

EXERCICES DE
GÉNIE INDUSTRIEL

(version 2.7 Révision 10 du 2015-06-23)

EXERCICE 1.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: tirage non-exhaustif

Énoncé:

Pour réaliser le montage d'un système électronique, nous disposons de résistances issues d'une production (très) importante, où nous savons que le pourcentage de pièces défectueuses est de 30% sur toute la production (retour d'expérience).

Nous devons envoyer 1'000 résistances à un client d'une production journalière de 100'000 unités.

1. Quelle est la probabilité d'en avoir exactement 250 de défectueuses?
2. Quelle est la probabilité d'en avoir moins ou égal à 324 de défectueuses?
3. Quelle est la probabilité d'avoir entre 250 et 320 pièces défectueuses?

Solution:

D'abord, il s'agit d'appliquer la loi binomiale (tirage non exhaustif, dont l'ordre n'est pas pris en compte) une première fois d'une probabilité et ensuite d'une probabilité cumulée.

Nous allons prendre pour ce calcul MS Excel:

S1. La probabilité d'en avoir exactement 250 de mauvaises dans les 1000 est de:

$$=LOI.BINOMIALE(250;1000;30\%;0)=0.0061\%$$

S2. La probabilité d'en avoir moins ou égal à 324 de mauvaises sur les 1'000 est de:

$$=LOI.BINOMIALE(324;1000;30\%;1)=95.38\%$$

Donc il y a aussi une probabilité cumulée de:

$$1-95.38\%=4.62\%$$

d'en avoir plus de 325 de mauvaises! Ce qui est très mauvais (on pouvait s'en douter)!

S3. La probabilité cumulée est alors de:

$$= LOI.BINOMIALE (320;1000;30\%;1)- LOI.BINOMIALE(250;1000;30\%;1)=92.05\%$$

EXERCICE 2.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: tirage non-exhaustif, loi binomiale cumulée

Énoncé:

Dans une usine de production un client externe souhaite un lot de 1'000 pièces.

Nous savons qu'il y a 30% de la production (très grosse production par hypothèse!) qui est défectueuse à cause de la grande complexité des pièces.

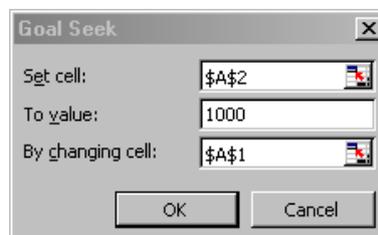
Combien de pièces devons-nous lui envoyer pour qu'il y ait 19 chances sur 20 d'avoir 1'000 pièces bonnes sachant que nous n'avons pas l'infrastructure ni le temps pour les tester toutes?

Solution:

Il s'agit dans MS Excel d'écrire la formule suivante:

	A
1	1000
2	=CRITERE.LOI.BINOMIALE(A1;1-30%;1-95%)

D'activer l'outil Valeur cible:



ce qui donne:

~1471 pièces

Ainsi, en envoyant 1'471 pièces au client, il a bien:

$$=LOI.BINOMIALE(471;1471;30%;1)=95.64\%$$

de probabilité cumulée de trouver entre 0 et 471 pièces défectueuses dans le lot d'envoi de 1471 pièces. Il n'aura donc que 5% de probabilité cumulée d'avoir moins de 1000 pièces bonnes au final!

Par ailleurs le responsable qualité chez le client sera donc sûr avec 95% de probabilité cumulée que l'acceptation du lot est ok s'il en vérifie au moins 470 pièces en faisant donc le calcul suivant:

$$=CRITERE.LOI.BINOMIALE(1471;30%;95%)=470$$

EXERCICE 3.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Analyse bayésienne

Énoncé:

D'un lot de 10 pièces dont le 30% est défectueux, nous prélevons sans remise un échantillon de taille 3. Quelle est la probabilité que la seconde pièce soit bonne (quelque soit la première)?

Solution:

Nous avons:

$$\begin{aligned} P(A) &= \sum_{i=1}^n P(A/B_i)P(B_i) = \sum_{i=1}^2 P(A/B_i)P(B_i) \\ &= P(A/B_1)P(B_1) + P(A/B_2)P(B_2) = \frac{7}{9} \frac{3}{10} + \frac{6}{9} \frac{7}{10} = 70\% \end{aligned}$$

où $P(A/B_1)$ est la probabilité que la deuxième soit bonne sachant que la première est mauvaise et $P(A/B_2)$ est la probabilité que la deuxième soit bonne sachant que la première est bonne. $P(B_1)$ est donc la probabilité que la première soit mauvaise, $P(B_2)$ la probabilité que la première soit bonne.

EXERCICE 4.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Panne, loi binomiale

Énoncé:

Nous souhaitons tester l'alternateur d'un groupe électrogène. La probabilité de défaillance à la sollicitation de ce matériel est estimée à 1 défaillance pour 1'000 démarrages.

Nous décidons d'effectuer un test de 100 démarrages.

E1. Quelle est la probabilité d'observer exactement 1 panne au cours de ce test?

E2. Quelle est la probabilité d'observer plus de 1 panne au cours de ce test?

Solution:

S1. Nous avons:

$$P(k = 1) = C_k^N \frac{N!}{k!(N-k)!} p^k q^{N-k} = \frac{100!}{1!(100-1)!} \left(\frac{1}{1000}\right)^1 \left(1 - \frac{1}{1000}\right)^{99} \cong 9\%$$

Soit dans MS Excel:

$$=LOI.BINOMIALE(1;100;1/1000;0)=9\%$$

S2. De même:

$$=1-LOI.BINOMIALE(1;100;1/1000;1)=0.46\%$$

EXERCICE 5.*Niveau: Gymnase (Lycée)**Auteur: Vincent ISOZ**Mots-clés: Boîtes à moustaches (box-plot, box-whisker plot)***Énoncé:**

Soit les données suivantes de deux ateliers 1 et 2 de production dont nous avons recueilli le nombre de pièces retournées chaque mois:

	B	C	D	E
1	Atelier	Défauts	Atelier	Défauts
2	1	65	2	60
3	1	68	2	45
4	1	72	2	46
5	1	55	2	50
6	1	64	2	47
7	1	70	2	55
8	1	66	2	49
9	1	74	2	52
10	1	75	2	50
11	1	70	2	46
12	1	93	2	50
13	1	64	2	52
14	1	60	2	52
15	1	62	2	48
16	1	70	2	52
17	1	80	2	63
18	1	61	2	53
19	1	60	2	54
20	1	62	2	54
21	1	68	2	54
22	1	70	2	53
23	1	72	2	55
24	1	71	2	57

Représentez avec MS Excel les boîtes à moustaches correspondantes pour comparer les deux ateliers.

Solution:

Avec MS Excel 2003 et antérieur:

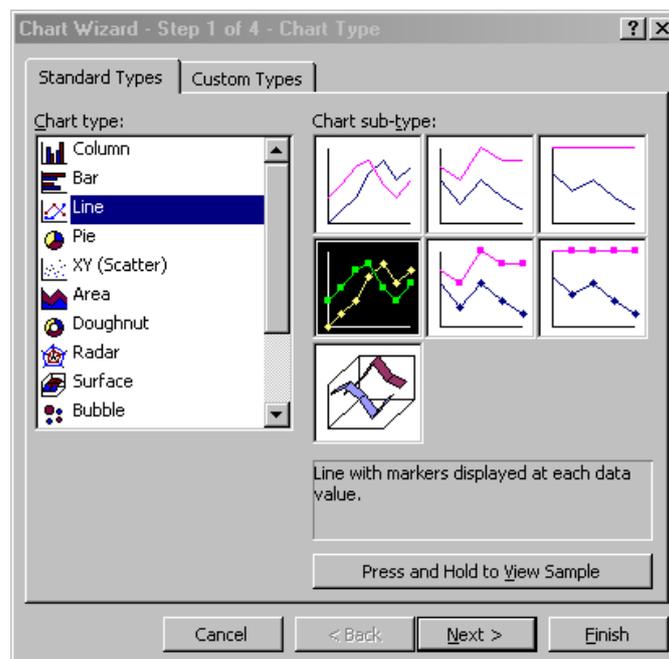
D'abord il nous faut créer la table résumant les statistiques nécessaires aux boîtes à moustache (**attention à respecter l'ordre des données telle que présenté ci-dessous!**):

	H	I	J
2	Valeurs	Atelier 1	Atelier 2
3	Q1	63	50
4	Min	55	45
5	Médiane	68	52
6	Max	93	63
7	Q3	72	54

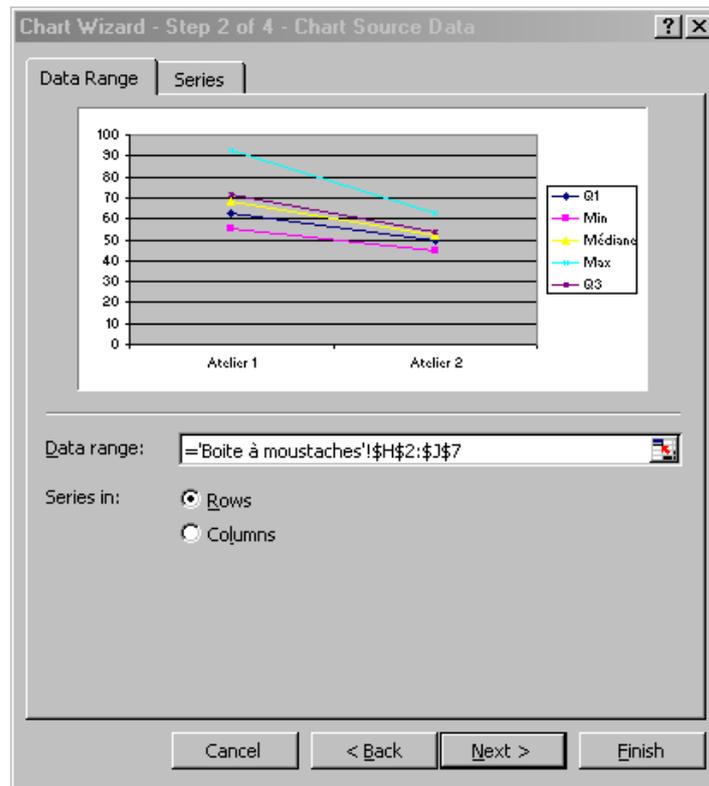
Avec les formules simples de MS Excel:

	H	I	J
2	Valeurs	Atelier 1	Atelier 2
3	Q1	=QUARTILE(C2:C24;1)	=QUARTILE(E2:E24;1)
4	Min	=MIN(C2:C24)	=MIN(E2:E24)
5	Médiane	=MEDIAN(C2:C24)	=MEDIAN(E2:E24)
6	Max	=MAX(C2:C24)	=MAX(E2:E24)
7	Q3	=QUARTILE(C2:C24;3)	=QUARTILE(E2:E24;3)

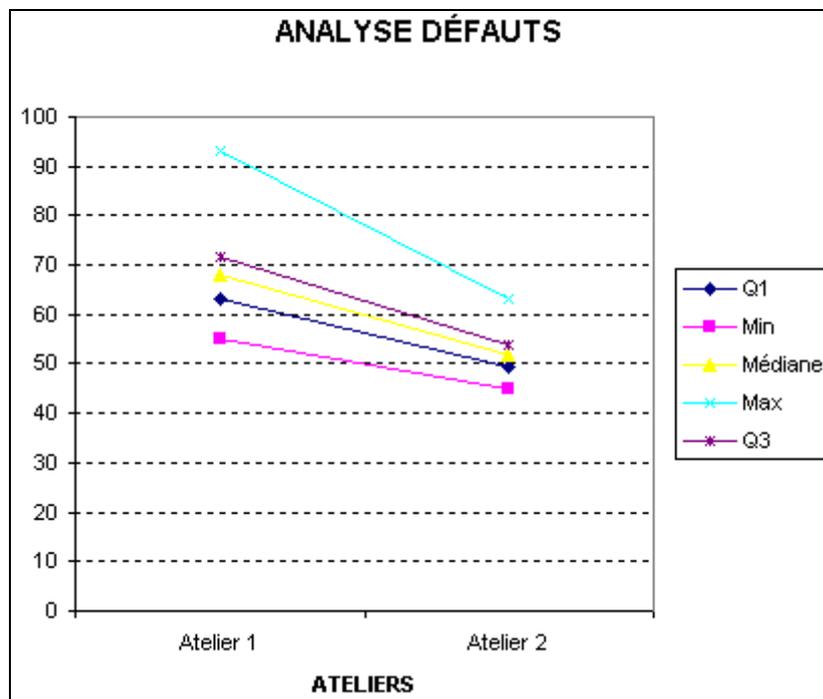
Quand nous avons ce tableau, nous allons pouvoir construire nos boîtes à moustache. En sélectionnant H2:J7 et en lançant l'assistant graphique:



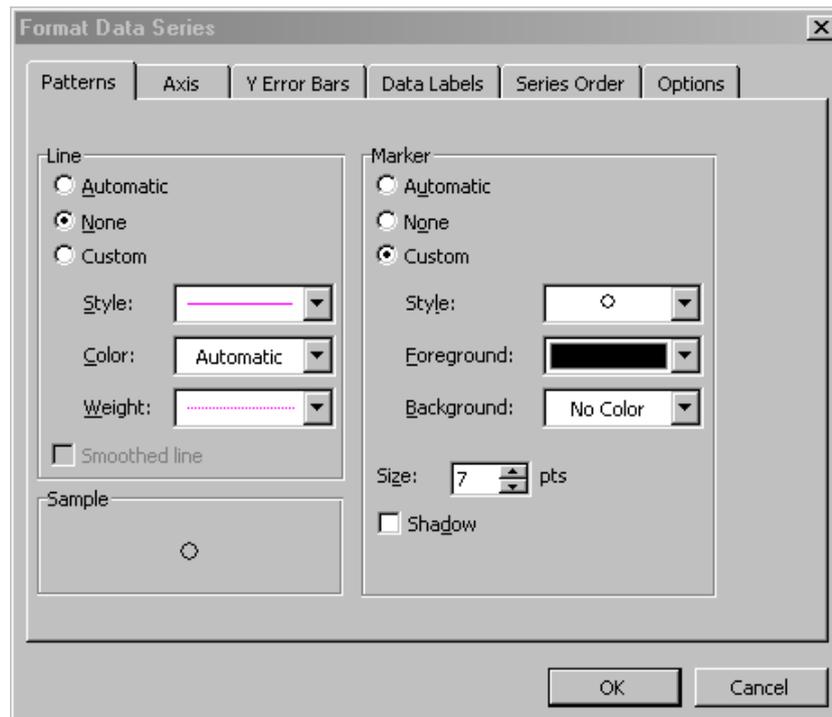
Prenez ensuite les séries en ligne plutôt qu'en colonne:



Si vous validez vous aurez à cette étape le résultat suivant après quelques manipulations évidentes:

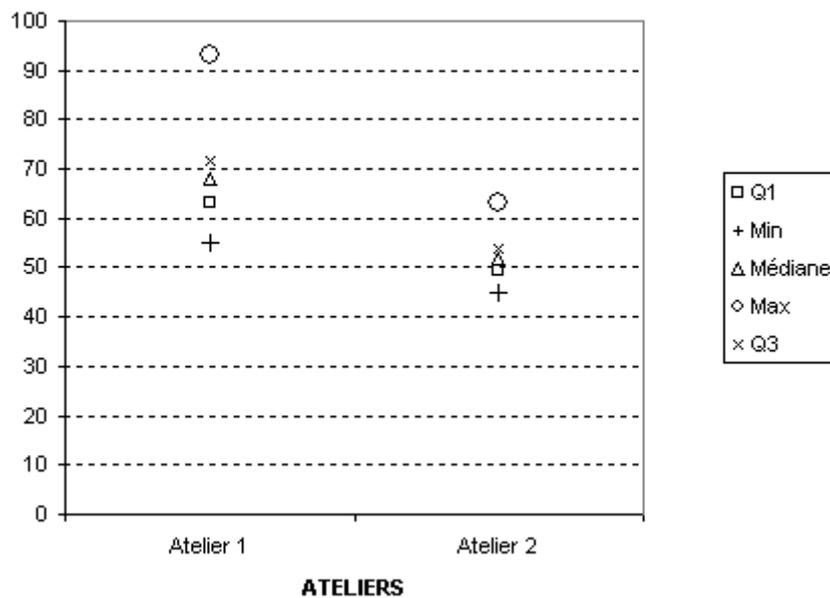


Pour chaque format de série, supprimez le trait (Trait : "Aucun"), choisissez la couleur de la marque (Noir) et supprimez l'arrière-plan:

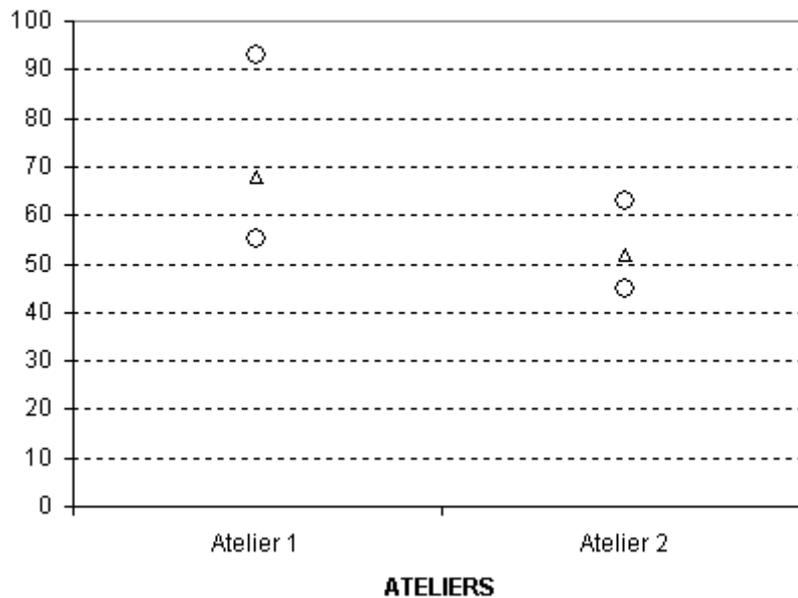


Vous obtiendrez alors:

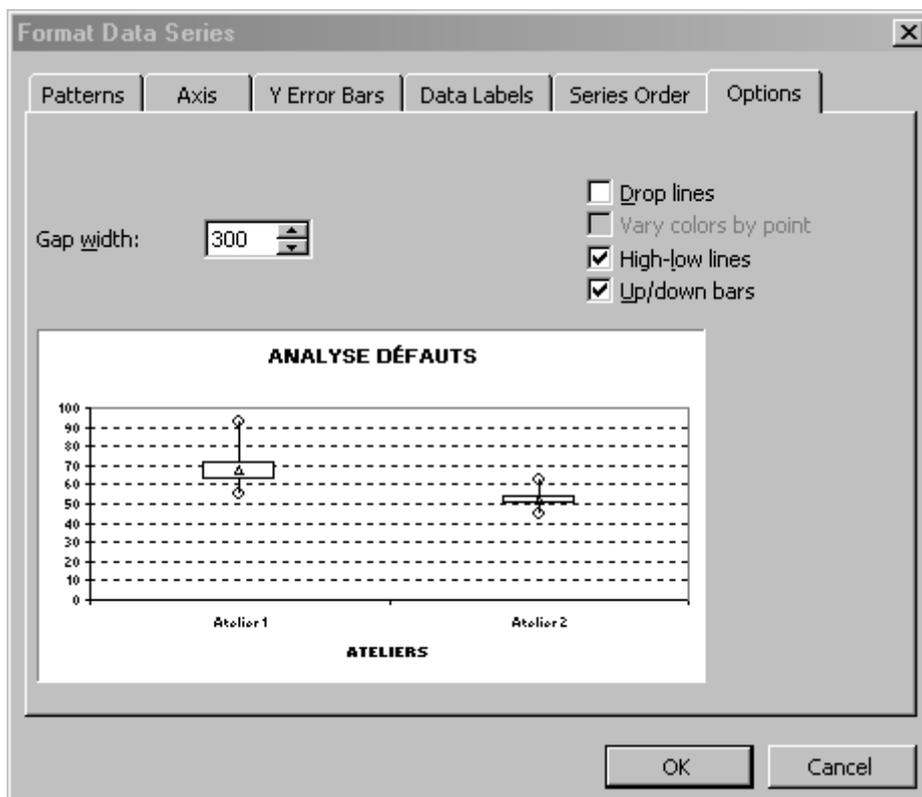
ANALYSE DÉFAUTS



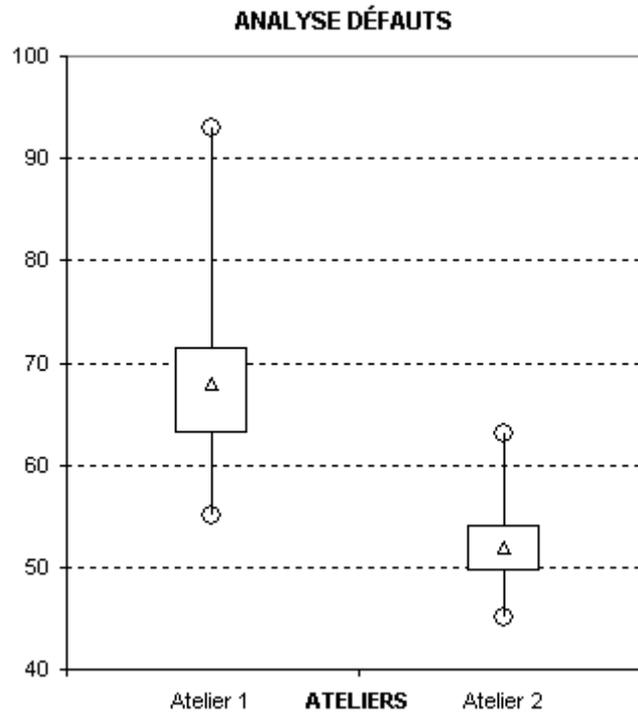
Ensuite, modifier le style de la marque à "Aucune" pour Q1 et Q3 et à "Cercle" de taille 7 (au lieu de 5) pour Min et Max. Nous avons alors:

ANALYSE DÉFAUTS

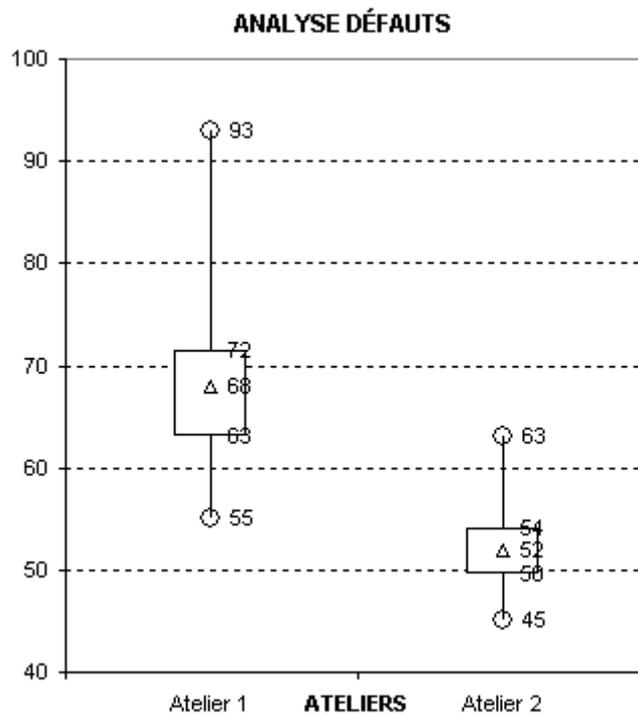
Sélectionner ensuite une des séries visibles (peu importe laquelle) et dans le Format de celle-ci dans l'onglet "Options", indiquer "300" à la place de "150" comme "Largeur de l'intervalle", et cocher "Lignes haut/bas" et "Barres haut/bas":



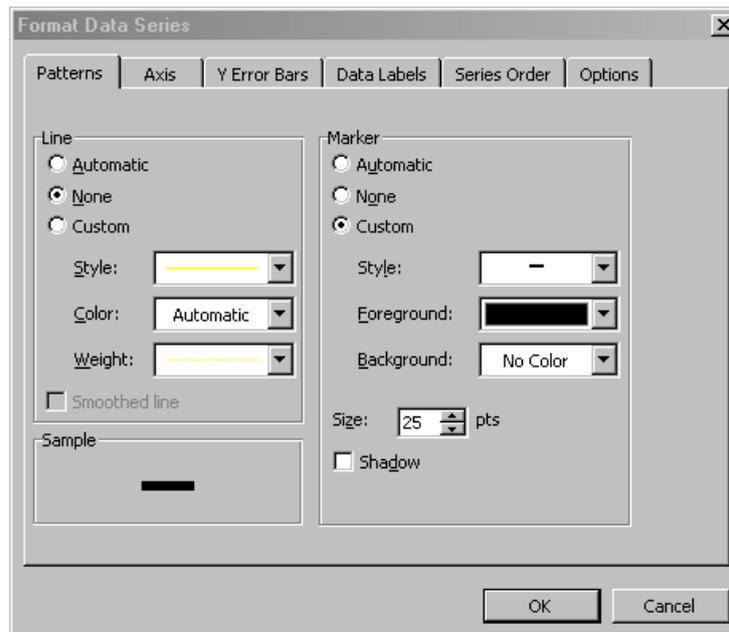
Après validation et redéfinition de la plage de l'axe des ordonnées, nous obtenons:



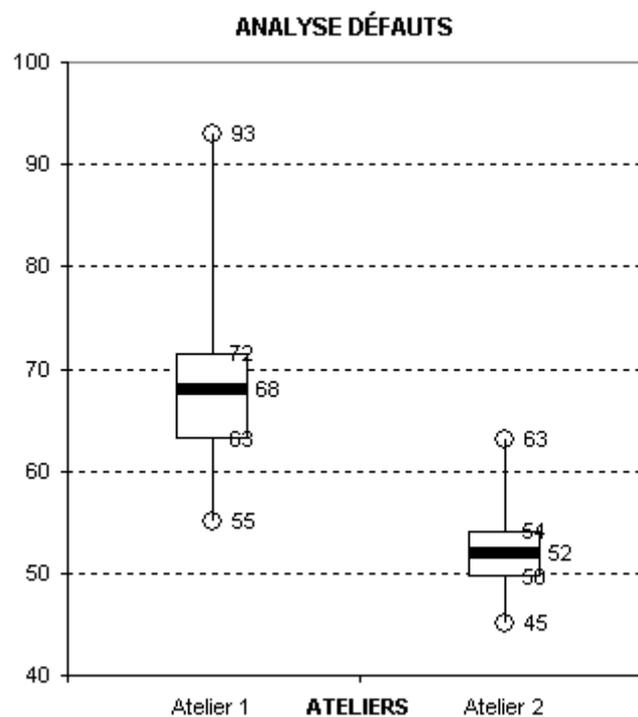
Si nous activons l'affichage des valeurs (par la méthode classique) nous avons:



Ensuite, pour la médiane il suffit d'aller dans ses propriétés et de changer sa marque comme suit:



Pour avoir finalement:



Avec MS Excel 2013 et ultérieur maintenant car la procédure a considérablement changé (et pas en bien...):

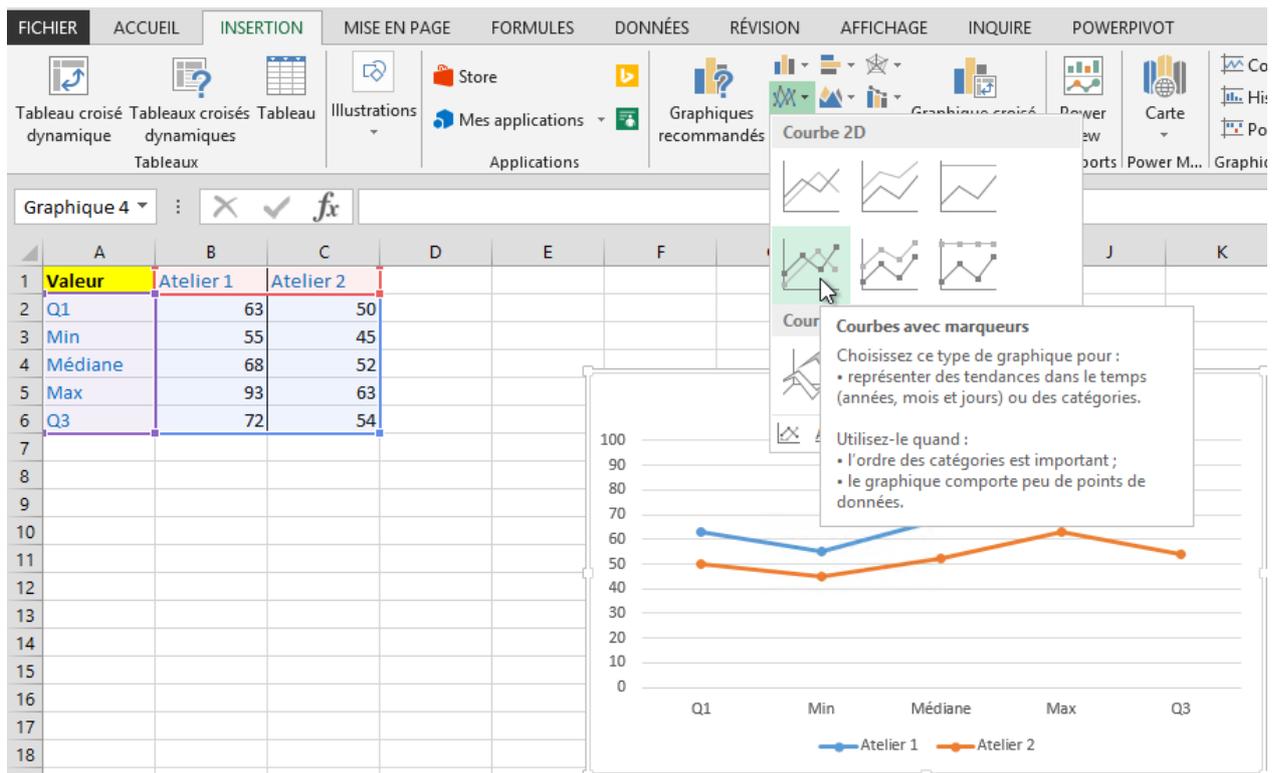
Nous repartons du même tableau (attention à respecter l'ordre des données tel qu'indiqué ci-dessous):

