

Théorie des prévisions

(Exercices inclus)

A partir du TETRAPRO, édité par Mr. H. Leijon, ITU



**UNION INTERNATIONALE DES TELECOMMUNICATIONS
INTERNATIONAL TELECOMMUNICATION UNION
UNION INTERNACIONAL DE TELECOMUNICACIONES**



Théories des Prévisions

1. SUPPOSITIONS GENERALES POUR LES PREVISIONS
 - 1.1 Pourquoi prévoir?
 - 1.1.1 *Densité téléphonique et le développement économique*
 - 1.1.2 *But des prévisions*
 - 1.1.3 *Prévoir, planification et programme de travail*
 - 1.2 Quand les prévisions sont nécessaires?
 - 1.3 Quoi prévoir?
 - 1.3.1 *Sujet des prévisions*
 - 1.3.2 *Envergure des prévisions*
 - 1.3.3 *Données nécessaires pour la planification*
 - 1.4 Comment faire les prévisions?
 - 1.4.1 *Exigences de base*
 - 1.4.2 *Surveillance de prévision*
 - 1.4.3 *Comment commencer?*
2. METHODES DE PREVISION
 - 2.1 Introduction
 - 2.2 Analyse statistique des demandes
 - 2.2.1 *Régression à 2 variables*
 - 2.2.2 *Régression multiple*
 - 2.3 Méthode de tendance
 - 2.3.1 *Analyse des séries chronologiques*
 - 2.3.2 *Extrapolation des tendances*
 - 2.4 Jugement individuel
 - 2.5 Autres méthodes
3. PREVISION POUR LA PLANIFICATION DES CENTRAUX
4. PREVISION POUR LA PLANIFICATION DES RESEAUX
 - 4.1 Introduction
 - 4.2 Modèle du trafic total
 - 4.3 Prévision point-à-point
 - 4.4 Méthode de Khruithof à double facteur
5. CONCLUSIONS
6. REFERENCES
7. EXERCICES

1. SUPPOSITIONS GENERALES DES PREVISIONS

1.1 Pourquoi prévoir ?

1.1.1 *Densité téléphonique et développement économique*

Les facilités des télécommunications du pays dépendent du niveau économique dans ce pays. Pour développer d'avantage l'économie, certaines facilités de base en télécommunications sont nécessaires. Cela est pris pour l'ensemble des pays. La corrélation entre le nombre de téléphones par habitants et le produit intérieur brut (PIB - GDP) par tête existe, comme montré dans la Figure 1.1/1. Certaines déviations arrivent comme le résultat des différences dans la structure économique des différents pays. Il est facilement compris que le pays avec une économie basée principalement sur l'agriculture n'a besoin que de peu de téléphones par tête contrairement aux pays hautement industrialisés.

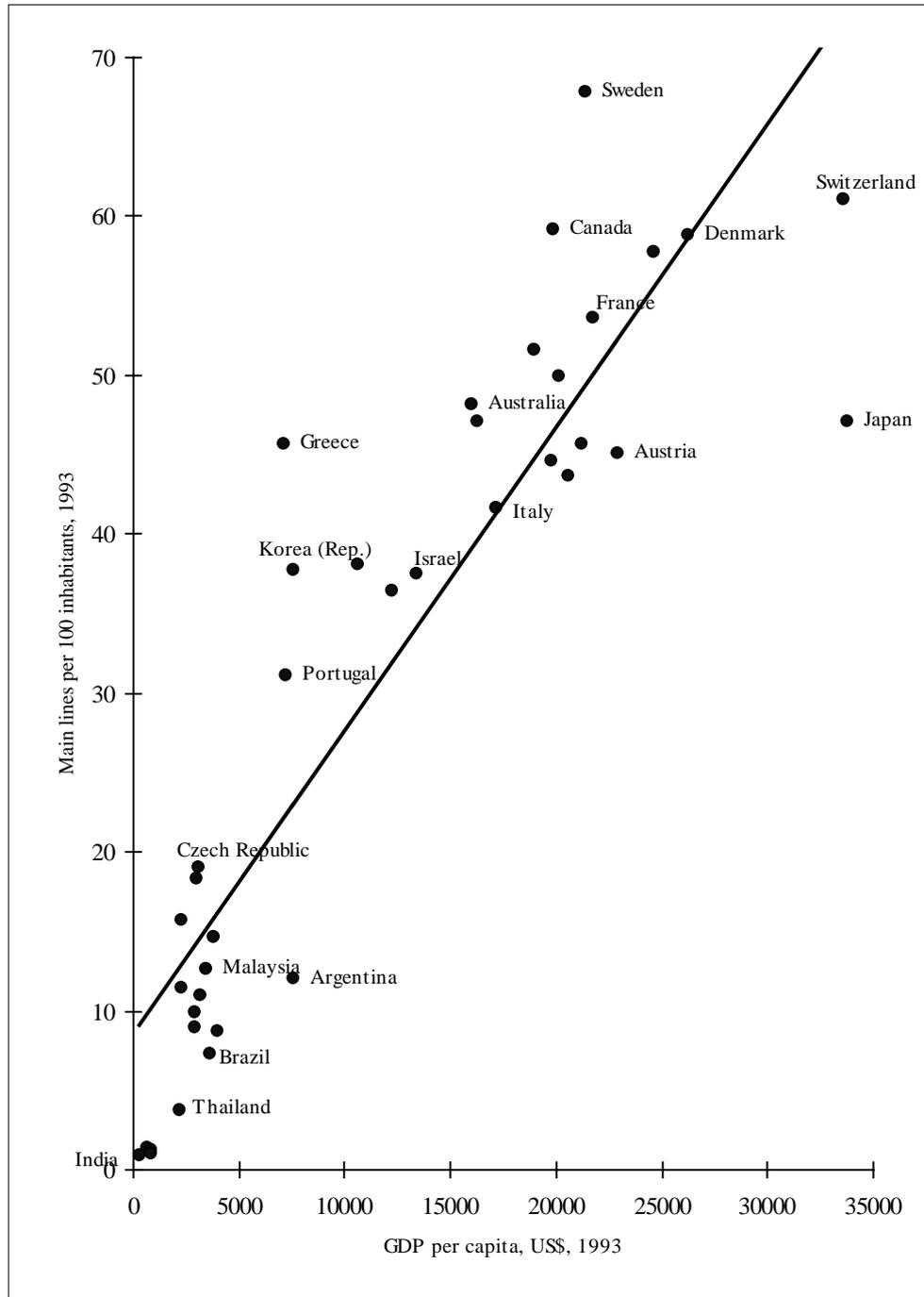


Figure 1.1/1 : Densité téléphonique et le produit intérieur brut (GDP) pour habitant dans quelques pays, 1993

La demande du téléphone dépend premièrement de la spécialité d'occupation de la population. Le besoin en téléphone est faible dans des villages qui se supportent eux même mais croit par le fait que les gens quittent le secteur agricole pour d'autres activités (industrie, commerce, services publiques, etc.). Le Gouvernement et les activités commerciales privées demandent le téléphone pour fonctionner. La concentration des gens de plus en plus dans les zones urbaines peut encore donner une demande croissante pour les téléphones résidentiels, mais cela concerne uniquement les personnes qui ont un certain niveau de vie. Une certaine demande de communications avec les zones rurales devrait aussi venir des zones urbaines. La demande est, bien sûr, dépendante aussi bien des habitudes et préférences nationales que des tarifs téléphoniques. Pour éviter l'influence de cette dernière, toutes les investigations et pronostiques devaient être faites sur les suppositions pour que les tarifs téléphoniques ne soient pas prohibitifs mais seulement suffisants pour supporter la maintenance adéquate et l'expansion du réseau téléphonique.

Pour estimer la future demande des télécommunications, le développement général espéré et, spécialement le développement économique doivent être pris en considération. Les données publiées dans les statues du courant économique et les informations relatives aux plans de développement économiques sont utiles.

1.1.2 *But des prévisions*

L'extension des services téléphoniques demande l'approvisionnement des appareils téléphoniques, équipements des lignes d'abonnés, équipement des centraux, les faisceaux de jonctions et de circuits. Cependant, il pourrait avoir un temps d'écologie entre l'identification des besoins - ou plutôt les besoins futurs - et le moment de réalisation. La durée de temps entre l'identification et l'approvisionnement pour ce besoin est considérable. Pour éviter de longues périodes d'attente et la congestion élevée, il est désirable de déterminer les besoins à l'avance. Cela rend possible l'extension du matériel au bon moment parce que l'action nécessaire peut être prise au temps approprié.

Une prévision devraient, cependant, produire premièrement l'estimation exacte des demandes futures pour les équipements des télécommunications.

1.1.3 *Prévision, planification et programme de travail*

Les prévisions donnent des bases pour les plans. Les plans sont considérés par la gestion qui fait leur décisions. Suite aux décisions, le programme de travail est formulé. Le programme de travail demande une planification détaillée, pour quels besoins doit être faite jusqu'à ce que tous les équipements deviennent opérationnels. Il y a la planification à long terme, à moyen terme et à court terme, chacune d'elles a ces propres exigences qui concernent le détail qui doit être spécifié à quel point, si le programme et l'emploi du temps devait être suivi. Chaque type de planification doit utiliser, plus ou moins, des prévisions détaillées pour les quantités demandées.

On doit utiliser les définitions qui suivent:

Prévision :	Une prévision est la prédiction de la demande future, généralement exprimée en quantités.
Plan :	Un plan est une proposition d'actions et estime le coût d'exécution du plan. Il peut inclure des lignes alternatives d'actions.
Décision :	Une gestion décide quel action devait être prise. Elle approuve le plan qui est alors convertis en programme de travail.
Programme de travail:	Un programme de travail est basé sur des plans approuvés et définit les actions à prendre comme de travail son timing.
Planning	Un programme de travail demande généralement le planning détaillé pour toutes les actions qui devraient être prises pour réaliser le programme de travail. Le timing de chaque action est d'une grande importance ici. Plus au moins un plan détaillé de travail doit être prêt au moment donné.

La relation entre les concepts ci-dessus est illustrée dans la Figure 1.1/2

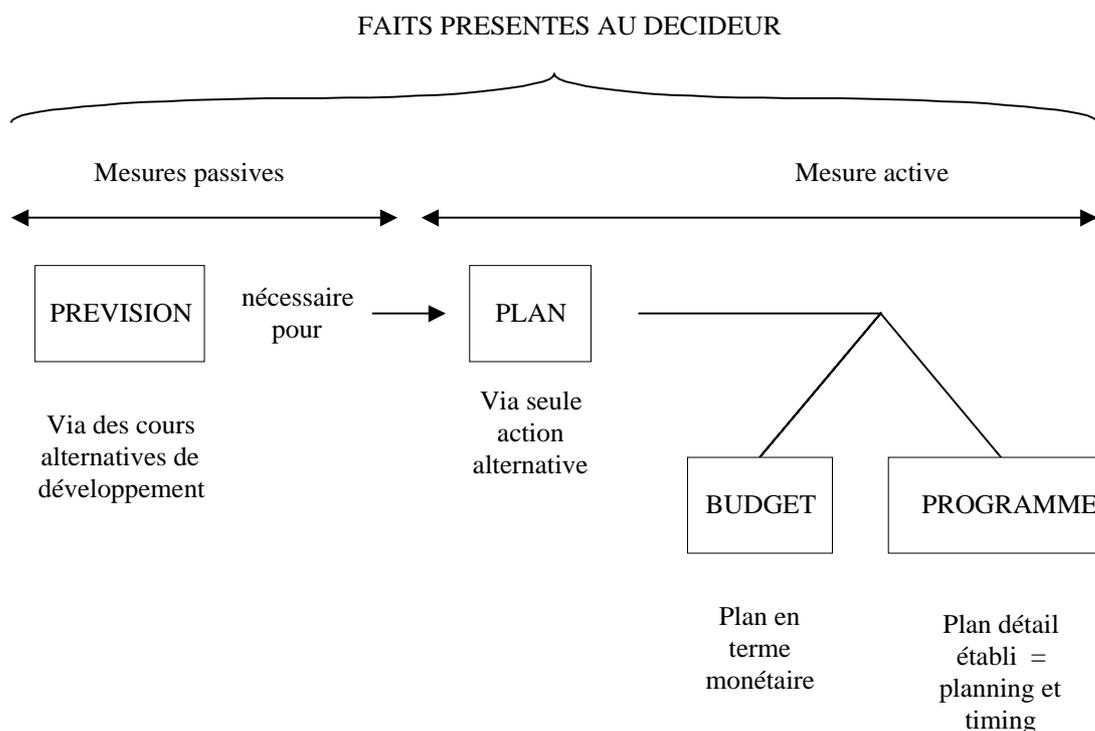


Figure 1.1/2 : Relation entre les prévisions et le plan

Puisque le programme de travail et les investissements sont basés sur les prévisions, il est très important à celles-ci d'être plus exactes que possible. Il est aussi désirable que le prévisionniste définisse le degré d'incertitude des prévisions. Cela permet au planificateur de savoir quand il peut être flexible.

La situation dans quelques pays en développement peut être telle que la demande est beaucoup supérieure à l'offre. Dans un tel cas, toutes les extensions doivent être utilisées immédiatement sans essentiellement réduire le trou entre l'offre et la demande. Il est très difficile de prévoir la demande dans de tels cas, puisque la demande existante ne peut pas être facilement mesurée ou estimée. La tâche donc du prévisionniste est d'estimer cette demande pour que le décideur puisse savoir l'effet de ses actions sur la situation offre/demande. Conclure une telle situation est aussi une matière relative à la politique suivie par l'administration. Est-ce qu'elle donne la priorité aux abonnés professionnels, au développement des zones locales, ou aux services de longue distance, etc.?

1.2 Quand la prévision est nécessaire ?

Les périodes de prévisions dépendent de périodes de planification. Qui dépendent à leur tour des temps de livraison des différents types d'équipements nécessaires. On donne quelques exemples ci-dessous:

Appareils d'abonnés	1-2 années
Équipement des centres de commutation	3-4 années
Câble d'abonnés	6-10 ans
Génie civil dans les réseaux locaux	10-15 ans
Constructions	10-20 ans
Sites pour construction	jusqu'aux 50 ans

La planification comme les prévisions est toujours divisée en:

Planification/prévision à long terme	plus de 10-15 ans
Planification/prévision à moyen terme	5-10 ans
Planification/prévision à court terme	moins de 5 ans

Les différents types de planification et de prévision concernent différents types d'équipements qui dépendent du moment d'utilisation de l'équipement. Clairement, les prévisions à court terme doivent être plus détaillées puisque elles alimentent la planification à court terme dans laquelle chaque détail du matériel doit être spécifié.

Une seule méthode de prévision ne doit pas être adéquate pour l'ensemble de périodes dans le temps. Pour cela, différentes méthodes devraient être utilisées pour les différentes périodes, et cela augmente l'ajustage du problème quand les prévisions pour des chevauchements ou non des périodes sont dérivées avec différentes méthodes, une fois que la divergence peut arriver entre les courbes de croissance.

Il n'y a pas de procédure standard pour faire face aux contradictions de prévisions. Dans la plus part des cas, le jugement basé sur l'expérience est utilisé et la courbe est tracée tout en faisant un lien entre les prévisions. Cela est montré dans la Figure 1.3/1.

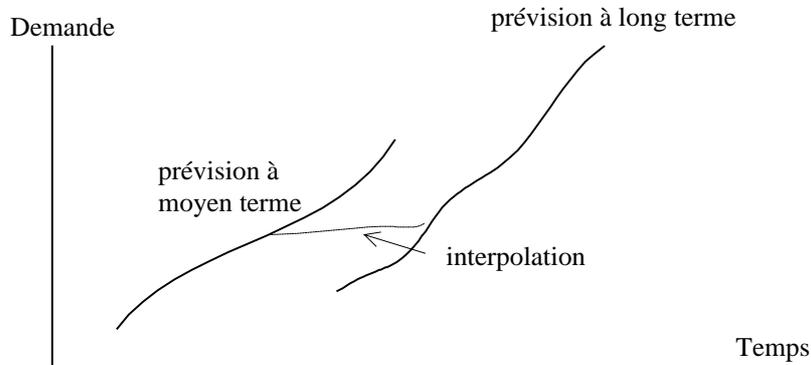


Figure 1.3/1 : Réduction des contradictions des prévisions

1.3 Prévoir quoi ?

1.3.1 *Sujet des prévisions*

Les prévisions concernent quatre domaines principaux:

Accommodation	Site Constructions
Lignes	Lignes d'abonnés direct au central (prévision du réseau d'abonnés) Réseau de jonction Réseau de faisceaux interurbains
Equipement	Appareils d'abonnés Equipement automatique de commutation Equipement des standards (des opératrices)
Numérotation	Plan de numérotation des centraux (stratégie à long - terme)

Les prévisions présentent généralement le nombre futur espéré de lignes directs de centre (DEL) ainsi que le trafic futur, ce dernier sera plus ou moins détaillé. Par conséquent, il y a:

- prévisions d'abonnés;
- prévisions du trafic.

1.3.2 *Envergure des prévisions*

Des prévisions peuvent concerner le total pour une entité donnée tel qu'un pays, une certaine partie du pays, zone métropolitaine, etc. Dans ce cas, elles sont appelées *prévisions totales* et contiennent peu de détail pour cette entité. Il est souvent suffisant uniquement pour les buts de planification à long terme. Une *prévision détaillée* fournit les données détaillées pour le développement futur espéré. Elle est séparément faite pour chaque unité ou le travail est valable.

Seulement les prévisions détaillées des abonnés spécifient la location exacte de chaque abonné.

Le détail des prévisions du trafic spécifie tous le trafic futur à l'intérieur de la zone concernée. Il est généralement basé sur les prévisions des abonnés et sur les mesures.

