

Contrat didactique du cours de statistique publique

Prérequis et articulation avec le cursus :

Connaissances élémentaires en analyse mathématique : notions de dérivation et d'intégration.
Connaissances élémentaires en statistique univariée et en calcul des probabilités : moyenne, variance, fonction de répartition, distributions usuelles

Compétences de formation développées

Les compétences sont à placer dans le contexte général de l'appropriation de la notion de modèle mathématique et de son utilisation pratique en gestion publique ce qui implique :

- L'appropriation du contenu des modèles statistiques de façon à les utiliser optimalement en gestion publique en toute connaissance de leur contenu et de leurs hypothèses implicites
- L'acquisition d'une certaine indépendance visant à chercher éventuellement d'autres modèles mieux appropriés

Objectifs de l'activité d'enseignement :

Le but est d'amener l'étudiant à poser un problème de gestion de bases de données publiques et à mettre en place différentes techniques de représentation de ces bases de données et également d'en calculer une paramétrisation interprétable et utilisable concrètement.

Il doit donc être capable de réaliser une analyse descriptive d'une série statistique bivariée, d'estimer ses paramètres et doit pouvoir les interpréter dans un cadre gestion publique.

Présentation du contenu

- Rappels en statistique élémentaire : représentation et paramétrisation de bases de données
- Analyse de séries statistiques bivariées
- Méthodes de régression linéaires (LSS), exponentielles et logarithmiques conduisant à l'estimation et à l'interprétation des paramètres d'un modèle
- Construction de composantes et interprétation de leur paramétrisation

Organisation du travail / attentes vis-à-vis des étudiants :

Cours magistral

Séances d'exercices

Séances d'exercices au labo d'informatique

Mise en place de modèles statistiques et de procédures de résolution de problèmes et de calculs utilisant un tableur

Modalités et critères d'évaluation :

Épreuve écrite comprenant trois types d'évaluation :

- Des questions théoriques évaluant le niveau de savoir approprié

- Des questions pratiques évaluant la compétence des étudiants à appliquer des techniques de calcul et à critiquer les ordres de grandeur des résultats obtenus
- Des questions d'interprétation portant sur des fichiers de bases de données, sur leur représentation, ou encore sur des fichiers de calcul

L'évaluation se fait sur base du contrat de confiance (EPCC) mis en place par le didacticien français André Antib

De séries de questions types pouvant être légèrement modifiées lors de l'examen sont proposées aux étudiants et reprennent 80% de l'examen

Support(s) de cours et bibliographie :

- Notes de cours fournies en support pdf
- La statistique par l'analyse, Daniel Justens, éditions du Céfal, 2004, 327 pages, ISBN 287130178-X
- Eléments de statistique - JJ Dreesbeke - Editions de l'Université de Bruxelles-2001
- Fichiers de bases de données en format xls
- Présentations powerpoint

Exemples de questions dans le cadre EPCC

Reprendre les séances de questions concernant régression, composantes et lois des grands nombres, déjà mises sur le site, ainsi que la première séance consacrée au théorème de Moivre

Interpréter la notion et le signe de la covariance.

Présenter succinctement la méthode dite « des moindres carrés ». Donner une application de cette méthode. Critiquer cette méthode

Mettre en équation le problème de régression linéaire simple avec une variable explicative et une constante. Détailler les calculs permettant l'obtention d'une équation introduisant tous les paramètres statistiques usuels. Interpréter le résultat obtenu. Déterminer la puissance explicative du modèle linéaire en termes de variance, en termes d'écarts-types.

Montrer comment on peut passer de la régression linéaire simple à la construction de composantes (dimension 2) dans le plan. Montrer pourquoi il est nécessaire, ici, de présenter deux modèles théoriques. Déterminer la puissance explicative de ce modèle linéaire en termes de variance, en termes d'écarts-types dans le cas d'un coefficient de corrélation positif (resp. négatif).

