Cours de contrôle de gestion appliqué à l'assurance

Partie 4

Cours complet, exercices d'application et bibliographie sur
www.cours-assurance.org

Partie 4 : Techniques de projection et applications aux calculs assurantiels

Plan de la partie :

- **Ajustements et régressions**
- **Séries chronologiques**

1. Techniques d'ajustement

1.1. Courbe d'ajustement

Soit un échantillon de taille *n* ayant comme mesures $x_1, x_2, \ldots, x_n$ et $y_1, y_2, \ldots, y_n$.
On peut reporter l'ensemble de ces points sur un graphique orthonormé et on obtient un nuage de points (*diagramme de dispersion*). Sur ce graphique, on peut souvent tracer une courbe épousant au mieux les données, c'est la *courbe d'ajustement*.
Au vu du diagramme de dispersion, on peut alors essayer de trouver une forme analytique de cette courbe d'ajustement.

1.2. La méthode des moindres carrés

1.2.1. Ajustement linéaire

Soit *n* points *Pi* de coordonnées (*xi*, *yi*). Il s'agit de trouver les coefficients *a* et *b* d'une droite *y* tel que la somme des carrés des distances verticales depuis les points *Pi* jusqu'à la droite *y* soit minimale.

L'écart de la droite *y* par rapport à un point *Pi* est donné par

$$e_i = y_i - (ax_i + b)$$

Si nous mesurons la distance le long de l'axe des *y* depuis le point *Pi* jusqu'à l'ordonnée du point *xi* sur la droite *y*. Un tel écart peut être négatif ou positif et c'est pourquoi nous ne minimisons pas la somme des écarts, mais la somme des carrés des écarts. On a ainsi

$$E = \sum_{i=1}^{n} {e_i}^2 = \sum_{i=1}^{n} [y_i - (ax_i + b)]^2 = \sum_{i=1}^{n} (y_i - ax_i - b)^2 \qquad (\#)$$

Le but est de minimiser *E* en fonction de *a* et de *b* ; *xi* et *yi* étant des données constantes.

Après calculs, on trouve :

$$a = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n}(x_i - \bar{x})^2} \qquad (*)$$

Et $b = \bar{y} - a\bar{x}$

La covariance et la variance étant définies ainsi :

$$Cov(x,y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})$$

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

L'expression (*) peut aussi s'écrire :

$$a = \frac{Cov(x,y)}{V(x)}$$

Pour résumer, la droite des moindre carrés de y en x passe par le point "moyen" de coordonnées$(\bar{x}, \bar{y})$, a pour pente $a = \frac{Cov(x,y)}{V(x)}$ et comme ordonnée à l'origine $b = \bar{y} - a\bar{x}$.

En pratique, on peut simplifier les calculs à l'aide de la formule suivante :

$$a = \frac{Cov(x,y)}{V(x)} = \frac{\overline{xy} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2}$$

Pour mesurer la qualité de l'ajustement, on peut calculer la somme des carrés des erreurs individuelles qui correspond au E détaillé à l'expression (#)

1.2.2. Ajustements non linéaires

Dans certains cas, l'ajustement à une fonction linéaire n'est pas adéquat : un ajustement des données à une fonction non linéaire doit être envisagé. Les deux cas que nous considérons sont ceux où on peut se ramener par simple transformation à un ajustement affine.

1.2.2.1. Ajustement à une fonction puissance

Supposons que les variables statistiques *x* et *y* sont liées par une relation de la forme :

$$y = bx^a$$

Dans ce cas, cette équation peut être transformée en prenant le logarithme :

$$ln(y) = ln(bx^a)$$

$$= ln(b) + ln(x^a)$$

$$= ln(b) + aln(x)$$

En effectuant les changements de variables suivants :

$\mathrm{Y} = \ln(y)$

$\mathrm{X} = \ln(x)$

$\mathrm{B} = \ln(b)$

Nous nous ramenons au cas étudié précédemment :