1.3.3 Données nécessaires pour la planification

Dépendants sur le besoins de planification, les résultats des prévisions devraient être présentés dans différentes manières. On donne les exemples suivants:

Distribution des abonnés à l'intérieur de la zone

- Localisation exacte des abonnés individuels.
- Nombre d'abonnés dans chaque grille de la carte à grilles.
- Nombre d'abonnés par zone de trafic dans la zone concernée.
- Nombre d'abonnés par catégorie d'abonné dans chaque zone de trafic.

Trafic

- Le trafic départ et arrivée par abonné.
- Le trafic départ et arrivée par abonné dans chaque zone de trafic.
- Le trafic départ et arrivée par abonné pour chaque catégorie d'abonnés.
- Le flux de trafic entre les zones.

On peut définir:

A = trafic total pour un groupe d'abonnés

N = nombre d'abonnés dans le groupe

α = trafic par abonné

y = intensité d'appels dans le groupe

h = temps moyen de maintien

Alors, $A = y h$, mais aussi $A = N \alpha$

Trafic A peut donc être prévu de manières différentes:

1. Directement, par exemple extrapolation de tendance des valeurs du trafic mesuré A pour les années passées.
2. Indirectement, par les prévision en premier soit de N ou de α , ou alternativement y et h .

Comme le trafic d'un groupe d'abonnés concerne les prévision, il est généralement favorable de faire les prévisions pour chaque catégorie d'abonné avant d'estimer le trafic total. La raison pour cela, est que la proportion entre les différentes catégories peut changer dans le futur.

1.4 Comment sont faites les prévisions ?

1.4.1 *Exigences de base*

Il y a deux exigences principales pour prévisions correctes:

- Une provision adéquate et informations exacte du passé devraient être valable. Cela devrait généralement se baser les rapports à partir des mesures sur les équipements existants (raccordement, appels, trafic, etc.) en plus des informations générales de fond.
- Utilisation systématique de ces données dans les prévisions.
- Une supposition éduquée comme pour le développement futur. Cette estimation du développement futur peut être une extrapolation du développement passé, par fois ajusté en prenant en considération des informations de fond valables. Le prévisionniste a besoin des données historiques exactes pour améliorer ses prévisions.

Par conséquent, les bases des prévision est l'étude du passé. Plus qu'on maîtrise le passé et décrive mathématiquement le passé, plus qu'on a la chance de faire des prévisions correctes.

Le degré d'incertitude des données des prévisions devrait être donné, pour que les utilisateurs de ces données puissent prendre cette incertitude en considération.

1.4.2 *Surveillance des prévisions*

Il y a toujours une raison critique pour les applications des prévisions. Les questions suivantes doivent être posées:

- Les prévisions sont elles pertinentes (valables) ?
- Les prévisions sont elles exactes ?
- Est ce que le résultat est crédible ?

La pertinence des prévisions dépend dans quelques cas sur l'utilisation correcte des suppositions pour le futur développement.

L'exactitude des prévisions peut dépendre sur la précision des statistiques des données historiques et sur les méthodes d'extrapolation utilisées. Une prévision crédible ou non est souvent une matière du jugement personnel. Ici l'esprit humain peut avoir des limites quand il vient d'imaginer l'accroissement futur. Quelques prévisions ont été cependant réduites dans la taille tant que le prévisionniste ne peut pas envisager un tel accroissement. Le développement de quelques réseaux a été entravé à cause du manque d'imagination.

Quand on juge la crédibilité des prévisions, il doit être rappelé que:

- Des prévisions concernant le développement de la population, de l'économie et de l'industrie, etc., peuvent être erronées.
- Les statistiques des télécommunications disponibles pour les périodes passées peuvent contenir des erreurs et peuvent ne pas être collectés sous les même conditions.
- Les données de développement historique disponibles peuvent couvrir uniquement une période limitée.
- Les relations passées et présentes entre les variables étudiées ne sont pas nécessairement vraies dans les futur.

1.4.3 *Comment peut on commencer?*

Le processus de prévision peut être divisé en parties comme suit:

Définition du problème

Le but des suppositions devait être déterminé.

Collection des données de base

Plusieurs sources de données de base sont examinées en détail. La population et la croissance économique sont étudiées. Des résultats des prévisions récents sont des faits essentiels.

Choix de la méthode de prévision

La méthode devait être choisie selon les informations disponibles et l'exactitude demandée, etc.

L'analyse et l'établissement des prévisions

L'analyse consiste en la préparation méthodologique des données de base et l'évaluation des résultats obtenus.

Documentation

Les prévisions devraient être présentées dans un format facile et compréhensible. Le résultat doit contenir des prévisions alternatives. À côté des prévisions les plus probables, il doit y avoir une prévision optimiste et une pessimiste, ainsi le planificateur est bien informé sur la position de la plus haute et la plus basse limite.

2. METHODES DE PREVISIONS

2.1 Introduction

Presque toutes les méthodes de prévision supposent que le futur devrait en quelque sorte ressembler au passé. Cela peut être interprété de façons différentes. Par exemple:

1. Pour les méthodes de prévision tendancielle il est supposé que le développement devrait suivre la courbe, qui a été ajustée à des données historiques existantes.
2. Quand on utilise les relations explicites entre la demande et divers facteurs déterminés, cela devrait rester le même dans le futur.
3. Comparer les diverses étapes du développement des télécommunications, il est supposé que le pays le moins développé (ou la zone) devrait développer au niveau de un plus développé.
4. Quand on applique le jugement personnel (subjectif) dans les prévisions, le futur devrait ressembler aux connaissances précédentes de la personne et à l'expérience du développement passé.

Même si le futur ressemble souvent au passé dans ce sens, il n'est jamais une reproduction exacte. Cependant les prévisions ne devraient jamais être acceptées sans critique. Cela veut dire que les prévisions - comme produit par quelques méthodes - sont fréquemment modifiées avant d'être acceptées pour les buts de planification. Pendant que de tels amendements peuvent apparaître subjectifs, ils sont nécessaires quand il y a une raison de croire que le futur devrait connaître quelques divergences du passé, sans tenir compte des méthodes utilisées. Les exemples de telles situations sont les changements de tarifs, améliorations radicales au service, etc.

Les suppositions utilisées dans les prévisions, devraient être revues et ce, conformément aux changements possibles dans l'environnement dans lequel le système téléphonique opère.

L'importance des données exactes et détaillées ne peuvent pas être surestimées. La qualité des prévisions dépend directement de la qualité des données historiques utilisées. Si les données fiables ne sont pas disponibles, alors le prévisionniste devrait commencer d'abord par un processus de collecte de données avant d'appliquer les méthodes de prévision. Plusieurs méthodes discutées dans ce chapitre pourraient être difficiles, si non impossible à utiliser sans données historiques adéquates.

2.2 Analyse des demandes avec statistique

On peut supposer que le développement du nombre de lignes téléphoniques dans un pays suivent certains modèles. Par exemple, il faudrait qu'il soit dépendant du nombre d'habitants, vivant un développement standard, technique et économique etc. Si quelques variables sont logiquement liées au développement téléphonique, ces variables peuvent être utilisées pour expliquer le développement. Dans ce qui suit, quelques variables sont décrites:

Taille de la population - plus qu'il y ait la croissance en matière de population, plus que le nombre de clients augmente et la demande des lignes d'abonnés augmente.

Niveau de vie - une mesure du niveau de vie est le produit intérieur brut (GDP) par habitant. Dans l'économie centralisée planifiée, le produit Matériel Net par habitant peut être utilisé. Les deux valeurs peuvent être converties au niveau de prix. Quand le niveau de vie augmente, la demande en lignes d'abonnés augmente généralement.

Activités de construction - comme les mesures de changement des structures dans la société sont dû à l'urbanisation, niveau de vie élevé, développement de la population etc., on peut utiliser le nombre des nouvelles administrations et départements. L'activité de construction peut être utilisée pour estimer la demande future pour les lignes d'abonnés.

Prix - Prix élevé pour l'utilisation du téléphone (c'est à dire les honoraires d'admission, coût d'abonnement et les charges d'appels) peut être supposé réduire la demande en lignes d'abonnés.

Le choix des variables explicatives peut dépendre de la disponibilité de leurs prévisions fiables. La population et le développement économique sont les variables communes les plus explicatives pour le nombre de lignes d'abonnés dans le pays.

Figure 2.2/1 montre la densité téléphonique et le GDP par habitant pour la période 1900 - 1965 en Suède.

Sur l'échelle logarithmique, la relation est approximativement linéaire, durant la période 1900-1915 et 1920-1965.

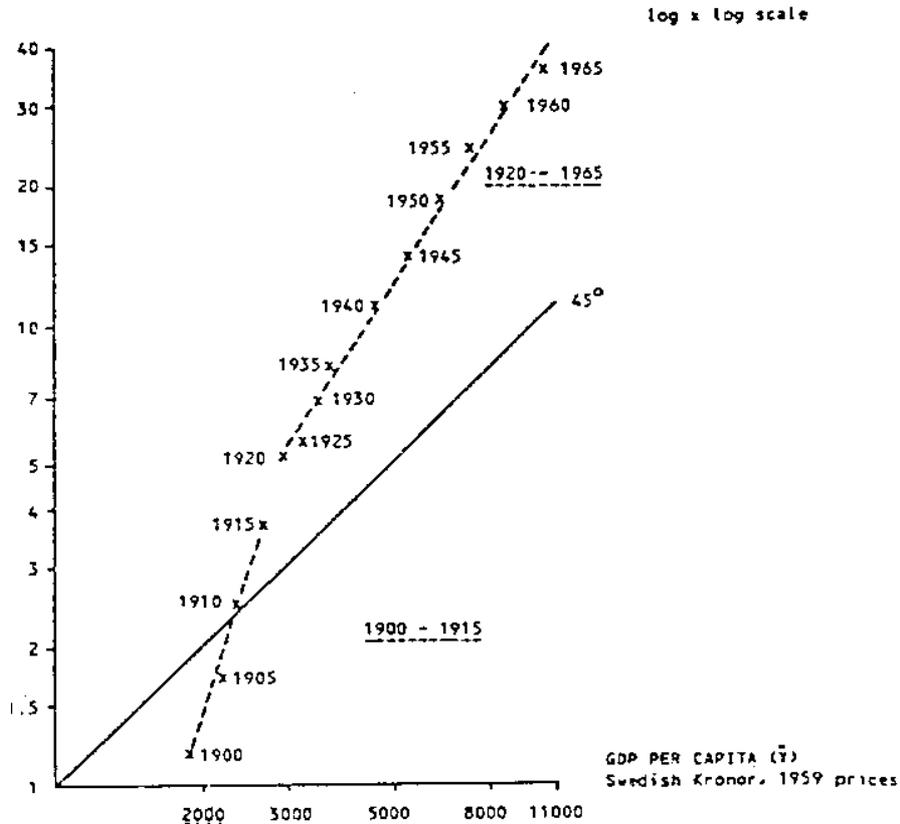


Figure 2.2/1 : Densité téléphonique et le GDP par habitant pour la période 1900-1965 en Suède

L'analyse de la demande statistique par *le modèle économétrique* est la technique tracée pour obtenir une relation explicite quantitative entre l'intérêt et d'autres variables qui peuvent influencer ce dernier.

Le modèle économétrique de demande est une représentation mathématique des comportement des clients, et il dépend de plusieurs facteurs qui pour lesquels les données peuvent être obtenues. Quelques facteurs, même s'ils ont une relation claire avec les variables de prévision, sont elles même difficile à prévoir, et d'autres facteurs peuvent être interdépendants, alors leur influence précise n'est pas claire.

Il y a deux classes principales de modèles : les modèles de *régression simple* et les modèles de *régression multiple*. Le premier inclut modèles dans lesquels la variables dépendante est supposée être une fonction linéaire de plusieurs variables explicatives indépendantes.

La seconde classe de modèles, le modèle de régression multiple, qui nous donne la possibilité d'estimer simultanément l'inter-relation entre l'ensemble des variables. La faiblesse du modèle multi-équations est que les variables explicatives peuvent être indépendantes. Le retour ("feed-back") possible entre les variables dépendantes et les variables explicatives peut être négligé.

2.2.1 Régression à 2 variables

Supposons qu'on peut étudier la relation entre les variables, x et y . Dans le but de décrire cette relation statistiquement, on a besoin d'un nombre d'observations des paires de valeurs, une de chaque variable:

$$(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$$

Si on trace ces quantités, on peut trouver qu'elles montrent un modèle consistant, et il est cependant possible de dessiner la courbe qui représente ces points. Deux problème donc se posent:

- a) décider quel type de courbe utiliser;
- b) choisir la “meilleure” courbe de ce type.

Le premier problème est toujours résolu par un regard sur les représentation graphique obtenues à partir des paires de valeurs observées (x_i, y_i) et choisir sur vue le type simple de courbe. Les lignes droites sont les plus populaires dans cette discipline.

Le second problème, choisir les valeurs optimales des paramètres des courbes, peut être résolu seulement après la prise de décision sur les critères retenus qui nous informent sur la qualité de la courbe. Le critère des moindres carrés est le plus utilisé. Ceci est pour se appel intuitive et ses propriétés mathématiques. On l'utilisera également ici. Selon le critère des moindre carrés la courbe est bonne si S , la somme des carrés des distances entre les points données et la courbe, est petite. La meilleure courbe est celle pour la quelle S est petite.

Le second problème peut maintenant être ré-formulé comme : *déterminer les paramètres des valeurs qui minimisent S .*

Supposons que nous voulons déterminer la ligne droite qui ajuste les observations. Cela signifie que nous supposons une relation linéaire entre x , variable indépendante, et y , variable dépendante. Le fait que les observations (x_i, y_i) ne sont pas alignées sur la ligne, est attribué à l'influence d'un terme d'erreur.

En bref, on suppose:

$$y = \alpha + \beta \cdot x + \varepsilon \tag{2.2.1}$$

Les paramètres α et β ne sont pas connus. L'erreur ε est aléatoire mais a une valeur espérée égale à zéro.

Laissons a et b être les estimés respectifs de α et β . La valeur de calcul pour y pour une valeur donnée de x_i utilisant ses estimations:

$$\hat{y}_i = a + b \cdot x_i \tag{2.2.2}$$

Cependant, on observe les paires (x_i, y_i) , alors $y_i - \hat{y}_i$ est la déviation de ce point particulier à partir de la ligne. Ainsi, on a, à chercher les a et b estimés qui minimise:

$$S = \sum_{i=1}^n (y_i - \hat{y}_i)^2$$

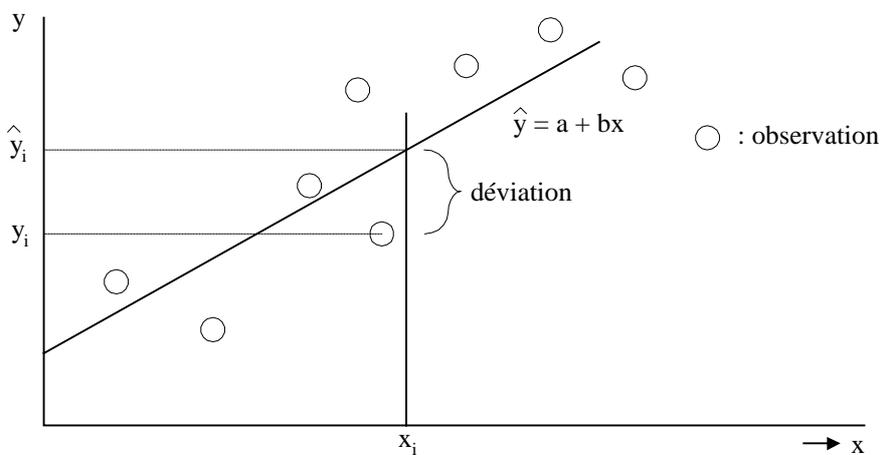


Figure 2.2/1 : Les observations et la ligne

Notons qu'on est intéressé seulement aux déviations verticales. Cela ne nous surprend pas comme le but ultime est de prédire les valeurs de y quand une nouvelle valeur x est donnée. En générale, on a choix de trois possibilités:

- V: Distances verticales
- H: Distances horizontales
- P: Distances perpendiculaires à la lignes

La figure qui suit donne les résultats obtenus, le commencement se fait à partir des mêmes observations, par ces trois cas.

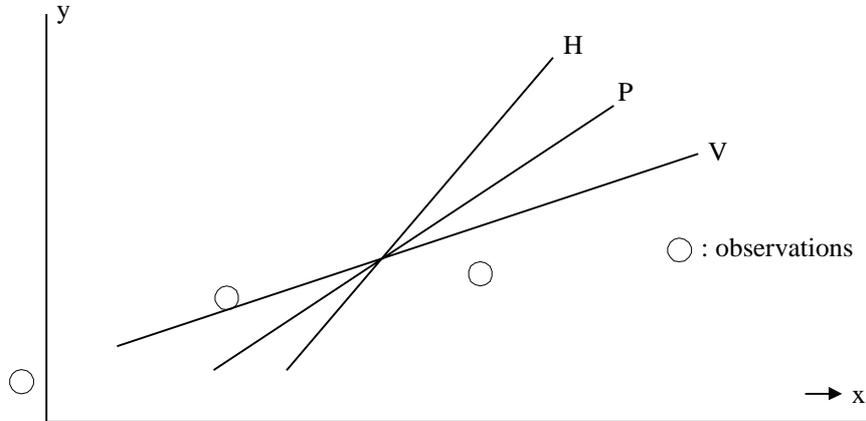


Figure 2.2/2 : Influence de la “distance” choisie

Les estimés a et b sont obtenus par la résolution de:

$$\frac{\partial S}{\partial a} = 0 \quad \text{et} \quad \frac{\partial S}{\partial b} = 0$$

où

$$S = \sum_{i=1}^n (y_i - a - b \cdot x_i)^2$$

Le résultat est:

$$b = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \tag{2.2.3}$$

$$a = y - b \cdot x \tag{2.2.4}$$

Pour juger la qualité produite de la ligne, on peut calculer le coefficient de corrélation r .

$$r^2 = \frac{\sum (\bar{y}_i - \bar{y})^2}{\sum (y_i - \bar{y})^2} = \frac{\text{Variance expliquée}}{\text{Variance totale}} \tag{2.2.5}$$

r a le même signe que b et appartient à $[-1, +1]$. $r = 0$ indique le manque de corrélation, en d’autres termes : la ligne droit ne peut pas représenter la relation entre y et x .

