

# Régression linéaire

Didier CHAUVÉAU, Jean-François DELMAS, Tony LELIÈVRE

October 10, 2017

## Contents

<b>1</b>	<b>Les données</b>	<b>2</b>
<b>2</b>	<b>Fonction régression</b>	<b>2</b>
2.1	Régression linéaire simple . . . . .	3
2.2	Régression linéaire multiple . . . . .	3

La séance de TP se fait sous environnement Linux. Pour commencer la séance, ouvrir un shell, créer un répertoire 'tp\_scilab' (`mkdir tp_scilab`) puis se placer sous ce répertoire (`cd tp_scilab`). Lancer un navigateur, par exemple Mozilla (`mozilla &`), et aller sur la page web suivante :

```
http://cermics.enpc.fr/scilab_new/site/Tp/Probabilites/tp_proba_reg
```

Lancer ensuite Scilab depuis ce répertoire 'tp\_scilab' (`scilab`) et ouvrir une fenêtre d'un éditeur, par exemple emacs (`emacs &`).

Pour chacune des questions, vous pouvez soit utiliser un "copier-coller", soit télécharger le programme correspondant.

Pour un rappel sur les opérations élémentaires de Scilab, on renvoie à l'*Introduction générale à Scilab pour les travaux pratiques à l'ENPC* :

Introduction à Scilab / Manipulations vectorielles

Introduction à Scilab / Graphiques, fonctions Scilab, programmation, saisie de données

On rappelle que pour avoir de l'aide sur une commande Scilab, il suffit de taper sur la ligne de commande Scilab :

```
help nom_de_la_commande;
```

# 1 Les données

On dispose de données concernant l'âge ( $r^1$ ), le kilométrage en milliers de kms ( $r^2$ ), et le prix en milliers d'euros ( $X$ ) pour un échantillon de voitures d'occasion d'un même type:

$r^1$	$r^2$	$X$
5	92	7.8
4	64	9.5
6	124	6.4
5	97	7.5
5	79	8.1
5	76	9.0
6	93	6.1
6	63	8.7
2	13	15.4
7	111	6.4
7	143	4.4

Si l'on veut étudier la liaison entre les variables potentiellement explicatives et la variable à expliquer (le prix  $X$ ), on peut commencer par visualiser les nuages de points ( $r^1, X$ ) et ( $r^2, X$ ):

Télécharger `tp_reg_donnees.sce`

**Question 1** *Que suggèrent les deux nuages?*

## 2 Fonction régression

La fonction `regression(M,X,alpha)` calcule, à partir de la donnée du modèle sous la forme  $X = M\gamma + \varepsilon$ , où  $X$  est la variable à expliquer et  $M$  la matrice définissant le modèle, les éléments usuels:

- L'estimateur du paramètre  $\gamma$ :  $\mathbf{gamma} = (M^t M)^{-1} M^t X$ .
- L'estimateur de  $\sigma^2$ .
- Pour le test de Fisher de l'hypothèse nulle "aucun régresseur n'est significatif", la table d'analyse de la variance et la valeur critique associée au niveau `alpha`.

Pour obtenir le code de la fonction, télécharger `tp_reg_regression.sce`, le sauvegarder sous le nom `tp_reg_regression.sce` dans le répertoire où vous utilisez scilab. Pour charger la fonction, utiliser la commande: `exec tp_reg_regression.sce`.

## 2.1 Régression linéaire simple

On propose tout d'abord le modèle simple:

$$X = \beta_0 \mathbf{1} + \beta_1 r^1 + \varepsilon,$$

où  $\varepsilon$  est un échantillon de  $\mathcal{N}_1(0, \sigma^2)$ .

**Question 2** *L'âge a-t-il une influence significative sur le prix ?*

Pour le modèle simple, où on explique le prix en fonction de l'âge, on utilise les commandes suivantes:

Télécharger `tp_reg_mod_simp.sce`

On représente la droite de régression sur le nuage:

Télécharger `tp_reg_trace_mod_simp_1.sce`

**Question 3** *Même question pour le kilométrage en considérant le modèle*

$$X = \beta_0 \mathbf{1} + \beta_1 r^1 + \beta_2 r^2 + \varepsilon,$$

Pour le tracé, télécharger `tp_reg_trace_mod_simp_2.sce`

*Parmi les deux sous-modèles à un seul régresseur, lequel explique le mieux le prix ?*

## 2.2 Régression linéaire multiple

On propose d'essayer à présent le modèle complet

$$X = \beta_0 \mathbf{1} + \beta_1 r^1 + \beta_2 r^2 + \varepsilon.$$

**Question 4** *Donner, à l'aide de la fonction `regression` des estimations sans biais des paramètres  $(\beta_0, \beta_1, \beta_2)$  et  $\sigma^2$ .*

On représente le plan de régression sur le nuage:

Télécharger `tp_reg_trace_mod_mult.sce`

**Question 5** *Test de l'utilité d'un régresseur: tester, au niveau 5%, l'hypothèse nulle "le kilométrage n'a pas d'effet sur le prix" contre "c'est faux".*

La fonction `regression` n'effectue pas le test demandé (quel test effectue-t-elle ?). Calculer la statistique et la  $p$ -valeur du test demandé, et conclure.