THE TEACHING OF MATHEMATICS
2003, Vol. VI, 2, pp. 69–80

COMPONENTS OF SUCCESSFUL EDUCATION

R. M. Dimitrić

Abstract. I present a sample of a diagnostic test developed for testing students who enroll into elementary statistics courses. I use the test for observational studies about correlations of students' diagnostic test scores and their subsequent performance in the observed classes. Concrete numerical statistics are given that have proved to be highly reliable and stable to a degree over different classes. The test is more precise than the common generic diagnostic/entrance tests that are administered at universities (meant to gage students' mathematical background). The test has a direct practical value since it is used in everyday teaching practice and is of wider theoretical significance since it addresses one of the components of successful education, namely influence of preparation of students on their success.

ZDM Subject Classification: D65; *AMS Subject Classification*: 00 A 35

This is the first in a series of planned three papers addressing quantifications of the following four important components of education process: 1) Adequacy of students' background for a particular course, 2) Amount of students' work (efforts) in a particular course, 3) Standards in educating, examining and evaluating, 4) Quality of teaching in a course. These four components are not independent when related to the same course and this is an important point to bear in mind. This paper will concentrate only on the first component, namely conditional the other components be given or fixed. Consequently the statistical analysis will at first be only univariate, without regard to other components. In such context, the statistics obtained are invariant constants to a great extent, and these constants can be used as partial calibrating predictors of students' performance. I will make my point through a study performed in the past few years. Enrollment figures seem to be on most of the administrators' minds regardless of the educational considerations and one of the consequences is that the students often take courses haphazardly, without due preparation. One of the immediate uses of the present study is to give numerical values for chances of students' success, given their level of preparation; thus diagnostically, the instructor can direct a student to get the adequate preparation and then take the course, rather than waste the whole semester before realizing that his chances of passing the course might be rather small.

In this study the course in question was "Elementary statistics", as usually taught to undergraduate students not majoring in "hard sciences" like engineering or mathematics, but rather aimed at pre-med, economics, social science majors, or else those who needed to fulfill a requirement through such a course in order to enter a graduate school. This one-semester course is very lightly mathematical and for the most part involves the following material: exploration and charting of data, some probability, binomial distribution, normal distributions, estimating parameters and hypothesis testing, inferences from two samples, correlation and regression. I have not covered the whole material in the syllabus in any of the classes observed, rather, usually about 70%–80%.

I have used four different classes in this study, at four different colleges or universities in the San Francisco Bay Area, code named here as follows: CCSF, SFSU, CCCC, UCBX. The classes were 40, 38, 18, and 15 students in size respectively, at the commencement of the course. The same textbook was used, the same notes, and the same standards in educating, examination and evaluation were applied as much as feasible; none of these standards were very high.

I have previously observed that the scores on usual high school tests widely taken before entering university (such as SAT) did not give much relationship with the students' performance in my classes (correlation never exceeded 10%). This has been confirmed in the context of diagnostic tests for calculus and pre-calculus; in general some universities have noticed the same lack of correlation and have subsequently removed the SAT1 scores from the entrance criteria. I think that one of the main reasons for this lack of correlation is that there is a prolific "cottage industry" that specializes in preparing students in skills for passing the test. Thus, a considerable number of students who pass (even with high scores) have the test taking skills, but not necessarily the required knowledge. To gauge the students' preparation for this class I have made a diagnostic test that the students took at the first class meeting. They had to answer 25 short questions in 25 minutes, without a calculator or other aids. They were told that the performance on this test would have no effect on my grading their performance in the course, but that they should put as much effort in it as if this were an ordinary graded exam. Each question was assigned 2 points for the correct answer, and 0 for incorrect, although some questions had a "partial credit" of 1 point. This test is the course specific: I have constructed the test in view of what students certainly need to know in order to pursue the course without a burden of having to constantly wonder about basic algebra used throughout the course. Over the span of several years, I have recorded "tender spots" in students' background education that were great inhibitors of their learning advance; those without problems of this kind had (on the whole) no problems in mastering the course. The test is very elementary and not exhaustive; it tests rather basic material (in elementary algebra or arithmetic), much simpler than the material usually found in the "requirement" for this course, which was "Intermediate Algebra". To prevent the practice of drilling for the test I try to keep this test confidential as much as possible.

