyen offert aux jeunes diplômés des écoles d'ingénieurs . Un échantillon de 95 nouveaux ingénieurs a donné les résultats suivants :
m = 180000F SCE x = 2,432 10 10
En admettant que les salaires annuels sont distribués normalement, entre quelles limites peut se situer le salaire annuel moyen pour l'ensemble des jeunes ingénieurs diplômés si on utilise un niveau de confiance de 99% ? On supposera le tirage avec remise.

X: salaire d'un jeune diplômé
$E(X) = \mu$ et $\sigma^2(X)$ inconnue $L(X) = N(E(X); \sigma^2(x))$
Variable aléatoire: m_{95}: salaire moyen de 95 jeunes diplômés
Loi de la variable aléatoire: loi normale car n>30
Paramètres de la loi: $E(m) = E(X) = \mu$ avec $\hat{\mu} = m = 180000$
N>>n donc avec remise:
$\sigma^2(m) = \hat{\sigma}^2(x) / n = s^2{}_x / (n-1) = SCEx / n\,(n-1)$ et donc $\boldsymbol{\sigma(m) = 1650.27}$
Remarque : on peut noter $\sigma^2(m)$ ou $\hat{\sigma}^2(m)$, l'important est surtout de faire la différence entre $\sigma^2(x)$ (variance connue dans la population) et $\hat{\sigma}^2(x)$ (variance dans la population estimée = SCEx / (n-1) à partir des n données de l'échantillon)
Donc: $L(m_{95}) = N(\mu; 1650.27)$ et on peut donc appliquer:
$\mathbf{P\,[\,m - t\,(\nu)_{(1-\alpha/2)}\,\sigma_m < \mu < m + t\,(\nu)_{(1-\alpha/2)}\,\sigma_m\,] = (1 - \alpha)}$
$(1 - \alpha) = 0.99$ $t_{0.995} = 2.576$
P [175749 < μ < 184251] = 0.99

L'intervalle [175749;184251] a 1% de risque de ne pas contenir le vrai salaire moyen des jeunes ingénieurs.

Exercice 2s

Supposons que l'on étudie la résistance à la température d'une peinture utilisée sur des assiettes. Supposons que cette variable soit distribuée normalement dans la population. La fabrication est limitée à 80 assiettes dans un premier temps.
🕮 Sur 20 assiettes la température maximale supportée fut :

72	73	73	74	75
74	75	76	66	70
67	69	73	72	73
71	70	72	71	70

🕮 Une étude sur 40 assiettes a donné une moyenne de 72 et une variance de 6,5.

Calculer dans chaque cas l'estimation par intervalle de confiance au niveau 0,95 de la température moyenne de résistance de cette peinture et déterminer les précisions relative et absolue.

X : temps de résistance de la peinture d'une assiette à la température
L(X) = N (E(X) ; σ(X)) avec E(X) = μ et σ²(X) inconnue N = 80

1er cas: n = 20

Variable aléatoire: m_{20}: temps moyen de résistance de 20 assiettes
Loi de la variable aléatoire: loi de Student (ddl = 19) car L(X) = N et σ²(X) inconnue et n<30
Paramètres de la loi: E(m) = E (X) = μ avec $\hat{\mu}$ = m = 71.8
Taux de sondage = 100 * 20 / 80 = 25% (>10 %) donc sans remise :
$\sigma^2(m) = [\hat{\sigma}^2(x) / n]*[(N - n) / (N-1)] = (6.8 / 20) * (60 / 79) = 0.258$
ou encore $\sigma^2(m) = [s^2_x / (n-1)]*[(N - n) / (N-1)] = (6.46 / 19)*(60 / 79) = 0.258$
ou encore $\sigma^2(m) = [SCEx / n(n-1)]*[(N -n) / (N-1)] = (129.2 / (20*19))* (60 / 79) = 0.258$
et donc **σ(m) = 0.508**

<u>Remarque</u> : on peut noter σ²(m) ou $\hat{\sigma}^2$(m), l'important est surtout de faire la différence entre σ^2(x) (variance connue dans la population) et $\hat{\sigma}^2$(x) (variance dans la population estimée = SCEx / (n-1) à partir des n données de l'échantillon)

Donc: $L(m_{95})$ = St (19) (μ; 0.508) et on peut donc appliquer:

$$\mathbf{P [m - t (\nu)_{(1 - \alpha / 2)} \sigma_m < \mu < m + t (\nu)_{(1 - \alpha / 2)} \sigma_m] = (1 - \alpha)}$$

(1 – α) = 0.95 $t_{0.975}(19) = 2.093$
P [70.75 < μ < 72.85] = 0.95

L'intervalle [70.75;72.85] a 5% de risque de ne pas contenir le vrai temps de résistance moyen de la peinture à la température.

