Section 12.1 — Scatter Plots and Correlation

Chris Godbout

Outline

Introduction

Definitions

Definition (Scatter Plot)

A scatter plot is a graph on the xy-plane that contains one point for each pair of data.

Definitions

Definition (Scatter Plot)

A scatter plot is a graph on the xy-plane that contains one point for each pair of data.

Definition (Variables)

If there is some sort of relationships, we say that a change in variable — the explanatory variable — influences a change in the other variable — the response variable.

Grades

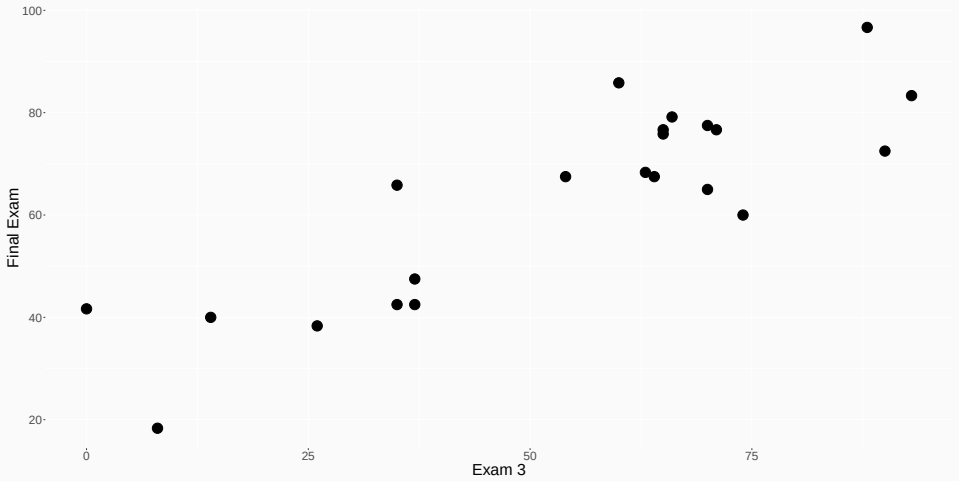
Hours Unsupervised	0	0	0.5	1.0	1.0	1.5
Overall Grade Average	96	91	88	92	94	91
Hours Unsupervised	2.0	3.0	3.0	4.0	4.5	5.6
Overall Grade Average	87	85	81	80	77	72

Grades

Exam Scores

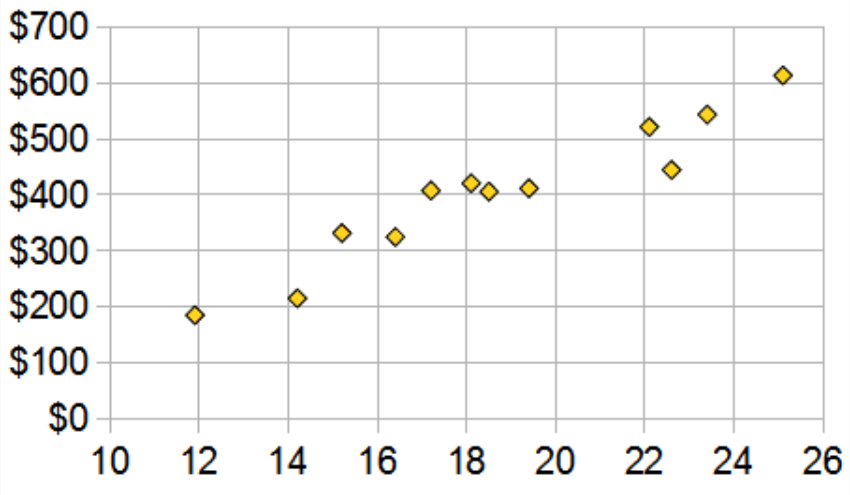
Exam 3	74	65	90	88	14	71	35	93	0	64	70
Final	60	77	73	97	40	77	43	83	42	68	78
Exam 3	65	66	37	70	54	60	35	26	37	63	8
Final	76	80	43	65	68	86	66	39	48	69	18

Exam Scores

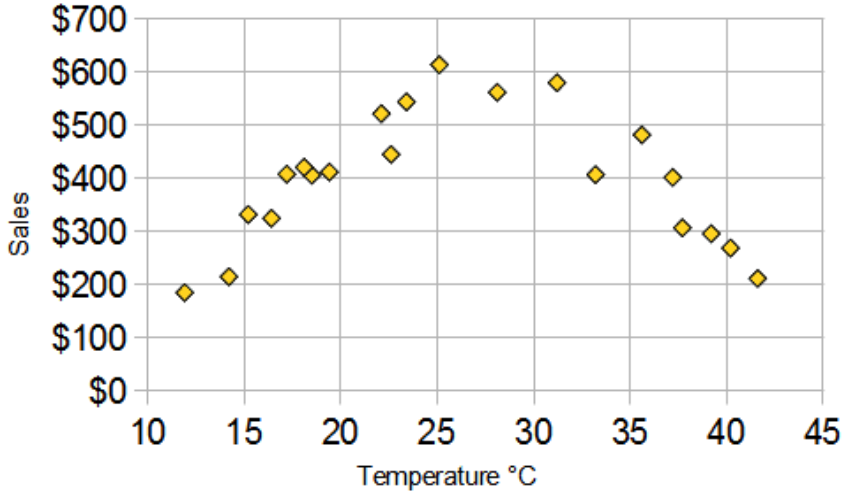


Correlations

Linear Correlation



Nonlinear Correlation



Pearson Correlation Coefficient

Definition (Correlation Coefficient)