Following are the questions from the test, with my brief comments added in brackets [...]. Percentages in parentheses are showing mean percentages of students in all four classes answering that particular question correctly.

Intro-Quiz 0

Answer the problems in order they appear, not skipping any. Do not use calculators; you do not need them.

0) What math courses, if any, have you taken in the past two years?

[The percentage of people having taken any mathematics recently was surprisingly low.]

1) Mark the y-axis in the following rectangular coordinate system: (59%)

Fig. 1

[Coordinate systems and graphing of the information is often used in the course, and the question is rather basic. The percentage of correct answers is surprisingly low.]

2) A number x is not smaller than 5 and is smaller than 10. Write this down using mathematical symbols, through inequalities. (14%)

[Basic interval handling and inequalities are essential. Low number of correct answers is due to the fact that many did not interpret well the phrase "not smaller than 5"; if they interpreted the inequalities to mean $5 < x < 10$, they earned 1 point out of 2.]

[Questions **3–6**) meant to test students' handling of basics of abstract algebra; there are 4 questions to minimize the effect of complete guessing of these "true-false" questions.]

3) Is it true that $\sqrt{a^2 + b^2} = ab$? (95%)

4) Is it true that $\sqrt{a^2 + b^2} = a + b$? (22%)

5) True or false: $\dfrac{a}{b+c} = \dfrac{a}{b} + \dfrac{a}{c}$? (62%)

6) True or false: $\dfrac{a+b}{c} = \dfrac{a}{c} + \dfrac{b}{c}$? (89%)

7) The m-th man standing in line is assigned the number $2m - 1$. What number does the 7th man have? (43%)

[This question was put in the test to see how students cope with elementary abstraction. Students who cannot overcome notational changes are greatly handicapped.]

8) Write down 0.75 as a fraction. (95%)

9) How many numbers are there in the sequence 5,6,7, ... ,97,98,99,100 ? (24%)

[How well can you count, with dots ... ? The most frequent answer was: 95. Is this a "trick" question?]

10) Express 0.333 as a percent. (62%) [Conversions of rational numbers into percentages and vice versa is an activity that continues throughout the course.]

11) Express 15% as a decimal number. (100%)

12) The graph of the equation $y = -2x+5$ is a line. Moving from left to right, is this line going up or down? (49%)

[Lines are basic, but half of the students do not seem to have these basics. Regression lines cannot be handled correctly without the general knowledge about lines in general.]

13) Simplify $(a-b)^2+(a+b)^2$. (35%)

[Things like this come in handy when doing variances.]

14) What is the square root of 64? (89%)

[Then, they need to do standard deviation.]

15) If $a>b$, write down the relation between $-a$ and $-b$, using math symbols. (43%)

[Inequalities and opposite numbers. They come in handy when handling the z-scores and symmetry.]

16) What is the set of z's satisfying $z \geq -1$ and $z<3$ at the same time? Write down this in mathematical terms, using the symbol z only once. (43%)

[Inequalities and intervals again.]

17) What is greater: 2 raised to the third power, or 3 raised to the 2nd power? (84%)

[Basic powers and their comparisons.]

18) Calculate $15-3\times 4$. (73%)

[In the course, students are allowed a non-graphing calculator, although, for every formula, I insist on them doing everything by hand, at least once. Just as well, calculators "know" precedence of operations.]

19) What is a chance of getting 2, 4 or 6, on a regular 6-face die? (62%)

[Common sense probability, before they had probability.]

[The next three questions deal with fractions and powers etc.]