L'amplitude de l'intervalle = 72.85 - 70.75 = 2.10
La marge d'erreur absolue = 2.10 / 2 = 1.05
La marge d'erreur relative = 100* 1.05 / 71.8 = 1.46%

2ème cas n = 40

Variable aléatoire: m_{40}: temps moyen de résistance de 40 assiettes
Loi de la variable aléatoire: loi Normale car n>30
Paramètres de la loi: E(m) = E (X) = μ avec $\hat{\mu}$ = m = 72
Taux de sondage = 100 * 40 / 80 = 50% (>10 %) donc sans remise:
$\sigma^2(m) = [s^2_x / (n-1)]*[(N - n) / (N-1)] = (6.5 / 39)*(40 / 79) = 0.0844$ et donc
σ(m) = 0.2905

<u>Remarque</u>: on peut noter σ²(m) ou $\hat{\sigma}^2$(m), l'important est surtout de faire la différence entre σ^2(x) (variance connue dans la population) et $\hat{\sigma}^2$(x) (variance dans la population estimée = SCEx / (n-1) à partir des n données de l'échantillon)

Donc: $L(m_{40}) = N (\mu; 0.2905)$ et on peut donc appliquer:

$$\mathbf{P [m - t_{(1 - \alpha / 2)} \sigma_m < \mu < m + t_{(1 - \alpha / 2)} \sigma_m] = (1 - \alpha)}$$

(1 – α) = 0.95 $t_{0.975} = 1.96$
P [71.43 < μ < 72.57] = 0.95

L'intervalle [71.43;72.57] a 5% de risque de ne pas contenir le vrai temps de résistance moyen de la peinture à la température.

L'amplitude de l'intervalle = 72.57 – 71.43 = 1.14
La marge d'erreur absolue = 1.14 / 2 = 0.57
La marge d'erreur relative = 100* 0.57 / 72 = 0.79%

Exercice 3s

Déterminer un intervalle de confiance à 95% pour la variance et l'écart-type de la population de l'exercice 1s.

Si on a n prélèvements indépendants et si on admet que la distribution de la variable suit une loi Normale dans la population.

$$P\left[\frac{\sum_{i=1}^{i=n}(xi-\bar{x})^2}{X^2_{1-\alpha/2}(\nu)} < \sigma^2(x) < \frac{\sum_{i=1}^{i=n}(xi-\bar{x})^2}{X^2_{\alpha/2}(\nu)}\right] = (1-\alpha) \qquad \nu = n-1$$

Pour l'exercice 1s où n = 95

$\hat{\sigma}^2(x)$ = SCEx / (n –1) = 2.432 10^{10} / 94 SCEx = 2.432 10^{10}
(1 – α) = 0.95 (1 – α / 2) = 0.975 ν = 94 (ν > 30)
On va approximer la loi du X² par la loi Normale et on obtient ainsi (voir les formules sur la table du X²):
$X^2_{0.025}(94) = [t_{0.025} + (2*94-1)^{0.5}]^2 / 2 = [-1.96 + (2*94-1)^{0.5}]^2 / 2 = 68.62$
$X^2_{0.975}(94) = [t_{0.975} + (2*94-1)^{0.5}]^2 / 2 = [+1.96 + (2*94-1)^{0.5}]^2 / 2 = 122.22$
P [198985436 < σ²(x) < 354415622] = 0.95
L'intervalle [198985436; 354415622] a 5% de risque de ne pas contenir la vraie valeur de la variance.
On en déduit :
P [14106 < σ(x) < 18826] = 0.95
L'intervalle [14106; 18826] a 5% de risque de ne pas contenir la vraie valeur de l'écart-type.