

# Projet - Controle Qualité

NGO Khoa Anh

23 janvier 2023

## Contents

<b>1 Partie 1: Test adéquation</b>	<b>2</b>
1.1 Principe du test Kolmogorov-Smirnov . . . . .	2
1.2 Test Kolmogorov-Smirnov sur la loi Gamma . . . . .	3
1.3 Principe du test Shapiro-Wilk . . . . .	3
<b>2 Partie 2: Carte de contrôle</b>	<b>9</b>
2.1 Carte pour la moyenne . . . . .	9
2.2 Carte pour l'étendue . . . . .	13
2.3 Carte pour variance . . . . .	17
<b>3 Partie 3: Détection un changement</b>	<b>21</b>
3.1 Méthode CUSUM . . . . .	21
3.2 Méthode EWMA . . . . .	23
<b>4 Partie 4: Détection de rupture</b>	<b>25</b>

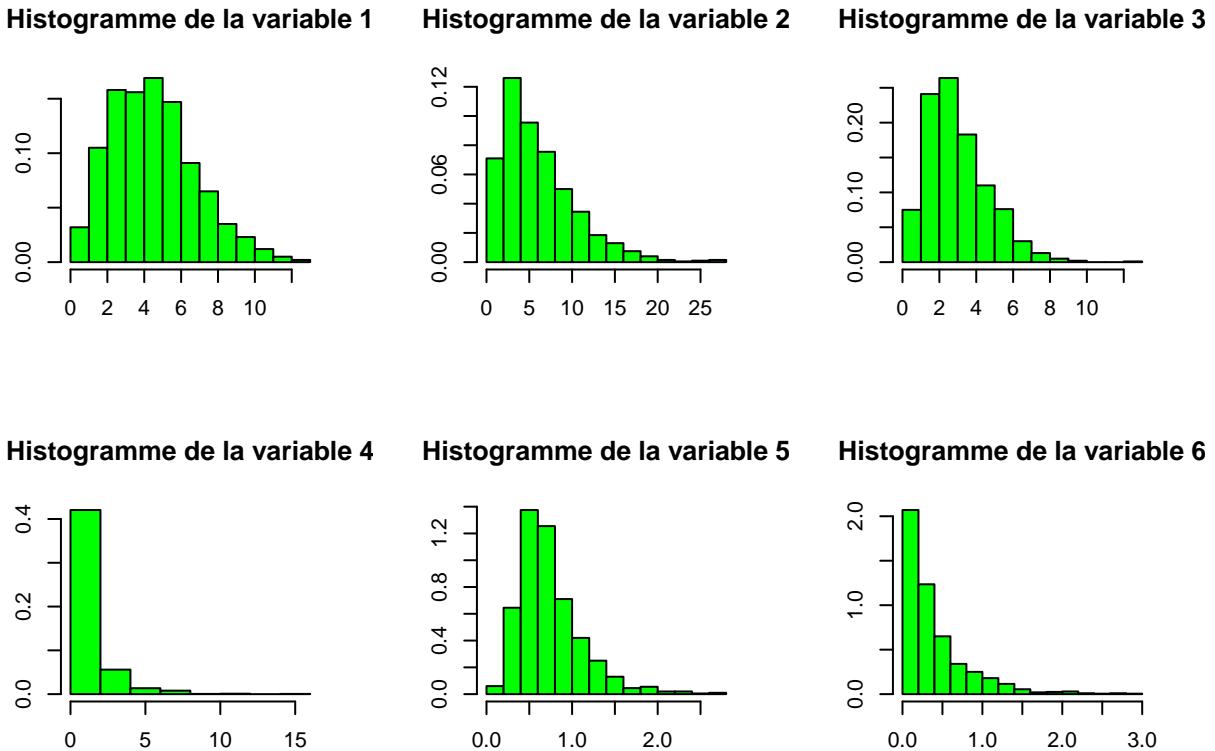
# 1 Partie 1: Test adéquation

On charge tout d'abord les données

```
## [1] 109
```

Nos données contiennent 109 variables quantitatives

On affiche l'histogramme des 6 premières colonnes pour trouver une loi usuelle:



En jetant un coup d'oeil sur les histogrammes, on trouve des lois comme: La loi Normale, loi de Gamma et loi exponentielle. Néanmoins, nos données suivent bien une loi continue, dans ce cas, on utilise donc le test adéquation Kolmogorov-Smirnov, avec un risque de 5% et le test Shapiro-Wilk pour la loi normale

## 1.1 Principe du test Kolmogorov-Smirnov

Le principe est de choisir comme statistique une distance entre fonctions que l'on applique à  $F_n$  et  $F_0$ .

$$D_n = \sup_{t \in R} |F_n(t) - F_0(t)| \text{ où}$$

$\sup_t$  : le supremum de l'ensemble des distances

$F_n(x)$  : la fonction de distribution empirique pour n observations de  $X_i$

## 1.2 Test Kolmogorov-Smirnov sur la loi Gamma

Le principe est pour détecter s'ils suivent une loi gamma ou non. On sait que l'espérance d'une loi  $Gamma(p, \lambda)$  est  $\frac{p}{\lambda}$  et sa variance est  $\frac{p}{\lambda^2}$ , donc il est facile d'estimer les paramètres à partir de nos données, avec la méthode des moments :

$$\begin{aligned}\mu &= \frac{p}{\lambda}, \sigma^2 = \frac{p}{\lambda^2} \\ \Leftrightarrow p &= \sigma^2 \lambda^2, \lambda = \frac{p}{\mu^2} \\ \Leftrightarrow p &= \sigma^2 \frac{p^2}{\mu^2} \\ \Leftrightarrow a &= \frac{\mu^2}{\sigma^2} \quad \lambda = \frac{\mu}{\sigma^2}\end{aligned}$$

## 1.3 Principe du test Shapiro-Wilk

Soit L'échantillon  $X_1, \dots, X_n$

$H_0$  : La loi de l'échantillon est normale

$H_1$  : La loi de l'échantillon n'est pas normale

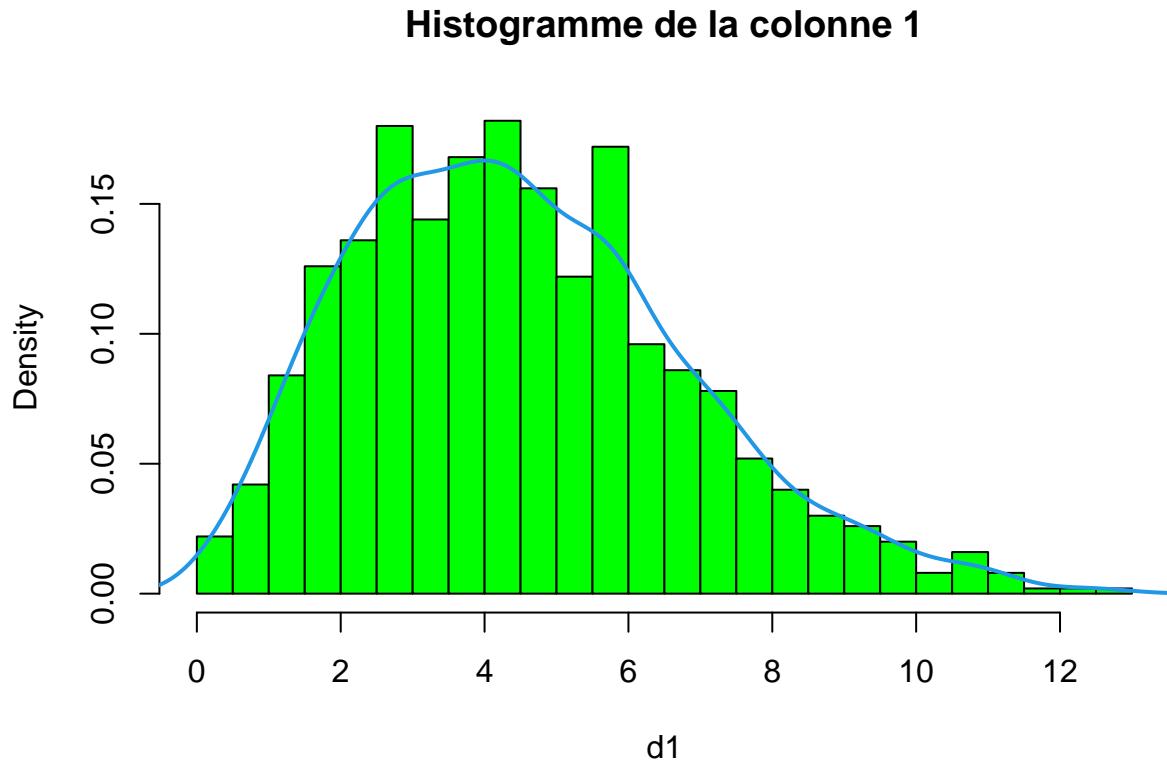
Le test de Shapiro-Wilk consiste à considérer le rapport entre l'estimation de la variance suivant la droite d'Henry et l'estimation de la variance par l'estimateur habituel.

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

ou les  $a_i$  sont des constantes tabulées. Si la distribution est bien normale, ce rapport doit être proche de 1.

On va donc maintenant visualiser chaque colonne et les adapter avec nos tests

### 1.3.1 Pour la 1ere colonne



Cette histogramme semble nous indiquer de la loi Normale. Pour vérifier cette hypothèse, on utilise le test de Shapiro-Wilk

```
##  
## Shapiro-Wilk normality test  
##  
## data: d1  
## W = 0.97547, p-value = 5.9e-12
```

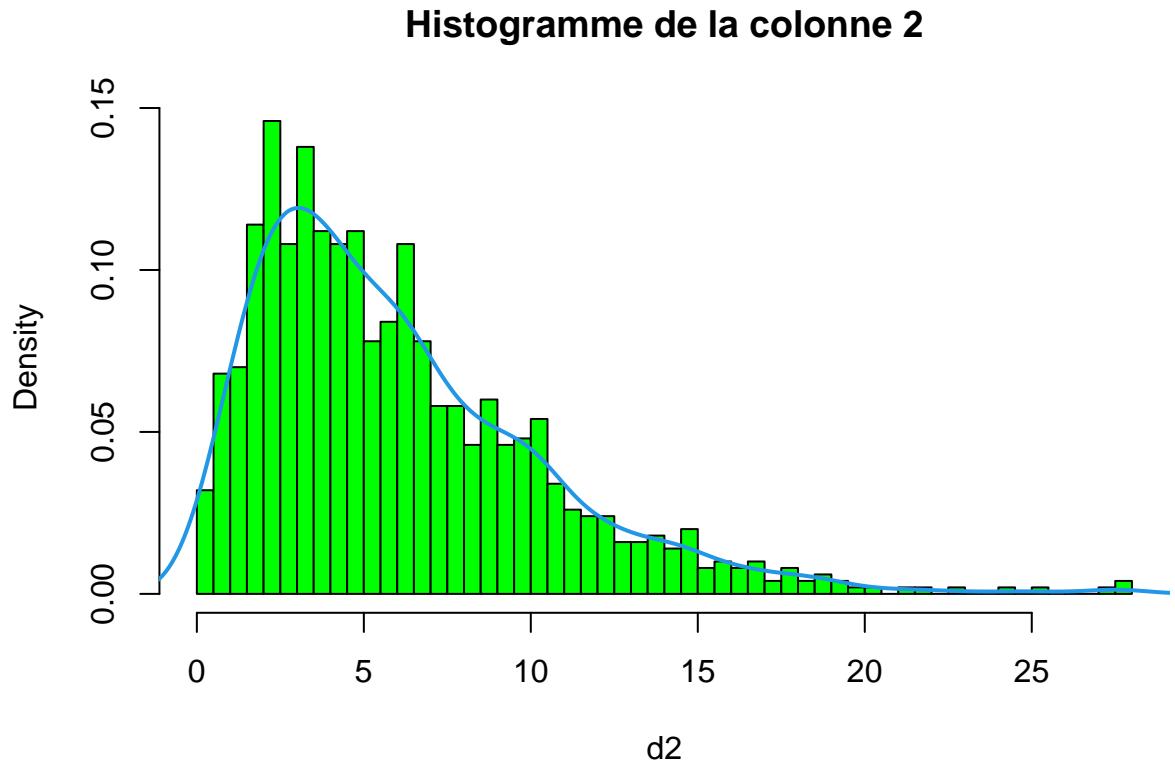
On trouve le **p-value** qui est inférieure à 0.05, cela indique de rejeter l'hypothèse nulle, ce veut dire que cette distribution ne suit pas de loi normale