	A	B	C
1	Valeur	Atelier 1	Atelier 2
2	Q1	63	50
3	Min	55	45
4	Médiane	68	52
5	Max	93	63
6	Q3	72	54

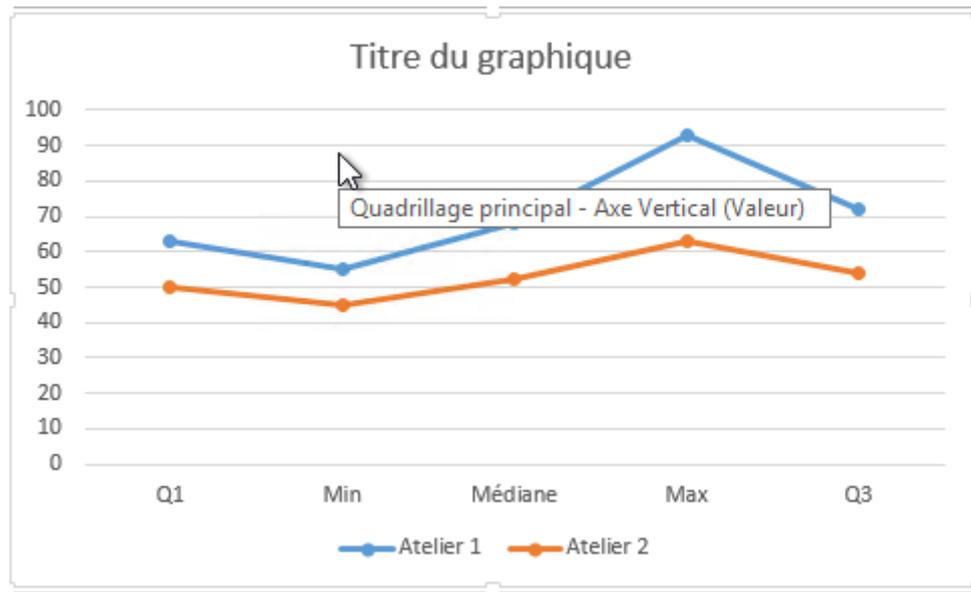
On sélectionne la plage suivante:

	A	B	C
1	Valeur	Atelier 1	Atelier 2
2	Q1	63	50
3	Min	55	45
4	Médiane	68	52
5	Max	93	63
6	Q3	72	54

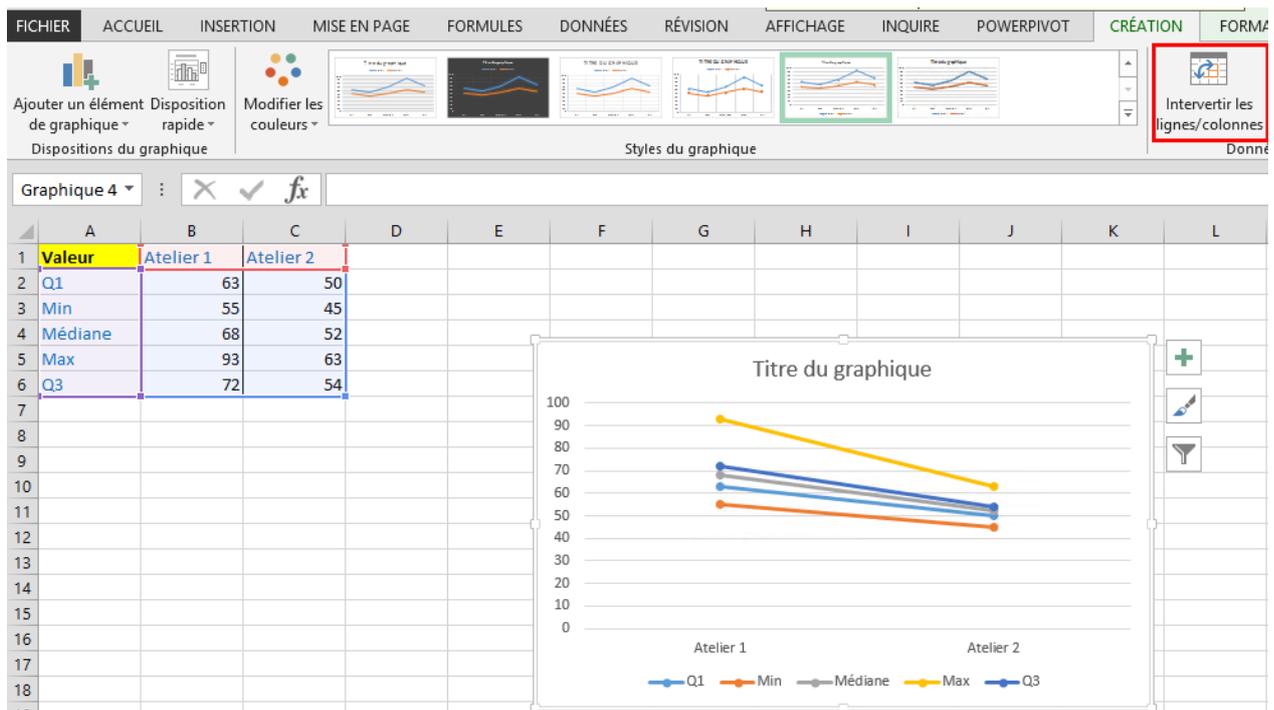
On va dans **Insertion/Courbes avec marqueurs**:



On obtient:

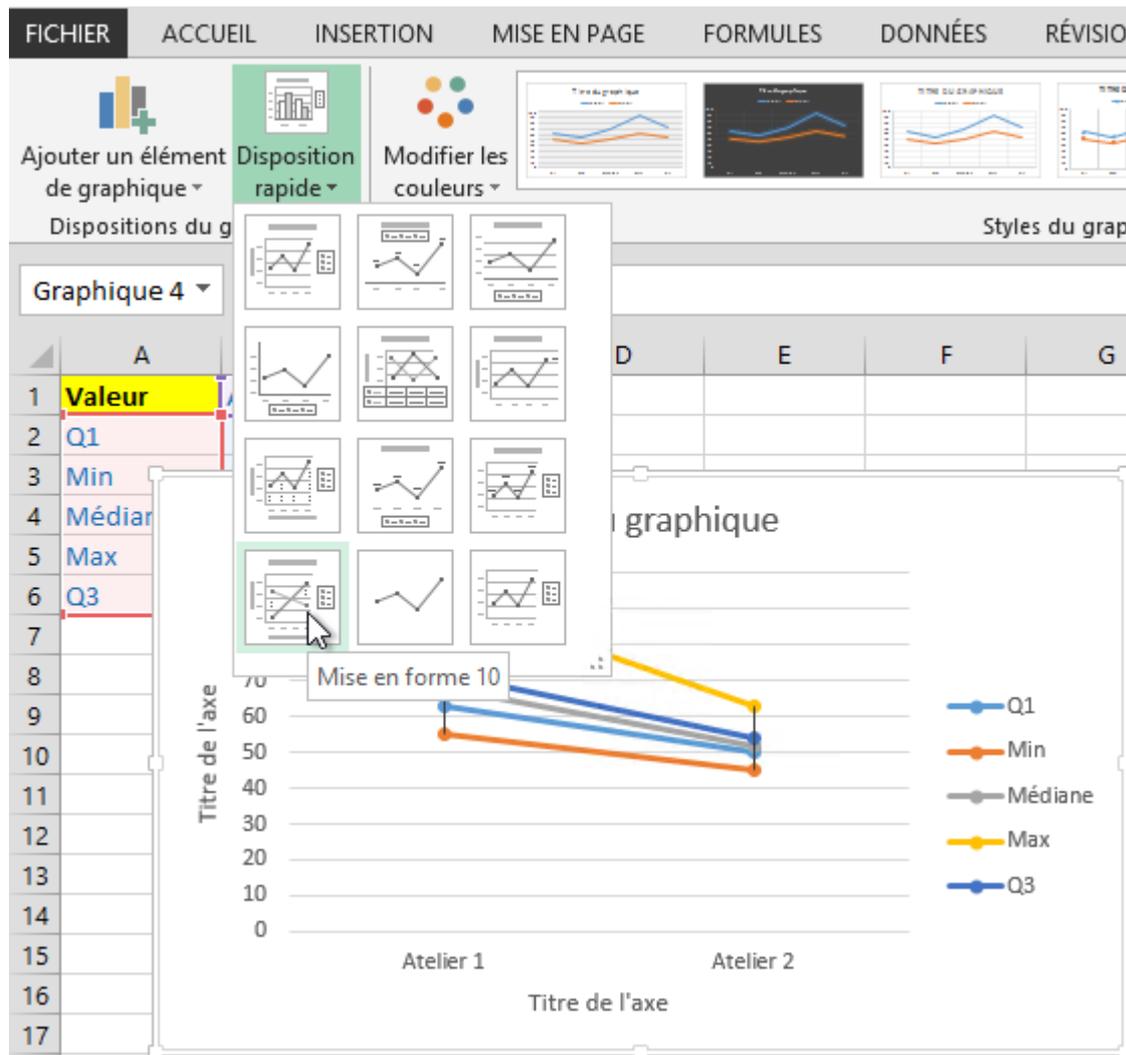


et en cliquant sur le bout **Intervertir lignes/colonnes** visible ci-dessous dans un encadré rouge, nous avons:

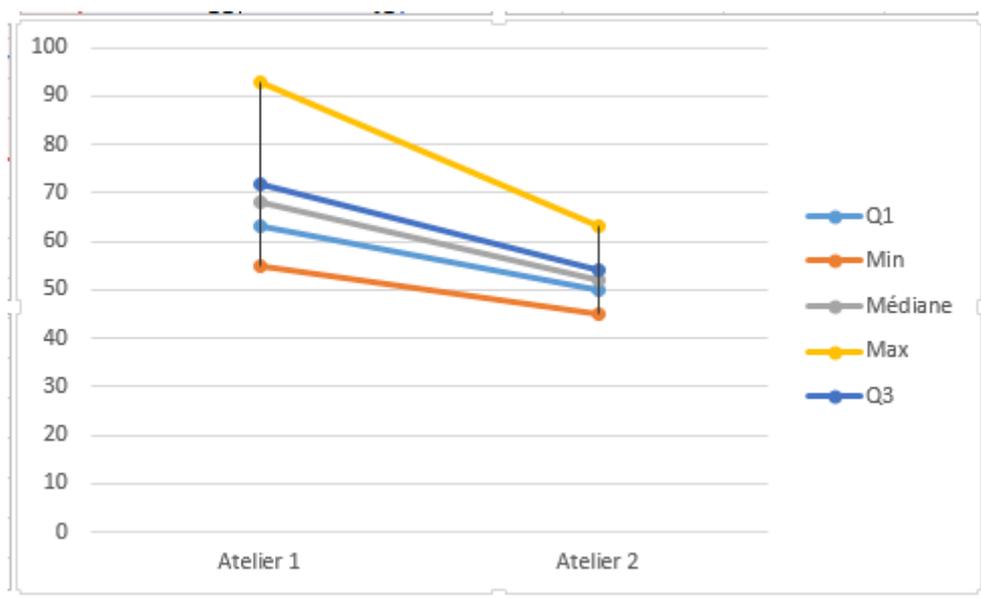


Ce qui donne:

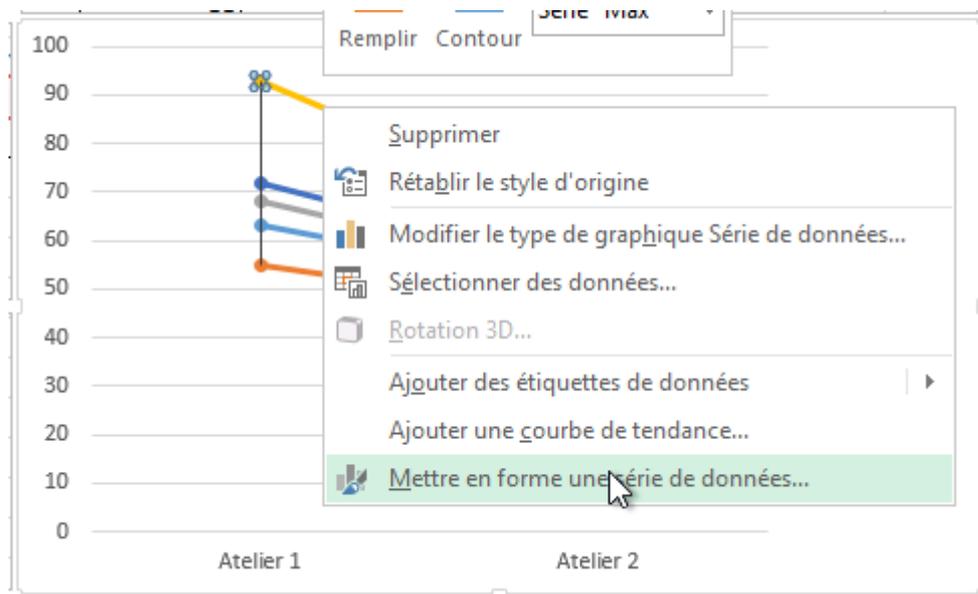
Ensuite, il faut aller dans **Disposition rapide** et prendre **Mise en forme 10** pour avoir les lignes verticales noires qui relient les points extrêmes de chaque future-box plot:



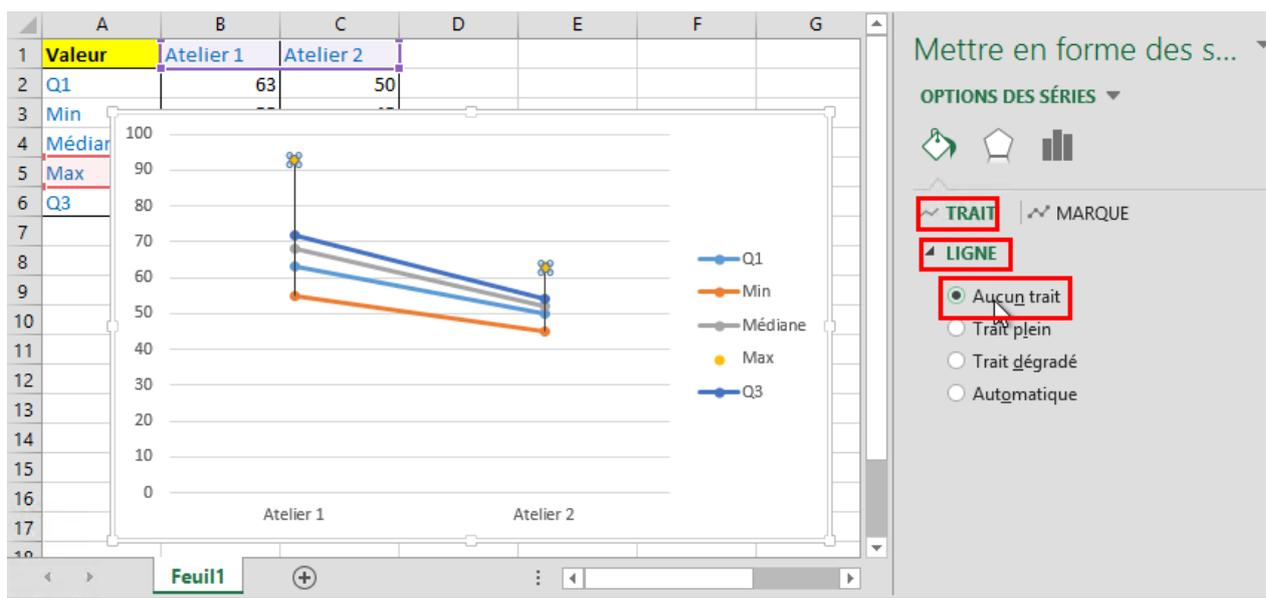
Ce qui donne après avoir supprimer les titres des axes et le titre principal du graphique:



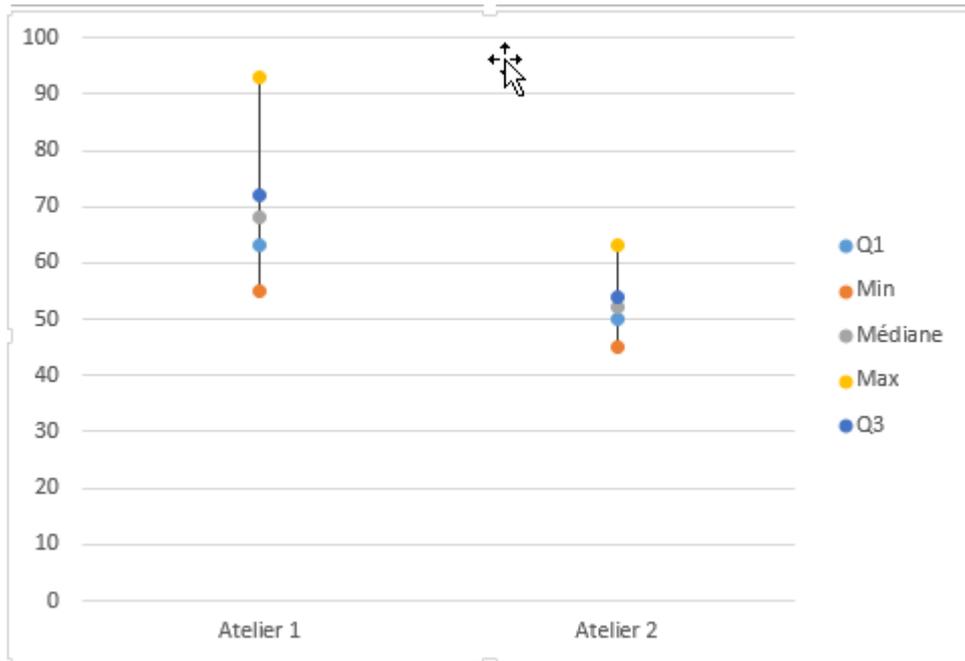
Une fois ceci fait, nous devons enlever les ligens reliant chaque couple de points via un clic droit et **Mettre en forme une série de données**:



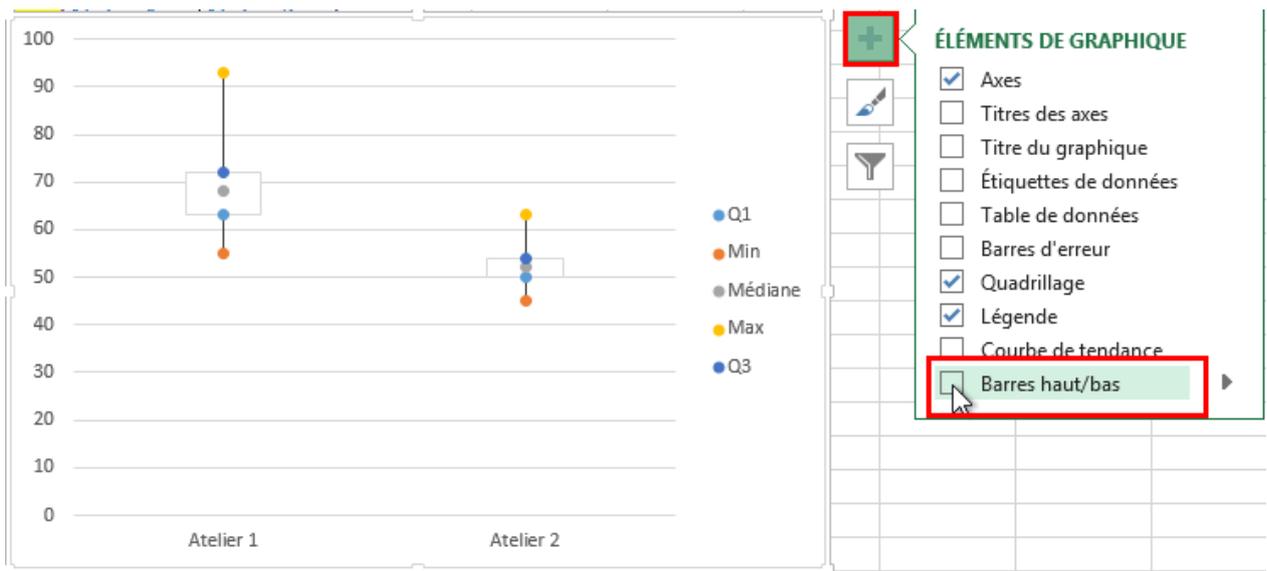
et ensuite dans al volet de mise en forme sélectionner **Aucun trait**:



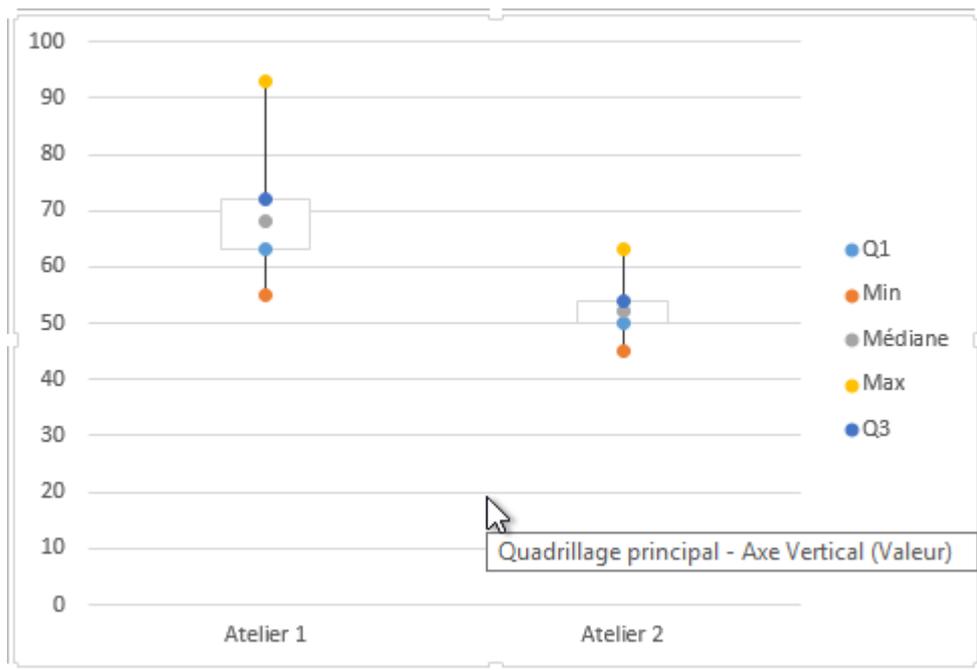
et répéter la manipulation pour chaque couple de points ce qui donne à la fin:



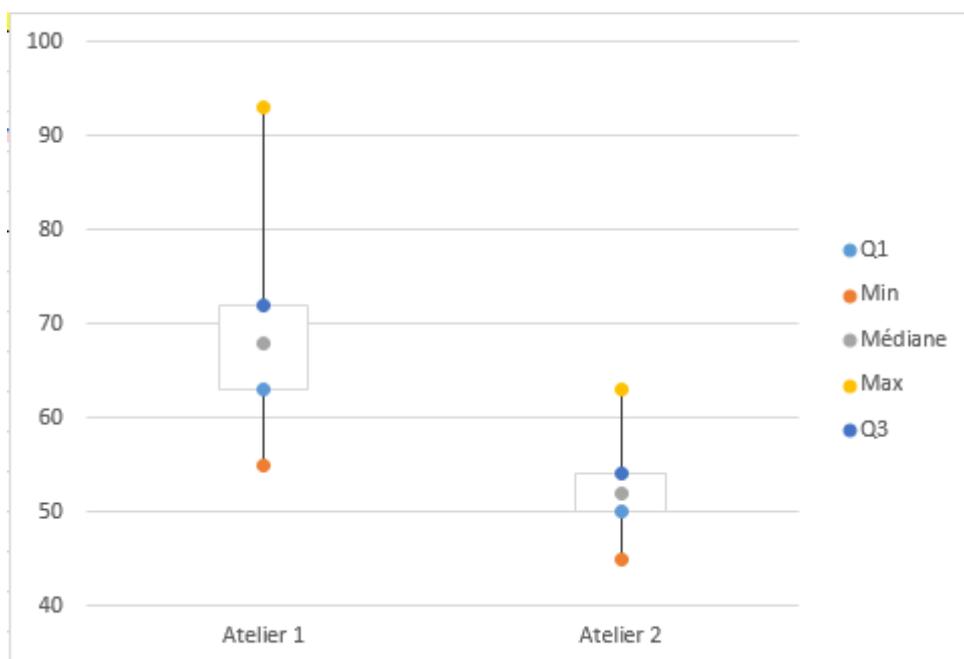
Une fois ceci, fait, vous allez:



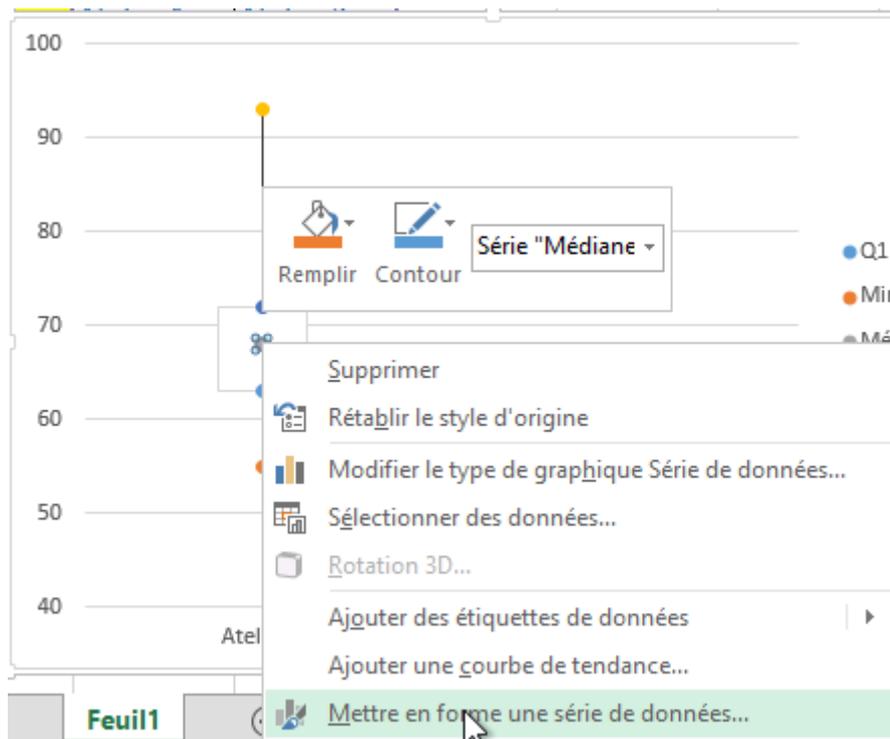
Ce qui donne donc:



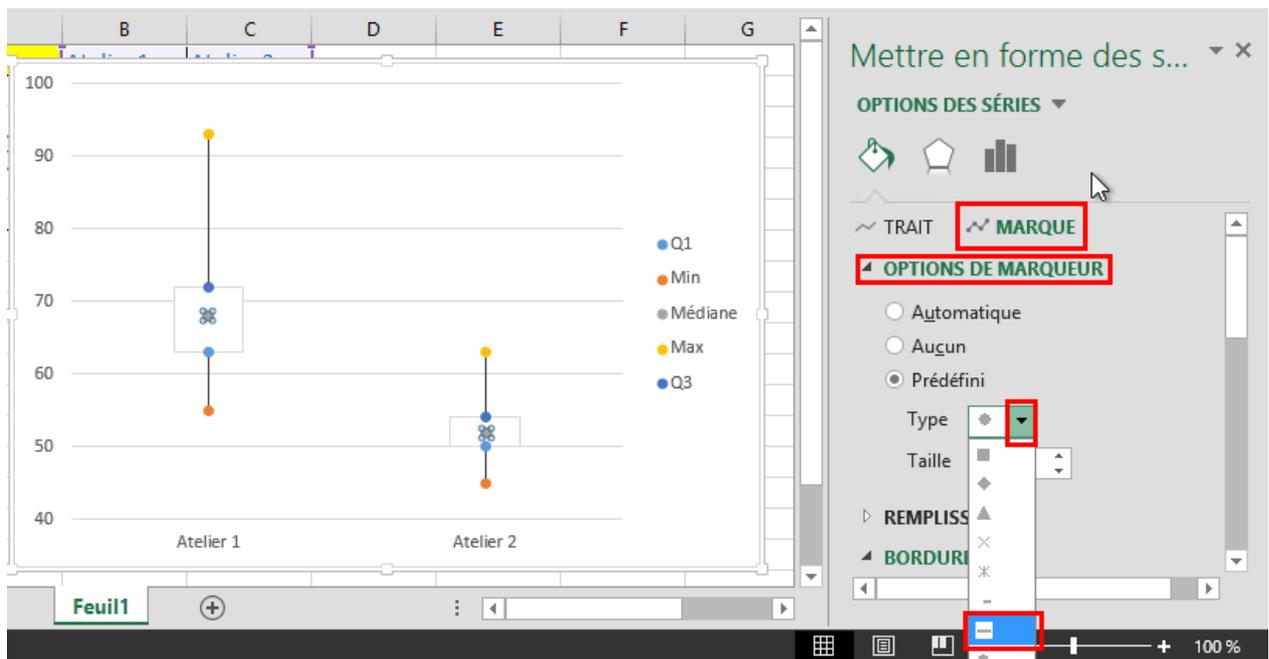
Nous allons changer la valeur minimale de l'axe Y afin de voir mieux les box-plot et pour pouvoir poursuivre:



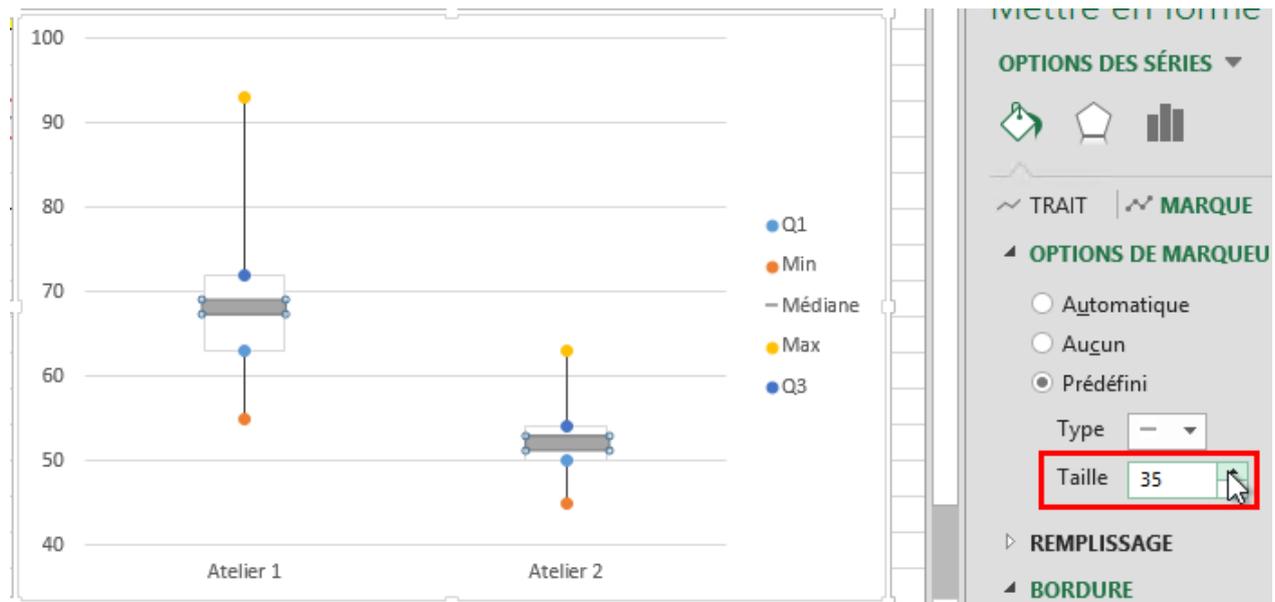
Il nous reste à modifier le point de la médiane en une ligne (en faisant au mieux car ce tableur est très limité à ce niveau là...). Pour cela nous faisons un clic droit sur l'un des points médians et nous cliquons sur **Mettre en forme une série de données**:



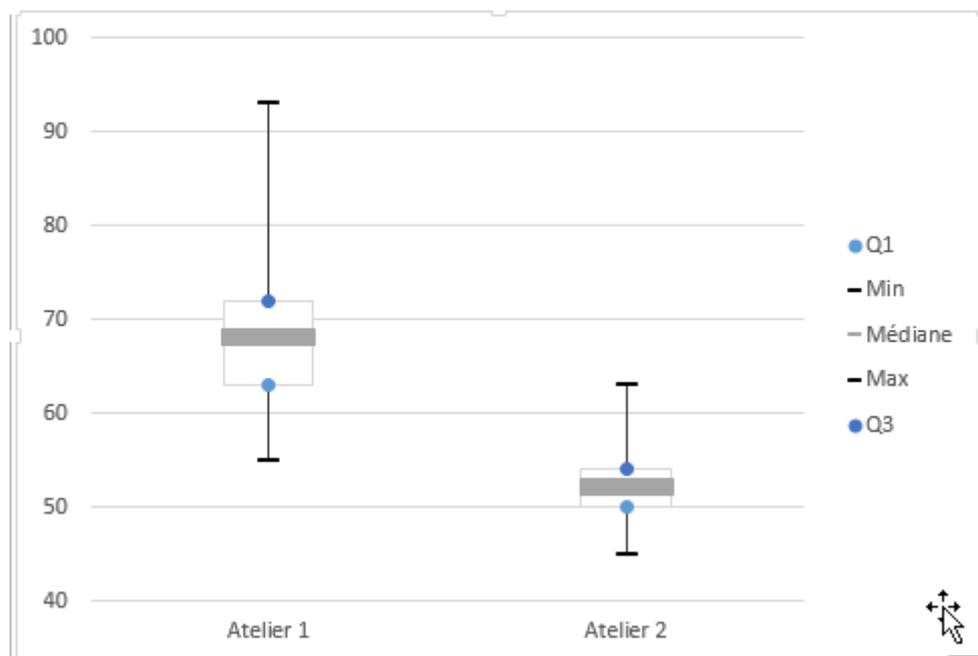
Ensuite, nous prenons:



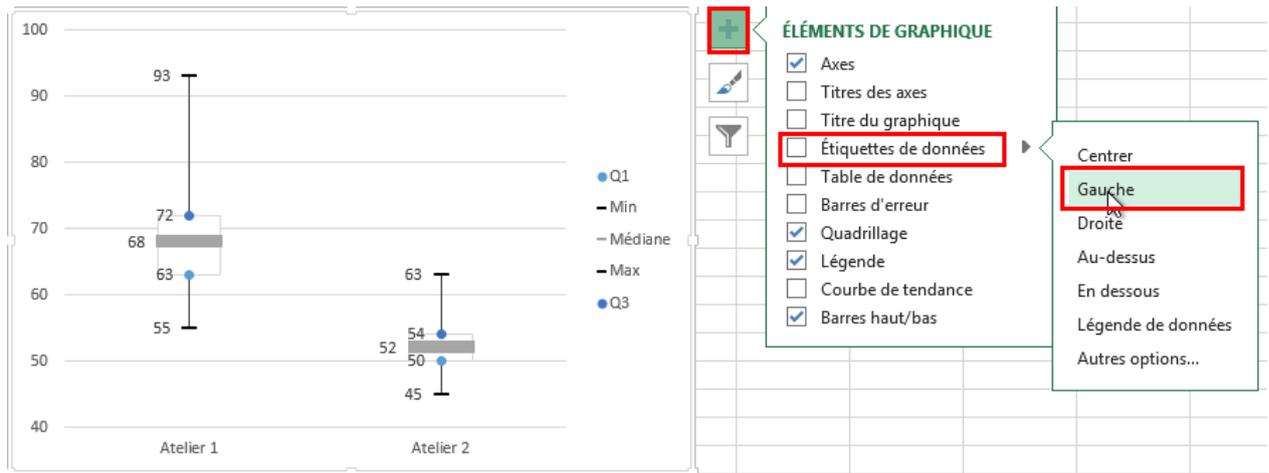
Et nous changeons la **Taille** à **35**:



Nous pouvons aussi changer le style des points extrêmes de la même façon ainsi que leur couleur pour obtenir:



Et enfin rajouter les étiquettes de données:



EXERCICE 6.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Centiles et rang

Énoncé:

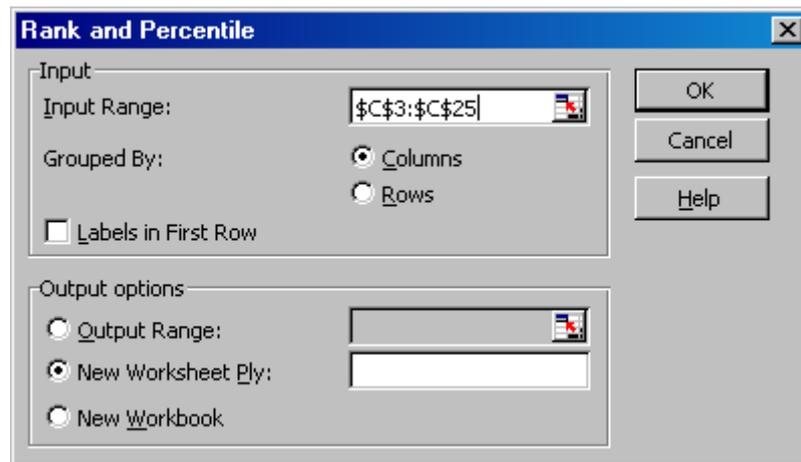
Soit les données suivantes de défauts:

	C
1	Défauts
2	65
3	68
4	72
5	55
6	64
7	70
8	66
9	74
10	75
11	70
12	93
13	64
14	60
15	62
16	70
17	80
18	61
19	60
20	62
21	68
22	70
23	72
24	71

Créez un tableau des centiles cumulés avec le rang des données dans MS Excel et déterminez la valeur modale.

Solution:

Il suffit d'aller dans le menu *Outils/Utilitaire d'analyse* et de sélectionner l'outil *Rang et centiles*:



et de valider par OK. Vient alors:

	A	B	C	D
1	<i>Point</i>	<i>Défauts</i>	<i>Rank</i>	<i>Percent</i>
2	11	93	1	100.00%
3	16	80	2	95.40%
4	9	75	3	90.90%
5	8	74	4	86.30%
6	3	72	5	77.20%
7	22	72	5	77.20%
8	23	71	7	72.70%
9	6	70	8	54.50%
10	10	70	8	54.50%
11	15	70	8	54.50%
12	21	70	8	54.50%
13	2	68	12	45.40%
14	20	68	12	45.40%
15	7	66	14	40.90%
16	1	65	15	36.30%
17	5	64	16	27.20%
18	12	64	16	27.20%
19	14	62	18	18.10%
20	19	62	18	18.10%
21	17	61	20	13.60%
22	13	60	21	4.50%
23	18	60	21	4.50%
24	4	55	23	0.00%

Les colonnes *A*, *B*, *C* sont évidents à comprendre. Ce qui va nous intéresser c'est de montrer la provenance des valeurs de la colonne *D*. Voici comment les obtenir simplement de deux manières différentes (c'est logique):

A	B	C	D	E	F	G
Point	Défauts	Rank	Percent	Via une fonction	A la main	
11	93	1	1	=PERCENTRANK(\$B\$2:\$B\$24;B2;5)	=F3+1/(\$C\$24-1)	
16	80	2	0.954	=PERCENTRANK(\$B\$2:\$B\$24;B3;5)	=F4+1/(\$C\$24-1)	
9	75	3	0.909	=PERCENTRANK(\$B\$2:\$B\$24;B4;5)	=F5+1/(\$C\$24-1)	
8	74	4	0.863	=PERCENTRANK(\$B\$2:\$B\$24;B5;5)	=F6+2/(\$C\$24-1)	
3	72	5	0.772	=PERCENTRANK(\$B\$2:\$B\$24;B6;5)	=F8+1/(\$C\$24-1)	
22	72	5	0.772	=PERCENTRANK(\$B\$2:\$B\$24;B7;5)	=F8+1/(\$C\$24-1)	
23	71	7	0.727	=PERCENTRANK(\$B\$2:\$B\$24;B8;5)	=F12+4/(\$C\$24-1)	
6	70	8	0.545	=PERCENTRANK(\$B\$2:\$B\$24;B9;5)	=F13+2/(\$C\$24-1)	
10	70	8	0.545	=PERCENTRANK(\$B\$2:\$B\$24;B10;5)	=F13+2/(\$C\$24-1)	
15	70	8	0.545	=PERCENTRANK(\$B\$2:\$B\$24;B11;5)	=F13+2/(\$C\$24-1)	
21	70	8	0.545	=PERCENTRANK(\$B\$2:\$B\$24;B12;5)	=F13+2/(\$C\$24-1)	
2	68	12	0.454	=PERCENTRANK(\$B\$2:\$B\$24;B13;5)	=F15+1/(\$C\$24-1)	Médiane
20	68	12	0.454	=PERCENTRANK(\$B\$2:\$B\$24;B14;5)	=F15+1/(\$C\$24-1)	
7	66	14	0.409	=PERCENTRANK(\$B\$2:\$B\$24;B15;5)	=F16+1/(\$C\$24-1)	
1	65	15	0.363	=PERCENTRANK(\$B\$2:\$B\$24;B16;5)	=F17+2/(\$C\$24-1)	
5	64	16	0.272	=PERCENTRANK(\$B\$2:\$B\$24;B17;5)	=F19+2/(\$C\$24-1)	
12	64	16	0.272	=PERCENTRANK(\$B\$2:\$B\$24;B18;5)	=F19+2/(\$C\$24-1)	
14	62	18	0.181	=PERCENTRANK(\$B\$2:\$B\$24;B19;5)	=F21+1/(\$C\$24-1)	
19	62	18	0.181	=PERCENTRANK(\$B\$2:\$B\$24;B20;5)	=F21+1/(\$C\$24-1)	
17	61	20	0.136	=PERCENTRANK(\$B\$2:\$B\$24;B21;5)	=F23+2/(\$C\$24-1)	
13	60	21	0.045	=PERCENTRANK(\$B\$2:\$B\$24;B22;5)	=F24+1/(\$C\$24-1)	
18	60	21	0.045	=PERCENTRANK(\$B\$2:\$B\$24;B23;5)	=F24+1/(\$C\$24-1)	
4	55	23	0	=PERCENTRANK(\$B\$2:\$B\$24;B24;5)	0	
		=COUNT(C2:C24)				

Ce qui donne bien le résultat généré avec l'utilitaire d'analyse mais avec plus de précisions:

	A	B	C	D	E	F	G
1	Point	Défauts	Rank	Percent	Via une fonction	A la main	
2	11	93	1	100.00%	100.000%	100.000%	
3	16	80	2	95.40%	95.454%	95.455%	
4	9	75	3	90.90%	90.909%	90.909%	
5	8	74	4	86.30%	86.363%	86.364%	
6	3	72	5	77.20%	77.272%	77.273%	
7	22	72	5	77.20%	77.272%	77.273%	
8	23	71	7	72.70%	72.727%	72.727%	
9	6	70	8	54.50%	54.545%	54.545%	
10	10	70	8	54.50%	54.545%	54.545%	
11	15	70	8	54.50%	54.545%	54.545%	
12	21	70	8	54.50%	54.545%	54.545%	
13	2	68	12	45.40%	45.454%	45.455%	Médiane
14	20	68	12	45.40%	45.454%	45.455%	
15	7	66	14	40.90%	40.909%	40.909%	
16	1	65	15	36.30%	36.363%	36.364%	
17	5	64	16	27.20%	27.272%	27.273%	
18	12	64	16	27.20%	27.272%	27.273%	
19	14	62	18	18.10%	18.181%	18.182%	
20	19	62	18	18.10%	18.181%	18.182%	
21	17	61	20	13.60%	13.636%	13.636%	
22	13	60	21	4.50%	4.545%	4.545%	
23	18	60	21	4.50%	4.545%	4.545%	
24	4	55	23	0.00%	0.00%	0.00%	
25			23				

Calculons maintenant la valeur Modale avec MS Excel:

=MODE(C2:C24)=70

Mais attention! Cette fonction de MS Excel renvoie, contrairement à ce qu'on ferait en statistiques, la plus petite valeur modale (si elle est à égalité avec une ou plusieurs valeurs ayant le même nombre d'occurrences). Pour s'assurer qu'elle renvoie la plus grande valeur, il faut trier la liste de données dans l'ordre décroissant.

EXERCICE 7.

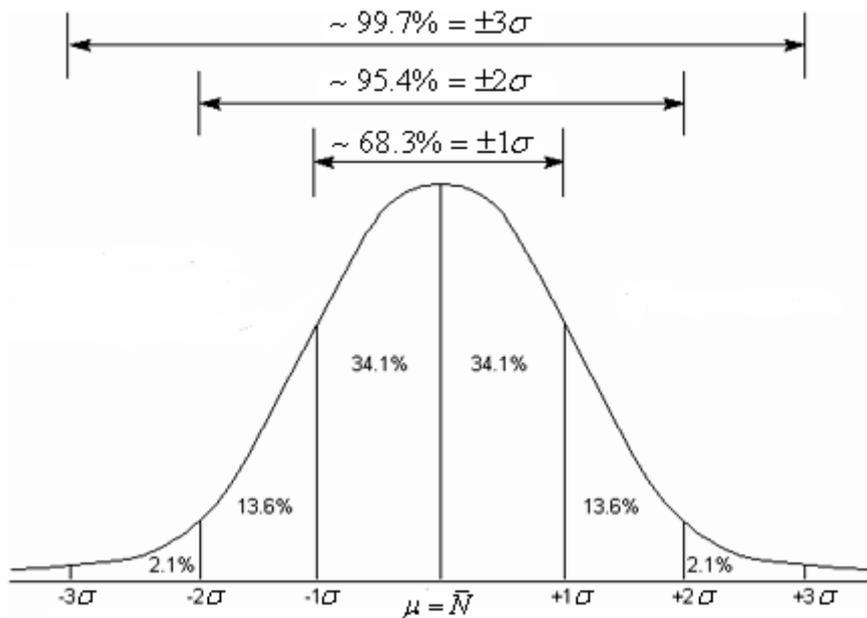
Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Loi multinomiale

Énoncé:

Dans une entreprise l'ensemble de la production a des variations de côtes suivant des lois Normales indépendantes et identiquement distribuées dont l'étendue dans l'acceptable (USL, LSL) est à un niveau de ± 3 sigma selon les % suivants:



[0,1] sigma	[1,2] sigma	[2,3] sigma	[3, et plus sigma]
68.3%	27.1%	4.3%	0.3%

Nous souhaiterons savoir sur un ensemble de 100 pièces commandées par un client qu'elle est la probabilité qu'il y en ait exactement 90 compris dans le premier intervalle, 5 dans le deuxième intervalle, 4 dans le troisième et 1 au-delà?

Solution:

Il nous faut utiliser la loi multinomiale:

$$P = \frac{n!}{\prod_{i=1}^m k_i!} p_i^{k_i} = \frac{100!}{90!5!4!1!} 0.683^{90} \cdot 0.271^5 \cdot 0.043^4 \cdot 0.003^1 \cong 4 \cdot 10^{-10}$$

Évidemment cette probabilité est extrêmement faible. Dans l'idéal, il faudrait calculer la probabilité cumulée mais c'est très long à calculer!

Il est facile d'obtenir ce résultat dans MS Excel avec deux fonctions:

	A	B	C
1	Nb	Prob.	p^k
2	90	68.30%	1.25275E-15
3	5	27.10%	0.00146166
4	4	4.30%	3.4188E-06
5	1	0.30%	0.003
6			
7		Prob:	4.09619E-10

De manière détaillée:

	A	B	C
1	Nb	Prob.	p^k
2	90	0.683	=B2^A2
3	5	0.271	=B3^A3
4	4	0.043	=B4^A4
5	1	0.003	=B5^A5
6			
7		Prob:	=MULTINOMIAL(A2:A5)*PRODUCT(C2:C5)

EXERCICE 8.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Tirage exhaustif

Énoncé:

Dans une petite production unique d'un lot de 1'000 pièces dont nous savons que 30% en moyenne sont mauvaises à cause de la complexité des pièces par retour d'expérience d'une fabrication précédente similaire.

E1. Nous savons qu'un client va en tirer 20 au hasard pour décider d'accepter ou de rejeter le lot. Il ne rejettera pas le lot s'il trouve zéro pièce défectueuse parmi ces 20. Quelle est la probabilité d'en avoir exactement 0 de défectueuse?

E2. Nous savons qu'un autre client va en tirer 20 au hasard pour décider d'accepter ou de rejeter le lot. Il ne rejettera pas le lot s'il trouve une quantité inférieure ou égale à 3 pièces défectueuses parmi ces 20. Quelle est la probabilité cumulée d'en trouver au moins ou égal à 3?

Solution:

S1. D'abord, il s'agit d'appliquer la loi hypergéométrique (tirage exhaustif, dont l'ordre n'est pas pris en compte)

Nous allons prendre pour ce calcul MS Excel:

$$=LOI.HYPERGEOMETRIQUE (0;20;300;1000)=0.073\%$$

et comme on exige un tirage nul nous avons vu dans le chapitre de Statistique que cela équivaut à calculer:

$$\frac{700}{1000} \cdot \frac{699}{999} \cdot \frac{698}{998} \cdot \frac{697}{997} \dots \cdot \frac{681}{981} = 0.073\%$$

Donc malheureusement c'est très mauvais pour le fournisseur car en d'autres termes, le client à plus de 99% de probabilité de tomber sur une ou plusieurs pièces défectueuses!

S2. Donc MS Excel 2007 et antérieur:

$$=LOI.HYPERGEOMETRIQUE (0;20;300;1000)+ LOI.HYPERGEOMETRIQUE (1;20;300;1000)+ LOI.HYPERGEOMETRIQUE (2;20;300;1000)+ LOI.HYPERGEOMETRIQUE (3;20;300;1000)=10.47\%$$

ou avec MS Excel 2010 et ultérieur:

$$=LOI.HYPERGEOMETRIQUE.N(3;20;300;1000;1)=10.47\%$$

Ce qui reste médiocre car cela signifie que cet autre client a en gros ~90% de probabilité cumulée de tomber sur plus de 3 pièces défectueuses.

EXERCICE 9.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Tirage exhaustif

Énoncé:

On souhaite mettre en production un petit développement informatique de 10'000 lignes de code. Le REX (Retour d'Expérience) montre que la probabilité de défaillance est de 1 bug pour 1'000 lignes de code (soit: 0.1% de 10'000 lignes).

1. On teste environ 15% des fonctions du logiciel au hasard avant l'envoi au client (soit l'équivalent de 1'500 lignes de code << 10'000). On demande quelle est la probabilité d'observer 1 bug et de comparer la valeur donnée par la loi binomiale et hypergéométrique.

2. On teste environ 50% des fonctions du logiciel au hasard avant l'envoi au client (soit l'équivalent de 5'000 lignes de code). On demande quelle est la probabilité d'observer 5 bugs et de comparer la valeur donnée par la loi binomiale et hypergéométrique.

Solution:

S1. Nous calculons d'abord avec la loi binomiale:

$$= \text{BINOMDIST}(1; 1500; 1/1000; 0) = 33.47\%$$

et ensuite avec la loi hypergéométrique:

$$= \text{HYPGEOMDIST}(1; 1500; 0.1\% * 10000; 10000) = 34.75\%$$

Nous voyons donc bien que conformément à la théorie, tant que le tirage n'est pas exhaustif les deux valeurs sont proches (approximation valable lorsque l'échantillon est environ 10 fois plus petit que la population).

S2. Nous calculons d'abord avec la loi binomiale:

$$= \text{BINOMDIST}(5; 5000; 1/1000; 0) = 17.55\%$$

et ensuite avec la loi hypergéométrique:

$$= \text{HYPGEOMDIST}(5; 5000; 10; 10000) = 24.62\%$$

Nous voyons donc bien que conformément à la théorie, quand le tirage est exhaustif les deux valeurs diffèrent.

EXERCICE 10.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Ecart-Moyen et écart-type

Énoncé:

Soit le tableau de production ci-dessous:

	A	B
1	Jour	Production
2	1	700
3	2	850
4	3	600
5	4	575
6	5	450
7	6	900
8	7	300
9	Total	4375
10	Moyenne	625

Déterminer avec MS Excel l'écart-moyen:

$$EM = \frac{\sum_{i=1}^N |x_i - \mu|}{N}$$

et l'estimateur de maximum de vraisemblance:

$$\sigma = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n-1}}$$

et donner leur interprétation.

Solution:

S1. Pour l'écart-moyen nous avons:

	A	B	C
1	Jour	Production	$ X-\mu $
2	1	700	75
3	2	850	225
4	3	600	25
5	4	575	50
6	5	450	175
7	6	900	275
8	7	300	325
9	Total	4375	1150
10	Moyenne	625	164.29

qui peut-être obtenu avec:

$$=ECART.MOYEN(B2:B8)=164.29$$

Soit en d'autres mots, en moyenne 164.29 éléments produits ont dévié de la moyenne chaque jour durant la semaine.

S2. L'estimateur de maximum de vraisemblance (estimateur sans biais) peut lui être obtenu par:

$$=ECARTYPE(B2:B8)=212.62$$

en termes statistiques nous avons alors que si les données sont distribuées normalement et que les erreurs sont dues à la somme d'une multitude de petits effets (loi normale) alors ~68% des valeurs de production se trouveront à ± 212.62 autour de la moyenne.

EXERCICE 11.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: *Qualité*

Énoncé:

Les tolérances sur les valeurs d'une résistance sont en 10^{-3} [Ω] pour la tolérance inférieure:

$$T_i = LSL = 420$$

et pour la tolérance supérieure:

$$T_s = USL = 720$$

Ces résistances ont une valeur qui suit une loi Normale de moyenne:

$$\mu = 600 \cdot 10^{-3} \text{ [Ω]}$$

et d'écart-type:

$$\sigma = 120 \cdot 10^{-3} \text{ [Ω]}$$

Quel est le pourcentage de bonnes résistances? Soit en d'autres termes, la probabilité d'être entre 420 et 720?

Solution:

D'abord, il s'agit d'appliquer la loi de Gauss.

Nous allons prendre pour ce calcul MS Excel:

$$=NORMDIST(720;600;120;1)-NORMDIST(420;600;120;1)=77.45\%$$

EXERCICE 12.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Cartes de contrôles de type *P*

Énoncé:

Soit le tableau suivant dans MS Excel:

	A	B	C
2			
3	Echantillon	Nombre de non conformités	Taille de l'échantillon
4	1	5	50
5	2	2	50
6	3	4	50
7	4	6	50
8	5	7	50
9	6	1	50
10	7	5	50
11	8	4	50
12	9	3	50
13	10	3	50
14	11	4	50
15	12	5	50
16	13	3	50
17	14	4	50
18	15	5	50
19	16	10	50
20	17	11	50
21	18	12	50
22	19	10	50
23	20	5	50
24	21	4	50
25	22	6	50
26	23	3	50
27	24	2	50
28	25	5	50
29	Total	129	1250

1. Calculez pour chaque ligne la proportion P en % du nombre d'éléments non-conformes par rapport aux conformes et ce aussi pour la ligne *Total*.
2. Calculez l'USL et la LSL à $\pm 3\sigma$ en faisant l'hypothèse que les non-conformités soient distribuées selon une loi binomiale.
3. Créez un graphique représentant la carte de contrôle.

4. Proposez une analyse graphique des données du tableau permettant de faire de l'inférence statistique avec un choix de classes qui vous semble le mieux adapté à la situation.

Solutions:

1. Les données demandées sont:

Echantillon	Nombre de non conformités	P
1	5	10.00%
2	2	4.00%
3	4	8.00%
4	6	12.00%
5	7	14.00%
6	1	2.00%
7	5	10.00%
8	4	8.00%
9	3	6.00%
10	3	6.00%
11	4	8.00%
12	5	10.00%
13	3	6.00%
14	4	8.00%
15	5	10.00%
16	10	20.00%
17	11	22.00%
18	12	24.00%
19	10	20.00%
20	5	10.00%
21	4	8.00%
22	6	12.00%
23	3	6.00%
24	2	4.00%
25	5	10.00%
Total	129	10.32%

où chaque échantillon est constitué de 50 éléments.

2. Si nous considérons que la variable aléatoire X est le total des non-conformités de tous les échantillons d'une période donnée alors nous avons:

$$E\left(\frac{X}{n}\right) = \frac{E(X)}{n}$$

avec n étant égal à $25 \cdot 50 = 1250$.

Si nous considérons notre variable aléatoire comme suivant une loi binomiale nous avons:

$$\frac{E(X)}{n} = \frac{nP}{n} = P$$

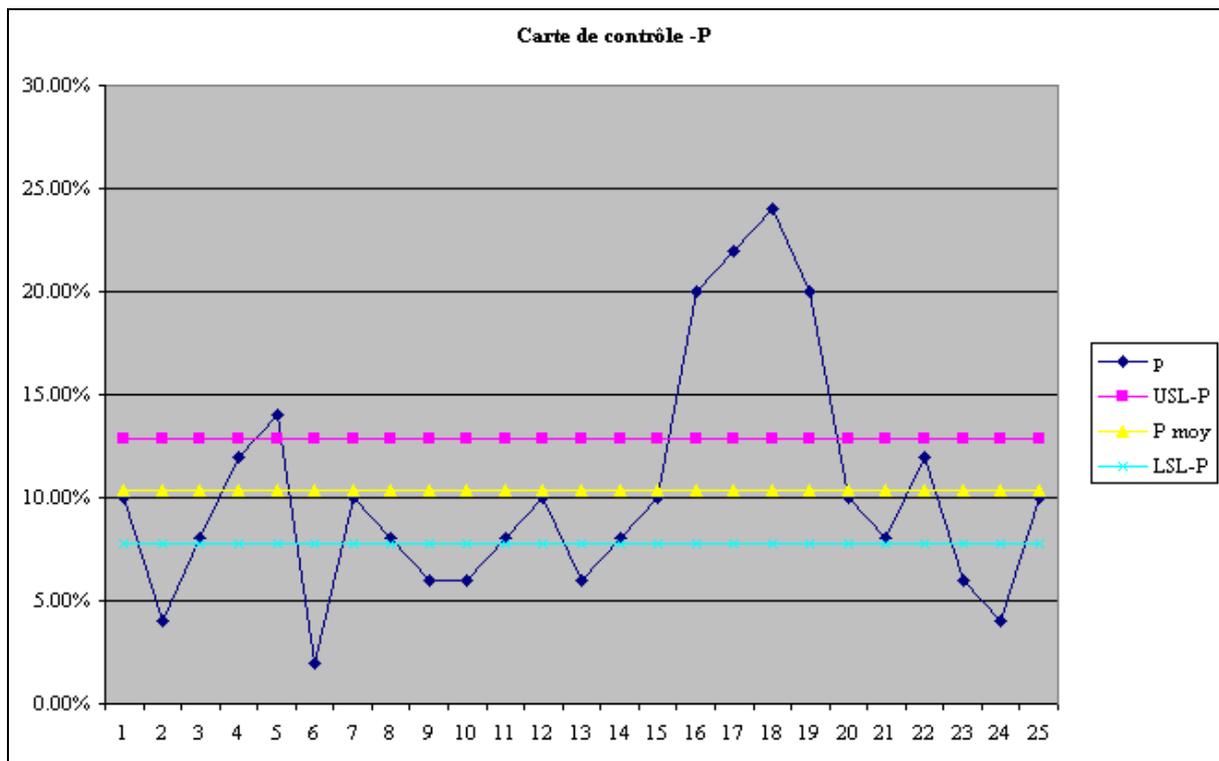
De même pour la variance:

$$V\left(\frac{X}{n}\right) = \frac{1}{n^2} V(X) = \frac{1}{n^2} nP(1-P) = \frac{P(1-P)}{n} \Rightarrow \sigma = \sqrt{\frac{P(1-P)}{n}}$$

Nous avons alors:

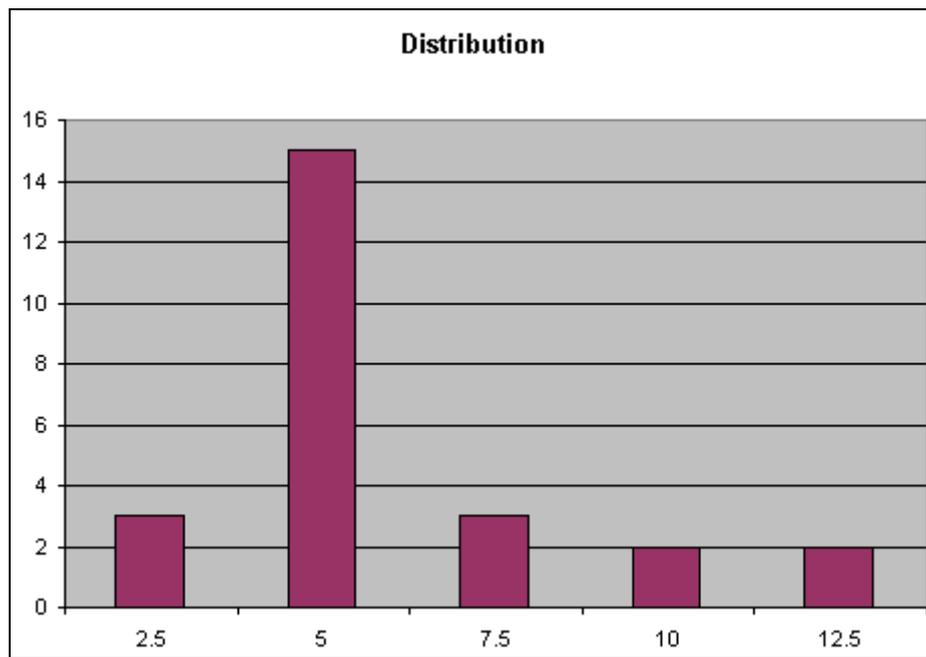
p moyen	10.32%
sigma	0.86%
USL	12.90%
LSL	7.74%

3. Voici le graphique:



4. Le graphique avec les classes correspondantes choisies au mieux:

Groupement par non-conformités	Quantité
2.5	3
5	15
7.5	3
10	2
12.5	2



EXERCICE 13.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Cartes de contrôles de type G (officiellement \bar{X})

Énoncé:

Soit le tableau suivant dans MS Excel:

	A	B	C
2			
3	Echantillon	Nombre de non conformités	Taille de l'échantillon
4	1	5	50
5	2	2	50
6	3	4	50
7	4	6	50
8	5	7	50
9	6	1	50
10	7	5	50
11	8	4	50
12	9	3	50
13	10	3	50
14	11	4	50
15	12	5	50
16	13	3	50
17	14	4	50
18	15	5	50
19	16	10	50
20	17	11	50
21	18	12	50
22	19	10	50
23	20	5	50
24	21	4	50
25	22	6	50
26	23	3	50
27	24	2	50
28	25	5	50
29	Total	129	1250

1. Calculez pour chaque ligne la proportion P en % du nombre d'éléments non-conformes par rapport aux conformes et ce aussi pour la ligne *Total*.
2. Calculez l'USL et la LSL à $\pm 3\sigma$ en faisant l'hypothèse que les non-conformités soient Normalement distribuées (selon une Gaussienne).
3. Créez un graphique représentant la carte de contrôle.

