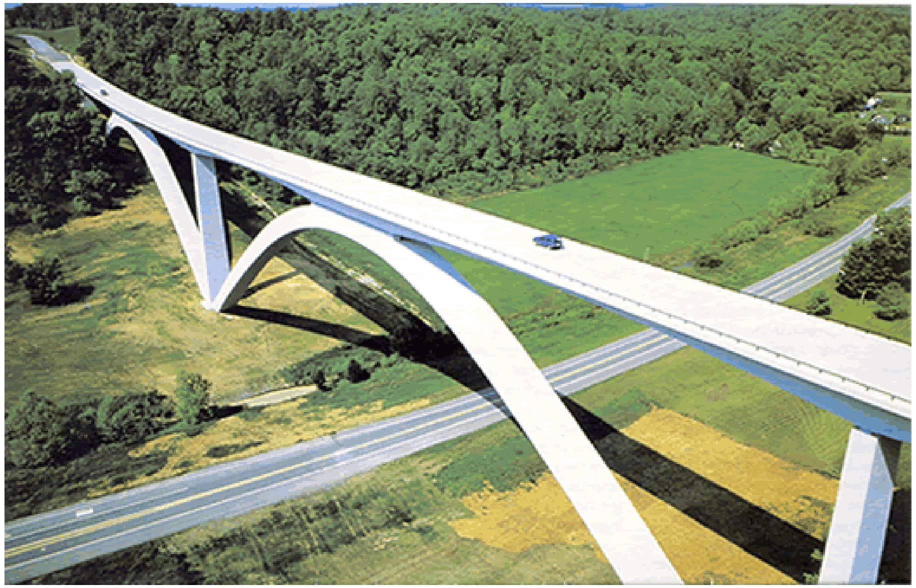
Quantitative Risk Analysis

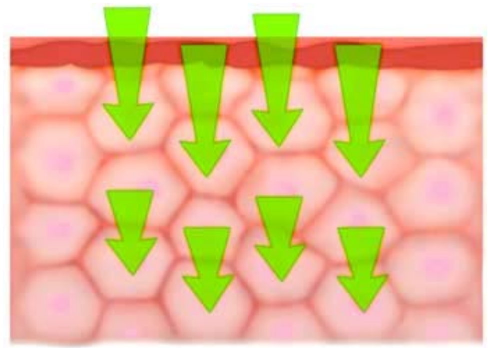
Using Correlation and Simple Linear Regression

COORDINATOR:
Dr Lihai Zhang
Infrastructure Engineering
Department (Room B307)
lihzhang@unimelb.edu.au



An environmental engineer is studying the rate of absorption of a certain chemical into skin. She obtains a series of experimental results as follows

Volume (mL)	Time (h)	Percent Absorbed
0.05	2	48.3
0.05	2	51.0
0.05	2	54.7
2.00	10	63.2
2.00	10	67.8
2.00	10	66.2
5.00	24	83.6
5.00	24	85.1
5.00	24	87.8



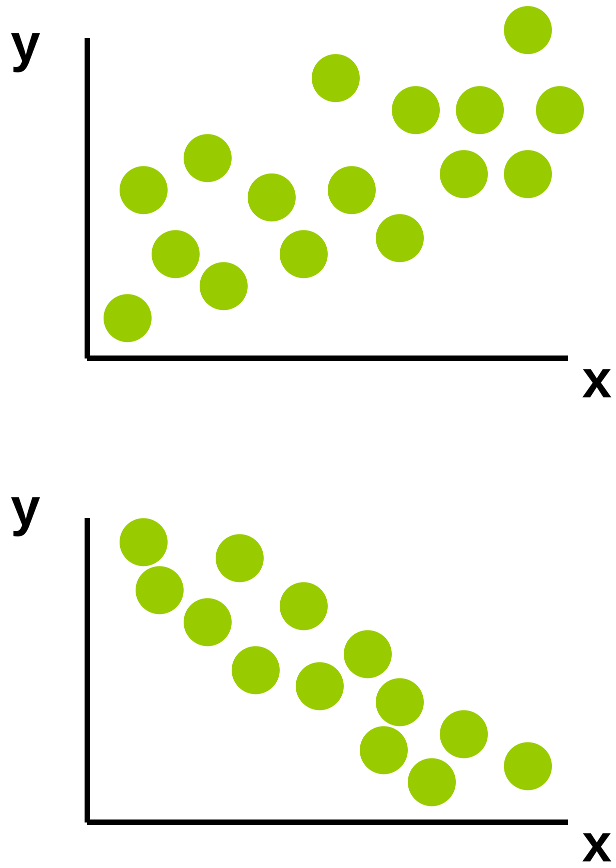
Is any correlation between time and absorption?