20) Calculate $3^2/\sqrt{81} =$ (84%)

21) Calculate $(1/2)^3 =$ (73%)

22) Calculate $\dfrac{3}{\frac{1}{3}} =$ (68%)

23) Write down $\mid z \mid < 3$, without using the absolute value sign, in terms of two inequalities. (14%)

[Absolute values and inequalities. Two sided intervals of estimation etc. It is apparent that students do not really understand the meaning of absolute value, and while most would say that "it is positive", rarely do they know a definition of absolute value.]

24) What is the negation of the statement $x>3$? Express it in math symbols. (11%)

[If they only knew what the word "negation" meant, there would be more correct answers. It is my strong feeling that a requirement for "Math" should also be "English".]

25) Calculate 0.25/2. (82%)

Data analysis

All four groups of students were given a few mid-term exams (three for most part), one final exam and students could earn up to 10% of credit for their active participation in the classroom. For CCSF and CCCC the midterms were each weighted 15% of the final grade and a comprehensive final exam weighted 45%; in addition, their consistent activity in class was counted as 10% of their final grade. For the SFSU group the weights were 17% for the midterms, 40% for the final exam and 9% for the activities. The UCBX group had two midterm exams weighted 25% each, with 40% for the final and 10% for the classroom activity. Only the scores of the students in the final roster who took all exams were considered. The data sets are to be found at the ending pages. The standardized (normalized) scores where then added with the appropriate weights and the resulting number was called a cumulative score.[1] Students' final grades were based on the absolute scale (PASS= $60\%-74\%$, B= $75\%-89\%$, A= 90%+) as well as on the relative scale (by ranking of the cumulative scores). My interest was to find out whether there was a relationship between students' diagnostic scores and their cumulative scores, as well as the outcome (passing or not passing) and to quantify this relationship, given other equal conditions.

The Diagrams 1–4 (coming from the data tables 1–4 provided) are linear scatter plots of this relationship with the lines of best fit placed to illustrate the extent of linear relationship. The squares of correlation coefficients (the percentage of variance "explained" by regression) are listed along.

The CCSF group. This was a rather heterogeneous group of students, in regards to socio-economic, education and family background. There were no requirements for enrollment, as long as there was space available for any particular class. Course descriptions ("official syllabi") list "prerequisites" for most courses, but this is not interpreted as a requirement. Most of the students worked, (but not in essential, long-term jobs) and they were taking the course as a prerequisite for often yet-undefined major they may choose (for those who transfer to a four-year college). Among the students on the final roster (after attrition of 50%), the diagnostic score range was 29–50 (same for the subset of students who passed) with the mean 39.33, median 39.5 and deviation of about 6.14 (Data 1). There is some positive linear relationship (Diag. 1) with the correlation coefficient $R \approx 0.30$ and the "explained" variance of about 9%. These numbers increase when we remove outliers in the lower portion of the graph. The standard in grading was lowered somewhat − thus the students who had cumulative scores around -0.5 were assigned a "C-" grade, whereas normally their performance would not warranty a passing grade.

[1] Thus the cumulative score formula for every student, for instance at CCSF, or CCCC was: CUM= $0.15\sum_{i=1}^{3}(T_i-\bar{x}_i)/s_i+0.1(a-\bar{a})/s_a+0.45(f-\bar{x}_f)/s_f$, where T_i denotes the student's score on the i-th test, $\bar{x}_i$ and s_i respectively the mean and the standard deviation on that test, a is the activity score with the corresponding statistics $\bar{a}$ and s_a, and the symbols that contain f denote the same things for the final exam.

Linear relationships show more prominently in three separate subgroups of this particular group. That is understandable, for heterogeneity does not allow for too strong a relationship ...

The SFSU group. This group of students was homogeneous in the sense that they were poorly prepared to take the class: the range of diagnostic scores (after 37% attrition) was from 26 to 46 (28-46 for the passing subset) with mean 35.7, median 36 and the standard deviation of 5.75 (Data 2). The correlation coefficient $R = 0.27$ and $R^2 = 0.0713$ give a positive linear correlation (Diag. 2), and when the outliers are removed, both the correlation and the explained percentage of the variance increase considerably.

The students in this group usually stayed in the same college throughout 4 years; most of them worked part-time jobs. This group did not have students with "stellar" performances in their class.

The CCCC group. Of my four samples, these were the students with poorest preparation and my grading standards were considerably altered in order to pass students who did not meet the preset grading standards. These are students who are most disadvantaged in their economic status, with sporadic employment and uncertain plans for their education path. The diagnostic scores of the students on the final roster (after 50% attrition) ranged from 17 to 44 (26-44 for the passing subset), with mean 31.86, median 31 and standard deviation 8.97 (Data 3). A strong linear correlation (Diag. 3) with $R \approx 80\%$ and the explained variation of about 63% enforce homogeneity in the aforementioned sense.

The UCBX group. The UCBX group of students consisted of those who already had had their undergraduate (and some of them master's) degrees and were taking this course as a requirement to enter a graduate school (most often a business school). This was a reasonably homogeneous group for purposes of the study. All but one student (who did not take the final exam and is not listed in the chart) passed the course. The diagnostic score range was 39–50 with mean 44.67, median 44.5 and standard deviation of 3.53 (Data 4). There was strong linear dependence (Diag. 4) between the diagnostic scores and the overall performance for the course, with the correlation coefficient $R \approx 0.70$ and the "explained" variance of the overall performance of about 49%. This strong dependence is undoubtedly due to the homogeneity of the group and confounding with other components. Thus these students are also likely to be hard workers and at the high maturity level (most of them were adults in their thirties and already substantially employed).

Conclusions and applications

Related work. Diagnostic tests seem to have become standard at many universities. One can see the trend as admittance of educational systems' failures to bring students to the required level. The reality of tests existence is also shaped by numbers of returning older students, expansion of vocational education, influx of foreign students, and other such factors contributing to heterogeneity. Diagnostic tests are mostly administered through two delivery mechanisms, namely paper-based (such as ours; see also [2], [3], [6]) and computer-based (such as the California

University system [1], or University of Manitoba [7] with a ratio of numbers of tests about 2 to 1. For ease of grading, these tests are almost invariably multiple choice, some are self-grading, and unproctored, some are taken after students were told what the tests will contain or after being prepared in the areas the tests would cover. These diagnostic/entrance tests are fairly general, in that they are meant in pre-qualifying for a number of courses of varying degrees of similarity. Admittedly, this author has not seen tests for a single specific course, or for a statistics course in particular.

Why a diagnostic test? It is likely that to a mathematician, the relationships studied are self-evident. The wider audience may not necessarily agree with this self-evidence. Thus, a professor of psychology has communicated to me that, if students achieve low scores on a diagnostic test it may be because the students want (even subconsciously) to give the instructor an impression of low achievement in order to avoid high standards applied in the course, etc. These kinds of claims are not rare among non-mathematicians, and this provides one more reason of having sound data on the subject.

The two immediate uses of diagnostic tests are to inform the professionals of the competence level and to inform the students as to the remedial work to be done. As much as they help the educator and the students they could also help the administration (which is often not mathematically knowledgeable). I say "could help the administrator", for often the administrators are only concerned with the numbers game, namely passing as many students as possible as a primary, and often the only goal. In this affair there is a chain reaction, for the instructors often lower standards beyond any decency to please the administrators who in turn claim that they are under pressure from students' parents to pass the students, as well as the legislators who finance the colleges on the basis of numbers of students who are socially promoted. To an outsider, it looks improbable that students score so low on rather elementary diagnostic tests. On the other hand it is clearly a consequence of this series of social promotions of students, where education perishes and the students, instructors, the profession, and the society as a whole stand on the losing side. An individual instructor is thus faced with a dilemma between honorably doing his job (which, under the circumstances would turn out to be rather hard) and participating in the anti-educational process of social promotion. Thus carefully designed diagnostic tests, and measurements related to them, should be an important part of education process.

Comparison with other tests. I believe that the form of the test does affect the data obtained. For instance, I believe that multiple choice tests do not offer as precise an answer of the level of preparation of the students, although through a case study in [2], the correlations discussed are shown to be approximately the same when it comes to multiple choice and non-multiple choice diagnostic tests. My students also neither took any preparation for the diagnostic test, nor was the content of this test known to them in advance. This, in my opinion adds to precision of the test. As already mentioned I am not aware of diagnostic tests designed for specific courses (I have constructed tests for all the courses I teach). Compared to a generic tests mentioned earlier the individual tests of the kind discussed here are certainly more accurate.

Reliability. It is logical to expect that a student's preparation influences his performance in a course. The question to answer is "What is the nature of that relationship?" One can see, in all courses, that, among the students who passed the class (with a grade of "C-" or higher) the lowest diagnostic test score was 26/50, and when the outliers or cases of lowering standards to pass more students are removed, that minimum is 31/50. Among the students who dropped the classes those with the diagnostic score below 30 predominate. There is always a positive linear dependence between the diagnostic and cumulative scores and the percentage of the variance of cumulative scores "explained" by the diagnostic score ranged from 7–63; when outliers are removed the remaining data sets changed this range to 22–63. Among the heterogeneous groups, there are subgroups with homogeneous tendencies of the same kind. The CCCC and the UCBX samples are homogeneous in their own way: the former in how poorly students were prepared and the latter in a sound preparation of the students. This homogeneity then results in a strong linear correlation between the diagnostic and the cumulative scores thus allowing for use of the diagnostic test score as a reliable predictor of a student's overall performance. Generally, in most sociological applications (and number crunching in diagnostic testing qualifies as such), correlation values higher than $+0.2$ or smaller than -0.2 are considered to be substantially important (see e.g. [5]).

Application. After looking through my students' diagnostic tests I advise them as to whether they need to act on their (lack) of preparation for the class. A score of below 30 points is a telling red light (a student is well prepared if his score is 45–50/50). Administratively "prerequisites" for every course should be enforced by having students *recently* (say within past year or maximum two) fulfill the requirements. This would save much energy on everybody's side and one of the beneficial effects is that the attrition rates would be considerably reduced. Another is that standards would not have to be so drastically compromised in the current education practice in order to "pass" students who otherwise do not meet even the basic standards. Some students who end up with substandard diagnostic score either take a preparatory course first, or decide to stay in the course hoping to improve. It is important not to prejudice them with the previous statistics of the test, however one should be realistic at the same time how much encouragement should be given to such students to continue.

Still to be done. It is rather complicated to conduct quantified research in this area, that would be meaningful and free of all bias. For instance, it is hard to control numerous, not well defined factors and their influence on the variables involved, in an observational study of this kind. It would be interesting to construct an experimental study and compare the results. As suggested by one of the referees, one may look into the (improved) predictive power of the diagnostic test scores when test questions with high (above 90%) correct response rate are removed. This certainly has value for correlational analysis, however it should not be forgotten that these tests define standards to a certain extent (better yet the standards in the course may define the lowest passing threshold of the diagnostic test, a posteriori). Further analysis may concentrate on same types (groups) of questions on the test and their predictive power. An additional interesting point is as follows: Paper [4] reports that there is no great correlation between their diagnostic test scores and the end-of-course scores. The author is satisfied with that, for he says that otherwise the remedial measures taken (such as weekly, one hour tutorials/classes for students designated as needing remediation) would have failed; this is under the assumption that lack of correlation is indeed due to remediation. In my own relevant courses there were no systematic remediations of this kind, in the duration of the statistics courses in question. After the diagnostic test, students either took a lower division prerequisite course first, or decision to seek help was left up to individual students, if they decided to stay in the course, despite substandard background. It is an interesting question to answer as to whether students with substandard background should be encouraged to stay and remediated as the course progresses, or else should take a prerequisite course, and then subsequently the statistics course in question. To iterate, what if students who took remediation (successfully) still did not pass the low threshold on the diagnostic, etc?

Finally, it is worth noting that similar results are obtained with diagnostic tests for other courses (such as elementary calculus, etc.) – see [2]; the dependencies and statistics are within the analogous ranges. This study will continue by examining roles of other components of education (listed in the introduction), and of special interest is to look into confounding effects. For instance, a student who has a high diagnostic score may be at the same time a student who devotes substantial time to studying – both components caused by strong working habits and intellectual esteem, etc.

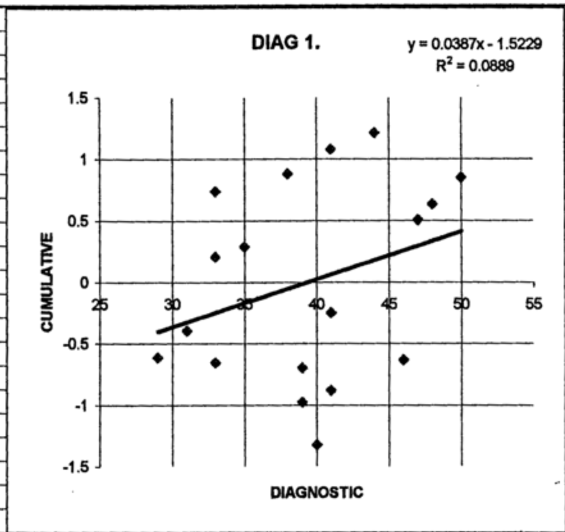
This study is primarily concerned with mathematics education, but there are a number of features that can be useful in other disciplines and education in general.

REFERENCES

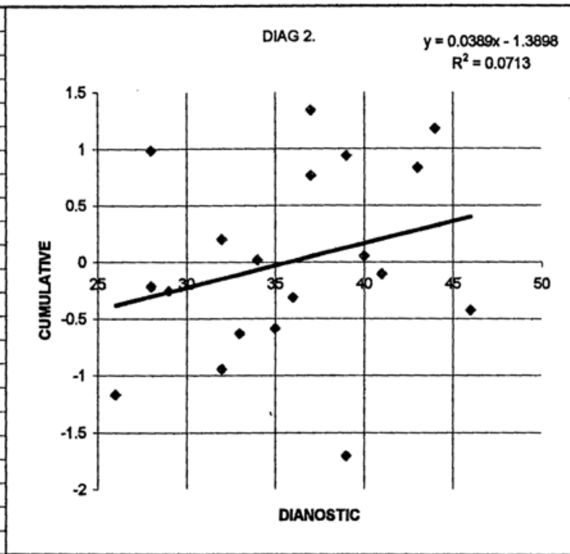
1. California Mathematics Diagnostic Testing Project. UC San Diego, 2004. `http://mdtp.ucsd.edu/test/intro2.htm`
2. R. M. Dimitrić, *Mathematics Education at TAMUG*, Preprint, 2003.
3. Edwards, P., *A Survey of Mathematics Diagnostic Testing on non-Specialist Mathematics Courses*, The Open Learning Foundation, UK, 1996.
4. Edwards, P., *Just How Effective is the Mathematics Diagnostic Test and Follow-up Support Combination?*, Journal of Teaching Mathematics and its applications, **16** (1997), No.3, 118–121.
5. Learning and Teaching Support Network (LTSN), Physical Sciences, MathsTEAM Project, Diagnostic Testing for Mathematics, 2003.
6. Valen E. Johnson, *Grade Inflation. A Crisis in College Education*, Springer: New York, 2003.
7. University of Manitoba Pre-Calculus Diagnostic Test, Department of Mathematics, University of Manitoba, Canada, 2004.

R. M. Dimitrić,
Texas A&M University, Department of Mathematics,
PO Box 1675, Galveston, TX 77553, USA
E-mail: `dimitric@tamug.edu`

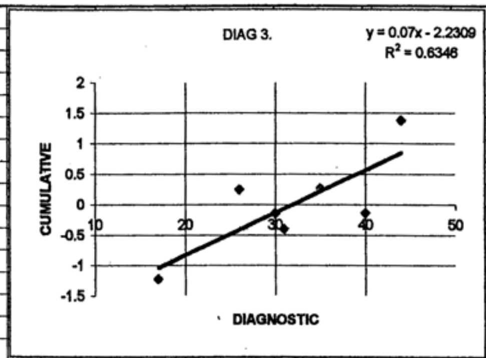
CCSF			DATA 1.				
	DIAG	T1	T2	T3	FINAL	ACTIV	CUM
	44	88	100	72	68	100	1.211
	41	70	99	92	71	70	1.076
	38	60	99	76	63	100	0.881
	50	84	78	92	65	50	0.851
	33	82	75	77	54	95	0.736
	48	87	81	56	58	70	0.632
	47	79	80	49	54	85	0.506
	35	88	81	59	40	60	0.29
	33	70	73	40	51	75	0.209
	41	80	82	32	28	45	-0.25
	31	86	37	62	15	75	-0.39
	29	68	50	52	15	60	-0.61
	46	43	36	47	40	40	-0.63
	33	57	62	30	22	60	-0.65
	39	62	37	38	17	85	-0.69
	41	63	51	29	11	60	-0.88
	39	46	56	26	19	40	-0.97
	40	38	59	20	5	40	-1.32
out of	50	100	100	100	100	100	2.134
mean	39.33	69.5	68.67	52.72	38.67	67.22	-0
stdev	6.136	16.13	21.16	22.29	25.16	20.16	0.797
med	39.5	70	74	50.5	40	65	-0.02



SFSU				DATA 2.			
	DIAG	T1	T2	T3	FINAL	ACTIV	CUM
	37	83	48	73	53	65	1.341
	44	80	44	53	51	90	1.177
	28	75	67	59	39	70	0.988
	39	81	52	52	39	90	0.941
	37	64	48	55	44	75	0.768
	43	88	49	19	43	85	0.836
	32	77	54	38	25	45	0.203
	40	61	27	32	36	65	0.054
	34	64	33	32	31	65	0.022
	41	73	40	2	32	40	-0.1
	28	53	36	10	34	55	-0.21
	29	63	14	6	38	55	-0.25
	36	65	35	10	29	35	-0.31
	46	63	36	8	25	40	-0.42
	35	63	59	7	11	35	-0.58
	33	72	30	8	15	45	-0.63
	32	56	24	6	13	50	-0.94
	26	51	33	7	6	35	-1.16
	39	31	6	7	4	50	-1.7
out of	50	100+1	100+2	100	100	100	
mean	35.74	66.47	38.68	25.47	29.89	57.37	2E-04
stdev	5.753	13.45	15.15	22.88	14.52	18.36	0.838
med	36	64	36	10	32	55	-0.1



CCCC				DATA 3.			
	DIAG	T1	T2	T3	FINAL	ACTIV	CUM
	44	77	86	88	81	90	1.381
	35	74	59	21	51	100	0.269
	26	69	82	34	38	100	0.249
	40	66	61	55	25	100	-0.14
	30	76	59	14	32	100	-0.14
	31	53	47	35	40	90	-0.39
	17	36	48	16	13	90	-1.22
out of	50	100	100	100	100	100	2.398
mean	31.86	64.43	63.14	37.57	40	95.71	1E-04
stdev	8.971	14.98	15.31	26.34	21.69	5.345	0.789
med	31	69	59	34	38	100	-0.14



UCBX				DATA 4.		
	DIAG	T1	T2	FINAL	ACTIV	CUM
	48	95	42	94	60	0.832
	49	83	74	81	95	0.816
	50	88	96	62	65	0.552
	43	91	63	67	80	0.455
	44	84	75	71	70	0.416
	45	78	70	65	95	0.202
	46	83	53	67	65	-0.04
	39	69	61	75	65	-0.18
	47	93	64	43	60	-0.23
	44	75	52	58	60	-0.58
	40	76	44	40	65	-1.07
	41	78	22	46	60	-1.18
out of	50	100	100	100	100	2.288
mean	44.67	82.75	59.67	64.08	70	4E-05
stdev	3.525	7.933	19.02	15.84	12.97	0.676
med	44.5	83	62	66	65	0.202