APPLIED REGRESSION ANALYSIS
Lecture 1

Ch1: Introduction

Gengxin li

Department of Mathematics and Statistics
University of Michigan dearborn

Simple Linear regression model: one explanatory variable

Figure shows clutch volume plotted against body size, with clutch volume on the *y*-axis and body size on the *x*-axis. Each point represents a single case. For this example, each case is one egg clutch for which both volume and body size (of the female that produced the clutch) have been recorded.

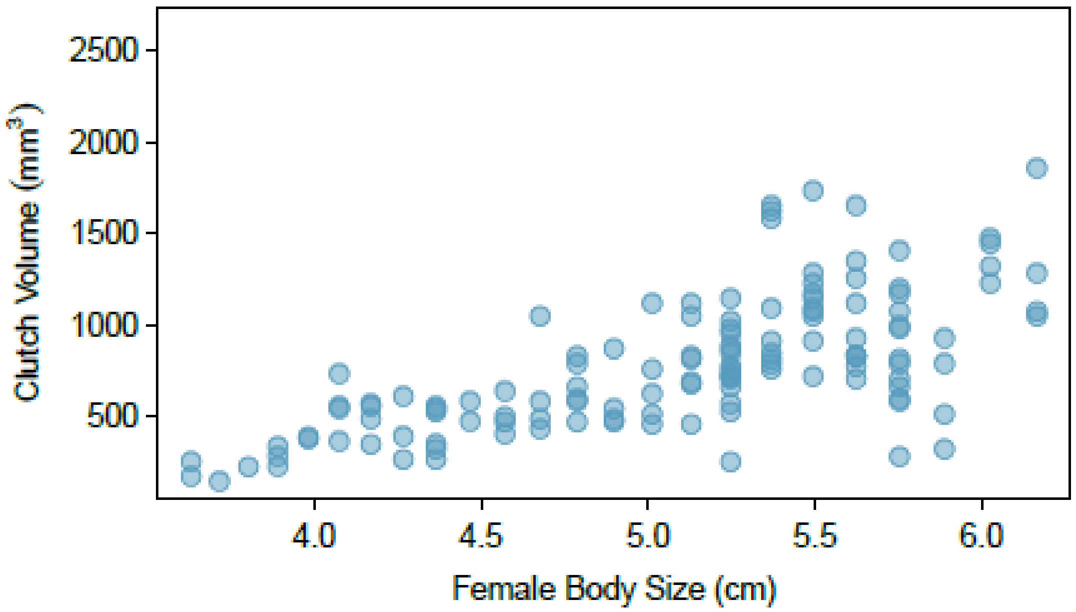

Figure 1.26: A scatterplot showing `clutch.volume` (vertical axis) vs. `body.size` (horizontal axis).

The plot shows that the points tend to lie in a straight line, which is indicative of a linear association. Two variables are **positively associated** if increasing values of one tend to occur with increasing values of the other

