

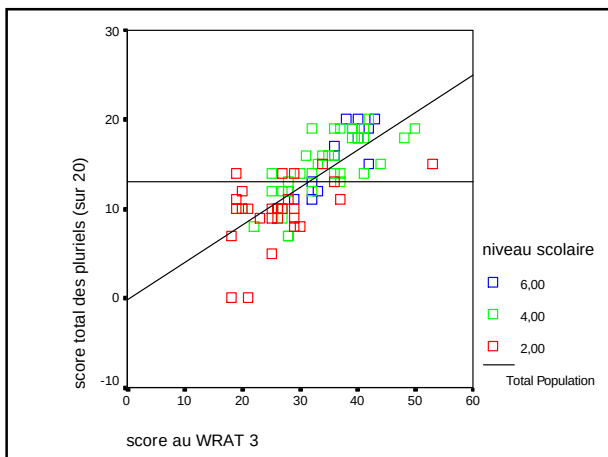
Objectifs du chapitre sur la régression

- ❖ Comprendre ce qu'est une régression simple et multiple
- ❖ Savoir calculer la RPE, l'ordonnée et la pente de régressions simple et double
- ❖ Savoir interpréter la RPE, l'ordonnée et la pente de régression simple et multiple
- ❖ Comprendre les conditions d'application de la régression

Une autre façon de représenter une association: la droite de régression

☐ Dans les graphiques de dispersion:

- ◆ un nuage de points
- ◆ à travers duquel peut passer
- ◆ cette droite passe par
- ◆ soit une autre façon d'exprimer
- ◆



La droite de régression (1)

- est une relation univoque:
contrairement à la corrélation
 $Y = f(X) = a + b(X)$

où X est conçu comme causant Y

- s'exprime par 2 paramètres:

ordonnée $a = \bar{Y} - b\bar{X}$

pende

$$b_{yx} = \frac{\text{Co var}_{xy}}{s_x^2} = \frac{\sum XY - \frac{\sum X \sum Y}{N}}{\sum X^2 - \frac{(\sum X)^2}{N}}$$

La droite de régression (2)

- permet de calculer des valeurs prédites

$$\hat{Y} = a + b(X)$$

(intra-

et extra-polation **danger!**)

Quand il y a une seule variable prédictrice (un seul X)

- la RPE (Réduction Proportionnelle de l'Erreur de prédiction) se détermine

soit par la corrélation
(au carré)

$$RPE = (r_{xy})^2$$

soit par la droite de régression
(conceptuellement)

$$\hat{Y} = a + b(X)$$

$$\frac{\sum (Y_i - \hat{Y})^2}{N-1}$$

$$RPE = \frac{SC_y - SC_{\text{résid.}}}{SC_y} = 1 - \frac{SC_{\text{résid.}}}{SC_y}$$

Alternative au calcul de la RPE

□ la RPE (Réduction Proportionnelle de l'Erreur de prédiction) se détermine

♦ soit par la droite de régression

$$RPE = \frac{SC_y - SC_{\text{résid.}}}{SC_y}$$

or $SC_y = SC_{\text{rég.}} + SC_{\text{résid.}}$ $\Rightarrow RPE = \frac{SC_{\text{rég.}}}{SC_y}$

$$SC_{\text{rég.}} = b \times \text{Covar}$$

Données du problème 9.1

% naiss. à risque	fertilité	% naiss. non mariées
6,1	43,0	9,2
7,1	55,3	12,0
7,4	48,5	10,4
6,3	38,8	9,8
6,5	46,2	9,8
5,7	39,9	7,7
6,6	43,1	10,9
8,1	48,5	9,5
6,3	40,0	11,6
6,9	56,7	11,6
Somme		

Les produits croisés

YX ₁	YX ₂	X ₁ X ₂
262,30	56,12	395,60
392,63	85,20	663,60
358,90	76,96	504,40
244,44	61,74	380,24
300,30	63,70	452,76
227,43	43,89	307,23
284,46	71,94	469,79
392,85	76,95	460,75
252,00	73,08	464,00
391,23	80,04	657,72
Somme		

Les carrés des données

Y^2	X_1^2	X_2^2
37,21	1849,00	84,64
50,41	3058,09	144,00
54,76	2352,25	108,16
39,69	1505,44	96,04
42,25	2134,44	96,04
32,49	1592,01	59,29
43,56	1857,61	118,81
65,61	2352,25	90,25
39,69	1600,00	134,56
47,61	3214,89	134,56

Calcul du problème 9.1-9.3 (1)

Y	X_1	YX_1	Y^2	X_1^2
67	460	3106,54	453,28	21516

$$\text{Co var}_{x1y} = \left(\sum X_1 Y - \frac{(\sum X_1 \sum Y)}{N} \right) =$$

$$S_{x1}^2 = \left(\sum X_1^2 - \frac{(\sum X_1)^2}{N} \right) =$$

$$\mathbf{b}_{yx1} = \frac{\sum_{j=1}^n y_j x_{1j}}{\sum_{j=1}^n x_{1j}^2}$$

$$\mathbf{a} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} = \begin{pmatrix} 1 \\ 2 \\ 3 \end{pmatrix} \cdot \begin{pmatrix} 1 \\ 0 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 2 \\ 0 \end{pmatrix} + \begin{pmatrix} 0 \\ 0 \\ 3 \end{pmatrix}$$

