

Introduction

Exercises ISLR – Ch.3

Marcelo Previato Simoes Nº 2367070

25/08/2025



Conceptual

Exercises: 5, 6

Exercise 5) Regression without intercept

$$\hat{y} = \hat{\beta}x \text{ replacing by } \hat{\beta} = \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2}$$

$$\Rightarrow \hat{y} = x \frac{\sum_{i=1}^n x_i y_i}{\sum_{i=1}^n x_i^2} = \frac{\sum_{i=1}^n x \cdot x_i y_i}{\sum_{i=1}^n x_i^2}$$

$$\Rightarrow \hat{y} = \sum_{i=1}^n a_i y_i \text{ in which } a_i = \frac{x \cdot x_i}{\sum_{i=1}^n x_i^2}$$

Exercise 6) Regression line through the means

Applying $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$ to the point \bar{x} :

$$\Rightarrow \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 \bar{x} \quad \text{replacing by } \hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$$

$\Rightarrow \hat{y} = \bar{y}$, therefore (\bar{x}, \bar{y}) belongs to the regression line



Applied

Exercise 15

Exercise 15a) Simple Regression (Boston)

Dependent Variable: crim (per capita crime rate per town)

All variables with coefficients statistically significant at 95%, except for Charles River proximity (chas)

Coefficient SIGNIFICANT (95%)	Coefficient NOT significant
zn	chas
lnd	
nox	
rm	
age	
dis	
rad	
tax	
ptratio	
black	
lstat	
medv	

Exercise 15a) Simple Regression (Boston)

Example: variable Nox

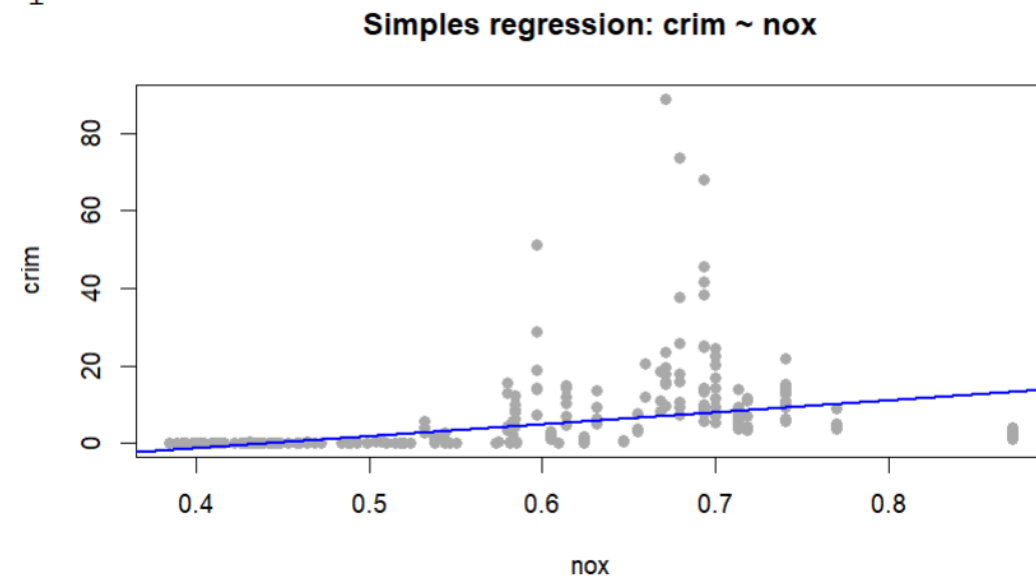
Significant, but it explains only 17.72% of the variance

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-13.720	1.699	-8.073	5.08e-15 ***
nox	31.249	2.999	10.419	< 2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.81 on 504 degrees of freedom
Multiple R-squared: 0.1772, Adjusted R-squared: 0.1756
F-statistic: 108.6 on 1 and 504 DF, p-value: < 2.2e-16





Exercise 15b) Multiple Regression

Dependent Variable: crim (per capita crime rate per town)

Only 5 variables remain significant: zn, dis, rad, black, medv

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	17.033228	7.234903	2.354	0.018949	*
zn	0.044855	0.018734	2.394	0.017025	*
indus	-0.063855	0.083407	-0.766	0.444294	
chas	-0.749134	1.180147	-0.635	0.525867	
nox	-10.313535	5.275536	-1.955	0.051152	.
rm	0.430131	0.612830	0.702	0.483089	
age	0.001452	0.017925	0.081	0.935488	
dis	-0.987176	0.281817	-3.503	0.000502	***
rad	0.588209	0.088049	6.680	6.46e-11	***
tax	-0.003780	0.005156	-0.733	0.463793	
ptratio	-0.271081	0.186450	-1.454	0.146611	
black	-0.007538	0.003673	-2.052	0.040702	*
lstat	0.126211	0.075725	1.667	0.096208	.
medv	-0.198887	0.060516	-3.287	0.001087	**

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 6.439 on 492 degrees of freedom

Multiple R-squared: 0.454, Adjusted R-squared: 0.4396

F-statistic: 31.47 on 13 and 492 DF, p-value: < 2.2e-16

Exercise 15c) Coefficients comparison



Relevant change in coefficients: 7 out of 13 variables changed had their coefficient sign changed

predictor	uni_coef	multi_coef
zn.zn	-0.07393498	0.044855215
indus.indus	0.50977633	-0.063854824
chas.chas	-1.89277655	-0.749133611
nox.nox	31.24853120	-10.313534912
rm.rm	-2.68405122	0.430130506
age.age	0.10778623	0.001451643
dis.dis	-1.55090168	-0.987175726
rad.rad	0.61791093	0.588208591
tax.tax	0.02974225	-0.003780016
ptratio.ptratio	1.15198279	-0.271080558
black.black	-0.03627964	-0.007537505
lstat.lstat	0.54880478	0.126211376
medv.medv	-0.36315992	-0.198886821

Exercise 15d) Non-linear association



Dependent Variable: crim + up to cubic association

Only 5 variables seem to have non-linear association

Non-linear Association Coefficient SIGNIFICANT (95%)	Non-linear Association Coefficient NOT significant
ind	chas
nox	zn
age	rm
dis	rad
medv	tax
	ptratio
	black
	lstat

Exercise 15d) Non-linear association



Example: variable Nox up to cubic term

Significant, and now it explains 29.7% of the variance

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	233.09	33.64	6.928	1.31e-11	***
nox	-1279.37	170.40	-7.508	2.76e-13	***
I(nox^2)	2248.54	279.90	8.033	6.81e-15	***
I(nox^3)	-1245.70	149.28	-8.345	6.96e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 7.234 on 502 degrees of freedom

Multiple R-squared: 0.297, Adjusted R-squared: 0.2928

F-statistic: 70.69 on 3 and 502 DF, p-value: < 2.2e-16

Cubic regression: $\text{crim} \sim \text{nox} + \text{nox}^2 + \text{nox}^3$