$\mathrm{Y} = \mathrm{aX} + \mathrm{B}$

1.2.2.2. Ajustement à une fonction exponentielle

Supposons que les variables statistiques *x* et *y* sont liées par une relation de la forme :

$y = be^{ax}$

Dans ce cas, cette équation peut être transformée en passant aux logarithmes :

$\ln(y) = \ln(be^{ax})$

$= \ln(b) + \ln(e^{ax})$

$= \ln(b) + \mathrm{ax}$

En effectuant les changements de variables suivants :

$\mathrm{Y} = \ln(y)$

$\mathrm{B} = \ln(b)$

Nous nous ramenons au cas étudié dans les paragraphes précédents :

$\mathrm{Y} = \mathrm{ax} + \mathrm{B}$

2. Série chronologique

Les exemples développés dans cette section sont totalement fictifs.

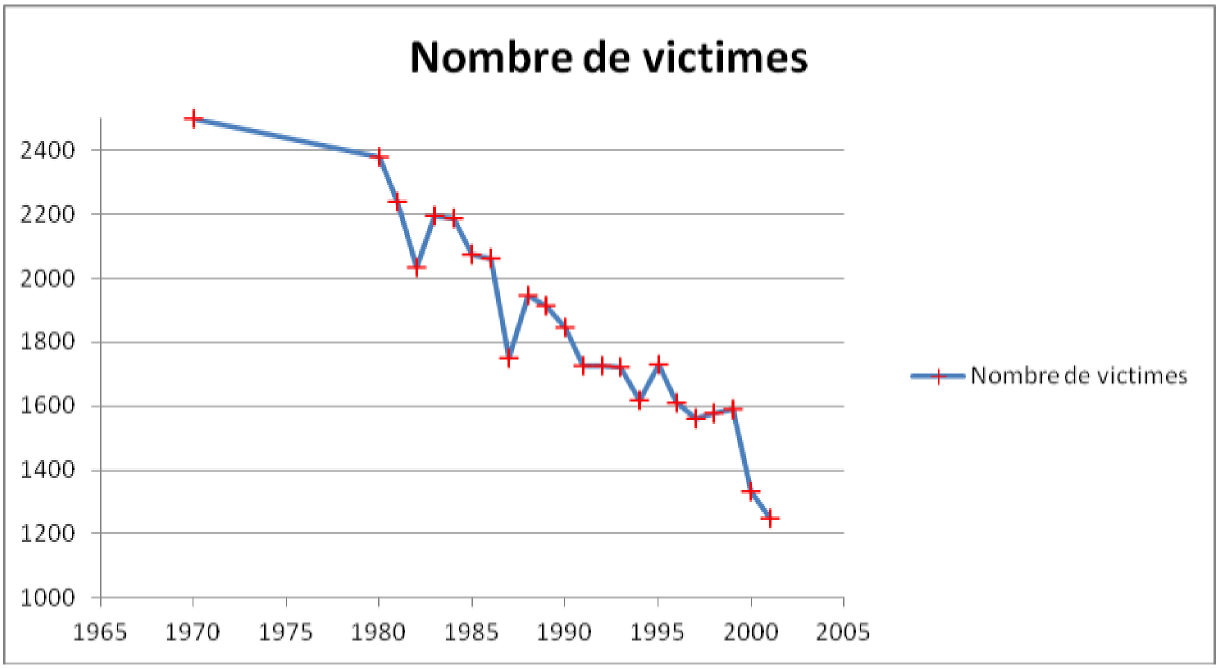
2.1. Introduction

Une *série chronologique* est une série statistique ordonnée en fonction du temps.

Exemples : le cours journalier en bourse (à la clôture) d'une action, le nombre de nouveaux assurés comptabilisés par une compagnie d'assurance chaque mois. L'étude de ces séries est intéressante car elle peut permettre de prévoir l'évolution du phénomène observé dans le temps.