[illegible]

Calcul alternatif de la RPE

 $RPE = \frac{SC_{\text{rég.}}}{SC_y}$

Y	X_1	YX_1	Y^2	X_1^2
67	460	3106,54	453,28	21516

$$SC_y = \left(\sum Y^2 - \frac{(\sum Y)^2}{N} \right) = 453,28 - \frac{67^2}{10} =$$

$$\mathbf{SC}_{\text{rég.}} = \mathbf{b} \times \mathbf{Covar}$$

Lien entre la régression simple et la corrélation bivariée pour le calcul

- pente

$$b_{yx} = \frac{\text{Cov}_{xy}}{s_x^2} = \frac{r_{xy} s_y}{s_x}$$

- pente standardisée

$$\beta_{yx} = r_{xy} = \frac{b_{yx} s_x}{s_y}$$

La droite de régression définit 2 sources de variation de Y

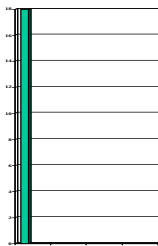
- la variance totale originale de Y:

$$\frac{\sum (Y_i - \bar{Y})^2}{N-1}$$

- la variance résiduelle de Y suite à sa prédiction par X:

$$\frac{\sum (Y_i - \hat{Y})^2}{N-1}$$

Pour chaque valeur de Y par rapport à l'axe des X

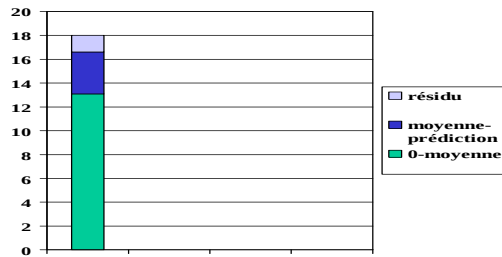


Soit un Y
observé de 18
(X = 40)

Sa moyenne
est 13,1

Sa valeur prédite est 16,6
selon l'équation de régression
Pluriel = -,201 + .420 Lecture

Chaque Y se décompose en 3 morceaux



Le calcul de régression ...

- ☐ détermine une droite qui
 - ◆ soit est identique à la droite de la moyenne $\bar{Y}$
 - ◆ soit en dévie de façon systématique
- ☐ ceci fait que plus $\frac{\sum (Y_i - \hat{Y})^2}{\sum (Y_i - \bar{Y})^2}$ sera petit par rapport à $\sum (Y_i - \bar{Y})^2$ meilleure sera la prédiction

Interprétation de la régression simple

- ☐ **RPE**
% d'explication ou de prédiction empirique de la variable dépendante Y
- ☐ Les éléments de la droite de régression
 - ◆ **a: ordonnée**
la valeur de Y quand X = 0
si cela signifie quelque chose
 - ◆ **b: pente**
quantité de changement de Y par unité de changement de X,
Note: quand il s'agit de β , l'unité de changement est l'écart-type

De la régression simple à la régression multiple...

- La RPE sera toujours égale à

$$RPE = \frac{SC_y - SC_{\text{résid.}}}{SC_y} = 1 - \frac{SC_{\text{résid.}}}{SC_y} = \frac{SC_{\text{rég.}}}{SC_y}$$

- La droite de régression (multiple)

$$a = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2 \dots - b_k \bar{X}_k$$

$$b_{yX_1X_2} = \frac{\text{Covar}_{X_1Y} s_{X_2}^2 - \text{Covar}_{X_1Y} \text{Covar}_{X_1X_2}}{s_{X_1}^2 s_{X_2}^2 - (\text{Covar}_{X_1X_2})^2}$$

Étapes du calcul de la régression double (1)

- Les formules

$$b_{yX_1X_2} = \frac{\text{Co var}_{X_1Y} s_{X_2}^2 - \text{Co var}_{X_2Y} \text{Co var}_{X_1X_2}}{s_{X_1}^2 s_{X_2}^2 - (\text{Co var}_{X_1X_2})^2}$$

$$b_{yX_2X_1} = \frac{\text{Co var}_{X_2Y} s_{X_1}^2 - \text{Co var}_{X_1Y} \text{Co var}_{X_1X_2}}{s_{X_1}^2 s_{X_2}^2 - (\text{Co var}_{X_1X_2})^2}$$

$$a = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2$$

Étapes du calcul de la régression double (2)

- Application de la formule de la pente

$$b_{yX_1X_2} = \frac{\left(\frac{\sum X_1 Y}{N} - \frac{(\sum X_1)(\sum Y)}{N^2} \right) \left(\frac{\sum X_2^2}{N} - \frac{(\sum X_2)^2}{N^2} \right) - \left(\frac{\sum X_2 Y}{N} - \frac{(\sum X_2)(\sum Y)}{N^2} \right) \left(\frac{\sum X_1 X_2}{N} - \frac{(\sum X_1)(\sum X_2)}{N^2} \right)}{\left(\frac{\sum X_1^2}{N} - \frac{(\sum X_1)^2}{N^2} \right) \left(\frac{\sum X_2^2}{N} - \frac{(\sum X_2)^2}{N^2} \right) - \left(\frac{\sum X_1 X_2}{N} - \frac{(\sum X_1)(\sum X_2)}{N^2} \right)^2}$$