Si r dévie de zéro, il y a difficulté de juger les meilleurs pas entre les observations et la ligne droite. Autres méthodes, cependant, devaient être appliquées.

Si on est satisfait de la valeur obtenue de r , on peut utiliser

$$y = a + b \cdot x$$

pour donner le point estimé de y , quelle que soit la valeur de x .

Il y a, cependant, d'autres méthodes pour vérifier que la ligne de régression donne une description fiable sur la façon de dépendance de y sur x . La première étape est d'estimer la variance de b et vérifier qu'elle est à l'intérieur avec des limites raisonnables. La deuxième étape est de vérifier contre les erreurs systématique avec le teste de Durban-Watson.

Le premier teste demande le calcul de la quantité

$$T = \frac{b \left(\sum (x - \bar{x})^2 \right)^{1/2}}{s} \quad (2.2.6)$$

où

$$s^2 = \frac{\sum y^2 - a \sum y - b \sum xy}{n - 2} \quad (2.2.7)$$

Pour une vérification exacte, il est nécessaire d'utiliser les tables statistiques pour la distribution de t , mais T devrait être inférieur à -2 ou supérieur à $+2$ et ce pour faire des analyses de régression acceptables.

Pour $-2 < T < 2$, la valeur de b ne peut être différente de zéro et on est pas sûre que notre ligne droite avec $b \neq 0$ représente réellement un changement de y avec x .

Si on veut dessiner les intervalles de confiance pour les paramètres α et β , ou pour nos valeurs y précitées, on doit faire les suppositions suivantes:

Les observations (x_i, y_i) à partir d'un échantillon aléatoire de la population distribuée Normale - bivariée:

$$\begin{aligned} E\{Y|X = x\} &= \alpha + \beta \cdot x \\ V\{Y|x\} &= \sigma^2 \end{aligned} \quad (2.2.8)$$

A partir d'aujourd'hui et sans perte de généralités, on devrait aussi supposer que $\bar{x} = \bar{y} = 0$; cela signifie que nous choisissons une nouvelle origine, alors que, avant de commencer les calculs on devrait soustraire \bar{x} de toutes les valeurs x_i et \bar{y} de toutes les valeurs y_i .

L'estimé de β peut maintenant être écrit comme

$$b = \frac{\sum x_i \cdot y_i}{\sum x_i^2} \quad (2.2.3a)$$

et

$$a = 0 \quad (2.2.4a)$$

La variance σ^2 est inconnue, mais on peut l'estimer par

$$s_e^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \bar{y})^2 \quad (2.2.8)$$

s_e est appelé l'erreur standard de l'estimé. s_e^2 est un estimateur impartial de σ^2 .

En plus,

$$\frac{(n-2)s_e^2}{\sigma^2} = \chi^2(n-2) \quad (\text{Chi -carré distribué avec } n-2 \text{ degré de liberté})$$

$$a \approx N\left(\alpha, \frac{\sigma^2}{n}\right) \quad \text{et} \quad b \approx N\left(\beta, \frac{\sigma^2}{\sum x_i^2}\right), \quad \text{ainsi} \quad \frac{a-\alpha}{\sigma} \sqrt{n} \quad \text{et} \quad \frac{b-\beta}{\sigma} \sqrt{\sum x_i^2}$$

sont normalement distribué, $N(0,1)$, (avec moyenne = 0 et déviation standard égale à 1).

Il peut être montré que a , b et s_e sont indépendant de tout autre, ainsi

$$\frac{a-\alpha}{\sigma} \cdot n \cdot \frac{\sigma^2}{s_e^2} \quad \text{et} \quad \frac{b-\beta}{\sigma} \cdot \sqrt{\sum x_i^2} \cdot \sqrt{\frac{\sigma^2}{s_e^2}}$$

sont t -distribué avec $n-2$ degré de liberté. Les intervalles de confiance peuvent être calculés et tester la performance des hypothèses. Pour concevoir l'intervalle de confiance pour la valeur y quand $x = x_o$, on doit regarder à la variance de $y_o = bx_o$. Dû à notre nouvelle origine, qui est différente de chaque échantillon, la variance de y_o inclue la variance de l'origine, $V(a)$, et est égale à:

$$V(y_o) = V(a) + x_o^2 \cdot V(b) = \sigma^2 \left(\frac{1}{n} + \frac{x_o^2}{\sum x_i^2} \right) \quad (2.2.9)$$

Ancore on arrive à $t(n-2)$ - distribution par l'élimination de l'inconnu σ^2 parce que

$$\frac{y_o - \alpha - \beta \cdot x_o}{\sigma \cdot \sqrt{\frac{1}{n} + \frac{x_o^2}{\sum x_i^2}}} \approx N(0,1)$$

et alors

$$\frac{y - \alpha - \beta \cdot x_o}{s_e \cdot \sqrt{\frac{1}{n} + \frac{x_o^2}{\sum x_i^2}}} \approx t(n-2)$$

Ci-dessus on s'est limité à une relation linéaire entre x et y . La même méthode, cependant, peut être appliquée dans certains cas où la relation non linéaire semble plus appropriée.

Pour cela, il est suffisant que la relation peut être menée dans une forme linéaire par une transformation appropriée des données.

Par exemple, si on suppose que la relation peut être exprimée par

$$y = \alpha \cdot e^{\beta \cdot x} \quad (2.2.10)$$

alors $\ln y = \ln \alpha + \beta \cdot x \quad (2.2.10a)$

ou $z = \gamma + \beta \cdot x \quad (2.2.10b)$

On determine les estimés c et b pour respectivement γ , et β , dans le cas usuel et prévoir avec

$$y = e^z = e^{c+\beta x} = e^c \cdot e^{\beta \cdot x} \quad (2.2.10c)$$

Si la transformation des données à quelques choses comme une relation linéaire qui est impossible, la courbe de répression linéaire peut être essayé.

L'idée fondamentale est que la relation actuelle entre x et y peut être représentée par

$$y = \beta_0 + \beta_1 \cdot x + \beta_2 \cdot x^2 + K + \beta_k \cdot x^k \quad (2.2.11)$$

Nos observations (x_i, y_i) sont le résultat de fluctuations aléatoires ϵ_i autour de cette ligne.

Appliquons le critère des moindres - carrés par l'estimation de $\beta_0, \beta_1, K, \beta_k$ signifie encore que $\sum \epsilon_i$ devrait être minimisé.

Alors, on a essayer de trouver le minimum de

$$\sum_{i=1}^n \left(y_i - (b_0 + b_1 \cdot x_i + b_2 \cdot x_i^2 + K + b_k \cdot x_i^k) \right)^2$$

où b_i devra estimer β_i .

Les dérivées partielles par rapport à b_0, b_1, K, b_k , et égaliser toutes les dérivées à zéro, donne le système d'équations linéaires:

$$\sum y_i = nb_0 + b_1 \sum x_i + b_2 \sum x_i^2 + K + b_k \sum x_i^k$$

$$\sum x_i y_i = b_0 \sum x_i + b_1 \sum x_i^2 + b_2 \sum x_i^3 + K + b_k \sum x_i^{k+1}$$

$$\sum x_i^k y_i = b_0 \sum x_i^k + b_1 \sum x_i^{k+1} + K + b_k \sum x_i^{2k}$$

La solution de ce système d'équations linéaires avec $k+1$ inconnus donne les estimés demandés. Il est conseillé de limiter le nombre de paramètres b_k utilisés.

2.2.2 Régression Multiple

Si on est pas satisfait avec la valeur de r , qui signifie que la variance résiduelle (= variance totale - variance explicative) est très large, on peut voir les résidus. La variance résiduelle est

$$(y_i - \bar{y})^2$$

et peut avoir une valeur supérieure seulement si au moins quelques résidus $y - \bar{y}$ sont large, cela signifie que les distances verticales entre les observations et la ligne $y = a + bx$ sont large.

Si, cependant, la variable étudiée y dépend probablement sur d'autres variables que x , on peut essayer de réduire les résiduels par l'introduction d'un ou plusieurs facteurs "explicatifs", x_1, x_2 , etc. Il y a deux méthode principales à appliquer, la *méthode itérative* et la *méthode directe*, qui devrait être expliquée par des exemples numériques.

Méthode itérative

Exemple : prévision du nombre d'appels locaux.

La méthode montrée au dessous est appelée la méthode itérative, parce qu'elle travaille par étape. Elle est préférable, quand les facteurs explicatifs à utiliser ne sont pas décidées à l'avance.

Il est à noter que même si la méthode entraîne un certain degré d'approximations, les résultats numériques devraient être suffisamment exacte pour des cas pratiques, comme celui donné ici.

Etape 1 Sélectionner un couple de facteurs explicatifs possibles, x_1, \dots, x_n pour les quels de bonnes prévisions sont disponibles (nombre d'abonnés, habitants, etc.).

Etape 2 Calculer les coefficients de corrélations entre la variable dépendante y (dans l'exemple ci-dessous le nombre d'appels locaux) et chacun des facteurs explicatifs x_1, \dots, x_n .

Etape 3 Sélectionner le facteur explicatif qui a le plus haut coefficient de corrélation avec y .

Etape 4 Si ce facteur est présumé être x_2 , faire votre première tentative avec la fonction de régression

$$y_1 = a_1 + b_1 \cdot x_2 \quad (2.2.12)$$

et calculer le paramètre a et b selon la méthode de la section 2.2.1.

Etape 5 Calculer les résiduels

$$r_1 = y_1 - (a_1 + b_1 \cdot x_2) \quad (2.2.13)$$

Etape 6 Calculer les coefficients de corrélation entre r_1 et chaque facteur explicatif restant, dans ce cas x_1, x_3, \dots, x_m

Etape 7 Sélectionner le facteur explicatif restant qui a le plus fort coefficient avec r_1 .

Etape 8 Si ce facteur est appelé x_1 , faire une tentative No. 2 par l'utilisation de la fonction de régression

$$r_1 = a_2 + b_2 \cdot x_1 \quad (2.2.14)$$

et calculer les paramètres a_2 et b_2 , selon la méthode dans la section 2.2.1.

Etape 9 Les deux fonctions à partir de la tentative 1, et la tentative 2 donnent

$$\begin{aligned} y_1 &= (a_1 + b_1 \cdot x_2) + r_1 = \\ &= (a_1 + b_1 \cdot x_2) + (a_2 + b_2 \cdot x_1) = \\ &= (a_1 + a_2) + b_2 \cdot x_1 + b_1 \cdot x_2 \end{aligned} \quad (2.2.15)$$

qui est une amélioration, comparé avec la fonction de régression à partir de la 1ère tentative.

Etape 10 Calculer les résiduels

$$r_2 = y_1 - [(a_1 + a_2) + b_2 \cdot x_1 + b_1 \cdot x_2] \quad (2.2.16)$$

Etape 11 Calculer le coefficient de corrélation entre r_2 et chacun des facteurs explicatifs restants, dans ce cas x_3, \dots, x_n

Etape 12 Aller avec la tentative 3 et continuer de la même façon jusqu'à ce que les résiduels calculés ne sont pas assez petits.

Le pas parfait ne voudrait pas dire que tous les résidus sont nuls. Cet étape devrait seulement être parvenu quand le nombre de paramètre est le même que le nombre d'observations dans le matériel statistique.

Calculs Numériques

Année de financement	Appels locaux y_1	Appels interurbains		
		Service manuel	Service automatique	Total y_2
1958/59	2305	110	125	235
1959/60	2423	103	165	268
1960/61	2605	95	220	315
1961/62	2852	84	267	351
1962/63	3040	78	322	400
1963/64	3376	73	372	445
1964/65	3642	55	445	500
1965/66	3700	38	530	568
1966/67	3884	21	609	630
1967/68	4059	13	650	663

Table 2.2/1 Appels téléphoniques (en millions)

Etape 1: Sélectionner les facteurs explicatifs possibles

Année 1er Jan.	Population x_1	Abonnés téléphoniques			Appareils téléphoniques		
		Central manuel	Central automatique	Total x_2	Central manuel	Central automatique	Total x_3
1959	7 436	352	1 712	2 064	408	2 118	2 526
1960	7 471	311	1 839	2 150	359	2 278	2 637
1961	7 499	271	1 973	2 244	313	2 448	2 761
1962	7 542	207	2 143	2 350	236	2 668	2 904
1963	7 581	174	2 294	2 468	198	2 856	3 054
1964	7 627	144	2 454	2 598	163	3 060	3 223
1965	7 695	101	2 622	2 723	114	3 273	3 387
1966	7 773	75	2 785	2 860	86	3 487	3 573
1967	7 844	41	2 956	2 997	46	3 711	3 757
1968	7 894	27	3 095	3 122	32	3 903	3 935

Table 2.2/2 Facteurs explicatifs (en milliers)

1:st Jan.	Population x_1	Total d'abonnés x_2	Appareils Total x_3
1969	7 964	3 214	4 111
1970	8 042	3 335	4 288
1971	8 120	3 453	4 458
1972	8 201	3 566	4 626
1973	8 281	3 672	4 793
1974	8 360	3 776	4 957
1975	8 430	3 879	5 120

Table 2.2/3 Préviation disponibles des facteurs explicatifs (en milliers)

Etape 2: Calculate the correlation coefficients. Note that the basic figures given in Tables 2.2/1-2 are rounded off in the following calculations.

Calcul des coefficients de corrélation entre le nombre d'appels locaux y et les facteurs explicatifs x_1 , x_2 et x_3 donne les résultats montrés. Les données sont prises des tableaux 2.2/1-2.

y_1 10^9	x_1 10^6	x_2 10^6	x_3 10^6
2.31	7.44	2.06	2.53
2.42	7.47	2.15	2.64
2.60	7.50	2.24	2.76
2.85	7.54	2.35	2.90
3.04	7.58	2.47	3.05
3.38	7.63	2.60	3.22
3.64	7.70	2.72	3.39
3.70	7.77	2.86	3.57
3.88	7.84	3.00	3.76
4.06	7.89	3.12	3.94
31.88	76.36	25.57	31.76

Table 2.2/4 Facteurs explicatifs données

Les sommes suivantes sont calculées:

$$\begin{array}{llll}
 \sum y_i = 31.88 & \sum x_1 = 76.36 & \sum x_2 = 25.57 & \sum x_3 = 31.76 \\
 \sum y_i^2 = 105.22 & \sum x_1^2 = 583.31 & \sum x_2^2 = 66.58 & \sum x_3^2 = 102.97 \\
 & \sum x_1 y_1 = 244.32 & \sum x_2 y_1 = 83.57 & \sum x_3 y_1 = 103.96
 \end{array}$$

Le coefficient de corrélation est calculé à partir de:

$$R(y_1 x_i) = \frac{D_i}{\sqrt{G_i \cdot H_i}}$$

$$D_i = n \sum x_i y_1 - \sum x_i \sum y_1$$

où $G_i = n \sum y_1^2 - (\sum y_1)^2$

$$H_i = n \sum x_i^2 - (\sum x_i)^2$$

ce qui donne

$$R(x_1y_1) = 0.979$$

$$R(x_2y_1) = 0.991$$

$$R(x_3y_1) = 0.989$$

Etape 3: Sélectionner le facteur explicatif qui devrait être utilisé en premier.

Dans cette première tentative on choisit le facteur explicatif avec le plus haut coefficient de corrélation avec y_1 . Ici c'est le nombre d'abonnés = x_2 .

Etape 4: Faire la tentative N° 1 avec la fonction de régression

$$y_1 = a_1 + b_1x_2$$

La formule de calcul de a et b dans la section 2.2 peut être écrite comme suite:

$$a_1n + b_1 \sum x_2 = \sum y_1$$

$$a_1 \sum x_2 + b_1 \sum x_2^2 = \sum x_2y_1$$

ou avec les valeurs numériques:

$$10.00 a_1 + 25.57 b_1 = 31.88$$

$$25.57 a_1 + 66.58 b_1 = 83.57$$

La solution est:

$$a_1 = - 1.196$$

$$b_1 = 1.715$$

Etape 5: Calculer le résiduel r_1 .

La fonction qui arrive à l'étape 4 (tentative 1) donne les valeurs essayées suivantes de y_1 :

Table 2.2/5 Valeurs ajustées de y à partir de l'étape 4

Année de financement	x_2	Valeurs ajustées - 1.13 + 1.677 x = y_1	Comparaison avec les valeurs données de y	
			Table 2.2/1 y	résiduel r_1
1958/59	2.06	2.34	2.31	- 0.03
1959/60	2.15	2.49	2.42	- 0.07
1960/61	2.24	2.64	2.60	- 0.04
1961/62	2.35	2.83	2.85	+ 0.02
1962/63	2.47	3.04	3.04	+ 0.00
1963/64	2.60	3.26	3.38	+ 0.12
1964/65	2.72	3.47	3.64	+ 0.17
1965/66	2.86	3.71	3.70	+ 0.01
1966/67	3.00	3.95	3.88	- 0.07
1967/68	3.12	4.15	4.06	- 0.09

Etape 6: Calculer le coefficient de corrélation avec r_1 .

