Stats Yr2 Chapter 1 :: Regression, Correlation & Hypothesis Tests

Chapter Overview

1:: Exponential Models

Recap of Pure Year 1. Using $y = ab^x$ to model an exponential relationship between two variables.

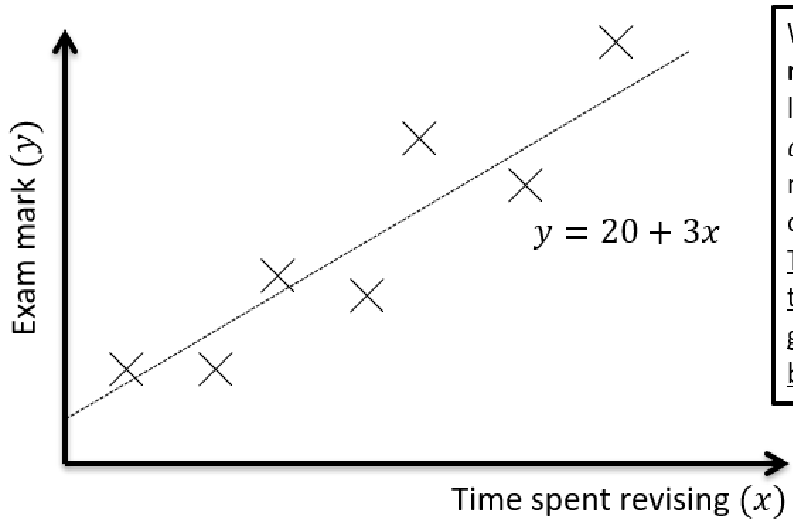
2:: Measuring Correlation

Using the Product Moment Correlation Coefficient (PMCC), r, to measure the strength of correlation between two variables.

3:: Hypothesis Testing for no correlation

We want to test whether two variables have some kind of correlation, or whether any correlation observed just happened by chance.

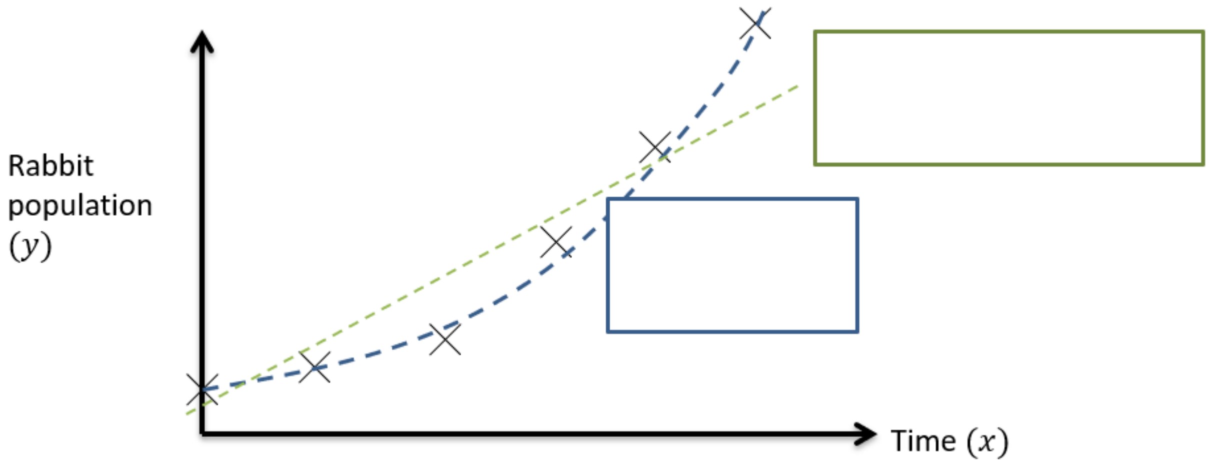
RECAP :: What is regression?



What we've done here is come up with a **model** to explain the data, in this case, a line $y = a + bx$. We've then tried to set a and b such that the resulting y value matches the actual exam marks as closely as possible.
The 'regression' bit is the act of setting the parameters of our model (here the gradient and y-intercept of the line of best fit) to best explain the data.

I record people's exam marks as well as the time they spent revising. I want to predict how well someone will do based on the time they spent revising. How would I do this?

Exponential Regression



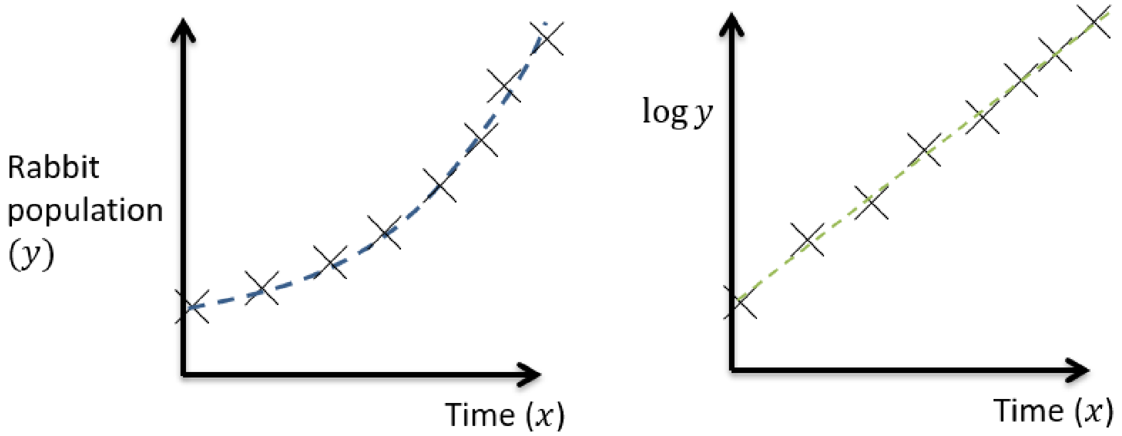
For some variables, e.g. population with time, it may be more appropriate to use an **exponential** equation, i.e. $y = ab^x$, where a and b are constants we need to fix to best match the data.

$y = ab^x$

In Year 1, what did we do to both sides to end up with a straight line equation?

If $y = kb^x$ for constants k and b then $\log y = \log k + x \log b$

Exponential Regression



Comparing the equations, we can see that if we log the y values (although leave the x values), the data then forms a straight line, with y-intercept $\log k$ and gradient $\log b$.

Example

[Textbook] The table shows some data collected on the temperature, in °C, of a colony of bacteria (t) and its growth rate (g).

Temperature, $\boldsymbol{t}$ (°C)	3	5	6	8	9	11
Growth rate, g	1.04	1.49	1.79	2.58	3.1	4.46

The data are coded using the changes of variable $x = t$ and $y = \log g$. The regression line of y on x is found to be $y = -0.2215 + 0.0792x$.
a. Mika says that the constant -0.2215 in the regression line means that the colony is shrinking when the temperature is 0°C. Explain why Mika is wrong
b. Given that the data can be modelled by an equation of the form $g = kb^t$ where k and b are constants, find the values of k and b.

a

b

Test Your Understanding

Robert wants to model a rabbit population P with respect to time in years t. He proposes that the population can be modelled using an exponential model: $P = kb^t$ The data is coded using $x = t$ and $y = \log P$. The regression line of y on x is found to be $y = 2 + 0.3x$. Determine the values of k and b.

Rabbit