Is any correlation between volume and absorption?

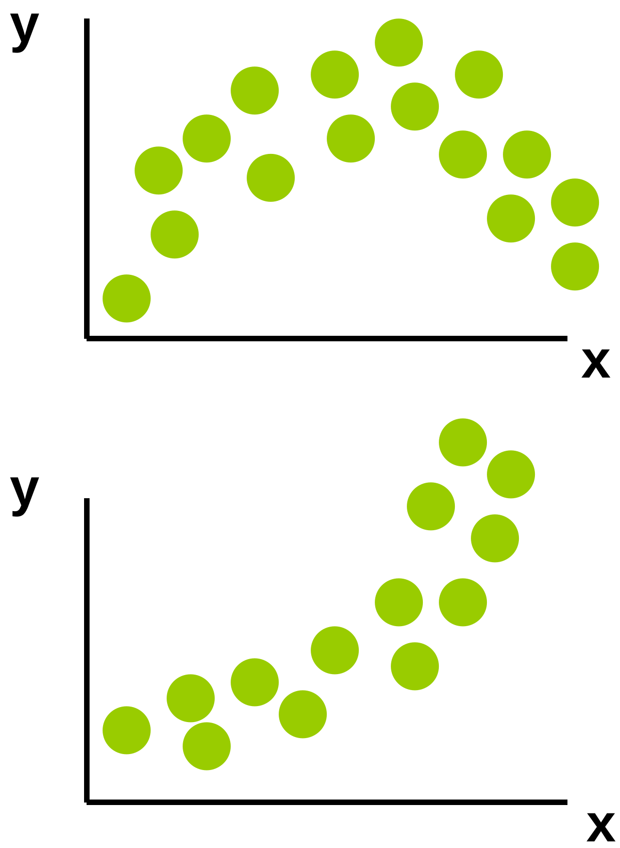
Scatter Plots

Scatter Plots are used to show the relationship between two variables.

