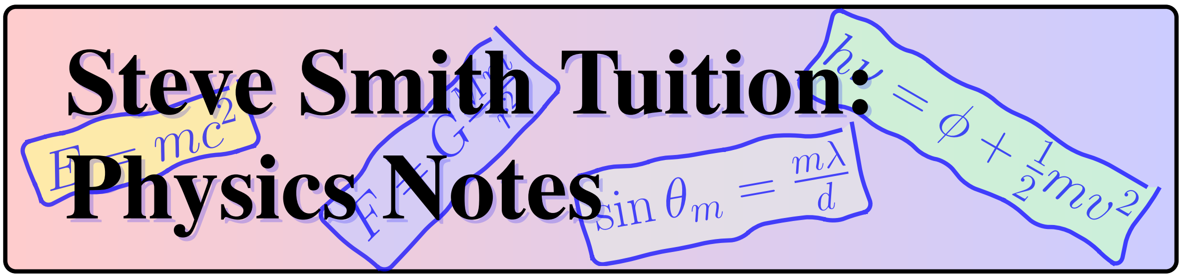


Coding in Statistics

Contents

1 Introduction **3**
1.1 Example 1 . . . 3

2 The Basic Idea **3**
2.1 Example 2 . . . 3
2.2 The Link to Graph Transformations . . . 4

Prerequisites

None.

Notes

None.

Document History

Date	Version	Comments
16th April 2013	1.0	Initial creation of the document.

1 Introduction

Coding is used in many different ways in statistics. You can use coding in correlations, discrete random variables, frequency tables, finding means, product moment correlations, regressions, standard deviations, variance...

In fact, wherever you do a statistical calculation, you can use coding. And the idea of coding is to make your calculation easier.

The basic idea is to use a cunning change of variable so that the numbers that you have in our problem are easier to manage. Here's an example.

1.1 Example 1

Find the mean of the following numbers: 98, 101, 97, 103, 101.

This is easy enough anyway, I know. All you'd have to do is to add up all the numbers and divide by 5:

$$mean \;=\; \frac{98+101+97+103+101}{5} \;=\; \frac{500}{5} \;=\; 100$$

So, here's what coding would do for you. You notice that all of the numbers are around 100. So you think to yourself: if I subtract 100 from all the numbers, and find the mean of those:

$$mean \;=\; \frac{-2+1-3+3+1}{5} \;=\; \frac{0}{5} \;=\; 0$$

then to find the mean of the original numbers, I just have to add on the 100 again:

$$mean\ of\ original\ numbers \;=\; 0+100 \;=\; 100$$

as before.

I know this is a simple example, but it's here just to show you the idea. Besides, seeing how this example works, you can translate the problem into Σ-notation:

$$mean \;=\; 100+\frac{\sum_{i=1}^{i=5}(x_i-100)}{5}$$

2 The Basic Idea

As I mentioned earlier, coding is essentially a change of variable. Usually the data we use are denoted by x_i, so we use x as our original variable. I'm gong to call the new variable X, and I'm going to define my new variable in terms of the old variable like this:

$$X_i \;=\; \frac{x_i-a}{b} \tag{1}$$

Now this transformation involves not only subtracting something (a) from each of the original values (we used 100 for that in Section 1.1), but we are also dividing the result by something (b).

Here's an example of this kind of thing.

2.1 Example 2

Find the mean of the following lengths: 5, 8, 11, 14, 17.

Let's say we want to perform the transformation (1) using 2 for the a value, and 3 for the b value. That is:

$$X_i \;=\; \frac{x_i-2}{3} \tag{2}$$

It's best to draw up a little table to see what the result of the transformation is (see Table 1).

Table 1: A simple example of coding

x_i	$\frac{x_i-2}{3}$	X_i
5	$\frac{5-2}{3}$	1
8	$\frac{8-2}{3}$	2
11	$\frac{11-2}{3}$	3
14	$\frac{14-2}{3}$	4
17	$\frac{17-2}{3}$	5

So by doing this transformation of variable, all we need to do is to work out the mean of the X_i values. This will simply be:

$$mean \;=\; \frac{1+2+3+4+5}{5} \;=\; \frac{15}{5} \;=\; 3$$

OK, but how do we use this mean of the X values to get the mean of the x values? Well, you might have twigged that we can just use the transformation (2) backwards. If

$$\frac{x_i - 2}{3} \;=\; X_i$$

then we can multiply both sides by 3,

$$x_i - 2 \;=\; 3X_i$$

and then add 2 to both sides,

$$x_i \;=\; 3X_i + 2$$

then you might well think that we could get the mean of the x values by finding the mean of the X values, and using the equation

$$x_{mean} \;=\; 3X_{mean} + 2$$

and if you did think that, you would be right! So,

$$x_{mean} \;=\; 3 \times 3 + 2 \;=\; 11$$

which does indeed turn out to be the mean of the x values.

2.2 The Link to Graph Transformations

Those eagle-eyed of you may well have spotted a similarity with our transformation (1) and those horrible graph transformations that you encountered in C1. Yes indeed - those graph transformations are very much the same sort of thing as we are doing here. Let's have a closer look at this similarity.

Let's take an example that could easily come out of a C1 graph transformation question. Let's say you had the function

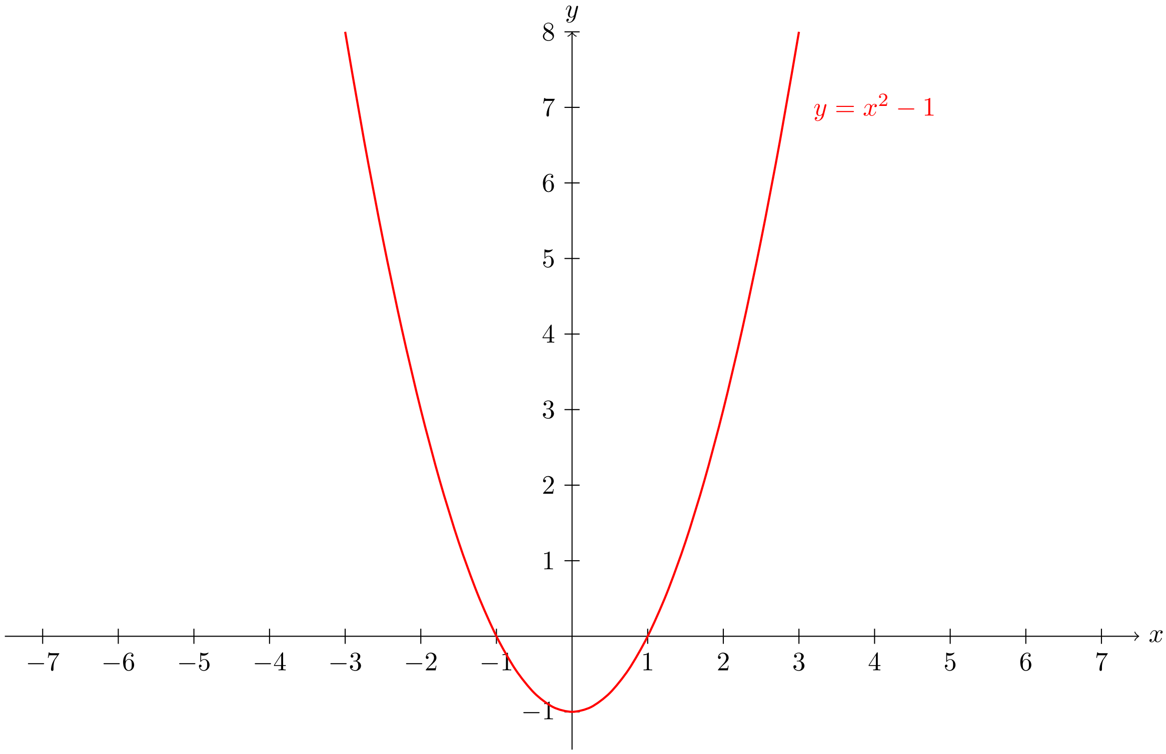
$$y \;=\; f(x) \;=\; x^2 - 1$$

and you are asked what the graph of the function

$$y \;=\; f\left(\frac{x-2}{3}\right)$$

would look like.

OK, well, let's see what these two graphs look like. Firstly, the graph of $y = x^2 - 1$



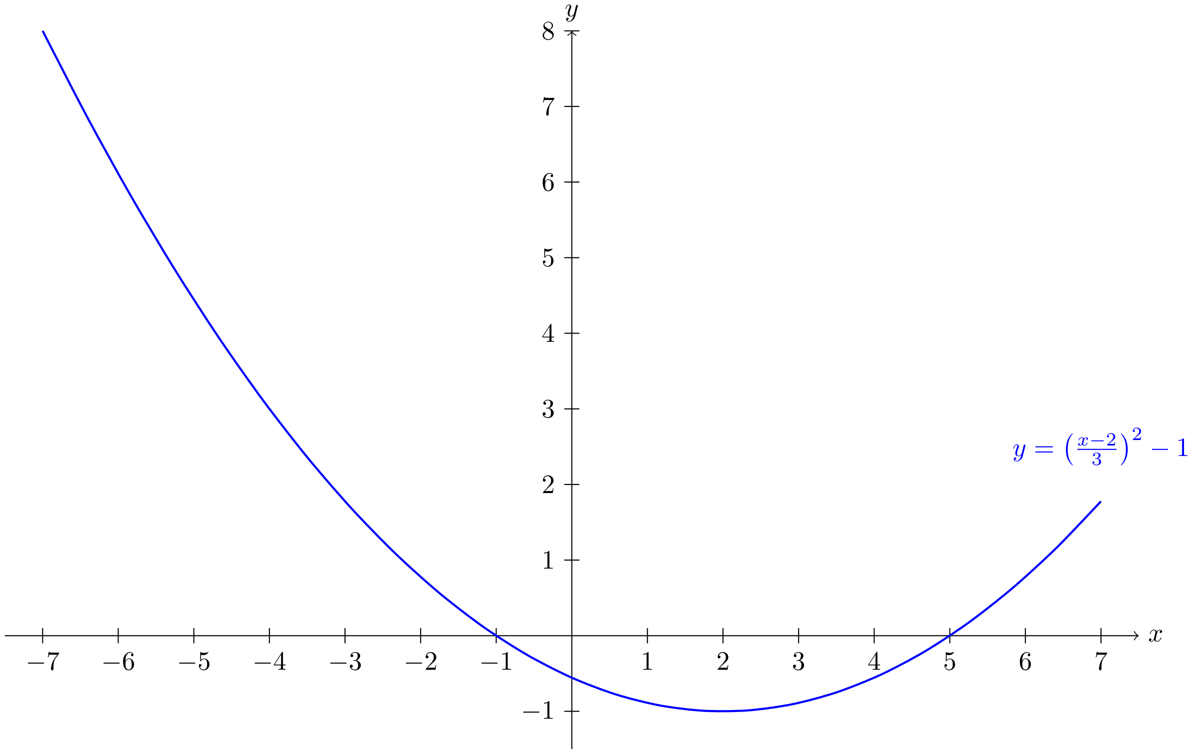
Now let's have a look at the graph of

$$y = f\left(\frac{x-2}{3}\right)$$

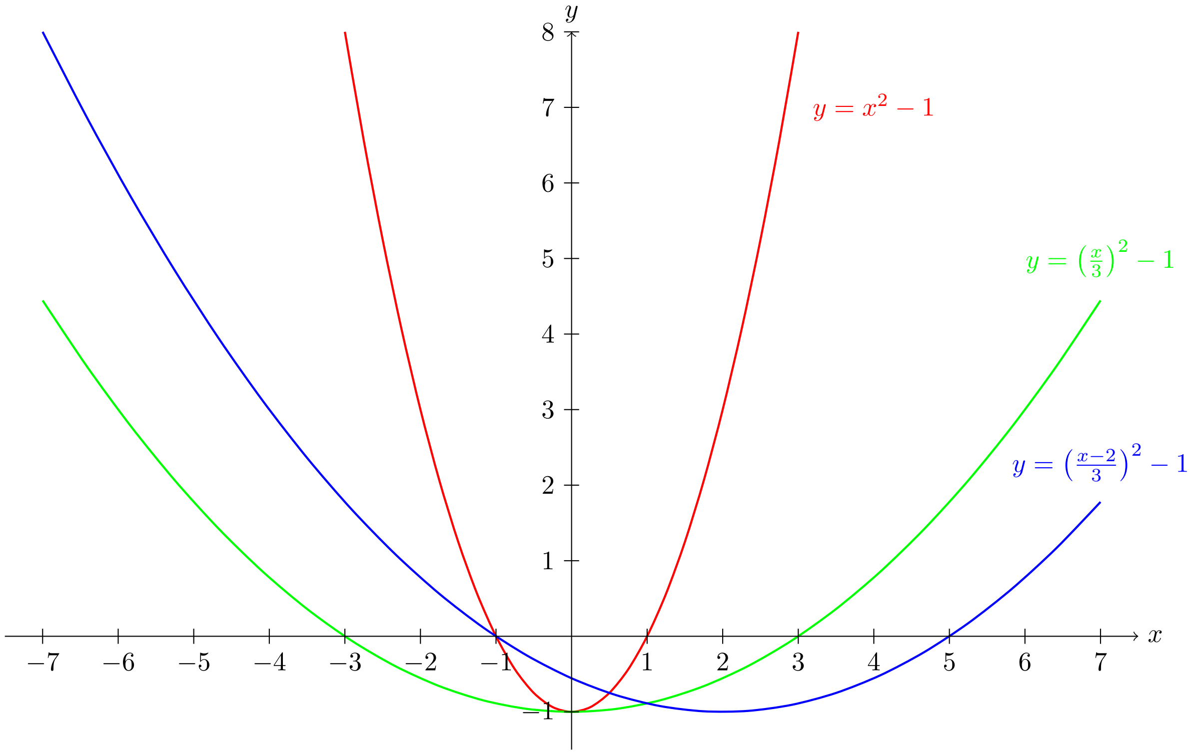
Now, since $f(x) = x^2 - 1$, this will be the graph of

$$y = \left(\frac{x-2}{3}\right)^2 - 1$$

and if you draw this, it should look something like



Now to see how to trasform the red curve onto the blue one, let's see them both on the same set of axes:



What's happening here is that there are *two* transformations going on.

Firstly, the red curve is stretched along the x-axis by a scale factor of 3. That transforms it into the green curve.

Secondly, the green curve is transformed onto the blue curve by shifting it 2 to the right.

That means that the transformation given by

$$y_i = \frac{x_i - a}{b}$$

stretches the x values by a factor of b, then shifts the resulting values a to the right.

Now if we were doing the reverse, that is, starting with the blue curve,

$$y = f\left(\frac{x-a}{b}\right)$$

then to transform this onto the red curve,

$$y = f(x)$$

then you would do the opposite transformation. And that is: firstly, shift the graph a to the left, then compress the graph by a scale factor of b in the x-direction.

Quick summary:

Table 2: Graph Transformations

Starting With:	Ending Up With:	Transformation:
x	$\frac{x-a}{b}$	Stretching the x values by a scale factor of b then shifting right by a
$\frac{x-a}{b}$	x	Shifting the x values left by a then compressing by a scale factor of b