Le tableau qui suit donne l'évolution du taux de mortalité des Belges de 40 ans de sexe féminin et masculin depuis la fin du 19^e siècle (source: Institut national de Statistique)

années	femmes	hommes	
1885	0,00985	0,01082	
1930	0,00541	0,00649	
1948	0,00314	0,00536	
1960	0,00184	0,00285	
1970	0,001545	0,00282	
1980	0,001457	0,002588	
moyennes			
1945,5	0,003873667	0,00515467	
variances			
984,583333	9,00232E-06	8,5452E-06	
covariances			
	-0,092855167	-0,09035067	8,68086E-06

Représenter les données graphiquement et justifier si possible les hypothèses menant à la régression dans les 3 cas suivants :

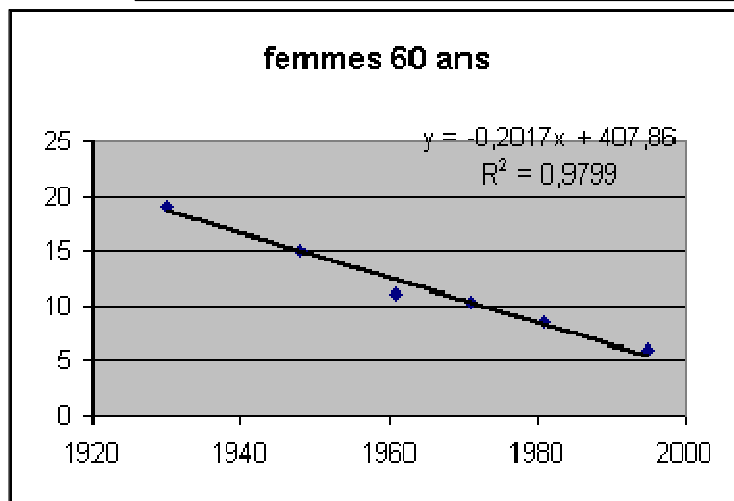
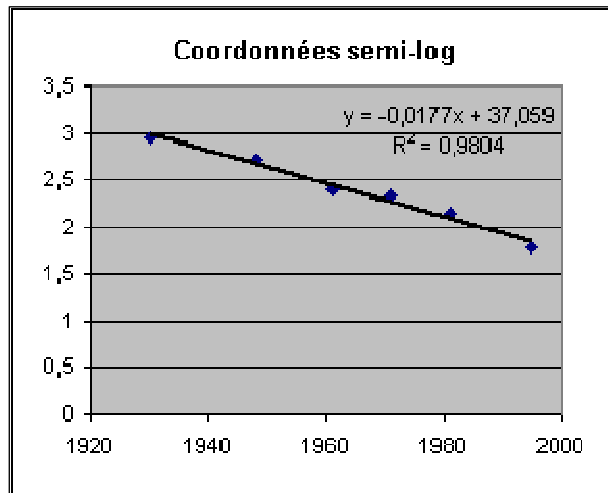
- Série chronologique des femmes pour les étudiantes
- Série chronologique des hommes pour les étudiants
- Série bivariée hommes-femmes pour tout le monde

Expliquer la notion de covariance et interpréter son signe. Dans les deux cas, calculer et interpréter intuitivement et quantitativement le coefficient de détermination.

Commenter et interpréter le fichier suivant :

Année	Taux	ln (taux)
1930	19,11	2,95021176
1948	15,12	2,71601837
1961	11,01	2,39880395
1971	10,303	2,33243511
1981	8,502	2,14030143
1995	5,899	1,77478285

moyenne x	1964,33333	
moyenne y	11,6573333	2,38542558
Var x	453,222222	
Var y	18,8166819	0,14403873
Cov	-91,4134444	-8,00005307



Détailler le raisonnement suivi et critiquer les deux modèles obtenus.
 Quel modèle retenez-vous et pourquoi ?