Linear relationships

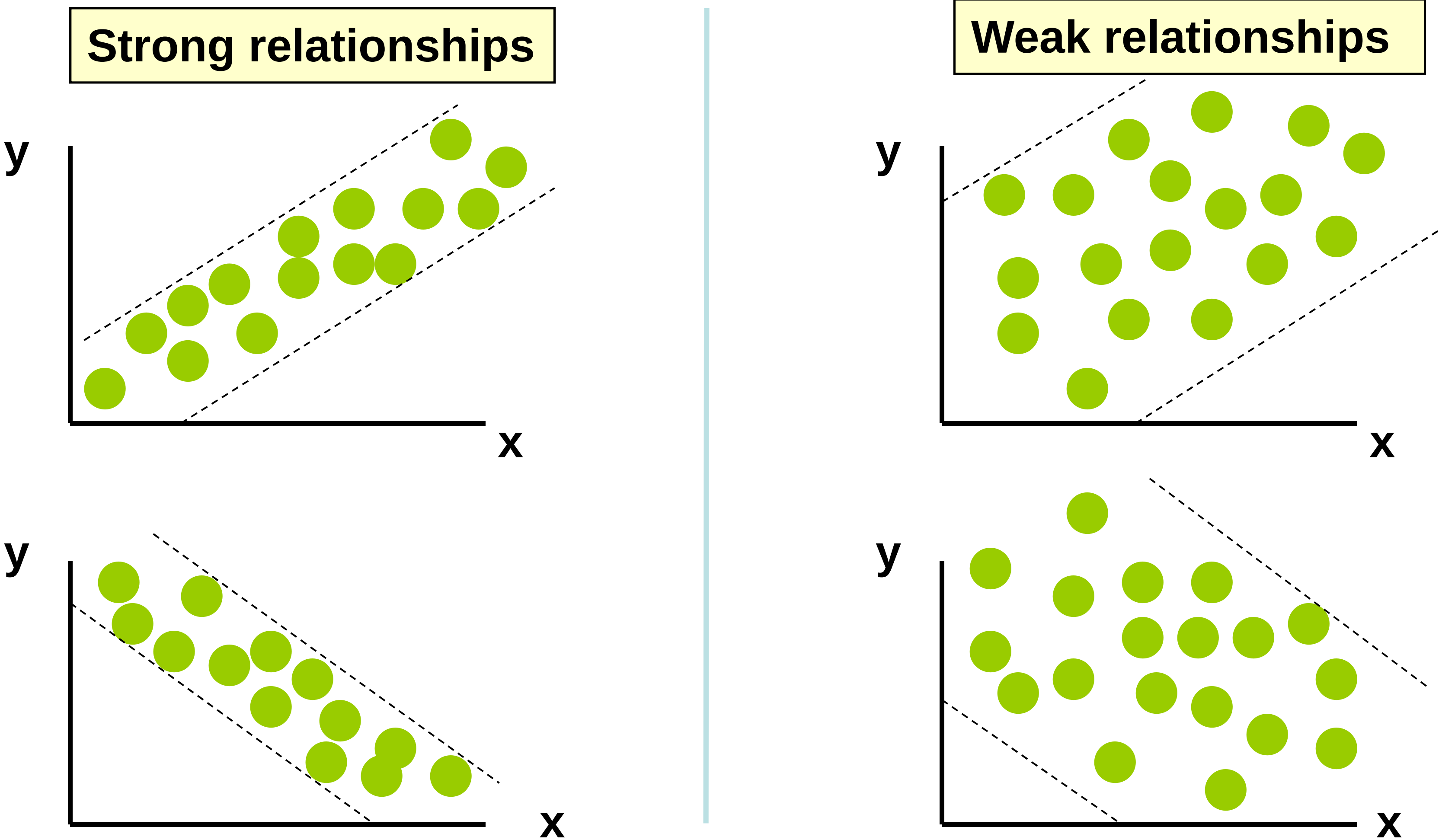


Nonlinear relationships



Correlation

Correlation analysis is used to measure strength of the association (linear relationship) between two variables

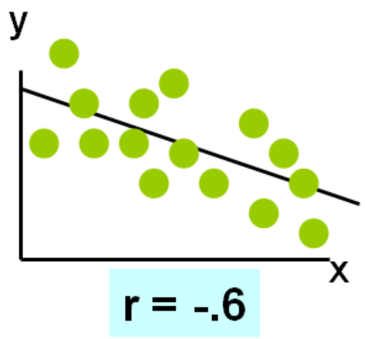


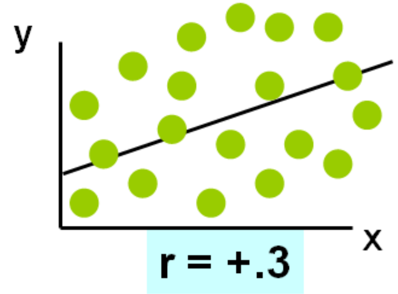
THE UNIVERSITY OF MELBOURNE

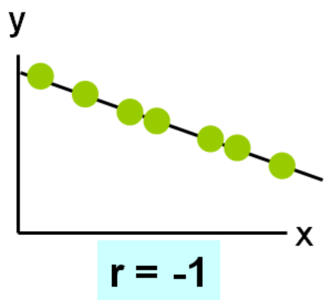
Correlation

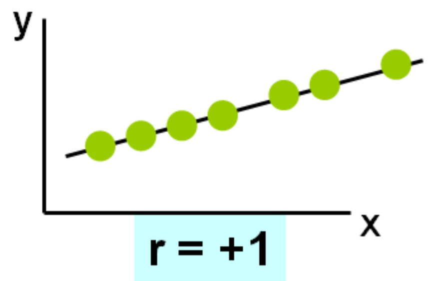
Correlation coefficient – A numerical measure of the strength of the linear relationship between two variables.