Néanmoins, notre distribution peut être semblable de loi Gamma. On vérifie notre point de vue avec le test de Kolmogorov-Smirnov sur la loi Gamma

```
##  
## One-sample Kolmogorov-Smirnov test  
##  
## data: d1  
## D = 0.039764, p-value = 0.08465  
## alternative hypothesis: two-sided
```

On trouve maintenant le **p-value** qui est supérieure à 0.05, ce veut dire que notre hypothèse est bonne, cette distribution suit la loi Gamma

### 1.3.2 Pour la 2eme colonne



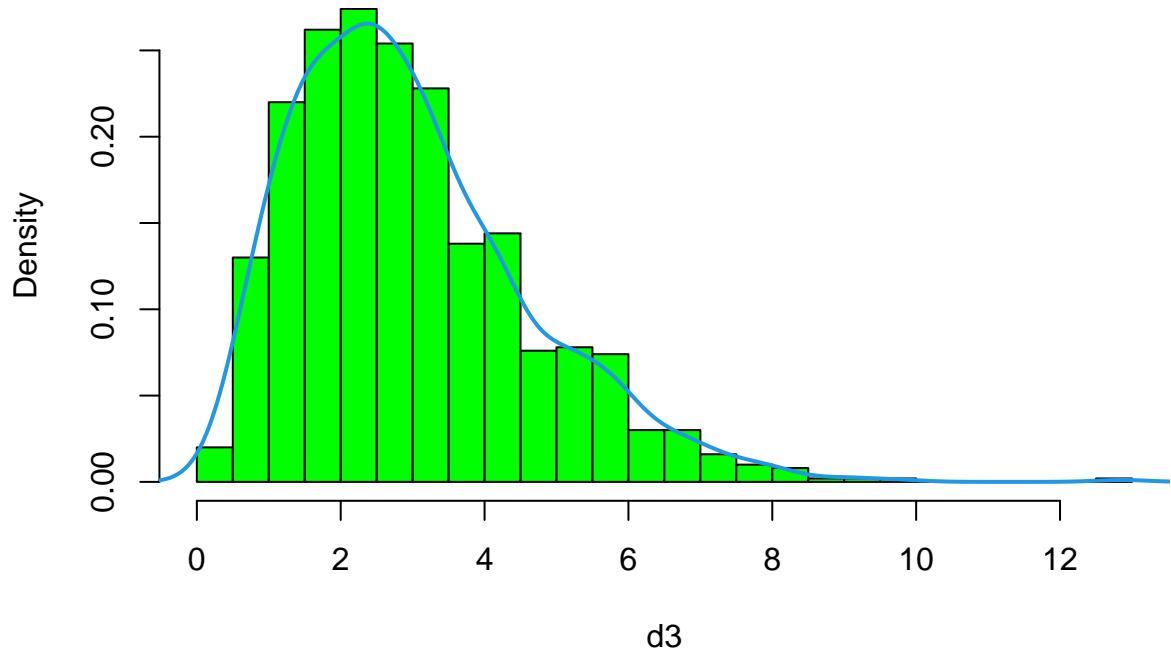
Par rapport à cette histogramme, on peut trouver facilement que la colonne 2 pourrait être suivre la loi Gamma comme sa queue est longue. Pour renforcer notre hypothese, on essaie avec le test de Kolmogorov-Smirnov sur la loi Gamma

```
##  
##  One-sample Kolmogorov-Smirnov test  
##  
## data: d2  
## D = 0.016637, p-value = 0.9448  
## alternative hypothesis: two-sided
```

On trouve maintenant le **p-value** qui est supérieure à 0.05, ca veut dire que notre hypothese est bon, cette distribution suit la loi Gamma

### 1.3.3 Pour la 3eme colonne

**Histogramme de la colonne 3**



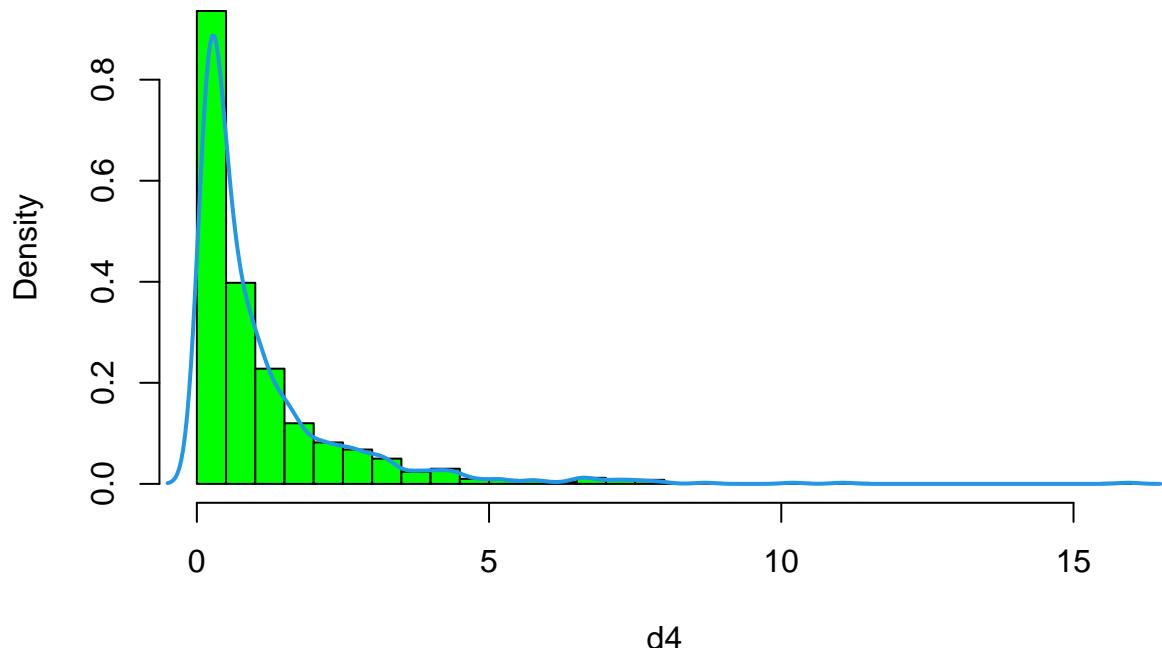
Pareil avec la dernière colonne, l'histogramme nous indique que cette distribution pourrait être suivre la loi Gamma

```
##  
##  One-sample Kolmogorov-Smirnov test  
##  
## data:  d3  
## D = 0.016671, p-value = 0.9439  
## alternative hypothesis: two-sided
```

Comme notre prévision, le **p-value** qui est supérieure à 0.05, ça veut dire que notre hypothèse est bonne, cette distribution suit la loi Gamma

### 1.3.4 Pour la 4eme colonne

**Histogramme de la colonne 4**



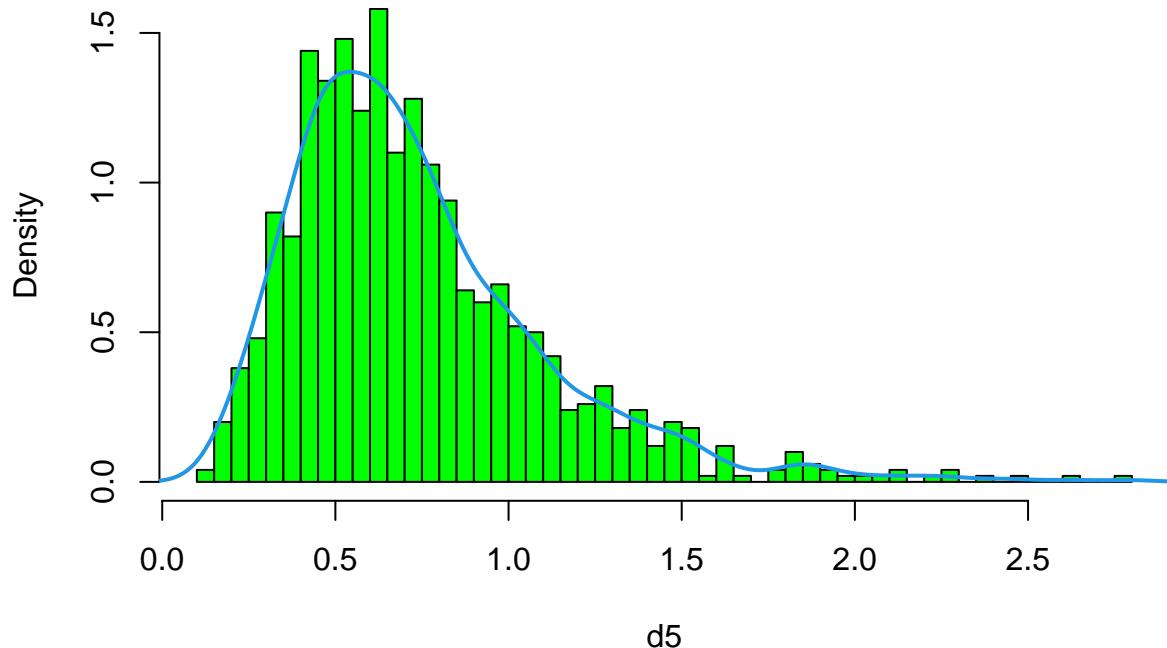
Cette histogramme semble de la loi exponentielle, on essaie donc avec le test de Kolmogorov-Smirnov sur la loi exponentielle

```
##  
##  One-sample Kolmogorov-Smirnov test  
##  
## data:  d4  
## D = 0.1043, p-value = 7.111e-10  
## alternative hypothesis: two-sided
```

Avec un **p-value** super petit, on rejette l'hypothèse de la loi exponentielle

### 1.3.5 Pour la 5eme colonne

**Histogramme de la colonne 5**



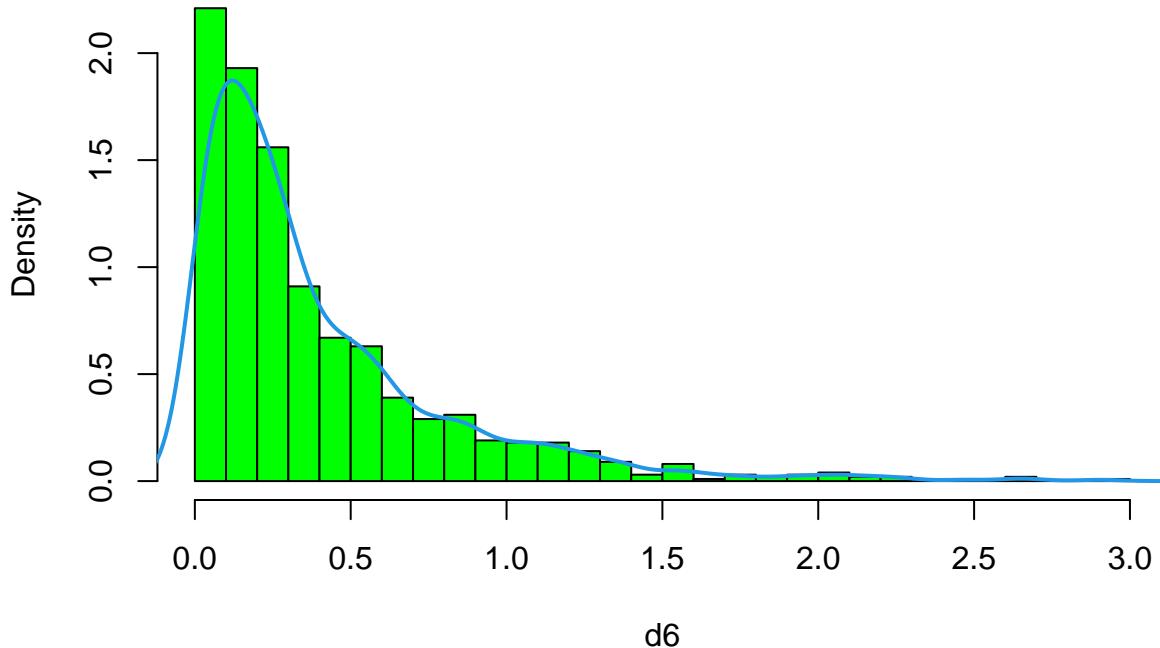
Cette distribution semble d'être de la loi Gamma. On essaie le test Kolmogorov-Smirnov pour renforcer notre hypothese

```
##  
##  One-sample Kolmogorov-Smirnov test  
##  
## data:  d5  
## D = 0.041411, p-value = 0.06478  
## alternative hypothesis: two-sided
```

Le **p-value** qui est supérieure à 0.05, ca veut dire que notre hypothese est bon, cette distribution suit la loi Gamma

### 1.3.6 Pour la 6eme colonne

#### Histogramme de la colonne 6



Avec une queue long, cette distribution semble d'être de la loi Gamma. On essaie le test de Kolmogorov-Smirnov sur la loi Gamma

```
##  
##  One-sample Kolmogorov-Smirnov test  
##  
##  data:  d6  
##  D = 0.031619, p-value = 0.2701  
##  alternative hypothesis: two-sided
```

Le **p-value** qui est supérieure à 0.05, ca veut dire que notre hypothese est bon, cette distribution suit la loi Gamma

## 2 Partie 2: Carte de controle

Dans cette partie, on construit les cartes de controle sur la moyenne, la variance et l'étendu pour les colonnes de 7 à 106 pour suivre les fluctuations d'une caractéristique mesurable ou dénombrable et d'en diagnostiquer les situations non maîtrisées

### 2.1 Carte pour la moyenne

Soient  $\bar{x} = \frac{1}{n} \sum_{i=1}^n X_i$  et  $E(X) = m$ ,  $var(X) = \sigma^2$

- Notons  $W = \frac{R}{\sigma}$  l'étendue relative, notons

$$d_2 = E(W), d_3 = \sqrt{\text{var}(W)}$$

- Avec un intervalle de  $3\sigma$ , on a:

$$\bar{x} \in [\bar{x} - \frac{3}{d_2\sqrt{n}}\bar{R}, \bar{x} + \frac{3}{d_2\sqrt{n}}\bar{R}]$$

On notera  $A_2 = \frac{3}{d_2\sqrt{n}}$

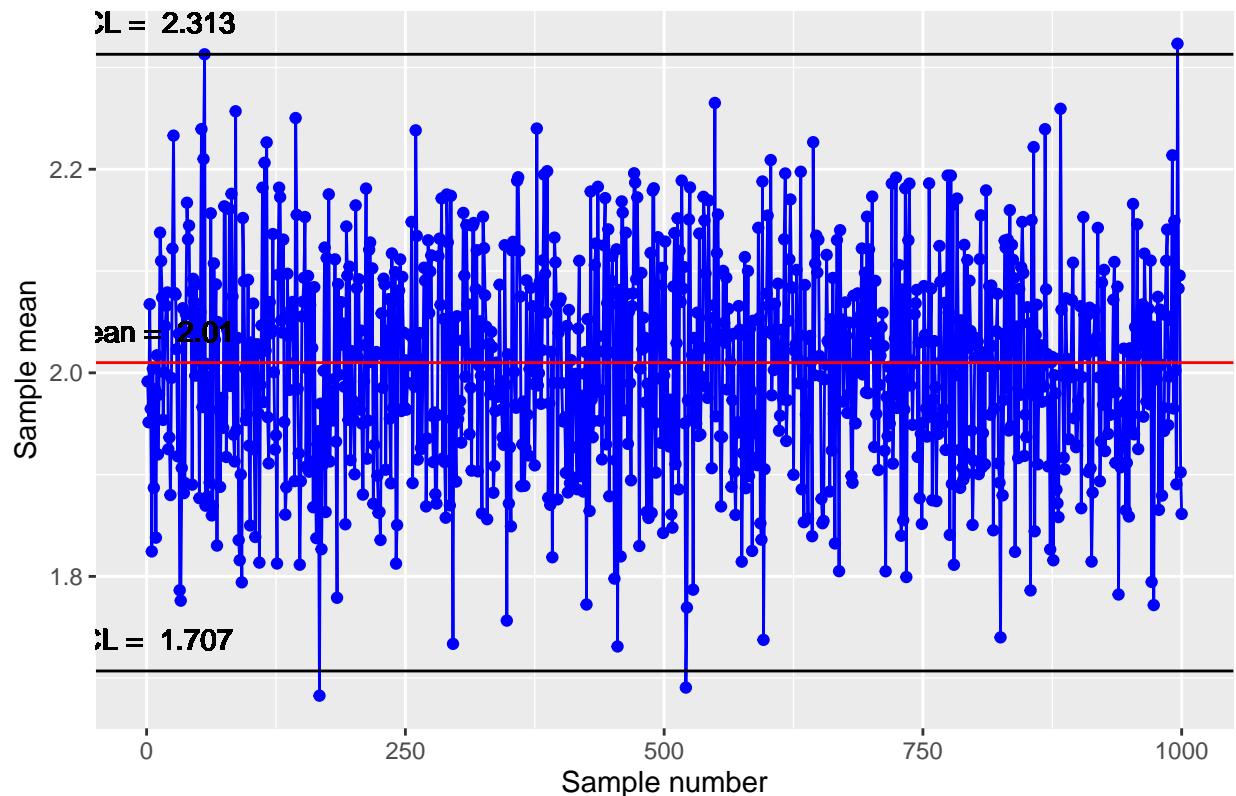
Donc,

$$UCL = \bar{x} + A_2\bar{R} \text{ et } LCL = \bar{x} - A_2\bar{R}$$

On a en total 100 sous groupes, de 7 à 106. On sépare sur les sous groupes de  $n=25$ , il y a donc 4 séparations, le premier est de 7 à 31, le 2eme 32 à 56, le 3eme est de 57 à 81 et le dernier est de 82 à 106

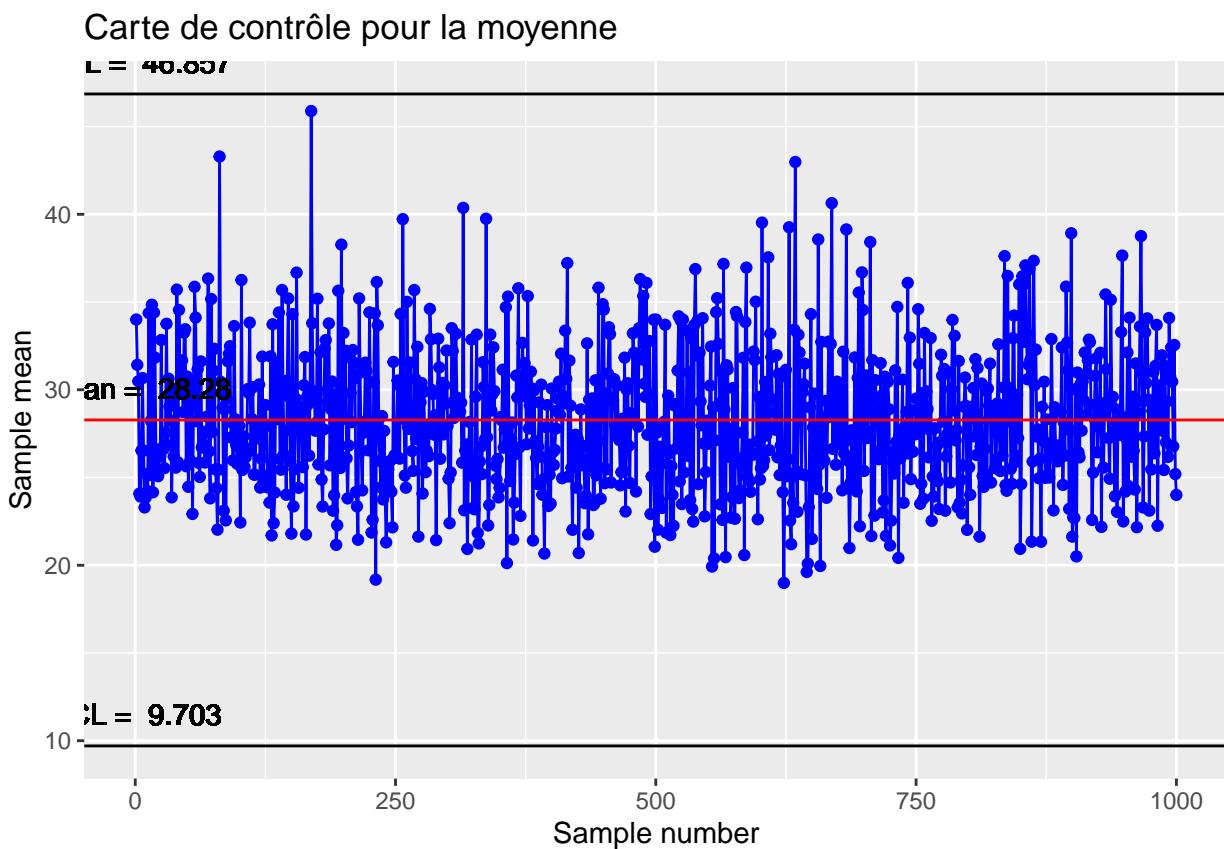
### 2.1.1 Pour les colonnes de 7 à 31

Carte de contrôle pour la moyenne



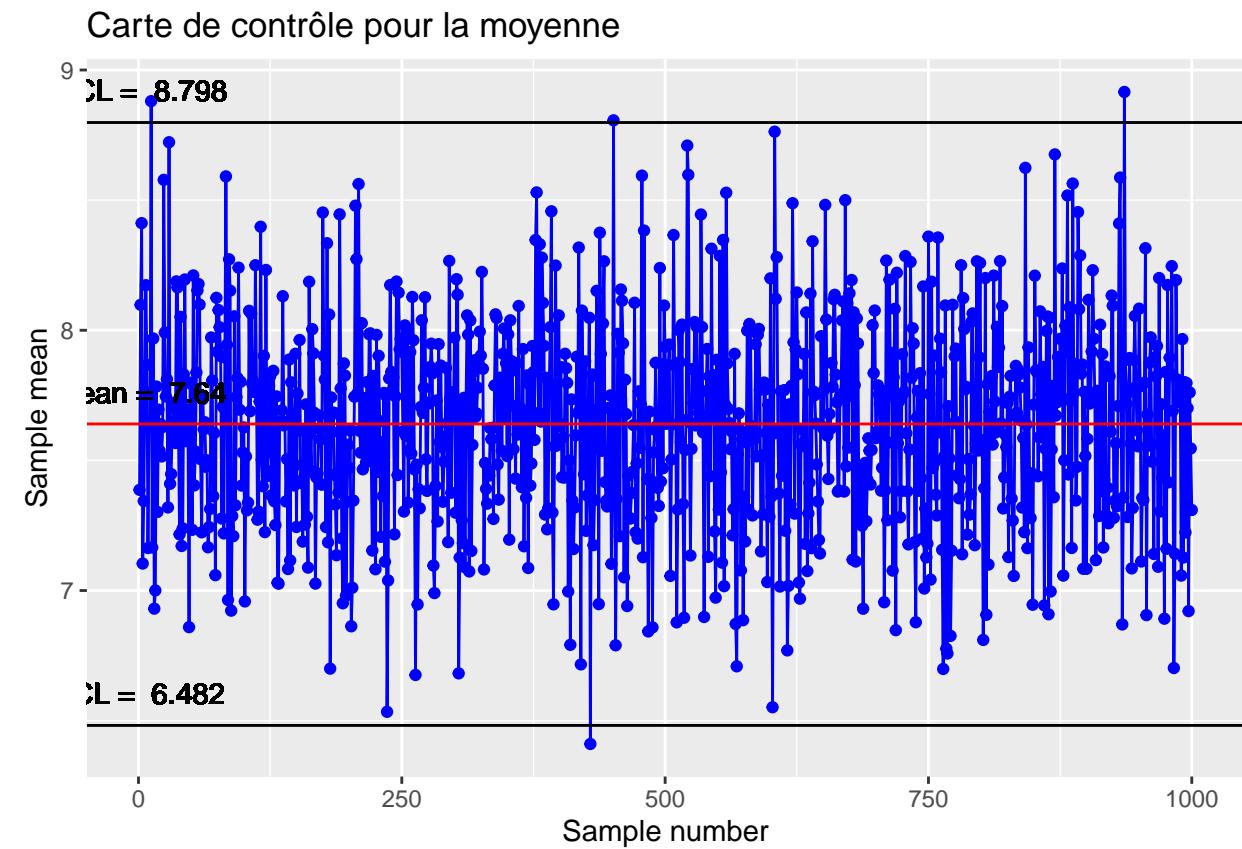
Il n'y a que 3 points hors de contrôle, avec un point qui se situe sur la borne supérieur, soit 0.3% de points non conformes

### 2.1.2 Pour les colonnes de 32 à 56



On trouve que ce processus est sous de contrôle

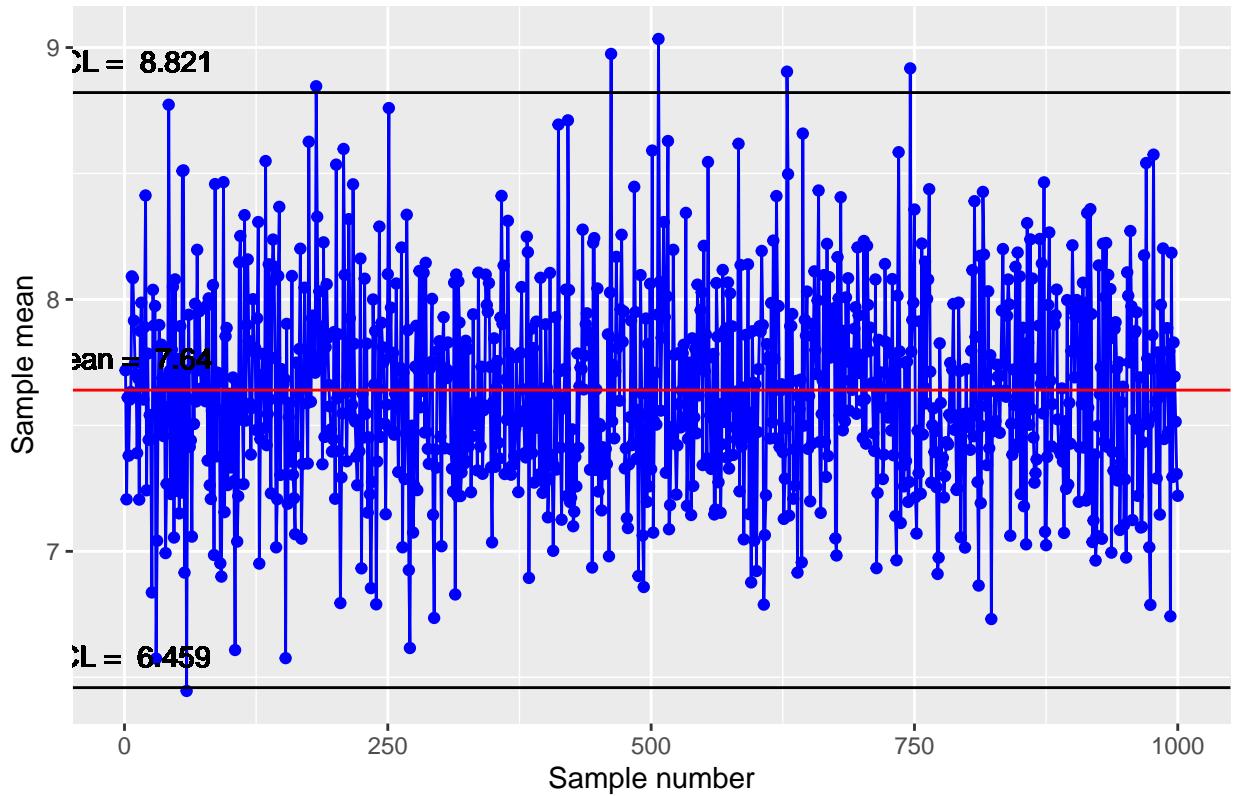
### 2.1.3 Pour les colonnes de 57 à 81



Pour ce processus, on trouve qu'il y a 3 points sont hors de contrôle, et 1 point est sur la borne supérieure, soit 0.3% de points non conformes

#### 2.1.4 Pour les colonnes de 82 à 106

Carte de contrôle pour la moyenne



Il y a 6 points hors de contrôle, soit 0.6% de points non conformes

## 2.2 Carte pour l'étendue

L'idée était de construire  $[\bar{R} - 3\sigma_R, \bar{R} + 3\sigma_R]$ , comme  $\sigma_R = \sigma d_3$  alors  $\hat{\sigma}_R = \frac{\bar{R}d_3}{d_2}$  avec les seuils suivants pour  $\bar{R}$

$$UCL = D_4 \bar{R}, LCL = D_3 \bar{R} \text{ avec } D_3 = 1 - \frac{3d_3}{d_2} \text{ et } D_4 = 1 + \frac{3d_3}{d_2}$$

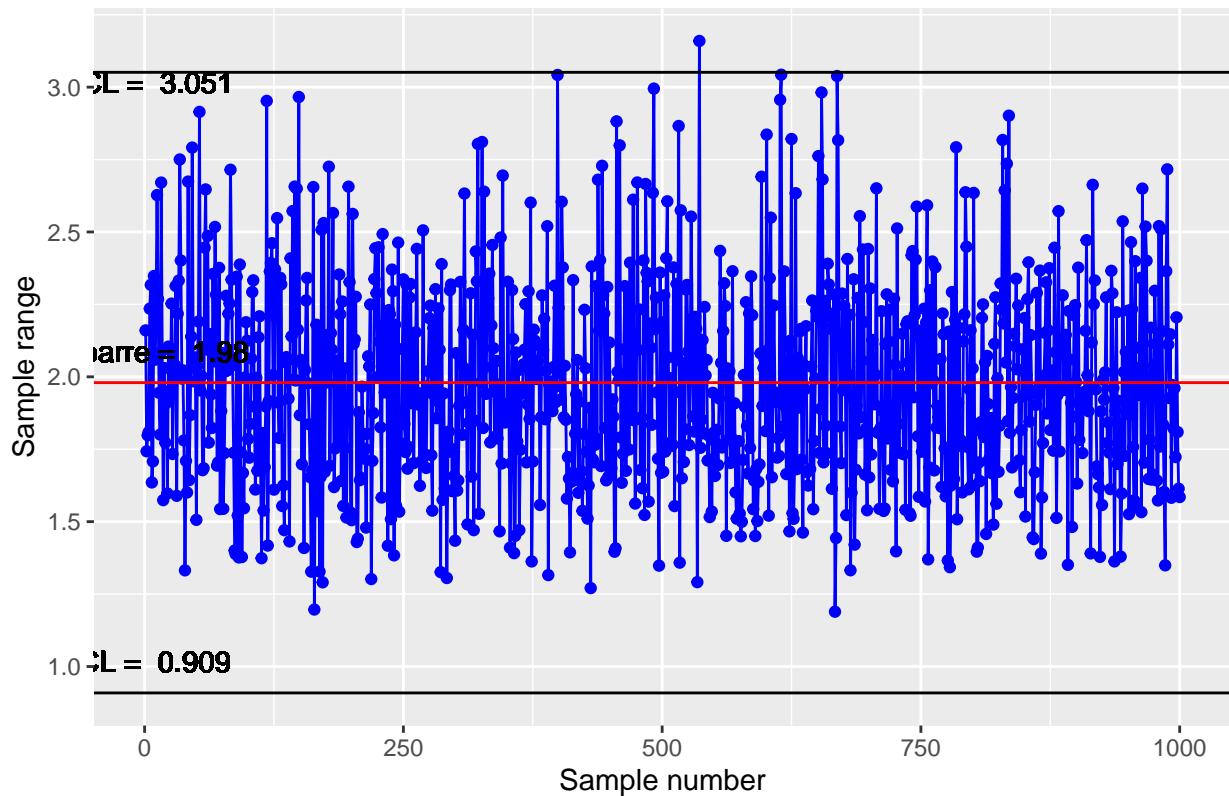
c'est à dire on est sous contrôle si

$$R_i \in [\bar{R} - D_4 \bar{R}, \bar{R} + D_4 \bar{R}]$$

Le principe est le même, on sépare sur les petits sous groupes comme on a fait pour la moyenne

### 2.2.1 Pour les colonnes de 7 à 31

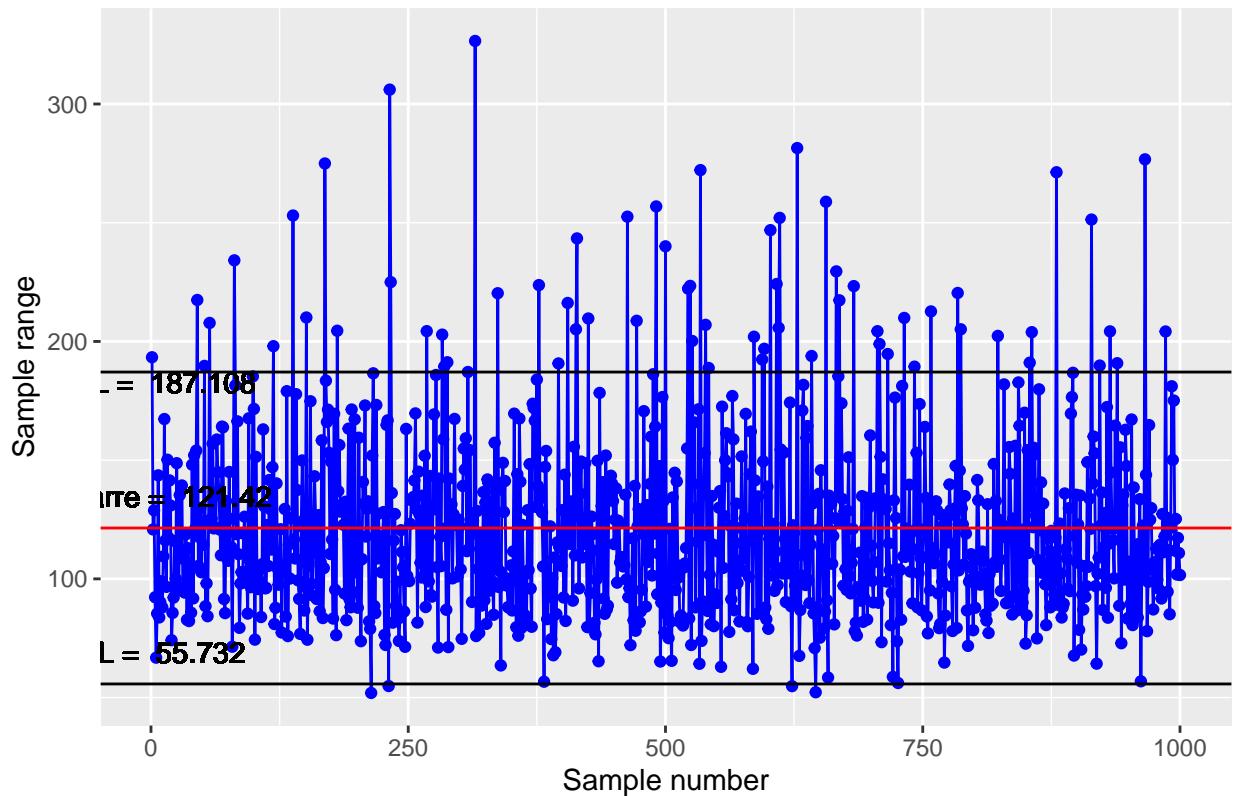
Carte de contrôle pour l'étendue



On trouve que pour ce processus, il y a 1 point hors de contrôle et 1 autre point qui est sur la borne supérieure, soit 0.1% de points non conformes

## 2.2.2 Pour les colonnes de 32 à 56

Carte de contrôle pour l'étendue



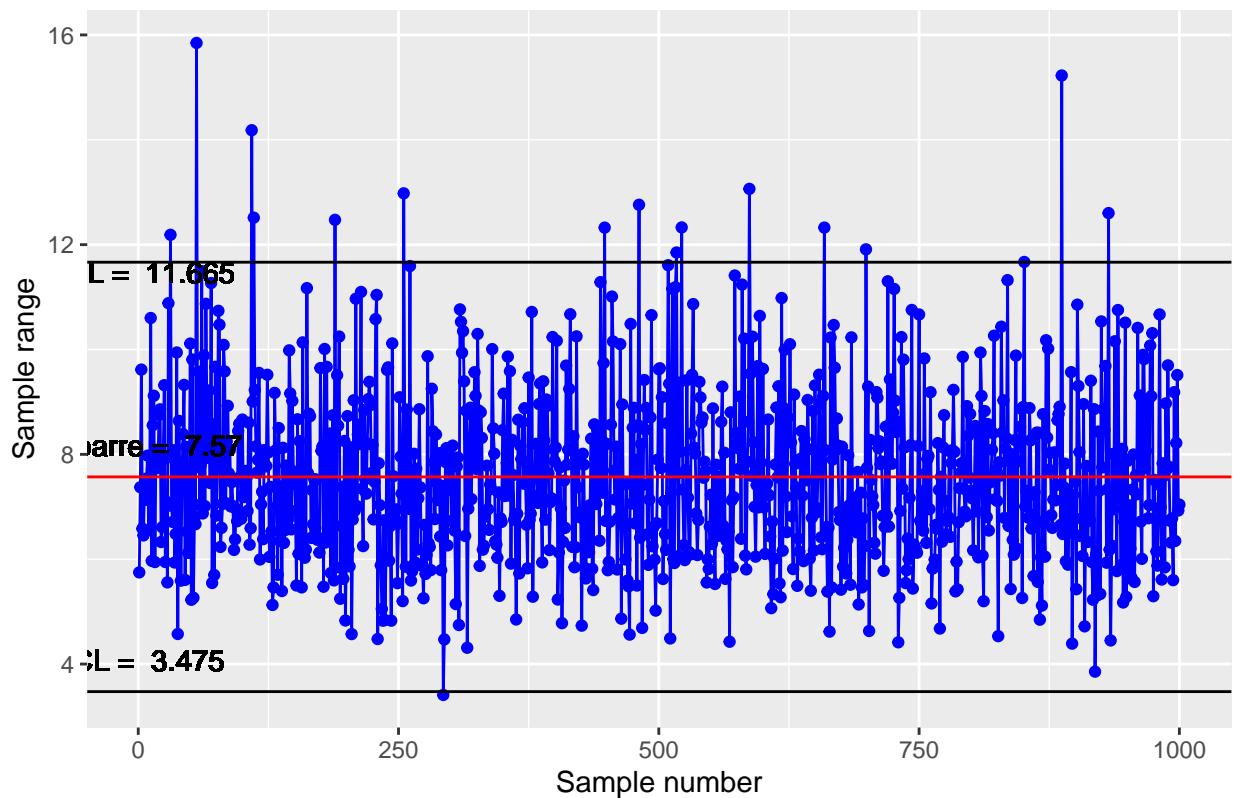
On trouve donc nombre d'états hors de contrôle

```
## [1] 70
```

Il y a donc 70 états hors de contrôles, soit 7% de points non conformes

### 2.2.3 Pour les colonnes de 57 à 81

Carte de contrôle pour l'étendue



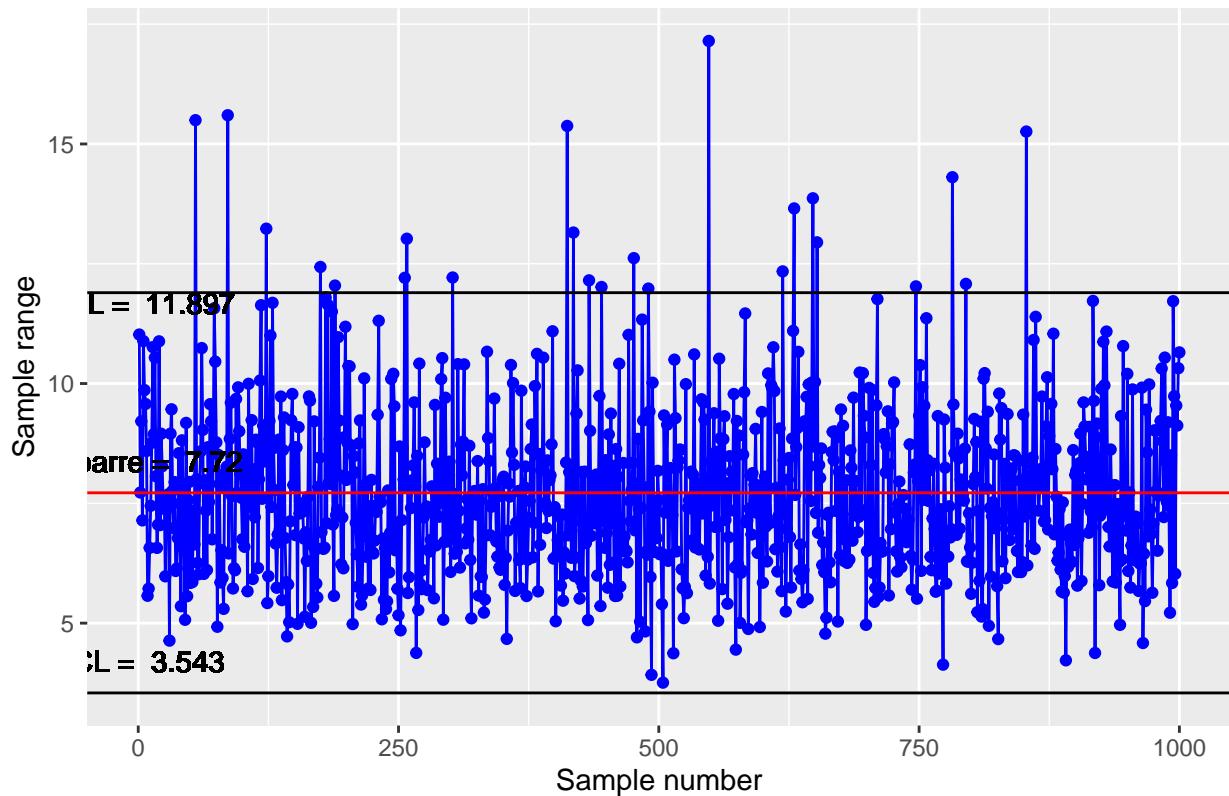
On trouve donc nombre d'états hors de contrôle

```
## [1] 17
```

Il y a donc 17 états hors de contrôles, soit 1.7% de points non conformes

#### 2.2.4 Pour les colonnes de 82 à 106

Carte de contrôle pour l'étendue



On trouve donc nombre d'états hors de contrôle

```
## [1] 23
```

Il y a donc 23 états hors de contrôles, soit 2.3% de points non conformes

### 2.3 Carte pour variance

La carte de  $\sigma^2$  sera basée sur les quantiles de la loi de chi deux

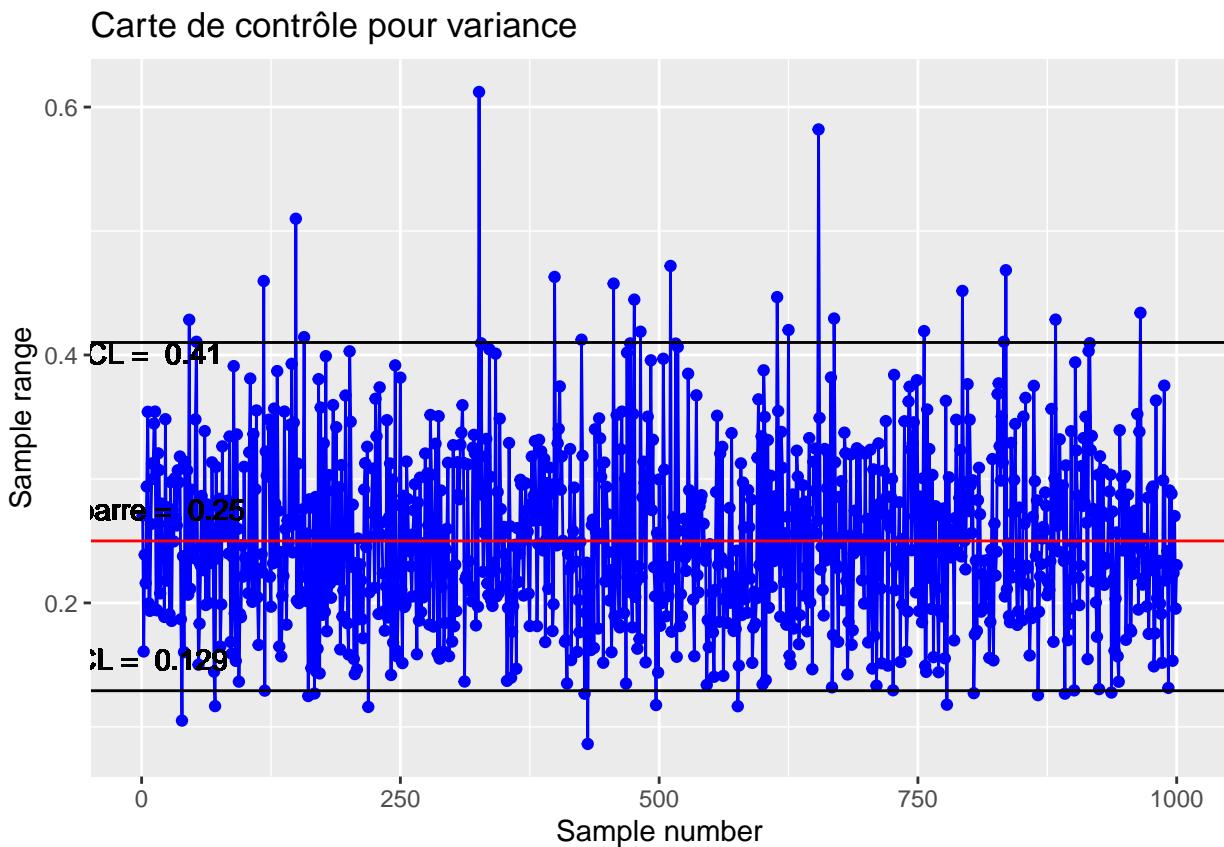
$$LCL = \frac{\bar{s}^2}{n-1} \chi^2_{1-\frac{\alpha}{2}, n-1} \text{ et } UCL = \frac{\bar{s}^2}{n-1} \chi^2_{\frac{\alpha}{2}, n-1}$$

et la ligne centrale  $\bar{s}^2$

Pour cette carte, de même principe et on prend des quantiles de la loi de Chi deux à  $n-1$  degrés de liberté et  $\alpha = 0.05$

Avec  $n = 25$ , on a  $\chi^2_{\frac{\alpha}{2}, n-1} = 39.37$  et  $\chi^2_{1-\frac{\alpha}{2}, n-1} = 12.4$

### 2.3.1 Pour les colonnes de 7 à 31



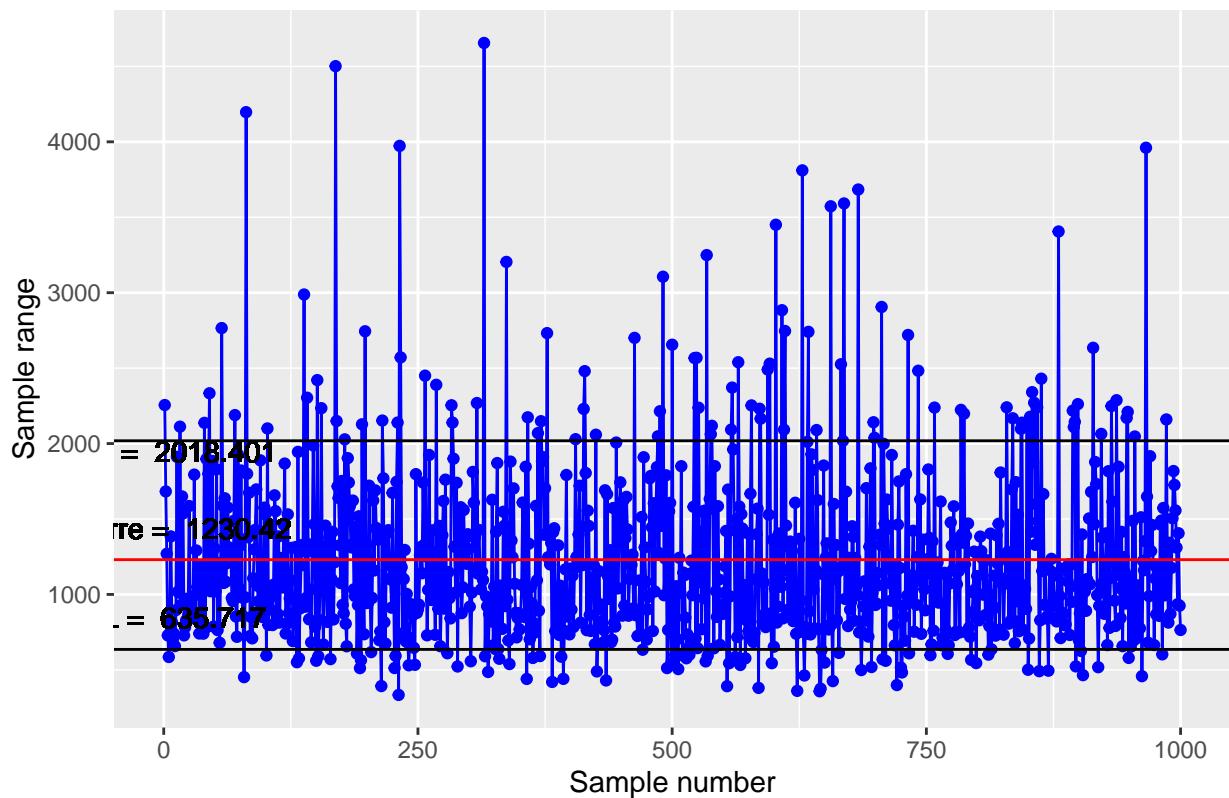
On trouve donc nombre d'états hors de contrôle

```
## [1] 36
```

Il y a donc 36 états hors de contrôles, soit 3.6% de points non conformes

### 2.3.2 Pour les colonnes de 32 à 56

Carte de contrôle pour variance



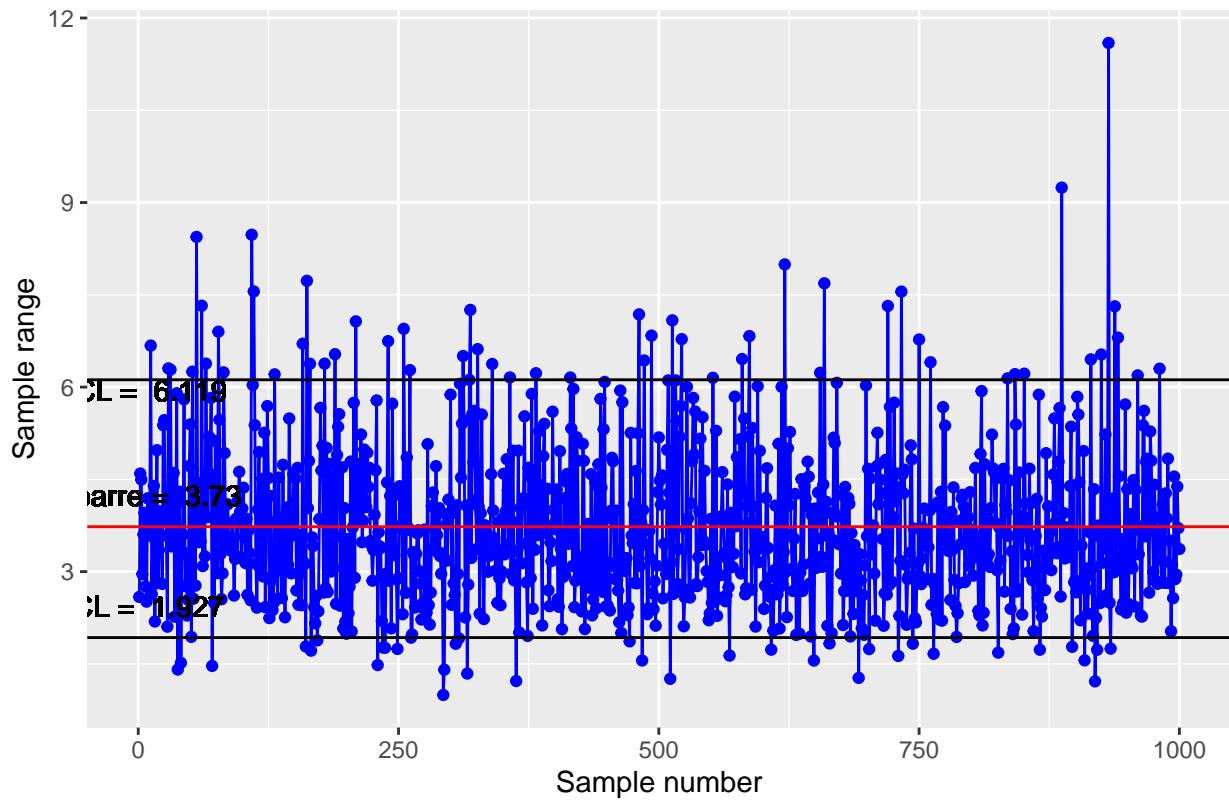
On trouve donc nombre d'états hors de contrôle

```
## [1] 180
```

Il y a donc 180 états hors de contrôles, soit 18% de points non conformes

### 2.3.3 Pour les colonnes de 57 à 81

Carte de contrôle pour variance

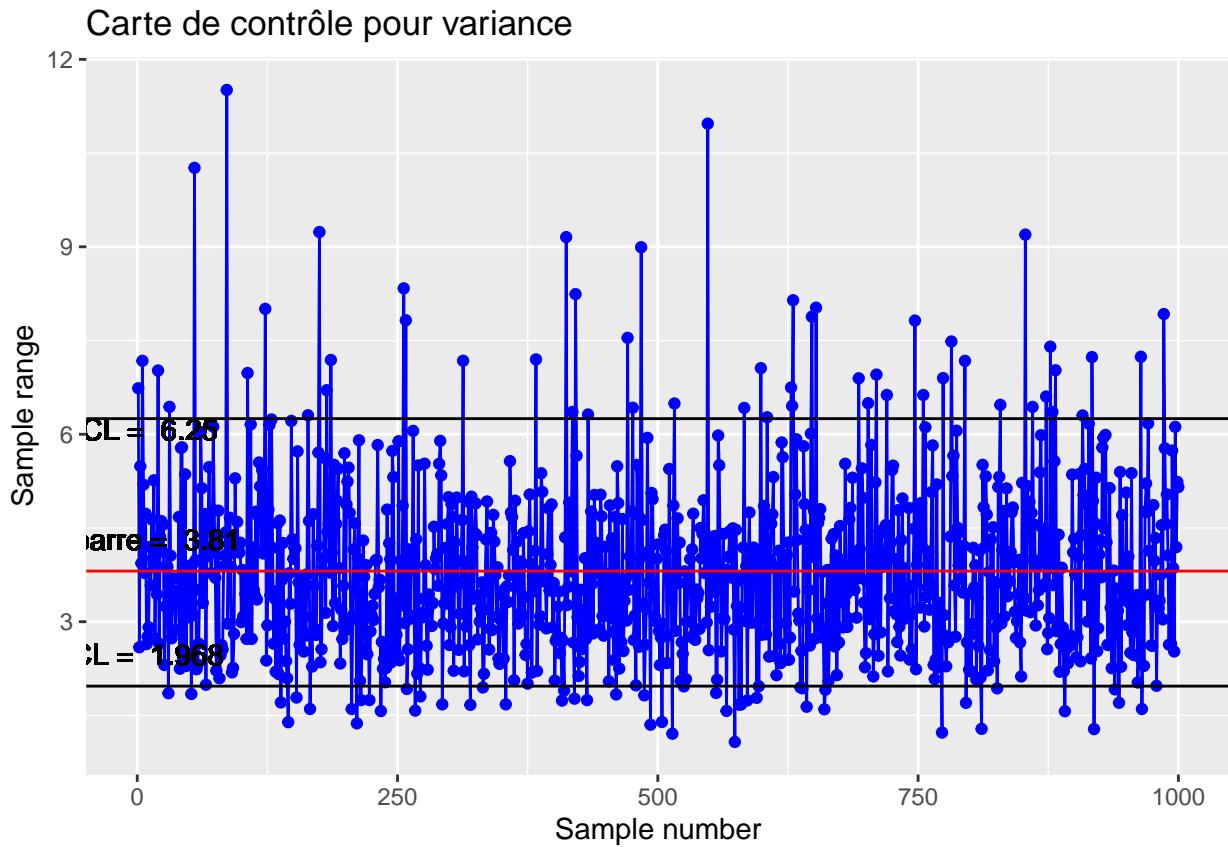


On trouve donc nombre d'états hors de contrôle

```
## [1] 90
```

Il y a donc 90 états hors de contrôles, soit 9% de points non conformes

### 2.3.4 Pour les colonnes de 82 à 106



On trouve donc nombre d'états hors de contrôle

```
## [1] 103
```

Il y a donc 103 états hors de contrôles, soit 10.3% de points non conformes

Commentaire: Pour les moyennes et étendues, leurs processus sont plus sous de contrôle (il y a très peu de points hors de contrôle) que pour les variances.

## 3 Partie 3: Détection un changement

Pour cette partie, on utilise le CUSUM et EWMA pour chaque colonne 107 et 108 pour détecter un changement par rapport à une moyenne égale à 2.

### 3.1 Méthode CUSUM

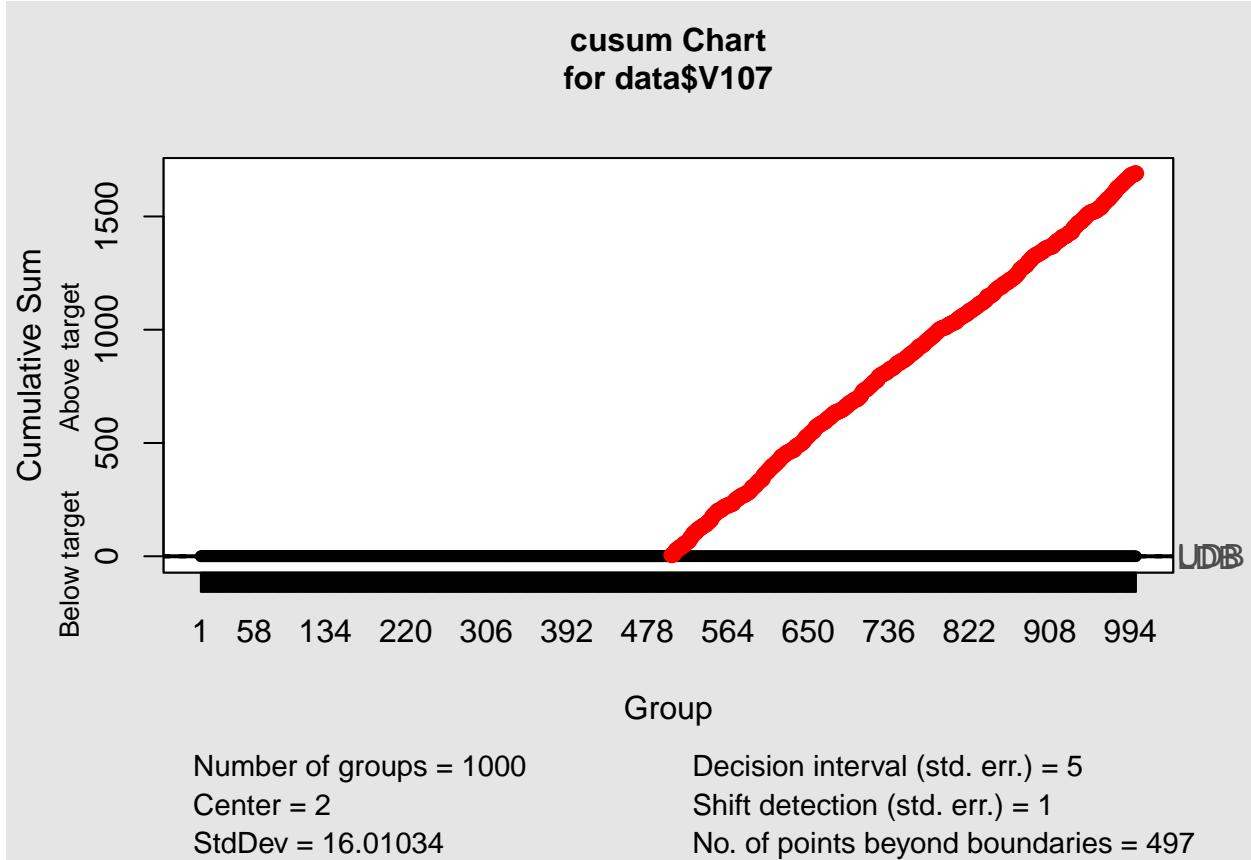
Une carte de somme cumulée (CUSUM) est un type de carte de contrôle utilisé pour surveiller les petites variations de la moyenne du processus. Il utilise la somme cumulée des déviations par rapport à une cible.

Le graphique CUSUM représente la somme cumulée des déviations par rapport à la cible pour les mesures individuelles ou les moyennes de sous-groupes.

Cette méthode utilise la statistique suivant:

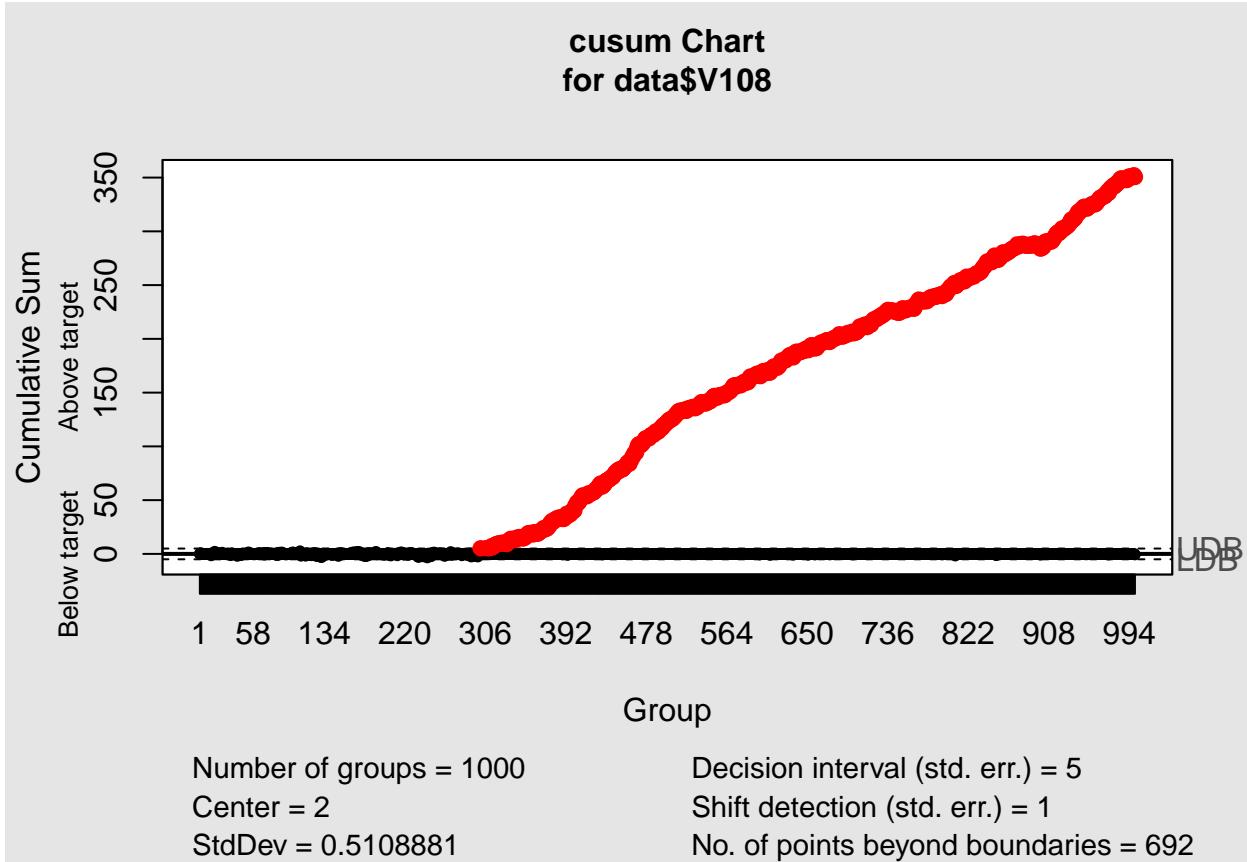
$$C_l = \sum_{j=1}^l (\bar{x}_j - \mu_0)$$

### 3.1.1 CUSUM pour colonne 107



Il y a 497 points hors de contrôle, donc 50% des points. Cette carte ne semble pas adaptée

### 3.1.2 CUSUM pour colonne 108



Pour cette colonne 108, on trouve que 692 points hors de contrôle, qui correspond avec 69.2% des points. Cette carte ne semble pas adaptée

## 3.2 Méthode EWMA

La Exponentially Weighted Moving Average (EWMA) est une statistique de suivi du processus qui établit la moyenne des données d'une manière qui donne de moins en moins de poids aux données à mesure qu'elles sont éloignées dans le temps. Pour la technique de contrôle EWMA, la décision dépend de la statistique EWMA, qui est une moyenne pondérée exponentiellement de toutes les données antérieures, y compris la mesure la plus récente. Par le choix du facteur de pondération,  $\lambda$ , la procédure de contrôle EWMA peut être rendue sensible à une dérive faible ou progressive du processus.

La statistique qui est calculée est :

$$EWMA_t = \lambda Y_t + (1 - \lambda) EWMA_{t-1}$$

pour  $t = 1, 2, \dots, n$  où

- $EWMA_0$  est la moyenne des données historiques
- $Y_t$  est l'observation au temps  $t$
- $n$  est le nombre d'observations à surveiller, y compris  $EWMA_0$
- $0 < \lambda \leq 1$  est une constante qui détermine la profondeur de mémoire de l'EWMA.

Le paramètre  $\lambda$  détermine le taux auquel les données “plus anciennes” entrent dans le calcul de la statistique EWMA. Une valeur de  $\lambda = 1$  implique que seule la mesure la plus récente influence l’EWMA. Ainsi, une grande valeur de  $\lambda$  (plus proche de 1) donne plus de poids aux données récentes et moins de poids aux données plus anciennes ; une petite valeur de  $\lambda$  (plus proche de 0) donne plus de poids aux données plus anciennes. La valeur de  $\lambda$  est généralement fixée entre 0.2 et 0.3 bien que ce choix soit quelque peu arbitraire.

La variance estimée de la statistique EWMA est d’environ

$$s_{EWMA}^2 = \frac{\lambda}{2 - \lambda} s^2$$

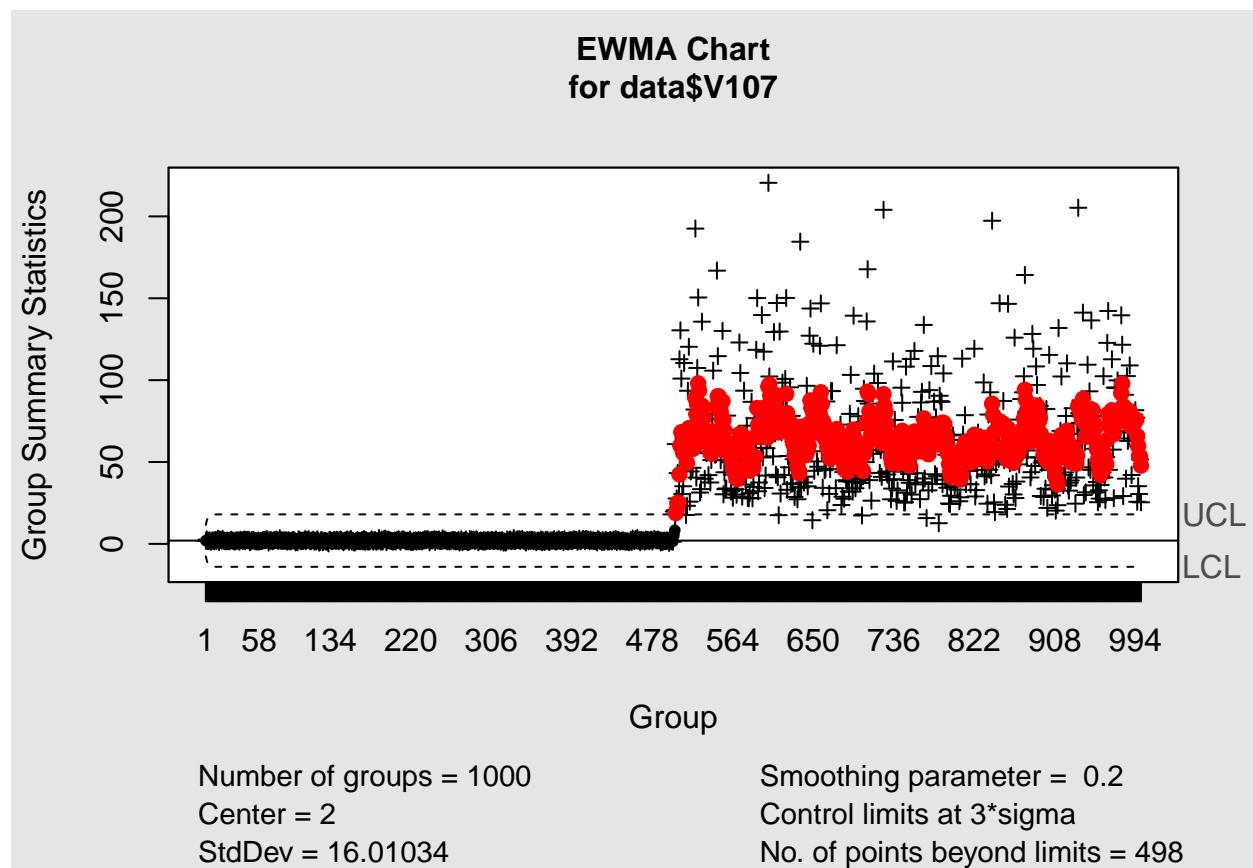
lorsque  $t$  n'est pas petit et où  $s$  est l'écart type calculé à partir des données historiques.

La ligne centrale de la carte de contrôle est la valeur cible ou  $EWMA_0$ . Les limites de contrôle sont :

$UCL = EWMA_0 + ks_{EWMA}$  et  $LCL = EWMA_0 - ks_{EWMA}$  où le facteur  $k$  est soit fixé à 3, soit choisi en utilisant les tables de Lucas et Saccucci (1990).

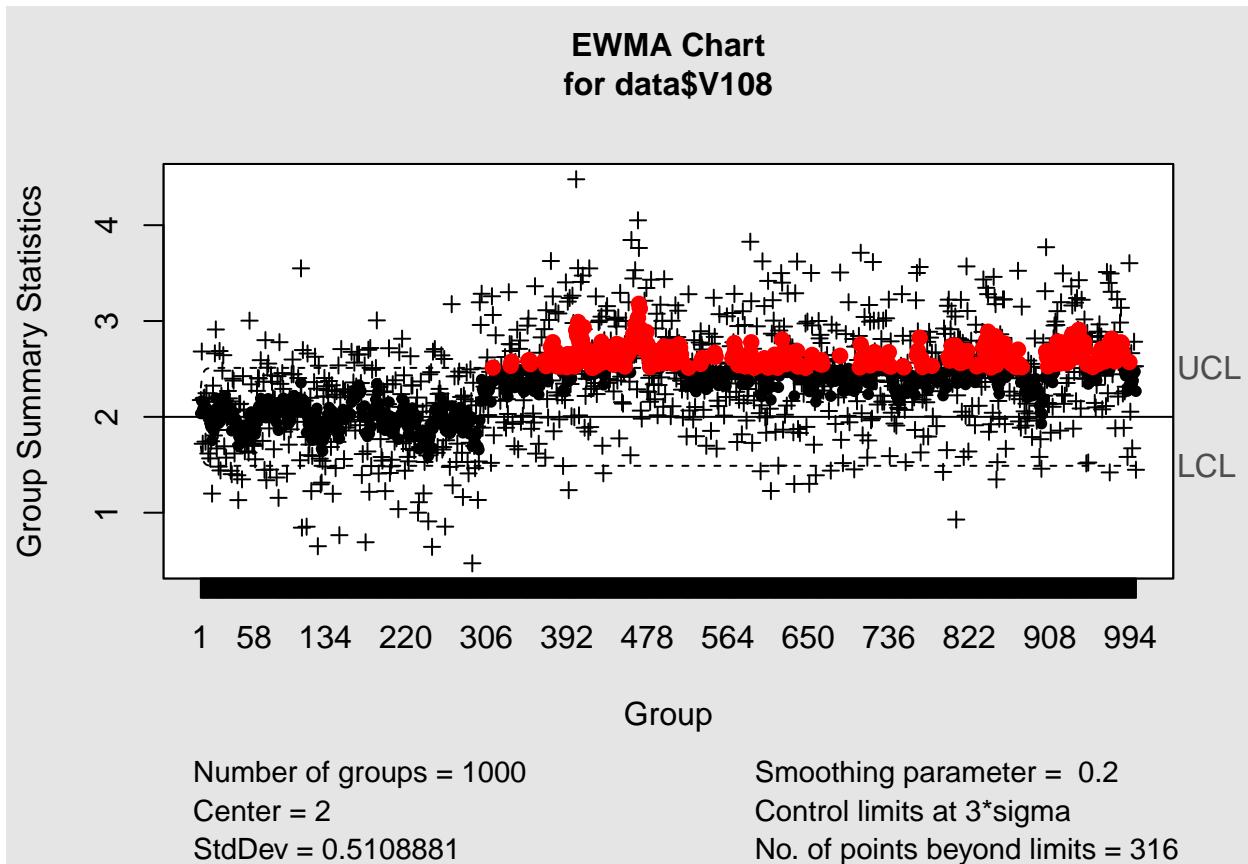
On applique maintenant la fonction EWMA :

### 3.2.1 Méthode EWMA pour la colonne 107



Le processus n'est pas sous de contrôle, avec 498 états qui correspondent à 49.8% points hors de contrôle.

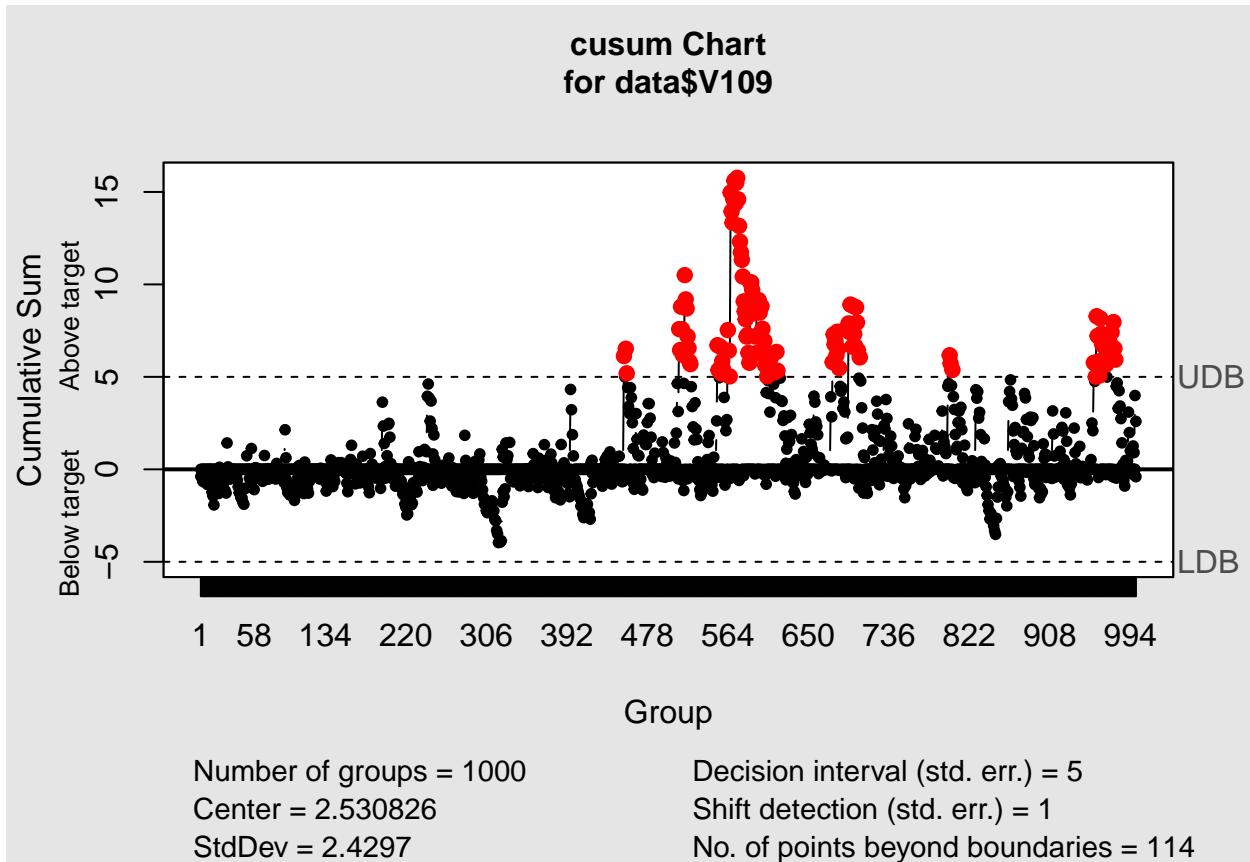
### 3.2.2 Méthode EWMA pour la colonne 108



Ce processus est aussi hors de contrôle, avec 316 points qui correspondent à 31.6% hors de contrôle.

## 4 Partie 4: Détection de rupture

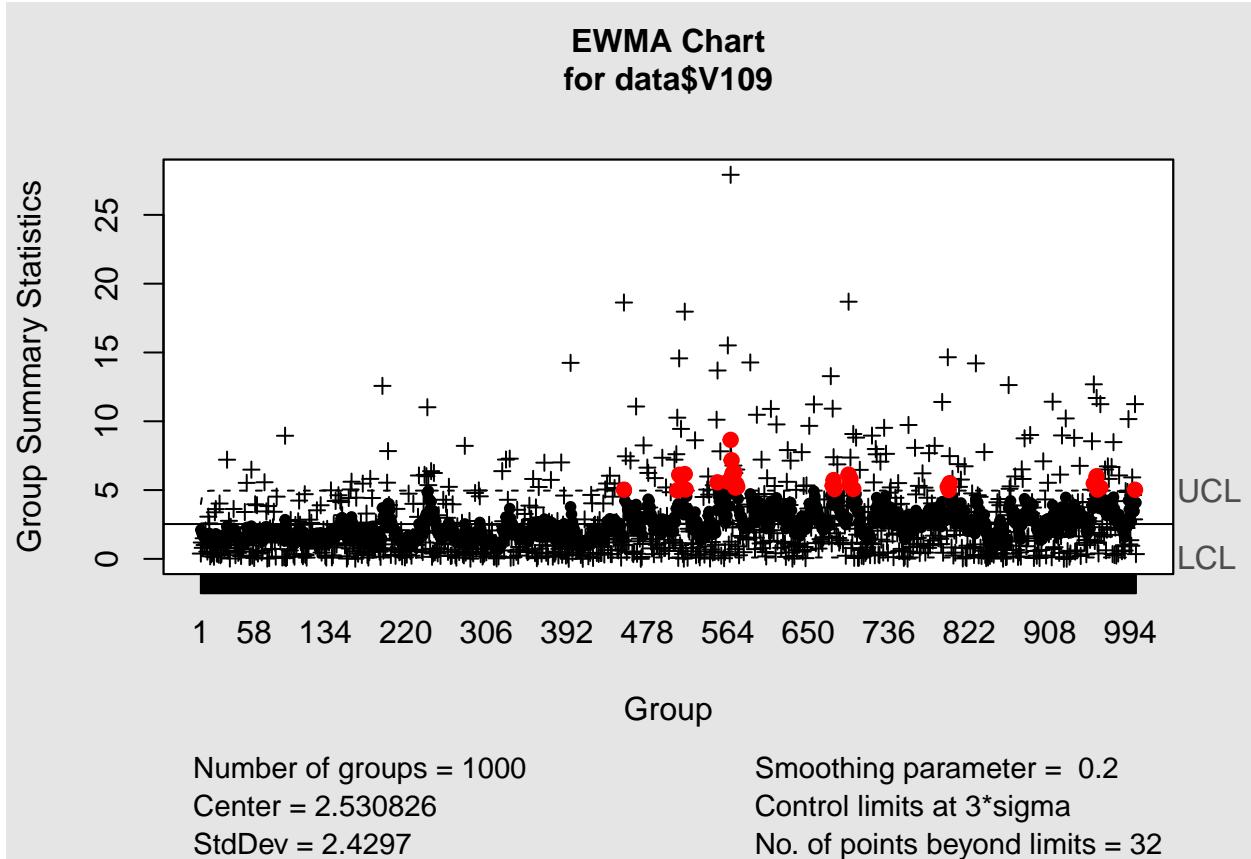
L'idée de cette partie est de trouver l'instant de rupture pour la colonne 109. On visualise tout d'abord nos données avec la carte CUSUM et EWMA



En regardant sur cette carte, on trouve les états qui passent la borne supérieur(UDB=5), ou à partir de cet instant, il y a un rupture. On trouve maintenant le premier point qui passe l'intervalle de contrôle

```
## $upper
## [1] 453 455 456 512 513 514 515 516 518 519 520 521 522 523 524 553 554 556
## [19] 557 558 559 564 565 566 567 568 569 570 571 572 573 574 575 576 577 578
## [37] 579 580 581 582 583 584 585 586 587 588 589 590 591 592 593 594 595 596
## [55] 597 598 599 600 601 602 603 604 606 610 611 612 614 616 617 676 677 678
## [73] 679 680 681 682 683 693 694 695 696 697 698 699 700 701 702 703 705 801
## [91] 802 804 955 957 958 959 960 961 962 963 964 965 966 967 968 970 971 972
## [109] 973 974 975 976 977 978
```

Alors l'état 453 qui est le premier point qui passe l'intervalle de contrôle



En regardant sur cette carte, on trouve aussi les états qui passent la borne supérieur, ou à partir de cet instant, il y a un rupture. On trouve maintenant le premier point qui passe l'intervalle de contrôle

```
## 453 510 512 513 514 518 519 553 564 567 568 569 570 571 572 573 574 676 677 678
## 453 510 512 513 514 518 519 553 564 567 568 569 570 571 572 573 574 676 677 678
## 693 694 695 698 799 800 801 955 958 959 962 999
## 693 694 695 698 799 800 801 955 958 959 962 999
```

On trouve, comme la carte CUSUM, l'état 453 est le premier point qui passe la borne supérieure(UCL=5)

Pour détecter la rupture, on utilise l'algorithme de détection: CUSUM.

- Problème de détection d'un changement dans la suite indépendante  $(X_k)_{k \geq 1}$

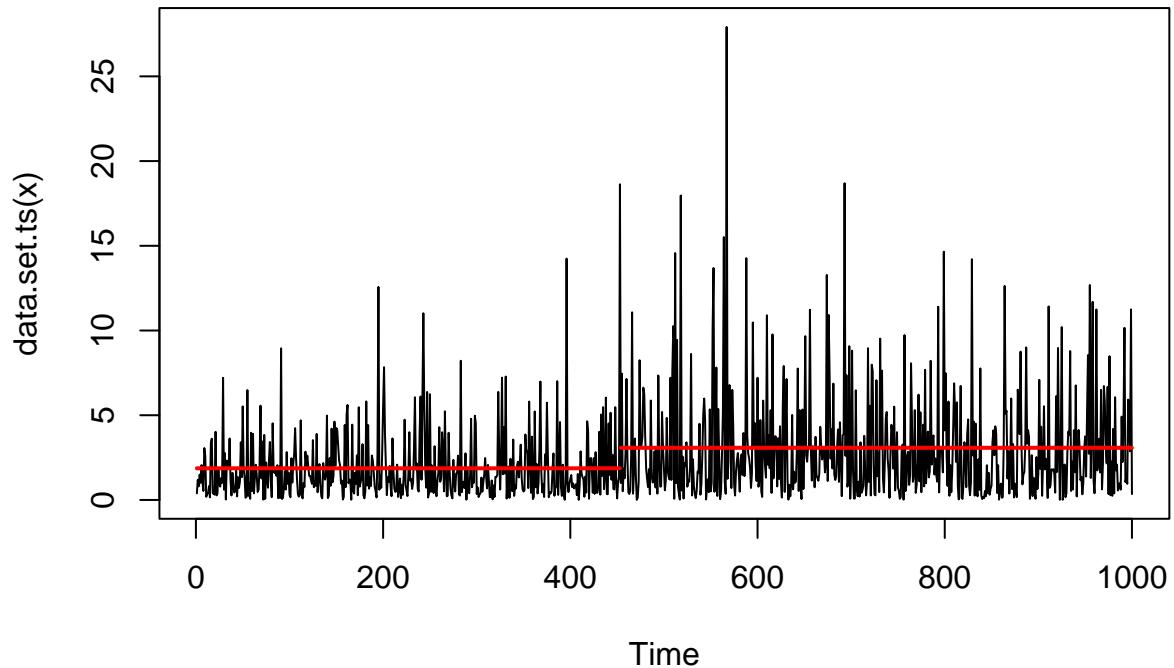
$$s_1^t - \min_{0 < k < t} S_1^k = \max_{1 < k < t} S_k^t, S_k^t = \sum_{i=k}^t \log \frac{p_{\theta_1}(X_i)}{p_{\theta_0}(X_i)}$$

- Accepter l'hypothèse de changement :

$$T_c = \inf(t > 1 : \max_{1 < k < t} S_k^t > h)$$

- Soit  $Z_i = \log \frac{p_{\theta_1}(X_i)}{p_{\theta_0}(X_i)}$ , donc on a:  $S_1^t = \sum_{i=1}^t Z_i$  qui change de signe après l'instant de rupture, c'est à dire

$$E_0(\log \frac{p_{\theta_1}(X_i)}{p_{\theta_0}(X_i)} | i < t_0) < 0 \text{ et } E_1(\log \frac{p_{\theta_1}(X_i)}{p_{\theta_0}(X_i)} | i > t_0) > 0$$



```
## [1] 453
```

En appliquant la méthode de Cusum dans la librairie changepoint, on trouve donc l'instant de rupture qui est à 453, qui est correspond avec ce qu'on trouve sur la carte contrôle de CUSUM et EWMA