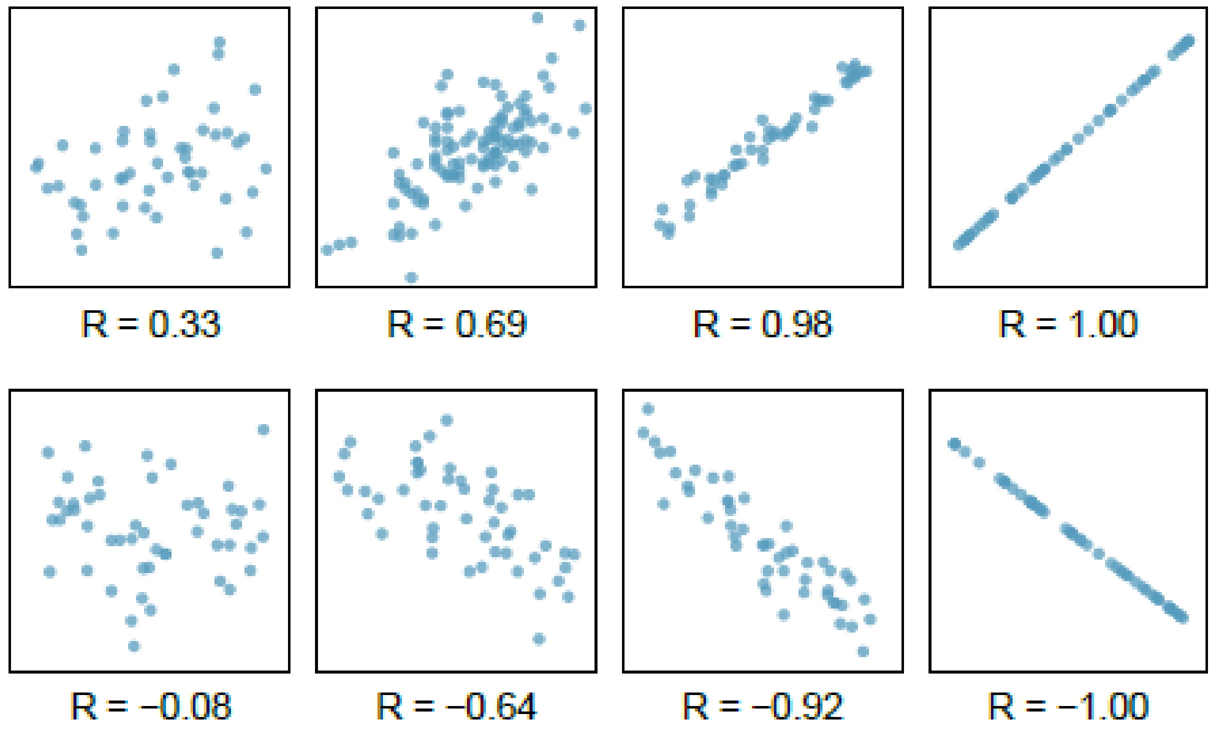

Figure 1.29: Scatterplots and their correlation coefficients. The first row shows positive associations and the second row shows negative associations. From left to right, strength of the linear association between x and y increases.

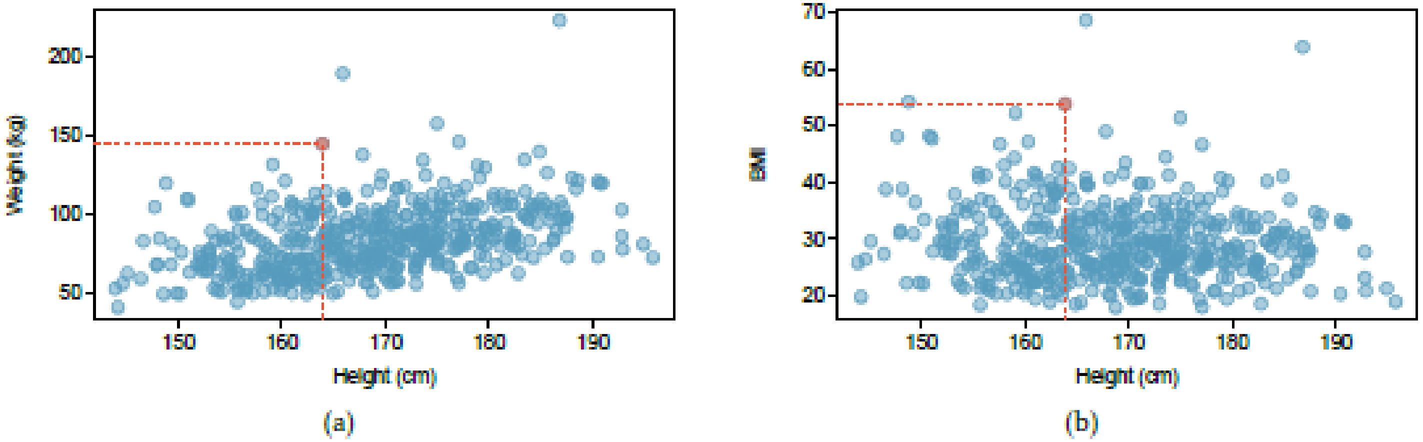

Figure 1.27: (a) A scatterplot showing height versus weight from the 500 individuals in the sample from NHANES. One participant 163.9 cm tall (about 5 ft, 4 in) and weighing 144.6 kg (about 319 lb) is highlighted. (b) A scatterplot showing height versus BMI from the 500 individuals in the sample from NHANES. The same individual highlighted in (a) is marked here, with BMI 53.83.

$$BMI = \frac{weight_{kg}}{height_m^2} = \frac{weight_{lb}}{height_{in}^2} \times 703.$$

Figure is a scatterplot of life expectancy versus annual per capita income for 165 countries in 2011.