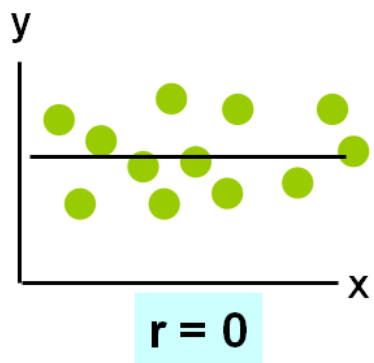
- Range between -1 and 1
- The closer to -1, the stronger the negative linear relationship
- The closer to 1, the stronger the positive linear relationship
- The closer to 0, the weaker the linear relationship

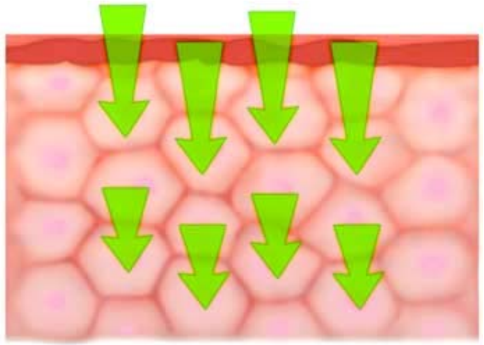












Correlation Coefficient

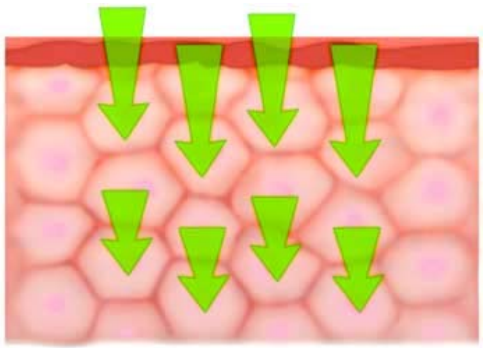
Sample correlation coefficient:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\left[\sum_{i=1}^{n}(x_i - \bar{x})^2\right]\left[\sum_{i=1}^{n}(y_i - \bar{y})^2\right]}}$$

where: r = Sample correlation coefficient
n = Sample size
x_i = Value of the independent variable
y_i = Value of the dependent variable

An environmental engineer is studying the rate of absorption of a certain chemical into skin. She obtains a series of experimental results as follows

Volume (mL)	Time (h)	Percent Absorbed
0.05	2	48.3
0.05	2	51.0
0.05	2	54.7
2.00	10	63.2
2.00	10	67.8
2.00	10	66.2
5.00	24	83.6
5.00	24	85.1
5.00	24	87.8



(a) Is any correlation between time and absorption?

(b) Is any correlation between volume and absorption?

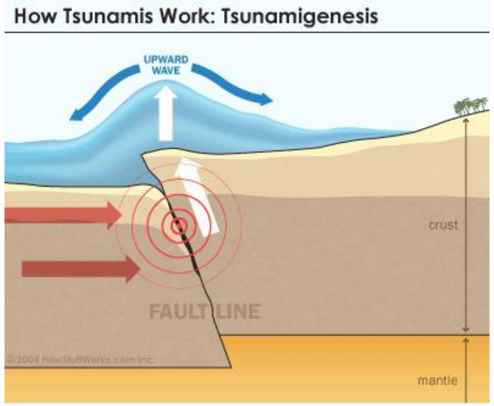
Solution:

Example (Earthquakes)

In a study of ground motion caused by earthquakes, the peak velocity (in m/s) were recorded for five earthquakes. The results are presented in the following table.

Velocity	**1.54**	**1.60**	**0.95**	**1.30**	**2.92**
Acceleration	**7.64**	**8.04**	**8.04**	**6.37**	**5.00**

Compute the correlation coefficient between peak velocity and peak acceleration. Is any correlation between them?



Solution:

Significance Test for Correlation

- Hypotheses

$$H_0: \rho = 0 \quad \text{(no correlation)}$$
$$H_1: \rho \neq 0 \quad \text{(correlation exists)}$$

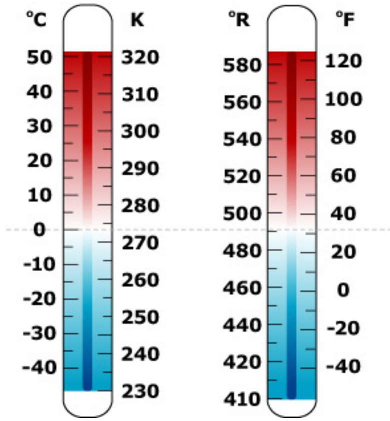
- Test statistic
  - $$t = \frac{r}{\sqrt{\dfrac{1-r^2}{n-2}}}$$ (with n – 2 degrees of freedom)

With consideration of sample size and standard deviation

Significance Test for Correlation Example

Is any correlation between temperature and number of bushfires?

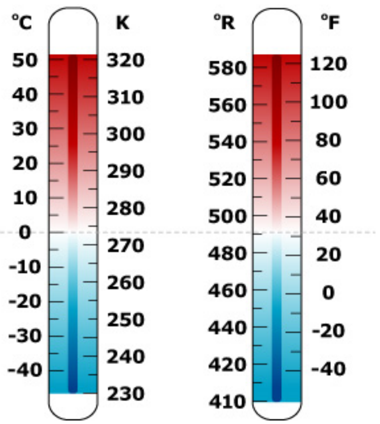
Temperature	Number of bushfires
y	**x**
35ºC	8
49ºC	9
27ºC	7
33ºC	6
60ºC	13
21ºC	7
45ºC	11
51ºC	12
Σ=321ºC	**Σ=73**



$$r = \frac{\sum (x-\bar{x})(y-\bar{y})}{\sqrt{\left[\sum (x-\bar{x})^2\right]\left[\sum (y-\bar{y})^2\right]}} = 0.886$$

Significance Test for Correlation Example

Is there evidence of a linear relationship between temperature and number of bushfires at the 0.05 level of significance?



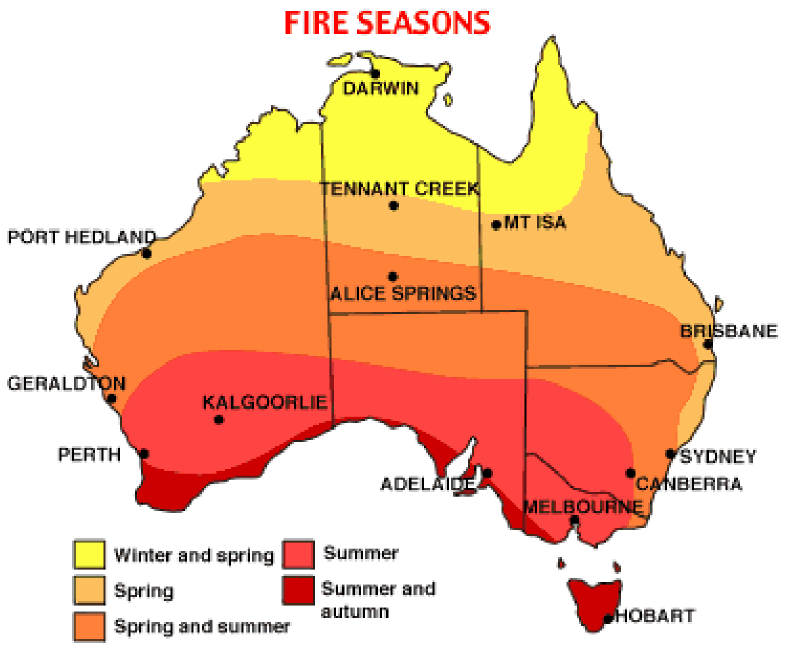
Significance Test for Correlation Example

H_0: ρ = 0 (No correlation)

H_1: ρ ≠ 0 (correlation exists)