The Pearson correlation coefficient, ρ, is the parameter that measures the strength of a linear relationship between two quantitative variables in a population. The correlation coefficient for a sample is denoted r. It is always between -1 and 1, inclusive.

$$-1 \leq r \leq 1$$

Formulas

Simple to remember

$$r = \frac{\sum(z_x z_y)}{n-1}$$

Formulas

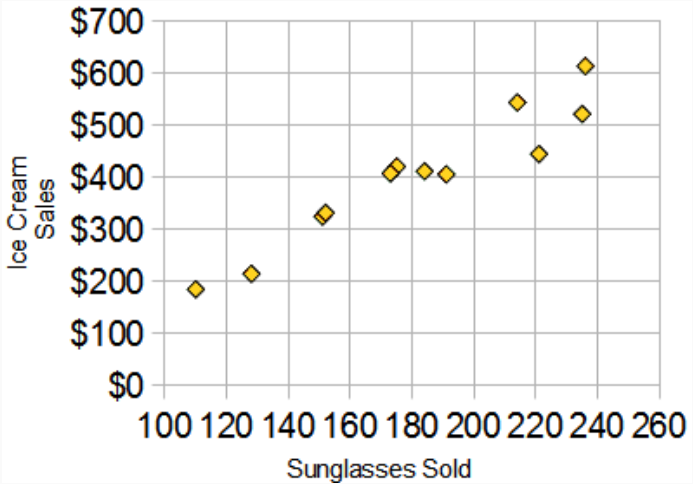
Simple to remember

$$r = \frac{\sum(z_x z_y)}{n-1}$$

Harder to remember

$$r = \frac{n(\sum xy) - (\sum x)(\sum y)}{\sqrt{n(\sum x^2) - (\sum x)^2}\sqrt{n(\sum y)^2 - (\sum y)^2}}$$

Correlation?



Significance of Correlation

Significance

A sample correlation coefficient, r, is significant if $|r| \geq r_\alpha$.

Hypothesis Test!

Hypotheses

Correlation	Negative Correlation	Positive Correlation
$H_0 : \rho = 0$	$H_0 : \rho \geq 0$	$H_0 : \rho \leq 0$
$H_a : \rho \neq 0$	$H_a : \rho < 0$	$H_a : \rho > 0$

Hypothesis Test!

Hypotheses

Correlation	Negative Correlation	Positive Correlation
$H_0 : \rho = 0$	$H_0 : \rho \geq 0$	$H_0 : \rho \leq 0$
$H_a : \rho \neq 0$	$H_a : \rho < 0$	$H_a : \rho > 0$

Test Statistic

$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}}$$

with $n - 2$ degrees of freedom.

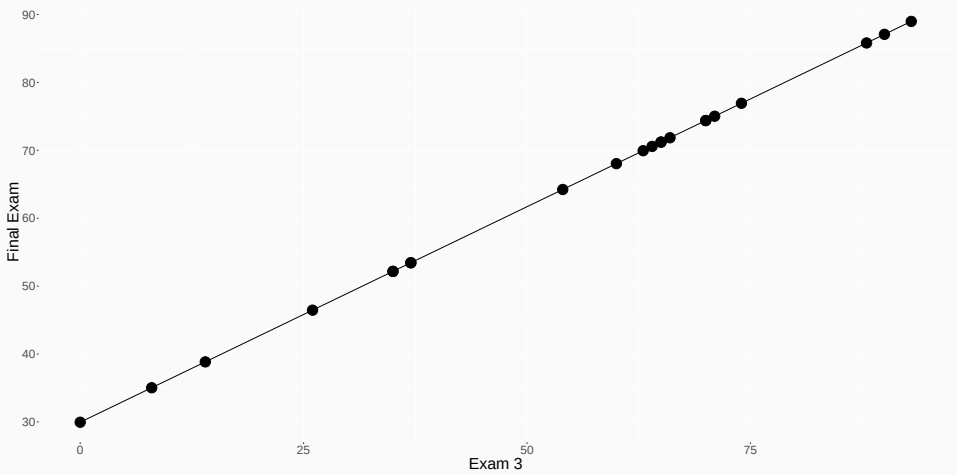
Coefficient of Determination

Definition

Definition (Coefficient of Determination)

The coefficient of determination, r^2, is the measure of the proportion of variation in the response variable that can be associated with the variation in the explanatory variable.

$r^2 = 1$



$r^2 < 1$