# Re: 'trend value' formula needed

***Source:*** http://sci.tech−archive.net/Archive/sci.math/2007−01/msg00914.html


- *From*: matt271829−news@xxxxxxxxxxx


- *Date*: 5 Jan 2007 18:18:18 −0800


khany wrote:

> matt271829−news@xxxxxxxxxxx wrote:
>
> > khany wrote:
> >
> > > matt271829−news@xxxxxxxxxxx wrote:
> > >
> > > > khany wrote:
> > > >
> > > > > Hi all,
> > > > >
> > > > > hope this isn't too off−topic
> > > > > but i could really do with
> > > > > some help
> > > > >
> > > > > i have, say, at least 5 records
> > > > > with 3 columns of data
> > > > > based on
> > > > > industry activity (so
> > > > > not pre−determined by me)
> > > > >
> > > > > the values of the first 2
> > > > > columns roughly reflect the
> > > > > value of the
> > > > > third.
> > > > >
> > > > > i want to be able to pass in 2
> > > > > arbitrary values for the first
> > > > > 2 columns
> > > > > and get a trend/predicted
> > > > > value for the third back.
> > > > >
> > > > > eg. some records I already
> > > > > have:

A:22000, B:24, C: 5000
A:36000, B:35, C: 3200
A:56000, B:43, C: 1800

so as A and B increase (not directly related) C decreases.

So if I pass in A:39000 and B:37 what would C be? more importantly, whats the formula?

thanks in advance

khany

So, you have two variables A and B, and you think that C is approximately determined by A and B. Based on the data you have, you want to find an equation "C = some calculation involving A and B" that gives the best prediction of C. There is no "formula" that will magically do this for you in all cases; instead there are techniques, which generally come under the name of "regression analysis" (Googling this should throw up plenty of info).

However, if you know, or can reasonably assume, that the "calculation involving A and B" approximately follows some particular form − for example C = p*A + q*B + r for some numbers p, q and r to be determined − then life is potentially easier. Do you have an idea about what the nature of relationship between A, B and C ought to be, or are you in the dark? (Knowing that C decreases as A and B increase is not really specific enough; it could decrease with A and B in any of innumerable ways.)

OK, I can see I need to pass more information to you all. I really

> appreciate this.
>
> The records represent cars that have sold and column A is the mileage
> of the car and B is the number of months old the car is at the point of
> sale. So based on existing records, I want to pass mileage and age and
> receive in return an approximate value or even value range (high|low).
>
> hope this helps

> I assume that you also know the make and model and, in particular, the original price of the car when it was new. (If not then any attempt to calculate value from age and mileage alone will result in completely meaningless numbers.)
>
> I think your best bet is to try to find out how the motor industry does this. There may be standard formulas that everyone uses. Then do a sanity check that these are an acceptable match with the data you have, and tweak if necessary.
>
> I really know nothing about the used car trade, but for an average−mileage vehicle, as a first guess, it seems not unreasonable that resale_price = original_price * (1 − yearly_depreciation_rate) ^ age. You would need to figure out the depreciation rate, which I suspect might vary significantly with model. You might find published estimates of depreciation rates, and/or infer them from your data, provided you have enough of it. Then you would need to adjust somehow for above−average or below−average mileage. Not quite sure how that would work. But, as I say, I'm just guessing, and finding out what method the experts use would be the way to go I think.

The problem with the industry figures i was using was that they were out, way out, and so i wanted to find a statistical way of determining the information. you're right about determining the depreciation index but i was hoping the data itself could establish that.

actually, that could be the key. most cars have a half life value of about 3 years. if the data could establish the dep index then perhaps your formula might work.

i see this as a three dimensional graph. x = age, y = mileage and z = value. any ideas how to apply a formula like this?

I'm not sure about the most sensible way to handle the mileage element,

but ignoring that for a moment, let p = price when new, v = resale value, r = annual depreciation rate (as a fraction; e.g. 10% = 0.1), and t = age in years. The standard formula would be v = p*(1 − r)^t. Taking logarithms we get

log(v) = log(p) + t*log(1 − r)

which means that log(v) is a linear function of t − i.e. the graph of log(v) against t is a straight line. The log(v)−axis intercept is log(p), and the slope of the line is log(1 − r), from which you can recover an estimate of r. Getting the best fit straight line to the data is an application of linear regression, and it would be usual to use "least−squares fitting". This is not hard: Google for "linear regression" and/or "least squares" and take your pick of the explanations (http://www.tufts.edu/~gdallal/slr.htm was one of the first I clicked on at random). If the data produces nothing like a straight line, and you're confident of the data, then this formula is not appropriate. One possible glitch is that the formula would give the value of a week−old car as very near the as−new price, whereas in practice cars can drop in value quite considerably as soon as the buyer drives out of the showroom. Remember also that this is per−model − each different model of car potentially has a different value of r, and most likely a different value of p.

I was idly wondering if there might be a similar concept of "mileage depreciation rate". This would mean, for example, that every additional ten thousand miles decreases the value by 5%, or whatever it might be. So, letting m be the mileage, and d be the "mileage depreciation rate", we would have something like v = p * (1 − r)^t * (1 − d)^m. Taking logs gives log(v) = log(p) + t*log(1 − r) + m*log(1 − d). This means that log(v) is a linear function of t and m (i.e. the 3D graph is a plane), and you could use a similar least−squares fitting technique to get the best−fit plane, and thereby estimate the quantities you want to know, namely r and d. Again, you'd need to check the data to see if this model is at all plausible.

.