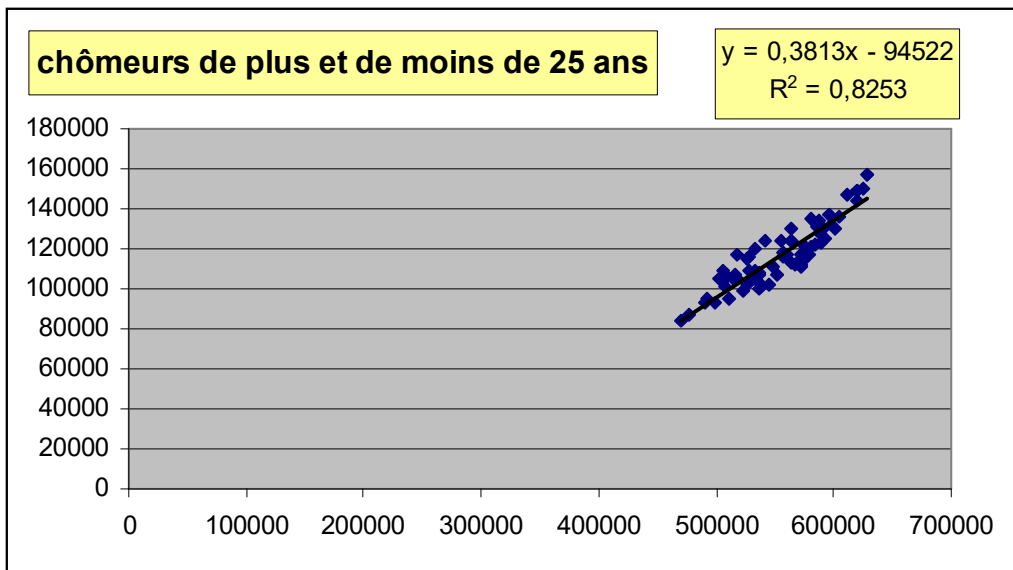
Le fichier qui suit donne la statistique du nombre de chômeurs de moins de 25 ans et de plus de 25 ans en Belgique entre 2005 et 2010 (données incomplètes sur l'image)

Chômeurs en Belgique		
Période	Plus de 25 ans	Moins de 25 ans
2005/01	590865	130322
2005/02	589379	126611
2005/03	584428	122401
2005/04	573100	117352
2005/05	567456	112135
2005/06	569028	112879
2005/07	620735	148729
2005/08	628884	157230
2005/09	628232	157221
2005/10	612117	146681
2005/11	595404	136524
2005/12	597141	132111
2006/01	593408	125326
2006/02	590072	122716
2006/03	579799	116615
2006/04	574205	113719
2006/05	572775	110737
2006/06	572942	112093
2006/07	619567	143565
2006/08	625809	149946
2006/09	619916	149147
2006/10	587625	134339
2006/11	565742	122683
2006/12	557274	115614
2007/01	560576	116566

Quelques paramètres statistiques de ces deux séries :

Moy-25	118004,3
Moy+25	557405,6
EcartType-25	16428,88
EcartType+25	39144,88

Et en voici une représentation graphique avec régression linéaire :



Pouvez-vous justifier la construction d'un modèle par régression linéaire ?

Interprétez le coefficient angulaire et le terme indépendant du modèle.

Interprétez qualitativement et quantitativement le coefficient de détermination

Pouvez-vous justifier la construction d'un modèle de type « composante » ?

Donnez l'équation de la composante principale

Interprétez ce modèle

Comparez les deux modèles et justifier les différences

Exercices supplémentaires relatifs au théorème de Moivre et au Théorème central limite

Le système de production que vous testez devrait donner un pourcentage de production de pièces correctes de l'ordre de 85%. Sur 500 pièces testées, vous constatez 88 pièces défectueuses. Testez l'optimalité du système de production et donnez un intervalle de confiance pour le taux de pièces correctes.

On estime la qualité du service population suffisante dès que le pourcentage d'erreurs est inférieur à 10%.

Énoncer le théorème de Moivre dans ce cadre précis

Tester la qualité de votre service si 32 erreurs ont été détectées sur 298 dossiers traités.

On estime la qualité du service production correcte lorsque plus de 90 % des pièces sortantes sont utilisables pendant plus d'un an.

Énoncer le théorème de Moivre. Tester la qualité de votre production si sur 500 pièces produites avant décembre 2009, 58 ont dû être remplacées à ce jour.

Un test de durée d'exécution des tâches est planifié dans votre service. Vous contrôlez ainsi l'exécution dans 375 cas pour obtenir les données observées suivantes :

Durée moyenne : 23 minutes

Écart-type des durées : 4 minutes

On demande :

- D'énoncer le théorème central limite
- De construire un intervalle de fiabilité 95 % des durées d'exécution. Quelle hypothèse doit-on faire ?
- De mesurer le niveau d'erreur sur la moyenne observée.
- De tirer des conclusions générales