



Régression

Plan du cours

Prérequis

Présentation  
générale

Objectif

Intérêt

Historique

Exemple

Validation

Données

Visualisation

Vocabulaire

# Modèle Linéaire - Régression

Thierry Dhorne

Institut Universitaire de Technologie de Vannes  
Université de Bretagne Sud

Année Universitaire 2014-2015



# Plan du cours

Régression

Plan du cours

Prérequis

Présentation  
générale

Objectif

Intérêt

Historique

Exemple

Validation

Données

Visualisation

Vocabulaire

1. Régression simple
2. Régression à deux variables explicatives
3. Régression multiple
4. Analyse de variance



# Pré-requis

## Régression

Plan du cours

### Pré-requis

Présentation générale

Objectif

Intérêt

Historique

Exemple

Validation

Données

Visualisation

Vocabulaire

- il y a quelques pré-requis
  - ▶ probabilités (couple de variables aléatoires)
  - ▶ statistique exploratoire (corrélation linéaire)
  - ▶ bases de statistique inférentielle (estimation et tests)
    - ▶ test de Student
    - ▶ test de Fischer
  - ▶ bases d'algèbre linéaire (pour les aspects théoriques généraux)



Régression

Plan du cours

Prérequis

Présentation  
générale

Objectif

Intérêt

Historique

Exemple

Validation

Données

Visualisation

Vocabulaire

# Modèle Linéaire - Régression

## Régression linéaire simple

Thierry Dhorne

Institut Universitaire de Technologie de Vannes  
Université de Bretagne Sud

Année Universitaire 2014-2015



# Objectif de la régression

## Régression simple

Plan du cours

Prérequis

Présentation générale

**Objectif**

Intérêt

Historique

Exemple

Validation

Données

Visualisation

Vocabulaire

## Pourquoi la régression linéaire simple ?

La régression linéaire simple est conçue pour expliquer (et éventuellement pour prédire)

- ▶ une variable quantitative
- ▶ à partir d'une autre variable quantitative

- par exemple

- ▶ expliquer le poids d'un individu par sa taille
- ▶ expliquer le prix d'un appartement par sa surface
- ▶ prédire la durée de vie d'une route en fonction de l'épaisseur de macadam



# Intérêt (pratique)

Pour expliquer

Régression  
simple

Plan du cours

Prérequis

Présentation  
générale

Objectif

**Intérêt**

Historique

Exemple

Validation

Données

Visualisation

Vocabulaire

- l'intérêt d'expliquer une variable par une autre est important
  - ▶ pour trouver l'origine d'un phénomène inconnu (facteurs polluants)
  - ▶ pour valider un mécanisme suspecté (radioprotection)
  - ▶ pour étudier une relation connue (allongement d'un ressort en fonction du poids)



# Intérêt de la régression linéaire simple

Pour la prédiction

## Régression simple

Plan du cours

Prérequis

Présentation générale

Objectif

**Intérêt**

Historique

Exemple

Validation

Données

Visualisation

Vocabulaire

- l'intérêt pratique de la prédiction est évident
  - ➔ si la variable à prédire
    - ▶ est connue plus tard que la variable prédictrice
    - ▶ est plus difficile à mesurer que la variable prédictrice
    - ▶ est destructrice
    - ▶ ou coûte beaucoup plus cher que la variable prédictrice



# Historique

Régression  
simple

Plan du cours

Prérequis

Présentation  
générale

Objectif

Intérêt

**Historique**

Exemple

Validation

Données

Visualisation

Vocabulaire



F.R. Galton fût le premier à tracer une droite reliant la taille de descendants à celle de leurs parents. Ceci lui permit d'observer une régression de la taille vers la moyenne d'où le nom de droite de régression toujours utilisé aujourd'hui.

G.U. Yule a placé, conjointement à Pearson, la découverte de





# Un exemple détaillé

## Régression simple

Plan du cours

Prérequis

Présentation  
générale

Objectif

Intérêt

Historique

**Exemple**

Validation

Données

Visualisation

Vocabulaire

- un exemple d'école qui permet de bien comprendre
  - ▶ un animal
  - ▶ un aliment
- la moule
  - ▶ l'animal : la coquille et le corps
  - ▶ l'aliment : la chair



# Un exemple

## Régression simple

Plan du cours

Prérequis

Présentation générale

Objectif

Intérêt

Historique

**Exemple**

Validation

Données

Visualisation

Vocabulaire

- la variable d'intérêt est le poids de chair
  - ➔ elle est assez difficile à obtenir
- la variable économique est le poids total
  - ➔ elle est assez facile à obtenir
- ★ c'est d'ailleurs elle que l'on paye
- on peut penser que la variable d'intérêt dépend fortement de la variable accessible
- on peut espérer élaborer un modèle statistique pertinent