Voici une série chronologique indiquant le nombre de victimes dans des accidents de la route au Luxembourg ayant été indemnisées par leur assureur :

Année	Nombre de victimes
1970	2497
1980	2380
1981	2239
1982	2034
1983	2193
1984	2185
1985	2075
1986	2062
1987	1749
1988	1946
1989	1914
1990	1846
1991	1725
1992	1725
1993	1720
1994	1616
1995	1730
1996	1609
1997	1559
1998	1575
1999	1588
2000	1331
2001	1246



Sur la représentation graphique d'une série chronologique, on peut distinguer les composantes fondamentales suivantes :

- le *mouvement de tendance générale* ou *trend* indiquant l'évolution générale du phénomène étudié
- les *mouvements cycliques* sur une grande période autour du trend. Ces mouvements peuvent être périodiques (exemple : récession et expansion économique, etc.)
- les *mouvements saisonniers* ou *variations saisonnières* sont des variations se reproduisant périodiquement à des moments bien déterminés (exemple : vente de contrat complémentaire santé à la rentrée scolaire, etc.)
- les *mouvements accidentels* ou *résiduels* sont dus à des facteurs exceptionnels pour la plupart imprévisibles (risque de guerre, etc.)

2.2. Estimation de la tendance

Il est clair qu'afin de pouvoir estimer la tendance, c'est-à-dire le mouvement d'un phénomène observé sur un grand intervalle de temps, il faut disposer d'une série statistique sur une longue période.

Disposant de ces données, le premier travail consiste à effectuer une représentation graphique adéquate permettant d'avoir une vue globale du phénomène étudié.

2.2.1. Ajustement analytique

Si les données sur le graphique se présentent sous une forme ressemblant à une courbe connue (droite, parabole, exponentielle, etc.), on peut essayer de dégager une forme analytique pour cette courbe.

Nous ajustons cette courbe avec les méthodes étudiées au chapitre précédent, ce qui permet de déterminer la tendance de la série chronologique.

2.2.2.Moyennes mobiles

Afin d'éliminer ou d'amortir les mouvements cycliques, saisonniers et accidentels, on peut aussi utiliser la technique des *moyennes mobiles*. On procède ainsi en quelque sorte au *lissage de la courbe*.

Le principe de cette méthode est de construire une nouvelle série obtenue en calculant des moyennes arithmétiques successives de longueur *p* fixe à partir des données originales. Chacune de ces moyennes obtenues correspondra au “milieu” de la période pour laquelle la moyenne arithmétique vient d'être calculée.

Exemple : le tableau ci-dessous contient des mesures d'un phénomène relevées à 9 instants différents.

t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
4	6	5	3	7	5	4	3	6

Si nous calculons les moyennes mobiles d'ordre 3, nous obtenons les valeurs suivantes :

t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
4	6	5	3	7	5	4	3	6
	5.00	4.67	5.00	5.00	5.33	4.00	4.33	

La moyenne mobile d'ordre 3 pour t_6 est : $\frac{7+5+4}{3} \approx 5.33$

On constate que, vu la façon de calculer ces moyennes, les deux valeurs extrêmes ont disparu, c'est-à-dire une de chaque côté.

En calculant les moyennes mobiles d'ordre 5, nous aurons :

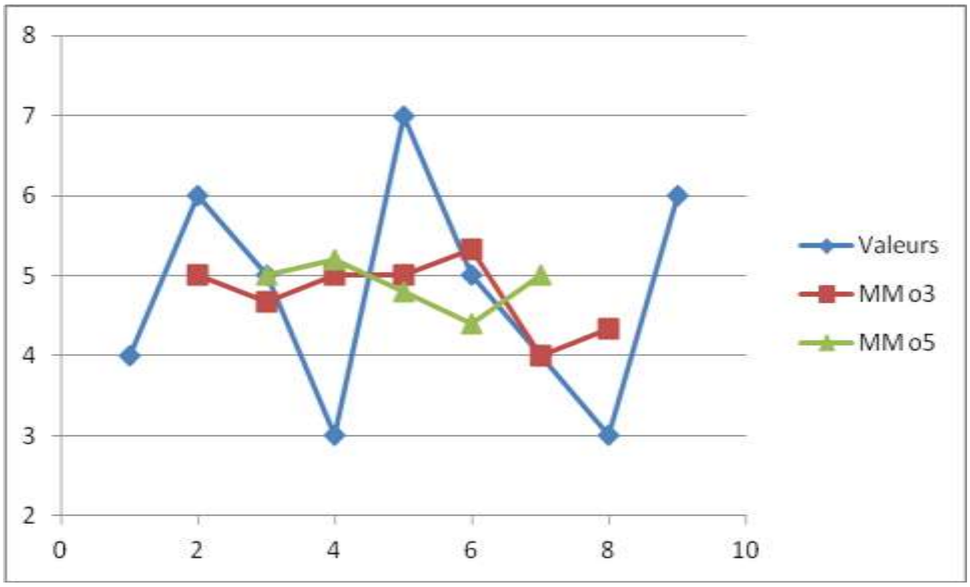
t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
4	6	5	3	7	5	4	3	6
		5	5.2	4.8	4.4	5		

La moyenne mobile d'ordre 3 pour t_4 est : $\frac{6+5+3+7+5}{5} \approx 5.2$

On constate que, vu la façon de calculer ces moyennes, les quatre valeurs extrêmes ont disparu, c'est-à-dire deux de chaque côté.

De manière générale, on constate que si *p* est impair, donc de la forme 2k + 1, à chaque extrémité *k* valeurs sont perdues.

En représentant graphiquement ces résultats, nous remarquons bien la tendance au “lissage” de la représentation originale avec l'utilisation de la technique des moyennes mobiles.



En choisissant *p* pair, nous sommes confrontés au problème que les moyennes obtenues ne correspondront pas à une abscisse existante, mais chevaucheront entre deux de ces valeurs.

Exemple : dans la série chronologique précédente, si nous calculons les moyennes mobiles d'ordre 4, nous obtenons des valeurs pour $t_{2,5}$, $t_{3,5}$, etc., ce qui n'est pas pratique. C'est pour cette raison qu'on calcule une moyenne sur 5 valeurs mais en prenant soin de pondérer les deux valeurs extrêmes par 1/2 au lieu de 1 pour les autres valeurs. Il faut quand même veiller à diviser par 4 (et non par 5) ! Ceci nous garantit que chaque valeur n'est prise en compte qu'une seule fois.

Nous aurons donc les moyennes mobiles d'ordre 4 suivantes :

t_1	t_2	t_3	t_4	t_5	t_6	t_7	t_8	t_9
4	6	5	3	7	5	4	3	6
		4.88	5.13	4.88	4.75	4.63		

Pour mieux comprendre, voici le calcul pour t_3 de la moyenne mobile d'ordre 4 :

$$\frac{\frac{1}{2}4+6+5+3+\frac{1}{2}7}{4} \approx 4.88$$

2.3. *Estimation de la nature des mouvements saisonniers*

Pour effectuer l'analyse des mouvements saisonniers, on essaie de déterminer si on est en présence d'une série dans laquelle pour une observation *O* donnée :