Étapes du calcul de la régression double (3)

□ Application de la formule de la pente en 5 morceaux

$$\begin{aligned} \text{Co var}_{x_1y} &= \left(\sum X_1 Y - \frac{(\sum X_1 \sum Y)}{N} \right) \quad 1 \\ s_{x_2}^2 &= \left(\sum X_2^2 - \frac{(\sum X_2)^2}{N} \right) \quad 2 \quad s_{x_1}^2 = \left(\sum X_1^2 - \frac{(\sum X_1)^2}{N} \right) \quad 3 \\ \text{Co var}_{x_2y} &= \left(\sum X_2 Y - \frac{(\sum X_2 \sum Y)}{N} \right) \quad 4 \\ \text{Co var}_{x_1x_2} &= \left(\sum X_1 X_2 - \frac{(\sum X_1 \sum X_2)}{N} \right) \quad 5 \end{aligned}$$

Étapes du calcul de la régression double (4)

□ La formule pour la pente de X_2 et Y

$$b_{y \cdot x_2 \cdot x_1} = \frac{\text{Co var}_{x_2y} s_{x_1}^2 - \text{Co var}_{x_1y} \text{Co var}_{x_1x_2}}{s_{x_1}^2 s_{x_2}^2 - (\text{Co var}_{x_1x_2})^2}$$

on ne fait qu'inverser la position de 1 et 2, mais ce sont les mêmes termes

Données

Chômage	Année	Prod. Industr.
3,1	1950	113
1,9	1951	123
1,7	1952	127
1,6	1953	138
3,2	1954	130
2,7	1955	146
2,6	1956	151
2,9	1957	152
4,7	1958	141
3,8	1959	159
Somme	28,2	1380

Les carrés des données

Chômage	Année	Prod. Industr.
9,61	0	12769
3,61	1	15129
2,89	4	16129
2,56	9	19044
10,24	16	16900
7,29	25	21316
6,76	36	22801
8,41	49	23104
22,09	64	19881
14,44	81	25281
Somme		

Les produits croisés des données

AxC	PlxC	AxPI
0	350,3	0
1,9	233,7	123
3,4	215,9	254
4,8	220,8	414
12,8	416,0	520
13,5	394,2	730
15,6	392,6	906
20,3	440,8	1064
37,6	662,7	1128
34,2	604,2	1431
Somme		

ΣY	ΣX_1	ΣX_2	ΣY^2	ΣX_1^2	ΣX_2^2	ΣYX_1	ΣYX_2	ΣX_1X_2
28,2	19545	1380	87,9	285	192354	144,1	3931,2	6570

Calcul de l'exemple (1)

$$\text{Co var}_{x_1y} = \left(\Sigma X_1 Y - \frac{(\Sigma X_1 \Sigma Y)}{N} \right) =$$

$$s_{x_2}^2 = \left(\Sigma X_2^2 - \frac{(\Sigma X_2)^2}{N} \right) =$$

$$s_{x_1}^2 = \left(\Sigma X_1^2 - \frac{(\Sigma X_1)^2}{N} \right) =$$

$$\text{Co var}_{x_2y} = \left(\Sigma X_2 Y - \frac{(\Sigma X_2 \Sigma Y)}{N} \right) =$$

$$\text{Co var}_{x_1x_2} = \left(\Sigma X_1 X_2 - \frac{(\Sigma X_1 \Sigma X_2)}{N} \right) =$$

Covar _{x1Y}	Var _{x1}	Var _{x2}	Covar _{x2Y}	Covar _{x1x2}
17,2	82,5	1914	39,6	360

☐ Appliquons les calculs dans la formule

$$b_{yx1.x2} = \frac{\text{Co var}_{x1y} s_{x2}^2 - \text{Co var}_{x2y} \text{Co var}_{x1x2}}{s_{x1}^2 s_{x2}^2 - (\text{Co var}_{x1x2})^2} =$$

Covar _{x1Y}	Var _{x1}	Var _{x2}	Covar _{x2Y}	Covar _{x1x2}
17,2	82,5	1914	39,6	360

☐ Oui, mais la pente de X₂ et Y?

$$b_{yx2.x1} = \frac{\text{Co var}_{x2y} s_{x1}^2 - \text{Co var}_{x1y} \text{Co var}_{x1x2}}{s_{x2}^2 s_{x1}^2 - (\text{Co var}_{x1x2})^2} =$$

$$b_{yx2.x1} =$$

on ne fait qu'inverser la position de 1 et 2,
mais ce sont les mêmes termes

Calcul de l'exemple (5)

☐ Le calcul final

$$b_{yx1.x2} = \frac{\text{Co var}_{x1y} s_{x2}^2 - \text{Co var}_{x2y} \text{Co var}_{x1x2}}{s_{x1}^2 s_{x2}^2 - (\text{Co var}_{x1x2})^2} =$$

$$b_{yx1.x2} =$$

$$b_{yx2.x1} = \frac{\text{Co var}_{x2y} s_{x1}^2 - \text{Co var}_{x1y} \text{Co var}_{x1x2}}{s_{x1}^2 s_{x2}^2 - (\text{Co var}_{x1x2})^2} =$$

$$b_{yx2.x1} =$$

$$a = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2 =$$

Calcul de l'exemple (6)

