ABSTRACT

Title : A method for comparing standard deviations precisely

Abstract : A specific relation often appears between the mean and standard deviation of variables collected in social sciences suveys and research. This article proposes a description and a possible correction of the phenomenon.

Key Words : Standard deviation, Mean, Correction.

A method for comparing standard deviations precisely

Problem

Although most of the research efforts in statistics concentrate on sophisticated multivariate analysis, a great part of the actual exploitation of surveys simply lies in basic univariate calculation, either because such data are easier to compute, or because they are easier to use as arguments in reports or studies. While newer multi-dimensional techniques developed by professional researchers may well be full of brilliant ideas and find relevant applications in particular fields, the manager of a small or medium-sized structure will often find it easier and more understandable to trust the main statistical indicators whose meaning he can "feel" : in the case of numerical variables, the mean and standard deviation (or variance).

Therefore, it is not wasted time to study what link may exist between those two indicators. When studying the statistics literature though, it is surprising to find no mention about a phenomenon that analysts often encounter when computing simple statistics on basic attitudinal data : a strong, curvilinear relation between the mean and standard deviation of the variables measured. The absence of any mention of this relation in textbooks or research papers is all the more stunning in that a large amount of work has been published on the subject of personality measurement for several decades (a basic review of the literature is given at the end of the article to support this idea).

A purely algebraic relation, of course, exists at first between the two indicators, as the formula for standard deviation appeals to that of the mean. But what deserves interest is that this relation goes much further, and therefore needs a more thorough definition.

To understand it briefly, let us imagine a survey in which a set of scaled opinions concerning various items (for instance, purchase intentions for different brands of foods) has been collected. The first reflex of the analyst will be to compute the mean and standard deviation of each item, so as to be able to classify them in rank order (in relation to the mean), and to check which are especially consensual (low standard deviation) or not (high standard deviation). The standard deviation will at that time be considered for its absolute value, which is not problematic as long as the univariate analysis remains to this point.

If the researcher starts to study in detail which items obtain the lower standard deviation though, he will probably tend to find them systematically at the extreme sides of the mean (that is, among the items either best or worst ranked). This is easy to justify intuitively: those items that get the most extreme mean values are also necessarily the most consensual -almost every evaluator ranks them in the same way (which both brings an extreme mean and a low standard deviation). At the same time, and for opposite reasons, the higher values for the standard deviation will usually correspond to medium values for the mean. Therefore, plotting the standard deviation against the mean will make a special pattern appear, which, far from being a neutral oval, will more or less look like a circumflex accent (high in the middle, low at the extremes).

Now this might cause some surprise to the 'naive' analyst, for such a relation, although common with unstandardized attitudinal variables, is barely if ever mentioned in the specialized literature. Hence it is neither explained, nor taken into account for the treatment of variables. Let us now give some empirical evidence of the phenomenon.

Example

A survey was conducted at Audencia Nantes Business School to evaluate the immediate opinion of students (positive or negative) about various animals (real or imaginary) and meats. Opinions were measured using five-point scales presented in a row, to be filled in manually. Respondents were 130 undergraduate French students in management, unaware of the objectives of the experiment. 250 items were evaluated, including 'horse', 'wasp' or 'shark' as real animals, 'Bambi' or 'Mickey Mouse' as imaginary animals, and 'lamb', 'chicken' or 'roastbeef' as meats.

The five points of the scale were accounted for 0, 0.25, 0.5, 0.75 and 1 to simplify the subsequent calculations. Here are an abstract of the results, and a diagram showing the relation between the mean (*m*) and standard deviation (*s*) (see table 1 and figure 1).

Table 1 - Mean, observed standard deviation and corrected standard deviation for 132 animals present in the Audencia survey

Figure 1 - Diagram showing the relation between m and s

The pattern on the diagram is explicit. As usual with this kind of data collection procedure, the 'circumflex accent' appears clearly, showing the lower standard deviation values for those items which bring the most extreme mean values (ie. 'Mosquito', among the worst ranked, or 'Dolphin', best ranked) and the higher standard deviation values for mildly evaluated items (for instance 'Shark'). Such a pattern might not be caused by a systematic misuse of the measurement scales. No major distortion from a normal distribution appears in the histogram of global frequencies of the five proposed categories (see Figure 2), and a detailed review does not bring any suspicion concerning the validity of the data.

Figure 2 - Histogram of frequencies of each category (for all items)

In absolute terms, it is not incorrect to state that the lower dispersion is to be found with the items that are extremely highly or badly evaluated. But such a statement does not take into account the fact that while this phenomenon is compulsory, it is not necessarily informative (in Shannon's acceptance of the term). As long as the data are not corrected, such items as 'Goat', for instance, would remain unnoticed from the point of view of standard deviation, while a deeper look at the question would show that they deserve some special consideration.

Correction

One can first observe that a definite maximum value for the standard deviation corresponds to every different value for the mean. For instance, a 0 value for an item's mean can only be obtained by a perfect consensus among the evaluators (every one ranking 0), consequently also bringing a 0 value for the standard deviation. The same assertion is true for a 1 value for the mean, only possibly a result of a perfect consensus and a 0 value for the standard deviation.

Now if we want to determine the maximum possible value for the standard deviation of an item that has a mean of 0.5, we just have to see that this maximum value is obtained if half of the evaluators give 0, and the other half give 1. In this case, the standard deviation is : $s = 0.5$. By extension, it can be shown that for any value m for the mean, the maximum possible value for the standard deviation is obtained when a proportion of m evaluators give 1, and a proportion of $(1-m)$ evaluators give 0. In that case :

$$s = (m.(1-m))^{1/2}$$

Therefore, it is possible to enrich our previous diagram with the curve showing the maximum possible values for the standard deviation all along the interval [0-1] (see figure 3).

Figure 3 - Diagram showing the relation between m and s

One can immediately see that the distribution of the points on the graph is grossly parallel to the upper curve, which appears as an extension limit to the cluster. It is then just natural to imagine a specific correction of the data that would eliminate the curve effect, inadequate for a quick comparative reading of the values shown on the graph. This correction consists in simply dividing for each item the absolute value *s* of its standard deviation by the maximum possible value corresponding to its mean : $(m.(1-m))^{1/2}$, which gives the formula :

$$s' = s / (m.(1-m))^{1/2}.$$

In that case, we obtain the following diagram (see figure 4).