# Validation du contexte d'application

## Régression simple

Plan du cours

Prérequis

Présentation générale

Objectif

Intérêt

Historique

Exemple

**Validation**

Données

Visualisation

Vocabulaire

- Une seule variable à prédire
- ➔ sinon
  - ▶ compliqué (hors programme)
  - ▶ variable par variable
- Variable à prédire quantitative
- ➔ sinon changer de modèle
- Une seule variable prédictrice = régression simple
- ➔ sinon vu plus tard (régression multiple)
- Variable prédictrice quantitative
- ➔ sinon Analyse de la variance (vu plus tard)



# Validation opérationnelle

## Régression simple

Plan du cours

Prérequis

Présentation générale

Objectif

Intérêt

Historique

Exemple

**Validation**

Données

Visualisation

Vocabulaire

- Lien entre les 2 variables
- Lien linéaire
- ➔ à une augmentation d'une unité de  $x$  correspond une augmentation constante de  $y$
- ➔ ceci définit une droite (en réalité affine)
- On s'assure du type de lien graphiquement



# Données

## Exemple

### Régression simple

	pds.an	pds.ch		pds.an	pds.ch		pds.an	pds.ch	
Plan du cours	1	3.75	1.51	21	7.58	2.55	41	9.20	3.73
	2	7.05	2.60	22	5.24	1.12	42	7.02	2.04
Prérequis	3	5.83	1.33	23	6.33	1.92	43	6.79	2.35
Présentation générale	4	9.58	3.49	24	4.92	1.64	44	8.06	2.13
	5	1.31	0.36	25	3.04	0.60	45	5.21	1.95
Objectif	6	8.58	3.57	26	6.49	2.00	46	4.81	1.97
	7	7.16	2.98	27	4.51	1.45	47	7.25	2.53
Intérêt	8	4.80	1.74	28	7.03	2.62	48	6.82	2.36
	9	6.32	3.12	29	5.67	1.63	49	9.17	3.03
Historique	10	6.90	3.08	30	1.37	0.53	50	12.09	4.30
Exemple	11	3.95	1.28	31	4.84	0.92	51	9.79	3.21
	12	4.13	1.33	32	5.59	1.72	52	6.03	1.89
Validation	13	6.70	2.84	33	5.25	2.23	53	7.43	2.48
Données	14	2.37	0.68	34	8.26	2.79	54	7.24	1.55
Visualisation	15	9.24	2.03	35	7.02	2.71	55	5.72	1.91
	16	7.71	2.03	36	4.54	1.53	56	5.48	2.09
Vocabulaire	17	7.68	2.66	37	8.13	2.31	57	5.63	1.56
	18	5.18	2.24	38	7.02	2.75	58	10.12	3.59
	19	5.02	1.83	39	6.89	2.81	59	8.16	3.05
	20	0.71	0.17	40	5.96	1.90			

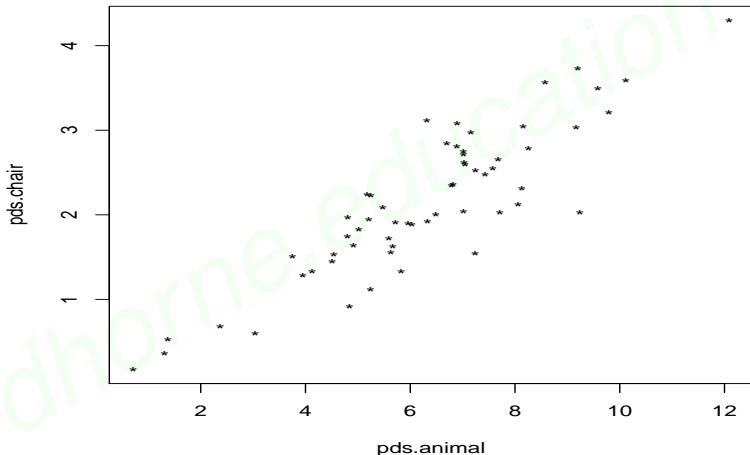


# Exemple

## Graphe croisé

Régression  
simple

- on peut réaliser un graphe des données pour visualiser le lien éventuel entre les variables





## Régression simple

Plan du cours

Prérequis

Présentation  
générale

Objectif

Intérêt

Historique

Exemple

Validation

Données

Visualisation

Vocabulaire

- le vocabulaire est relativement standardisé
  - ▶ variable à prédire et variable prédictrice
  - ▶ variable à expliquer et variable explicative
  - ▶ variable dépendante et variable indépendante
- ★ nous déconseillons ce dernier vocabulaire