Cognadev Technical Report Series

15

10th January, 2020

Raw score data transformations

This report provides the rationale, computational information, and graphical examples associated with score transformation methodologies in use by Cognadev, other test publishers, and data scientists.

Cognadev UK Ltd., 2 Jardine House, Harrovian Business Village, Harrow, Middlesex, HA1 3EX. Registered UK # 05562662

Contents

1. Why Transform Scores?

This document explains the various transformations that can be applied to raw test scores computed from psychological assessment scales. I'll be using the terms 'variable', "scale score' or simply "score" interchangeably throughout. Psychologists mostly rely upon the additive metric properties of integer (whole number) or real-valued numbers (decimal/fractional numbers) when working with test scores, irrespective of whether the psychological attribute being assessed does indeed vary quantitatively[1]. So, for this exposition, I'll work within that status-quo framework. And, as we shall see below, some transformations or re-expression of raw scores purposefully lose metric information in order to gain more informative interpretability of raw score magnitudes.

1.1 For Mathematical Purposes

1.1.1 Averaging multiple scores

When wanting to produce an average of scores of several different variables or scales, where say each variable assessing a facet of Conscientiousness possesses a different score range (say 0-20, 15-75, 0-40), it's mistaken to form an average score based upon the scores from each scale because the magnitude of each scale score might be the same numerically (say 20), but the psychological meaning is vastly different. E.g. 20/20 on the first scale is the maximum possible score, 20/75 on the second scale is not only near the minimum score, but also misleading as the minimum possible score is 15, not 0, and 20/40 on the third scale is the middle score. An average score of 20 (20+20+20)/3) may be possible to compute arithmetically but is quite literally meaningless. The individual is scoring maximum on one scale, near the minimum on another, and at the mid-point on the third scale. To produce a valid mean score, the individual scores first need to be transformed/re-expressed in a common metric, then averaged.

1.1.2 Correlating or computing the magnitude agreement between scores

When wishing to correlate scores between two attribute scales (e.g. assessing the monotonic relationship between scores on Extraversion and Anxiety scales using a Pearson correlation coefficient), the raw magnitude of scores expressed within different ranges on each scale confounds the assessment of that monotonicity. Likewise, if looking to index the agreement between two vectors of scale scores, where each scale possesses a different magnitude range, then any index of agreement will be confounded by the differing score range of each attribute scale. In the case of the Pearson correlation coefficient, both variables being correlated will be initially standardised.

---

[1] McGrane, J. A., & Maul, A. (2020). The human sciences, models and metrological mythology. *Measurement* (https://doi.org/10.1016/j.measurement.2019.107346), 152,107346, 1-9.

1.2 For Interpretation Purposes

1.2.1 Multi-attribute Score Profiles

When wanting to compare multiple scores on several different psychological attributes, perhaps say as a profile of scores for an individual, it is necessary to ensure each attribute possesses the same score range. If not, an individual may possess the same magnitude scale score on all attributes which, if plotted, would show a straight-line profile across all attributes. But, if those attribute ranges are completely different to one another, that plot would be highly misleading because as in section 1.1.1 above, whether a numerical score is high or low depends on the score-range within which it is expressed. E.g. 20/100 is low, 20/21 is extremely high – yet on a raw-score profile plot, both scores would be 20 with no sense of what a maximum or minimum score might be for any attribute.

1.2.2 Comparing an individual to a normative group

In this situation, we wish to compare an individual's score to how a group of individuals score (who are considered normative exemplars of some category or class of people). What we can do is re-express/transform a raw score as a percentile score, which informs us as to where the individual's raw score is located within the frequency distribution of scores comprising the normative group. Alternatively, we can transform the group scores in other ways to provide a common metric for raw score interpretation while also transforming the shape of the observed score distribution so as to conform to a normal distribution, with known distribution proportions of cases scoring at each raw-score magnitude.

Ultimately, whether to transform or not is context-dependent. For example, test publishers such as Hogan Assessments rely upon simple norm-group lookup tables, where the empirical frequency distribution of the percentage of cases associated with each raw score in that norm-group is used to interpret the relative magnitude of an individual's raw score. Others like Cognadev transform scores to a common metric and rely upon that metric to convey magnitude interpretation. Many use normalised-standardized T, sten, or stanine scores to convey an individual's relative score on a test.

The next three sections provide detailed expositions and examples of each popular methodology used by virtually all psychological assessment test publishers, including Cognadev.

2. Percentiles

Basically, no more than simply expressing the percentage of cases who score at or below a raw score. Sometimes a person's raw score is never disclosed, in favour of a statement such as: "this person scored at the 90th percentile" or "their percentile rank is 90". Pretty simple you'd think except for the fact that so many authors can't seem to decide upon whether a percentile includes the raw score or not. As the title of one of my earlier technical whitepapers from 2011 asks: "Percentiles and Percentile Ranks: Confused or What?"; when reviewing 22 definitions from websites and textbooks it is clear that confusion reigns until you consider the assumptions some make about scale scores. The whitepaper goes into explicit detail explaining how such competing definitions arise, along with the relevant equations, worked examples, and deeper consideration of a scale score as a discrete integer magnitude or continuous real-valued number (e.g. instead of a discrete magnitude of 2, we might assume 2 is actually the midpoint between a continuously varying range of scores between 1.5 and 2.499999999).

Let's take an example of some test scores … the Eysenck Personality Questionnaire-Revised (EPQR) Extraversion scale, with a 0-23 test score range…

Table 1: Frequency distribution of EPQR-E scale scores (UK males reference sample dataset)

	Frequency table: LONG_E (EPQR100M)			
Scale Score	Count	Cumulative Count	Percent	Cumulative Percent
0	9	9	1.48	1.48
1	12	21	1.97	3.44
2	13	34	2.13	5.57
3	17	51	2.79	8.36
4	16	67	2.62	10.98
5	12	79	1.97	12.95
6	15	94	2.46	15.41
7	16	110	2.62	18.03
8	22	132	3.61	21.64
9	26	158	4.26	25.90
10	32	190	5.25	31.15
11	31	221	5.08	36.23
12	36	257	5.90	42.13
13	31	288	5.08	47.21
14	29	317	4.75	51.97
15	33	350	5.41	57.38
16	39	389	6.39	63.77
17	35	424	5.74	69.51
18	29	453	4.75	**74.26**
19	31	484	5.08	**79.34**
20	34	518	5.57	84.92
21	39	557	6.39	91.31
22	33	590	5.41	96.72
23	20	610	3.28	100.00
Missing	0	610	0.00	100.00

Let's assume *(in line with the status-quo)* that the scale scores are integer 'mappings' onto a continuously varying quantitative attribute. What would be the 75th percentile score – that score below which 75% of the sample score? Well, we can see from the above table that it must be between 18 and 19 ... as this is where between 74.26% and 79.34% of the sample scores are found. Applying the standard formula ...

Eq.1

$$P_i = ll + \left(\frac{np - cf}{f_i} \right) \cdot w$$

where

$P_i =$ the i^{th} percentile

$ll =$ the exact lower limit of the interval containing the percentile point

$n =$ the total number of scores

$p =$ the proportion corresponding to the desired percentile

$cf =$ the cumulative frequency of scores below the interval containing the percentile point

$f_i =$ the frequency of scores in the interval containing the i^{th} percentile point

$w =$ the width of the class interval

Our scores in this case are single values – no range at all. So, our class intervals are in fact the scores themselves. E.g. 0-0, 1-1, 2-2, 3-3 etc.

The exact limits however correspond to ±0.5 around each class interval boundary score – the 0, 1, 2, 3, 4, 5, 6 etc. So, our exact limits are:

0 = -0.5 to +0.5

1 = +0.5 to +1.5

2 = +1.5 to +2.5

3 = +2.5 to +3.5

etc.

Let's re-label and slightly alter Table 1 above to correspond with our notation in the formula ...

Table 2: Frequency distribution of EPQR-E scale scores (UK males reference sample dataset) - augmented

Scale Score	Frequency table: LONG_E (EPQR100M)					
	Exact Limits	Midpoint	f	cf	Percent	Cumulative Percent
0	-0.5 to 0.5	0	9	9	1.48	1.48
1	0.5 to 1.5	1	12	21	1.97	3.44
2	1.5 to 2.5	2	13	34	2.13	5.57
3	2.5 to 3.5	3	17	51	2.79	8.36
4	3.5 to 4.5	4	16	67	2.62	10.98
5	4.5 to 5.5	5	12	79	1.97	12.95
6	5.5 to 6.5	6	15	94	2.46	15.41
7	6.5 to 7.5	7	16	110	2.62	18.03
8	7.5 to 8.5	8	22	132	3.61	21.64
9	8.5 to 9.5	9	26	158	4.26	25.90
10	9.5 to 10.5	10	32	190	5.25	31.15
11	10.5 to 11.5	11	31	221	5.08	36.23
12	11.5 to 12.5	12	36	257	5.90	42.13
13	12.5 to 13.5	13	31	288	5.08	47.21
14	13.5 to 14.5	14	29	317	4.75	51.97
15	14.5 to 15.5	15	33	350	5.41	57.38
16	15.5 to 16.5	16	39	389	6.39	63.77
17	16.5 to 17.5	17	35	424	5.74	69.51
18	17.5 to 18.5	18	29	453	4.75	**74.26**
19	18.5 to 19.5	19	31	484	5.08	**79.34**
20	19.5 to 20.5	20	34	518	5.57	84.92
21	20.5 to 21.5	21	39	557	6.39	91.31
22	21.5 to 22.5	22	33	590	5.41	96.72
23	22.5 to 23.5	23	20	610	3.28	100.00
Missing			0	610	0.00	100.00

where

$P_{75} =$ the 75^{th} percentile

$ll = 18.5 =$ the exact lower limit of the interval containing the percentile point

$n = 610 =$ the total number of scores

$p = 0.75 =$ the proportion corresponding to the desired percentile (note this is nothing more than the percentile expressed as a proportion $(75 \div 100)$

$cf = 453 =$ the cumulative frequency of scores below the interval containing the percentile point

$f_i = 31 =$ the frequency of scores in the interval containing the i^{th} percentile point

$w = 1.0 =$ the width of the class interval

feeding these values into the formula we obtain …

$$P_i = ll + \left(\frac{np - cf}{f_i}\right) \cdot w$$

$$P_{75} = 18.5 + \left(\frac{610 \cdot 0.75 - 453}{31}\right) \cdot 1.0$$

$$P_{75} = 18.5 + \left(\frac{457.5 - 453}{31}\right) \cdot 1.0$$

$$P_{75} = 18.5 + 0.1452 \cdot 1.0 = 18.645$$

Or in MathCad:

$percentile := 0.75$ $ll := 18.5$ $cf := 453$ $f := 31$ $w := 1$ $N := 610$

$$pscore := ll + \left(\frac{(N \cdot percentile) - cf}{f}\right) \cdot w = 18.645$$

So, the 75th percentile is a score of 18.645. This is the score at which 75% of observations will be observed to be below. BUT – the score is unattainable as this is an integer scored test. What we actually observe is that 74.26% scores will lie at or below 18, with 79.34% of scores at 19 or below. IF we wish to use exact percentiles – then we must accept that our scores are estimates of hypothetical real-valued continuous numbers, hence a score of 18.645 is perfectly valid under these conditions, and the definition of a percentile is most correctly defined as:
the value below which P% of the values fall.

2.1 Percentile Ranks

If we decide to transform our raw scores into percentile ranks, reporting these instead of the raw score to an individual, how do we do this?

For example, when an individual scores 19 on a test, what do we conclude? Here we need to compute the percentile rank of the score – which is just the reverse of computing the score for a particular percentile. Now we know the score =19, but need to compute the percentile for it.

The formula is:

Eq. 2

$$PR_x = \left[\frac{\left(cf + \left(\frac{x - ll}{w} \right) \cdot f_i \right)}{n} \right] \cdot 100.0$$

where

$PR_x =$ the percentile rank of score x

$ll =$ the exact lower limit of the interval containing the score x

$n =$ the total number of scores

$cf =$ the cumulative frequency of scores below the interval containing the score x

$f_i =$ the frequency of scores in the interval containing x

$w =$ the width of the class interval

So, for a score of 19, the exact percentile rank is:

$cf := 453 \quad score := 19 \quad f := 31 \quad ll := 18.5 \quad w := 1 \quad N := 610$

$$PR := \frac{cf + \left(\left(\frac{score - ll}{w} \right) \cdot f \right)}{N} \cdot 100.0 = 76.803$$

a score of 19 is at the 76.8th percentile – the score at which 76.8% of scores will be found to be below this score.

But, using actual scores means that only certain % values can be provided – based upon the exact number of frequencies observed for each score. So, there can be no 75th percentile for our observed frequency distribution – only a 74.26th or 79.34th percentile. So …

❶ If you want to assign exact percentile ranks to scores, then you must use the formulae above and assume each integer score is actually a point-estimate from an interval of possible scores. Here, the definition of a percentile is the value **below** which P% of the values fall.

❷ Alternatively, if you simply prefer to state the frequency of people who score at or below an observed test score, then you use the actual frequencies of scores in your normative data. Here, the percentile is the point **at or below** which a given percentage of scores is observed.

The first statement assumes that scores are theoretically continuous but can only be observed as integers; the second assumes scores are simply discrete integers.

3. Standardisation

This refers specifically to transforming raw test scores into what's called 'standard scores or z-scores'. A routine statistical procedure which retains the essential magnitude variation in a scale, while re-re-expressing scores in a common metric. In this way, any standard score can be expressed in a convenient metric, as the equation below shows:

Eq. 3

$$\text{z-score} = \frac{(\text{raw score} - \bar{x})}{s} \text{...}\textit{where}\text{...}$$

$\bar{x} =$ the norm sample mean

$s =$ the norm sample standard deviation

then we can apply a

$$\text{transformation equation} = (\text{z-score} \cdot \text{new\_Std.dev}) + \text{new\_mean}$$

and re-express our standard scores as:

$\text{Stanine} = (\text{z-score} \cdot 2) + 5$ .... (range 1-9: integers)

$\text{Sten} = (\text{z-score} \cdot 2) + 5.5$ .... (range 1-10: integers)

$\text{T-score} = (\text{z-score} \cdot 10) + 50$ .... (nominal range 20-80: integers)

This is the transformation applied (standardisation of each variable's observations) when computing a Pearson correlation. The advantage to such a transformation is that variables with different measurement ranges can be transformed into a common metric, prior to using, comparing, or correlating their magnitudes.

In some Cognadev assessments for example, the raw scores for final reported variables have vastly different score ranges. To ensure interpretability, we create a representative norm-group with these raw scores, then transform these raw scores to T-scores. Subsequent respondent raw-scores are then identified in a look-up table created by the norm-group score transformation. That is, a new raw score is identified in the normative group raw-to-T-score-table, which provides its T-score equivalent.

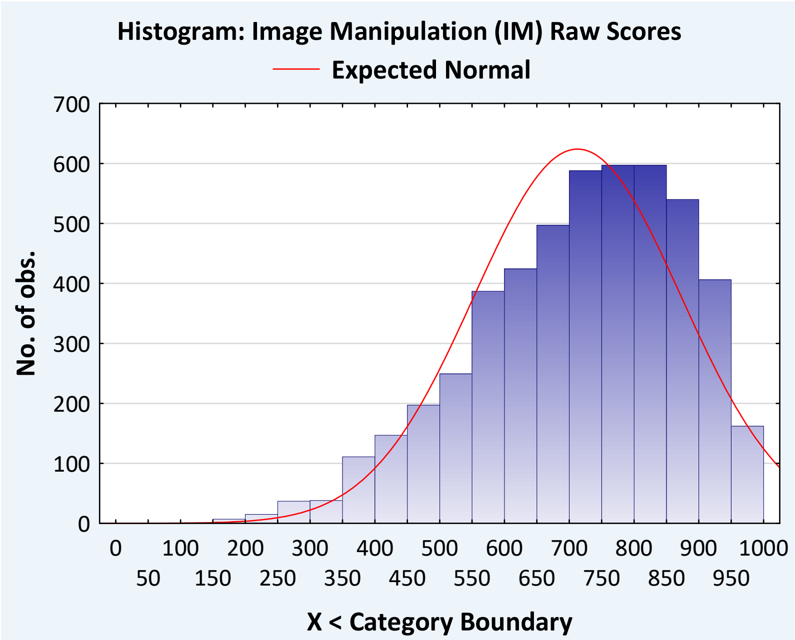
However, it is important to note that the distribution 'shape' of the raw scores is not affected by the transformation except where the compression or expansion of the variability of scores is concerned. For example, a sample of hypothetical raw scores for two variables have been generated below, differing in their magnitudes and variabilities. Both are transformed to T-scores. The histograms and descriptive statistics show the effect of the standardizing transformation.

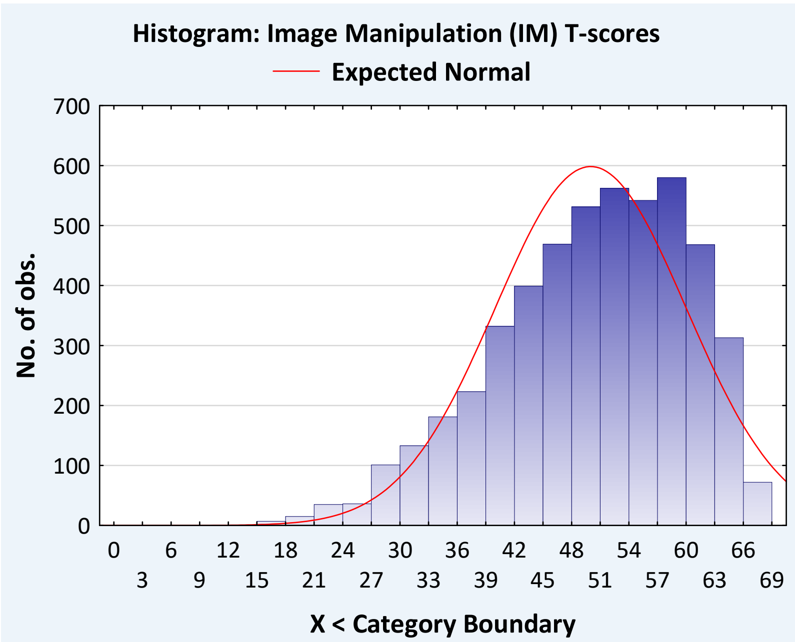
Variable 1: Image Manipulation Skill

Table 3: Raw and T-scores for variable Image Manipulation

Variable	Descriptive Statistics (Tech Report #15, sample data, n=5000 cases, moderate beta					
	Valid N	Mean	Median	Minimum	Maximum	Std.Dev.
Image Manipulation (IM)	5000	712.81	732	134.0	997.0	159.84
T-score (IM)	5000	50.00	51	13.8	67.8	10.00

Figure 1: Histograms of Raw and T-scores for Image Manipulation



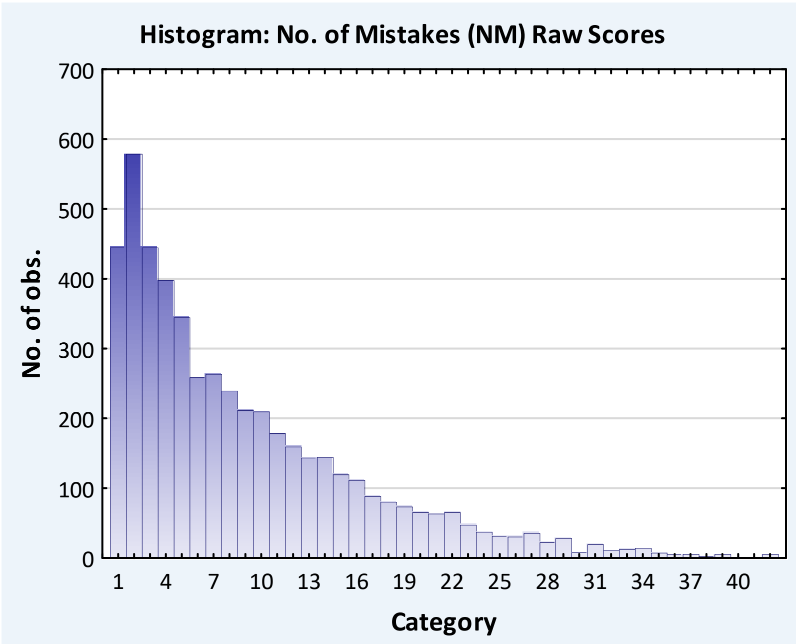


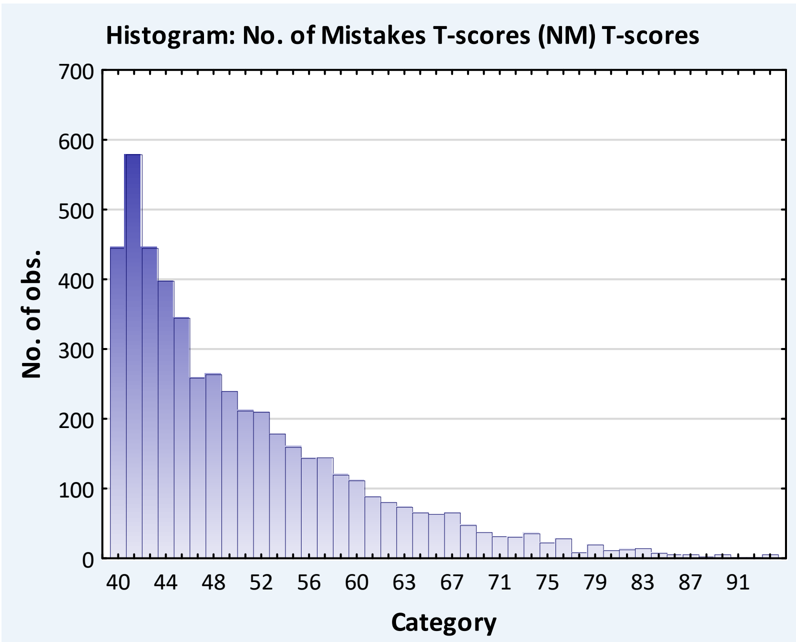
Variable 2: No. of Mistakes

Table 4: Raw and T-scores for variable No. of Mistakes

Variable	Descriptive Statistics (Tech Report #15, sample data, n=5000 cases, J-shaped beta, 8-J					
	Valid N	Mean	Median	Minimum	Maximum	Std.Dev.
No. of mistakes (NM)	5000	8.87	7.00	1.00	42.00	7.51
T-scores (NM)	5000	50.00	47.51	39.52	94.14	10.00

Figure 2: Histograms of Raw and T-scores for No. of Mistakes





The distribution shape is maintained for both variables … although the variability is expanded for Variable 2 as a result of the T-score transformation.

4. Normalisation

With respect to data transformations, there are two distinct meanings to this word.

There is the normalisation routinely undertaken in applications which transform data as a function of some reference magnitude (e.g. in orthogonal or parallel coordinate vector comparisons, sometimes it is useful to transform all vector magnitudes into a common metric varying between 0 and 1). Given this kind of 'normalisation' is solely concerned with re-expressing data in a common metric, I refer to this transformation as a "common-metric" transformation, in order to keep its meaning distinct from the second 'normalising' transformation.

The second meaning is that associated with a unit-normal distribution. That is, we transform data magnitudes such that the new magnitudes are more normally distributed, in line with the expected proportions observed for each magnitude within a perfect normal distribution. The data are first standardized then 'normalized', hence referred to as normalised-standardized scores.

4.1 Common-Metric Transformation (CM)

When wishing to display and compare data from variables whose measurement metric is not the same, it is useful to rescale each variable's values into a convenient common metric. This is especially the case when displaying data with different ranges in multiple line-graphs and scatterplots, and for comparing multiple variables with coefficients such as the Gower agreement index.

Also, with specific regard to the Gower and other relative magnitude agreement indices, using two variable vectors with differing minimum and maximum values produces incorrect agreement measures because the relative change in one variable vector is not equal to the relative change in another (by reason of the inequality which exists between the respective minima and maxima).

Unlike conventional standardization, which transforms a variable's values by subtracting each observed value from the mean of all observed values and dividing this difference by the standard deviation of the values, the rescaling implemented here preserves the relativity between each variables' observations while rescaling the magnitudes into a common metric. CM normalisation is simply a linear transformation of raw scores into a new metric, preserving exact magnitude relations and raw-score variability.

The formula is:

Eq.4

$$x_{cm} = \left\{ \left[\frac{(x_{observed\_score} - x_{range\_min})}{(x_{range\_max} - x_{range\_min})} \right] * (cm_{new\_range\_max} - cm_{new\_range\_min}) \right\} + cm_{new\_range\_min}$$

where

x_{range_min} = minimum possible value of raw score range for variable x

x_{range_max} = maximum possible value of raw score range for variable

$cm_{new\_range\_min}$ = the common-metric minimum possible rescaled value of x_{cm}

$cm_{new\_range\_max}$ = the common metric maximum possible rescaled value of x_{cm}

A worked example

old scale

$x_{observed\_score} := 246 \quad x_{range\_min} := 0 \quad x_{range\_max} := 495$

new scale

$cm_{new\_range\_min} := 0 \quad cm_{new\_range\_max} := 100$

$$x_{cm} := \left(\left(\left(\frac{(x_{observed\_score} - x_{range\_min})}{(x_{range\_max} - x_{range\_min})} \right) \cdot (cm_{new\_range\_max} - cm_{new\_range\_min}) \right) + cm_{new\_range\_min} \right) = 49.697$$

$$x_{cm} := \mathrm{round}\left(\left(\left(\frac{(x_{observed\_score} - x_{range\_min})}{(x_{range\_max} - x_{range\_min})} \right) \cdot (cm_{new\_range\_max} - cm_{new\_range\_min}) \right) + cm_{new\_range\_min} \right) = 50$$

4.1.1 A comparative example of T-score vs CM-score transformations

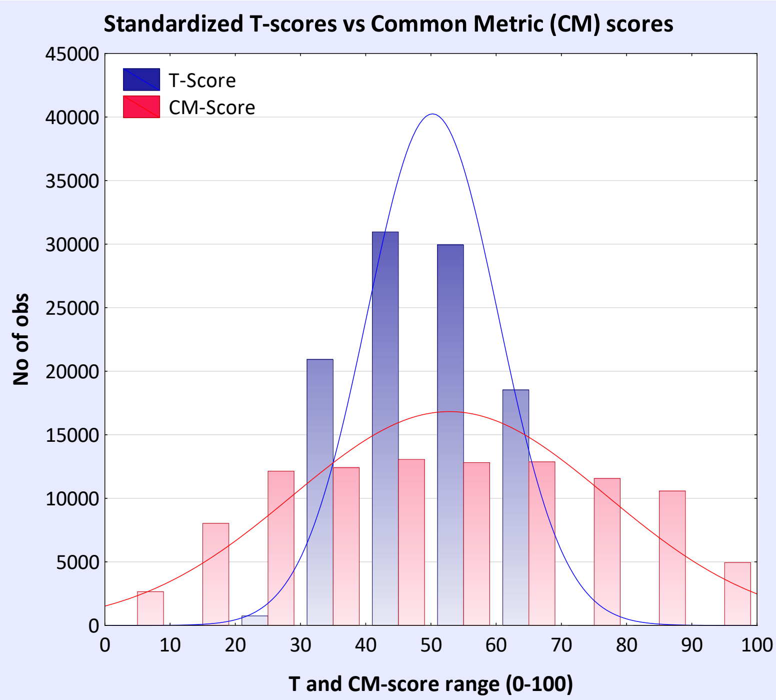
For a CPP T-*attribute* External Focus, the descriptive statistics for the raw sum-scores, the T-score transformed scores, and CM scores, are:

Table 5: Descriptive statistics for raw, T, and CM scores; CPP attribute External Focus

	Descriptive Statistics (T-01 External Focus Data.sta)						
Variable	Valid N	Mean	Median	Minimum	Maximum	Std.Dev.	Skewness
Raw Rule Score	101063	210.31	210	0	400	95.835	-0.003
T-Score	101063	50.00	50	28	70	10.017	-0.009
CM-Score	101063	52.58	52	0	100	23.968	-0.003

With T and CM-score histogram:

Figure 3: Standardized T-scores and CM score histograms



Note how the T-scores compress the scores into a much smaller range.

Monotonically, they are virtually equivalent, as seen in Table 6. But, as Table 5 above shows, in terms of actual magnitude range, they are very different.

Table 6: Pearson Correlations between raw, T, and CM scores; CPP attribute External Focus

	Correlations (T-01 External Focus Data.sta), N=101063		
Variable	Raw Rule Score	T-Score	CM-Score
Raw Rule Score	1.0000	0.9996	0.9999
T-Score	0.9996	1.0000	0.9995
CM-Score	0.9999	0.9995	1.0000

If we look at another variable, comparing standardized to CM scores, we can see the transformation effect from a different perspective. Table 7 shows the descriptive statistics for the raw, z, T-, and CM-scores.

Table 7: Descriptive statistics for raw, z, T, and CM scores for a rule variable

	Descriptive Statistics (T-12 Metacognitive monitoring.sta)						
Variable	Valid N	Mean	Median	Minimum	Maximum	Std.Dev.	Skewness
Raw Rule Score	101063	421.13	435.00	57.000	700.00	130.422	-0.308
z-score	101063	-0.00	0.11	-2.792	2.14	1.000	-0.308
T-score	101063	50.00	51.00	22.000	71.00	10.003	-0.309
CM-Score	101063	60.16	62.00	8.000	100.00	18.636	-0.307

Note the difference in means and especially the highlighted standard deviations between the standardized T-scores and CM-scores. Figure 4 shows the frequency distribution the raw scores.

Figure 4: A raw score histogram

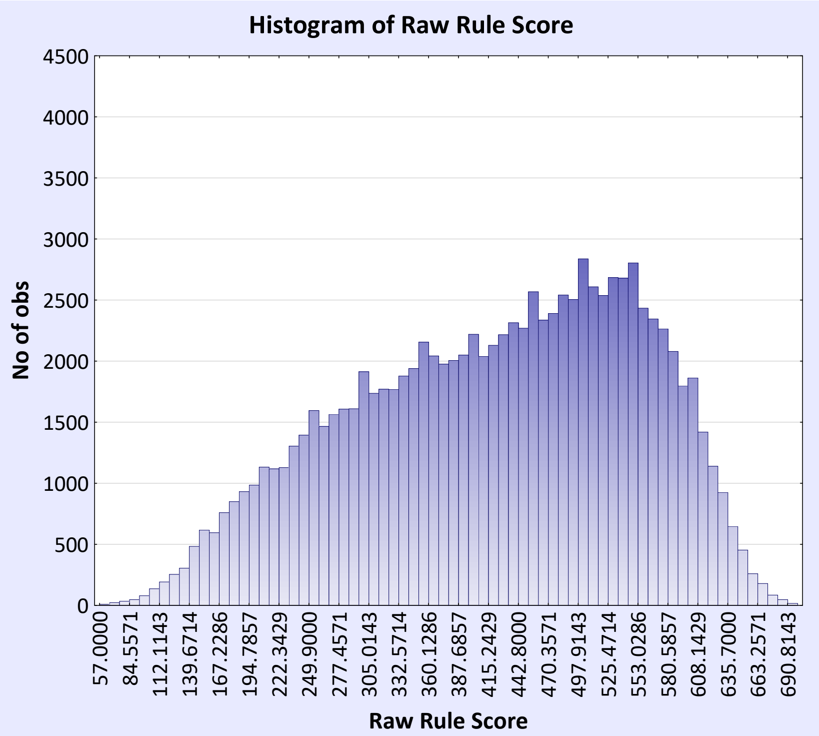
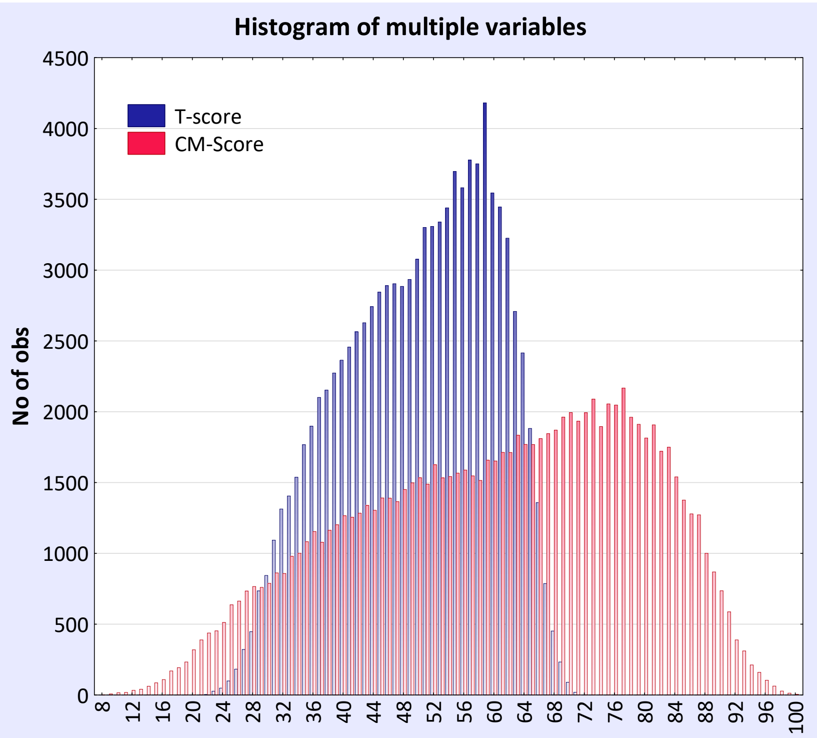


Figure 5: The standardized T and SM Transformed score histograms



The CM transformation preserves the exact shape and spread of raw sum scores, but now within a 0-100 metric. Whereas the T-score transformation compresses the scores into a narrower range. Monotonically the scores are all equivalent to one another, as shown in Table 8.

Table 8: Pearson Correlations between raw, T, and CM scores; CPP attribute External Focus

Variable	Correlations (T-12 Metacognitive monitoring.sta) Marked correlations are significant at p < .05000 N=101063 (Casewise deletion of missing data)		
	Raw Rule Score	T-score	CM-Score
Raw Rule Score	1.0000	0.9996	0.9999
T-score	0.9996	1.0000	0.9995
CM-Score	0.9999	0.9995	1.0000

4.2 Normalised-Standardized Scores (NST)

Normalized scores are specifically constructed to provide proportions of cases at each score level according to the expected frequencies of a perfect normal distribution. The mean and standard deviation of the raw data are the key parameters which enable standard scores and the normalization to be created. In this way, one can refer to a suitable-metric transformed integer score, with a very good idea just how many people obtained that score. Obviously, this can still end up with some inaccuracies where observed score ranges have big gaps between them, and frequencies are not spread very well across the whole score range. In other words, it's not a magic bullet to fix up badly skewed distributions of scores!

If we refer to Eq.3 again:

$$\text{z-score} = \frac{(\text{raw score} - \bar{x})}{s} \text{...}\textit{where}\text{...}$$

$\bar{x} =$ the norm sample mean

$s =$ the norm sample standard deviation

then we can apply a

$$\text{transformation equation} = (\text{z-score} \cdot \text{new\_Std.dev}) + \text{new\_mean}$$

and re-express our standard scores as:

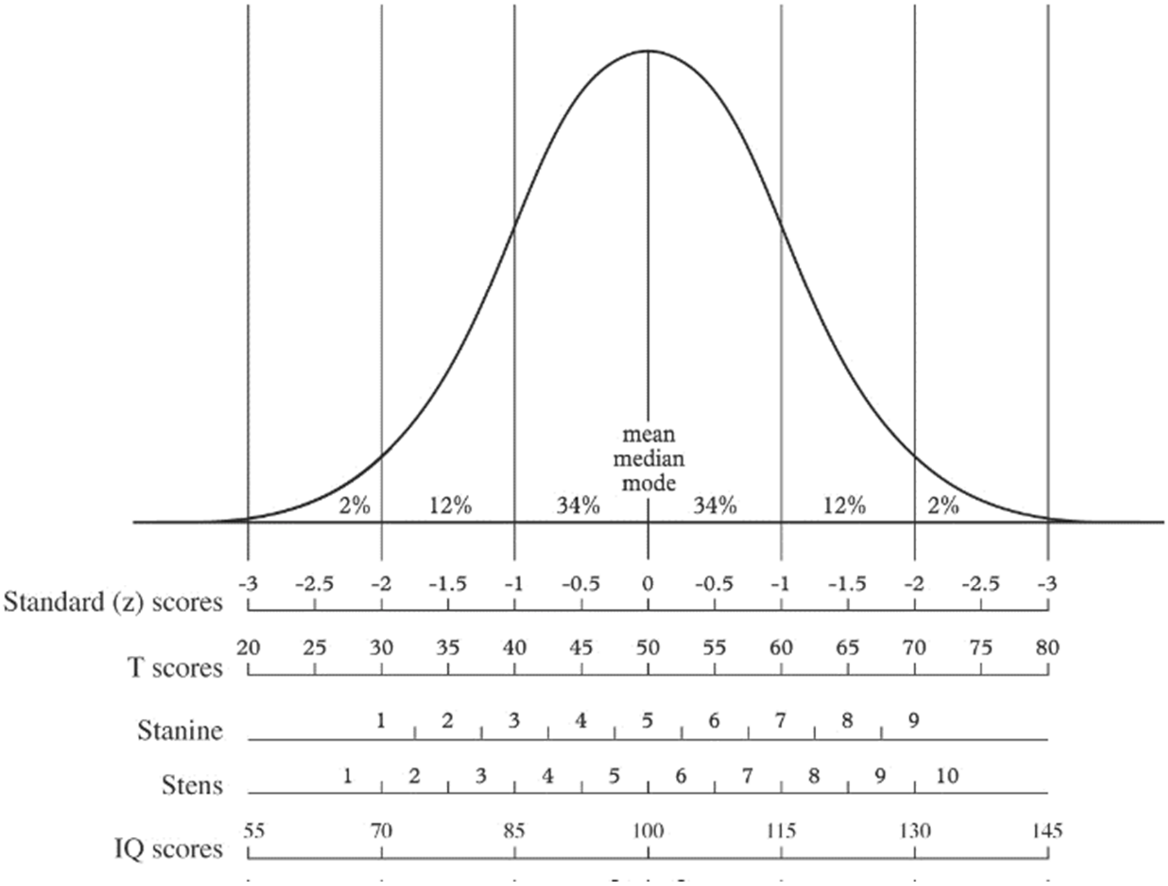
$$\text{Stanine} = (\text{z-score} \cdot 2) + 5 \text{ .... (range 1-9: integers)}$$

$$\text{Sten} = (\text{z-score} \cdot 2) + 5.5 \text{ .... (range 1-10: integers)}$$

$$\text{T-score} = (\text{z-score} \cdot 10) + 50 \text{ .... (nominal range 20-80: integers)}$$

Then not only do we standardize scores as explained in section 3, but we now transform them according to how many such scores would be expected to be found at each transformed score, or at each raw-score magnitude. Figure 6 provides the expected normal distribution percentage of cases, relative to the magnitudes of each kind of transformed score.

Figure 6: The normal distribution expected frequencies



The 7-steps for computing the Normalised-standard (NST) scores

1. Work out the cumulative proportion (CP) for each raw score – doing steps 2-5 below …

2. Prepare a frequency distribution of the scores

3. Compute the Cumulative Frequency (CF) per score, where CF = the sum of the frequencies observed up to and including the "current score".

4. Compute the CF to the mid-point of each score interval by adding the CF at a score to half the numbers of cases observed at the same score.

5. Divide the mid-point CF (from #4) by the total number of cases (N). This is the CP (Cumulative proportion).

6. Obtain the *theoretical* Normal Distribution z-score for each CP *(the inverse normal calculation)* – here we find the z-score from a given proportion of the area under the curve.

7. Use the conventional transformation formula to convert the z-score into a sten, stanine, or T-score.

Clearly, this is computationally intensive. Fortunately, anyone interested in producing NST scores and the *raw_score*-to-*NST-lookup* table for one or more scales and transformed score types can download my free software to do this (Stanscore 5).

A hypothetical variable using 25k cases: the score differences

Table 9: The frequency distributions of some transformed scores. grouped into 10 classes:

(a) CM scores

	Frequency table: CM Score (A_16 - Coherence CM Scores, 25k dataset)			
From To	Count	Cumulative Count	Percent	Cumulative Percent
0 <=x<10	0	0	0.00	0.00
10 <=x<20	2	2	0.01	0.01
20 <=x<30	7	9	0.03	0.04
30 <=x<40	3096	3105	12.38	12.42
40 <=x<50	9188	12293	36.75	49.17
50 <=x<60	12569	24862	50.28	99.45
60 <=x<70	138	25000	0.55	100.00
70 <=x<80	0	25000	0.00	100.00

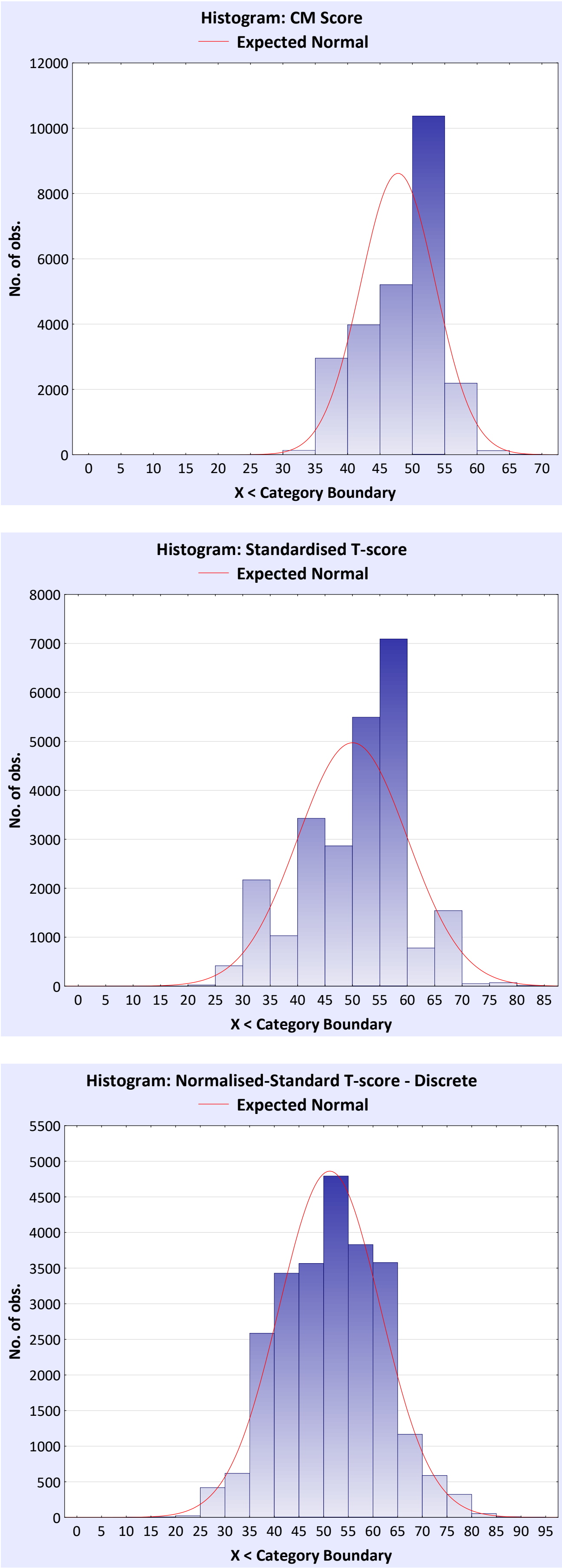
(b) Standardised T-scores

	Frequency table: Standardised T-score (A_16 - Coherence CM Scores, 25k dat			
From To	Count	Cumulative Count	Percent	Cumulative Percent
0 <=x<10	2	2	0.01	0.01
10 <=x<20	6	8	0.02	0.03
20 <=x<30	442	450	1.77	1.80
30 <=x<40	3204	3654	12.82	14.62
40 <=x<50	6295	9949	25.18	39.80
50 <=x<60	12587	22536	50.35	90.15
60 <=x<70	2324	24860	9.30	99.45
70 <=x<80	125	24985	0.50	99.95
80 <=x<90	13	24998	0.05	100.00
90 <=x<100	0	24998	0.00	100.00

(c) Normalised-Standardised T-scores

	Frequency table: Normalised-Standard T-score - Discrete (A_16 - Coherence CM Scores, 25k data			
From To	Count	Cumulative Count	Percent	Cumulative Percent
0 <=x<10	0	0	0.00	0.00
10 <=x<20	12	12	0.05	0.05
20 <=x<30	440	452	1.76	1.81
30 <=x<40	3204	3656	12.82	14.62
40 <=x<50	6995	10651	27.98	42.60
50 <=x<60	8625	19276	34.50	77.10
60 <=x<70	4745	24021	18.98	96.08
70 <=x<80	913	24934	3.65	99.74
80 <=x<90	62	24996	0.25	99.98
90 <=x<100	4	25000	0.02	100.00

Figure 7: The score histograms for the transformed "Coherence" scores



Moving from CM to Normalised-Standardised T-scores, we can see how the normalisation procedure 'forces' scores to be distributed more 'normally'.

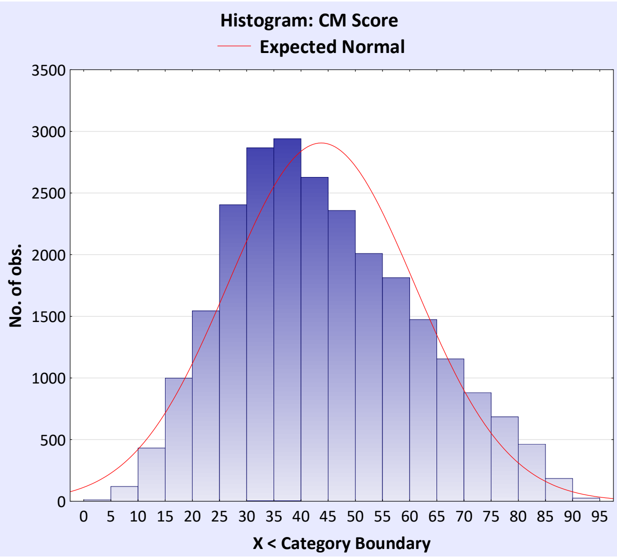
The CM scores show the actual distribution shape.

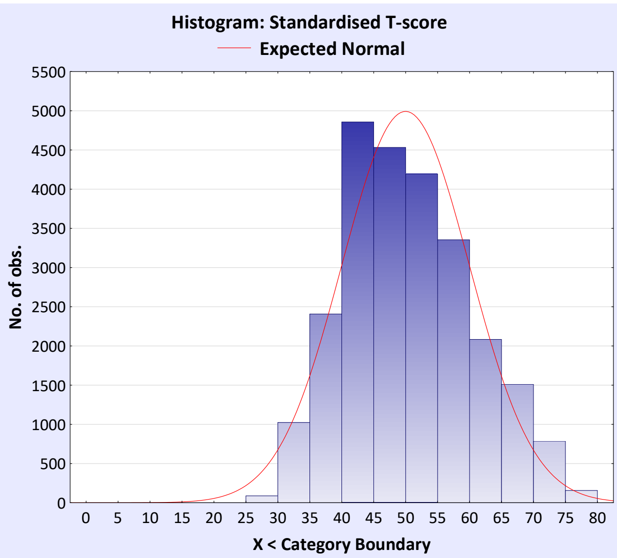
Note also that the CM scores possess much less variability than standardised T-scores.

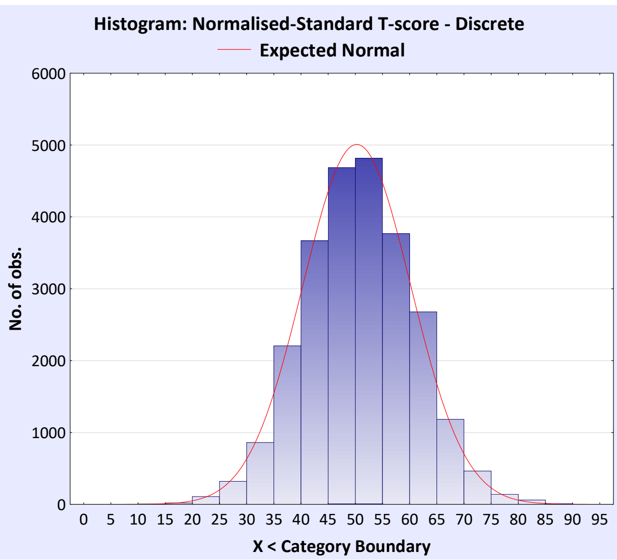
	Descriptive St
Variable	Std.Dev.
CM Score	5.78
Standardised T-score	10.04
Normalised-Standard T-score - Discrete	10.26

Another hypothetical variable – more variability, using 25k cases

Figure 8: The score histograms for more variable attribute scores







Moving from CM to Normalised-Standardised T-scores, we can see how the normalisation procedure 'forces' scores to be distributed more 'normally'.

The CM scores show the actual distribution shape.

Note also that the CM scores possess much more variability than T-scores.

But this is an exceptional hypothetical variable whose CM scores are more variable than the standardised t-scores.

	Descriptive Sta
Variable	Std.Dev.
CM Score	17.16
Standardised T-score	9.99
Normalised-Standard T-score - Discrete	9.95