Figure 4 - Diagram showing the relation between m and $s' = s / (m.(1-m))^{1/2}$

Hence it is possible to make a clearer comparison between the high and the low values for the standard deviations of various items. The most extreme points from the point of view of the mean do not any longer show any particular inclination for small standard deviations. Now while some of the higher values for standard deviations are slightly modified, the most important change comes from the bottom center of the cluster, where such points as 'Goat' appear to be the more consensual, which was not the case before.

From a general point of view, what is the benefit of such a correction ? Simply (but it is extremely valuable) to allow the researcher to compare standard deviations free of trend, which otherwise would be impossible.

In what other cases is such a phenomenon susceptible to appear, and such a correction appropriate to use ?

Basically, in any case where a large number of opinions or attitudes has been collected with the use of rating scales, whatever the coarseness of the scales. A significant amount of research in marketing, especially in the field of consumer behavior or methodological issues, but also in psychology, sociology, or in the social sciences in general, makes extensive use of such scales, and may therefore benefit from this correction before any analysis of the standard deviations.

Here are five examples, taken from various fields, where the graphs mean/standard deviations both with and without correction are shown, to illustrate how large its range of application might be.

Literature example 1

Source: Riker (1944).

Field: Attitude measurement in the 1940's (psychology), based on research of Thurstone and Likert.

Variables measured: 28 assertions about political and social issues (for example, 'The Negro should be given the same educational advantage as the white man').

Collecting technique: 11 category-verbal intensity scales and graphic scales.

Methodological note: Original data reported in the article, quoted 1 to 11, have been standardized to interval [0-1] to facilitate comparison with other works.

Major changes brought by the correction: Item considered most consensual; global consensuality of best-ranked items.

Figure 5 - Relation between m and s in Literature Example 1 (Uncorrected Original Data)

Figure 6 - Relation between m and s in Literature Example 1 (Corrected Data)

Literature example 2

Source: Thomas and al. (1997).

Field: Administrative science

Variables measured: 23 assertions about organizational culture measured in the purchase division of a company.

Collecting technique: 7-point Likert scales.

Methodological note: Original data reported in the article, quoted 1 to 7, have been standardized to interval [0-1] to facilitate comparison with other works.

Major changes brought by the correction: Item considered most consensual.

Figure 7 - Relation between m and s in Literature Example 2 (Uncorrected Original Data)

Figure 8 - Relation between m and s in Literature Example 2 (Corrected Data)

Literature example 3

Source: Morokoff and al. (1997).

Field: Personality measurement

Variables measured: 36 assertions about sexual behavior and attitudes.

Collecting technique: 5 point-verbal intensity scales (from 'strongly agree' to 'strongly disagree').

Methodological note: Original data reported in the article, quoted 1 to 5, have been standardized to interval [0-1] to facilitate comparison with other works.

Major changes brought by the correction: Items considered most and least consensual; considerable changes among the best-ranked items.

Figure 9 - Relation between m and s in Literature Example 3 (Uncorrected Original Data)

Figure 10 - Relation between m and s in Literature Example 3 (Corrected Data)

Literature example 4

Source: Aaker (1997).

Field: Marketing research.

Variables measured: 42 brand personality traits (for example, 'sincere', 'unique', 'masculine', ...)

Collecting technique: 5 point-Likert scales.

Methodological note: Original data reported in the article, quoted 1 to 5, have been standardized to interval [0-1] to facilitate comparison with other works.

Major changes brought by the correction: 3 items considered least consensual; item considered most consensual.

Figure 11 - Relation between m and s in Literature Example 4 (Uncorrected Original Data)

Figure 12 - Relation between m and s in Literature Example 4 (Corrected Data)

Literature example 5

Source: Conway et al. (1996).

Field: Cognitive psychology.

Variables measured: 58 encoding and retrieval performances about true and false memories.

Collecting technique: 5 point and 3 point-scales (1=low to 3 or 5=high).

Methodological note: Original data reported in the article, quoted 1 to 3 or 1 to 5, have been standardized to interval [0-1] to facilitate comparison with other works.

Major changes brought by the correction: Items considered most and least consensual; Considerable change among the lowest-ranked items.

Figure 13 - Relation between m and s in Literature Example 5 (Uncorrected Original Data)

Figure 14 - Relation between m and s in Literature Example 5 (Corrected Data)

Conclusion

Although all the examples listed above are taken from the social sciences, the suggested correction could be useful as well for data coming from other fields, including natural science, assuming that the measures studied are bound between two precise limits (which means that they fundamentally have the nature of a proportion). In climatology for instance, let us imagine that we examine the number of rainy days in September in different parts of the world throughout the century. If we observe a mean of 0,5 rainy days in Windhoeck (Namibia) and a mean of 30 days in Darjeeling (India), we can be sure that the corresponding standard deviations will be extremely low, whereas a mean of 10 days in Paris (France) or Tokyo (Japan) can be associated with either a moderate or a large standard deviation. In such a case, the proposed correction only could lead to an objective comparison of the items' consensuality.

In that respect, the range of application of the proposed technique is very wide and goes past the usual calculation of consumer, citizen, social subject or personality measurement. It concerns virtually every research where elementary statistics have an interest in themselves. It might therefore be considered in any case where numerous standard deviations are to be studied and compared.

References

Historical bibliography :

Cox, E. P. (1980) The Optimal Number of Response Alternatives for a Scale: a Review. *Journal of Marketing Research*, 17, pp. 407-422.

Green, P. E., Rao, V. R. (1970) Rating Scales and Information Recovery - How Many Scales and Response Categories to Use. *Journal of Marketing*, 34, pp. 33-39.

Likert, R. (1932) A technique for the measurement of attitude. *Archives of psychology*, 140, 1-55.

Osgood, C. E., Suci, G. J., Tannenbaum, P. (1957) *The measurement of Meaning* (Chicago: Chicago University Press).

Thurstone L. L., Chave, E. J. (1929) *The measurement of attitude* (Chicago: Chicago University Press).

Torgerson, W. J. (1958) *Theories and methods of scaling* (New York :John Wiley and sons).

References of the experiments used in the article :

Aaker, J.L. (1997) Dimensions of Brand Personality. *Journal of Marketing Research*, 34, pp. 347-356

Conway M.A., Collins A.F., Gathercole S.E., Anderson S.J. (1996) Recollections of True and False Autobiographical Memories. *Journal of Experimental Psychology: General*, 125, 1, pp. 69-95

Morokoff P.J., Quina K., Harlow L.L., Whitmire L., Grimlay D.L., Gibson P.R. and Burkholder G.J. (1997) Sexual Assertiveness Scale (SAS) for Women : Development and Validation. *Journal of Personality and social Psychology*, 73, 4, p. 79

Riker B. L. (1944) A Comparison of Methods Used in Attitude Research. *Journal of Abnormal Social Psychology*, 39, pp. 24-42

Thomas G., Hult M. and Ferrell O. C. (1997) Global Organizational Learning Capacity in Purchasing: Construct and Measurement. *Journal of Business Research*, 40, 2, p. 111

	m	s	s'		m	s	s'
Bee	0,49	0,31	0,62	Rabbit	0,80	0,20	0,51
Lamb	0,74	0,22	0,51	Greyhound	0,57	0,25	0,50
Eagle	0,66	0,31	0,65	Lizard	0,41	0,32	0,65
Donkey	0,50	0,26	0,52	Dragonfly	0,54	0,28	0,56
Spider	0,16	0,26	0,72	Hare	0,65	0,21	0,44
Maggot	0,16	0,21	0,59	Lion	0,72	0,30	0,67
Ostrich	0,56	0,23	0,47	Wolf	0,55	0,34	0,69
Whale	0,73	0,23	0,53	Mammoth	0,47	0,28	0,57
Ram	0,52	0,24	0,47	Mackerel	0,33	0,23	0,49
Big dog	0,54	0,33	0,67	Marmot	0,76	0,24	0,55
Doe	0,77	0,21	0,50	Jelly fish	0,13	0,19	0,57
Bison	0,50	0,24	0,49	Whiting	0,46	0,25	0,51
Boa	0,28	0,31	0,69	Sparrow	0,58	0,24	0,48
Ox	0,53	0,24	0,49	Fly	0,17	0,20	0,54
Ewe	0,64	0,25	0,52	Seagull	0,51	0,30	0,59
Duck	0,63	0,26	0,53	Mouffette	0,42	0,22	0,45
Poodle	0,43	0,33	0,68	Mussel	0,44	0,28	0,57
Camel	0,64	0,22	0,46	Mosquito	0,09	0,16	0,56
Cat	0,71	0,30	0,66	Sheep	0,63	0,25	0,52
Kitten	0,82	0,27	0,69	Goose	0,48	0,24	0,48
Bat	0,28	0,29	0,66	White bear	0,79	0,24	0,59
Caterpillar	0,26	0,24	0,55	Black bear	0,71	0,26	0,57
Horse	0,80	0,22	0,55	Panda	0,87	0,20	0,59
Goat	0,57	0,22	0,45	Peacock	0,65	0,26	0,55
Roe deer	0,66	0,21	0,44	Butterfly	0,75	0,25	0,58
Chimpanzee	0,64	0,26	0,55	Night butterfly	0,45	0,33	0,65
Puppy	0,83	0,22	0,58	Partridge	0,55	0,24	0,47
Stork	0,68	0,22	0,47	Parrot	0,71	0,24	0,54
Ladybird	0,77	0,21	0,50	Seal	0,74	0,25	0,57
Pig ('cochon')	0,37	0,27	0,56	Octopus	0,29	0,29	0,64
Big Hamster	0,46	0,31	0,63	Pigeon	0,44	0,30	0,61
Cocker spaniel	0,65	0,31	0,65	Penguin	0,76	0,21	0,50
Dove	0,73	0,24	0,55	Guinea-fowl	0,41	0,24	0,49
Cock	0,52	0,25	0,50	Gold fish	0,61	0,26	0,54
Crow	0,27	0,23	0,53	Pony	0,71	0,26	0,56
Crab	0,46	0,27	0,54	Pig ('porc')	0,34	0,25	0,54
Toad	0,22	0,25	0,60	Foal	0,80	0,21	0,52
Shrimp	0,56	0,27	0,55	Hen	0,52	0,28	0,55
Crocodile	0,32	0,27	0,58	Chicken	0,54	0,29	0,59
Swan	0,70	0,27	0,59	Chick	0,77	0,27	0,63
Dolphin	0,92	0,15	0,55	Ray	0,41	0,26	0,53
Turkey	0,46	0,25	0,50	Rat	0,14	0,23	0,66
Dinosaur	0,49	0,34	0,68	Fox	0,57	0,29	0,59
Squirrel	0,82	0,21	0,56	Shark	0,34	0,36	0,75
Elephant	0,76	0,21	0,50	Rhinoceros	0,47	0,27	0,54
Snail	0,47	0,28	0,55	Dog ('Saint-Bernard')	0,83	0,24	0,63
Starfish	0,58	0,26	0,53	Boar	0,43	0,27	0,54
Pheasant	0,54	0,24	0,47	Sardine	0,45	0,28	0,56
Fawn	0,71	0,27	0,59	Salmon	0,66	0,27	0,57
Stone marten	0,33	0,27	0,58	Scorpion	0,22	0,31	0,74
Ant	0,40	0,30	0,61	Sole	0,54	0,27	0,55
Gazelle	0,76	0,20	0,46	Mouse	0,37	0,29	0,60
Girafe	0,72	0,19	0,42	White mouse	0,47	0,31	0,61
Gorilla	0,48	0,26	0,52	Bull	0,50	0,26	0,53
Frog	0,43	0,29	0,58	Tadpole	0,34	0,26	0,55
Cheetah	0,66	0,31	0,65	Tuna	0,46	0,27	0,54
Wasp	0,18	0,24	0,62	Tiger	0,69	0,31	0,67
Hamster	0,53	0,27	0,54	Turtle	0,69	0,24	0,52
Hedgehog	0,62	0,28	0,57	Trout	0,63	0,24	0,49
Owl	0,50	0,29	0,57	Vulture	0,34	0,30	0,63
Hippopotamus	0,55	0,28	0,56	Calf	0,63	0,22	0,47
Swallow	0,65	0,24	0,49	Silkworm	0,41	0,30	0,60
Lobster	0,55	0,26	0,52	Earthworm	0,22	0,21	0,52
Oyster	0,42	0,29	0,58	Tapeworm	0,09	0,17	0,61
Kangaroo	0,79	0,19	0,47	Viper	0,14	0,24	0,70
Koala	0,84	0,20	0,54				

Table 1 - Mean, observed standard deviation and corrected standard deviation for 132 animals present in the Audencia survey

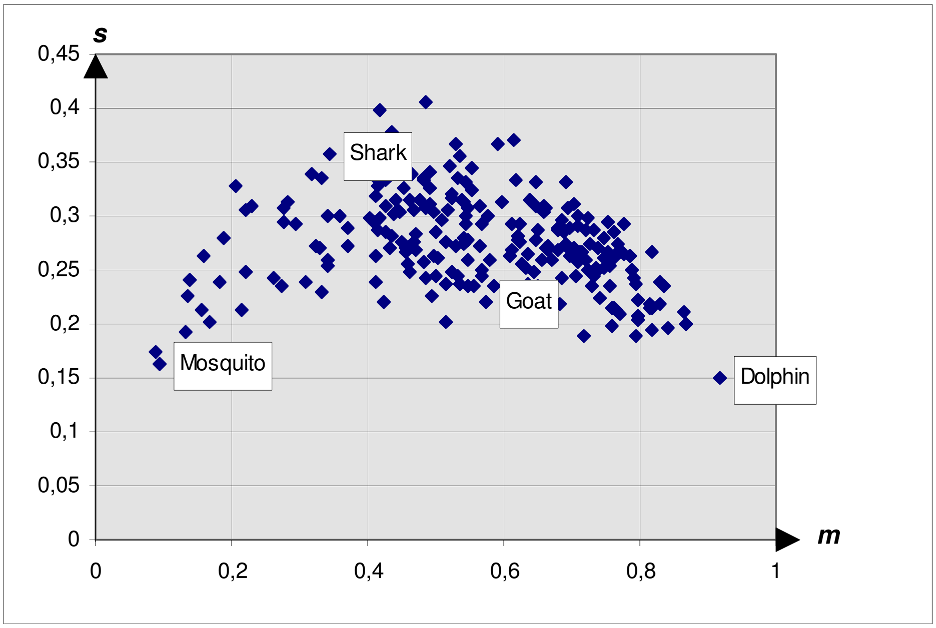

Figure 1 - Diagram showing the relation between m and s

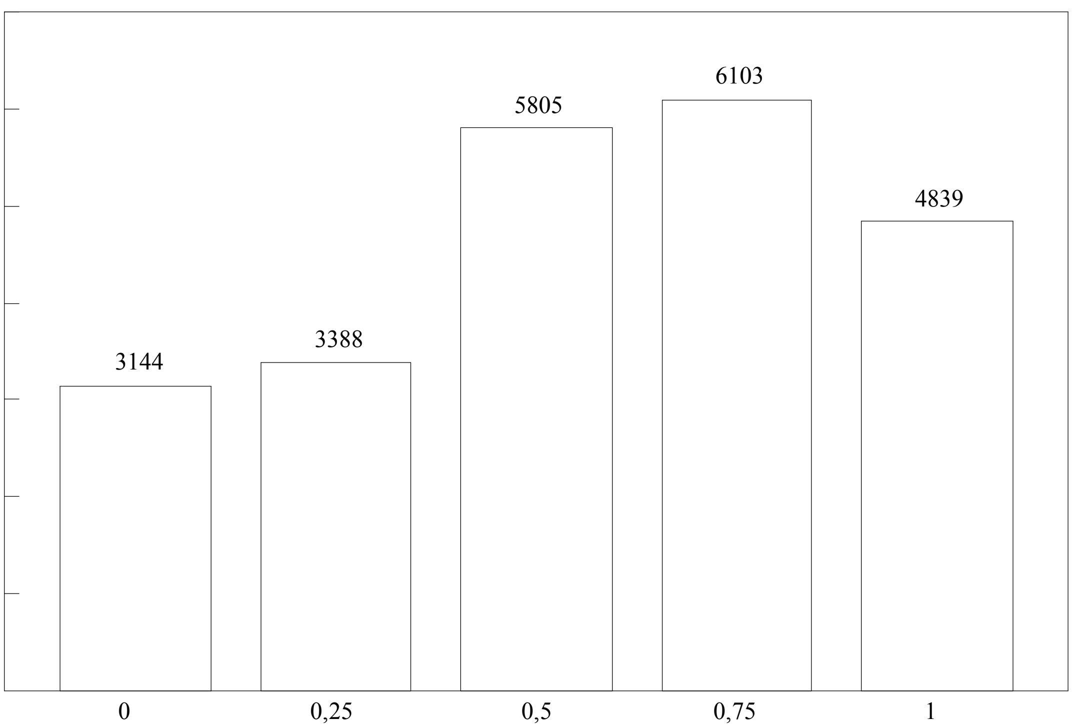

Figure 2 - Histogram of frequencies of each category (for all items)

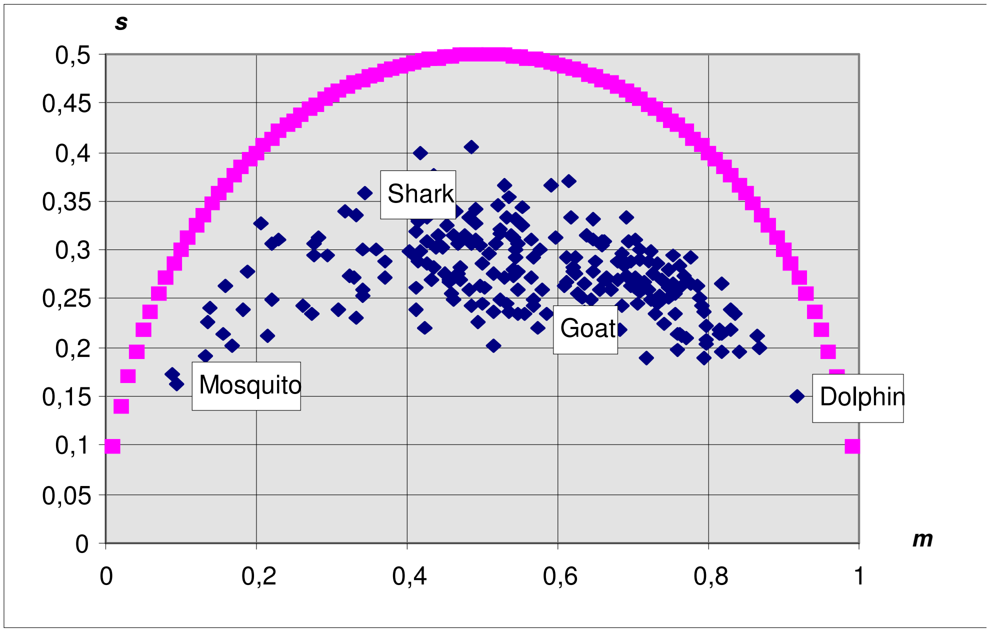

Figure 3 - Diagram showing the relation between m and s

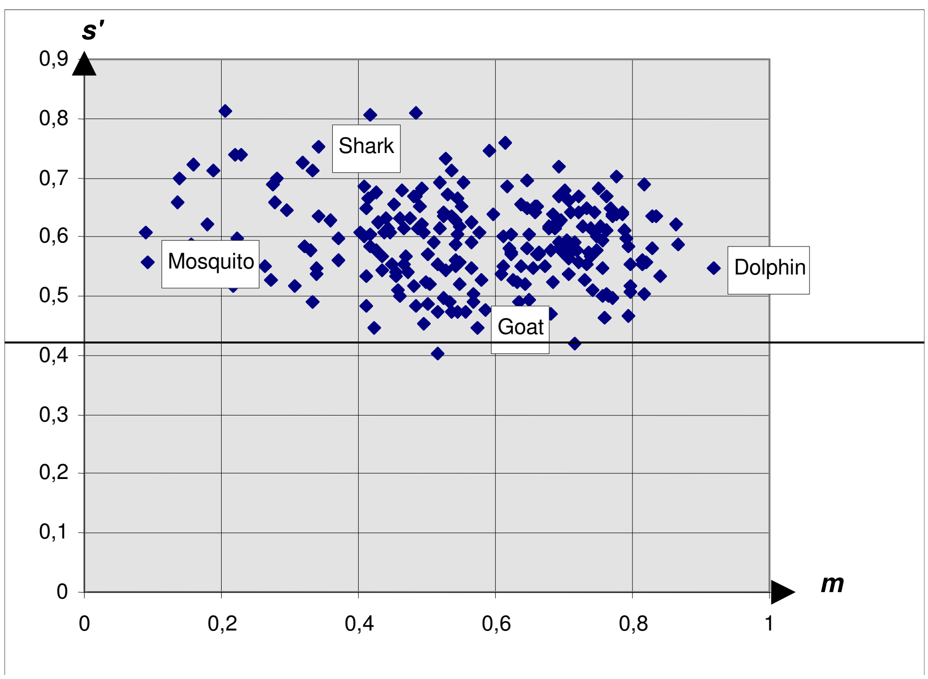

Figure 4 - Diagram showing the relation between m and $s' = s \,/\, (m.(1-m))^{1/2}$

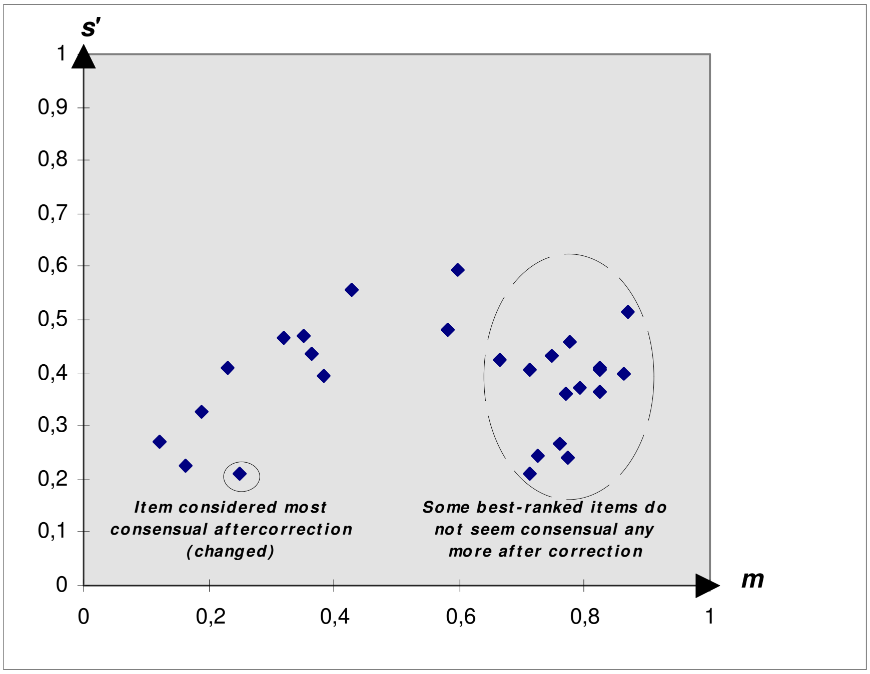

Figure 5 - Relation between m and s in Literature Example 1 (Uncorrected Original Data)

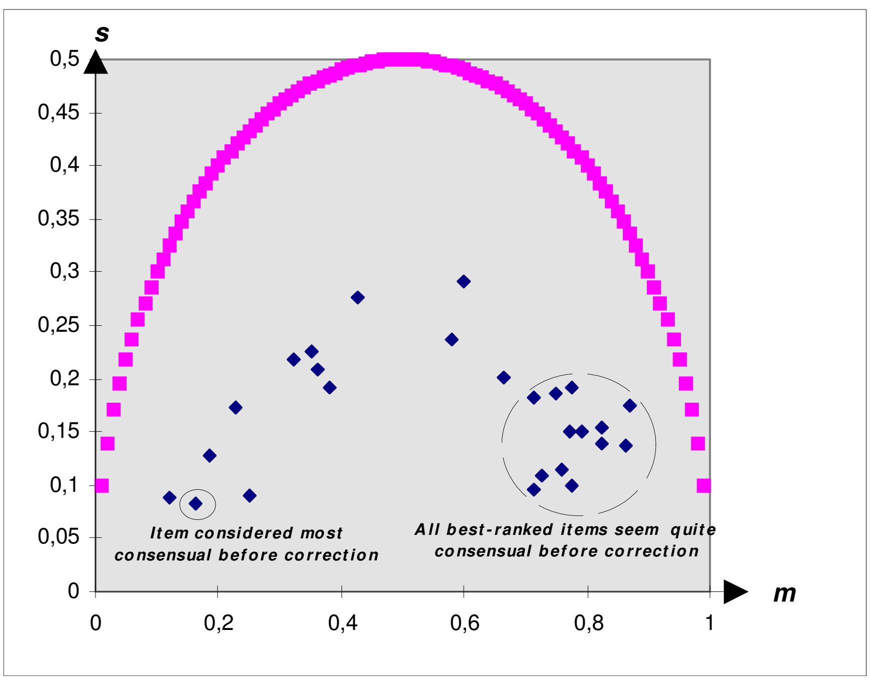

Figure 6 - Relation between m and s in Literature Example 1 (Corrected Data)

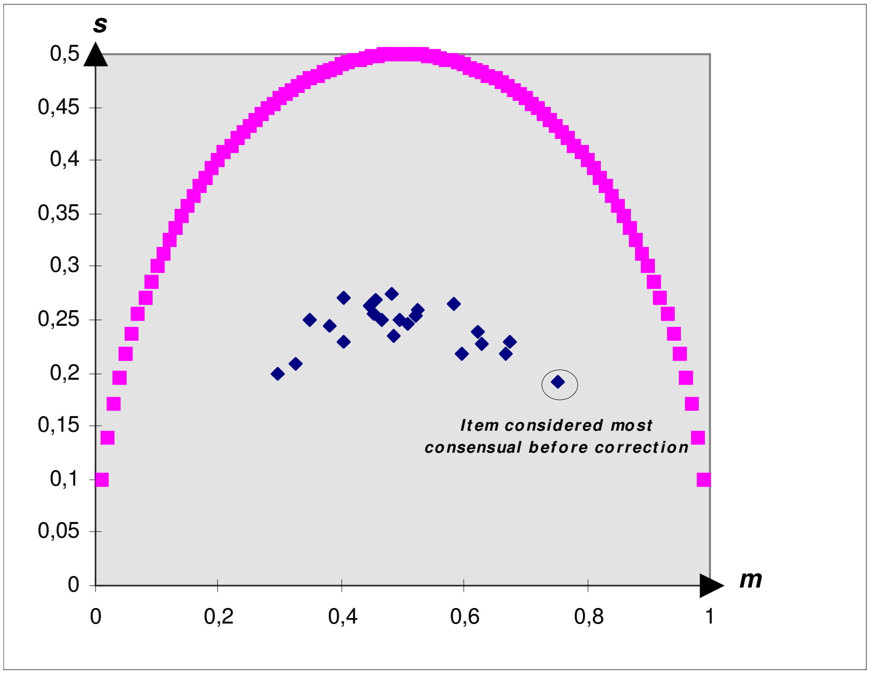

Figure 7 - Relation between m and s in Literature Example 2 (Uncorrected Original Data)

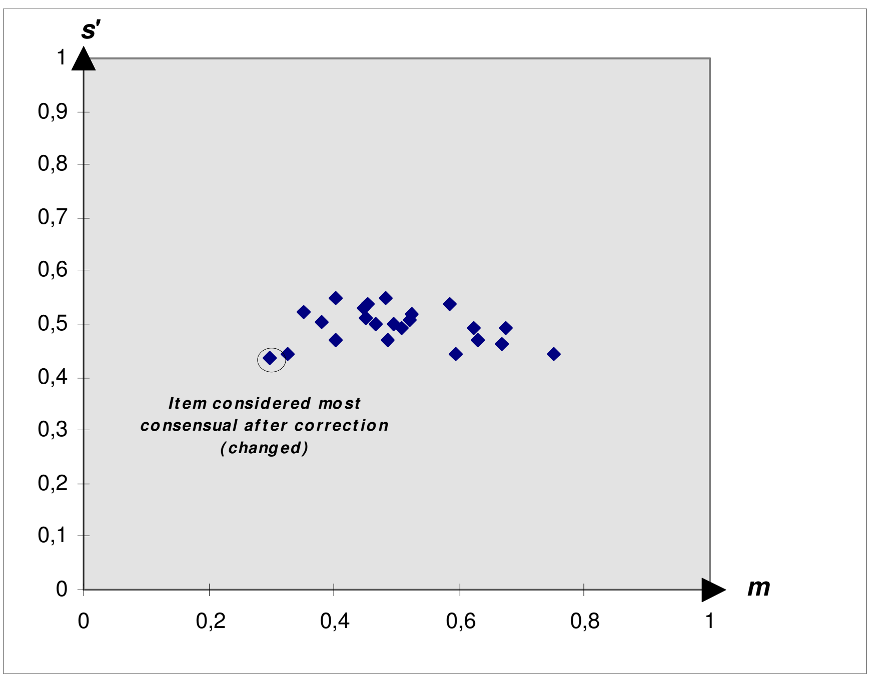

Figure 8 - Relation between m and s in Literature Example 2 (Corrected Data)

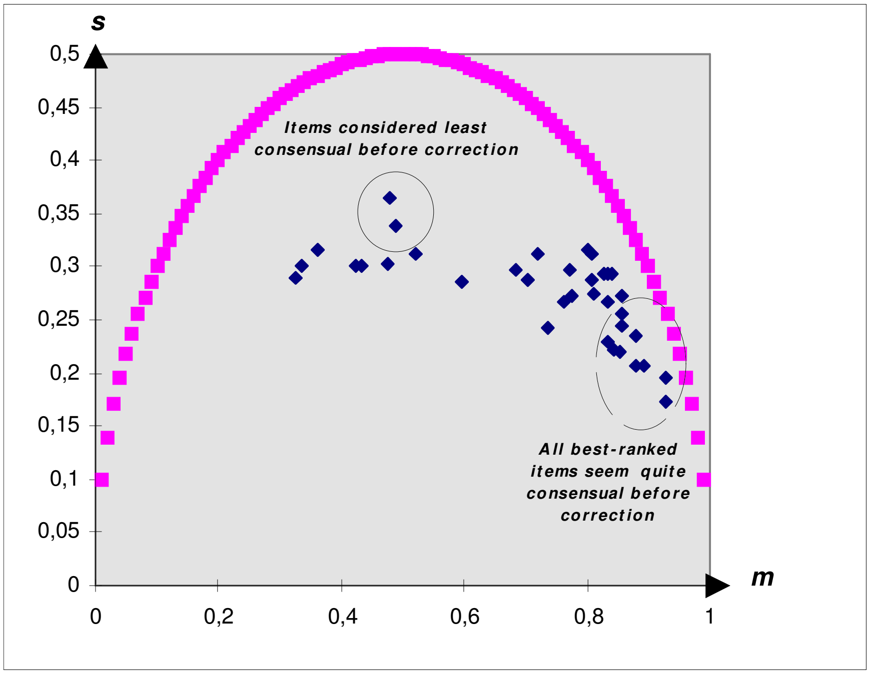

Figure 9 - Relation between m and s in Literature Example 3 (Uncorrected Original Data)

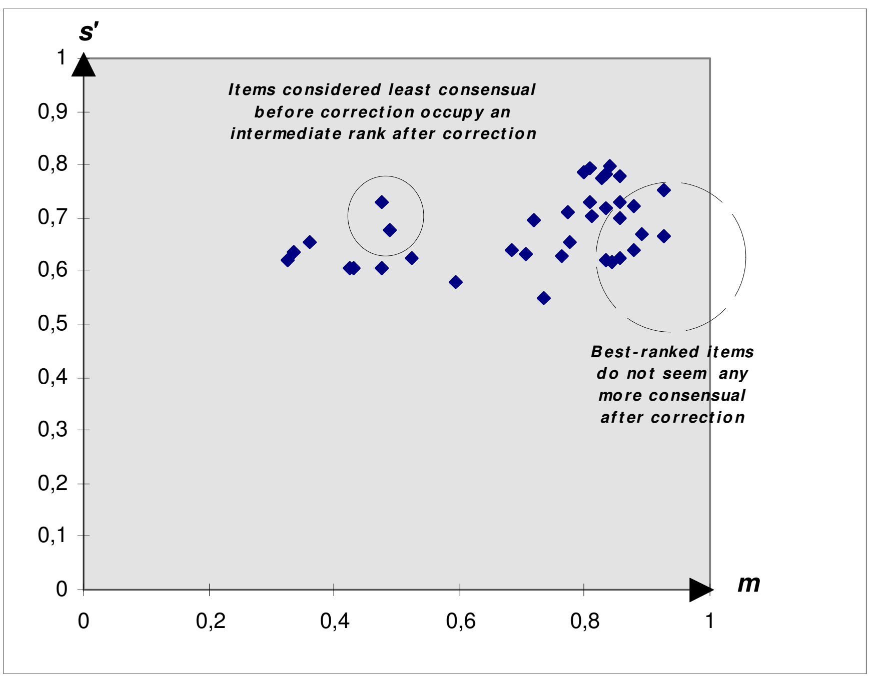

Figure 10 - Relation between m and s in Literature Example 3 (Corrected Data)

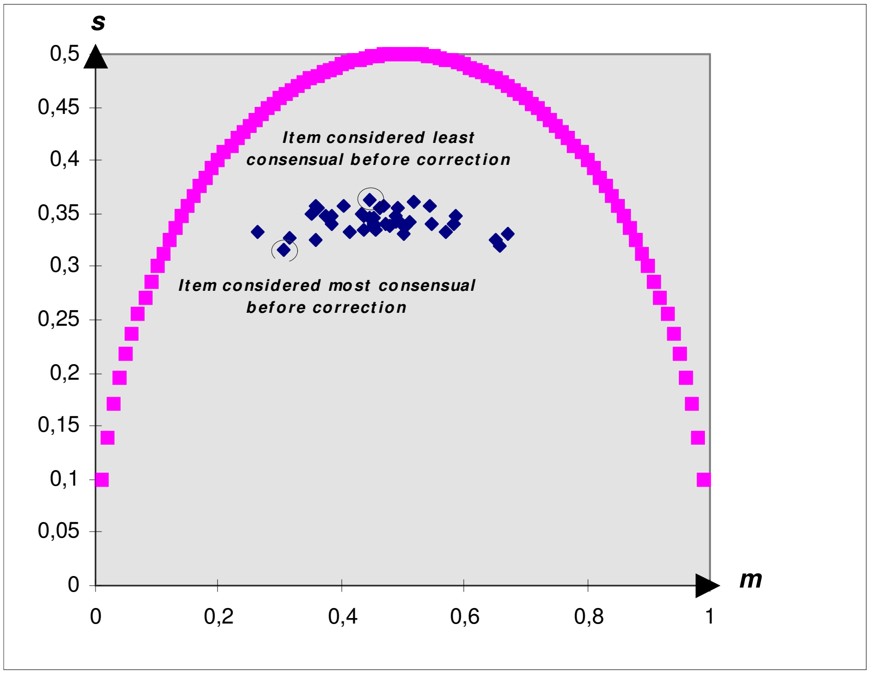

Figure 11 - Relation between m and s in Literature Example 4 (Uncorrected Original Data)

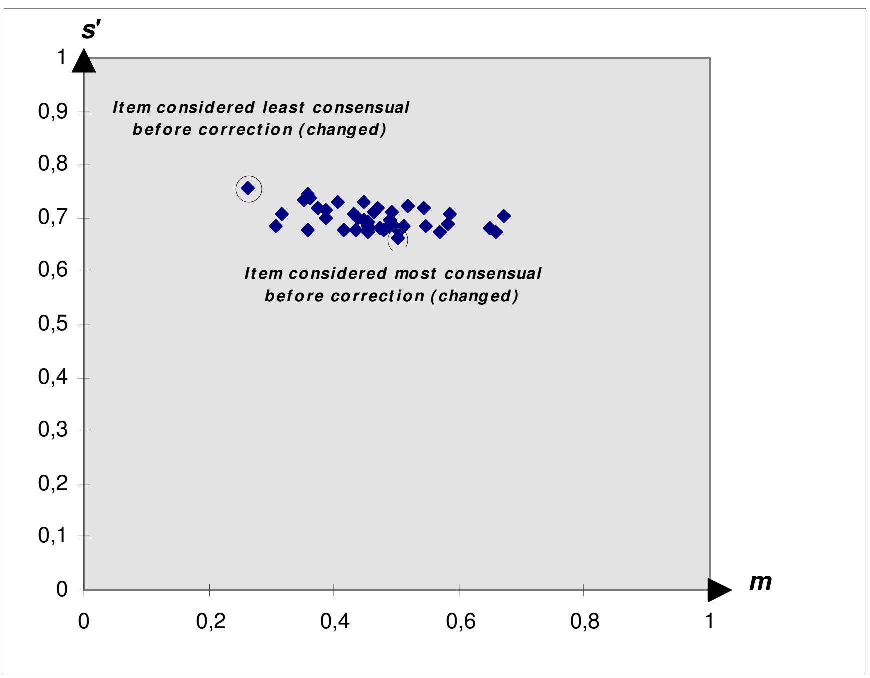

Figure 12 - Relation between m and s in Literature Example 4 (Corrected Data)

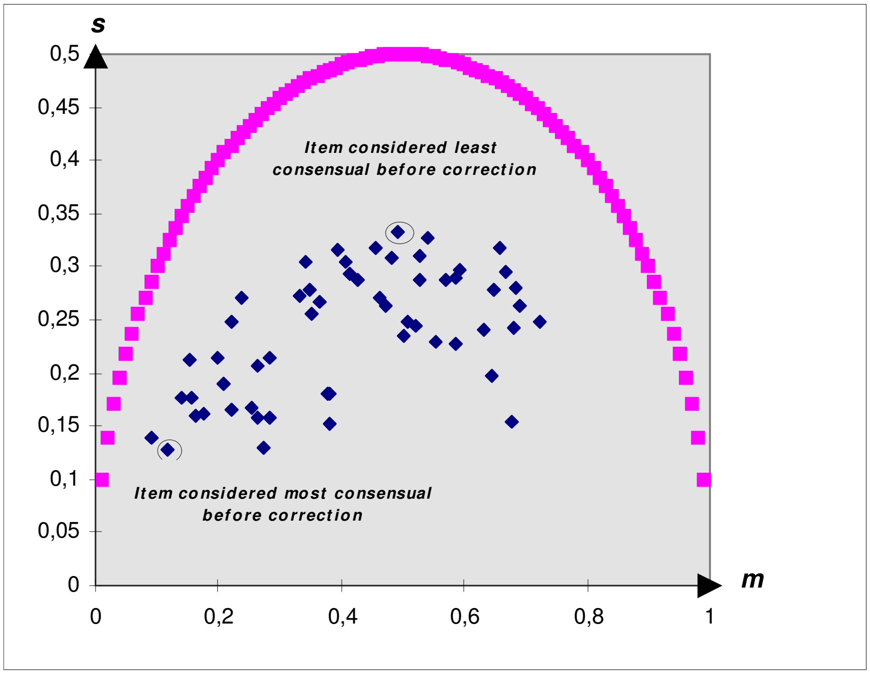

Figure 13 - Relation between m and s in Literature Example 5 (Uncorrected Original Data)

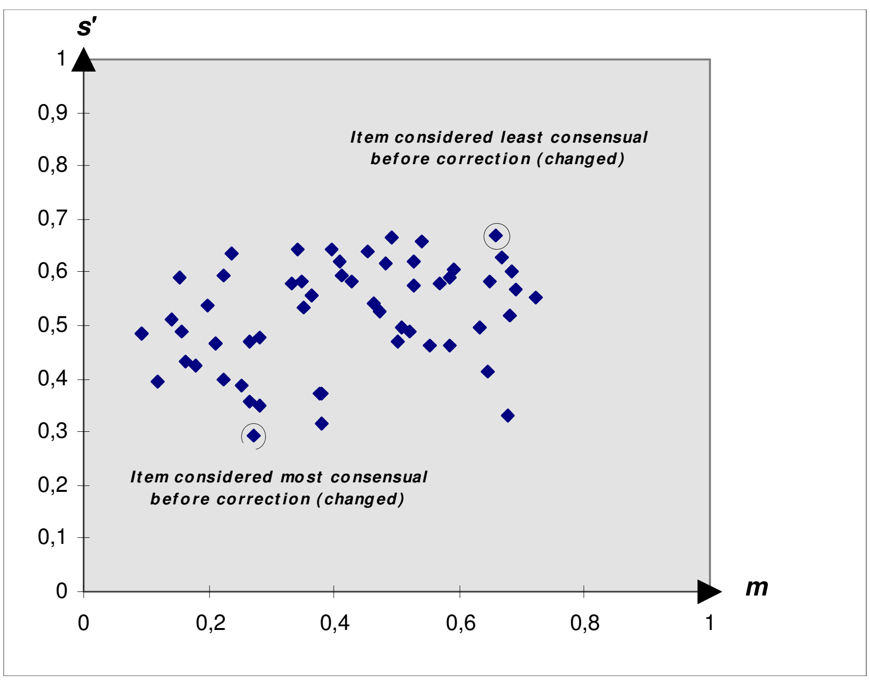

Figure 14 - Relation between m and s in Literature Example 5 (Corrected Data)