$\alpha = .05$, df = 8 - 2 = 6

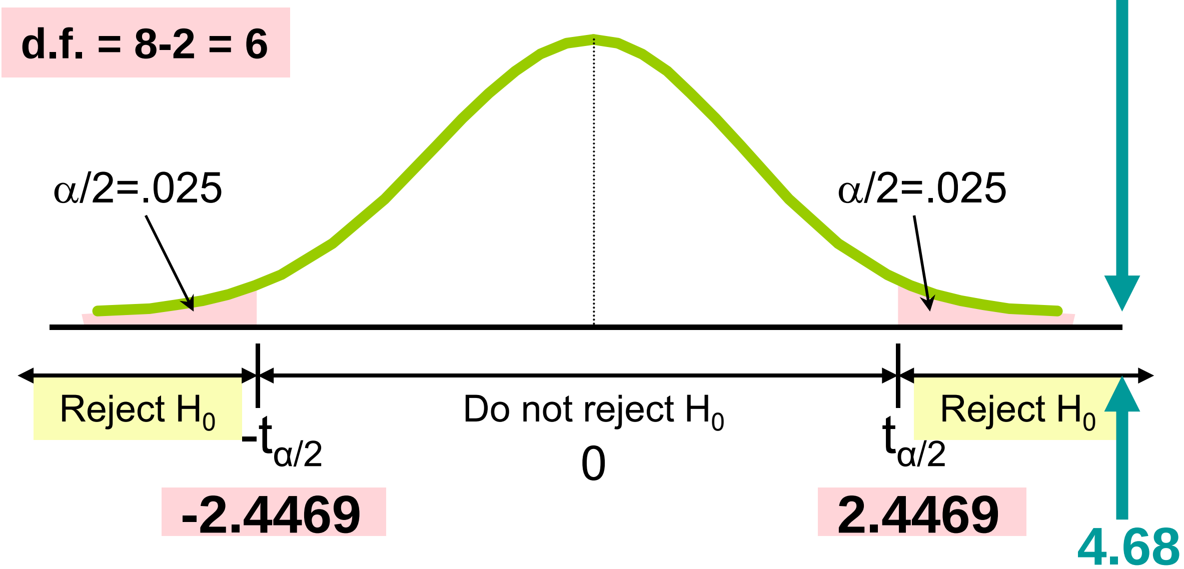
$$t = \frac{r}{\sqrt{\frac{1-r^2}{n-2}}} = \frac{.886}{\sqrt{\frac{1-.886^2}{8-2}}} = 4.68$$



THE UNIVERSITY OF MELBOURNE

Significance Test for Correlation Example

$$t = \frac{r}{\sqrt{\dfrac{1-r^2}{n-2}}} = \frac{.886}{\sqrt{\dfrac{1-.886^2}{8-2}}} = 4.68$$



Decision:
Reject H_0

Conclusion:
There is evidence of a linear relationship at the 5% level of significance

The Least-Squares Line

- When two variables have a linear relationship, the scatterplot tends to be clustered around a line known as **the least-squares line**

Example:

An accelerated test, steel structures are operated under extreme conditions until failure

Two variables:
Lifetime vs Temperature

Temperature (^oC)	Lifetime (hours)
40	851
45	635
50	764
55	708
60	469
65	661
70	586
75	371
80	337
85	245
90	129
95	158