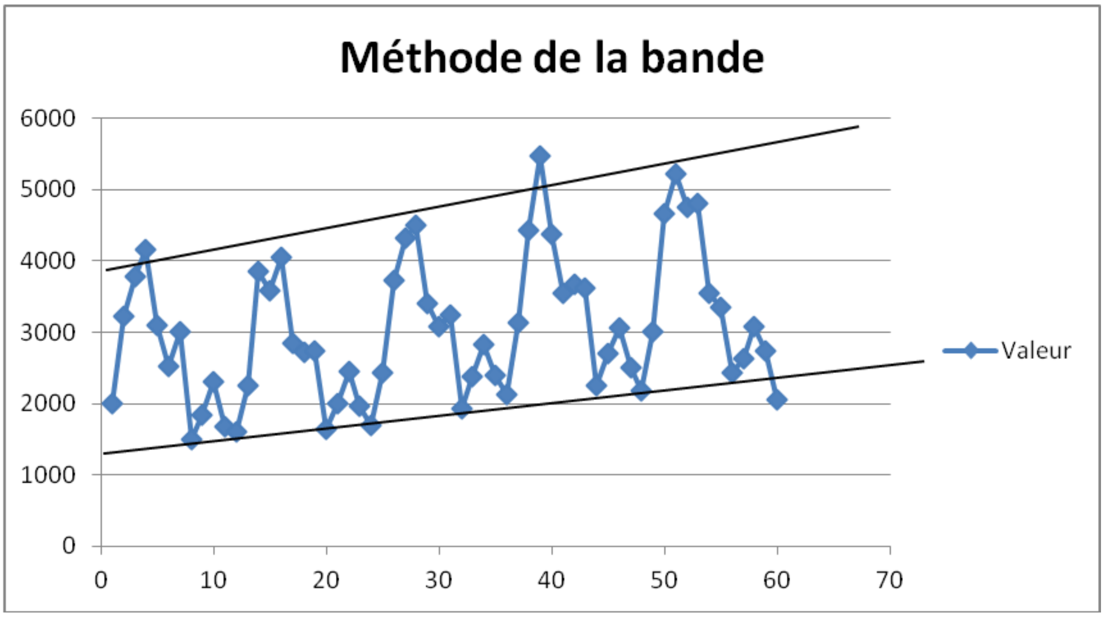
- la variation saisonnière *S* s'ajoute simplement à la résultante des autres composantes *R* ; O = R+S
  C'est le *modèle additif*
- la variation saisonnière *S* est proportionnelle à la résultante des autres composantes *R* : S=cR et alors O=R+S=R+cR= (1+c)R
  C'est le *modèle multiplicatif.*

2.3.1. Méthode graphique :

Afin de faire cette distinction, on peut se baser sur une méthode graphique : on fait un graphique représentant la série chronologique, puis on trace une droite passant respectivement par les minima et par les maxima de chaque saison. Si ces deux droites sont parallèles, nous sommes en présence d'un modèle additif. Dans le cas contraire, c'est un modèle multiplicatif.

Exemple :
Nous étudions ces méthodes sur un exemple concret en nous basant sur la série chronologique "Nouveaux contrats d'assurances souscrits selon le mois"

Année	Janvier	Février	Mars	Avril	Mai	Juin	Juillet	Août	Septembre	Octobre	Novembre	Décembre
1996	2006	3224	3789	4153	3100	2527	3015	1504	1847	2314	1673	1602
1997	2247	3862	3586	4047	2838	2727	2730	1648	2007	2450	1966	1695
1998	2433	3723	4325	4493	3399	3083	3247	1928	2377	2831	2388	2126
1999	3127	4437	5478	4384	3552	3678	3611	2260	2699	3071	2510	2182
2000	3016	4671	5218	4746	4814	3545	3341	2439	2637	3085	2737	2055



Sur cet exemple, nous constatons que les deux droites ne sont pas parallèles, nous sommes donc en présence d'un modèle multiplicatif.

2.3.2. Méthode analytique :

On calcule les moyennes et écarts-types pour chacune des périodes considérées et on calcule la droite des moindres carrés $\sigma = a\bar{x} + b$. Si *a* est nul, c'est un modèle additif, si $\neq 0$, le modèle est multiplicatif.

Période	moyenne	Écart-type
1996	2562.8	850.7

1997	2650.3	782.3
1998	3029.4	803.6
1999	3415.8	946.6
2000	3525.3	1023.4

En calculant la droite des moindres carrés, on obtient a = 0.195 et b = 289.037, ce qui confirme encore une fois que pour cet exemple, nous sommes bien en présence d’un modèle multiplicatif.