Solutions:

1. Les données demandées restent:

Echantillon	Nombre de non conformités	P
1	5	10.00%
2	2	4.00%
3	4	8.00%
4	6	12.00%
5	7	14.00%
6	1	2.00%
7	5	10.00%
8	4	8.00%
9	3	6.00%
10	3	6.00%
11	4	8.00%
12	5	10.00%
13	3	6.00%
14	4	8.00%
15	5	10.00%
16	10	20.00%
17	11	22.00%
18	12	24.00%
19	10	20.00%
20	5	10.00%
21	4	8.00%
22	6	12.00%
23	3	6.00%
24	2	4.00%
25	5	10.00%
Total	129	10.32%

où chaque échantillon est constitué de 50 éléments.

2. Si nous considérons le cas Gaussien nous avons démontré que:

$$\mu_{\bar{X}} = \mu$$

et pour la variance:

$$\sigma_{\bar{X}} = \frac{\sigma}{\sqrt{n}}$$

où σ sera pris comme étant l'estimateur de maximum de vraisemblance.

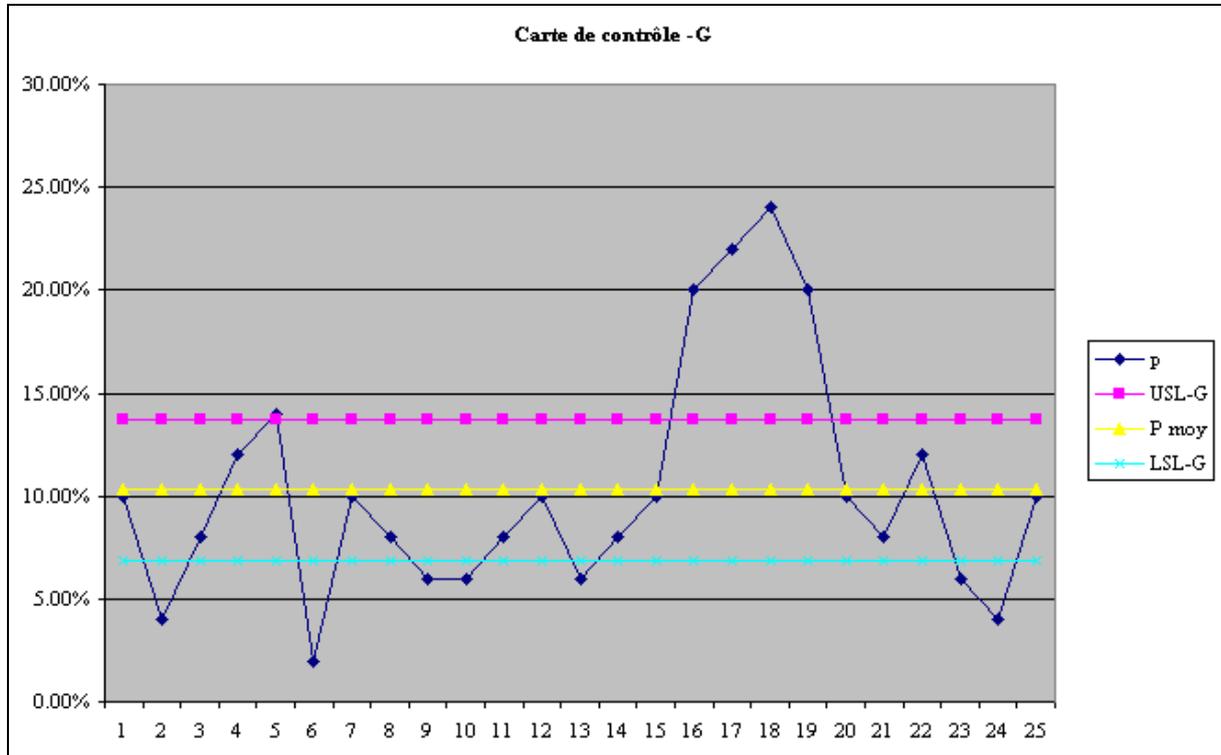
Nous avons alors (à comparer avec le précédent!):

p moyen	10.32%
sigma	1.141%
USL	13.74%
LSL	6.90%

avec l'écart-type étant l'écart-type de la moyenne obtenu donc par:

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \text{ECARTTYPE}(\dots) / \text{RACINE}(\text{COMPTE}(\dots))$$

3. Voici le graphique:



EXERCICE 14.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Cartes de contrôles de type NP

Énoncé:

Soit le tableau suivant dans MS Excel:

	A	B	C
2			
3	Echantillon	Nombre de non conformités	Taille de l'échantillon
4	1	5	50
5	2	2	50
6	3	4	50
7	4	6	50
8	5	7	50
9	6	1	50
10	7	5	50
11	8	4	50
12	9	3	50
13	10	3	50
14	11	4	50
15	12	5	50
16	13	3	50
17	14	4	50
18	15	5	50
19	16	10	50
20	17	11	50
21	18	12	50
22	19	10	50
23	20	5	50
24	21	4	50
25	22	6	50
26	23	3	50
27	24	2	50
28	25	5	50
29	Total	129	1250

1. Calculez l'USL et la LSL à $\pm 3\sigma$ en faisant l'hypothèse que les non-conformités soient distribuées selon une loi binomiale.
2. Créez un graphique représentant la carte de contrôle.

Solutions:

1. Les données nécessaires restent toujours:

Echantillon	Nombre de non conformités	P
1	5	10.00%
2	2	4.00%
3	4	8.00%
4	6	12.00%
5	7	14.00%
6	1	2.00%
7	5	10.00%
8	4	8.00%
9	3	6.00%
10	3	6.00%
11	4	8.00%
12	5	10.00%
13	3	6.00%
14	4	8.00%
15	5	10.00%
16	10	20.00%
17	11	22.00%
18	12	24.00%
19	10	20.00%
20	5	10.00%
21	4	8.00%
22	6	12.00%
23	3	6.00%
24	2	4.00%
25	5	10.00%
Total	129	10.32%

où chaque échantillon est constitué de 50 éléments.

Si nous considérons notre variable aléatoire comme suivant une loi binomiale nous avons:

$$E(X) = nP = 50 \cdot 25 \cdot 10.32\% = 129$$

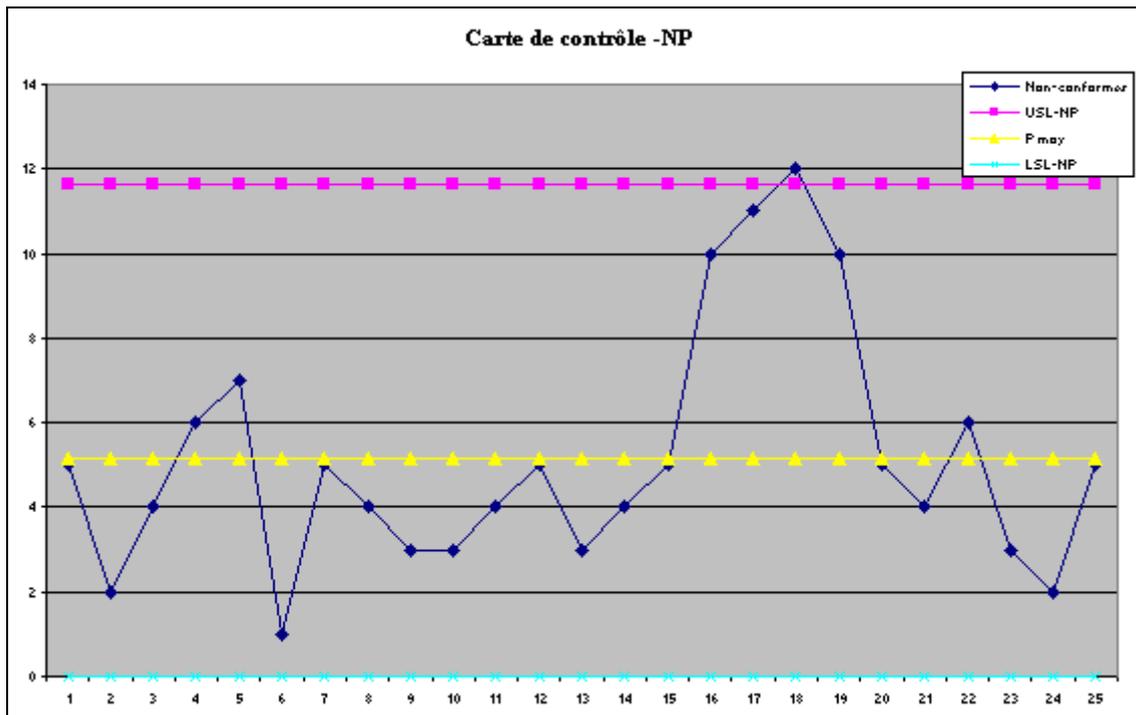
et pour la variance:

$$V(X) = nP(1-P) \Rightarrow \sigma = \sqrt{nP(1-P)}$$

Nous avons alors:

p moyen	5.16
sigma	2.15
USL	11.61
LSL	0.00

3. Voici le graphique:



Ce qui est médiocre mais qui s'explique par le fait que l'estimateur de maximum de vraisemblance de la loi binomiale est comme nous l'avons démontré:

$$\hat{p} = \frac{k}{N}$$

ce qui correspondant à l'analyse faite par la carte de contrôle de type $-P$. Donc l'usage d'une carte de contrôle de type NP est alors erronée alors que la loi Normal même sur les quantités plutôt que sur les proportions reste meilleure.

EXERCICE 15.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Cartes de contrôles de type C

Énoncé:

Soit le tableau suivant dans MS Excel (qui suffit):

	A	B
2		
3	Echantillon	Nombre de non conformités
4	1	5
5	2	2
6	3	4
7	4	6
8	5	7
9	6	1
10	7	5
11	8	4
12	9	3
13	10	3
14	11	4
15	12	5
16	13	3
17	14	4
18	15	5
19	16	10
20	17	11
21	18	12
22	19	10
23	20	5
24	21	4
25	22	6
26	23	3
27	24	2
28	25	5
29	Total	129

1. Calculez l'USL et la LSL à $\pm 3\sigma$ en faisant l'hypothèse que les non-conformités soient distribuées selon une loi de Poisson.
2. Créez un graphique représentant la carte de contrôle.

Solutions:

1. Si nous considérons notre variable aléatoire comme suivant une loi de Poisson nous avons selon notre calcul de l'estimateur de maximum de vraisemblance de la moyenne:

$$E(X) = \frac{1}{n} \sum_{i=1}^n x_i = 5.16$$

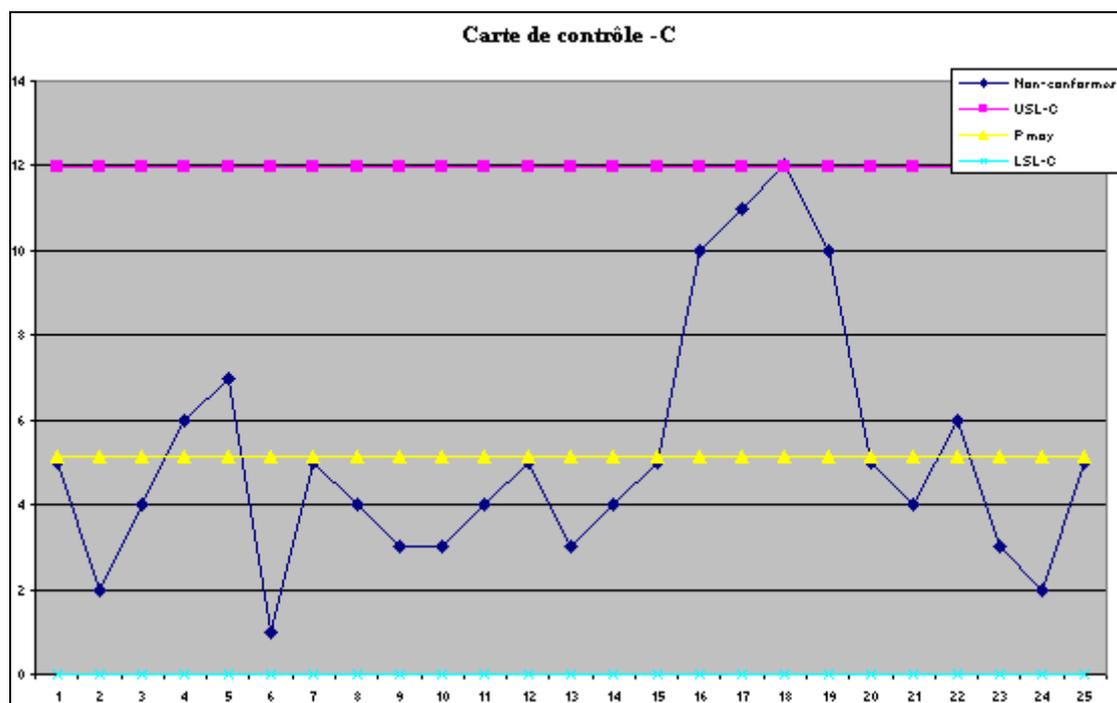
et pour la variance:

$$V(X) = E(X) \Rightarrow \sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n x_i}$$

Nous avons alors:

Moyenne	5.16
sigma	2.27
USL	11.97
LSL	0.00

2. Voici le graphique:



Nous voyons que le résultat n'est pas bon ce qui est normal puisque la loi binomiale ne peut être approximée par une loi de Poisson que si la probabilité est tend vers zéro. Or 10% ce n'est pas vraiment cela...

Par contre comme nous l'avons démontré en cours, elle peut bien être approximée par une loi Normale qui n'impose pas cette hypothèse.

EXERCICE 16.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Échantillons

Énoncé:

Si 5% de la production annuelle d'une entreprise est défectueuse, qu'elle est la probabilité qu'en prenant un échantillon de 75 pièces de la ligne de production que 2% ou moins soit défectueux?

Solution:

Nous savons que:

$$Z = \frac{k - np}{\sqrt{npq}}$$

suit une loi Normale centrée réduite où np correspond à l'espérance de la loi binomiale et \sqrt{npq} à son écart-type.

Puisque nous travaillons dans cet exercice en % du lot total, nous travaillerons non pas avec la variable binomiale k mais avec:

$$k' = \frac{k}{n}$$

et de par la propriété de l'espérance, nous avons:

$$E\left(\frac{k}{n}\right) = \frac{1}{n}E(k) = \frac{1}{n}np = p$$

et de par la propriété de la variance nous avons:

$$V\left(\frac{k}{n}\right) = \sigma^2 = \frac{1}{n^2}V(k) = \frac{1}{n^2}npq = \frac{pq}{n}$$

Nous avons alors:

$$Z = \frac{k' - p}{\sqrt{\frac{pq}{n}}}$$

Nous avons dès lors:

$$Z = \frac{0.02 - 0.05}{\sqrt{\frac{0.05 \cdot 0.95}{75}}} \cong -1.19$$

La probabilité cumulée correspondante à cette valeur de la variable aléatoire est:

$$= \text{NORMSDIST}(-1.19) = 11.66\%$$

Nous avons donc une probabilité cumulée de 11.66% qu'en prenant 75 échantillons de la ligne de production que nous en ayons 2% qui soient défectueux.

Ce résultat peut parfois paraître déroutant à cause du fait que souvent les gens pensent que nous avons une probabilité de 100% d'avoir 2.5% d'éléments défectueux si l'échantillon correspond à toute la production de l'entreprise.

Or, quand $n \rightarrow +\infty$, nous avons k' qui tend vers 5%. Nous avons alors:

$$Z \cong 0$$

Nous avons alors:

$$= \text{NORMSDIST}(0) = 50\%$$

ce qui montre et rappelle que pour la loi Normale centrée réduite nous avons par construction 50% de probabilité cumulée que la moyenne soit de 5% et non 100% de probabilité cumulée.

Ce qui explique le résultat de 11.6% obtenu plus haut.

EXERCICE 17.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Défaillance

Énoncé:

Soit un matériel électronique réparable, nous connaissons le taux de défaillance $\lambda = 0.00001/h$. (soit une MTBF de 100'000 heures).

Trouver la probabilité de défaillance entre $t_1 = 200h$. et $t_2 = 300h$. sachant que celle-ci suit une loi exponentielle dans le cas le plus général qui soit.

Solution:

Nous savons que:

$$F(t) = 1 - \exp\left(-\int_0^t \lambda(t) dt\right) = 1 - e^{-\lambda t}$$

La fonction de distribution est alors:

$$f(t) = \frac{dF(t)}{dt} = \lambda e^{-\lambda t} \frac{n!}{r!(n-r)!}$$

Elle suit donc une loi exponentielle!

Nous avons alors la probabilité de défaillance:

$$P(t_1 \leq D \leq t_2) = \int_{t_1}^{t_2} \lambda e^{-\lambda t} dt = e^{-\lambda t_1} - e^{-\lambda t_2} = 0.001 = 0.1\%$$

Soit avec MS Excel:

$$=EXPONDIST(300;0.00001;1) - EXPONDIST(200;0.00001;1)=0.1\%$$

EXERCICE 18.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Défaillance

Énoncé:

La durée de vie de bielles d'une voiture suit un loi log-normale de paramètres : $\mu = 5$ et $\sigma = 1.4$.

Le taux de défaillance suit ainsi une loi du type:

$$R(t) = 1 - \frac{1}{\sigma\sqrt{2\pi}} \int_0^t \frac{1}{x} e^{-\frac{1}{2}\left(\frac{\ln(x)-\mu}{\sigma}\right)^2} dx$$

et rappelons que le taux de fiabilité est donné par:

$$F=1-R$$

- A. Calculez la fiabilité F et la défaillance R pour $T=300$ heures
- B. Calculez son MTBF, ou MUT et dites si c'est bon ou mauvais

Solution:

A. Nous allons calculer ceci avec MS Excel:

$$F=\text{LOGNORMDIST}(300;5;1.4)=0.69$$

et donc:

$$R=1-F=0.31$$

ce qui est mauvais.

B. Puisque nous savons que l'espérance est donnée par:

$$MTBF \cong MUT = E(X) = e^{\mu + \frac{\sigma^2}{2}} = 395.44h.$$

où nous supposons le temps de réparation négligeable car la durée de vie d'une bielle est beaucoup plus grande que le temps de réparation de celle-ci.

EXERCICE 19.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Défaillance

Énoncé:

Un composant électrique non réparable à un taux de défaillance constant de:

$$\lambda = 2 \cdot 10^{-6} / h.$$

Calculer son MTTF (Mean Time To Failure: donc il s'agit d'un élément non réparable) dans les 500 premières heures.

Solution:

Puisque son taux de défaillance est constant, nous devons utiliser la loi exponentielle dont l'espérance (MTTF) est donnée par:

$$MTTF = \frac{1}{\lambda} = 5 \cdot 10^5 h.$$

EXERCICE 20.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Défaillance

Énoncé:

Un processus de fabrication produit en moyenne 3.84 éléments défectueux par jour (processus constant/sous contrôle dont la moyenne est fixe même si le nombre d'essais est très grand). Quelle est la probabilité d'avoir 5 éléments défectueux dans un jour pris au hasard?

Solution:

Étant donné les informations il s'agit d'utiliser la loi de Poisson:

$$P(5) = \frac{\mu^k}{k!} e^{-\mu} = \frac{3.84^5}{5!} e^{-3.84}$$

Nous avons alors avec MS Excel:

$$=POISSON(5;3.84;0)= 14.95\%$$

EXERCICE 21.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Défaillance

Énoncé:

Sur une machine de production, on sait qu'il y a en moyenne 1 panne tous les 4 mois (20 jours ouvrés par mois)? Quelle est la probabilité cumulée qu'il y ait 0 panne lors d'une journée de démonstration:

Solution:

Etant donné les informations il s'agit d'utiliser la loi de Poisson.

Nous avons d'abord:

$$\mu = \frac{1}{4 \cdot 20} = 0.0125 \text{ [panne / jour]}$$

Donc il vient:

$$P(0) = \frac{\mu^k}{k!} e^{-\mu} = \frac{0.0125^0}{0!} e^{-0.0125}$$

Nous avons alors avec MS Excel:

$$=POISSON(0;0.0125;0)= 98.75\%$$

EXERCICE 22.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Défaillance

Énoncé:

La probabilité de trouver une erreur par un auditeur dans une chaîne de production est de 1%.

- A. Quelle est la probabilité que la première erreur soit trouvée à la 70^{ème} pièce vérifiée?
- B. Quelle est la probabilité pour que plus que 50 pièces doivent être auditées pour trouver une erreur.

Solution:

A. Il suffit d'appliquer la loi géométrique:

$$P(k) = q^{k-1} p = 0.01 \cdot 0.99^{70-1} = 0.4998\%$$

MS Excel n'intègre pas la distribution géométrique....

B. La probabilité est simplement donnée par:

$$P(k > 50) = 0.99^{50} = 60.5\%$$

EXERCICE 23.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Défaillance

Énoncé:

Supposons que le temps en mois entre l'arrêt de lignes de montage d'une production ait une valeur MTBF de 5 mois.

E1. Quelle est la constante λ de la loi exponentielle qui modélise le mieux le fonctionnement de l'atelier.

E2. Quelle est la probabilité que la ligne de production s'arrête à nouveau après 15 mois et après 5 mois de bon fonctionnement?

E3. Quelle est la probabilité que la ligne de fabrication tombe en panne dans moins de 20 mois.

E4. Quelle est la probabilité que la ligne de production tombe en panne entre le 10^{ème} et le 20^{ème} mois?

E5. Quelle est la valeur de σ, μ et trouvez la probabilité que le temps jusqu'à ce que la ligne de production tombe à nouveau en panne soit comprise entre $[\mu \pm 3\sigma]$

Solution:

S1. Nous avons:

$$MTBF = 5 = \frac{1}{\lambda} \Rightarrow \lambda = 0.2 / \text{mois}$$

S2. Puisque le système suit une loi exponentielle:

$$P(X < 15) = e^{-15\lambda} = 4.98\%$$

Soit dans MS Excel:

$$=1-EXPONDIST(15;0.2;1)=4.98\%$$

et pour les cinq mois:

$$=1-EXPONDIST(5;0.2;1)=36.7\%$$

S3. C'est simple:

$$P(X < 20) = 1 - P(X > 20) = 1 - e^{-20\lambda} = 98.17\%$$

Soit dans MS Excel:

$$=EXPONDIST(20;0.2;1)=98.17\%$$

S4. Il suffit de calculer:

$$\begin{aligned} P(10 < X < 15) &= P(X < 15) - P(X < 10) = (1 - e^{-15\lambda}) - (1 - e^{-10\lambda}) \\ &= 95.02\% - 86.46\% = 8.55\% \end{aligned}$$

Soit dans MS Excel:

$$=EXPONDIST(15;0.2;1)-EXPONDIST(10;0.2;1)=8.55\%$$

S5. Calcul l'écart-type et la moyenne d'une loi exponentielle est chose aisée:

$$\sigma = \mu = \frac{1}{\lambda} = 5$$

Donc:

$$\begin{aligned} \mu + 3\sigma &= 20 \\ \mu - 3\sigma &= -10 \end{aligned}$$

Nous devons donc trouver $P(-15 \leq X \leq 20)$ ce qui est égal à $P(0 \leq X \leq 20)$ ou encore plus simplement $P(X \leq 20)$. Donc:

$$P(X < 20) = 1 - P(X > 20) = 1 - e^{-20\lambda} = 98.17\%$$

EXERCICE 24.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Défaillance

Énoncé:

Le nombre moyen de produits défectueux dans une ligne de production est de 10.5 par mois avec un écart-type de 2.5. Quelle est la probabilité suivant une loi Normale que le nombre d'éléments défectueux choisis aléatoirement sera inférieur à 15.

Solution:

Il suffit de calculer avec MS Excel:

$$=NORMDIST(15;10.5;2.5;1)=96.47\%$$

EXERCICE 25.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Défaillance Weibull

Énoncé:

Nous avons dans MS Excel les données suivantes représentant les durées de vies de sept machines données en années:

	A	B	C
12			
13	Moyenne	18.684286	
14	Ecart-Type	7.4008848	
15			
16		Machine 1	8.5
17		Machine 2	12.54
18		Machine 3	13.75
19		Machine 4	19.75
20		Machine 5	21.46
21		Machine 6	26.34
22		Machine 7	28.45

Nous avons calculé dans B13 la moyenne arithmétique (l'espérance) et dans B14 l'estimateur de maximum de vraisemblance de l'écart-type de la loi Normale... (hmm...):

	A	B
12		
13	Moyenne	=AVERAGE(C16:C22)
14	Ecart-Type	=STDEV(C16:C22)

Nous voulons modéliser la durée de vie par une loi de Weibull de paramètres:

$$\gamma = 0, \eta = ?, \beta = ?$$

Nous demandons:

1. De déterminer η, β en utilisant la méthode de Lloyd et Lipov
2. Déterminer $MUT = MTBF$ (suppose temps d'arrêt pour réparation négligeable)
3. De calculer la probabilité cumulée R que la machine tienne le coup plus de 20 mois.
4. La probabilité cumulée F que la machine tombe en panne entre le 15 et 30^{ème} mois.
5. La durée limite de garantie minimale assurant une probabilité cumulée de fonctionnement (fiabilité) de 95%.

Solution:

S1. Nous avons démontré dans le chapitre Technique de Gestion que pour $\gamma = 0$:

$$E(X) = \eta \Gamma_{-1} \left(1 + \frac{1}{\beta} \right)$$

et :

$$\sigma^2 = \eta^2 \left(\Gamma_{-1} \left(\frac{\beta+2}{\beta} \right) - \Gamma_{-1} \left(\frac{\beta+1}{\beta} \right)^2 \right)$$

Nous créons donc une zone dans la feuille Excel avec la structure suivante:

	D	E	F	G	H
1					
2	beta				
3	nu				
4					
5				Objectif	
6	Espérance	#DIV/0!	=	18.68	Espérance
7	Morceaux \		0		
8		#DIV/0!			
9		#DIV/0!			
10	Variance	#DIV/0!			
11	Ecart-Type	#DIV/0!	=	7.401	Ecart-Type

Avec les formules suivantes:

	D	E	F	G	H
1					
2	beta				
3	nu				
4					
5				Objectif	
6	Espérance	=E3*EXP(GAMMALN(1+(1/E2)))	=	=B13	Espérance
7	Morceaux Variance	=E3^2			
8		=EXP(GAMMALN((1+(2/E2))))			
9		=EXP(GAMMALN((1+(1/alpha))))			
10	Variance	=E7*(E8-E9^2)			
11	Ecart-Type	=SQRT(E10)	=	=B14	Ecart-Type

La fonction Gamma (à ne pas confondre avec la loi Gamma!) n'existant pas dans MS Excel nous devons prendre l'exponentielle de la fonction du logarithme népérien de la fonction Gamma qui elle est disponible O_o.

Une fois ceci fait, nous allons demander à l'aide du solveur de MS Excel de déterminer les valeurs de E2 et E3 afin de faire correspondre l'espérance et l'écart-type de la loi de Weibull aux valeurs obtenues dans B13 et B14.

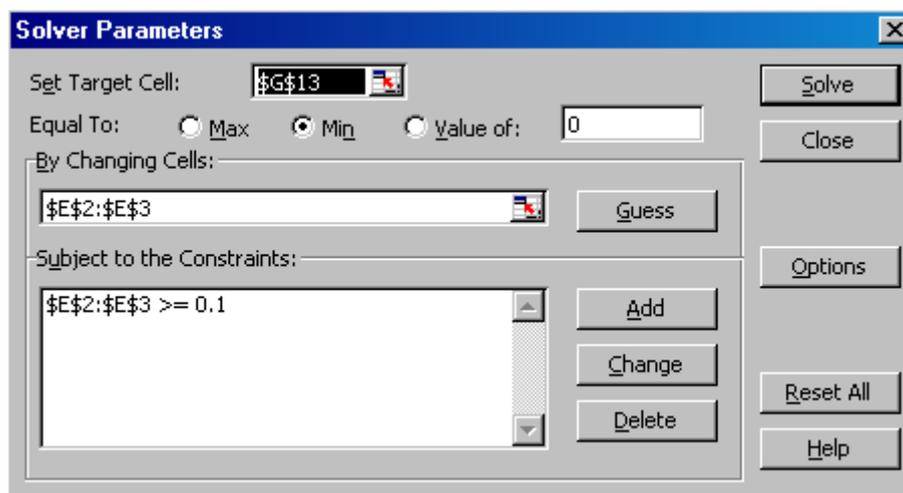
Mais le solveur a besoin d'un objectif et il peut en avoir qu'un. Nous devons donc jouer avec des paramètres supplémentaires qui en tant qu'objectif unique imposeront la détermination de E2 et E3. Nous allons pour cela jouer avec la variation entre l'espérance calculée par le solveur et l'espérance déterminée expérimentalement et idem avec l'écart-type. Cela nous donne:

	G	H
5	Objectif	Delta
6	18.68429	#DIV/0!
7		
8		
9		
10		Delta
11	7.400885	#DIV/0!
12		
13	#DIV/0!	Objectif 0

avec les formules suivantes:

	G	H
5	Objectif	Delta
6	=B13	=(G6-E6)^2
7		
8		
9		
10		Delta
11	=B14	=(G11-E11)^2
12		
13	=SUM(H6;H11)	Objectif 0

Nous pouvons maintenant lancer le solveur avec les paramètres:



L'exécution du solveur donne:

$$E2 = \beta = 3.59$$

$$E3 = \eta = 20.7$$

S2. L'exécution du solveur donne aussi:

$$E(X) = MUT = MTTF = \eta \Gamma_{-1} \left(1 + \frac{1}{\beta} \right) = 18.65$$

S3. La probabilité que la machine tienne le coup plus de 20 mois est alors donnée par:

$$R = 1 - \text{WEIBULL}(20; \beta; \eta; 1) = 58.67\%$$

et donc qu'elle tienne le coup au moins 20 mois:

$$=WEIBULL(20;beta;nu;1)=41.32\%$$

S4. La probabilité que la machine tombe en panne entre le 15 et 30^{ème} mois:

$$F=WEIBULL(30;beta;nu;1)- WEIBULL(15;beta;nu;1)=70.78\%$$

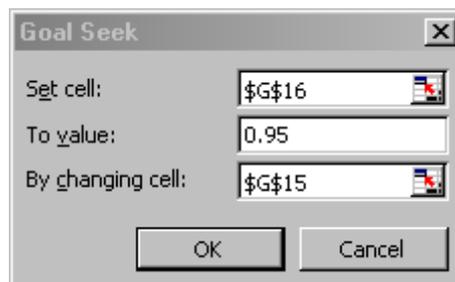
S5. Pour répondre à cette question nous reprenons les cellules suivantes nous d'abord les cellules suivantes dans MS Excel nous ayant servir pour le calcul de la deuxième partie de la solution 3 (S3):

	F	G
15		20
16	Prob>=20 hours	0.413261729

Soit avec les formules:

	F	G
15		20
16	Prob>=20 hours	=1-WEIBULL(G15;beta;nu;1)

Maintenant nous utilisons l'outil cible de MS Excel dans Outils/Outil cible:



En validant par OK, nous obtenons alors à l'écran:

$$G15=9.07$$

années. Donc nous pouvons garantir nos machines 9 années sans avoir peu de perdre trop d'argent en les commercialisant.

EXERCICE 26.

Niveau: Gymnase (Lycée)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Défaillance

Énoncé:

Un fabricant d'ampoules suite à une étude observe que ces ampoules ont une durée moyenne de vie de 250 heures avec un écart-type de 15 heures suivant une loi Normale. Il souhaiterait sur ses emballages indiquer le temps de ses ampoules tel que seulement 5% de celles-ci se trouvent en-dessous de l'indication. Quelle est la durée à indiquer?

Solution:

Il suffit de calculer avec MS Excel:

$$=NORMINV(5\%;250;15)=225.325$$

EXERCICE 27.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Défaillance

Énoncé:

Le nombre de pièces sortant d'une usine donnée en l'espace d'une semaine est une variable aléatoire d'espérance 50. Nous supposons en plus que la variance de la production hebdomadaire est de 25. Nous cherchons à majorer la probabilité que la production de la semaine prochaine soit comprise entre 40 et 60 pièces.

Solution:

Pour calculer ceci il faut d'abord se souvenir que l'inégalité de Bienaymé-Tchebychev est basée en parties sur le terme $|X - E(X)|$ donc nous avons:

$$|40 - 50| = |60 - 50| = 10$$

donc l'inégalité de BT nous permet bien de travailler sur des intervalles égaux en valeur absolue ce qui s'écrit aussi:

$$-10 \leq X - E(X) \leq 10$$

Ensuite, ne reste plus qu'à appliquer simplement l'inégalité numériquement:

$$P(|X - E(X)| \geq \varepsilon) = P(|X - E(X)| \geq 10) \leq \frac{V(X)}{\varepsilon^2} = \frac{25}{10^2} = 0.25 = 25\%$$

EXERCICE 28.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Intervalle

Énoncé:

Un échantillon de boulons a été pris d'une chaîne de production avec une moyenne de 2 cm de diamètres et un écart-type de 1.5 cm. Quel pourcentage minimum de boulons seront compris dans un intervalle de $\pm 1.75\sigma$.

Solution:

Pour répondre, nous utilisons la deuxième formulation de l'inégalité de Bienaymé-Tchebychev:

$$P(|X - E(X)| \geq t\sigma) = P(|X - E(X)| \geq 1.75\sigma) \leq \frac{1}{t^2} = \frac{1}{1.75^2} = 0.327$$

Donc nous avons:

$$1 - 0.327 = 0.673 = 67.3\%$$

de boulons qui sont dans l'intervalle spécifié.

EXERCICE 29.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Défaillance

Énoncé:

Une société fabriquant des cartes mères sous contrôle statistique sait que le nombre moyen d'imperfections par carte mère est de $\mu = 5$ avec un écart-type de $\sigma = 2.34$.

Chaque échantillon pouvant être vu comme une variable aléatoire indépendante (...) et identiquement distribuée.

Sous les hypothèses du théorème central limite, nous souhaiterions savoir si un échantillon aléatoire de $n = 36$ pièces sont prises pour une inspection avec une moyenne $\bar{x} = 6$ imperfections par unité quelle est la probabilité cumulée d'avoir $\bar{x} \leq 6$ si le processus est sous contrôle?

Solution:

Nous allons donc utiliser le théorème central limite puisque la moyenne mesurée à 6 peut être considérée comme un somme pondérée des variables aléatoires. Elle est ainsi distribuée selon une loi normale centrée réduite. Et nous savons que ce théorème s'utilisait avec l'écriture suivante de la variable aléatoire:

$$Z_n = \frac{S_n - n\mu}{\sigma\sqrt{n}} = \frac{M_n - \mu}{\sigma/\sqrt{n}}$$

En l'occurrence ici c'est la deuxième relation qui nous intéresse.

Nous avons donc:

$$Z_n = \frac{M_n - \mu}{\sigma/\sqrt{n}} = \frac{6 - 5}{2.34/\sqrt{36}} = 2.56$$

et pour finir nous devons utiliser MS Excel:

$$=NORMSDIST((6-5)/(2.34/SQRT(36)))=99.48\%$$

EXERCICE 30.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Moyenne et variance de multiples variables normalement distribuées

Énoncé:

Les chaînes de production d'une société ont produit des pièces pendant 4 semaines. Chaque semaine, la moyenne arithmétique et l'écart-type sans biais d'une côte d'une pièce ont été calculées sur la base de distribution normale. A la fin des 4 semaines le tableau suivant a été obtenu:

N° Semaine	Moyenne μ_i	Ecart-type σ_i
1	20.5	0.3
2	20.6	0.2
3	20.5	0.1
4	20.4	0.5

Calculez la moyenne globale et l'écart-type global.

Solution:

Nous avons démontré dans le chapitre de Statistiques que:

$$N(\mu, \sigma) = N\left(\frac{1}{n} \sum_{i=1}^n \mu_i, \frac{1}{n} \sqrt{\sum_{i=1}^n \sigma_i^2}\right)$$

il vient alors immédiatement:

$$\mu = \frac{20.5 + 20.6 + 20.5 + 20.4}{4} = 20.50$$

et:

$$\sigma = \frac{1}{4} \sqrt{0.3^2 + 0.2^2 + 0.1^2 + 0.5^2} = 0.156$$

EXERCICE 31.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Sondages (test Z à 1 échantillon)

Énoncé:

Un sondage a été conduit dans des entreprises qui utilisent des panneaux solaires comme source primaire d'électricité. La question qui a été posée était la suivante: Quel est le % de votre électricité provenant du solaire?

Un échantillon aléatoire de 55 réponses donne une moyenne arithmétique de 45 [kW]. Supposons que l'écart-type standard pour cette question est de 15.5 [kW] (écart-type théorique supposé connu!).

Quel est l'intervalle de confiance représentant 95% de valeurs.

Solution:

Nous avons démontré que:

$$\bar{X} - Z_{0.025} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{0.025} \frac{\sigma}{\sqrt{n}}$$

Et dans le cas de cet exercice nous avons donc:

$$n = 55, \bar{X} = 45, \sigma = 15.5, Z_{\alpha/2} = 1.96$$

Avec MS Excel nous obtenons justement simplement:

$$=ABS(NORMSINV(5\%/2))=1.96$$

Nous avons donc:

$$45 - 1.96 \frac{15.5}{\sqrt{55}} \leq \mu \leq 45 + 1.96 \frac{15.5}{\sqrt{55}}$$

Donc nous avons:

$$P(40.9 \leq \mu \leq 49.1) = 95\%$$

Signalons que la valeur:

$$1.96 \frac{15.5}{\sqrt{55}}$$

peut s'obtenir avec la fonction:

=CONFIDENCE(5%;15.5;55)

EXERCICE 32.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Intervalle de confiance de la moyenne (test t à 1 échantillon)

Énoncé:

Un gestionnaire d'une société de location de voitures souhaiterait estimer le nombre moyen de fois que des véhicules de luxe sont loués par mois. Il prend pour cela un échantillon aléatoire de 19 voitures de luxes et obtient la suite suivante de nombre de voitures louées:

3,7,12,5,9,13,2,8,6,14,6,1,2,3,2,5,11,13,5

E1. Calculez la moyenne arithmétique

E2. Déterminez l'intervalle à 95% de la moyenne.

Solution:

S1. La moyenne arithmétique vaut:

$$\bar{X} = \frac{3+7+12+5+9+13+2+8+6+14+6+1+2+3+2+5+11+13+5}{19} = 6.68$$

S2. Nous avons démontré que:

$$\bar{X} - \frac{S}{\sqrt{n}} T_{\alpha/2}(n-1) \leq \mu \leq \bar{X} + \frac{S}{\sqrt{n}} T_{\alpha/2}(n-1)$$

avec:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Nous avons alors:

$$\bar{X} = \frac{3+7+12+5+9+13+2+8+6+14+6+1+2+3+2+5+11+13+5}{19} = 6.68$$

$$S = \text{ECARTYPE}(3;7;12;5;9;13;2;8;6;14;6;1;2;3;2;5;11;13;5) = 4.23$$

$$n = 19$$

$$df = n - 1 = 18$$

$$T_{0.025}(18) = \text{LOLSTUDENT.INVERSE}(5\%;18) \cong 2.1$$

Nous calculons:

$$6.68 - \frac{4.23}{\sqrt{19}} 2.1 \leq \mu \leq 6.68 + \frac{4.23}{\sqrt{19}} 2.1$$

Soit 95% de probabilité cumulée que la moyenne se trouve dans l'intervalle:

$$4.64 \leq \mu \leq 8.71$$

EXERCICE 33.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Intervalle de confiance de la proportion

Énoncé:

Dans une production de 300 pièces nous en avons trouvées 8 de défectueuses. Quelle est l'intervalle de confiance à 95% de la proportion de pièces défectueuses?

Solution:

Avec $\alpha = 5\%$, nous avons alors:

$$\hat{p} - Z_{0.025} \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \leq p \leq \hat{p} + Z_{0.025} \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

$$\hat{p} - 1.96 \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}} \leq p \leq \hat{p} + 1.96 \sqrt{\frac{\hat{p} \cdot (1 - \hat{p})}{n}}$$

Nous avons d'une production de 300 éléments trouvés 8 qui étaient défectueux. Nous vérifions d'abord avec :

$$\hat{p} = \frac{8}{300} \cong 2.66\%$$

que:

$$n \cdot \hat{p} = 300 \cdot \frac{8}{300} = 8 \geq 5$$

Donc il est acceptable d'utiliser l'intervalle de confiance par la loi Normale. Nous avons dès lors:

$$\frac{8}{300} - 1.96 \sqrt{\frac{\frac{8}{300} \cdot \left(1 - \frac{8}{300}\right)}{300}} \leq p \leq \frac{8}{300} + 1.96 \sqrt{\frac{\frac{8}{300} \cdot \left(1 - \frac{8}{300}\right)}{300}}$$

$$0.84\% \leq p \leq 4.49\%$$

EXERCICE 34.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Nombre d'échantillons pour une proportion

Énoncé:

Nous souhaiterions savoir le nombre d'échantillons à prendre d'un lot de production sachant que la proportion de défectueux est imposée à 30% avec une erreur tolérée d'environ 5% entre la proportion réelle et empirique et ce afin d'obtenir un intervalle de confiance à un niveau de 95% du résultat:

Solution:

Nous prenons la relation démontrée dans le cours de statistiques:

$$n \cong \frac{Z_{\alpha} p(1-p)}{\delta^2} = \frac{Z_{5\%} 30\%(1-30\%)}{5\%^2} = \frac{1.96 \cdot 30\% \cdot 70\%}{5\%^2} = 165$$

EXERCICE 35.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Comparaison de deux proportions

Énoncé:

Dans le cadre d'un plan d'échantillonnage nous avons prélevé sur un premier lot de 50 échantillons, 48 en parfait états. Dans un second lot de 30 échantillons, 26 étaient en bon état.

Nous avons donc:

$$\hat{p}_X = \frac{48}{50} = 96\%, \hat{p}_Y = \frac{26}{30} = 86.66\%$$

Nous souhaiterions donc savoir si la différence est significative avec une certitude de 95% ou simplement due au hasard?

Solution:

Nous utilisons alors:

$$\hat{p} = \frac{n_1 \hat{p}_1 + n_2 \hat{p}_2}{n_1 + n_2} = \frac{48 + 26}{50 + 30} = 92.50\%$$

et:

$$Z \cong \frac{\hat{p}_X - \hat{p}_Y}{\sqrt{\hat{p} \cdot (1 - \hat{p}) \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{0.96 - 0.8666}{\sqrt{0.925 \cdot (1 - 0.925) \left(\frac{1}{50} + \frac{1}{30} \right)}} \cong 1.545$$

Ce qui correspond à une probabilité cumulée en utilisant MS Excel de :

$$= \text{NORMSDIST}(1.545) = 93.88\%$$

Donc la différence est dû au hasard (ceci dit c'est presque in extremis...).

EXERCICE 36.*Niveau:* Universitaire (Faculté)*Auteur:* Vincent ISOZ (isozv@hotmail.com)*Mots-clés:* Intervalle de confiance de l'écart-type**Énoncé:**

Un échantillon de 9 vis a été tiré d'une ligne de production et la mesure de leur diamètre en [mm] a été reportée ci-dessous:

13,13,12,12.55,12.99,12.89,12.88,12.97,12.99

E1. Calculez l'écart-type**E2.** Estimez l'intervalle de confiance de l'écart-type à 95%.**Solution:****S1.** L'écart-type vaut (à mettre dans MS Excel):

$$=ECARTTYPE(13;13;12;12.55;12.99;12.89;12.88;12.97;12.99)=0.335$$

S2. Nous avons démontré que:

$$\frac{(n-1)S^2}{\chi_{\alpha/2}^2(n-1)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi_{1-\alpha/2}^2(n-1)}$$

avec:

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X})^2$$

Nous avons alors:

$$S^2 = ECARTYPE(13;13;12;12.55;12.99;12.89;12.88;12.97;12.99)^2=0.112$$

$$n = 9$$

$$df = n - 1 = 8$$

$$\chi_{0.025}^2(8)=KHIDEUX.INVERSE(0.025;8)=17.534$$

$$\chi_{0.975}^2(8)=KHIDEUX.INVERSE(0.975;8)=2.179$$

Donc:

$$\frac{8 \cdot 0.112}{17.534} \leq \sigma^2 \leq \frac{8 \cdot 0.112}{2.179}$$

soit:

$$0.051 \leq \sigma^2 \leq 0.412$$

Soit 95% de probabilité cumulée que l'écart-type se trouve dans l'intervalle:

$$0.226 \leq \sigma \leq 0.642$$

EXERCICE 37.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Nombre d'échantillons (relatif au test Z à 1 échantillon)

Énoncé:

Un gestionnaire responsable d'une chaîne de production veut savoir combien de temps passe en moyenne un employé à l'usinage sur une pièce donnée. Celui-ci aimerait être dans une estimation se trouvant dans les ± 2 minutes par rapport à la moyenne et il est connu que l'écart-type d'usinage est de 3 minutes.

Quelle taille d'échantillon de mesures doit-il prendre pour s'assurer un intervalle de confiance à 95% de la moyenne qu'il mesurera?

Solution:

Nous prenons la relation démontrée dans le cours de statistiques:

$$n \cong \frac{Z^2 \sigma^2}{\delta^2} = \left(\frac{\text{LOI.NORMALE.STANDARD.INVERSE}((1-95\%)/2) \cdot 3}{2} \right)^2 \cong 8.64$$

EXERCICE 38.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Test d'hypothèse Z bilatéral

Énoncé:

Un ancien sondage a montré que le salaire moyen de manager des 500 plus grandes compagnies était de 80'000.-/an. Un enquêteur veut tester si ce chiffre est toujours valide quelques années plus tard. Pour cela, il prend un échantillon aléatoire de 150 managers et il obtient une moyenne de 78'000.- avec un écart-type sur toute la population assumé à 15'000. Le niveau de confiance imposé est de 5%. Devons-nous accepter la valeur de 80'000 ou la rejeter?

Déterminez aussi avec l'outil cible de MS Excel la p -value de l'intervalle de confiance.

Solution:

S1. Nous avons donc:

$$\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

avec les données suivantes:

$$\bar{X} = 78'000$$

$$\sigma = 15'000$$

$$n = 150$$

$$Z_{\alpha/2} = 1.96$$

Il vient alors immédiatement:

$$75'559.49 \leq \mu \leq 80'400.50$$

Comme 78'000 est dans l'intervalle, nous pouvons accepter l'hypothèse sous les contraintes données! En d'autres termes, nous pouvons aussi dire que l'échantillon est avec 95% de confiance représentatif de la population.

S2. Pour trouver la p -value, nous construisons dans MS Excel le tableau suivant:

	A	B
1	Ecart-Sup	=78000-NORMSINV(B3/2)*15000/SQRT(150)
2	Hypothèse	80000
3	Seul de confiance (p-value)	0.05
4	Niveau de confiance:	=100%-B3
5	Z correspondant:	=NORMSINV(B4)

qui donne dans l'état actuel des choses:

	A	B
1	Ecart-Sup	80400.45584
2	Hypothèse	80000
3	Seul de confiance (p-value)	5%
4	Niveau de confiance:	95.000%
5	Z correspondant:	1.6449

Ensuite, nous lançons l'outil valeur cible avec les paramètres suivantes:

et nous trouvons:

	A	B
1	Ecart-Sup	80000.00
2	Hypothèse	80000.00
3	Seul de confiance (p-value)	10.25%
4	Niveau de confiance:	89.753%
5	Z correspondant:	1.2676

Donc la p -value, valeur limite de validité du seuil de confiance de l'hypothèse (pour rappel), est de 10.25% (0.102). Ce qui correspond donc à un seuil de confiance limite de ~89.75% comme nous pouvons le voir.

Il est bien sûr possible de faire cette même manipulation pour tout type de statistique inférentielle!

EXERCICE 39.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Test d'hypothèse Z bilatéral

Énoncé:

Une entreprise fabriquant des processeurs mesure chaque jour le nombre de pièces défectueuses. Historiquement celle-ci obtient une moyenne de 3 pièces par jour. Un consultant qualité doit venir faire un audit pour voir si cette information est juste. Etant donné son coût à la journée il ne vient que 10 jours prendre les mesures et obtient le tableau suivant:

	A
1	3
2	6
3	7
4	8
5	6
6	5
7	4
8	2
9	1
10	9

et obtient avec MS Excel une moyenne de 5.1 et un écart-type de 2.6.

Qu'elle est l'intervalle de confiance à 95% des ces mesures. Faites le calcul à la main et avec MS Excel.

Solution:

Nous avons donc:

$$\bar{X} - Z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

avec les données suivantes:

$$\bar{X} = 5.1$$

$$\sigma = 2.6$$

$$n = 10$$

$$Z_{\alpha/2} = 1.96$$

Il vient alors immédiatement:

$$3.4899 \leq \mu \leq 6.711$$

En rappelant à nouveau la valeur:

$$1.96 \frac{2.6}{\sqrt{10}}$$

peut s'obtenir avec la fonction:

$$=CONFIDENCE(5\%;2.6;10)$$

Donc selon les données obtenues, il doit rejeter l'information qualité obtenue par son client.

Avec MS Excel nous pouvons utiliser la fonction suivante qui nous donne le seuil de confiance limite:

$$=ZTEST(A1:A10;3)=0.534\%$$

Soit:

$$\alpha = 0.534\% \cdot 2 = 1\%$$

Donc un intervalle de confiance de 99%. Ce qui est un peu abusé quand même...

EXERCICE 40.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Test d'hypothèse t-student bilatéral

Énoncé:

Une machine utilisée pour produire des joints a été stable et opérante sous contrôle statistique pendant plusieurs années. Mais récemment, l'épaisseur des joints semble être plus petite qu'ils ne l'étaient au préalable.

L'épaisseur moyenne était historiquement de $\mu = 0.070$. Un responsable qualité (AQ: Quality Assurance) veut déterminer si l'âge des machines est la cause de cette variation. Pour cela, il prend un échantillon de 10 joints pour le test et trouve une moyenne arithmétique de l'échantillon de 0.074. et un écart-type sans biais de 0.008.

Nous souhaiterions tester l'hypothèse comme quoi la machine fonctionne correctement avec un intervalle de confiance de 95%!

Solution:

Comme nous ne connaissons pas la variance théorique, nous devons utiliser le test-t.

Nous avons donc:

$$\bar{X} - T_{\alpha/2} \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} + T_{\alpha/2} \frac{S}{\sqrt{n}}$$

avec les données suivantes pour MS Excel 2007 et antérieur:

$$\bar{X} = 0.074$$

$$s = 0.008$$

$$n = 10$$

$$df = n - 1 = 9$$

$$T_{0.025}(9) = \text{TINV}(2.5\%;9) = 2.69$$

Soit:

$$0.074 - 2.69 \frac{0.008}{\sqrt{10}} \leq \mu \leq 0.074 + 2.69 \frac{0.008}{\sqrt{10}}$$

Ce qui donne:

$$0.0595 \leq \mu \leq 0.142$$

Comme la moyenne historique se trouve dans l'intervalle de confiance, il s'ensuit que nous devons rejeter l'hypothèse comme quoi la variation est due au vieillissement de la machine.

EXERCICE 41.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Test d'hypothèse du χ^2 bilatéral

Énoncé:

Une société a dans ses études des délais de ses projets un écart-type historique de 2.78 jours. Ces derniers 17 jours (c'est donc bref par rapport à la donnée précédente!) l'écart-type des délais a été de 3.01 jours.

Avec un niveau de confiance de 95% pouvons-nous dire si la variance augmente ou si c'est un effet temporaire?

Solution:

Nous savons que:

$$\frac{(n-1)S^2}{\chi^2_{\alpha/2}(n-1)} \leq \sigma^2 \leq \frac{(n-1)S^2}{\chi^2_{1-\alpha/2}(n-1)}$$

Nous avons les informations suivantes:

$$S^2 = 3.01^2 = 9.0601$$

$$\sigma^2 = 2.78^2 = 7.729$$

$$n = 17$$

$$df = n - 1 = 16$$

$$\chi^2_{0.025}(16) = \text{CHIINV}(0.025;16) = \text{KHIDEUX.INVERSE}(0.025;16) = 28.845$$

$$\chi^2_{0.975}(16) = \text{CHIINV}(0.975;16) = \text{KHIDEUX.INVERSE}(0.975;16) = 6.907$$

Soit:

$$5.025 \leq \sigma^2 \leq 20.986$$

Nous ne devons donc pas rejeter cette hypothèse!

EXERCICE 42.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Test Z bilatéral sur la différence de deux moyennes

Énoncé:

En décembre, la moyenne de la production par employé à la société Sénégal-Electrique a été de 150 machines par heures avec une déviation de 15 machines. Pour le même mois, la production moyenne à la société Cazamance-Electromotrice a été de 135 machines par heure, avec une déviation de 9 machines.

Calculez:

1. La probabilité cumulée que la différence entre les deux moyennes soit supérieure à 20.
2. Si nous devons rejeter ou non l'hypothèse comme quoi la différence est égale à 20 avec un intervalle de confiance de 95% ?

Nous avons les données suivantes à disposition:

	A	B
1	Sen-El.	Caz-El.
2	133	135
3	170	116
4	148	142
5	168	144
6	138	154
7	135	119
8	137	139
9	151	115
10	163	122
11	155	143
12	140	146
13	150	133
14	163	135
15	151	132
16	162	128
17	139	138
18	139	153
19	164	153
20	144	118

Nous demandons d'expliciter les calculs avec MS Excel et de faire l'analyse avec l'utilitaire d'analyse de MS Excel.

Solution:

Nous allons utiliser la relation:

$$Z^{calc.} = \frac{(\bar{X}_2 - \bar{X}_1) - (\mu_2 - \mu_1)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$$

qui permet de se reporter à un test bilatéral.

Donc nous avons les données suivantes:

$$\bar{X}_1 = \text{MOYENNE}(A2:A20) = 150$$

$$\bar{X}_2 = \text{MOYENNE}(B2:B20) = 135$$

$$\mu_2 - \mu_1 = 20$$

$$V_1 = \sigma_1^2 = 15^2 = 225$$

$$n_1 = 19$$

$$V_2 = \sigma_2^2 = 9^2 = 81$$

$$n_2 = 19$$

Soit:

$$Z^{calc.} = \frac{(150 - 135) - 20}{\sqrt{\frac{225}{19} + \frac{81}{19}}} = -1.24591$$

S1. Ce qui correspond à une probabilité cumulée de:

$$=\text{NORMSDIST}(-1.24591)=10.63\%$$

Si les deux moyennes $(\bar{X}_2 - \bar{X}_1) = 15$ et 20 étaient égales, nous aurions 50% pour le résultat du calcul précédent (le numérateur valant zéro) et donc 50% de probabilité que la différence soit supérieure à 20.

Il s'ensuit puisque nous avons 10.63% dans le calcul précédent, que nous avons:

$$50\% - 10.63\% \cong 40\%$$

de probabilité cumulée d'arriver à la limite des 50% qui représente le point à partir duquel la moyenne sera supérieure à 20.

S2. Maintenant pour voir si c'est dans l'intervalle de confiance de 95% nous devons avoir:

$$|Z^{calc.}| \leq |Z_{\alpha/2}|$$

Soit:

$$1.24591 \leq \text{ABS}(\text{NORMSINV}(5\%/2))$$

Ce qui donne:

$$1.24591 \leq 1.96$$

Avec l'utilitaire d'analyse de MS Excel cela donne:

Nous obtenons alors le tableau suivant:

	A	B	C
1	z-Test: Two Sample for Means		
2			
3		<i>Variable 1</i>	<i>Variable 2</i>
4	Mean	150	135
5	Known Variance	225	81
6	Observations	19	19
7	Hypothesized Mean Difference	20	
8	z	-1.24591	
9	P(Z<=z) one-tail	0.106399	
10	z Critical one-tail	1.644854	
11	P(Z<=z) two-tail	0.212798	
12	z Critical two-tail	1.959964	

Nous y retrouvons bien les données calculées plus haut pour le test bilatéral (two-tail).

Nous y voyons aussi la valeur classique du Z critique en bilatéral:

$$=NORMSINV(97.5\%)=ABS(NORMSINV(2.5\%))=1.959964$$

Nous y voyons aussi la valeur en probabilité cumulée bilatérale correspondante au Z de -1.24591:

$$=NORMSDIST(-1.24591)*2=0.212798=21.2798\%$$

Pour conclure, nous ne pouvons donc pas rejeter l'hypothèse comme quoi la moyenne est différente de 20 à un niveau de 95%.

EXERCICE 43.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Test *t* de Student bilatéral sur 1 échantillon

Énoncé:

Nous devons étalonner un appareil à l'aide d'une pièce étalon dont nous savons que le diamètre réel est de 15 microns. Nous avons obtenu les mesures suivantes:

	A
1	Mesures
2	15.0809
3	15.0873
4	14.9679
5	15.0423
6	15.1029
7	14.9803
8	15.1299
9	15.0414
10	15.0351
11	15.0559
12	15.0793
13	15.0753
14	15.0483
15	15.0515
16	15.0962
17	15.0654
18	14.9759
19	15.0507

Nous souhaiterions vérifier un intervalle de confiance de 95% que nous sommes proches la moyenne vraie (pas trop différent).

Solution:

Il s'agit donc d'appliquer:

$$\frac{\bar{X} - \mu}{S / \sqrt{n}} = T(n-1)$$

en se rappelant que la loi distribution de Student est symétrique et centrée en zéro (d'où le choix du test bilatéral comme étant le choix le plus fréquent).