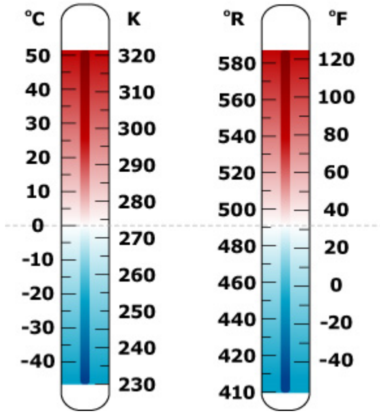
The Least-Squares Line

Dependent variable: the variable we wish to explain

Independent variable: the variable used to explain the dependent variable

Lifetime vs Temperature

Dependent variable **Independent variable**



The Least-Squares Line

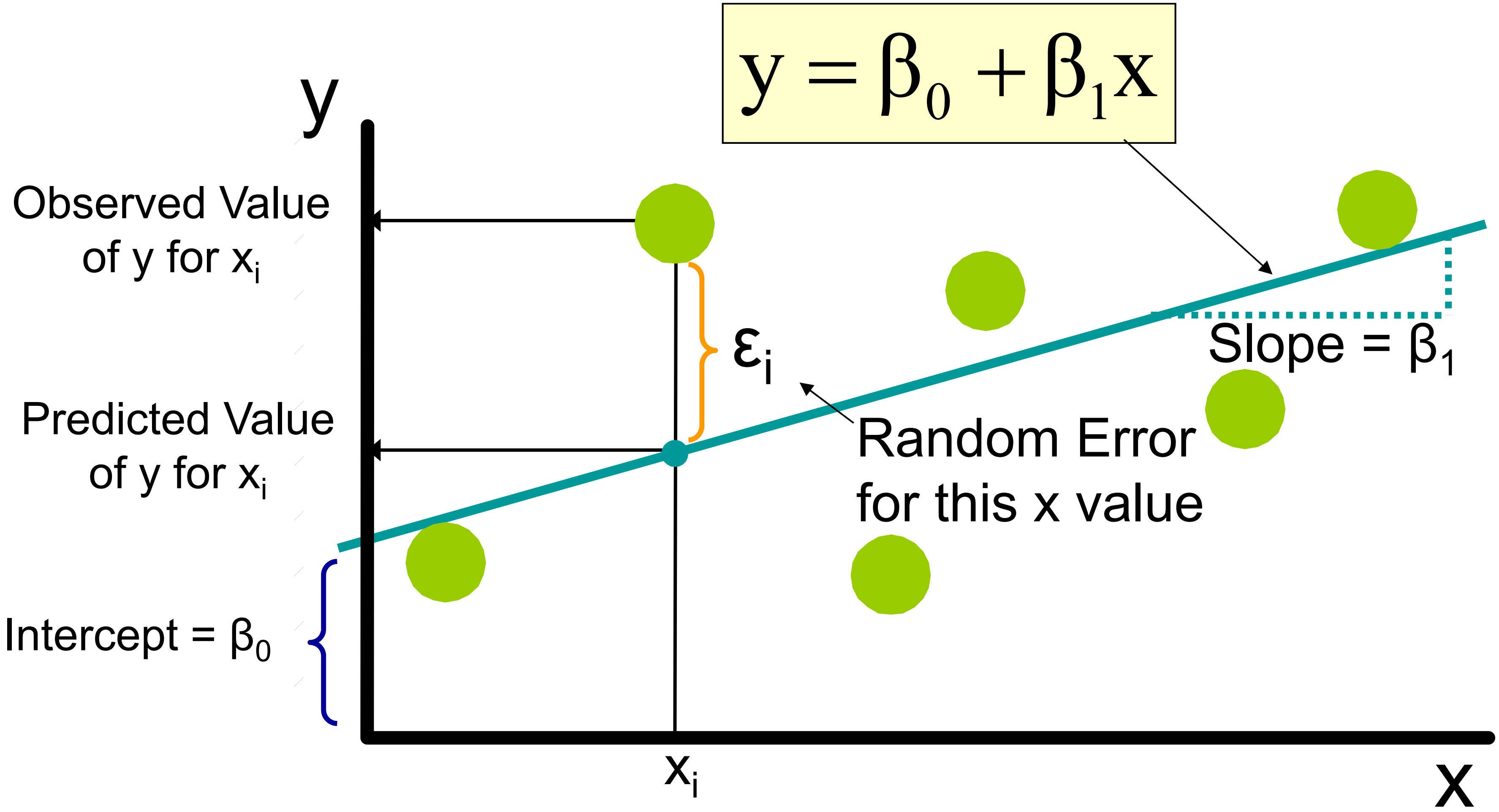
Independent Variable
(e.g. Temperature)

Dependent Variable
(e.g. Lifetime)

Regression Coefficient

$$\boxed{y = \beta_0 + \beta_1 x}$$

The Least-Squares Line



The Least-Squares Line

$$\boxed{y = \beta_0 + \beta_1 x}$$

$$\beta_1 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

$$\beta_0 = \bar{y} - \beta_1 \bar{x}$$

The Least-Squares Line

Example:

An accelerated test, steel structures are operated under extreme conditions until failure

(a) Compute the least-squares line for predicting life-time from temperature.

(b) Predict the lifetime for a temperature of 73ºC.

Temperature (ºC)	Lifetime (hours)
40	851
45	635
50	764
55	708
60	469
65	661
70	586
75	371
80	337
85	245
90	129
95	158

Solution:

Example (Fuel economy)

Inertial weight (in tons) and fuel economy (in mi/gal) were measured for a sample of seven diesel trucks. The results are presented in the following table

Weight	Mileage
8.00	7.69
24.50	4.97
27.00	4.56
14.50	6.49
28.50	4.34
12.75	6.24
21.25	4.45

(a) Compute the least-squares line for predicting mileage from weight.

(b) Predict the mileage for trucks with a weight of 15 tons.

Solution: