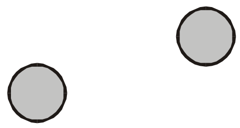
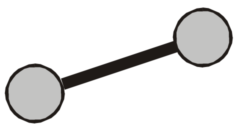
Degrees of Freedom

Imagine I throw two balls in the air

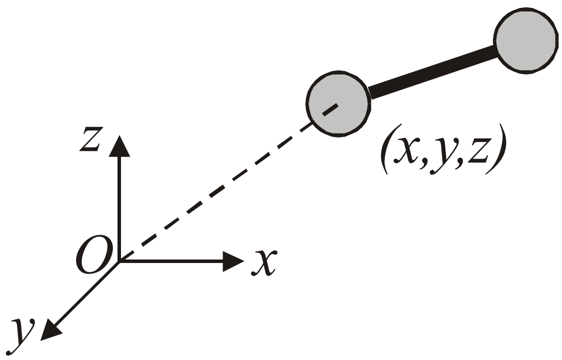


I wish to describe the position of each ball. The balls are travelling independently of one another so I require three coordinates for each ball- that is six coordinates in all.

Now imagine that I attach a rod between the balls.



This rod acts as a constraint on the system and reduces the freedom of the balls to move independently of one another, the position of one of the balls is fixed.



Then the position of the second ball is such that it must lie on a sphere whose centre is the position of the first ball. So I no longer require three coordinates to describe the position of the second ball – two will do.

The constraint has caused a reduction in the total degrees of the system by one.

In statistics there are frequent instances of where a constraint causes a reduction of the number of degrees of freedom of a system.
Suppose I take a sample of size 100 and construct a grouped frequency table with six intervals.

Interval	I_1	I_2	I_3	I_4	I_5	I_6
Frequency	F_1	F_2	F_3	F_4	F_5	F_6

As we fill up the table the frequency F_1 does not constrain F_2 to any particular value, but if we have values for F_1 F_2 F_3 F_4 F_5 then the constraint that the total frequency is 100. $F_1 + F_2 + F_3 + F_4 + F_5 + F_6 = 100$ means that the value of F_6 is determined by the other five values.

$$F_6 = 100 - (F_1 + F_2 + F_3 + F_4 + F_5)$$

So with this grouped frequency table with six columns has only five degrees of freedom, since the requirement that the total frequency = Sample size reduces the degrees of freedom by one. In general a frequency table with n rows or columns will have n-1 degrees of freedom.

Degrees of freedom in a contingency table

A contingency table shows the results from taking a sample where each individual object in the sample can have two or more properties. For clarity we concentrate on just two properties. For example athletes should be considered for their muscle fibre and their type of sporting event.
Suppose, for simplicity, athletes can have either slow twitch or fast twitch muscle fibres and be entered for either explosive or endurance sports

		Sport type	
		Explosive	Endurance
Fibre	Slow Twitch		
	Fast Twitch		

This is a 2×2 contingency table. Now imagine that 120 athletes are sampled, of which 53 have slow twitch and 56 participate in endurance sports. These values determine the other row and column totals.

		Sport type		
		Explosive	Endurance	Totals
Fibre	Slow Twitch			53
	Fast Twitch			67
	Totals	56	64	

The circled values are constrained by the other value(s) and the sample size.
This also means that filling in any one box in the contingency table determines al other values.

Sport Type

		Sport type		
		Explosive	Endurance	Totals
Fibre	Slow Twitch	5	48	53
	Fast Twitch	51	16	67
	Totals	56	64	120

If, for example, 5 athletes with slow twitch fibres participate in explosive sports then the other frequencies are determined by the constraints that the sum of the values in each row has to equal the row total, and likewise for the columns.

So a 2×2 contingency table with n columns and m rows has

Degrees of freedom $= \nu = (n-1)(m-1)$