Et nous avons dans le cas présent:

$$\mu = 15$$

qui est imposé et:

$$\bar{X}_1 = \text{MOYENNE}(A2:A19)=15.05369$$

$$n = \text{NB}(A2:A19)=18$$

$$S = \text{VAR}(A2:A19)=0.00192$$

Cela donne:

$$\frac{\bar{X} - \mu}{S/\sqrt{n}} = 118.6562$$

Or, nous avons avec MS Excel:

$$T_{2.5\%}(n-1)=\text{T.INV}(2.5\%;17)=-2.1098$$

$$T_{97.5\%}(n-1)=\text{T.INV}(2.5\%;17)=+2.1098$$

Nous sommes donc très très loin du compte et nous allons devoir rejeter l'hypothèse nulle comme quoi la moyenne est égale à 15 en faveur de l'hypothèse alternative (comme quoi elle est bien différente!).

Il serait intéressant calculer la p -value correspondant à 118.6562. Nous avons alors:

$$=\text{T.DIST}(K11;17;\text{TRUE})=1.0000000$$

Donc in extenso la p -value est nulle!

Il serait aussi intéressant de savoir quel aurait été l'intervalle de confiance toléré. Il vient alors:

$$-2.1098 = \frac{\bar{X} - \mu}{S/\sqrt{n}} \Rightarrow \bar{X}_{\text{inf}} = -2.1098 \frac{S}{\sqrt{n}} + 15 = 14.99904526$$

et:

$$+2.1098 = \frac{\bar{X} - \mu}{S/\sqrt{n}} \Rightarrow \bar{X}_{\text{sup}} = 2.1098 \frac{S}{\sqrt{n}} + 15 = 15.00095474$$

Donc évidemment avec une moyenne

EXERCICE 44.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Test t de Student homoscédastique bilatéral d'égalité de la moyenne

Énoncé:

Le débit d'eau en litre par minutes entre plusieurs points équidistants de deux pipelines 1 et 2 a été mesuré. Le tableau suivant a été obtenu:

	A	B
1	Pipeline 1	Pipeline 2
2	163	167
3	150	157
4	171	149
5	155	145
6	186	135
7	145	157
8	154	135
9	173	167
10	152	154
11	150	165
12	143	170
13	138	165
14	166	154
15	193	176
16	158	155
17	175	157
18	167	134
19	150	156
20	158	147

Un ingénieur souhaite déterminer s'il y a statistiquement une différence significative entre les deux pipeline à un niveau de confiance de 95% en supposant les variances égales. Faites le calcul de manière détaillée avec MS Excel et ensuite avec l'Utilitaire d'Analyse.

Solution:

Nous devons donc utiliser lorsque les variances sont égales (homoscédasticité):

$$T^{calc.}(n-2) = \frac{(\bar{X}_2 - \bar{X}_1) - (\mu_2 - \mu_1)}{\sqrt{V_R \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{(\bar{X}_2 - \bar{X}_1)}{\sqrt{V_R \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}}$$

Donc:

$$|T^{calc.}(n-2)| \leq |T_{\alpha/2}(n-2)|$$

Avec MS Excel il est aisé d'obtenir selon le tableau donné les différents paramètres:

$$\bar{X}_1 = \text{MOYENNE}(A2:A20)=160.369$$

$$\bar{X}_2 = \text{MOYENNE}(B2:B20)=155.00$$

$$n_1 = \text{NB}(A2:A20)=19$$

$$n_2 = \text{NB}(B2:B20)=19$$

$$S_1^2 = \text{VAR}(A2:A20)=211.246$$

$$S_2^2 = \text{VAR}(B2:B20)=145.00$$

$$V_R = \frac{(n_1 - 1)S_1^2 + (n_2 - 1)S_2^2}{n_1 + n_2 - 2} = 178.228$$

L'estimation de la différence est donc de:

$$|\bar{X}_2 - \bar{X}_1| = 5.369$$

Soit:

$$T^{calc.}(36) = \frac{(\bar{X}_2 - \bar{X}_1) - (\mu_2 - \mu_1)}{\sqrt{V_R \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{(\bar{X}_2 - \bar{X}_1)}{\sqrt{V_R \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = 1.2397$$

Maintenant pour voir si c'est dans l'intervalle de confiance de 95% nous devons avoir:

$$|T^{calc.}(36)| \leq |T_{\alpha/2}(36)|$$

Soit:

$$1.2397 \leq \text{TINV}(0.05;36)$$

Ce qui donne:

$$1.2397 \leq 2.028$$

Avec l'utilitaire d'analyse de MS Excel cela donne:

Nous obtenons le tableau suivant:

	A	B	C
1	t-Test: Two-Sample Assuming Equal Variances		
2			
3		<i>Variable 1</i>	<i>Variable 2</i>
4	Mean	160.3684	155
5	Variance	211.2456	145
6	Observations	19	19
7	Pooled Variance	178.1228	
8	Hypothesized Mean Difference	0	
9	df	36	
10	t Stat	1.239791	
11	P(T<=t) one-tail	0.111537	
12	t Critical one-tail	1.688298	
13	P(T<=t) two-tail	0.223074	
14	t Critical two-tail	2.028094	

Nous y retrouvons bien la valeur calculée précédemment de 1.2397.

De plus, nous avons la valeur limite de confiance ($\alpha = 22.31\%$) qui est elle difficile à obtenir formellement et mais qui peut indirectement être obtenue dans MS Excel avec la relation suivante:

$$=TTEST(A2:A20;B2:B20;2;2)=TDIST(1.239791;36;2)=0.22307$$

Calculons aussi les bornes inférieure et supérieure à 95% de la différence (il est dommage qu'Excel ne l'indique pas). Nous avons pour la borne inférieure:

$$=-TINV(0.025;36)*RACINE(178.228*(1/19+1/19))+5.369$$

$$=-2.028*RACINE(178.228*(1/19+1/19))+5.369=-3.415$$

et pour la borne supérieure:

$$=TINV(0.025;35)*RACINE(178.228*(1/19+1/19))+5.369$$

$$=2.028*RACINE(178.228*(1/19+1/19))+5.369=14.153$$

EXERCICE 45.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Test *t* de Student de données appariées

Énoncé:

Le débit d'eau en litre par minutes sur un point précis d'un pipeline a été mesuré avant après réparation. Nous avons obtenu les données suivantes:

	A	B
1	Pipeline 1	Pipeline 2
2	163	167
3	150	157
4	171	149
5	155	145
6	186	135
7	145	157
8	154	135
9	173	167
10	152	154
11	150	165
12	143	170
13	138	165
14	166	154
15	193	176
16	158	155
17	175	157
18	167	134
19	150	156
20	158	147

Un ingénieur souhaite déterminer s'il y a statistiquement une différence significative entre les deux pipeline à un niveau de confiance de 95% en supposant les variances égales. Faites le calcul de manière détaillée avec MS Excel.

Solution:

Nous devons donc utiliser:

$$\begin{aligned}
 T^{(n-1)} &= \frac{\bar{X} - \mu}{S/\sqrt{n}} = \frac{\bar{X}_D - \mu_D}{S_D/\sqrt{n}} = \frac{\bar{D} - \delta_0}{S_D/\sqrt{n}} = \frac{\frac{1}{n} \sum_i d_i - \delta_0}{S_D/\sqrt{n}} = \frac{\frac{1}{n} \sum_i (X_{i2} - X_{i1}) - \delta_0}{S_D/\sqrt{n}} \\
 &= \frac{\left(\frac{1}{n} \sum_i X_{i2} - \frac{1}{n} \sum_i X_{i1} \right) - \delta_0}{S_D/\sqrt{n}} = \frac{(\bar{X}_2 - \bar{X}_1) - \delta_0}{S_D/\sqrt{n}}
 \end{aligned}$$

avec:

$$S_D = \sqrt{\frac{1}{n-1} \sum_i (d_i - \bar{D})^2}$$

Avec MS Excel il est aisé d'obtenir selon le tableau donné les différents paramètres:

$$\bar{X}_1 = \text{MOYENNE}(A2:A20)=160.369$$

$$\bar{X}_2 = \text{MOYENNE}(B2:B20)=155.00$$

$$n_1 = \text{NB}(A2:A20)=19$$

$$n_2 = \text{NB}(B2:B20)=19$$

$$S_1 = \text{RACINE}(\text{VAR}(A2:A20))=\text{ECARTTYPE}(A2:A20)=14.534$$

$$S_2 = \text{RACINE}(\text{VAR}(B2:B20))=\text{ECARTTYPE}(B2:B20)=12.041$$

$$\text{ErT}_1 = \text{ECARTTYPE}(A2:A20) / \text{NB}(A2:A20) = 3.33$$

$$\text{ErT}_2 = \text{ECARTTYPE}(B2:B20) / \text{NB}(B2:B20) = 2.76$$

où ErT est l'abréviation d'erreur-type (nous l'indiquons car les logiciels de statistique le font souvent aussi).

L'estimation de la différence est donc de:

$$|\bar{X}_2 - \bar{X}_1| = 5.369$$

Nous créons ensuite une colonne supplémentaire dans la feuille Excel avec les différences:

	A	B	C
1	Pipeline 1	Pipeline 2	Différence
2	163	167	4.00
3	150	157	7.00
4	171	149	-22.00
5	155	145	-10.00
6	186	135	-51.00
7	145	157	12.00
8	154	135	-19.00
9	173	167	-6.00
10	152	154	2.00
11	150	165	15.00
12	143	170	27.00
13	138	165	27.00
14	166	154	-12.00
15	193	176	-17.00
16	158	155	-3.00
17	175	157	-18.00
18	167	134	-33.00
19	150	156	6.00
20	158	147	-11.00

et nous avons alors l'écart-type de la différence:

$$S_D = \text{ECARTYPE}(C2:C20) = 19.639$$

Soit:

$$T^{calc.}(18) = \frac{(\bar{X}_2 - \bar{X}_1) - \delta_0}{S_D / \sqrt{n}} = \frac{\bar{X}_2 - \bar{X}_1}{S_D / \sqrt{n}} = \frac{6.369}{19.639 / \sqrt{19}} = 1.119$$

Maintenant pour voir si c'est dans l'intervalle de confiance de 95% nous devons avoir:

$$|T^{calc.}(18)| \leq |T_{\alpha/2}(18)|$$

Soit:

$$1.119 \leq \text{TINV}(0.05;35)$$

Ce qui donne:

$$1.119 \leq 2.100$$

EXERCICE 46.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Test *t* de Student hétéroscédastique bilatéral d'égalité de la moyenne (test de Welch)

Énoncé:

Le débit d'eau en litre par minutes entre plusieurs points équidistants de deux pipeline 1 et 2 a été mesuré. Le tableau suivant a été obtenu:

	A	B
1	Pipeline 1	Pipeline 2
2	163	167
3	150	157
4	171	149
5	155	145
6	186	135
7	145	157
8	154	135
9	173	167
10	152	154
11	150	165
12	143	170
13	138	165
14	166	154
15	193	176
16	158	155
17	175	157
18	167	134
19	150	156
20	158	147

Un ingénieur souhaite déterminer s'il y a une statistiquement une différence significative entre les deux pipeline à un niveau de confiance de 95% en supposant les variances non-égales. Faites le calcul de manière détaillée avec MS Excel et ensuite avec l'utilitaire d'analyse.

Solution:

Nous devons donc utiliser lorsque les variances sont égales (hétéroscédasticité):

$$T^{calc.} (n-2) = \frac{(\bar{X}_2 - \bar{X}_1) - (\mu_2 - \mu_1)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

Donc:

$$|T^{calc.}(n-1)| \leq |T_{\alpha/2}(n-1)|$$

Avec MS Excel il est aisé d'obtenir selon le tableau donné les différents paramètres:

$$\bar{X}_1 = \text{MOYENNE}(A2:A20)=160.369$$

$$\bar{X}_2 = \text{MOYENNE}(B2:B20)=155.00$$

$$n_1 = \text{NB}(A2:A20)=19$$

$$n_2 = \text{NB}(B2:B20)=19$$

$$S_1^2 = \text{VAR}(A2:A20)=211.246$$

$$S_2^2 = \text{VAR}(B2:B20)=145.00$$

Soit:

$$T^{calc.}(36) = \frac{(\bar{X}_2 - \bar{X}_1) - (\mu_2 - \mu_1)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = \frac{(\bar{X}_2 - \bar{X}_1)}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} = -1.2397$$

Maintenant pour voir si c'est dans l'intervalle de confiance de 95% nous devons avoir (remarquez que nous n'utilisons pas 36 mais 35 degrés de liberté suite à la démonstration du test de Welch):

$$|T^{calc.}(35)| \leq |T_{\alpha/2}(35)|$$

Effectivement:

$$f_n \cong \frac{\left(\sum_{i=1}^n \frac{S_i^2}{n_i}\right)^2}{\sum_{i=1}^n \frac{(S_i^2 / n_i)}{n_i - 1}} = 34.79674$$

Donc la valeur entière la plus proche est bien 35.

Soit:

$$1.2397 \leq \text{TINV}(0.05;35)$$

Ce qui donne:

$$1.2397 \leq 2.0301$$

Avec l'utilitaire d'analyse de MS Excel cela donne:

Nous obtenons le tableau suivant (à comparer avec le précédent):

	A	B	C
1	t-Test: Two-Sample Assuming Unequal Variances		
2			
3		<i>Variable 1</i>	<i>Variable 2</i>
4	Mean	160.3684	155
5	Variance	211.2456	145
6	Observations	19	19
7	Hypothesized Mean Difference	0	
8	df	35	
9	t Stat	1.239791	
10	P(T<=t) one-tail	0.11165	
11	t Critical one-tail	1.689572	
12	P(T<=t) two-tail	0.223301	
13	t Critical two-tail	2.030108	

Nous y retrouvons bien la valeur calculée précédemment de 1.2397.

De plus, nous avons la valeur limite de confiance ($\alpha = 22.33\%$) qui est elle difficile à obtenir formellement et mais qui peut indirectement être obtenue dans MS Excel avec la relation suivante:

$$=TTEST(A2:A20;B2:B20;2;3)=TDIST(1.239791;36;2)=0.2233$$

EXERCICE 47.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Test F unilatéral et bilatéral de Fisher

Énoncé:

La société Kolda Automation reçoit des joints d'étanchéité de deux fournisseurs. Le gestionnaire AQ (assurance qualité) de la société veut comparer la variance de l'épaisseur des joints avec un seuil $\alpha = 0.05$.

Pour cela, il prend un échantillon de 10 joints du fournisseur A et de 12 joints du fournisseur B. Voici le tableau avec l'épaisseur des joints:

	A	B
1	A	B
2	3.20	3.00
3	3.30	3.10
4	3.10	3.20
5	3.20	3.10
6	3.10	3.20
7	3.10	3.08
8	3.15	3.20
9	3.33	3.07
10	3.11	3.20
11	3.10	3.30
12		3.20
13		3.30

- Déterminez si l'hypothèse d'égalité est acceptable ou non en unilatéral avec MS Excel.
- Déterminez si l'hypothèse d'égalité est acceptable ou non en bilatéral avec MS Excel (bilatéral autour de la valeur Modale).

Solution:

S1. Nous devons donc utiliser:

$$\frac{\sigma_X^2}{\sigma_Y^2} = \frac{1}{F_{n-1,m-1}(x)} \frac{S_X^2}{S_Y^2}$$

puisque la moyenne théorique (espérance) n'est pas connue.

L'objectif de notre hypothèse est l'égalité des variances donc:

$$1 = \frac{1}{F_{n-1,m-1}(x)} \frac{S_X^2}{S_Y^2} \Rightarrow \frac{S_X^2}{S_Y^2} = F_{n-1,m-1}$$

Dans notre cas, nous avons:

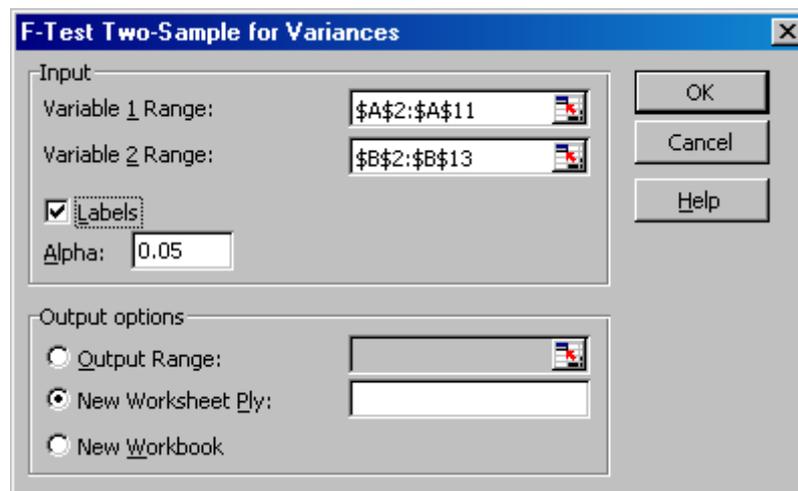
$$S_X^2 = \text{VAR}(A2:A11)=0.007543$$

$$S_Y^2 = \text{VAR}(B2:B13)=0.008584$$

Le rapport donne:

$$F_{n-1,m-1} = \frac{S_X^2}{S_Y^2} = 0.8787$$

Avec l'utilitaire d'analyse de MS Excel:



Ce qui donne:

F-Test Two-Sample for Variances		
	Variable 1	Variable 2
Mean	3.169	3.1625
Variance	0.007543	0.008584
Observations	10	12
df	9	11
F	0.878757	
P(F<=f) one-tail	0.430011	
F Critical one-tail	0.322322	

Nous retrouvons dans le tableau généré par l'outil les valeurs calculées plus haut.

Nous avons donc une probabilité cumulée de 43% (inférieure donc à 95%) que la variable aléatoire soit plus grande que F ce qui se calcule avec MS Excel:

$$=1-\text{FDIST}(0.878757;9;11)=0.430011$$

Remarquons que la valeur critique de F est bien de 0.322 (inférieur à la valeur calculée). Cela peut-être vérifié avec MS Excel:

$$=\text{FINV}(0.95;9;11)=0.322$$

ou autrement vu:

$$=1-\text{FDIST}(0.322;9;11)=0.05$$

Remarque: Si les variances étaient égales, le F-Test vaudrait donc 1.

S2. Le test bilatéral à la valeur modale sera donnée par le même tableau où nous saisissons (c'est une astuce) un alpha de 0.025. Ce qui donne:

F-Test Two-Sample for Variances		
	A	B
Mean	3.169	3.1625
Variance	0.007543	0.008584
Observations	10	12
df	9	11
F	0.878757	
P(F<=f) one-tail	0.430011	
F Critical one-tail	0.255619	

Si nous multiplions par 2 la valeur P(F<=f) nous avons:

$$0.430011*2=0.860022$$

qui peut être obtenu via:

$$=\text{FTEST}(A2:A11;B2:B13)=0.860022$$

Remarque: Nous appliquons l'option unilatérale quand on ne s'intéresse qu'au dépassement dans un seul sens et l'option bilatérale quand on retient le dépassement en plus ou en moins. C'est à choisir d'après l'expérience qu'on étudie.

EXERCICE 48.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Test d'ajustement du khi-deux (loi de Poisson)

Énoncé:

Nous avons les mesures suivantes des déviations d'une valeur nominale en valeur absolue:

Ecart-en microns	F(i)
0	5
1	12
2	37
3	56
4	72
5	75
6	53
7	45
8	28
9	11
10	6

Veillez déterminer avec MS Excel si les données ci-dessus peuvent être modélisées par une loi de Poisson avec une moyenne $\mu = 5$ (ceci dit si le lecteur calcule la moyenne des écarts il tombera sur la valeur de 4.8325).

Donnez le seuil de confiance de l'ajustement et donnant le maximum d'informations.

Tracez le graph. Prob.-Prob. des données obtenues.

Solution:

D'abord, nous voyons que certaines classes ont moins de 5 échantillons or c'est insuffisant selon le modèle du test d'ajustement du khi-deux. Nous devons donc regrouper les deux classes extrêmes.

Nous pouvons facilement calculer avec la fonction SOMME() de MS Excel que nous avons mesuré 400 échantillons.

Nous allons dans un premier temps, comparer les observations expérimentales O aux théoriques E . Pour cela, nous utilisons la fonction:

$$=LOI.POISSON(X;5;0)$$

intégrée à MS Excel et nous obtenons le tableau suivant:

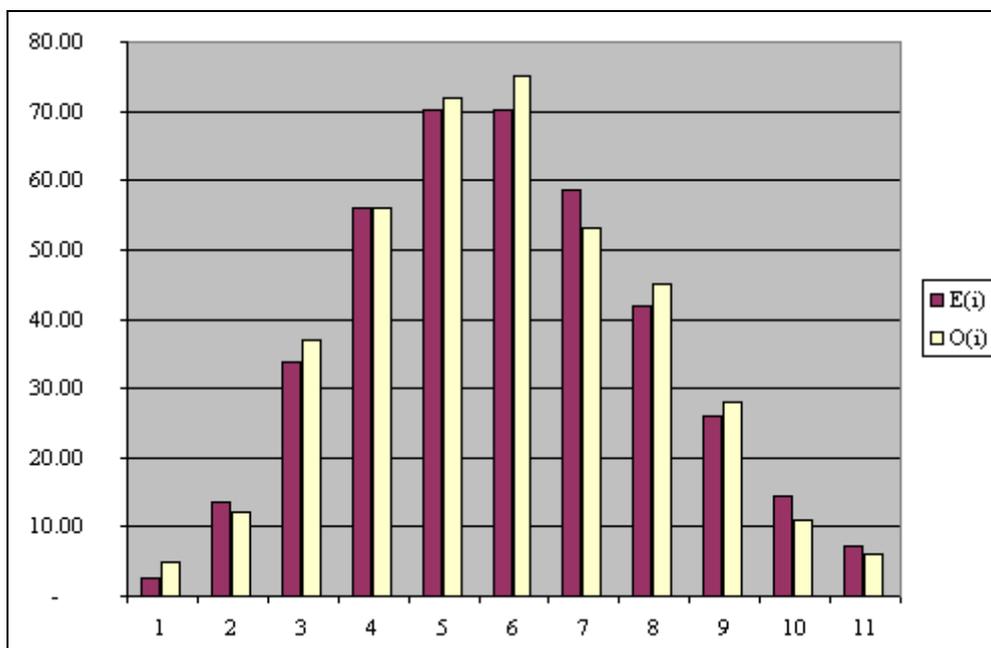
Ecart absolu en microns	$E(i)$	$O(i)$
0	2.70	5.00
1	13.48	12.00
2	33.69	37.00
3	56.15	56.00
4	70.19	72.00
5	70.19	75.00
6	58.49	53.00
7	41.78	45.00
8	26.11	28.00
9	14.51	11.00
10	12.73	6.00
	400.00	400.00

Attention!! le dernier terme 12.73 a été obtenu en faisant le calcul:

$$=(1-LOI.POISSON(9;5;1))*400$$

C'est un piège auquel il faut faire très attention lors de l'utilisation des techniques d'ajustement. Si nous n'avions pas fait cela, la somme de la colonne $E(i)$ ne vaudrait pas 400.

Nous avons aussi le graphique suivant:



Nous regroupons les classes maintenant qui ont moins de 5 échantillons pour obtenir:

Ecart absolu en microns	E(i)	O(i)
1	16.17	17.00
2	33.69	37.00
3	56.15	56.00
4	70.19	72.00
5	70.19	75.00
6	58.49	53.00
7	41.78	45.00
8	26.11	28.00
9	14.51	11.00
10	12.73	6.00
	400.00	400.00

Nous calculons maintenant la distance du khi-deux:

$$D = \sum_{i=1}^N \frac{(O - E)^2}{E}$$

Nous avons alors:

Ecart absolu en microns	E(i)	O(i)	$(O(i) - E(i))^2/E(i)$
1	16.17	17.00	0.0425
2	33.69	37.00	0.3253
3	56.15	56.00	0.0004
4	70.19	72.00	0.0468
5	70.19	75.00	0.3301
6	58.49	53.00	0.5151
7	41.78	45.00	0.2485
8	26.11	28.00	0.1366
9	14.51	11.00	0.8475
10	12.73	6.00	3.5589

et donc 10-1 degrés de liberté et la somme de la colonne orange vaut:

$$D = \sum_{i=1}^N \frac{(O - E)^2}{E} = 6.052$$

Dans MS Excel, calculons la probabilité cumulée de cette valeur. Nous utilisons pour cela la fonction:

$$=CHIDIST(6.052;9)=73.47\%$$

La valeur intéressante est surtout:

$$=1-CHIDIST(6.052;9)=26.5\%$$

Donc la valeur est bien dessous des 95% cumulés habituels et l'ajustement par une loi de Poisson est donc accepté

Il est aussi possible d'obtenir cette valeur en utilisant une autre fonction MS Excel:

$$=1-\text{CHITEST}(O(i);E(i))= 26.5\%$$

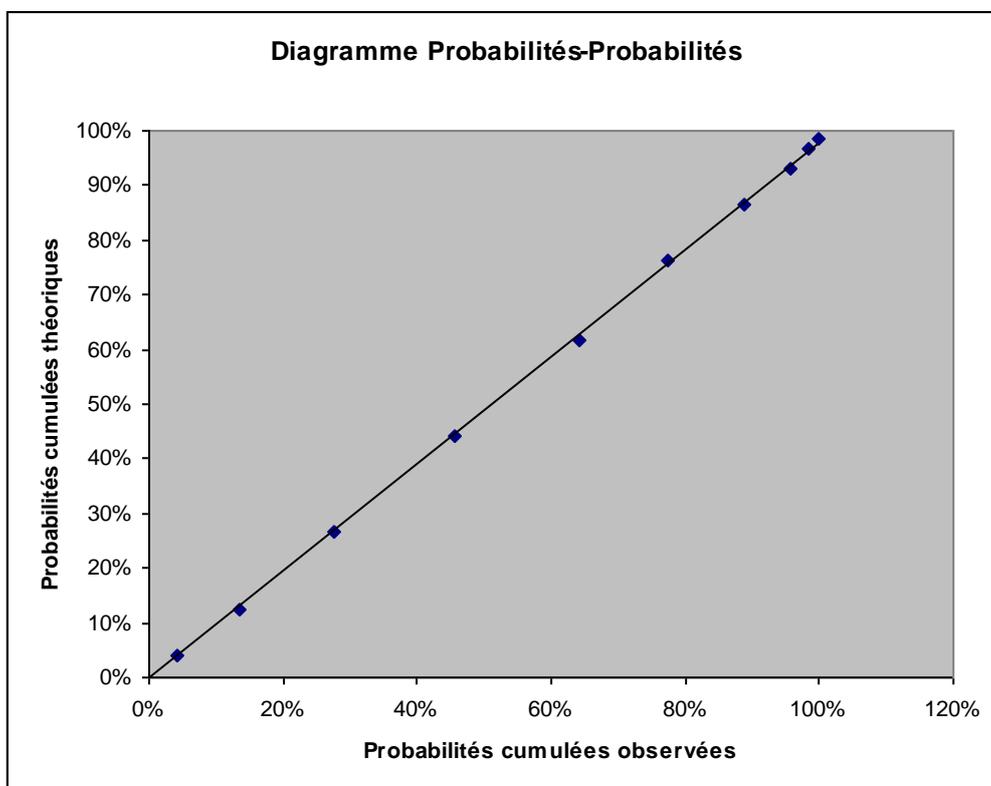
Nous pouvons vérifier quelle est la valeur limite de D correspondant aux 95% cumulés en écrivant:

$$=\text{CHIINV}(0.05;9)=16.92$$

Nous avons par ailleurs aussi:

$$=\text{CHIINV}(73.47\%;9)=6.052$$

Le graphique Prob-Prob est lui donné par:



EXERCICE 49.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Test d'ajustement du khi-deux (loi exponentielle)

Énoncé:

Une société vend des systèmes électromécaniques mis sous fortes contraintes et étudie leur fiabilité sur une durée de 4 ans maximum. Un échantillon test de 90 appareils a été mis sur le banc d'essai.

Les résultats suivants ont été obtenus:

	A	B
1	Mois	Survivants
2	3	32
3	6	16
4	9	18
5	12	9
6	15	4
7	18	4
8	21	2
9	24	2
10	27	1
11	30	1
12	33	1
13	36	0

Nous cherchons à déterminer si l'hypothèse d'une loi exponentielle à paramètre constant est acceptable avec un niveau de confiance de 95%.

Solution:

D'abord, nous savons pour la loi exponentielle que:

$$\lambda = \frac{1}{\mu} = \frac{1}{\sigma}$$

et que l'estimateur de maximum de vraisemblance de l'espérance de la loi exponentielle est donnée par la moyenne arithmétique.

Nous obtenons alors simplement avec MS Excel:

$$=\text{MOYENNE}(B2:B13)=7.5$$

Nous pouvons alors construire directement le tableau suivant:

	A	B	C	D	E
1	Mois	Survivants O(i)	Loi exponentielle Cumulé	Valeur associée E(i)	$(O(i)-E(i))^2/E(i)$
2	3	32	0.33	30	0.18
3	6	16	0.55	20	0.76
4	9	18	0.70	13	1.63
5	12	9	0.80	9	0.00
6	15	4	0.86	6	0.66
7	18	4	0.91	4	0.00
8	21	2	0.94	3	0.18
9	24	2	0.96	2	0.02
10	27	1	0.97	1	0.04
11	30	1	0.98	1	0.04
12	33	1	0.99	1	0.38
13	36	0	0.99	0	0.36

obtenu avec les formules suivantes:

	A	B	C	D	E
1	Mois	Survivants O(i)	Loi exponentielle Cumulé	Valeur associée E(i)	$(O(i)-E(i))^2/E(i)$
2	3	32	=EXPONDIST(A2;1/7.5;1)	=C2*90	=(D2-B2)^2/D2
3	6	16	=EXPONDIST(A3;1/7.5;1)	=(C3-C2)*90	=(D3-B3)^2/D3
4	9	18	=EXPONDIST(A4;1/7.5;1)	=(C4-C3)*90	=(D4-B4)^2/D4
5	12	9	=EXPONDIST(A5;1/7.5;1)	=(C5-C4)*90	=(D5-B5)^2/D5
6	15	4	=EXPONDIST(A6;1/7.5;1)	=(C6-C5)*90	=(D6-B6)^2/D6
7	18	4	=EXPONDIST(A7;1/7.5;1)	=(C7-C6)*90	=(D7-B7)^2/D7
8	21	2	=EXPONDIST(A8;1/7.5;1)	=(C8-C7)*90	=(D8-B8)^2/D8
9	24	2	=EXPONDIST(A9;1/7.5;1)	=(C9-C8)*90	=(D9-B9)^2/D9
10	27	1	=EXPONDIST(A10;1/7.5;1)	=(C10-C9)*90	=(D10-B10)^2/D10
11	30	1	=EXPONDIST(A11;1/7.5;1)	=(C11-C10)*90	=(D11-B11)^2/D11
12	33	1	=EXPONDIST(A12;1/7.5;1)	=(C12-C11)*90	=(D12-B12)^2/D12
13	36	0	=EXPONDIST(A13;1/7.5;1)	=(C13-C12)*90	=(D13-B13)^2/D13

Si nous faisons la somme des distance nous avons pour:

$$=SOMME(E2:E13)=4.27$$

et donc:

$$=1-CHIDIST(4.27;11)=1-CHITEST(B2:B13;D2:D13)=1-96.14%=3.85\%$$

avec la p-value:

$$=CHIINV(0.05;11)=19.67$$

Comme le test est unilatéral et que nous avons la valeur 3.85% qui est bien inférieur aux 95%. Nous acceptons l'hypothèse.