Avant notre deuxième tentative, on calcule le coefficient de corrélation entre les résiduels de la tentative 1 et les facteurs explicatifs restants x_1 et x_3 .

Table 2.2/6 Calcul des coefficients de corrélation.

Résiduels à partir de la tentative N° 1 r_1	x_1	x_3	Sommes calculées	
	à partir de la Table 2.2/2		Variable	Valeur
- 0.03	7.44	2.53	$\sum r_1^2$	= 0.0642
- 0.07	7.47	2.64	$\sum x_1^2$	= 583.31
- 0.04	7.50	2.76		
- 0.02	7.54	2.90	$\sum x_3^2$	= 102.97
- 0.00	7.58	3.05		
+ 0.12	7.63	3.22	$\sum r_1 x_1$	= - 0.0073
+ 0.17	7.70	3.39		
- 0.01	7.77	3.57	$\sum r_1 x_3$	= - 0.1199
- 0.07	7.84	3.76		
- 0.09	7.89	3.94		
Total	76.36	31.76		

Les coefficient de corrélation sont:

$$R(r_1, x_1) = - 0.029$$

$$R(r_1, x_3) = - 0.019$$

Cela indique une très faible corrélation.

Etape 7: Sélectionner le facteur explicatif suivant.

Le plus fort coefficient de corrélation est celui entre r_1 et x_1 , alors notre seconde tentative devrait utilisé le facteur explicatif x_1 .

Etape 8: Faire la tentative 2 avec la fonction

$$r_1 = a_2 + b_2 x_1$$

Les paramètres a_1 et b_2 sont donnés à partir des équations suivantes:

$$a_2 n + b_2 \sum x_1 = \sum r_1$$

$$a_2 \sum x_1 + b_2 \sum x_1^2 = \sum r_1 x_1$$

ou $10.00 a_2 + 76.36 b_2 = - 0.04$

$$76.36 a_2 + 583.31 b_2 = - 0.3089$$

La solution est alors:

$$a_2 = 0.18 \quad 0.113$$

$$b_2 = - 0.023 \quad - 0.0154 \quad (\text{cor} - 0.028822)$$

Par conséquent $r_1 = 0.113 - 0.0154 x_1$

Etape 9: Les résultats de tentatives 1 et 2 ci-dessus devaient maintenant donner la fonction de régression pour les appels locaux:

$$y_1 = \begin{matrix} (-1.196 + 1.715 x_2) \\ \uparrow \qquad \qquad \uparrow \\ \text{(à partir de tentative 1)} \\ \downarrow \qquad \qquad \downarrow \\ = (-1.196 + 1.715 x_2) \end{matrix} + \begin{matrix} r_1 \\ \\ \text{(à partir de tentative 2)} \\ \\ = (0.113 - 0.0154 x_1) \end{matrix} =$$

par conséquent

$$y_1 = -1.083 - 0.0154 x_1 + 1.715 x_2$$

Ce résultat est comparé avec les valeurs données de y_1 dans le tableau 2.2/1.

Etape 10: Calculer les résiduels r_2

Année financière	Valeurs ajustées pour les appels locaux - 1.083 - 0.0154 x_1 + 1.715 x_2 = y_1	Valeurs données de y	
		Tableau 2.2/1 x_2	Résiduel r_2
1958/59	2.34	2.31	- 0.03
1959/60	2.49	2.42	- 0.07
1960/61	2.64	2.60	- 0.04
1961/62	2.83	2.85	+ 0.02
1962/63	3.04	3.04	0.00
1963/64	3.26	3.38	+ 0.12
1964/65	3.46	3.64	+ 0.18
1965/66	3.70	3.70	0.00
1966/67	3.94	3.88	- 0.06
1967/68	4.15	4.06	- 0.09

A partir de cette fonction de régression, $y(x_1, x_2)$, on trouve le nombre futur suivant d'appels locaux (en millions):

	x_1	x_2	$y(x_1, x_2)$	$y(x_2)$
1968/69	7.96	3.21	4300	4310
1969/70	8.04	3.34	4520	4530
1970/71	8.12	3.45	4710	4720
1971/72	8.20	3.57	4910	4930
1972/73	8.28	3.67	5080	5100
1973/74	8.36	3.78	5270	5290
1974/75	8.43	3.88	5440	5460

Le résultat est donnée dans la figure 2.2/3

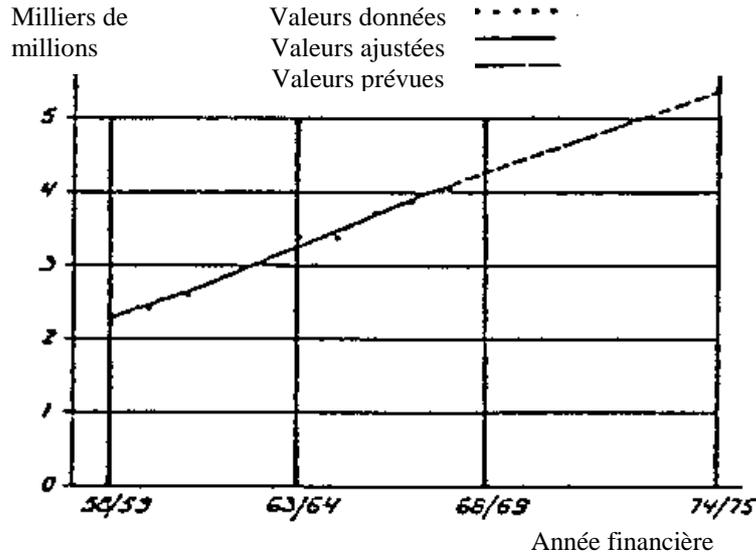


Figure 2.2/3

A partir des résiduels r_2 il arrive que la tentative 2 donne particulièrement aucune différence composée avec l'alternative 1, $y(x_2)$ ci-dessus. Cela suivi déjà à partir plutôt du petit coefficient de corrélation $R(r_1, x_1)$.

On devrait cependant, stopper à l'étape 8 et utiliser la fonction de régression.

$$y_1 = -1.083 - 0.0154 x_1 + 1.715 x_2$$

Méthode directe

Pour cette méthode on suppose à partir du commencement que la variable à prévoir dépend du nombre de variables explicatives.

Exemple: *Prévision du nombre d'appels interurbains.*

Le nombre d'appels interurbains, y_2 , augmente généralement avec l'introduction du service automatique (STD). Il est fréquemment observé que l'introduction du STD entre deux régions génère le double du nombres d'appels.

Cependant, il semble approprié d'utiliser le facteur explicatif "degré d'automatisation" (dénnoté par x_4), pour prévoir le nombre des appels interurbains. Ce degré est défini comme la proportion entre le nombre d'appels interurbains et tous les appels longue distance.

Il semble aussi probable que nombre d'appareils téléphoniques (dénnotés par x_3) pourrait être un facteur explicatifs utilisable. Après ces considérations, on sélectionne à priori les deux facteurs explicatifs:

- nombre d'appareils téléphoniques = x_3
- degré d'automatisation = x_4

Il est présumé que nous n'avons pas besoins de facteurs explicatifs supplémentaires, alors on peut appliquer "la méthode directe" pour le calcul des paramètres dans la fonction de régression:

$$y_2 = a + b_3 \cdot x_3 + b_4 \cdot x_4 \tag{2.2.17}$$

Année financière	Appels interurbains en milliers de millions y_2	Appareils téléphoniques en millions x_3	Degré d'automatisation x_4
1958/59	0.235	2.53	0.53
1959/60	0.268	2.64	0.62
1960/61	0.315	2.76	0.70
1961/62	0.351	2.90	0.76
1962/63	0.400	3.05	0.81
1963/64	0.445	3.22	0.84
1964/65	0.500	3.39	0.89
1965/66	0.568	3.57	0.93
1966/67	0.630	3.76	0.97
1967/68	0.663	3.94	0.98
Total	4.375	31.76	8.03

Tableau 2.2./8 Appels interurbains et facteurs explicatifs

Année financière		Appareils téléphoniques en millions x_3	Degré d'automatisation x_4
1968/69		4.11	0.99
1969/70		4.29	0.99
1970/71		4.46	1.00
1971/72		4.63	1.00
1972/73		4.79	1.00
1973/74		4.96	1.00
1974/75		5.12	1.00

Table 2.2/9 Prévisions donnés des facteurs explicatifs:

A partir des tableaux, pour la période 1958 - 68 on calcule les totaux suivant:

$$\begin{aligned} \sum x_3^2 &= 102.97 & \sum x_3 x_4 &= 26.13 \\ \sum x_4^2 &= 6.65 & \sum y_2 x_3 &= 14.55 \\ & & \sum y_2 x_4 &= 3.71 \end{aligned}$$

Les paramètres a , b_3 et b_4 sont estimés à partir du système d'équations:

$$\begin{aligned} a n + b_3 \sum x_3 + b_4 \sum x_4 &= \sum y_2 \\ a \sum x_3 + b_3 \sum x_3^2 + b_4 \sum x_3 x_4 &= \sum y_2 x_3 \\ a \sum x_4 + b_3 \sum x_3 x_4 + b_4 \sum x_4^2 &= \sum y_2 x_4 \end{aligned}$$

ou

$$\begin{aligned} 10.00 a + 31.76 b_3 + 8.03 b_4 &= 4.38 \\ 31.76 a + 102.97 b_3 + 26.13 b_4 &= 14.55 \\ 8.03 a + 26.13 b_3 + 6.65 b_4 &= 3.71 \end{aligned}$$

La solution du système d'équations est:

$$a = -0.481$$

$$b_3 = 0.222$$

$$b_4 = 0.267$$

Par conséquent, la fonction de régression devrait être:

$$y_2 = -0.481 + 0.222 x_3 + 0.267 x_4$$

Le résultat est compté avec les valeurs de y_2 données dans le tableau suivant:

Année financière	$y_2 =$ $-0.481 + 0.222 x_3$ $+ 0.267 x_4$	Comparaison avec les valeurs données de y_2	
		A partir du tableau 2.2/8	Résiduel r
1958/59	0.223	0.235	+0.012
1959/60	0.271	0.268	-0.003
1960/61	0.319	0.315	-0.004
1961/62	0.366	0.351	-0.015
1962/63	0.412	0.400	-0.012
1963/64	0.458	0.445	-0.013
1964/65	0.510	0.500	-0.010
1965/66	0.560	0.568	+0.008
1966/67	0.613	0.630	+0.017
1967/68	0.656	0.663	+0.007

Tableau 2.2/10 Valeurs ajustées des appels interurbains (y_2)

Extrapolant ces chiffres dans le futur, on trouve les prévisions suivantes pour le nombre d'appels interurbains en milliers de millions:

1968/69	0.695
1969/70	0.735
1970/71	0.776
1971/72	0.814
1972/73	0.887
1974/75	0.923

Le résultat est donné à la Figure 2.2/4.

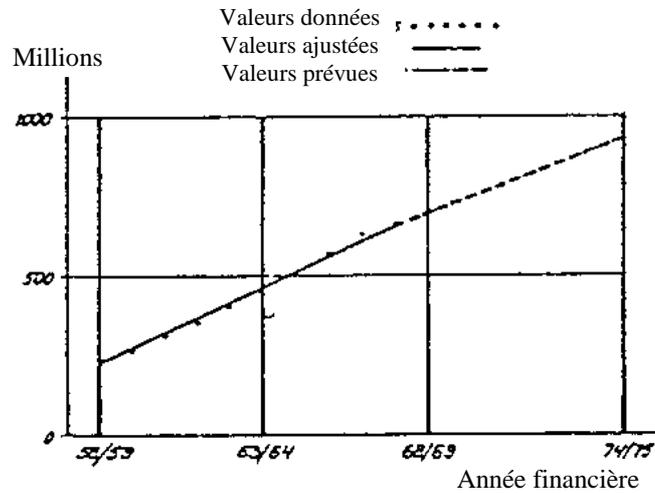


Figure 2.2/4

2.3 Méthodes de tendance

Une série chronologique est un ensemble d'observations d'une quantité comme une fonction de temps. Les observations peuvent suivre un certain comportement et le prévisionniste doit trouver comment les quantités observées seront développées dans le futur, c à d estimer la tendance.

Cette section traitera quelques courbes de développement, telles que la tendance linéaire, la tendance exponentielle et les tendances relatives au niveau de saturation. La section sur les séries chronologiques traite avec les méthodes expliquant une variation observée et un comportement croissant comme un résultat de quatre facteurs: tendance, variations cycliques, variations saisonnières et variations aléatoires.

Si des données historiques correctes sont disponibles, les méthodes de tendances peuvent être utilisées pour prévoir le développement futur sur une supposition importante: la même dépendance entre les variables devrait exister dans le futur. C'est la même supposition comme il est fait pour toutes les prévisions basées sur l'extrapolation des courbes décrivant le passé.

2.3.1 Analyse des série chronologiques`

L'analyse de séries chronologiques est le nom d'un nombre de méthodes demandés pour trouver une explication pour les fluctuations dans valeurs étudiées (y_t).

Toujours, les valeurs de y sont crues être le résultat de quatre facteurs:

- T. La tendance T est le processus fondamental de croissance (cette tendance peut être évidemment négative ou zéro).
- C. Les variations cycliques C sont des quantités sinusoïdales avec une fréquence très faible. Les cycles d'affaires, sept années de grasse suivies par les sept années de maigrisse, sont des exemple de ça.
- S. Les variations saisonnières S oscillent également, mais avec une fréquence élevée, et sont généralement plus prononcées que les variations cycliques. Les modèles journaliers, hebdomadaires ou annuels sont toujours incomprise dans S .
- I. Les mouvements irréguliers ou aléatoires, (I) sont dus aux événements de chance (ou de hasard) tels que les grevés, tremblements de terre, changements dans les tarifs, etc. et peuvent être vus comme résiduels, inexpliqués par T , C et S .

Dans la fig.2.3/1, les effets de T , C et S sont montrés.

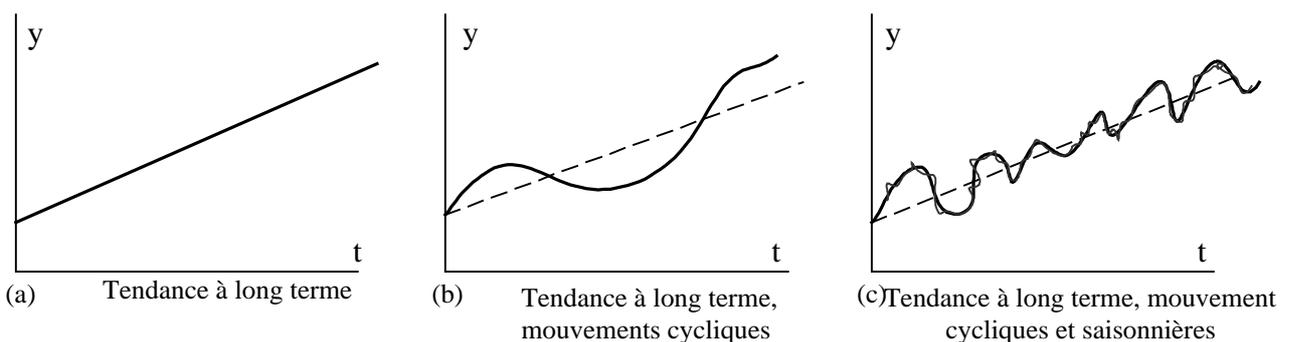


Figure 2.3/1

Influence des variations cycliques et saisonnières sur la tendance, dans une série chronologique idéale

Les trois premier facteurs sont étudiés séparément. Facteur I ne peut pas être exprimé comme une fonction du temps et agir comme une variations résiduelle.

On peut exprimer d'effet combiné des quatre facteurs comme

$$Y = T + S + C + I \quad (2.3.1)$$

ou par multiplication, exprimée par la formule

$$Y = T \times S \times C \times I \tag{2.3.2}$$

La combinaison de l'addition et la multiplication est aussi possible, c.a. dire

$$Y = T + S \times C \times I \tag{2.3.3}$$

Elle est toujours suffisent pour les prévisions à long terme de assembler les données dans les chiffres annuels et de tenter de tracer une ligne de tendance à travers ces chiffres.

Une variété des courbes mathématiques peut être tracée pour l'analyse des données des séries chronologiques, rangées à partir de l'extrapolation à main libre du graphe des séries, tracer une simple ligne droite, tracer une ligne droite pour des données des séries transformer, l'utilisation des courbes polynomiales, et les méthodes plus compliquées.

Tendance linéaire

N'importe quelle ligne droite donnant la relation entre 2 variables peut être exprimée comme:

$$y = a + bt \tag{2.3.4}$$

Quand *a* et *b* ont leur valeurs spécifiées, la ligne est uniquement définir et la valeur de *y* peut être calculée pour chaque valeur de *t*. Un nombre des paires des valeurs enregistrées de *y* et *t* sont disponibles pour le previsionniste. Ces valeurs ne doivent pas en général être exactement tout au lange de la ligne droite. Tâche est d'estimer *a* et *b*. Supposons que *n* enregistrement sont disponibles, les valeurs de *a* et *b* calculées sous le critère des moindres carrées sont obtenues par la résolution de paires d'équations simultanées:

$$\sum y = na + b \sum t \tag{2.3.5}$$

$$\sum t \cdot y = a \sum t + b \sum t^2 \tag{2.3.6}$$

où *n* = nombre de paires des valeurs (*t*, *y*);

Méthode

Calculer les valeurs suivantes:

$$\sum t y = t_1 y_1 + t_2 y_2 + \dots + t_n y_n$$

$$\sum t^2 = t_1^2 + t_2^2 + \dots + t_n^2$$

$$\sum t = t_1 + t_2 + \dots + t_n$$

$$\sum y = y_1 + y_2 + \dots + y_n$$

n = nombre de paires = nombre de valeurs *t* = nombre de values *y*

Ici: $\bar{t} = \frac{\sum t}{n}$ $\bar{y} = \frac{\sum y}{n}$

Finalement, calculer *a* et *b* pour le système d'équations (2.3.5) et (2.3.6):

$$b = \frac{\sum t \cdot y - n \cdot \bar{t} \cdot \bar{y}}{\sum t^2 - n \bar{t}^2} \tag{2.3.7}$$

et

$$a = \bar{y} - b \bar{t} \tag{2.3.8}$$

Pour chaque valeur t , la valeur y est donnée à partir de:

$$y = a + b \cdot t$$

qui est l'équation de la tendance de la ligne droite que nous avons cherché.

Tendance exponentielle

La croissance exponentielle peut être exprimée par l'équation

$$y = a \cdot e^{b \cdot t} \tag{2.3.9}$$

où y et t sont les variables, et a et b sont des paramètres de la courbe, et e = base pour le logarithme naturel.

Prenant le logarithme pour toute l'équation:

$$\ln y = \ln a + b \cdot t \tag{2.3.10}$$

Introduire

$$z = \ln y \quad \text{et} \quad c = \ln a$$

L'équation peut être écrite

$$z = c + b \cdot t \tag{2.3.11}$$

On voit que cette équation est une ligne droite.

Méthode

Prendre le logarithme de toutes les valeurs y et appeler les, valeurs z .

Alors calculer:

$$\sum tz \quad \sum t^2 \quad \sum t \quad \sum z$$

$$\bar{t} = \frac{\sum t}{n} \quad \bar{z} = \frac{\sum z}{n} \quad n \bar{t}^2$$

comme avant !

On a donc:

$$b = \frac{\sum t \cdot z - n \cdot \bar{t} \cdot \bar{z}}{\sum t^2 - n \cdot \bar{t}^2} \tag{2.3.12}$$

et

$$c = \bar{z} - b \cdot \bar{t} \tag{2.3.13}$$

$$z = c + b \cdot t$$

Pour chaque valeur t la valeur y est donc calculée comme

$$y = \text{anti-logarithme de } z \tag{2.3.14}$$

Vérifiant le modèle de régression

A. Testant la signification des paramètres du temps

Cette approche de régression dépend de la supposition que la quantité en considération (chaque trafic mesuré au taux d'appels) est corrélé avec le temps. Cela peut ou ne peut pas être vrai, les valeurs peuvent être peu indépendantes du temps. Les formules données ci-dessus devraient, cependant, produire toujours la valeur numérique de b . Il est, cependant, nécessaire de tester si cette valeur est arrivée par pure chance et si sa principale vérité devrait être zéro. Cela peut être fait par le calcul de la valeur

$$H = \frac{b \sum (t - \bar{t})^2}{s} \quad \text{où} \quad s^2 = \frac{\sum y^2 - a \sum y - b \sum t \cdot y}{n - 2} \quad (2.3.15)$$

Calculer en premier s^2 , calculer alors s = la racine carré de s^2 .

Les notations ci-dessus sont valables pour la ligne droite. Pour le cas exponentiel, calculer s^2 comme

$$s^2 = \frac{\sum z^2 - c \sum z - b \sum t \cdot z}{(n - 2)} \quad (2.3.16)$$

Pour une vérification exacte, il est nécessaire d'utiliser les tableaux statistiques, mais comme guide approximatif, H devrait être supérieur à 2 ou inférieur à - 2 pour la description linéaire qui doit être valable. Si les valeurs dans l'intervalle -2 à +2 sont obtenues, alors la valeur de b n'est pas significativement différente de zéro, cela veut dire que la quantité de trafic n'a pas changé systématiquement avec le temps et la description analytique n'est pas appropriée. La meilleure prévision pour les valeurs futurs de y est alors qu'il n'y a pas de changement du tout.

B. Examen des erreurs systématique

Même si le paramètre b passe le test ci-dessus, le modèle peut être invalide, cela est dû à l'existence des erreurs systématiques en cours la "bonne" ligne est tracée au points de données. Les procédures suivantes devraient être utilisées pour vérifier la forme systématique des erreurs:

Dénoter que les valeurs sur la ligne tracée correspondent aux valeurs enregistrée de y par \bar{y} , et plutôt dénotée les premières lectures par y_1, \bar{y}_1 , la seconde par y_2, \bar{y}_2 , et ainsi de suite.

Calculer

$$w = \sum_{i=1}^{n-1} (y_i - \bar{y}_i)(y_{i+1} - \bar{y}_{i+1}) \quad (2.3.17)$$

et
$$v = \sum_{i=1}^n (y_i - \bar{y}_i)^2 \quad (2.3.18)$$

Calculer la valeur $2 - 2 \frac{w}{v}$

Les statistiques, connus comme *statistique de Durbin-Waston*, devaient être entre 1.5 et 2.5 et préférablement entre 1.7 et 2.3. Si la valeur est en dehors de cet intervalle, les données devraient être examinées pour l'indication des déviations systématiques à partir de la courbe tracée. Des causes communes sont à la base des déformations des courbes (voir figure 2.3/2, graphe 3), ou la discontinuité des données (voir la même figure, graphe 4). Si les statistiques de *Durbin-Waston* donnent un résultat de test insatisfaisant, un ajustement approprié doit être fait aux données et l'analyse doit être refaite. Autrement les prévisions peuvent être sujet des erreurs sérieuses.

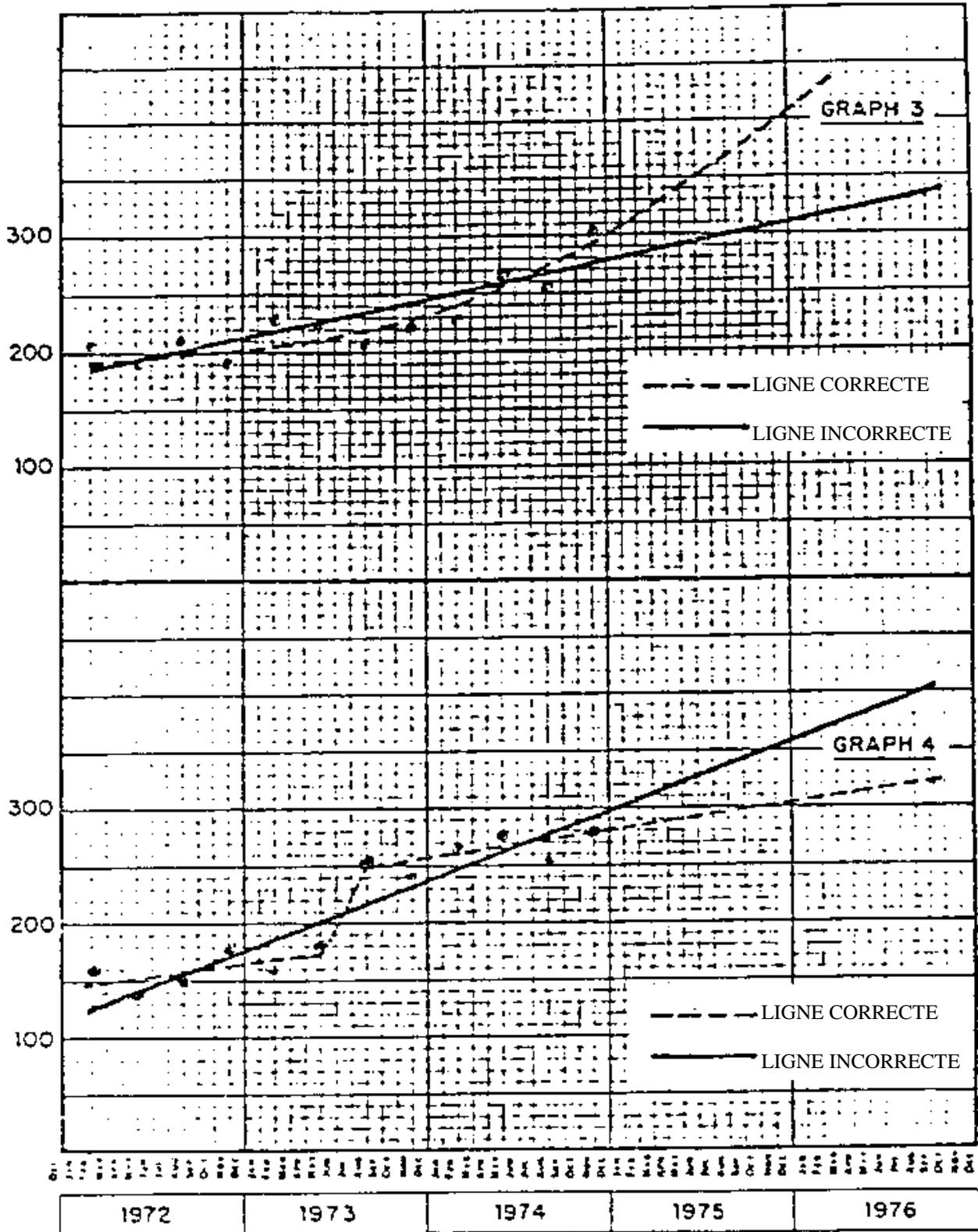


Figure 2.3/2

Prévision des niveaux futurs

Supposons que la description analytique du passé, passe le test décrit ci-dessus impeccablement, l'étape qui suit est de donner les prévisions demandées.

Tous cela est demandé pour calculer la valeur de y à partir de

$$y = a + b \cdot t$$

(Si la courbe de croissance exponentielle est tracée pour des données historiques, le niveau actuel des prévisions est obtenu par la prise de l'anti -logarithme naturel de la valeur de y.)

L'idée générale des précisions des prévisions peut être obtenue par désigner l'intervalle de confiance pour la valeur prévue, à l'intérieur duquel elle peut être fixé avec un degré de probabilité prédéterminer avec lequel la valeur future devait être.

Pour l'instant, si les prévisions pour le temps futur t_0 est y_0 , alors il y a approximativement une probabilité de 95 % que la valeur achevée de y devrait être à l'intérieur de l'intervalle.

$$y_0 \pm 2\sqrt{u}$$

où
$$u = s^2 \left[1 + \frac{1}{n} + \frac{(t_0 - \bar{t})^2}{\sum (t - \bar{t})^2} \right] \tag{2.3.19}$$

et s^2 est calculée à partir de (2.3.15).

Il devait être signalé (voir fig 2.3/3) que si la prévision est fait plus dans dans le futur, l'intervalle de confiance est plus large, en conséquence les prévisions deviennent plus incertaines.

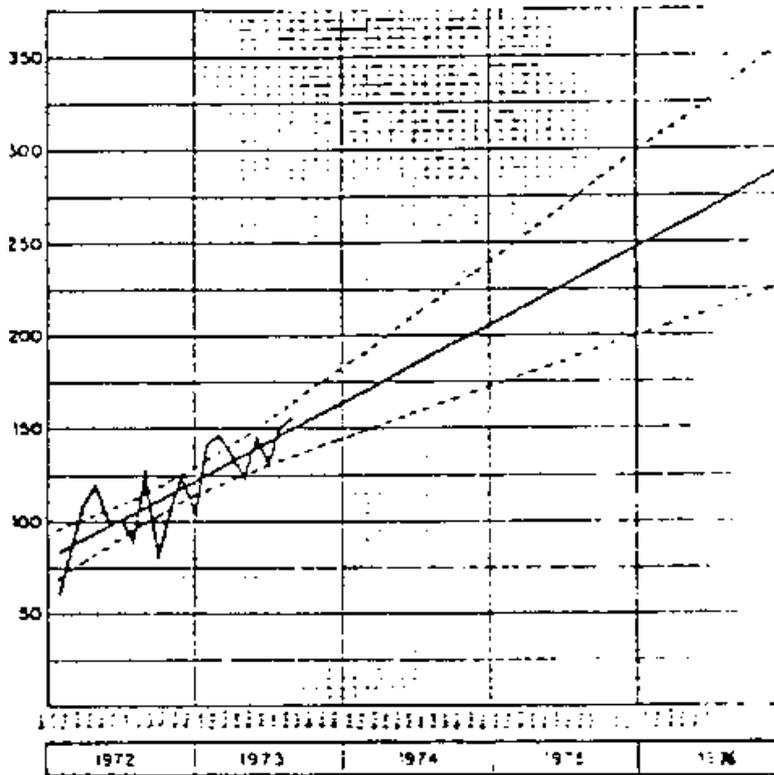


Figure 2.3/3 : L'intervalle de confiance (95 %) pour développement linéaire prévue

Si les quantités tracées dans le graphe montrent un modèle consistant dans le temps, il y a possibilité de tracer la courbe du meilleurs pas. Les valeurs futures sont donc prévues par l'extrapolation de cette courbe.

Le premier problème est de chercher l'équation de la courbe qui décrit le mieux la relation entre la quantité et le temps.

Pour plusieurs types d'équipements techniques, l'historique de son usage peut être décrit en trois phases. La première phase, on observe un usage accéléré qui change dans la seconde phase à une croissance linéaire dans le temps. Dans la troisième phase l'usage devient lent et éventuellement arrive au niveau de saturation. L'exemple d'un tel phénomène peut être le nombre de ménages avec radios, télévisions et téléphones.

Pour d'autres cas, le niveau de saturation peut devenir zéro. Par exemple: l'utilisation des taxis pour chevaux à London, l'utilisation du service télégrammes, l'utilisation des bateaux conventionnels pour le transport des biens, l'utilisation des haches à pierre dans la bataille, etc.

On devrait pas traiter avec ce deuxième type de courbes de développement.

Pour la quantité qui devrait atteindre le niveau de saturation dans le futur, il peut être possible de décrire toute l'histoire avec une expression mathématique. Il peut cependant être simple et plus exacte de décrire séparément chaque phase.

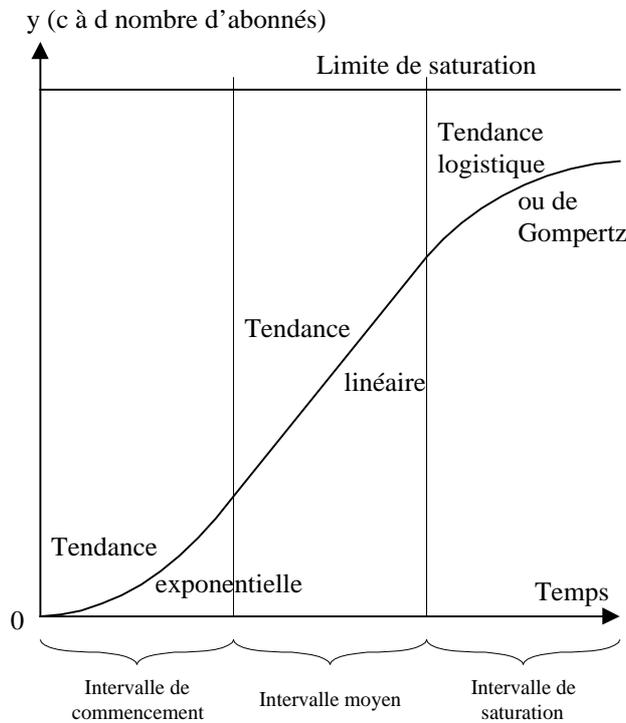


Figure 2.3/4
Développement dans le temps des services des télécommunications

Les courbes décrivant la tendance du développement dans le temps sont souvent appelées courbes des croissance, même si la "croissance" décroît parfois en quantité. Il y a quelque types communs de courbes des tendance:

Linéaire: $y = a + b t$ (2.3.20)

Parabolique: $y = a + b t + c t^2$ (2.3.21)

Exponentielle: $y = a e^{bt}$ (2.3.22)

Gompertz: $y = e^{a-b \cdot r(t)}$ (2.3.23)

Notations

- $t =$ point du temps (variable indépendante)
- $a, b, c, r =$ paramètres à calculer à partir des données historiques
- $y =$ l'article à prévoir (variable dépendante)
- $e =$ la base du logarithme naturel

Quelques exemples numériques des tendance:

1) *Tendance linéaire* $y = a + b \cdot t$

La formule a des paramètres inconnus a et b , qui pouvaient être calculés à partir des données historiques. Pour calculer deux paramètres on a besoin des deux équations.

Ces deux équation peuvent être obtenues à partir de deux points dans le diagramme à travers lequel la ligne droite devrait passer.

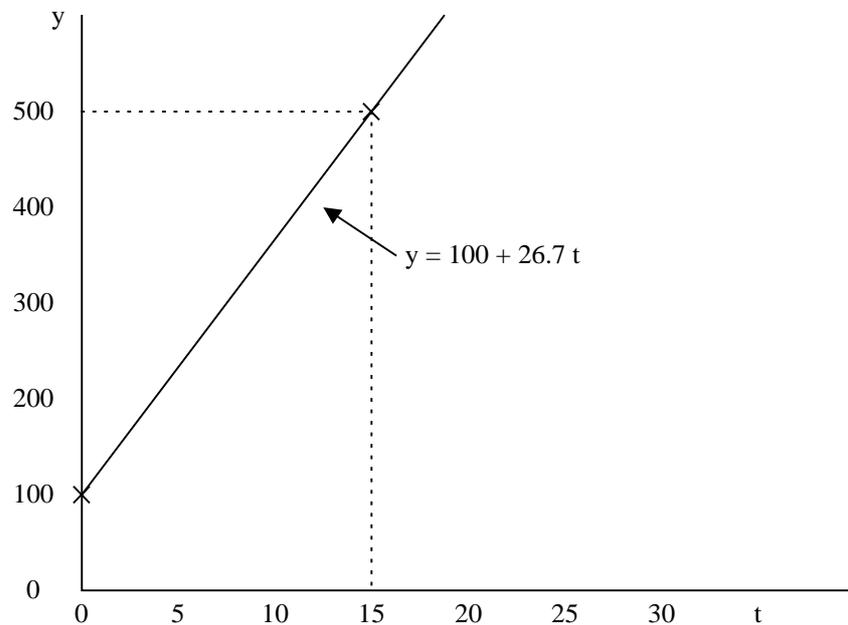


Figure 2.3/5
Tendance linéaire, estimées à partir de deux points

Supposons les deux points: $t = 0$ $t = 15$
 $y = 100$ $y = 500$

L'insertions des deux points dans l'équation de ligne droite (2.3.20) donne:

$$100 = a + b \cdot 0$$

$$500 = a + b \cdot 15$$

ce qui donne $a = 100$ et $b = \frac{500 - 100}{15} = 26.7$

Cela donne la tendance:

$$y = 100 + 26.7 \cdot t$$

2) *Tendance exponentielle* $y = a e^{b \cdot t}$

Supposons les mêmes points que le cas précédent:

$$t = 0 \qquad t = 15$$

$$y = 100 \qquad y = 500$$

Les deux équations demandées devraient être

$$100 = a e^{b \cdot 0}$$

$$500 = a e^{b \cdot 15}$$

Alors: $a = 100$ et b est calculé à partir de:

$$500 = 100 e^{b \cdot 15}$$

$$5 = e^{b \cdot 15}$$

$$\ln 5 = 1.609 = 15 \cdot b$$

ce qui donne

$$b = 0.1073$$

La tendance est: $y = 100 e^{0.1073 t}$

La courbe est tracée dans le diagramme ci-dessous:

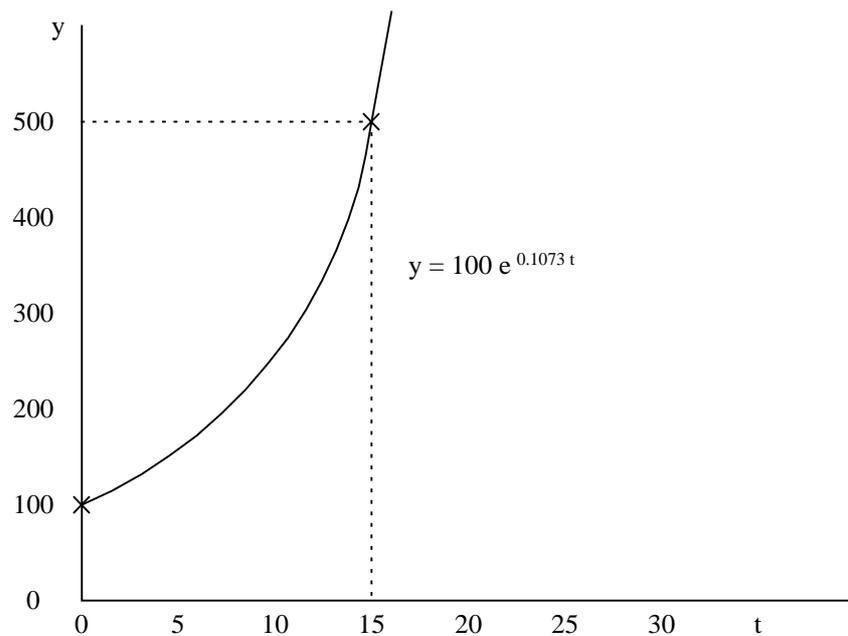


Figure 2.3/6
Tendance exponentielle, déterminée à partir de deux points

3) *Tendance de Gompertz* $y = e^{a - b r^t}$

Dans ce cas on a trois paramètres a , b et r , et le calcul nécessite trois équations pour la détermination des paramètres. On utilise les deux points comme avant:

$$t = 0 \qquad t = 15$$

$$y = 100 \qquad y = 500$$

La troisième équation peut être obtenue par le supposition d'une valeur de saturation à l'infini, c'est à dire le point:

$$t = \infty \quad y = 3000 \quad (\text{c'est à dire la valeur 3000 devrait être atteinte après un nombre infini d'années})$$

Les paramètres sont calculés à partir de:

$$100 = e^{a-b \cdot r^0} \quad 500 = e^{a-b \cdot r^{15}}$$

La troisième équation donne pour $r < 1$ et $t = \infty$

$$3000 = e^a \quad \text{donc} \quad a = \ln 3000 = 8.006$$

Si $a = 8.006$ est insère dans la première équation,

$$4.605 = 8.006 - b; \quad \text{ce qui donne} \quad b = 3.401$$

Alors r est calculé à partir de la seconde équation pour

$$a = 8.006 \quad \text{et} \quad b = 3.401$$

$$6.215 = 8.006 - 3.401 r^{15}$$

ou
$$r = \left(\frac{8.006 - 6.215}{3.401} \right)^{1/15} \quad r = 0.958$$

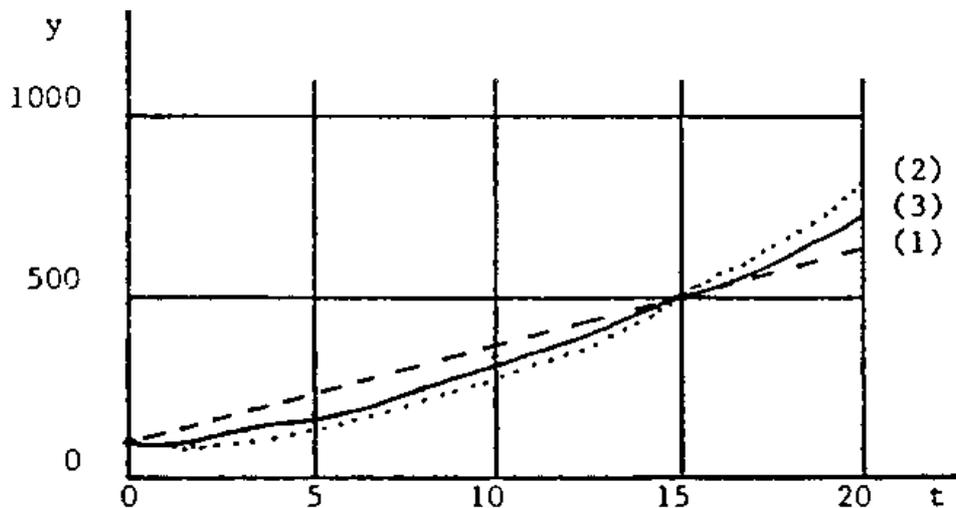


Figure 2.3/7 : Exemple numérique des tendances

- 1) Tendence linéaire: $y = 100 + 26.7 \cdot t$
- 2) Tendence exponentielle: $y = 100 \cdot e^{0.1073 \cdot t}$
- 3) Tendence de Gompertz: $y = e^{8.006 - 3.401(0.958)^t}$
(valeur de saturation = 3000)

Autres courbes de croissance:

On considère

$$y = M - a \cdot e^{-t} \quad (2.3.24)$$

Il y a quatre configuration de cette courbe, et ce selon les signes de a et de b . Elles sont illustrées ci-dessous.

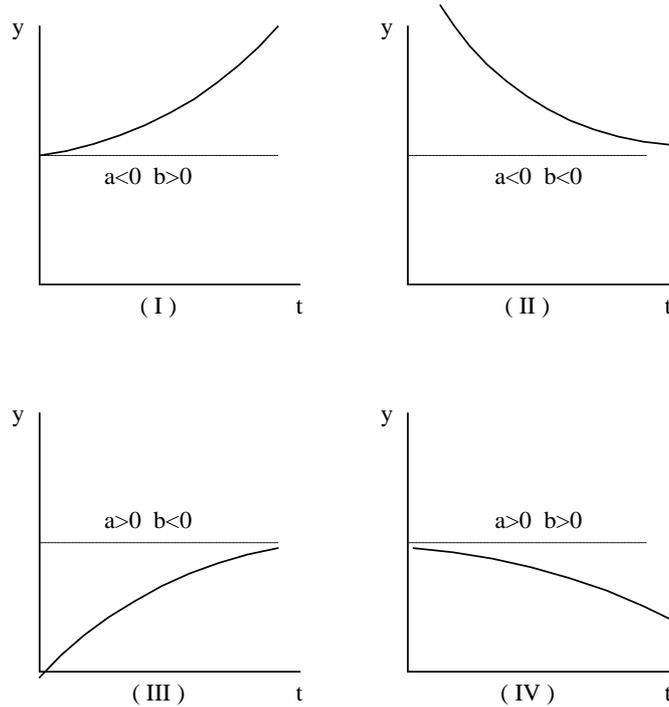


Figure 2.3/8 : Courbes exponentielles modifiées

Quand il s'agit des prévisions de demande en télécommunications, seulement (ii) et (iii) sont applicables pour les prévisions à long terme, (ii) est moins fréquemment utilisée. Un déclin vers le minimum comme montré en (ii) peut être utilisé pour de telles variables comme le nombre d'appels manuels par abonnés ou le nombre des télégramme par habitant. Cependant peu de services des télécommunications sont caractérisés par l'étude du déclin. Ainsi (iii) est plus applicable. Evidement l'approche sous-entendue au niveau de saturation comme montré à (iii) peut couvrir une période sur plusieurs décennies. Il est possible de l'appliquer pour les prévisions de la densité téléphonique dans les pays à forte pénétration, mais elle n'est pas applicable où le développement téléphonique est dans sa phase primaire. Une alternative à la courbe de Gompertz est la courbe logistique, comme décrite ci-dessous.

Courbe logistique

La courbe logistique simple est donné par le formule:

$$y = F + \frac{M - F}{1 + a \cdot e^{-bt}} \tag{2.3.25}$$

La courbe a une forme en S avec une valeur inférieur F et un maximum M . La courbe croit à son maximum dans le temps pour $a > 0, b > 0$.

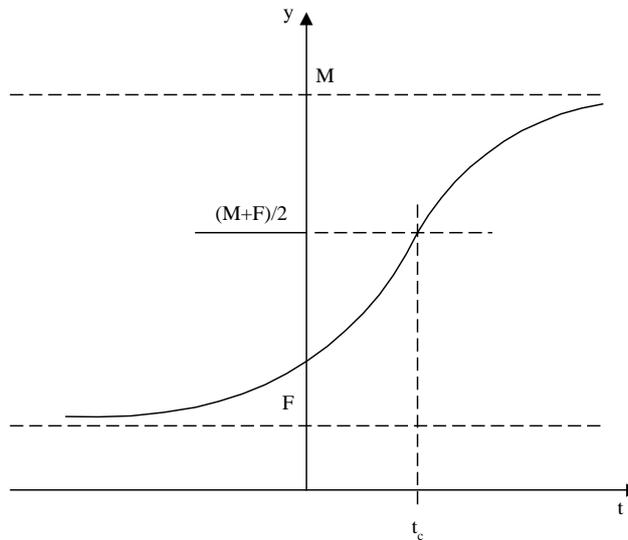


Figure 2.3/9 : Courbe logistique

La courbes a les propriété suivantes:

- i. La croissance augmente d'une manière stable jusqu'au sommet et puis décroît.
- ii. Le point d'inflexion au niveau dans lequel le taux de croissance absolue change de la croissance jusqu'à décroissante au niveau entre l'apogée et le périgée, au temps $t = t_c$, où t_c est donné par $1/b \cdot \ln(1/a)$.

Commentaires

Le but de l'analyse des tendances peut ne pas être toujours à découvrir comment la moyenne de toutes les valeurs de trafic change avec le temps. Pour les buts de dimensionnement, on est particulièrement intéressé à comment la partie supérieure de la distribution du trafic annuel change avec le temps, et ce dans le but d'avoir une valeur représentative du trafic à l'heure chargée. Ce trafic à l'heure chargée détermine donc le nombre de circuits et les autocommutateurs nécessaires dans le futur. Pour mettre en pratique l'extrapolation significative des séries chronologiques, il est, cependant essentiel que les données utilisées se réfèrent réellement aux valeurs représentatives pour notre but. L'ensemble des données historiques basées sur les minutes taxées (payées) par mois, par exemple, pourraient certainement ne pas être une valeur des efforts pour chaque analyse sérieuse des séries chronologiques. Il est ainsi, parce que la relation entre le trafic à l'heure chargé et les minutes payées durant le mois n'est pas généralement établie par expérience, mais devrait être calculée comme le produit de plusieurs facteurs inexactes.

Un autre cas où les séries chronologiques ne peuvent pas fonctionner, est quand les conditions opérationnelles subissent des changements. Par exemple, si le service est soudainement amélioré d'une manière remarquable, le trafic change toujours énergiquement. De tels sauts d'amélioration de services peut signifier une augmentation en matière de trafic à 100 % ou plus. Pour estimer la grandeur d'une telle croissance brusque, en trafic, l'information de performance de service est nécessaire avant et après le changement, combien de trafic a été supprimé avant le changement, etc. De telles données ne sont pas généralement disponibles ou sont plutôt difficiles à collecter.

2.4 Jugement individuel

La méthode de prévision la plus simple mathématiquement est le jugement individuel. La prévision est basée sur l'expérience et les informations collectées. Aucune analyse systématique n'est faite.

Cependant, la qualité de prévision dépend directement de la qualité des données de base. A moins que les données des base fiables sont disponibles, le prévisionniste doit établir un processus de collecte de données adéquat avant de commencer ses prévisions, ou plutôt, de fixer les données les plus fiables pour son futur travail.

Un exemple typique de prévisions basé sur le jugement individuel est l'estimation du nombre futur d'abonnés dans une zone donnée. Ces prévisions sont généralement liées aux futurs bâtiments. Chaque type de bâtiment a sa densité téléphonique typique. Les prévisions à court terme peuvent concerner les bâtiments existants, bâtiment en construction et planifiés. Les prévisions à moyen et long terme peuvent concerner les différents types de zones comme, quand les zones avec les bâtiments administratifs demanderont plus de téléphones que dans les zones résidentielles. Le

prévisionniste peut, comme aide, avoir des tableaux montrant le nombre moyen d'abonnés dans les différents types de bâtiments et de zones. Quand cette étude est faite, l'information correcte de la planification des villes est essentielle, comme l'expérience passée.

Un exemple du nombre moyen d'abonnés pour les différents types de bâtiments est montré au tableau 2.4/1. Le tableau devrait naturellement être ajusté conformément à la situation du pays ou les villes concernées. Pour les prévisions à moyen et long terme, la table 2.4/2 montre les valeurs typiques pour les différents types de zones d'agglomération. Les chiffres devraient également être ajustés conformément au pays concerné. Les chiffres donnés dans les tableaux devraient être mis à jour quand la densité téléphonique augmente.

Type de bâtiment	Nombre d'abonnés
Bâtiments officiels et offices, banques, compagnies d'assurance, grands hôtels, clubs, grands restaurants, hôpitaux, grands magasins de commerce.	a étudier
Petits hôtels, restaurants, magasins alimentaires, maisons de pensions	1 - 2
Chemistes, docteurs, avocats, etc	1 - 1.5
Magasins	0.5 - 1
Grande usine	examiner
Petites usines, atelier	0.5 - 1.5
Cinéma, stations d'essence	1 - 2
Maisons privées pour la haute classe	1
Maisons privées pour la basse classe	0.3 - 0.5
Maisons en terrasses	0.3 / appartement
Bloc d'appartements, haute classe	0.5 - 1 / appartement
Bloc d'appartements, basse classe	0.2 / appartement

Table 2.4/1 Estimation du nombre d'abonnés par différents types de bâtiment. Cet exemple est valable pour les pays en développement avec 4-8 abonnés par 100 habitants.

Type de zone construite	Abonnés par hectare
A. Quartiers pauvres	0.25
B. Parcs, jardins, etc.	0.50
C. Ancienne maison privée avec grand jardin	1.00
D. Districts résidentiel des pauvres travailleurs	1.50
E. Districts résiduelles pour les travailleurs de bonne classe	2.00
F. Maison moderne privée avec grand jardin	3.00
G. Districts modernes résidentiels pour les travailleurs	4.00
H. Zones industriels	5.00
I. Maisons modernes privées avec petits jardins	7.00
J. Maisons non-détachées de type ancien	8.00
K. Zones pour la classe ouvrière et petits magasins	10.0
L. Maisons modernes non détachées	13.0
M. Bâtiments résidentiels non détachés 1-2 étages et petits magasins	18.0
N. Blocs d'immeubles jusqu'à 4 étages	25.0
O. blocs d'immeubles et magasins à 4 étages	28.0
P. Centre d'affaires dans une zone résidentielle	30.0
Q. Blocs d'immeuble supérieur à 4 étages	40.0
R. Bâtiments des offices à plus de 3 étages	80.0
S. Bâtiments des offices de 4 à 6 étages	150
T. Bâtiments des offices de plus de 6 étages	250

Table 2.4/2 Estimation du nombre d'abonnées par hectare dans les zones faites des différents types.

Cet exemple est valable pour les pays en développement avec 4-8 abonnés pour 100 abonnés.

2.5 Autres méthodes

Ce qui suit, est vue générale de quelques autres méthodes utilisées pour les prévisions.

Comparaisons analytiques

Cette méthode relaye des données à partir d'autres, préférablement les zones comparables où le développement est plus avancé. La prévision est donc faite par comparaison et supposition que le développement devrait être le même.

Cette méthode est utilisée quand les séries de données historiques adéquates sont très courtes pour permettre l'analyse fiable de tendance. C'est fréquemment le cas dans les pays en développement. Il est donc supposé que le développement devrait suivre le même comportement comme dans les pays ou les zones les plus développés.

Le tableau 2.5/1 montre l'accroissement en matière de pénétration téléphonique dans les période 1960-1975, exprimé par le nombre de téléphones pour 100 habitants pour certains pays. Il montre aussi durant cette période le développement des Royaumes-Unies au même taux et combien d'années été nécessaire pour aller de la pénétration de 1960 à celle de 1975.

Pays	Téléphone par 100 habitants		Pénétration correspondante pour UK (RU)	
	1960	1975	Dates	Période
Argentine	5.99	9.41	12 années	1937-1949
Chili	2.43	4.26	8 années	1923-1931
Costa Rica	1.34	5.02	25 années	1910-1935
Chiper	2.95	10.65	25 années	1925-1950
Fiji	1.71	4.53	18 années	1914-1932
Grèce	2.30	20.71	45 années	1922-1967
Irlande	5.13	12.78	20 années	1935-1955
Japon	5.21	37.88	41 années	1935-1976
Mexique	1.46	4.37	20 années	1911-1931
Espagne	5.47	12.46	18 années	1936-1954

Table 2.5/1 Nombre de téléphones pour 100 habitants pour certains pays.

Le tableau 2.5/1 montre qu'il n'y a pas de claire connexion entre le temps qu'il prend un pays particulier pour développer ses valeurs de 1960 à 1975, et le temps qu'il faut aux Royaume-Uni pour arriver au même degré de développement. Il n'est pas cependant recommandé de comparer les pays avec de très grande différence en matière de développement. Le tableau 2.5/2 donne la même information que le tableau précédent, mais seulement pour les pays qui ont le même développement économique.

Pays	Téléphone pour 100 habitants		Dates correspondantes pour le R-U (UK)	
	1960	1975		
Australie	20.88	37.49	9 ans	1969-1976
Danemark	22.17	42.48	10 ans	1968-1978
Finlande	12.89	35.78	20 ans	1955-1975
Allemagne (F.R.)	9.98	30.25	24 ans	1949-1973
Iceland	21.42	40.41	10 ans	1967-1977
Pays bas	13.15	34.41	19 ans	1955-1974
Norvège	19.50	33.90	8 ans	1966-1974

Table 2.5/2 Pénétration téléphonique 1960-1975 pour quelques pays avec approximativement le même développement économique.

Ce tableau montre une bonne conformité avec le développement aux Royaumes Unies plus que le tableau précédent.

On peut maintenant conclure ce qui suit:

1. La projection à court terme pour une densité téléphonique des pays à partir d'un niveau donné, par comparaison avec la croissance de quelques pays plus avancés qui était au même niveau auparavant, ne peut pas être recommandée toujours puisque des circonstances spéciales peuvent influencer le développement à ne pas suivre la même tendance. Les quelles circonstances peuvent être: récession économique ou l'expansion, limitation des provision, différentes politiques tarifaires, etc.
2. Pour les projections à long terme, il est nécessaire de choisir un pays qui est au moins avancé de 15 ans et ce dans le but d'avoir des séries suffisamment longues pour un accroissement futur possible.
3. Si une telle comparaison est faite, cependant, on suppose que la comparaison entre la performance d'un pays dans les quinze dernières années et un autre pays dans les quinze années précédentes est valable. Cela n'est pas probable, du fait du changement rapide dans l'économie mondiale et par ce qu'il y a des événements perturbateurs tels que les guerres et les grandes récessions qui faussent probablement la comparaison.

Méthodes supplémentaires:

La méthode oculaire

La méthode est que le prévisionniste inspecte simplement les données en ordre pour avoir une idée de quelques tendances possibles ou relations entre les variables et donc manuellement essayer de tracer les courbes correspondantes.

Moyenne mobile

La nouvelle valeur est calculée comme la moyenne du nombre de valeurs observées. Le but de cet approche est de réduire les irrégularités causées, pour l'instant, par les variations saisonnières.

Glissement exponentiel

Similaire à la méthode de la moyenne mobile, mais le poids est donné aux observations de telle manière que les observations récentes ont plus de poids.

Recherche du marché

Cette méthode donne une étude du marché par des questions à poser aux clients potentiels.

Méthodes combinées (incluant l'analyse structurelle)

Une combinaison de différentes méthodes telles que recherche du marché, courbes de croissances, projections et modèles économétriques.

3. PREVISION POUR LA PLANIFICATION DES CENTRAUX

Pour la planification des centraux, la demande devrait être prévue pour tous les articles particuliers pour chaque central séparément. Cela inclue ce qui suit, par abonné et par catégorie:

- le nombre des lignes directes d'abonnés (DEL)
- trafic de départ
- trafic d'arrivée
- trafic local
- trafic longue distance
- trafic international
- services spéciaux

Une fois ces prévisions sont faites, les chiffres correspondants à partir des différents centres peuvent être ajoutés pour produire le résultat des grandes zones, ou pour le pays en entier. Cela est connu comme prévisions "bottom up" ou la méthode ascendante.

Quand la prévision pour une zone devrait être faite, toutes les informations disponibles relatives au futur changement dans cette zone devraient être collectées. Les informations pertinentes obtenues dans ce sens peuvent la faire possible de voir les discontinuités dans le développement de la demande. En d'autre terme, les fluctuations des quantités d'intérêt sont relativement large quand on a, à faire à une petite zone. Le nombre de rapports historiques fiables est aussi souvent très limité.

Ces deux facteurs combinent de faire la prévision des petites zones avec moins de fiabilité. Ajoutant les résultats pour obtenir les chiffres correspondants pour la zone élargie peut laisser les erreurs s'aggraver. Pour faire face à ça, la prévision "bottom up" est comparée avec la prévision "top down".

Commençant à partir de la prévision du pays entier, ou pour une zone large, on peut arriver aux prévisions pour les petites région par scinder les quantités prévues. Toujours, les données plus chère et fiables sont disponibles pour les prévisions pour les zones de grande taille. Cependant, la prévision "top down" est généralement vue comme plus faible que la prévision "bottom up".

Il y a deux façons dans lesquelles l'accord entre les deux prévisions peut être atteint:

1. En changeant la division de la quantité totale sur les différents zones.
2. En réduisant les prévisions locales par un facteur donné, soit uniformément ou après quelque système.

Clairement, cette procédure d'assortiment des deux prévisions est plutôt subjective. Il est aussi claire qu'une zone qui a été soumise à de grandes prévisions devrait réduire ses chiffres jusqu'à une raison satisfaisante pour cette grandeur de prévision peut être donnée. Une telle raison peut être un plan de développement connu de cette zone, qui motive un accroissement soudain de la demande. La prévision de la grande zone ne prend pas souvent en compte de telles particularités locales.

4. PREVISION POUR LA PLANIFICATION DES RESEAUX

4.1 Introduction

La planification du réseau téléphonique est basée sur l'estimation du trafic futur nécessaire. La prévision à long terme est nécessaire pour le plan de développement et ce, pour assurer l'extension coordonnée sur une période de 15 à 25 ans. Approximativement chaque 2 à 4 ans ce plan de développement devrait être mis à jour.

A l'intérieur du plan de développement, les prévisions à court terme sont nécessaires pour fournir les données de base pour planifier les étapes actuelles d'extension. Elles devraient contenir l'estimation du trafic nécessaire pour les quatre ou les six prochaines années pour s'approvisionner au moment opportun, etc. La prévision à court terme devrait être mise à jour chaque année.

Les trafics nécessaires sont exprimés en Erlang. Pour prévoir ces besoins, le trafic entre chaque paires de centraux est généralement estimé séparément pour chaque direction. Le trafic à l'intérieur de chaque zone de centre est aussi estimé.

4.2 Matrice de trafic

Pour spécifier les besoins en trafic dans une région à n centre, n^2 valeurs de trafic sont nécessaires. Une façon standard pour spécifier ces trafics est de les présenter dans une matrice, qui est appelée matrice de trafic.

de	vers	1	i	j	n	SO
1		A(11)			A(1n)	O(1)
i			A(ii)	A(ij)		O(i)
j			A(ji)	A(jj)		O(j)
n		A(n1)			A(nn)	O(n)
ST		T(1)	T(i)	T(j)	T(n)	A(11)

Ici:

$A(ij)$ est le trafic de i vers j ;

$A(ji)$ est le trafic de j vers i ;

$A(ii)$ est le trafic local dans le central i ;

$O(i)$ est la somme de tous les trafics départ de i ;

$T(j)$ est la somme de tous les trafics arrivée à j .

La somme de tous les totaux de lignes $O(i)$, c'est à dire les entrées dans la colonne SO (somme du trafic départ) donne le trafic total $A..$. Le même résultat est obtenu par l'addition de tous les totaux des colonnes $T(j)$, c'est à dire les entrées de la ligne ST (somme de trafic d'arrivée) En bref:

$$\sum_i O(i) = \sum_j T(j) = A..$$

Pour qu'il n'y ait pas de confusion, le symbole $A(i, j)$ peut être utilisé. Cependant, il devrait fréquemment être nécessaire de distinguer entre le trafic présent de i à j $A(i, j/0)$ et le trafic estimé à une date future t : $A(i, j/t)$. Evidemment,

$$O(i/t) = \sum_j A(ij/t) \quad \text{et} \quad T(j/t) = \sum_i A(ij/t)$$

4.3 Prévision point-à-point

Il existe de différentes méthodes pour estimer $A(i,j/t)$ basée sur l'accroissement espéré en matière du nombre d'abonnés dans les zones (i) et (j), changement espéré dans le trafic par abonné, etc. La matrice de trafic peut donc être complétée par l'addition des entrées des lignes pour obtenir le $O(i/t)$ et des colonne pour obtenir le $T(j/t)$.

Pour l'estimation des trafics point-à-point dans le réseau, on base souvent les calculs sur, l'accroissement prévu des lignes d'abonnées et de la matrice actuelle de trafic. Différentes formules sont utilisées, pour lesquelles les plus communes devraient être données ci-dessous. On ne peut pas dire qu'une formule est plus exacte que l'autre. Seulement les informations à partir des rapports futurs peuvent indiquer quelle formule est la meilleur pour un cas particulier. Quand cela est trouvé, il n'y pas cependant, de garantie que cela devrait être valable pour toujours.

Estimation du trafic total

Prenant en considération que différentes catégories d'abonnés produisent différents quantités de trafic, il peut être possible parfois d'estimer le trafic futur à partir de:

$$A(t) = N_1(t) \cdot \alpha_1 + N_2(t) \cdot \alpha_2 + K \quad (4.3.1)$$

où $N_1(t)$, $N_2(t)$, etc., sont les nombres d'abonnés prévus pour les catégories 1,2, etc, et α_1 , α_2 , etc., sont le trafic par abonné des catégories 1, 2, etc.

S'il n'est pas possible de séparer les abonnés en catégories avec différent trafic, le trafic futur peut simplement être estimé comme:

$$A(t) = A(0) \cdot \frac{N(t)}{N(0)} \quad (4.3.2)$$

où $N(t)$ et $N(0)$ sont les nombres d'abonnés pour les temps t et zéro.

Estimation du trafic point-à-point

Pour l'estimation du trafic d'un centre vers un autre, plusieurs formules peuvent être utilisées. L'idée principale est de prendre en compte l'accroissement en abonnés dans les deux centres et d'appliquer certaines facteurs de pondération pour ces accroissement.

$$A_{ij}(t) = A_{ij}(0) \cdot \frac{W_i G_i + W_j G_j}{W_i + W_j} \quad (4.3.3)$$

où W_i et W_j sont des poids et G_i est l'accroissement en abonnés dans le centre i , et G_j dans le centre j .

$$G_i = \frac{N_i(t)}{N_i(0)} \quad G_j = \frac{N_j(t)}{N_j(0)} \quad (4.3.4)$$

Pour W_i et W_j il existe différentes méthodes.

Formule 1 de Rapp

$$W_i = N_i(t) \quad W_j = N_j(t) \quad (4.3.5)$$

La supposition ici est que le trafic par abonné du centre i au centre j est proportionnel au nombre d'abonné au centre j .

Formule 2 de Rapp

$$W_i = N_i(t)^2 \quad W_j = N_j(t)^2 \quad (4.3.6)$$

Cette formule suppose que le changement du trafic départ et arrivée par abonné est aussi petit que possible.

Formule des télécommunications d'Australie

$$W_i = \frac{N_i(0) + N_i(t)}{2} \quad W_j = \frac{N_j(0) + N_j(t)}{2} \quad (4.3.7)$$

Cette formule est une modification de la première formule de Rapp.

Une quatrième formule est dérivée à partir des suppositions suivantes: le trafic par un abonné au centre i vers tous les abonnés au centre j est constant.

$$\frac{A_{ij}(t)}{N_i(t) \cdot N_j(t)} = \frac{A_{ij}(0)}{N_i(0) \cdot N_j(0)} \quad (4.3.8)$$

$$A_{ij}(t) = A_{ij}(0) \cdot G_i \cdot G_j$$

Ces quatre formules peuvent être plutôt ajustées par l'introduction de N dans les facteurs de pondération.

Modèle de gravité

Le trafic entre deux centres peut être exprimé comme:

$$A_{ij} = K(d_{ij}) \cdot N_i \cdot N_j \quad (4.3.9)$$

où $K(d_{ij})$ = facteur d'intérêt communautaire.

Il a été trouvé que ce facteur dépend, à un certain degré, de la distance entre les centraux d_{ij} . On peut, cependant, écrire $K(d_{ij})$ comme suit:

$$K(d_{ij}) = e^{-\gamma \cdot d_{ij}} \quad (4.3.10)$$

ou
$$K(d_{ij}) = d_{ij}^{-g} \tag{4.3.11}$$

La valeur des paramètre γ ou g peut être calculée à partir de la matrice de trafic connue. Il peut être nécessaire d'ajuster l'expression de A_{ij} pour les paires de centres avec les relations spéciales à chaque autre; par exemple, une grande usine dans une partie du pays et la maison mère dans une autre partie.

Les formules données ici sont toutes déduites sous certaines suppositions. Une supposition est que le trafic par abonné reste constant, une autre est qu'un abonné a le même intérêt d'appeler chaque autre abonné. Il est, cependant, trouvé dans les pays développés que les nouveaux abonnés peuvent avoir un faible trafic que les anciens. Il est partiellement pris en compte dans la formule d'Australie.

En principe il est impossible de confirmer qu'une formule est plus fiable d'une autre. Le prévisionniste devrait essayer toutes les différentes formules et trouver laquelle donne les résultats les plus crédibles pour son cas particulier.

4.4 Méthode de Kruithof à double facteur

La méthode de Kruithof nous rend capable d'estimer les valeurs futures individuels du trafic, où $A(i,j)$ dans la matrice de trafic. Les valeurs, à présent, sont supposées être connues ainsi que les sommes futures des lignes et des colonnes.

La procédure est d'ajuster l'individuel $A(i,j)$ conformément aux nouvelles sommes des lignes et des colonnes, c'est à dire

$$A(i, j) \text{ est changé en } A(i, j) \cdot \frac{S_1}{S_0}$$

où S_0 est la somme actuelle et S_1 est la nouvelle somme pour la ligne ou la colonne individuelle.

Si on commence à l'ajustement de $A(i, j)$, selon les nouvelles sommes des lignes, S_j , ces sommes devraient coïncider, mais ce n'est pas le cas pour les sommes des colonnes. L'étape suivante est donc d'ajuster les valeurs trouvées $A(i, j)$ de manière à être conformes avec les sommes des colonnes. Cela rend les sommes des lignes non conforme, alors l'étape suivante est d'ajuster les nouvelles valeurs $A(i, j)$ conformément à les sommes des colonnes. La procédure continue jusqu'à ce qu'une exactitude suffisante pour les sommes des lignes et des colonnes soit atteinte. L'itération est plutôt rapide et donne en général des résultat satisfaisants après environ trois corrections.

La procédure est mieux comprise à partir de l'exemple numérique donné à dessous.

La méthode est applicable quand le changement des proportions entre des $A(i, j)$ individuels n'est pas espéré changer beaucoup, et pour les cas où il n'est pas possible de prévoir les $A(i, j)$ par d'autres méthodes.

Exemple d'utilisation de la méthode de Kruithof à double facteur

Considérons un réseau téléphonique avec deux centres:

Données:

1. L'intérêts de trafic actuel $A_{ij}(0)$

i	j	1	2	somme
1		10	20	30
2		30	40	70
somme		40	60	100

(4.4.4)

2. Prévoir le trafic total de départ et d'arrivée par central:

$A_i(t)$ et $A_j(t)$:

i	j	1	2	somme
1				45
2		?		105
somme		50	100	150

(4.4.5)

Problème:

Estimer les valeurs et trafic $A(i, j/t)$ avec la méthode de Kruithof.

Solution:

Itération 1: Multiplication des lignes.

A_i est distribué comme donné par l'intérêt de trafic actuel.

Résultat:

i	j	1	2	somme
1		15	30	45
2		45	60	105
somme		60	90	150

$$A_{ij}(1) = \frac{A_{ij}(0)}{A_i(0)} \cdot A_i(t) \quad (4.4.6)$$

Après la multiplication des lignes, la somme des colonnes diffère de ce qui était prévu. La prochaine itération devrait être la multiplication des colonnes.

Itération 2: Multiplication des colonnes.

$A_j(2)$ est distribué comme donné par l'itération 1.

Résultat:

i	j	1	2	somme
1		12.5	33.33	45.83
2		37.5	66.67	104.17
somme		50	100	150

$$A_{ij}(2) = \frac{A_{ij}(1)}{A_j(1)} A_j(t) \quad (4.4.7)$$

Après la multiplication des colonnes, les sommes des lignes diffèrent des valeurs prévues. La prochaine itération devrait être la multiplication des lignes.

Itération 3: Multiplication des lignes.
 A_i est distribué comme donné par l'itération 2.

Résultat:

i	j	1	2	somme
1		12.27	32.73	45
2		37.80	67.20	105
somme		50.07	99.93	150

$$A_{ij}(3) = \frac{A_{ij}(2)}{A_i(0)} A_i(t) \tag{4.4.8}$$

Itération 4: Multiplication des colonnes.
 A_j est distribué comme donné par l'itération 3.

Résultat:

i	j	1	2	somme
1		12.25	32.75	45
2		37.75	67.25	105
somme		50	100	150

$$A_{ij}(4) = \frac{A_{ij}(3)}{A_j(0)} A_j(t) \tag{4.4.9}$$

Après 4 itérations les sommes des lignes et des colonnes sont égales aux valeurs prévues. On peut écrire:

$$A_{ij}(t) = A_{ij}(4) \tag{4.4.10}$$

5. CONCLUSIONS

Ce chapitre a été concerné principalement avec des mathématiques derrière les méthodes de prévision. Cela devrait le faire possible pour le lecteur d'être capable d'assister dans le travail des prévision, spécialement en ce qui concerne le traitement préliminaire des données historiques.

A côté de l'explication, quand et pourquoi les prévision sont nécessaire, ce chapitre contient quelques méthode fréquemment utilisées pour la description des données historiques en formules mathématiques. Il devrait être accentué encore que le *traitement des données historiques ne signifie pas les prévisions*. Il signifie uniquement le traitement des données historiques. Les prévision, en d'autres termes, est l'art de deviner le développement futur. Seulement dans certains cas, il doit être jugée si les conditions futures et les suppositions sont les mêmes pour le passé. Les ajustements des valeurs extrapolées peuvent être nécessaires pour différentes raisons. L'amélioration du service peut doubler le trafic pendant un ou deux mois. Les nouvelles activités dans la société (p.e. industrie) peuvent causer d'énormes changement dans la dispersion du trafic, etc.

La prévisions concernent généralement la *demande* des souscriptions et trafic causée par différents services des télécommunications. Depuis que les quantités de demande sont difficiles ou impossibles à mesurer, les données historiques mises à la disposition du prévisionniste sont généralement des quantités mesurées, tel que le trafic écoulé. Cependant, le prévisionniste devrait avant tout transférer ces données en données de demande. Si elles concernent le pays et le service où la demande est bien satisfaite, la demande peut être estimée avec une exactitude acceptable.

Si la demande est plus loin que la provision, il est très difficile d'estimer la demande réelle. Le prévisionniste devrait prendre en compte cette difficulté tant que les conditions ne permettent pas un équilibre entre la provision et la demande.

Une autre difficulté dans l'utilisation des données historiques est qu'il peuvent ne pas être suffisamment fiables pour permettre une analyse. Une raison pour cela est que les données ont été collectées par différents principes, pour différents buts et aux moments différents. Il peut également arriver que les données historiques ne sont pas disponibles du tout. Les prévisions devraient donc être basées sur d'autres indications indirectes, qui n'améliorent pas généralement la fiabilité des prévisions.

6. REFERENCES

- BEAR, D. Some Theories on Telephone Traffic Distribution; A critical survey (ITC-paper, Stockholm, 1973)
- LEIJON, K.H. Introduction to Practical Teletraffic Engineering. Section F. Forecasting (ITU REM/72/038, Benghazi, 1976)
- CCITT (GAS 5) Methods Used in Long-term Forecasting of Domestic Telecommunication Demand and Required Resources (1980)
- LONNQUIST, I. Theory and Methods for Traffic Forecasts (TELE, Stockholm, 1968)
- LONNQUIST, I. Forecasting. Teletraffic Engineering Seminar, Istanbul, 1980 (ITU)
- LONNSTROM, M.M. A Telephone Development Project, Stockholm, 1967
- MARKLUND, F.
- MOO, I.
- NYGAARD, J. An Integrated Telephone Traffic Forecasting Tool for Use in Planning, I. (Teleteknik, Vol. XIV, 1970, No. 1)
- RAPP, Y. Some Economic Aspects on the Long-term Planning of Telephone Networks (Stockholm, 1969)
- TURNER, W.M. Forecasting. ITU/SIDA Seminar, New Delhi, 1975. SEM/ASIA/75
- CCITT (GAS 5) Economic Studies at the National Level in the Field of Telecommunications

Note: Ce chapitre a été édité à partir d'un manuel préparé par K.H. Leijon, L.M. Ericsson, Sweden.

7. EXERCISES

TECHNIQUES DE PREVISION

1. Un objet est à prévoir. Le développement historique a été comme suit:

Année (à la fin de)	Echelle de temps (t)	Stock (y)	Croissance absolue %	
1968	1	583		
1969	2	615	32	5.5
1970	3	646	31	5.0
1971	4	697	51	7.9
1972	5	738	41	5.9
1973	6	802	64	8.7
1974	7	844	42	5.2

Faites des prévisions pour le stocke espéré pour cinq et dix années à venir dans les quatre différentes manières:

- a. Supposons un accroissement absolu inchangé et constant par année, basé sur la croissance moyenne historique.
- b. Supposons un accroissement en pourcentage inchangé et constant par année, basé sur la croissance total des six années passées.
- c. Par adaptation d'une ligne de tendance linéaire aux données historiques.
- d. Par adaptation d'une tendance exponentielle aux données historiques.

Dessiner les données historiques dans un diagramme et montrer dans le même diagramme les deux lignes de tendance c) et d) étendues à l'année 16(6 + 10), et aussi les deux autres prévisions a) et b).

Comparer les résultats des prévisions.

Surveiller les lignes de tendance comme concernant la fiabilité statistiques (coefficient de corrélation, T - test et test de Durbin - Waston)

Discuter la crédibilité des prévisions.

2. Calculer tous les raccordements prévus pour l'année 1983 (31.3.83) dans un réseau téléphonique local

Données:

CROISSANCE DE DEMANDE DANS UN RESEAU TELEPHONIQUE LOCAL

Au	Capacité équipé	Lignes en service	Liste d'attente	Demande totale
31/3/72	11800	9825	1901	11,726
31/3/73	12000	11114	1781	12,925
31/3/74	12100	11458	3555	15,013
31/3/75	12500	11530	5106	16,636
30/9/75	12500	11653	5662	17,315
31/3/76	12800	12275	2587	14,862
31/3/77	14150	13683	1624	15,307
31/3/78	15500	14437	1893	16,330

3. Prévoir le trafic total de départ pour Décembre 1986 dans un central téléphonique local.

Données:

- a. Trafic mesuré pour jan. 1979- déc. 1981 (tableau A)
- b. Taux d'appels mesuré pour jan. 1979 - Déc. 1981 (tableau B)
- c. Prévision des raccordements pour décembre 1986 = 2969
(1170 abonnés en janvier 1979 et 1600 en décembre 1981)

A. TRAFIC DEPART TOTAL

MOIS	1979	1980	1981
JANVIER	38.6	39.4	45.6
FEVRIER	37.9	43.7	46.2
MARS	42.1	48.7	47.2
AVRIL	40.6	43.8	46.2
MAI	40.1	40.2	45.6
JUIN	38.1	42.6	48.5
JUILLET	37.7	41.1	44.4
AOUT	39.9	44.2	47.4
SEPTEMBRE	40.4	41.0	49.1
OCTOBRE	40.7	43.8	48.7
NOVEMBRE	40.8	41.8	45.0
DECEMBRE	42.2	49.5	49.5

B. TAUX D'APPELS TOTAL DE DEPART (Erlang/abonnes)

MOIS	1979	1980	1981
JANVIER	.033	.031	.033
FEVRIER	.033	.034	.033
MARS	.036	.038	.034
AVRIL	.035	.033	.033
MAI	.034	.030	.032
JUIN	.032	.032	.035
JUILLET	.031	.031	.032
AOUT	.033	.032	.034
SEPTEMBRE	.034	.030	.035
OCTOBRE	.034	.032	.032
NOVEMBRE	.033	.030	.028
DECEMBRE	.034	.036	.031

4. Considérons un réseau téléphonique avec deux centres.

Donnée:

a. L'intérêt de trafic actuel $[A_{ij}(0)]$

i \ j	1	2	$O_2(o)$
*	10	20	30
2	30	40	70
$T_j(o)$	40	60	100

b. Les valeurs prévues dans le future en matière de trafic de départ, et trafic d'arrivée par central $[O_i(t); T_j(t)]$

i \ j	1	2	$O_i(E)$
1	?	?	45
2	?	?	105
$T_j(t)$	50	100	150

Estimer les valeurs de trafic $A_{ij}^{(t)}$ par l'utilisation de la méthode de Kruithof.

5. Calculer l'intensité de trafic entre deux centres dans une zone locale.

Donnée:

- a. Une zone locale est divisée en zones de trafic No 1,2,3,4... Les futurs trafics entre toutes les zones de trafic sont prévues.
- b. La zone locale doit être divisée en zones de centraux. Les zones des centres ne doit pas coïncider avec les zones de trafic.
- c. Essayer de calculer le trafic espéré futur du centre A au centre B.
- d. On a les informations suivantes:

Le centre A devrait avoir

- 5000 lignes d'abonnés à partir de la zone de trafic I, qui a au total 10,000 lignes d'abonnés.

- 8000 lignes d'abonnés à partir de la zone de trafic II, qui a au total 12,000 lignes d'abonnés.

Le centre B devrait avoir

- 9000 lignes d'abonnés à partir de la zone de trafic III, qui est l'ensemble des lignes d'abonnés qui sont la bas.

- 2000 lignes d'abonnés à partir de la zone de trafic IV, qui a au total 6000 lignes d'abonnés.

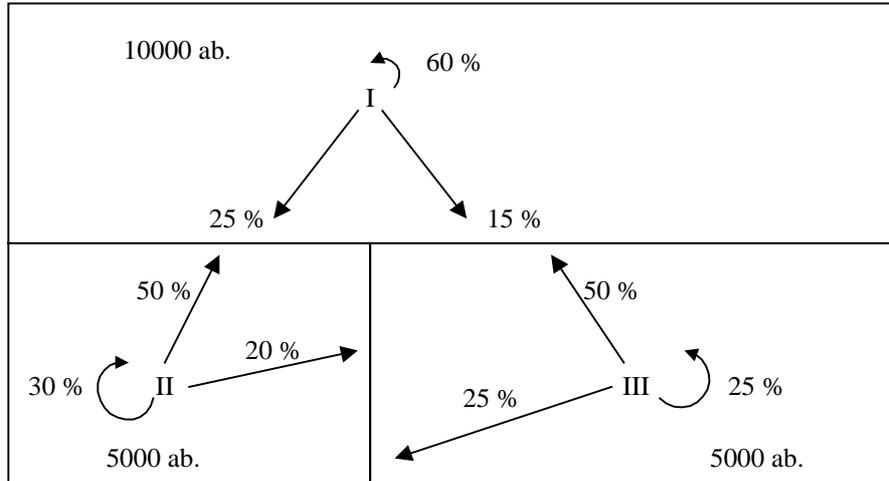
- e. On sait également à partir des prévisions, le trafic total espéré de chacun, vers chaque autre de la zone de trafic entière comme suit:

De la zone de trafic No.	Vers zone de trafic No.	Trafic total espéré
I	III	100 erl.
I	IV	90 erl.
II	III	105 erl.
II	IV	95 erl.

6. L'estimation de l'intérêt de trafic entre les centres : l'objectif de l'exemple est d'illustrer comment une matrice de trafic pour les centraux peut être calculée à partir d'une matrice de trafic donnée des zones de trafic.

Donnée:

Une zone constituée de trois zone de trafic: I, II, III:



Zone de trafic I

Nombre d'abonnés	10,000
Trafic total de départ/abonné	0.06 erl.
Distribution de ce trafic	60 % I 25 % II 15 % III

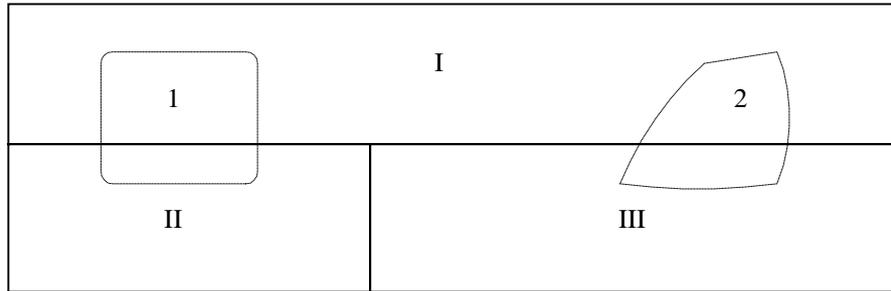
Zone de trafic II

Nombre d'abonnés	5,000
Trafic total de départ/abonné	0.05 erl.
Distribution de ce trafic	50 % I 30 % II 20 % III

Zone de trafic III

Nombre d'abonnés	5,000
Trafic total de départ/abonné	0.04 erl.
Distribution de ce trafic	50 % I 25 % II 25 % III

La zone est desservie par un nombre de centraux 1, 2, ... etc.



Centre 1

Nombre total d'abonnés	8,000
Nombre d'abonnés appartenant à la zone de trafic I	5,000
Nombre d'abonnés appartenant à la zone de trafic II	3,000

Centre 2

Nombre total d'abonnés	6,000
Nombre d'abonnés appartenant à la zone de trafic I	4,000
Nombre d'abonnés appartenant à la zone de trafic III	2,000

Tâche:

Calculer le flux de trafic total espéré du centre 1 vers le centre 2.

Conseil:

Commencer par le calcul du trafic *d'un* abonné dans la zone de trafic I vers *un* abonné dans la zone de trafic II, etc.

7. La matrice de trafic actuelle $A_{ij}^{(0)}$ a été estimée:

De	Vers			Somme
	Centre No. j			
	1	2	3	
1	25	30	45	100
2	35	55	110	200
3	60	85	155	300
Somme	120	170	310	600

Le nombre de lignes principales par central pour l'année t a été prévu:

Centre No.	$N_i(0)$	$N_i(t)$
1	2000	3000
2	3500	3500
3	6800	7500

Les lignes principales n'ont pas été classifiées par différentes catégories vue que la proposition des différents abonnés est espérée être la même dans le futur.

Le trafic total départ est arrivée par central est cependant prévue à partir des formules:

$$A_i^{(t)} = N_i^{(t)} \frac{A_i^{(0)}}{N_i^{(0)}}$$

$$A_j^{(t)} = N_j^{(t)} \frac{A_j^{(0)}}{N_j^{(0)}}$$

Centre No.	$A_i^{(t)}$	$A_j^{(t)}$
1	150.0	180.0
2	200.0	170.0
3	331.9	341.9
Somme	681.9	691.9

Puisque la somme des $A_i^{(t)}$ et $A_j^{(t)}$ diffère, on peut utiliser la valeur moyenne de ces sommes comme une estimations de $A_i^{(t)}$ et ajuster $A_i^{(t)}$ et $A_j^{(t)}$. Cela devrait donner:

Centre No.	$A_i^{(t)}$	$A_j^{(t)}$
1	151.1	178.7
2	201.5	168.8
3	334.3	339.4
Somme	686.9	686.9

- Maintenant tirer la matrice de trafic pour l'années future t et remplir les trafic totaux calculés ci-dessus.
- Calculer les différents trafics point-à-point par l'utilisations de la méthode de croissance pondérée, à partir des données connues relatives au nombre de lignes principales, actuels et futures, et le trafic point-à-point actuels. Le type des poids dépend de votre propre choix.
- Remplir ces valeur dans la nouvelle matrice de trafic, et calculer les sommes des colonnes et des lignes. Vous allez voir que ces sommes ne sont pas conforme avec les valeurs de votre première matrice.

- d. Si on considère que le trafic par ligne principale est constant durant la période des prévisions, alors le trafic point-à-point obtenu doit être corrigé, de façon à ce que les sommes soient conformes avec les trafic totaux prévus.

Si le temps le permet, faite cette correction , en utilisant la méthode de Kruithof a double facteur.

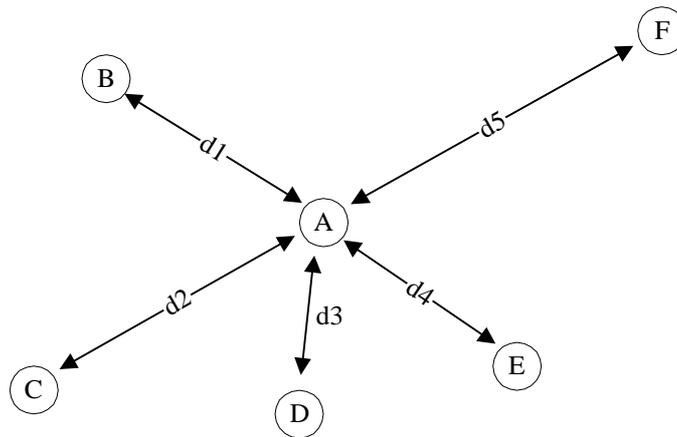
8. Supposons que le trafic interurbain prévu durant l'heur chargé pour le centre "A" est de 75 Erlgs, est distribué sur 5 routes vers les centres "B", "C", "D", "E" et "F", et pour les quelles aucune donnée de trafic n'est disponible supposons aussi que les raccordements prévues pour ces centraux sont:

A	10 000	(c ₀)
B	5 000	(c ₁)
C	7 000	(c ₂)
D	4 000	(c ₃)
E	2 000	(c ₄)
F	10 000	(c ₅)

Et que les distances à partir de "A" sont:

B	20 miles	(d ₁)
C	30 miles	(d ₂)
D	10 miles	(d ₃)
E	10 miles	(d ₄)
F	50 miles	(d ₅)

La situation est illustrée dans le diagramme ci-dessous.



Estimer la distribution interurbain total sur cinq différentes routes A vers B, A vers C, ..., etc.

Suggestion:

Le trafic prévu pour une route peut être obtenu à partir de

$$t_i = \frac{c_0 \cdot c_i}{d_i^2} \Bigg/ \sum_{j=1}^n \frac{c_0 \cdot c_j}{d_j^2}$$

où T est le trafic total.