Life expectancy is measured as the expected lifespan for children born in 2011

Income is adjusted for purchasing power in a country.

Describe the relationship between life expectancy and annual per capita income;

do they seem to be linearly associated?

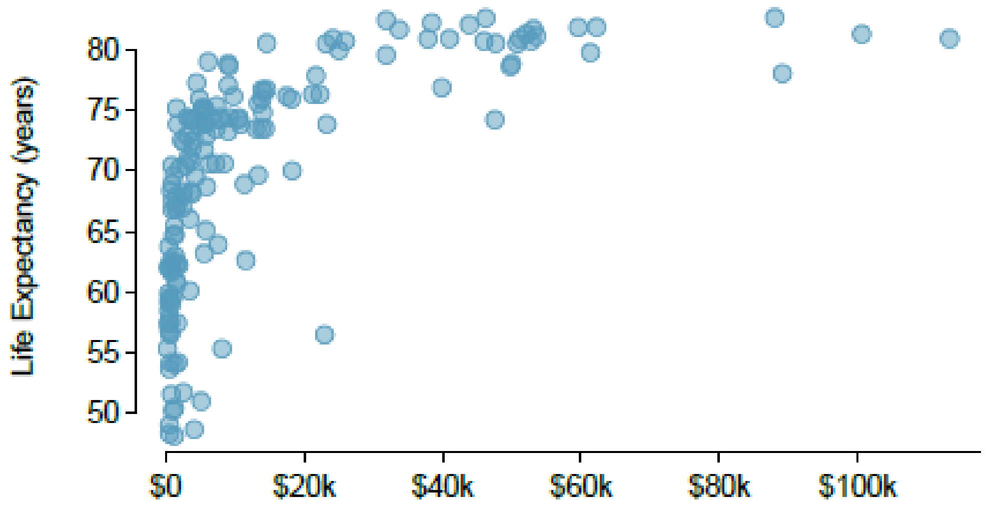

Figure 1.28: A scatterplot of life expectancy (years) versus annual per capita income (US dollars) in the wdi.2011 dataset.

Figure is a scatterplot of life expectancy versus annual per capita income for 165 countries in 2011.

Life expectancy is measured as the expected lifespan for children born in 2011

Income is adjusted for purchasing power in a country.

Describe the relationship between life expectancy and annual per capita income;

do they seem to be linearly associated?

small increases in income are associated with relatively **LAGRE** increases in life expectancy

small increases in income are associated with **SMALL** increases in life expectancy

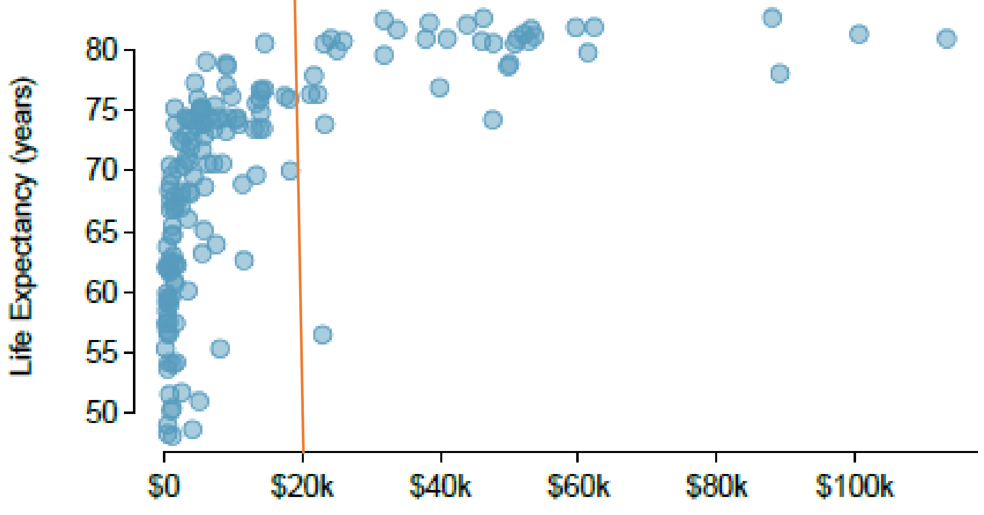

Figure 1.28: A scatterplot of life expectancy (years) versus annual per capita income (US dollars) in the wdi.2011 dataset.

Conclusion: nonlinear relationship

i. Life expectancy and annual per capita income are positively associated;
higher per capita income is associated with longer life expectancy. However, the two variables are not linearly associated.

When income is low, small increases in per capita income are associated with relatively large increases in life expectancy. However, once per capita income exceeds approximately $20,000 per year, increases in income are associated with smaller gains in life expectancy.

ii. In a linear association, change in the *y*-variable for every unit of the *x*-variable is consistent across the range of the *x*-variable.

for example, a linear association would be present if an increase in income of $10,000 corresponded to an increase in life expectancy of 5 years, across the range of income.

Transforming data:
Figure shows an approximately linear relationship;
As calculated from statistical software, $r = 0{:}79$, which is indicative of a strong linear relationship

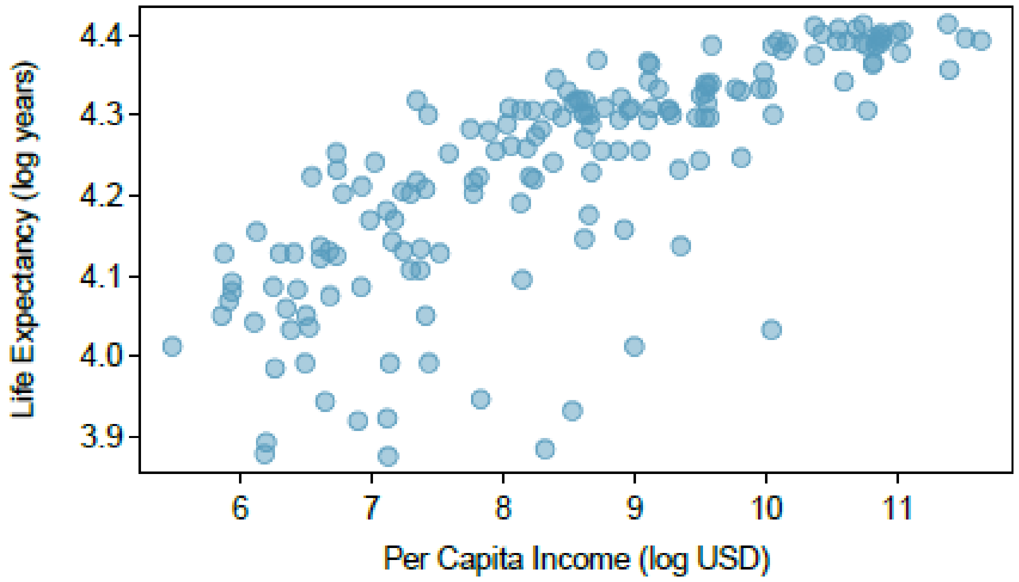

Figure 1.31: A scatterplot showing `log(income)` (horizontal axis) vs. `log(life.expectancy)` (vertical axis).

<u>Example</u>: **Multiple regression case**, several explanatory variables.
The Functional polymorphisms Associated with human Muscle Size and Strength study (FAMuSS) measured a variety of demographic, phenotypic, and genetic characteristics for about 1,300 participants. The famuss dataset is a subset of the data for 595 participants.

Sex, age, race, height, weight, actn3.r577x: independent variable, explanatory variables

	sex	age	race	height	weight	actn3.r577x	ndrm.ch
1	Female	27	Caucasian	65.0	199.0	CC	40.0
2	Male	36	Caucasian	71.7	189.0	CT	25.0
3	Female	24	Caucasian	65.0	134.0	CT	40.0
595	Female	30	Caucasian	64.0	134.0	CC	43.8

Y: Dependent variable, study variable, response

Figure 1.6: Four rows from the famuss data matrix.

variable	description
sex	Sex of the participant
age	Age in years
race	Race, recorded as African Am (African American), Caucasian, Asian, Hispanic or Other
height	Height in inches
weight	Weight in pounds
actn3.r577x	Genotype at the location r577x in the ACTN3 gene.
ndrm.ch	Percent change in strength in the non-dominant arm, comparing strength after to before training

Figure 1.7: Variables and their descriptions for the famuss dataset.