EXERCICE 50.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Méthode ABC et courbe de Lorenz

Énoncé:

Soit les données suivantes:

	A	B	C
1	N° Machine	Arrêt [h.]	Nb. de pannes
2	1	100	4
3	2	32	5
4	3	50	4
5	4	19	2
6	5	4	3
7	6	30	4
8	7	40	12
9	8	80	15
10	9	55	8
11	10	150	8
12	11	160	14
13	12	5	8
14	13	10	3
15	14	20	3

Déterminez les machines engendrant 80% des coûts horaires cumulés et approchez les résultats à une loi de probabilité de Pareto.

Solution:

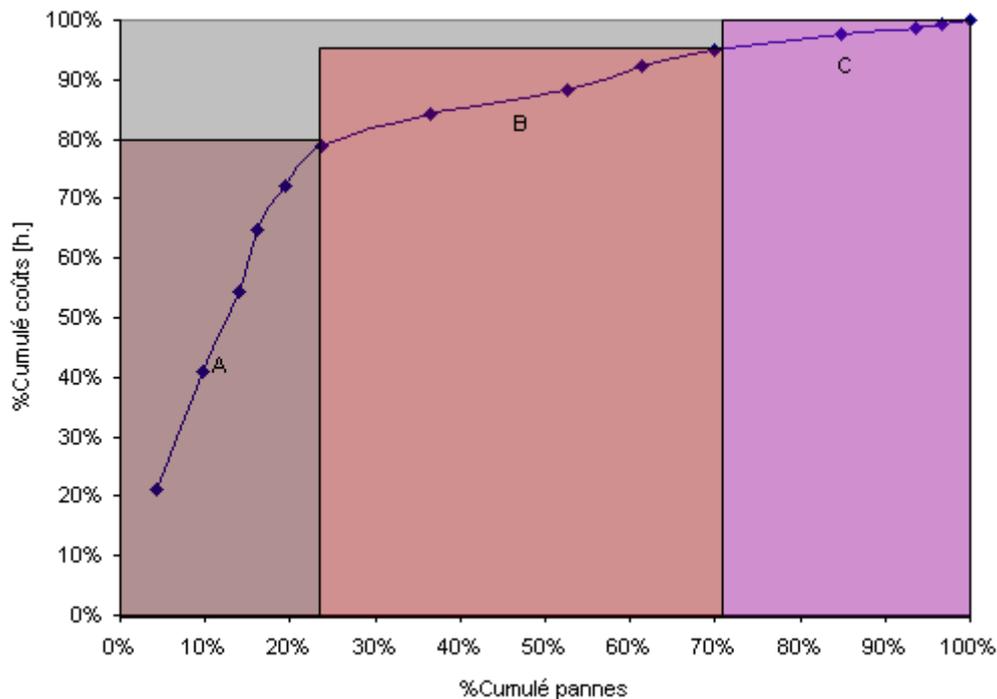
Dans MS Excel ou ailleurs nous devons d'abord (aisément) établir le tableau suivant:

	A	B	C	D	E	F	G
1	N° Machine	Arrêt [h.]	Cumul coûts [h.]	%Cumulé	Nb. de pannes	Cumul pannes	%Cumulé
2	11	160	160	21.19%	4	4	4.30%
3	10	150	310	41.06%	5	9	9.68%
4	1	100	410	54.30%	4	13	13.98%
5	8	80	490	64.90%	2	15	16.13%
6	9	55	545	72.19%	3	18	19.35%
7	3	50	595	78.81%	4	22	23.66%
8	7	40	635	84.11%	12	34	36.56%
9	2	32	667	88.34%	15	49	52.69%
10	6	30	697	92.32%	8	57	61.29%
11	14	20	717	94.97%	8	65	69.89%
12	4	19	736	97.48%	14	79	84.95%
13	13	10	746	98.81%	8	87	93.55%
14	12	5	751	99.47%	3	90	96.77%
15	5	4	755	100.00%	3	93	100.00%

avec les relations suivantes:

	A	B	C	D	E	F
1	N° Machine	Arrêt [h.]	Cumul coûts [h.]	%Cumulé	Nb. de pannes	Cumul pannes
2	11	160	=B2	=C2/\$C\$15	4	=E2
3	10	150	=B3+C2	=C3/\$C\$15	5	=F2+E3
4	1	100	=B4+C3	=C4/\$C\$15	4	=F3+E4
5	8	80	=B5+C4	=C5/\$C\$15	2	=F4+E5
6	9	55	=B6+C5	=C6/\$C\$15	3	=F5+E6
7	3	50	=B7+C6	=C7/\$C\$15	4	=F6+E7
8	7	40	=B8+C7	=C8/\$C\$15	12	=F7+E8
9	2	32	=B9+C8	=C9/\$C\$15	15	=F8+E9
10	6	30	=B10+C9	=C10/\$C\$15	8	=F9+E10
11	14	20	=B11+C10	=C11/\$C\$15	8	=F10+E11
12	4	19	=B12+C11	=C12/\$C\$15	14	=F11+E12
13	13	10	=B13+C12	=C13/\$C\$15	8	=F12+E13
14	12	5	=B14+C13	=C14/\$C\$15	3	=F13+E14
15	5	4	=B15+C14	=C15/\$C\$15	3	=F14+E15

Nous avons alors graphiquement:



où les zones A, B et C sont arrondis à des points existants. Ainsi, la zone critique A compte les machines 11, 10, 1, 8, 9 et 3. Une amélioration de la fiabilité de ces machines peut donc procurer jusqu'à 78.8% du temps total cumulé sur les pannes.

Maintenant déterminons les paramètres de la loi de répartition de Pareto:

$$P(X \leq x) = P_i = 1 - \left(\frac{x_m}{x_i} \right)^k$$

Nous devons déterminer x_m et k les autres paramètres nous sont donnés par nos mesures (le tableau).

Nous pouvons jouer de la manière suivante:

$$P_i - P_{i-1} = - \left(\frac{x_m}{x_i} \right)^k + \left(\frac{x_m}{x_{i-1}} \right)^k = x_m^k \left(\frac{1}{x_{i-1}} - \frac{1}{x_i} \right)^k$$

d'où:

$$\log(P_i - P_{i-1}) = \log \left(x_m^k \left(\frac{1}{x_{i-1}} - \frac{1}{x_i} \right)^k \right) = k \log(x_m) + k \log \left(\frac{1}{x_{i-1}} - \frac{1}{x_i} \right)$$

Donc à l'aide de:

$$\log(\Delta P) = k \log\left(\frac{1}{x_{i-1}} - \frac{1}{x_i}\right) + k \log(x_m)$$

on doit pouvoir déterminer les deux paramètres recherchés en considérant l'expression comme l'équation d'une droite dont k est la pente et $k \log(x_m)$ l'ordonnée à l'origine.

$$\log(\Delta P) = ax + b$$

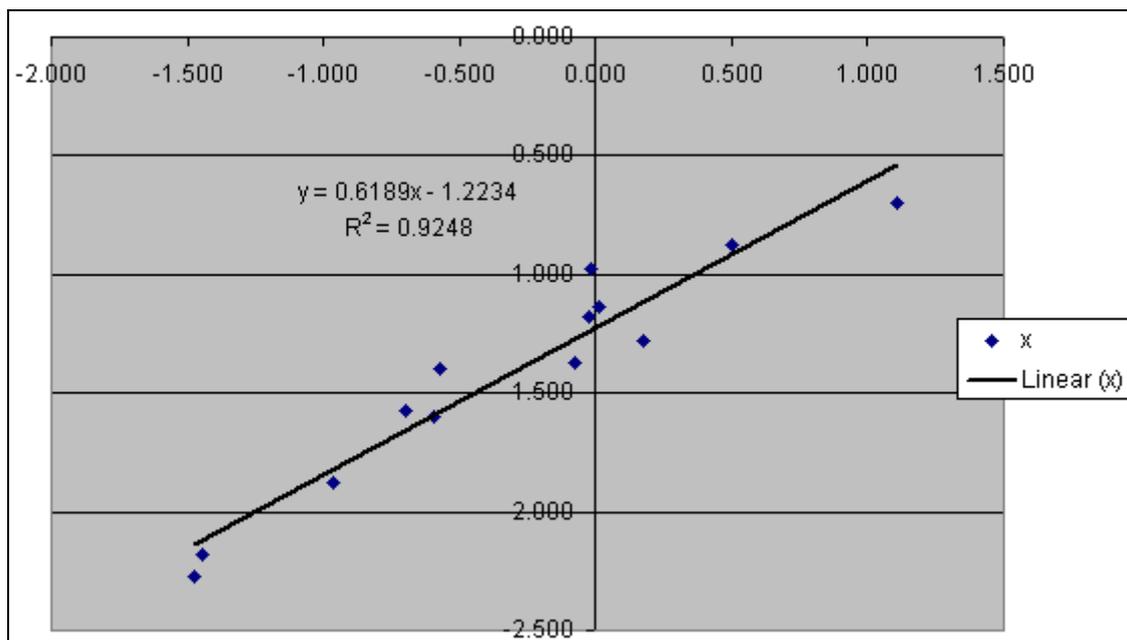
Pour cela dans MS Excel nous rajoutons deux colonnes:

	H	I
1	Log(Delta(P))	x
2	-0.702	1.111
3	-0.878	0.502
4	-0.975	-0.021
5	-1.138	0.014
6	-1.179	-0.027
7	-1.276	0.174
8	-1.373	-0.077
9	-1.401	-0.574
10	-1.577	-0.697
11	-1.599	-0.596
12	-1.878	-0.966
13	-2.179	-1.448
14	-2.276	-1.477
15	#NUM!	#DIV/0!

avec les formules suivantes:

	H	I
1	Log(Delta(P))	x
2	=LOG(D3-D2)	=LOG(1/G2-1/G3)
3	=LOG(D4-D3)	=LOG(1/G3-1/G4)
4	=LOG(D5-D4)	=LOG(1/G4-1/G5)
5	=LOG(D6-D5)	=LOG(1/G5-1/G6)
6	=LOG(D7-D6)	=LOG(1/G6-1/G7)
7	=LOG(D8-D7)	=LOG(1/G7-1/G8)
8	=LOG(D9-D8)	=LOG(1/G8-1/G9)
9	=LOG(D10-D9)	=LOG(1/G9-1/G10)
10	=LOG(D11-D10)	=LOG(1/G10-1/G11)
11	=LOG(D12-D11)	=LOG(1/G11-1/G12)
12	=LOG(D13-D12)	=LOG(1/G12-1/G13)
13	=LOG(D14-D13)	=LOG(1/G13-1/G14)
14	=LOG(D15-D14)	=LOG(1/G14-1/G15)
15	=LOG(D16-D15)	=LOG(1/G15-1/G16)

Une régression linéaire simple (cf. chapitre de Méthodes Numériques) nous donne alors dans un graphique Excel:



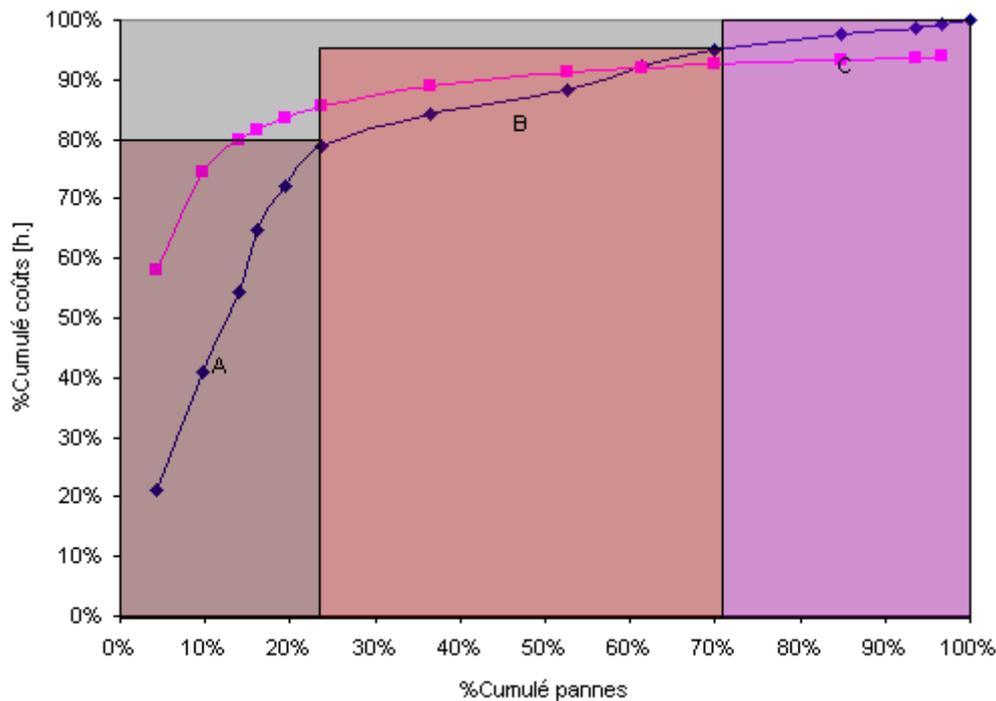
Nous avons donc:

$$k = 0.6189 \text{ et } k \log(x_m) = -1.2234$$

donc:

$$x_m = 10^{\frac{-1.2234}{k}} = 0.0105$$

Nous pouvons alors obtenir la vraie courbe de Pareto correspondante facilement dans MS Excel:



En ayant rajouté la colonne suivante:

	J	J
1	Pareto	Pareto
2	0.580937	=1-(\$I\$23/G2)^\$I\$22
3	0.746304	=1-(\$I\$23/G3)^\$I\$22
4	0.797943	=1-(\$I\$23/G4)^\$I\$22
5	0.815068	=1-(\$I\$23/G5)^\$I\$22
6	0.834802	=1-(\$I\$23/G6)^\$I\$22
7	0.854095	=1-(\$I\$23/G7)^\$I\$22
8	0.888555	=1-(\$I\$23/G8)^\$I\$22
9	0.911114	=1-(\$I\$23/G9)^\$I\$22
10	0.919056	=1-(\$I\$23/G10)^\$I\$22
11	0.925375	=1-(\$I\$23/G11)^\$I\$22
12	0.933862	=1-(\$I\$23/G12)^\$I\$22
13	0.937695	=1-(\$I\$23/G13)^\$I\$22
14	0.938988	=1-(\$I\$23/G14)^\$I\$22

La différence entre la courbe expérimentale et théorique est non négligeable... Par ailleurs, comme k est inférieur à 1, alors comme nous l'avons vu démontré dans le chapitre de Statistiques, la loi de Pareto n'a ni espérance, ni variance.

Il faut faire attention au fait que dans le domaine de la gestion la loi de Pareto est utilisée un peut à torts et à travers alors que une autre loi de probabilité pourrait être beaucoup plus adaptée.

Par ailleurs, un simple test d'ajustement du Khi-Deux (cf. chapitre de Statistiques) nous montre qu'il faut rejeter l'approximation par une loi de Pareto. Par ailleurs des logiciels spécialisés comme @Risk rejettent l'approximation par Pareto au-delà des 20 meilleurs ajustements, le meilleur ajustement étant selon ce même logiciel une loi log-normale.

EXERCICE 51.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Analyse de la variance à un facteur (ANOVA)

Énoncé:

Imaginons une entreprise faisant les trois huit. Nous avons trois équipes qui travaillent sur une même machine. Nous souhaitons vérifier avec un seuil de confiance de 95% s'il y a une différence de productivité moyenne entre les trois équipes sur une semaine de travail.

Nous possédons les données suivantes:

	A	B	C	D
1		Equipe 1	Equipe 2	Equipe 3
2	Lundi	78	77	88
3	Mardi	88	75	86
4	Mercredi	90	80	79
5	Jeudi	77	83	93
6	Vendredi	85	87	79
7	Samedi	88	90	83
8	Dimanche	79	85	79

Nous demandons de détailler les étapes du calcul de l'analyse de la variance de l'hypothèse des égalités et d'utiliser l'outil intégré à MS Excel et d'en tirer les conclusions adéquates

Solution:

Commençons par la méthode détaillée. Pour cela, nous avons vu dans la partie théorique du site que l'ensemble des calculs pouvaient se résumer au tableau suivant:

Source	Somme des carrés	ddl du χ^2	Moyenne des carrés	F	Valeur critique F
Inter-Classe	$Q_A = \sum_i n_i (\bar{x}_i - \bar{\bar{x}})^2$	k-1	$MSE_k = \frac{Q_A}{k-1}$	$\frac{MSE_k}{MSE}$	$P(F > F_{k-1, N-k})$
Intra-Classe	$Q_R = \sum_{\check{y}} (x_{\check{y}} - \bar{x}_i)^2$	N-k	$MSE = \frac{Q_R}{N-k}$		
Total	$Q_T = \sum_{\check{y}} (x_{\check{y}} - \bar{\bar{x}})^2$	N-1			

Nous allons donc nous mettre à la tâche. D'abord calculons les sommes des carrés:

$$Q_A = \sum_i n_i (\bar{x}_i - \bar{\bar{x}})^2 = \sum_{i=1}^3 n_i (\bar{x}_i - \bar{\bar{x}})^2$$

Nous avons dans le cas qui nous concerne:

$$n_1 = n_2 = n_3 = 7$$

Le calcul de la moyenne des tous les échantillons donne immédiatement:

$$\bar{\bar{x}} = 83.28571$$

et nous avons:

$$\bar{x}_1 = 83.571 \quad \bar{x}_2 = 82.429 \quad \bar{x}_3 = 83.847$$

Nous avons alors:

$$(\bar{x}_1 - \bar{\bar{x}}) = 0.286 \quad (\bar{x}_2 - \bar{\bar{x}}) = -0.857 \quad (\bar{x}_3 - \bar{\bar{x}}) = 0.571$$

et:

$$(\bar{x}_1 - \bar{\bar{x}})^2 = 0.082 \quad (\bar{x}_2 - \bar{\bar{x}})^2 = 0.735 \quad (\bar{x}_3 - \bar{\bar{x}})^2 = 0.327$$

et:

$$Q_A = \sum_{i=1}^3 n_i (\bar{x}_i - \bar{\bar{x}})^2 = 7 \cdot 1.142 \cong 8$$

Ensuite, calculons:

$$Q_R = \sum_{ij} (x_{ij} - \bar{x}_i)^2 = \sum_{i=1}^3 \sum_{j=1}^7 (x_{ij} - \bar{x}_i)^2$$

Nous avons d'abord:

$$(x_{ij} - \bar{x}_i)$$

Qui donnera:

	A	B	C	D
11		Equipe 1	Equipe 2	Equipe 3
12	Lundi	-5.571	-5.429	4.143
13	Mardi	4.429	-7.429	2.143
14	Mercredi	6.429	-2.429	-4.857
15	Jeudi	-6.571	0.571	9.143
16	Vendredi	1.429	4.571	-4.857
17	Samedi	4.429	7.571	-0.857
18	Dimanche	-4.571	2.571	-4.857

Ensuite, nous mettons tout cela au carré:

$$(x_{ij} - \bar{x}_i)^2$$

Qui donnera:

	A	B	C	D
21		Equipe 1	Equipe 2	Equipe 3
22	Lundi	31.041	29.469	17.163
23	Mardi	19.612	55.184	4.592
24	Mercredi	41.327	5.898	23.592
25	Jeudi	43.184	0.327	83.592
26	Vendredi	2.041	20.898	23.592
27	Samedi	19.612	57.327	0.735
28	Dimanche	20.898	6.612	23.592

Nous avons alors (le calcul est donné par chaque colonne):

$$Q_R = \sum_{i=1}^3 \sum_{j=1}^7 (x_{ij} - \bar{x}_i)^2 = 177.714 + 175.714 + 176.857 = 530.285$$

Indiquons que les trois termes ci-dessus peuvent directement être obtenus avec la fonction DEVSQ() de MS Excel.

Maintenant nous avons le dernier:

$$Q_T = \sum_{ij} (x_{ij} - \bar{\bar{x}})^2 = Q_R = \sum_{i=1}^3 \sum_{j=1}^7 (x_{ij} - \bar{\bar{x}})^2$$

Rappelons que nous avons obtenu:

$$\bar{\bar{x}} = 83.28571$$

Nous avons la table suivante alors pour:

$$(x_{ij} - \bar{\bar{x}})$$

	A	B	C	D
31		Equipe 1	Equipe 2	Equipe 3
32	Lundi	-5.286	-6.286	4.714
33	Mardi	4.714	-8.286	2.714
34	Mercredi	6.714	-3.286	-4.286
35	Jeudi	-6.286	-0.286	9.714
36	Vendredi	1.714	3.714	-4.286
37	Samedi	4.714	6.714	-0.286
38	Dimanche	-4.286	1.714	-4.286

et la table suivante pour:

$$(x_{ij} - \bar{\bar{x}})^2$$

	A	B	C	D
41		Equipe 1	Equipe 2	Equipe 3
42	Lundi	27.939	39.510	22.224
43	Mardi	22.224	68.653	7.367
44	Mercredi	45.082	10.796	18.367
45	Jeudi	39.510	0.082	94.367
46	Vendredi	2.939	13.796	18.367
47	Samedi	22.224	45.082	0.082
48	Dimanche	18.367	2.939	18.367

Nous avons alors (le calcul est donné par chaque colonne):

$$Q_T = \sum_{i=1}^3 \sum_{j=1}^7 (x_{ij} - \bar{\bar{x}})^2 = 178.286 + 180.857 + 179.143 = 538.286$$

et nous avons aussi:

$$k - 1 = 2 \quad N - k = 21 - 3 = 18 \quad N - 1 = 20$$

Nous avons alors:

$$MSk = \frac{Q_A}{k - 1} = \frac{8}{2} = 4$$

$$MSE = \frac{Q_R}{N - k} = \frac{530.285}{18} = 29.4603$$

Nous avons pour la fonction de masse de Fisher (fonction de distribution):

$$F_{k-1, N-k} = \frac{MSk}{MSE} = 0.1358$$

Maintenant que nous avons cette valeur, nous devons la comparer en test unilatéral à la loi de Fisher avec les degrés de liberté 2 et 18. Les tables de Fisher donnent une valeur de 3.55 pour un seuil de confiance de 95% (probabilité cumulée) et MS Excel nous donne la même chose:

$$=FINV(1-95\%;2;18)=FINV(5\%;2;18)=3.5545$$

Donc nous avons 95% de probabilité cumulée d'être entre 0 et 3.5545, ce dernier chiffre étant la valeur critique du test de Fisher appelé "*F-critique*"!

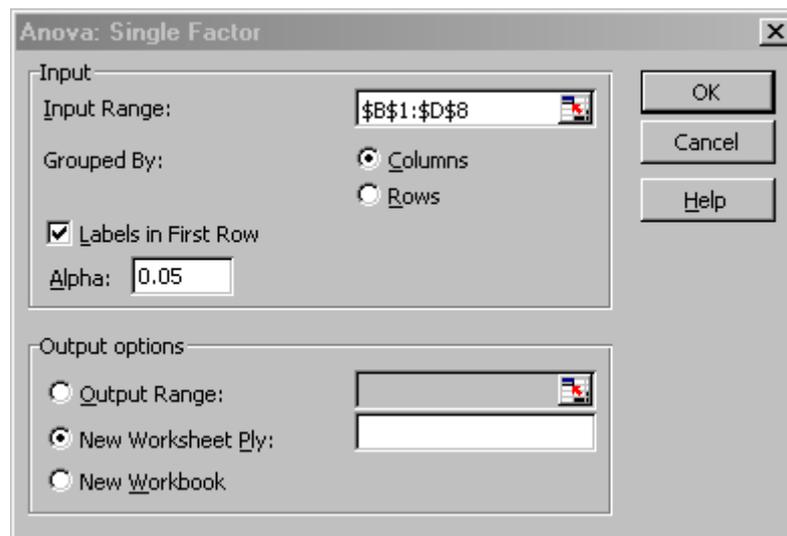
Il s'agit donc de la valeur au-dessus de laquelle il faut être pour ne pas avoir à rejeter l'hypothèse!

Remarquons que la probabilité cumulée associée à 0.1358 est donné dans MS Excel 2007 et antérieur (la fonction a enfin été corrigée pour donner immédiatement la probabilité cumulée dans MS Excel 2010 sans avoir à faire le 1-...) par:

$$=1-FDIST(0.1358;2;18)=1-0.8739=1-87.39\%=12.609\%$$

Donc... nous ne pouvons pas dans le cas présent rejeter l'hypothèse comme quoi les moyennes des productions sont égales.

Avec l'utilitaire d'Analyse de MS Excel cela donne:



Ce qui génère:

	A	B	C	D	E	F	G
1	Anova: Single Factor						
2							
3	SUMMARY						
4	<i>Groups</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>		
5	Equipe 1	7	585	83.57143	29.61905		
6	Equipe 2	7	577	82.42857	29.28571		
7	Equipe 3	7	587	83.85714	29.47619		
8							
9							
10	ANOVA						
11	<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
12	Between Groups	8	2	4	0.135776	0.873924	3.554557
13	Within Groups	530.2857	18	29.46032			
14							
15	Total	538.2857	20				

Il est préférable de se baser sur la p -value qui résume dans le cas présent, le fait que la probabilité cumulée est inférieure à 95%. Comme à l'habitude, si la p -value est inférieure à 5%, il faudra donc rejeter l'hypothèse.

EXERCICE 52.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Analyse de la variance à deux facteurs sans répétitions (ANOVA)

Énoncé:

Considérons une entreprise ayant à sa disposition deux équipes de travail et trois machines différentes pour la fabrication de même pièces. Sous l'hypothèse qu'il n'y a pas d'interactions Machines*Équipes nous souhaiterions étudier les effets des deux facteurs.

Voici la production obtenue:

	A	B	C
1		Equipe 1	Equipe 2
2	Machine 1	256	242
3	Machine 2	246	238
4	Machine 3	254	246

Solution:

Pour appliquer le tableau vu dans le cours théorique:

Source	Somme des carrés	ddl χ^2	Moyenne des carrés	F	Valeur critique F
Inter-Classe1	$Q_{A1} = r \sum_i (\bar{x}_i - \bar{\bar{x}})^2$	r-1	$MSk1 = \frac{Q_{A1}}{r-1}$	$\frac{MSk1}{MSE}$	$P(F > F_{r-1, (k-1)(r-1)})$
Inter-Classe2	$Q_{A2} = k \sum_j (\bar{x}_j - \bar{\bar{x}})^2$	k-1	$MSk1 = \frac{Q_{A2}}{k-1}$	$\frac{MSk2}{MSE}$	$P(F > F_{k-1, (k-1)(r-1)})$
Intra-Classe	$Q_R = \sum_{ij} (x_{ij} - \bar{x}_i - \bar{x}_j + \bar{\bar{x}})^2$	(k-1)(r-1)	$MSE = \frac{Q_R}{(k-1)(r-1)}$		
Total	$Q_T = \sum_{ij} (x_{ij} - \bar{\bar{x}})^2$	N-1			

Nous allons donc calculer les moyennes arithmétiques des colonnes \bar{x}_j , des lignes \bar{x}_i et globale moyenne $\bar{\bar{x}}$. Ce qui donne:

	A	B	C	D
1		Equipe 1	Equipe 2	Moyenne
2	Machine 1	256	242	249
3	Machine 2	246	238	242
4	Machine 3	254	246	250
5	Moyenne	252	242	247

et nous avons dans le tableau ci-dessus le nombre de lignes $k = 3$ et le nombre de colonnes $r = 2$.

Nous avons donc pour les machines:

$$Q_{A1} = r \sum_i (\bar{x}_i - \bar{\bar{x}})^2 = 2 \sum_{i=1}^3 (\bar{x}_i - 247)^2 = 2 \cdot \left((249 - 247)^2 + (242 - 247)^2 + (250 - 247)^2 \right) = 76$$

et pour les équipes:

$$Q_{A2} = k \sum_j (\bar{x}_j - \bar{\bar{x}})^2 = 3 \sum_{j=1}^2 (\bar{x}_j - 247)^2 = 3 \cdot \left((252 - 247)^2 + (242 - 247)^2 \right) = 150$$

et pour l'intra-classe:

$$Q_R = \sum_{ij} (x_{ij} - \bar{x}_i - \bar{x}_j - \bar{\bar{x}})^2$$

Source	Somme des carrés	ddl χ^2	Moyenne des carrés F	Valeur critique F
Inter-Classe1 (machines)	76	2	38	6.333 19
Inter-Classe2 (équipes)	150	1	150	$P(F > F_{1,2})$
Intra-Classe	12	2	6	
Total	238	5		

Maintenant que nous avons ces valeurs, nous devons les comparer en test unilatéral à la loi de Fisher avec les degrés de liberté respectifs.

MS Excel nous donne pour les machines:

$$=FINV(1-95\%;2;2) = FINV(5\%;2;2) = 19$$

Donc nous avons 95% de probabilité cumulée d'être entre 0 et 19, ce dernier chiffre étant la valeur critique du test de Fisher appelé "*F-critique*"!

Il s'agit donc de la valeur en-dessous de laquelle il faut être pour ne pas avoir à rejeter l'hypothèse!

Remarquons que la probabilité cumulée associée à 6.333 est donné dans MS Excel 2007 et antérieur (puisque la fonction a enfin été corrigé pour donner immédiatement la probabilité cumulée dans MS Excel 2010 sans avoir à faire le 1-...):

$$=1-FDIST(6.333;2;2)=1-0.1364=1-13.64\%=86.36\%$$

Donc... nous ne pouvons pas dans le cas présent rejeter l'hypothèse comme quoi les moyennes entre les machines sont égales (donc le type de machine n'influence pas la production).

MS Excel nous donne pour les équipes:

$$=FINV(1-95\%;1;2)=FINV(5\%;2;2)=18.512$$

Donc nous avons 95% de probabilité cumulée d'être entre 0 et 18.512, ce dernier chiffre étant la valeur critique du test de Fisher appelé "*F-critique*"!

Il s'agit donc de la valeur en-dessous de laquelle il faut être pour ne pas avoir à rejeter l'hypothèse!

Remarquons que la probabilité cumulée associée à 25 est donné dans MS Excel 2007 et antérieur (puisque la fonction a enfin été corrigé pour donner immédiatement la probabilité cumulée dans MS Excel 2010 sans avoir à faire le 1-...):

$$=1-FDIST(25;1;2)=1-0.0377=1-3.77\%=96.23\%$$

Donc... nous devons dans le cas présent rejeter l'hypothèse comme quoi les moyennes entre les équipes sont égales (donc le numéro d'équipe influence la production).

Avec l'utilitaire d'Analyse de MS Excel cela donne:



Ce qui génère:

	A	B	C	D	E	F	G
1	Anova: Two-Factor Without Replication						
2							
3	<i>SUMMARY</i>	<i>Count</i>	<i>Sum</i>	<i>Average</i>	<i>Variance</i>		
4	Machine 1	2	498	249	98		
5	Machine 2	2	484	242	32		
6	Machine 3	2	500	250	32		
7							
8	Equipe 1	3	756	252	28		
9	Equipe 2	3	726	242	16		
10							
11							
12	ANOVA						
13	<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
14	Rows	76	2	38	6.333333	0.136364	19
15	Columns	150	1	150	25	0.03775	18.51282
16	Error	12	2	6			
17							
18	Total	238	5				
19							

Il est préférable de se baser sur la P -value qui résume dans le cas présent, le fait que la probabilité cumulée est inférieure à 95% pour les lignes et supérieure à 95% pour les colonnes. Comme à l'habitude, si la P -value est inférieure à 5%, il faudra rejeter l'hypothèse.

Ainsi:

1. La P -value pour les lignes (machines) étant de ~13% les machines n'influencent pas la production de manière significative.
2. La P -value pour les colonnes (équipes) étant de ~3% les équipes influencent la productivité de manière significative.

EXERCICE 53.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Analyse de la variance à deux facteurs avec répétitions (ANOVA)

Énoncé:

Considérons une entreprise ayant à sa disposition deux équipes de travail et trois machines différentes pour la fabrication de pièces identiques. Nous souhaiterions étudier les effets des deux facteurs à l'aide de mesures répétées sur deux journées.

Voici la production obtenue (plan équilibré):

	A	B	C	D
1		Equipe 1	Equipe 2	
2	Machine 1	256	242	Jour 1
3		251	240	Jour 2
4	Machine 2	246	238	Jour 1
5		248	240	Jour 2
6	Machine 3	254	246	Jour 1
7		251	246	Jour 2

Et nous utilisons le tableau démontré dans le cours théorique:

Somme des carrés	ddl χ^2	Moyenne des carrés	F	Valeur critique F
$Q_B = nr \sum_i (\bar{x}_i - \bar{\bar{x}})^2$	r-1	$MSkB = \frac{Q_B}{r-1}$	$\frac{MSkB}{MSE}$	$P(F > F_{r-1, N-kr})$
$Q_A = nk \sum_j (\bar{x}_j - \bar{\bar{x}})^2$	k-1	$MSkA = \frac{Q_A}{k-1}$	$\frac{MSkA}{MSE}$	$P(F > F_{k-1, N-kr})$
$Q_{A \times B} = n \sum_{ij} (\bar{x}_{ij} - \bar{x}_i - \bar{x}_j + \bar{\bar{x}})^2$	(k-1)(r-1)	$MSkAB = \frac{Q_{A \times B}}{(k-1)(r-1)}$	$\frac{MSkAB}{MSE}$	$P(F > F_{(k-1)(r-1), N-kr})$
$Q_R = \sum_{mij} (x_{mij} - \bar{x}_{ij})^2$	N-kr	$MSE = \frac{Q_R}{N-kr}$		
$Q_T = \sum_{mij} (x_{mij} - \bar{\bar{x}})^2$	N-1			

Nous allons donc calculer les moyennes arithmétiques des colonnes \bar{x}_j , des lignes d'échantillons \bar{x}_i , des mesures intra-échantillons \bar{x}_{ij} et enfin de la moyenne globale $\bar{\bar{x}}$. Ce qui donne:

	A	B	C	D
1		Equipe 1	Equipe 2	Moyenne
2	Machine 1	256.00	242.00	
3		251.00	240.00	
4	Moyenne	253.50	241.00	247.25
5	Machine 2	246.00	238.00	
6		248.00	240.00	
7	Moyenne	247.00	239.00	243.00
8	Machine 3	254.00	246.00	
9		251.00	246.00	
10	Moyenne	252.50	246.00	249.25
11	Moyenne Générale	251.00	242.00	246.50

soit explicitement:

	A	B	C	D
1		Equipe 1	Equipe 2	Moyenne
2	Machine 1	256	242	
3		251	240	
4	Moyenne	=AVERAGE(B2:B3)	=AVERAGE(C2:C3)	=AVERAGE(B4:C4)
5	Machine 2	246	238	
6		248	240	
7	Moyenne	=AVERAGE(B5:B6)	=AVERAGE(C5:C6)	=AVERAGE(B7:C7)
8	Machine 3	254	246	
9		251	246	
10	Moyenne	=AVERAGE(B8:B9)	=AVERAGE(C8:C9)	=AVERAGE(B10:C10)
11	Moyenne Générale	=AVERAGE(B4;B7;B10)	=AVERAGE(C4;C7;C10)	=AVERAGE(B11:C11)

et nous avons dans le tableau ci-dessus le nombre de lignes répétitions $n = 2$ (deux jours à chaque fois) et le nombre de colonnes $r = 2$ (deux équipes) et le nombre d'échantillons $k = 3$ (3 machines).

Nous allons exceptionnellement pas calculer la table de cette ANOVA à la main. Donc si nous le faisons explicitement avec les formules du logiciel MS Excel, nous obtenons:

	A	B	C	D	E	F	G
1		Equipe 1	Equipe 2	Moyenne			
2	Machine 1	256.00	242.00				
3		251.00	240.00				
4	Moyenne	253.50	241.00	247.25			
5	Machine 2	246.00	238.00				
6		248.00	240.00				
7	Moyenne	247.00	239.00	243.00			
8	Machine 3	254.00	246.00				
9		251.00	246.00				
10	Moyenne	252.50	246.00	249.25			
11	Moyenne Générale	251.00	242.00	246.50			
12							
13		SS	ddl	MS	F	p -value	F crit
14	Q _A	81.50	2	40.75	10.63043478	0.0107	5.14325285
15	Q _B	243.00	1	243.00	63.39130435	0.0002	5.987377607
16	Q _{AxB}	19.500	2	9.75	2.543478261	0.1585	5.14325285
17	Q _γ	23.00	6	3.83			

Avec les détails de la table de l'ANOVA calculée à la main d'abord pour les variances:

	A	B
13		SS
14	Q _A	=2*2*(D4-D11)^2+(D7-D11)^2+(D10-D11)^2
15	Q _B	=2*3*(B11-D11)^2+(C11-D11)^2
16	Q _{AxB}	=2*(B4-D4-B11+D11)^2+(C4-D4-C11+D11)^2+(B7-D7-B11+D11)^2+(C7-D7-C11+D11)^2+(B10-D10-B11+D11)^2+(C10-D10-C11+D11)^2
17	Q _γ	=(B2-B4)^2+(C2-C4)^2+(B3-B4)^2+(C3-C4)^2+(B5-B7)^2+(C5-C7)^2+(B6-B7)^2+(C6-C7)^2+(B8-B10)^2+(C8-C10)^2+(B9-B10)^2+(C9-C10)^2

Et enfin pour les colonnes restantes:

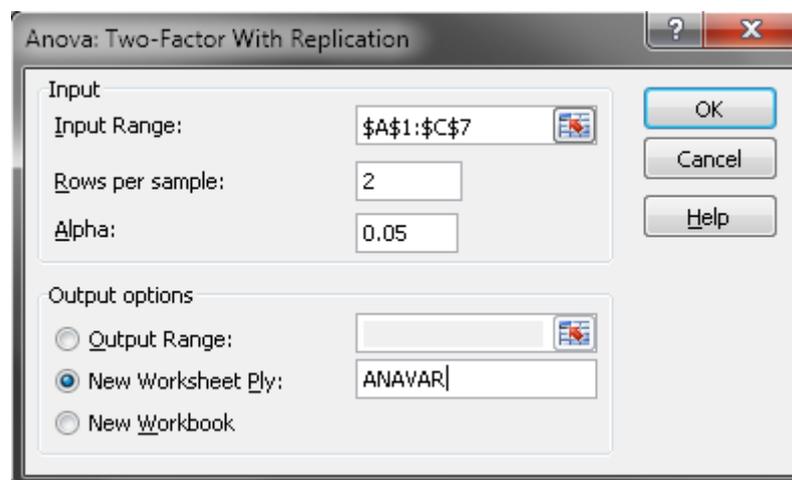
	C	D	E	F	G
13	ddl	MS	F	p -value	F crit
14	=COUNTA(A2:A5;A8)-1	=B14/C14	=D14/\$D\$17	=FDIST(E14;C14;\$C\$17)	=FINV(1-95%;C14;\$C\$17)
15	=COUNTA(B1:C1)-1	=B15/C15	=D15/\$D\$17	=FDIST(E15;C15;\$C\$17)	=FINV(1-95%;C15;\$C\$17)
16	=C15*C14	=B16/C16	=D16/\$D\$17	=FDIST(E16;C16;\$C\$17)	=FINV(1-95%;C16;\$C\$17)
17	=COUNT(B2:C3;B5:C6;B8:C9)-3*2	=B17/C17			

Au vu des résultats obtenus (les valeurs des *p*-values). Donc... nous rejettons l'hypothèse nulle comme quoi les moyennes sont égales pour l'influence des équipes seules ou des machines seules (ce qui est bon signe pour le patron!). Par contre, en ce qui concerne le terme d'interaction, nous ne pouvons pas rejeter le fait que le choix de la machine influence le travail des équipes.

Voyons si nous obtenons la même chose avec l'Utilitaire d'Analyse. Nous partons toujours du tableau suivant:

	A	B	C
1		Equipe 1	Equipe 2
2	Machine 1	256	242
3		251	240
4	Machine 2	246	238
5		248	240
6	Machine 3	254	246
7		251	246

Nous exécutons dans l'Utilitaire d'Analyse l'ANOVA à deux facteurs avec réplication avec les paramètres suivants:



et nous obtenons:

	A	B	C	D
1	Anova: Two-Factor With Replication			
2				
3	SUMMARY	Equipe 1	Equipe 2	Total
4	<i>Machine 1</i>			
5	Count	2	2	4
6	Sum	507	482	989
7	Average	253.5	241	247.25
8	Variance	12.5	2	56.91667
9				
10	<i>Machine 2</i>			
11	Count	2	2	4
12	Sum	494	478	972
13	Average	247	239	243
14	Variance	2	2	22.66667
15				
16	<i>Machine 3</i>			
17	Count	2	2	4
18	Sum	505	492	997
19	Average	252.5	246	249.25
20	Variance	4.5	0	15.58333
21				

	A	B	C	D	E	F	G
22	<i>Total</i>						
23	Count	6	6				
24	Sum	1506	1452				
25	Average	251	242				
26	Variance	13.6	11.2				
27							
28							
29	ANOVA						
30	<i>Source of Variation</i>	<i>SS</i>	<i>df</i>	<i>MS</i>	<i>F</i>	<i>P-value</i>	<i>F crit</i>
31	Sample	81.5	2	40.75	10.63043	0.010662	5.143253
32	Columns	243	1	243	63.3913	0.000209	5.987378
33	Interaction	19.5	2	9.75	2.543478	0.158495	5.143253
34	Within	23	6	3.833333			
35							
36	Total	367	11				

nous retrouvons donc les mêmes résultats que ceux calculés à manuellement et démontrés dans le chapitre de Statistiques.

EXERCICE 54.

Niveau: Universitaire (Faculté)

Auteur: Vincent ISOZ (isozv@hotmail.com)

Mots-clés: Maîtrise Statistique des Procédés (SPC) et Six Sigma

Énoncé:

Considérons une nouvelle petite production de $n = 50$ pièces par lot de 10 (afin d'ajuster en cours de production). La mesure de côtes de 5 pièces chaque heure pendant 10 heures avec une tolérance de 10 ± 0.07 soit en termes de centièmes un étendue de:

$$E = 2 \cdot 7 = 14$$

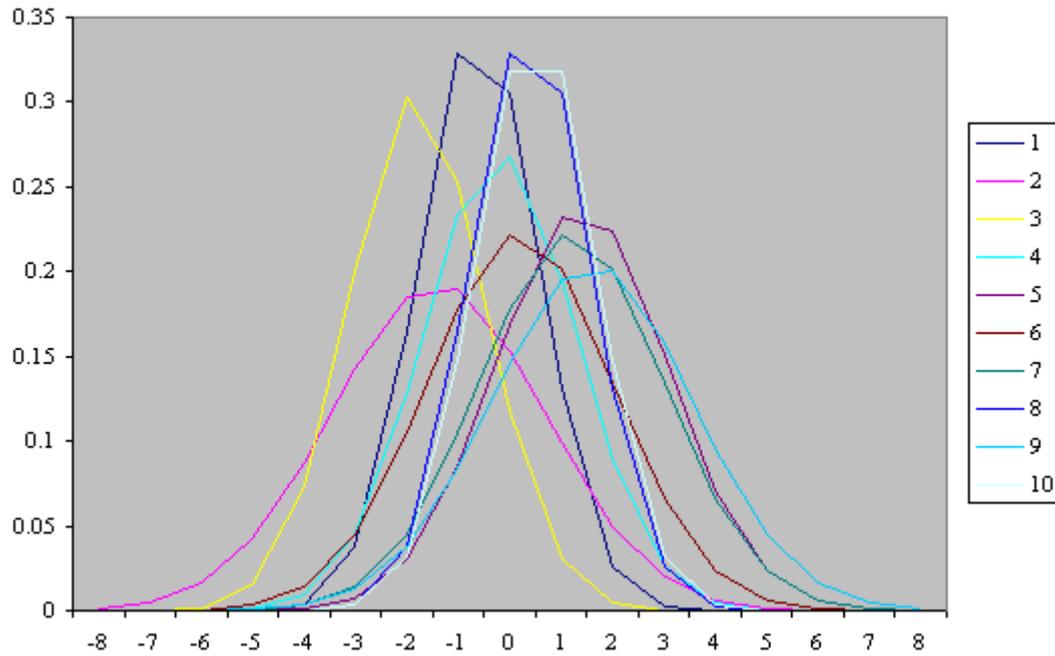
et une cible de $T = 0$ (en termes d'écart). Nous avons les données suivantes:

	1	2	3	4	5	6	7	8	9	10
1	-2	-4	-1	0	4	0	3	0	1	-1
2	0	-3	0	-2	1	-2	0	1	-1	2
3	-1	0	-3	-1	0	0	-1	-1	3	1
4	1	1	-2	2	2	0	1	0	4	0
5	-1	-1	-3	0	0	3	3	2	1	0
μ	-0.6	-1.4	-1.8	-0.2	1.4	0.2	1.2	0.4	1.6	0.5
σ	1.14	2.07	1.30	1.48	1.67	1.79	1.79	1.14	1.95	1.14

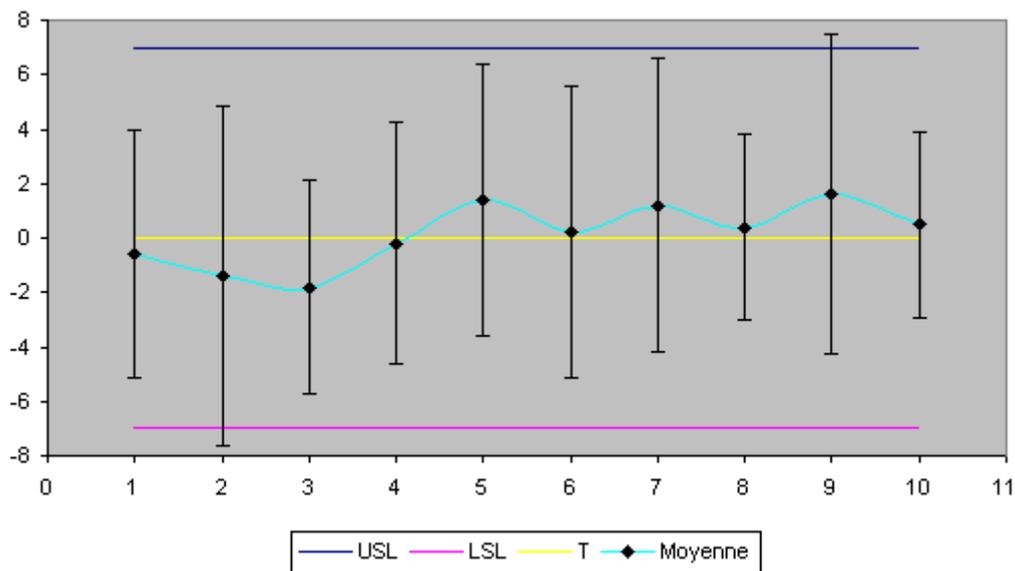
- Déterminez graphiquement les distributions Normales de chacun des lots selon leur écart-type (estimateurs) et leur moyenne expérimentale.
- Représentez la carte de contrôle $\mu - \sigma$ avec la cible T et les limites inférieures et supérieures USL et LSL .
- Calculez \bar{X} , S_{CT} , S_{LT} et leur intervalle de confiance à 95%.
- Calculez $C_{p,observé}$, $P_{p,observé}$, $C_{p,vrai}$, $P_{p,vrai}$, C_{pk} , P_{pk} , C_{pm} , P_{pm} , PPM_{tot}^{LT} , PPM_{tot}^{CT} , PPM_{tot}^{LT}
- Donnez votre avis sur les valeurs obtenues

Solution:

Nous voyons immédiatement que le processus de fabrication a été non-stationnaire pendant cette première production il faudra donc apporter des corrections à l'avenir:



ou sous forme de carte de contrôle avec la représentation d'un écart-type de 3σ (ce qui est suffisant pour des petites quantité des pièces bon marché à fabriquer):



Donc on devine quand même que le processus est limite....

Remarque: Une chose intéressante c'est que l'on peut analyser aussi ce graphique en utilisant les outils mathématiques de l'analyse des séries temporelles aussi (cf. chapitre d'Économétrie).

D'abord, si nous voulons faire une étude statistique pertinente des différentes données ci-dessus nous pouvons calculer la moyenne générale des écarts qui sous l'hypothèse d'une distribution Normale est la moyenne arithmétique (cf. chapitre de Statistique):

$$\bar{X} = 0.12$$

Ensuite l'écart-type des données de toutes les pièces est de:

$$S_{LT} = 1.80$$

en utilisant l'estimateur de maximum de vraisemblance de la variance de la loi Normale:

$$S_{LT} = \sqrt{\frac{\sum_{i,j} (X_{ij} - \bar{X})^2}{n}}$$

nous avons une valeur supérieur à 1 ce qui est donc non-conforme à ce que Six Sigma exige dans son niveau de qualité.

Donc l'erreur-standard (l'estimateur de l'écart-type de la moyenne) est de:

$$\sigma_{\mu} = \frac{S_{LT}}{\sqrt{n}} = 0.25$$

Donc l'intervalle de confiance à 95% de la moyenne est de (cf. chapitre de Statistique):

$$\bar{X} - 1.96 \frac{S_{LT}}{\sqrt{n}} \leq \mu \leq \bar{X} + 1.96 \frac{S_{LT}}{\sqrt{n}}$$

Soit dans notre cas:

$$-0.379 \leq \mu \leq 0.619$$

Et l'inférence statistique avec notre écart-type long terme utilisant le test d'hypothèse bilatéral du χ^2 donne (cf. chapitre de Statistique):

$$\frac{(n-1)S_{LT}^2}{\chi_{\alpha/2}(n-1)} \leq \sigma_{LT}^2 \leq \frac{(n-1)S_{LT}^2}{\chi_{1-\alpha/2}(n-1)}$$

Ce qui nous donne dans notre cas:

$$2.268 \leq \sigma_{LT}^2 \leq 5.047$$

soit:

$$1.506 \leq \sigma_{LT} \leq 2.246$$

Nous remarquons alors que sur une analyse long terme avec les intervalles:

$$-0.644 \leq \mu \leq 0.884$$

$$1.506 \leq \sigma_{LT} \leq 2.246$$

Calculons maintenant la performance globale du procédé long terme (si supposé centré donc!). Nous avons:

$$P_p = \frac{USL - LSL}{6S_{LT}} = \frac{E}{6S_{LT}} = \frac{14}{6 \cdot 1.80} = 1.294$$

Mais avec un instrument ayant un P_{pc} de 4, cela correspond réellement à:

$$P_{p, \text{vrai}} = \frac{P_{p, \text{observé}}}{\sqrt{1 - \frac{P_{p, \text{observé}}}{P_{pc}}}} = 1.574$$

De plus, indiquons que comme nous savons faire un calcul d'intervalle de confiance pour σ_{LT} (voir le calcul fait précédemment), il est alors aisé d'en avoir un pour P_p aussi!

Si l'analyse de la performance globale du procédé long terme est non centrée (ce qui est le cas ici) nous utilisons donc:

$$P_{pk} = \min \left\{ \frac{USL - \bar{X}}{3S_{LT}}, \frac{\bar{X} - LSL}{3S_{LT}} \right\} = \min \left\{ \frac{7 - 0.12}{3 \cdot 1.80}, \frac{0.12 - (-7)}{3 \cdot 1.80} \right\} = \min \{1.271, 1.316\} = 1.271$$

et nous savons encore une fois qu'à cause de l'instrument, cette valeur est un peu sous évaluée! Nous avons bien évidemment:

$$P_{pk} \neq P_p$$

donc le processus n'est pas centré (on s'en doutait...). Alors il faut calculer la capacité potentielle décentrée moyenne long terme du procédé P_{pm} selon les relations déterminées plus haut:

$$P_{pm} \cong 1.291$$

Bref, que ce soit de la valeur de P_{pk} , P_p ou P_{pm} , nous voyons que les valeurs sont toutes limites capable.

Si nous faisons alors nos calculs de PPM selon les relations obtenues plus haut avec la valeur de S_{LT} et de \bar{X} obtenues, nous avons alors:

$$PPM_{tot}^{LT} = PPM_{UL} + PPM_{LL} \cong 114$$

Ensuite dire ce chiffre est bon ou mauvais cela est difficile car il nous manque l'information de savoir quel est le coût de production, le coût de revient et de réparation d'un produit et le tout est lui-même dépendant de la quantité total fabriquée! Mais nous pouvons utiliser aussi le modèle de Taguchi pour connaître la valeur des paramètres (moments) calculés qu'il serait préférable de ne pas dépasser!

Calculons maintenant les indices de capabilité court terme! Pour cela, il nous faut l'estimateur de la moyenne de l'ensemble en considérant chaque échantillon comme une variable aléatoire. Nous savons (cf. chapitre de Statistique) que cette moyenne est aussi la moyenne arithmétique dans le cas d'une loi Normale et elle est strictement égale à celle que l'on calcule en considérant l'ensemble des échantillons comme une seule et unique variable aléatoire. Donc il vient que:

$$\bar{X} = 0.12$$

En ce qui concerne l'écart-type par contre ce n'est pas pareil. Mais nous savons (cf. chapitre de Statistiques) que la loi Normal est stable par la somme. Par exemple, nous avons démontré qu'étant donné deux variables aléatoires indépendantes et distribuées selon une loi Normale (en imaginant que chaque variable représente deux de nos dix échantillons), nous avons donc pour leur somme:

$$X + Y = N(\mu_{X+Y}, \sigma_{X+Y}) = N(\mu_X + \mu_Y, \sqrt{\sigma_X^2 + \sigma_Y^2})$$

Or nous avons aussi démontré dans le chapitre de Statistiques que de par la propriété de linéarité de l'espérance, nous avons:

$$E(Z) = E\left(\frac{X+Y}{2}\right) = \frac{1}{2}E(X+Y) = \frac{1}{2}(\mu_X + \mu_Y) = \frac{\mu_X + \mu_Y}{2}$$

ce qui est conforme à notre remarque précédent pour la variance:

$$V(Z) = V\left(\frac{X+Y}{2}\right) = \frac{1}{4}V(X+Y) = \frac{1}{4}\sqrt{\sigma_X^2 + \sigma_Y^2} = \sqrt{\frac{\sigma_X^2 + \sigma_Y^2}{2}}$$

Donc in extenso:

$$S_{CT} = \sqrt{\frac{\sum_{i=1}^k S_i^2}{k}}$$

et dans notre cas particulier:

$$\sigma_{CT} = \sqrt{\frac{\sum_{i=1}^5 \sigma_i^2}{5}} = 1.583$$

nous avons une valeur supérieur à 1 ce qui est donc non-conforme à ce que Six Sigma exige dans son niveau de qualité.

Donc l'erreur-standard (l'estimateur de l'écart-type de la moyenne) est de:

$$\sigma_{\mu} = \frac{S_{CT}}{\sqrt{k}} = 0.707$$

Donc l'intervalle de confiance à 95% de la moyenne est de (cf. chapitre de Statistique):

$$\bar{X} - 1.96 \frac{S_{CT}}{\sqrt{k}} \leq \mu \leq \bar{X} + 1.96 \frac{S_{CT}}{\sqrt{k}}$$

Soit dans notre cas:

$$-1.267 \leq \mu \leq 1.507$$

Nous remarquons donc qu'en court terme, l'intervalle est beaucoup plus large qu'en long terme ce qui est normal étant donné la faible valeur de k (qui vaut donc 5 dans notre exemple).

Et l'inférence statistique avec notre écart-type long terme utilisant le test d'hypothèse bilatéral du χ^2 donne (cf. chapitre de Statistique):

$$\frac{(n-1)S_{CT}^2}{\chi_{\alpha/2}(n-1)} \leq \sigma_{CT}^2 \leq \frac{(n-1)S_{CT}^2}{\chi_{1-\alpha/2}(n-1)}$$

Ce qui nous donne dans notre cas:

$$0.899 \leq \sigma_{CT}^2 \leq 20.692$$

soit:

$$0.948 \leq \sigma_{CT} \leq 4.549$$

Nous remarquons alors que sur une analyse long terme avec les intervalles:

$$-1.267 \leq \mu \leq 1.507$$

$$0.948 \leq \sigma_{CT} \leq 4.549$$

Les variations peuvent donc être énormes avec une probabilité cumulée de 95% et il faudra prendre garde dans un cas pratique d'apporter des réglages au plus vite afin de diminuer au maximum les moments!

Calculons maintenant la capabilité potentielle du procédé court terme (si supposé centré donc!). Nous avons:

$$C_p = \frac{USL - LSL}{6S_{CT}} = \frac{E}{6C_{LT}} = \frac{14}{6 \cdot 1.58} = 1.473$$

Donc nous avons:

$$C_p > P_p$$

ce qui est normal car si les mesures que nous avons étaient vraiment faites sur une longue période alors ce serait très problématique alors que sur une courte période c'est déjà un peu plus normal. D'où cette relation d'ordre entre les deux indices.

Mais avec un instrument ayant un C_{pc} de 4, cela correspond réellement à:

$$C_{p,\text{vrai}} = \frac{C_{p,\text{observé}}}{\sqrt{1 - \frac{C_{p,\text{observé}}}{C_{pc}}}} = 1.854$$

De plus, indiquons que comme nous savons faire un calcul d'intervalle de confiance pour σ_{CT} (voir le calcul fait précédemment), il est alors aisé d'en avoir un pour C_p aussi!

Si l'analyse de la capabilité potentielle du procédé court terme est non centrée (ce qui est le cas ici) nous utilisons donc:

$$C_{pk} = \min \left\{ \frac{USL - \bar{X}}{3S_{CT}}, \frac{\bar{X} - LSL}{3S_{CT}} \right\} = \min \left\{ \frac{7 - 0.12}{3 \cdot 1.583}, \frac{0.12 - (-7)}{3 \cdot 1.583} \right\} = \min \{1.44, 1.49\} = 1.44$$

et nous savons encore une fois qu'au cause de l'instrument, cette valeur est un peu sous évaluée! Nous avons bien évidemment:

$$C_{pk} \neq C_p$$

donc le processus n'est pas centré (on s'en doutait...). Alors il faut calculer la capabilité potentielle décentrée moyenne court terme du procédé C_{pm} selon les relations déterminées plus haut:

$$C_{pm} \cong 1.47$$

Bref, que ce soit de la valeur de C_{pk} , C_p ou C_{pm} , nous voyons que les valeurs sont toutes limites capable.

Si nous faisons alors nos calculs de PPM selon les relations obtenues plus haut avec la valeur de S_{CT} et de \bar{X} obtenues, nous avons alors:

$$PPM_{tot}^{CT} = PPM_{UL} + PPM_{LL} \cong 10$$

Ensuite dire si ce chiffre est bon ou mauvais est difficile car il nous manque l'information de savoir quel est le coût de production, le coût de revient et de réparation d'un produit et le tout est lui-même dépendant de la quantité total fabriquée! Mais nous pouvons utiliser aussi le modèle de Taguchi pour connaître la valeur des paramètres (moments) calculés qu'il serait préférable de ne pas dépasser!

Le lecteur intéressé pour comparer ces données avec celles fournies par Minitab:

