

Florent ARNAL

ESTIMATION & CARTES DE CONTRÔLE

Université de Bordeaux

Adresse électronique : florent.arnal@u-bordeaux.fr

Site internet : <http://flarnal.e-monsite.com>

2017

Sommaire

I	Rappels	2
II	Estimations ponctuelles	3
II.1	Estimation ponctuelle de la moyenne μ d'une population	3
II.2	Estimation ponctuelle de la variance σ^2 d'une population	3
II.3	Estimation ponctuelle de l'écart-type σ d'une population	4
II.4	Estimation ponctuelle de la proportion p d'un caractère dans une population	5
III	Estimation par intervalle de confiance	5
IV	Carte de contrôle de la moyenne	6
IV.1	Introduction	6
IV.2	Principe des cartes de contrôle de Shewhart	6
IV.3	Mise en place d'une carte de contrôle	6
IV.4	Utilisation d'une carte de contrôle	7
V	Résumé sur les cartes de contrôle avec R	8
V.1	Introduction des données	8
V.2	Autour de la moyenne et de la variabilité	8
V.3	Autour de critères qualitatifs	9
VI	Annexe : Valeurs des coefficients d_2	9

Introduction

Ce chapitre propose une initiation à l'utilisation des méthodes statistiques permettant d'estimer une moyenne, une variance ou une fréquence (proportion) à partir des résultats observés sur un échantillon.

On abordera, dans la première partie, le problème de l'estimation ponctuelle alors que dans la deuxième partie, on introduit la notion d'intervalle de confiance.

On suppose que les prélèvements sont effectués de manière aléatoire et avec remise ou peuvent être considérés comme tels.

I Rappels

Définition :

Soit $(X_i)_{1 \leq i \leq n}$ n variables aléatoires indépendantes de même loi, d'espérance μ et d'écart-type σ .

On rappelle que la variable aléatoire égale à la moyenne est définie par :

$$\bar{X} = \frac{X_1 + X_2 + \dots + X_n}{n}$$

La variable aléatoire égale à la variance est définie par

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

Définition :

Enfin, considérons des échantillons de taille n extraits d'une même population sur lesquels on s'intéresse à un caractère qualitatif.

La variable aléatoire égale à la fréquence d'apparition de ce caractère sur un échantillon est définie par

$$F = \frac{X}{n}$$

où X est la variable aléatoire égale au nombre d'apparitions de ce caractère.

Théorème :

Si le caractère X est distribuée suivant la loi normale $\mathcal{N}(\mu; \sigma)$ alors

$$\bar{X} \hookrightarrow \mathcal{N}\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$$

Théorème : (Théorème central limite)

Si n est grand alors la loi de $\bar{X}$ se rapproche de la loi $\mathcal{N}\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$. Ce théorème est à utiliser lorsqu'on connaît pas l'écart-type de X et que l'on considère de grands échantillons ($n \geq 30$).

Théorème : (Distribution de Student)

Si le caractère X est distribuée suivant la loi normale $\mathcal{N}(\mu; \sigma)$ alors

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

Ce théorème est à utiliser lorsqu'on ne connaît l'écart-type de X .

Théorème : Soient n variables aléatoires indépendantes $X_1, \dots, X_n$ suivant la loi normale $\mathcal{N}(\mu; \sigma)$.

$$\frac{nS^2}{\sigma^2} \hookrightarrow \chi^2(n-1)$$

L'utilisation de ce théorème nécessite que la distribution dans la population soit normale.

Théorème : Soit une population qui contient une proportion p d'individus présentant un caractère qualitatif donné. Si n est grand alors la loi de F se rapproche de la loi $\mathcal{N}\left(p; \sqrt{\frac{p(1-p)}{n}}\right)$. Ce théorème est à utiliser lorsque l'on considère de grands échantillons ($n \geq 30$).

II Estimations ponctuelles

Définition : Soit α un paramètre d'un modèle.

Une variable Y est un estimateur sans biais de θ si $\mathbb{E}(Y) = \theta$.

Une estimation ponctuelle de ce paramètre θ , noté $\hat{\theta}$, est alors y , réalisation de la variable Y sur l'échantillon considéré.

II.1 Estimation ponctuelle de la moyenne μ d'une population

On suppose que les n variables aléatoires $(X_i)_{1 \leq i \leq n}$ sont indépendantes, de même loi, d'espérance μ et d'écart-type σ .

$$\mathbb{E}(\bar{X}) = \mathbb{E}\left(\frac{X_1 + X_2 + \dots + X_n}{n}\right) = \frac{1}{n} \sum_{i=1}^n \mathbb{E}(X_i) = \frac{1}{n} \times n\mu = \mu.$$

Théorème : Une estimation ponctuelle de la moyenne d'une population est donnée par

$$\hat{\mu} = \bar{x}$$

II.2 Estimation ponctuelle de la variance σ^2 d'une population

On suppose que les n variables aléatoires $(X_i)_{1 \leq i \leq n}$ sont indépendantes, de même loi normale $\mathcal{N}(\mu; \sigma)$.

Théorème : L'espérance de la loi $\chi^2(n)$ est égale à n .

Il en résulte que $\mathbb{E}\left(\frac{nS^2}{\sigma^2}\right) = n-1$ ce qui conduit à la relation suivante : $\mathbb{E}\left(\frac{n}{n-1}S^2\right) = \sigma^2$.

Théorème : Une estimation ponctuelle de la variance d'une population est donnée par

$$\hat{\sigma}^2 = \frac{n}{n-1}s^2 = \frac{SCE}{n-1}$$

Remarque : On considère fréquemment qu'une estimation de l'écart-type de la population est $\hat{\sigma} = \sqrt{\frac{n}{n-1}} s$ mais il faut savoir qu'il s'agit d'une estimation biaisée.
 Pour obtenir une estimation de l'écart-type, on peut utiliser les étendues comme présenté ci-dessous.

II.3 Estimation ponctuelle de l'écart-type σ d'une population

Soit $(X_i)_{1 \leq i \leq n}$ n variables aléatoires indépendantes de même loi $\mathcal{N}(\mu; \sigma)$.

Les variables centrées réduites seront notées X_i^* de densité f et de fonction de répartition Φ .

L'étendue R (Range en anglais) est définie ainsi

$$R = \max_i(X_i) - \min_i(X_i)$$

On introduit W définie par

$$W = \frac{\max_i(X_i) - \mu}{\sigma} - \frac{\min_i(X_i) - \mu}{\sigma} = \frac{R}{\sigma}$$

Considérons $M_n = \frac{\max_i(X_i) - \mu}{\sigma} = \max_i(X_i^*)$ et $m_n = \frac{\min_i(X_i) - \mu}{\sigma} = \min_i(X_i^*)$.

La fonction de répartition de M_n est donnée pour tout réel t par

$$P(M_n \leq t) = P\left(\bigcap_i (X_i^* \leq t)\right) = \prod_i P(X_i^* \leq t) = [\Phi(t)]^n$$

Ainsi, la fonction de densité de M_n , notée f_{M_n} et obtenue par dérivation, est donnée par

$$f_{M_n}(t) = n(f(t)) \times [\Phi(t)]^{n-1}$$

En outre, $m_n = \min_i(X_i^*) = -\max_i(-X_i^*)$ où les X_i^* sont centrées réduites. Ainsi

$$\mathbb{E}(m_n) = -\mathbb{E}(M_n)$$

On en déduit que $\mathbb{E}(W) = \mathbb{E}\left(\frac{R}{\sigma}\right) = 2\mathbb{E}(M_n)$.

On convient de noter

$$d_2 = 2\mathbb{E}(M_n)$$

A l'aide de logiciel de calcul formel, en calculant $d_2 = 2 \times \int_{-\infty}^{+\infty} nt(f(t)) \times [\Phi(t)]^{n-1}$, on obtient :

pour $n = 2$, $d_2 = 1,128$;

pour $n = 3$, $d_2 = 1,693$ (cf. Annexes).

On obtient $\mathbb{E}(W) = E\left(\frac{R}{\sigma}\right) = d_2$ soit $\mathbb{E}\left(\frac{R}{d_2}\right) = \sigma$. Ainsi :

$$\hat{\sigma} = \frac{\bar{R}}{d_2}$$

avec $\bar{R}$ = moyenne des étendues.

II.4 Estimation ponctuelle de la proportion p d'un caractère dans une population

Soit une population qui contient une proportion p d'individus présentant un caractère qualitatif donné.

La variable aléatoire égale à la fréquence d'apparition de ce caractère sur un échantillon est définie par $F = \frac{X}{n}$ où X est la variable aléatoire égale au nombre d'appartenance de ce caractère.

$X \hookrightarrow \mathcal{B}(n; p)$ donc $\mathbb{E}(X) = np$ ce qui conduit à $\mathbb{E}\left(\frac{X}{n}\right) = \frac{1}{n}E(X) = p$.

Theoreme : Une estimation ponctuelle de la proportion dans une population est donnée par

$$\hat{p} = f$$

où f représente la proportion observée sur l'échantillon.

III Estimation par intervalle de confiance

Nous avons vu précédemment qu'il était possible d'avoir une estimation ponctuelle de la moyenne μ de la population en considérant $\bar{x}$.

Cependant, différents échantillons nous donneraient différentes estimations de cette moyenne.

L'estimation ponctuelle étant trop liée à l'échantillon choisi, il est souvent intéressant de déterminer des intervalles liés à cette moyenne qui sont appelés intervalles de confiance.

Définition : Estimer un paramètre θ par intervalle, au niveau de confiance $1 - \alpha \in]0; 1[$, à partir d'un estimateur Y revient à déterminer un réel positif η tel que

$$P(Y - \eta \leq \theta \leq Y + \eta) = 1 - \alpha$$

L'intervalle $[y - \eta; y + \eta]$ est une estimation de θ par intervalle de confiance au niveau $1 - \alpha$.

Déterminons une estimation par intervalle de confiance de la moyenne μ lorsque la distribution dans la population est normale, d'espérance μ et d'écart-type σ connu.

Notons u_a le réel tel que $P(U \leq u_a) = a$. On a : $u_a = \Phi^{-1}(a)$ où $U \hookrightarrow \mathcal{N}(0; 1)$.

Ainsi : $P(-u_{1-\frac{\alpha}{2}} \leq U \leq u_{1-\frac{\alpha}{2}}) = 1 - \alpha$.

D'après ce qui précède, on a : $\bar{X} \hookrightarrow \mathcal{N}\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$ donc $\frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \hookrightarrow \mathcal{N}(0; 1)$.

Ainsi : $P\left(-u_{1-\frac{\alpha}{2}} \leq \frac{\bar{X} - \mu}{\frac{\sigma}{\sqrt{n}}} \leq u_{1-\frac{\alpha}{2}}\right) = 1 - \alpha$ soit $P\left(-u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \bar{X} - \mu \leq u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$

ce qui conduit à $P\left(\bar{X} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \leq \mu \leq \bar{X} + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}\right) = 1 - \alpha$.

Théorème : (Intervalle de confiance d'une moyenne)

Lorsque la distribution dans la population est normale, d'espérance μ et d'écart-type σ connu, une estimation par intervalle de confiance au niveau $1 - \alpha$ de la moyenne μ est donnée par

$$I_{1-\alpha} = \left[\bar{x} - u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}}; \bar{x} + u_{1-\frac{\alpha}{2}} \frac{\sigma}{\sqrt{n}} \right]$$

Exemple : Une estimation par intervalle de confiance au niveau 0,95 est donnée par :

$$I_{0,95} = \left[\bar{x} - 1,96 \frac{\sigma}{\sqrt{n}}; \bar{x} + 1,96 \frac{\sigma}{\sqrt{n}} \right]$$

Il est très vraisemblable que la moyenne de la population μ appartienne à cet intervalle ...

IV Carte de contrôle de la moyenne

IV.1 Introduction

Dans le cadre d'une démarche qualité, il faut surveiller, par exemple, le bon fonctionnement d'appareils de mesure ou la stabilité d'une fabrication.

On met donc en place des contrôles en cours de fabrication sur le nombre de défectueux, la moyenne, l'écart-type ou l'étendue. Ils permettront de surveiller la fabrication et d'enregistrer les variations au cours du temps de la qualité de la fabrication.

On dit qu'un processus industriel est sous contrôle lorsque, pour le ou les caractères pris en considération, la moyenne et la variabilité des valeurs observées (ou la proportion de défectueux) restent stables. Une fabrication peut faire l'objet d'un contrôle sans être sous contrôle.

Une carte de contrôle est un graphique sur lequel on reporte, dans l'ordre chronologique, les valeurs d'une statistique calculée sur des échantillons, en général de même effectif, issus de la fabrication. Chaque point a pour abscisse le no de l'échantillon et pour ordonnée la valeur de la statistique calculée sur cet échantillon (moyenne, écart-type,...). Une telle carte comporte une ligne centrale ainsi que des limites de surveillance et de contrôle tracées à l'avance.

Nous allons présenter ici la carte de contrôle mise en place par Shewhart pour le contrôle de la moyenne. Elle peut être complétée par une carte de contrôle de l'écart-type ou de l'étendue pour permettre de contrôler la dispersion (toute augmentation de cette dernière conduit à une détérioration de la qualité de la production).

IV.2 Principe des cartes de contrôle de Shewhart

On prélève à intervalles réguliers des échantillons d'effectif fixe n . Pour la carte de contrôle de la moyenne, on calcule pour chaque échantillon la moyenne des valeurs observées (on peut également calculer l'étendue ou l'écart-type pour contrôler la dispersion).

IV.3 Mise en place d'une carte de contrôle

La ligne centrale d'une carte de contrôle correspond à la valeur cible qui correspond à l'espérance de la statistique associée aux grandeurs mesurées.

Ainsi, pour la carte de contrôle de la moyenne, la valeur cible est la moyenne μ de la population considérée (estimation ou valeur de référence). Il est à noter que cette étape est plus complexe pour la carte de contrôle de l'écart-type ...

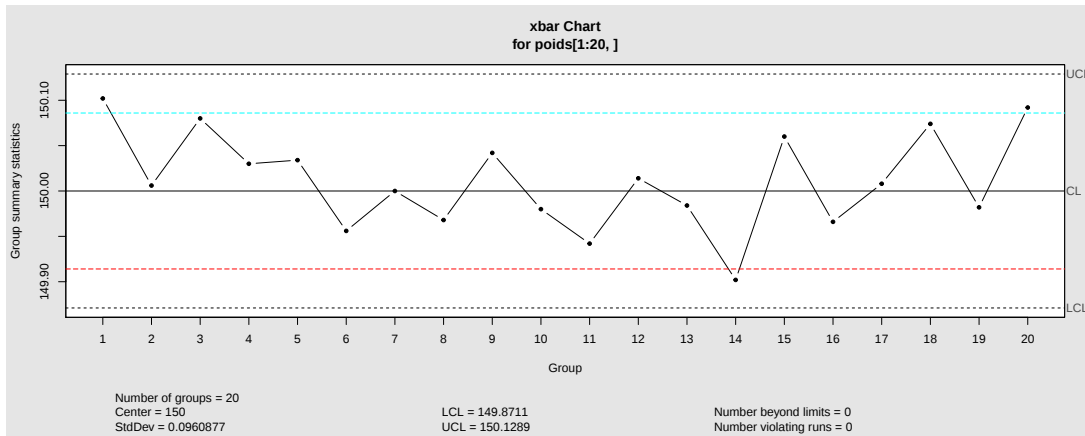
Définition :

La limite supérieure de surveillance (LSS) et la limite inférieure de surveillance (LIS) sont situées à 2 écarts-types (de la statistique utilisée) de la valeur cible.

La limite supérieure de contrôle (LSC ou UCL en anglais) et la limite inférieure de contrôle (LIC ou

LCL en anglais) situées à 3 écarts-types de la valeur cible. On a donc :

$$LIS = \mu - 2\frac{\sigma}{\sqrt{n}}; LSS = \mu + 2\frac{\sigma}{\sqrt{n}}; LCL = LIC = \mu - 3\frac{\sigma}{\sqrt{n}}; UCL = LSC = \mu + 3\frac{\sigma}{\sqrt{n}}.$$



La taille n assez faible des échantillons prélevés en cours de fabrication sous-entend que la distribution de X (égale, par exemple, à la masse des éléments de la production) est normale. Cette hypothèse est absolument nécessaire pour qu'une éventuelle validation de l'estimation de l'écart-type soit cohérente.

Étant donné que $\bar{X}$ est distribuée suivant la loi normale $\mathcal{N}\left(\mu; \frac{\sigma}{\sqrt{n}}\right)$, on en déduit que :

$$P\left(\mu - 2\frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 2\frac{\sigma}{\sqrt{n}}\right) \simeq 0,954 \text{ et } P\left(\mu - 3\frac{\sigma}{\sqrt{n}} \leq \bar{X} \leq \mu + 3\frac{\sigma}{\sqrt{n}}\right) \simeq 0,9973.$$

Conclusion : Le risque de fausse alarme est de 0,27% pour le dépassement des limites de contrôle et limites de contrôle et de 4,6% pour celui des limites de surveillance.

IV.4 Utilisation d'une carte de contrôle

Voici quelques règles de décision en cours de fabrication :

- Si la valeur est entre les limites de surveillance, le résultat est satisfaisant.
- Si la valeur est entre une limite de surveillance et une limite de contrôle, on procède immédiatement à une nouvelle prise. Si le résultat ne se trouve pas entre les limites de surveillance, on procède à un réglage.
- Si la valeur n'est pas entre les limites de contrôle, on procède à un réglage.

Régulièrement, il faut s'assurer que la carte de contrôle mise en place donne des résultats satisfaisants. Si tel n'est pas le cas (sortie fréquente des limites), il faut essayer de trouver la cause d'une éventuelle défaillance ou essayer de redéfinir les limites de cette carte. On peut signaler qu'une carte de contrôle peut également être utilisée pour un contrôle à posteriori en faisant apparaître par exemple un phénomène de dérive (plusieurs points consécutifs vers le bas, plusieurs points consécutifs situés d'un même côté de la cible, ..).

V Résumé sur les cartes de contrôle avec R

V.1 Introduction des données

Pour réaliser des cartes de contrôle avec R, il faut utiliser le package **qcc**. La saisie des valeurs dans un vecteur (nommé *valeurs* ici) et des numéros d'échantillons peut se faire ainsi :

```
> # Création d'un vecteur de 20 échantillons de 5 répétitions
> valeurs = c(...)
> echantillon = rep((1:20), each = 5)
```

On peut également importer un fichier csv.

V.2 Autour de la moyenne et de la variabilité

Plusieurs types de cartes peuvent être obtenues :

- **Carte de contrôle de la moyenne** avec estimation de l'écart-type par les étendues :

```
> library("qcc")
> masses <- qcc.groups(valeurs, echantillon)
> cible = 125
> carte = qcc(masses, center=cible, type="xbar")
```

Création de la carte de contrôle avec les limites de surveillance

```
> LS <- limits.xbar(center=cible, std.dev=carte$std.dev, sizes=carte$sizes, conf=2)
> plot(carte, restore.par = FALSE)
> abline(h = LS, lty = 5, col = "red")
```

- **Carte de contrôle de l'étendue**

```
> qcc(valeurs, type="R")
```

- **Carte de contrôle de l'écart-type**

```
> qcc(valeurs, type="S")
```

V.3 Autour de critères qualitatifs

- Carte de contrôle de la proportion de défectueux

```
> qcc(valeurs, type="p")
```

- Carte de contrôle du nombre de défectueux

```
> qcc(valeurs, type="np")
```

- Carte de contrôle du nombre de défauts

```
> qcc(valeurs, type="c")
```

VI Annexe : Valeurs des coefficients d_2

Table 1: Control Chart Coefficients

Subgroup Size n	d_2	D_1	D_2	D_3	D_4	A_2
2	1.128	0	3.686	0	3.267	1.880
3	1.693	0	4.358	0	2.575	1.023
4	2.059	0	4.698	0	2.282	0.729
5	2.326	0	4.918	0	2.115	0.577
6	2.534	0	5.078	0	2.004	0.483
7	2.704	0.205	5.203	0.076	1.924	0.419
8	2.847	0.387	5.307	0.136	1.864	0.373
9	2.970	0.546	5.394	0.184	1.816	0.337
10	3.078	0.687	5.469	0.223	1.777	0.308
11	3.173	0.812	5.534	0.256	1.744	0.285
12	3.258	0.924	5.592	0.284	1.716	0.266
13	3.336	1.026	5.646	0.308	1.692	0.249
14	3.407	1.121	5.693	0.329	1.671	0.235
15	3.472	1.207	5.737	0.348	1.652	0.223
20	3.735	1.548	5.922	0.414	1.586	0.180
25	3.931	1.804	6.058	0.459	1.541	0.153