□ L'équation de prédiction

$$\hat{Y} = a + b_1(X_1) + b_2(X_2) = \text{taux..de..ch\^omage..}$$

$$= 13,65 + 0,66 \cdot \text{ann\^ee} - 0,10 \cdot \text{production..industrielle}$$

Covar _{X1Y}	Var _{X1}	Var _{X2}	Covar _{X2Y}	Covar _{X1X2}
17,2	82,5	1914	39,6	360

$$SC_{\text{regr. } Y \rightarrow X1X2} = (b_{YX1X2} \times \text{Covar}_{YX1}) + (b_{YX2X1} \times \text{Covar}_{YX2}) =$$

Données des problèmes 9.1-9.3

% naiss. à risque	fertilité	% naiss. non mariées
6,1	43,0	9,2
7,1	55,3	12,0
7,4	48,5	10,4
6,3	38,8	9,8
6,5	46,2	9,8
5,7	39,9	7,7
6,6	43,1	10,9
8,1	48,5	9,5
6,3	40,0	11,6
6,9	56,7	11,6
Somme		

Les produits croisés

YX ₁	YX ₂	X ₁ X ₂
262,30	56,12	395,60
392,63	85,20	663,60
358,90	76,96	504,40
244,44	61,74	380,24
300,30	63,70	452,76
227,43	43,89	307,23
284,46	71,94	469,79
392,85	76,95	460,75
252,00	73,08	464,00
391,23	80,04	657,72
Somme		

Les carrés des données

Y^2	X_1^2	X_2^2
37,21	1849,00	84,64
50,41	3058,09	144,00
54,76	2352,25	108,16
39,69	1505,44	96,04
42,25	2134,44	96,04
32,49	1592,01	59,29
43,56	1857,61	118,81
65,61	2352,25	90,25
39,69	1600,00	134,56
47,61	3214,89	134,56
Somme		

Somme des	% naiss. à risque (Y)	Fertilité (X ₁)	% naiss. non mariées (X ₂)
valeurs	67	460	102,5
carrés	453,28	21515,98	1066,35
produits croisés	3106,54 (YX ₁)	689,62 (YX ₂)	4756,09 (X ₁ X ₂)

	(Y)	(X ₁)	(X ₂)
valeurs	67	460	102,5
carrés	453,28	21515,98	1066,35
prod. croisés	3106,54 (YX ₁)	689,62 (YX ₂)	4756,09 (X ₁ X ₂)

$$\text{Co var}_{x_1y} = \left(\sum X_1 Y - \frac{(\sum X_1 \sum Y)}{N} \right) =$$

$$s_{x_2}^2 = \left(\sum X_2^2 - \frac{(\sum X_2)^2}{N} \right) =$$

$$s_{x_1}^2 = \left(\sum X_1^2 - \frac{(\sum X_1)^2}{N} \right) =$$

$$\text{Co var}_{x_2y} = \left(\sum X_2 Y - \frac{(\sum X_2 \sum Y)}{N} \right) =$$

$$\text{Co var}_{x_1x_2} = \left(\sum X_1 X_2 - \frac{(\sum X_1 \sum X_2)}{N} \right) =$$

Calcul du problème 9.1-9.3 (2)

Co var_{x1y} =

s_{x2}^2 =

s_{x1}^2 =

Co var_{x2y} =

Co var_{x1x2} =

Calcul du problème 9.1-9.3 (3)

Covar _{x1y}	Var _{x1}	Var _{x2}	Covar _{x2y}	Covar _{x1x2}
24,54	355,98	15,725	2,87	41,09

☐ Appliquons les calculs dans la formule

$$b_{yx1x2} = \frac{\text{Covar}_{x1y} s_{x2}^2 - \text{Covar}_{x2y} \text{Covar}_{x1x2}}{s_{x1}^2 s_{x2}^2 - (\text{Covar}_{x1x2})^2} =$$

$b_{yx1x2} =$

Calcul du problème 9.1-9.3 (4)

Covar _{x1y}	Var _{x1}	Var _{x2}	Covar _{x2y}	Covar _{x1x2}
24,54	355,98	15,725	2,87	41,09

☐ Oui, mais la pente de X_2 et Y?

$$b_{yx2x1} = \frac{\text{Co var}_{x2y} s_{x1}^2 - \text{Co var}_{x1y} \text{Co var}_{x1x2}}{s_{x2}^2 s_{x1}^2 - (\text{Co var}_{x1x2})^2} =$$

on ne fait qu'inverser la position de 1 et 2,
mais ce sont les mêmes termes

Calcul du problème 9.1-9.3 (5)

☐ Le calcul final

$$b_{yx1.x2} = \frac{\text{Co var}_{x1y} s_{x2}^2 - \text{Co var}_{x2y} \text{Co var}_{x1x2}}{s_{x1}^2 s_{x2}^2 - (\text{Co var}_{x1x2})^2} =$$

$$b_{yx1.x2} =$$

$$b_{yx2.x1} = \frac{\text{Co var}_{x2y} s_{x1}^2 - \text{Co var}_{x1y} \text{Co var}_{x1x2}}{s_{x1}^2 s_{x2}^2 - (\text{Co var}_{x1x2})^2} =$$

$$b_{yx2.x1} =$$

$$a = \bar{Y} - b_1 \bar{X}_1 - b_2 \bar{X}_2 =$$

$$a =$$

Calcul du problème 9.1-9.3 (6)

☐ L'équation de prédiction

$$\hat{Y} = a + b_1(X_1) + b_2(X_2) = \text{\%naissances.à.risque}$$

$$= 3,5 + 0,07 \text{taux.de.fertilité} + 0,003 \text{\%femmes..non.mariées}$$

Calcul du problème 9.1-9.3 (7* nouvelle)

☐ La RPE par la voie rapide

$$SC_{\text{regr. } Y=X1X2} = (b_{YX1X2} \times \text{Covar}_{YX1}) + (b_{YX2X1} \times \text{Covar}_{YX2}) =$$

$$RPE_{Y=X1X2} = \frac{SC_{\text{reg}}}{SC_Y} =$$

Vérification du calcul du problème 9.1-9.3

Y	X ₁	X ₂	Y' = a + b ₁ X ₁ + b ₂ X ₂	Y'	Y - Y'	(Y - Y') ²
6,1	43,0	9,2				
7,1	55,3	12,0				
7,4	48,5	10,4				
6,3	38,8	9,8				
6,5	46,2	9,8				
5,7	39,9	7,7				
6,6	43,1	10,9				
8,1	48,5	9,5				
6,3	40,0	11,6				
6,9	56,7	11,6				

Ceci permet de calculer la RPE

□ La RPE égale

$$RPE = \frac{SC_y - SC_{résid.}}{SC_y} = 1 - \frac{SC_{résid.}}{SC_y}$$

$$SC_y = \sum (Y - \bar{Y})^2 = \sum Y^2 - \frac{(\sum Y)^2}{N} = 453,28 - \frac{(67)^2}{10} = 453,28 - \frac{4489}{10} = 4,38$$

$$RPE = \frac{SC_y - SC_{résid.}}{SC_y} = 1 - \frac{SC_{résid.}}{SC_y} = 1 - \frac{2,69}{4,38} = 0,386$$

Comment savoir si RPE égale 0 ou non?

□ rappel test t pour la corrélation:

$$t_{(dl)} = \frac{r_{xy} \sqrt{N-2}}{\sqrt{1-r_{xy}^2}}$$

avec N-2 comme ddl

□ pour la RPE, test F avec k-1 et N-k comme ddl et où k = ordonnée + pentes

$$F_{(ddl,ddl)} = \frac{RPE(N-k)}{(k-1)(1-RPE)}$$

Calcul des sommes de carrés des pentes

$$SC_{\text{totale}} = SC_{\text{régression}} + SC_{\text{résiduelle}}$$

$$SC_{\text{régression}} = \sum SC_{b_i}$$

$$SC_{b_{y_{xi}.x_i + 1}} = b_{y_{xi}.x_i + 1} \times \text{Covar}_{x1y}$$

$$RPE = \frac{SC_y - SC_{\text{résid.}}}{SC_y} = 1 - \frac{SC_{\text{résid.}}}{SC_y} = \frac{SC_{\text{régression}}}{SC_y}$$

Calcul des sommes de carrés des pentes

$$SC_{b_{y_{xi}.x_i + 1}} = b_{y_{xi}.x_i + 1} \times \text{Covar}_{x1y}$$

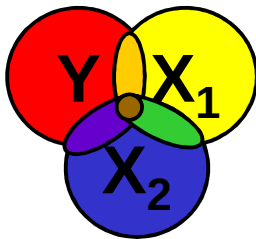
$$SC_{\text{régression}} = \sum SC_{b_i}$$

$$RPE = 1 - \frac{SC_{\text{résid.}}}{SC_y} = \frac{SC_y - SC_{\text{résid.}}}{SC_y}$$

Interprétation de la régression multiple

- ☐ **RPE: même chose**
% d'explication ou de prédiction empirique de la variable dépendante Y
- ☐ **Les éléments de la droite de régression**
 - ◆ **a: ordonnée**
la valeur de Y quand tous les $X_i = 0$
si cela signifie quelque chose
 - ◆ **b: pente**
quantité de changement de Y par unité de changement de X_i **b** $_{yx\ 1..x\ 2}$
lorsque les autres $X_i = 0$

Interprétation graphique de la régression multiple (diagrammes de Venn)



Différentes approches de régression multiple

☐ Régression hiérarchique:

- ◆ comparaison de modèles inclusifs

☐ Régression en étape (séquentielle):

- ◆ selon les patrons de (cor)relations entre les variables

◆ plusieurs approches pas à pas:

- élimination (backward)
- sélection (forward)
- stepwise

Vrai ou Faux?

$$\begin{array}{r} 3 \\ +2 \\ \hline 6 \end{array}$$

Table 2
Summary of Regression Analyses for Type I Addition Problems

Equation	R^2	df	F	RMS
True, nontie problems				
$RT = 926 + 7.91 (\text{PROD})$	.671	1, 88	179.28	106.1
$RT = 879 + 2.09 (\text{SUM})$	.657	1, 88	168.21	108.3
$RT = 908 + 64.1 (\text{MIN})$	.601	1, 88	132.53	116.8
False, nontie problems				
$RT = 1057 + 10.53 (\text{PROD})$	.640	1, 88	156.70	151.1
$RT = 998 + 2.74 (\text{SUM})$	.609	1, 88	146.94	157.6
$RT = 1027 + 87.3 (\text{MIN})$	.600	1, 88	132.04	159.4
Combined set, nontie problems				
$RT = 926 + 7.91 (\text{PROD}) + 131 (\text{TRUTH})$ + 2.62 ($\text{PROD} \times \text{TRUTH}$) Partial F s: 118.36, 22.17, 6.52	.702	3, 176	138.50	130.6
$RT = 682 + 80 (\text{ENCODE}) + 6.25 (\text{PROD})$ + 133 (TRUTH) + 2.72 ($\text{PROD} \times \text{TRUTH}$) Partial F s: 6.89, 43.01, 23.54, 7.20	.714	4, 175	109.08	128.4
$RT = 712 + 66 (\text{ENCODE}) + 7.26 (\text{PROD})$ + 93 (TRUTH) + 66 (REENC) Partial F s: 13.42, 97.32, 8.99, 13.42	.713	3, 176	145.96	128.2
Combined set, tie problems				
$RT = 971 + 121 (\text{TRUTH})$	.240	1, 18	5.68	113.3
Combined set, tie and nontie problems				
$RT = 726 + 68 (\text{ENCODE}) + 5.32 (\text{PROD})$ + 81 (TRUTH) + 68 (REENC)	.561	3, 196	83.81	154.8
$RT = 760 + 50 (\text{ENCODE}) + 7.72 (\text{PROD})$ + 106 (TRUTH) + 50 (REENC)	.702	3, 196	153.88	127.7

Note. Mean $RT = 1,169.40$ ms. PROD stands for product of addends; PROD^2 is same as PROD except that values for tie problems are set to zero; $\text{SUM}^2 = \text{squared sum of addends}$; $\text{MIN} = \text{smaller addend}$; $\text{TRUTH} = \text{correct (0) or incorrect (1) stated sum}$; $\text{ENCODE} = \text{number of items encoded (3 or 4)}$; and $\text{REENC} = \text{reencoding of digits in incorrect stated sums (1 or 2)}$.

Table 3
Summary of Regression Analyses for Type II Addition Problems

Equation	R^2	df	F	RMS
$RT = 776 + 108 (\text{ENCODE}) + 11.60 (\text{UNITPROD})$ + 150 (CARRY) + 11.60 (TENPROD) + 111 (TRUTH) + 108 (REENC) Partial F s: 38.03, 142.43, 9.61, 142.43, 13.68, 38.03	.743	4, 195	140.93	200.0
$RT = 724 + 118 (\text{ENCODE}) + 11.66 (\text{UNITPROD})$ + 147 (CARRY) + 11.66 (TENPROD) + 236 (TRUTH) Partial F s: 34.66, 140.62, 8.93, 140.62, 49.30	.740	4, 195	138.19	201.5
$RT = 749 + 104 (\text{ENCODE}) + 96.3 (\text{UNITMIN})$ + 142 (CARRY) + 96.3 (TENMIN) + 117 (TRUTH) + 104 (REENC) Partial F s: 35.31, 141.77, 8.38, 141.77, 15.30, 35.31	.741	4, 195	140.56	200.6

Note. Mean $RT = 1,682.96$ ms. ENCODE stands for number of items (digits) encoded (either 3, 5, or 6 in this set of problems); UNITMIN and TENMIN for minimum addend in units and tens columns, respectively; UNITPROD and TENPROD for product of addends in units and tens columns, respectively; CARRY for presence (1) or absence (0) of a carry from the units to tens column; TRUTH for correct (0) or incorrect (1) stated sum; and REENC for the reencoding of digits in incorrect stated sums (1 or 2).

Do Phonological Tests Predict Reading and Spelling?

Dependent variable	Outcome measure (fixed order multiple regressions: 6 years 7 months)			
	Schonell spelling	Schonell reading	France reading	Arithmetic
Steps 1-4 (R^2 change)				
Age	.04	.04	.00	—
Mother's educational level	.42**	.38**	.42***	.37**
BPVS	.00	.01	.02	.00
IQ	.14***	.17***	.13***	.06*
Step 5 (R^2 change)				
Rhyme oddity (4 years 7 months)	.05**	.09***	.08***	.03
Alliteration oddity (4 years 7 months)	.11***	.17***	.15***	.05*
Joint rhyme/alliteration choice (5 years 7 months)	.07**	.09***	.10***	.02
Phoneme deletion, first sound (5 years 11 months)	.05**	.08***	.07**	.03
Phoneme deletion, end sound (5 years 11 months)	.01	.03*	.03*	.01
Phoneme tapping (5 years 11 months)	.05**	.04*	.03*	.08**

Note. BPVS = British Picture Vocabulary Scale.

* $p < .05$. ** $p < .01$. *** $p < .001$.

TABLE 11.4 CORRELATION MATRIX OF VARIABLES USED IN THE STEPWISE REGRESSION ANALYSIS*

	GPA	GRE-Q	GRE-V	MAT	AR
GPA	1.000	.611	.581	.604	.621
GRE-Q		1.000	.468	.267	.508
GRE-V			1.000	.425	.405
MAT				1.000	.525
AR					1.000

*GPA = Grade-Point Average; GRE-Q = Graduate Record Examination-Quantitative; GRE-V = Graduate Record Examination-Verbal; MAT = Miller Analogies Test; AR = Average Rating. Original data in Table 11.3.

TABLE 11.5 SUMMARY OF STEPWISE REGRESSION ANALYSIS FOR DATA OF TABLE 11.3. $N = 30$ *

Variables in Equation					Variables Not in Equation			
Step	Variable Entered	R	df	F*	F to Remove*	Variable	Partial Correlation	F to Enter*
1	AR	.6207	1/28	17.550	17.550	GRE-Q	.43814	1/27 6.414
2	GRE-V	.7180	2/27	14.363		GRE-V	.46022	1/27 7.256
3	MAT	.7562	3/26	11.577		MAT	.41727	1/27 5.692
4	GRE-Q	.8003	4/25	11.134		GRE-Q	.34032	1/26 3.406
5	AR-REMOVED	.7855	3/26	13.964		MAT	.34113	1/26 3.424
	GRE-V		1/26		2.311	GRE-Q	.40029	1/25 4.770
	MAT		1/25		8.389			
	GRE-Q		1/25		8.389			

*See footnote a, Table 11.4, for the names of the independent variables.

*The F ratio for the overall R at each step. For example, at step 2 the F ratio for $R = .7180$ is 14.363 with 2 and 27 degrees of freedom.

Calculs de régression pas à pas

Étape	Ajout	RPE	ΔRPE
1.	AR	,3853	,3853
2.	GRE-V	,5155	,1302
3.	MAT	,5718	,0563
4.	GRE-Q	,6408	,0690
5. (retrait)	-AR	,6170	,0238

Calcul de ΔR^2 (1)

$$F_{(f+1, N-f)} = \frac{(N-f)(RPE_i - RPE_{i-1})}{(f-r)(1-RPE_i)} = \frac{(N-f)(R_i^2 - R_{i-1}^2)}{(f-r)(1-R_i^2)}$$

Calcul de l'ajout de la première variable

Avertissement:

les données du tableau 11.5 sont des R et non des R^2

$$F_{(f+1, N-f)} = \frac{(N-f)(R_i^2 - R_{i-1}^2)}{(f-r)(1-R_i^2)} = \frac{(30-2)(0.3853 - 0^2)}{(2-1)(1-0.3853)}$$

$$= \frac{28(0.3853)}{1(0.6147)} = \frac{10.7875}{0.6147} = 17.55$$

Calcul de ΔR^2 (2)

Calcul de l'ajout de la deuxième variable

$$F_{(f-r, N-f)} = \frac{(N-f)(R_i^2 - R_j^2)}{(f-r)(1-R_j^2)} = \frac{(30-3)(0.5155 - 0.3853)}{(3-2)(1-0.5155)} =$$

$$= \frac{27(0.5155 - 0.3853)}{1(0.4845)} = \frac{27(0.1302)}{0.4845} = \frac{3.5169}{0.4845} = 7.26$$

Vous pouvez faire les autres maintenant 😊
pour l'autre tableau de données (celui sur la lecture),
vous pouvez faire les calculs avec l'information
suivante: $N = 64$; voir aussi Bryant et al. (1990).
Developmental Psychology, 26, 429-438.

Rappel

Le calcul visant à déterminer si la RPE du
modèle global est significative
(-ment différente de 0)
revient à comparer le changement par rapport à
rien

$$F_{(f-r, N-f)} = \frac{(N-f)(R_i^2 - R_j^2)}{(f-r)(1-R_j^2)} = \frac{(30-3)(0.5155 - 0^2)}{(3-1)(1-0.5155)} =$$

$$= \frac{27(0.5155)}{2(0.4845)} = \frac{13.9192}{0.9690} = 14.365$$

ce qui revient à appliquer
la formule standard

$$F_{(p(N-p))} = \frac{(N-p)RPE}{(p-1)(1-RPE)} = \frac{(N-p)R^2}{(p-1)(1-R^2)}$$

$$F_{(0.130)} = \frac{(30-3) \times 0.5155}{(3-1)(1-0.5155)} = \frac{(27) \times 0.5155}{2(0.4845)} = \frac{13.9192}{0.9690} = 14.365$$

Notions particulières

☐ Corrélation partielle:

- ♦ corrélation entre deux variables pour laquelle les (cor)relations avec une troisième variable (ou plus) ont été extraites sur les deux variables en jeu

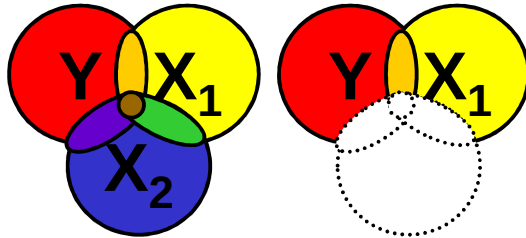
☐ Corrélation semi-partielle:

- ♦ corrélation entre deux variables pour laquelle les (cor)relations avec une troisième variable (ou plus) ont été extraites sur l'une des deux variables en jeu

Notions particulières



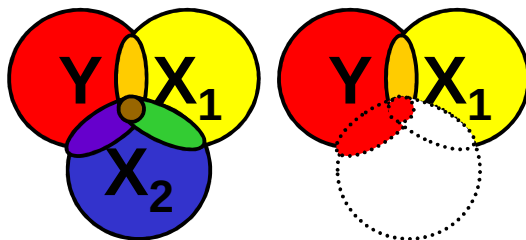
Corrélation partielle:



Notions particulières



Corrélation semi-partielle:



Retour sur la notion de
contrôle

Notion particulière (2) 2 sortes de contrôle

• contrôle expérimental:

- ❖ action planifiée pour annihiler un effet ou le répartir dans toutes les conditions

• contrôle statistique:

- ❖ calcul pour soustraire des relations (corrélations, covariances) d'une relation multivariée

Formules pour la RPE multiple (double)

□ Formule 1:

$$R^2_{yx1.x2} = \frac{r^2_{yx1} + r^2_{yx2} - 2r_{yx1}r_{yx2}r_{x1x2}}{1 - r^2_{x1x2}}$$

□ Formule 2:

$$R^2_{yx1.x2} = \beta_{yx1.x2}r_{yx1} + \beta_{yx2.x1}r_{yx2}$$

□ Formule 3:

$$R^2_{yx1.x2...p} = r^2_{yx1} + r^2_{y(x2.x1)} + \dots + r^2_{y(xp.x1x2...xp-1)}$$

Conditions et facteurs de la corrélation et de la régression

□ Conditions:

- ◆ **Homoscédasticité:**
homogénéité des variances autour de la droite de régression
- ◆ **Normalité des données**

□ Facteurs

- ◆ **restriction de l'étendue des valeurs:**
diminue la taille de r et de (b)
- ◆ **hétérogénéité des sous-échantillons:**
augmente l'erreur

RAPPEL

$$r_{xy} = \frac{\text{Covar}_{xy}}{s_x s_y}$$

$$b_{yx} = \frac{\text{Covar}_{xy}}{s_x^2}$$

$$b_{yx \cdot x2} = \frac{\text{Co var}_{x1y} s_{x2}^2 - \text{Co var}_{x2y} \text{Co var}_{x1x2}}{s_{x1}^2 s_{x2}^2 - (\text{Co var}_{x1x2})^2}$$

$$\sum X^2 - \frac{(\sum X)^2}{N} \quad \sum XY - \frac{\sum X \sum Y}{N}$$

Extra

Calcul de ΔR^2 (1)

$$F_{f-r, N-f} = \frac{(N-f)(RPE - RPE)}{(f-r)(1-RPE)} = \frac{(N-f)(R_i^2 - R_i^2)}{(f-r)(1-R_i^2)}$$

Calcul de l'ajout de la première variable

Avertissement:

les données du tableau 11.5 sont des R et non des R^2

$$F_{f-r, N-f} = \frac{(N-f)(R_i^2 - R_i^2)}{(f-r)(1-R_i^2)} = \frac{(30-2)((.6207)^2 - 0^2)}{(2-1)(1-0.3853)}$$

$$= \frac{28(0.3853)}{1(0.6147)} = \frac{10.7875}{0.6147} = 17.55$$

Calcul de ΔR^2 (2)

Calcul de l'ajout de la deuxième variable

$$F_{(f-r, N-f)} = \frac{(N-f)(R_t^2 - R_r^2)}{(f-r)(1-R_r^2)} = \frac{(30-3)((0.7180)^2 - (0.6207)^2)}{(3-2)(1-0.5155)} =$$

$$= \frac{27(0.5155 - 0.3853)}{1(0.4845)} = \frac{27(0.1302)}{0.4845} = \frac{3.5169}{0.4845} = 7.26$$

Vous pouvez faire les autres maintenant 😊
 pour l'autre tableau de données (celui sur la lecture),
 vous pouvez faire les calculs avec l'information
 suivante: $N = 64$; voir aussi Bryant et al. (1990).
Developmental Psychology, 26, 429-438.

Rappel

Le calcul visant à déterminer si la RPE du modèle global est significative (-ment différente de 0)

revient à comparer le changement par rapport à rien

$$F_{(f-r, N-f)} = \frac{(N-f)(R_t^2 - R_r^2)}{(f-r)(1-R_r^2)} = \frac{(30-3)((0.7180)^2 - 0^2)}{(3-1)(1-0.5155)} =$$

$$= \frac{27(0.5155)}{2(0.4845)} = \frac{13.9192}{0.9690} = 14.365$$

ce qui revient à appliquer la formule standard

$$F_{(p-N, p)} = \frac{(N-p)RPE}{(p-1)(1-RPE)} = \frac{(N-p)R^2}{(p-1)(1-R^2)}$$

$$F_{(0.188)} = \frac{(30-3) \times 0.5155}{(3-1)(1-0.5155)} = \frac{(27) \times 0.5155}{2(0.4845)} = \frac{13.9192}{0.9690} = 14.365$$

Calcul de l'exemple (2)

Co var $x_1 y$ =

S x_2^2 =

S x_1 =

Co var $x_2 y$ =

Co var $x_1 x_2$ =