SECONDARY ANALYSIS OF PUBLISHED RESEARCH IN THE BEHAVIORAL SCIENCES[1]

Leroy Wolins, Iowa State University

Least-square regression procedures, including analysis of covariance and partial correlation, may be thought of as directed toward two conceptually different goals. The first is prediction. In this case the purpose is to predict future events from earlier ones where there is little or no concern about underlying variables or what causes what. The predictor variables represent nothing other than the operationally defined measurement procedures. A typical example of this is the prediction of Grade-Point-Average from test information.

The second goal is more theoretical. The purpose is to attribute causation. In this case the variables represent constructs and the analysis is directed toward answering questions about these constructs rather than the operationally defined measures of them.

When these predictor variables are orthogonal, these regression procedures tend to be robust and informative. When these predictor variables are not orthogonal the analysis for the second goal is likely to be misleading for a variety of reasons. One reason such analyses are misleading has been well documented in the statistical literature (Cochran, 1970; DeGracie and Fuller, 1972) and in the literature in other fields. Measurement error in the predictor variables tends to produce significant partial regression coefficients for correlated measures not necessarily because these measures are tapping different constructs but possibly because each of these measures tap the same construct but with error. As a result of this error a composite made up of these fallible measures will predict the dependent variable better than a single variable will.

However, measurement error in predictor variables is not the only reason the analysis for this second goal may be misleading. Other reasons for misleading results are less familiar but can be no less important. Many of these can be divided into two general categories: non-addivity or lack of interval measurement and irrelevant variance in the predictors.

In regard to non-addivity, imagine an analysis of covariance situation with two groups, men and women in an academic setting, a dependent variable, Salary, and a covariate, Number of Publications. Suppose it happened that at the lower salary levels that Salary was highly dependent on Publications but at the higher salary levels there was little relationship. That is, the relationship between number of publications (X) and salary (Y) is monotonic but negatively accelerated. Also suppose that males exceed the females both in terms of Salary and Publications but all the data lies on a single curved line. Now if you fit these data using the usual covariance analysis, the partial regression coefficient for Sex, controlling for Publications will be significant, indicating to the naive researcher that males are paid more than females. That is, even if sex was irrelevant in nature, the conventional covariance analysis would produce the semblance of an effect. That semblance is attributable entirely to the nonadditive character of the results, a character which is not reflected by the usual analysis.

Of course, such systematic non-addivity is not likely to occur in the social sciences and one reason it is not likely to occur is that the intervals in our scales are arbitrary in size relative to the construct to which they are directed. For a faculty member in an academic institution, Publications is an indicator of a vaguely defined construct, the extent and the quality of an individual's research. Although we may imagine individual faculty members have a true and invariant location on this conceptual dimension, we have no way of knowing how this indicator is related to this dimension. The dependent variable, Salary, should also be functionally related to this true variable if the indicator of the true variable is a proper control variable. However, the only basis we have for determining whether the dependent variable is related to this true dimension <u>in the same way</u> as that dimensions indicator (Publications) is by evaluating the linearity of the relationship between Publications and Salary.

In addition to non-addivity, most variables used in the social sciences reflect more (or less) than their target construct. For example, it seems perfectly reasonable to suppose that Number of Publications is an indicator of the construct, Research Productivity. However, it is equally evident that all publications are not equal. I believe administrators in universities, and committees assigned the task of making decisions about promotions and salary increases make serious efforts to judge the merit of each publication as well as counting papers or articles. As a result, if a group of faculty were evaluated in terms of number of publications for two adjacent five year periods, one might find a high correlation. However this high correlation might represent, in part, the fact that individuals tend to publish in the same area and that certain areas require large scale, involved research whereas in other areas short research efforts are publishable.

This variability that is reliably measured by these control variables which is not reflected in the dependent variable, does exactly the same thing measurement error does but it is much more difficult to accommodate.

This brings us to the data I have at hand.

In the May 1975 issue of Science there appeared an article entitled "Sex Differentials in the Academic Reward System". The authors, Alan E. Bayer and Helen S. Astin, have been extremely cooperative in every way. They not only sent me their data but have provided other information when I requested it and admonitions concerning various intracies about the data I did not request but should have. This kind of cooperation deserves special commendation, especially since they knew my purpose, criticism, and other authors whose data I requested tended to beg off, dismiss the matter perfunctorily, or not respond at all. I do intend to pursue the problems with these other authors but today unless there is extra time, I will confine my comments to the Bayer-Astin article.

Table 1 contains Salary data based on groups homogeneous with respect to Sex, Academic Rank and Departmental Affiliation. These data are based on about 2000 Males and 2000 Females who filled in their questionnaires properly. Since fewer females than males occur in the academic population, females were sampled more intensively than the males in order to obtain approximately equal numbers of each. This explains some of the peculiar results in Table 2. In the last two columns of Table 2 are the numbers for each sex occuring at each Rank and Departmental Affiliation. For example, females tend to predominate at the lower ranks for this sample whereas males tend to predominant at all ranks in the academic population.

In Figure 1, I have plotted the mean against the standard deviation of salary for each group with data for twenty-five or more individuals. It is quite evident from this figure that these two statistics are not independent which suggests non-addivity. A logarithmic transformation is suggested by the fact that merit increases seem to be based on a percentage of previous salary. Also I tried a square-root transformation and it did not work as well as the logarithmic transformation. In Figure 2 is depicted the same data expressed on the log scale. The apparent relationship between $\bar{X}$ and S is considerably reduced on the log scale.

In Figure 3, only groups containing at least 25 males and females at each Rank and Departmental Affiliation are depicted. For example, for $R = 1$ and $D = 9$, only the females have more than 25 cases so this group is not depicted in Figure 3. From the Figure, it is also evident that all three variables, Rank, Departmental Affiliation and Sex are related to Salary. Moreover it appears that the sex differential increases with salary.

In Figure 4 is depicted the same data on the log scale. Here the differential appears more nearly constant.

(By the way, the two outliers are for the "Health" group. For the males about 80% hold doctorates, for the females about 30% hold doctorates. For the males more than 80% of these doctorates are professional degrees where less than 50% are professional degrees for the females. These results are similar for all three health groups but less marked for the full professors in this health group.)

From these results it appears evident that we are better off, from a statistical point of view if further analysis is based on log salary rather than salary.

In Figure 5, log salary is plotted against number of articles published. Number of articles published is a coded variable but is a monotonic function of the actual number of publications. In this Figure there appears to be six outliers. All six are affiliated with departments of Biology. Except for these six outliers, which appear to reward the females more than the males, no sex differential is evident.

In Figure 6, the number of articles published for Males and Females are plotted. Again it is evident there is a sizable differential favoring the Males for most ranks and departmental affiliations.

In Figure 7 is plotted Log Dollars against Number of Books. Like Number of Publications, this is a coded variable. The three outliers are for the Male "Health" groups. These groups as all ready noted, contain disproprortionate numbers of Ph.D.'s and professional doctorates as compared to females in these "Health" groups comparable in rank.

Bayer and Astin, from their regression analysis, attribute a salary differential from $600 to $1000 per year to sex, favoring males. It is my contention that this result would be expected for this type of analysis because the independent variables used for control are in some sense fallible or the dependent variable is not linearly related to them, or both. The data I have presented seem to indicate people of the same Academic Rank, who publish the same amount, receive the same pay.

As made clear by Bayer and Astin, and very evident from the data presented here, there are sex differences. However, it is not clear that the institution responsible for these differences is the academic one. For example, I note that males at all ranks have more children than females but this differential increases with increasing rank. This suggests that children impede the progress of females more so than males. On the other hand, it appears that males and females report that their academic careers were interrupted about equally often at the lower ranks but males at the full professor rank report more career interruptions than their female counterparts, probably because of military service during World War II.

Although these data do not support a sex differential in the academic reward system, it is essential to understand that these data do not indicate that men and women are equally rewarded either. For example, if one chose to use number of publications as the dependent variable and regressed this variable on the remaining ones, including salary, it is likely that the sex variable would be significant, erroneously indicating men publish more than women when salary etc., is controlled. By exploiting this regression phenomena, one can use data to support whatever position one wishes, under null conditions. Of course, if a strong sex effect were present, then this regression procedure would find it, however, so would this rather casual approach used here.

Bayer and Astin claim these estimates of the sex differential are under estimates for various reasons. I claim they are overestimates for other reasons. In a sense, we agree that these are not good estimates. On these bases I claim analysis of data like this should not exceed in complexity what I have done here. If clear sex differentials are not evident from examining the means for large homogeneous groups, then the data are not adequate for inferring a differential. Seeking a precise probability estimate or an unbiased estimate of a difference for data like this is futile. There is no reasonable way to obtain such precision given fallible measurement.

The reasons that such analyses are never appropriate to adjust for group differences on covariates are as follows:

1) Social science variable do not have interval properties. As a result the construct measured by the covariate may not be functionally related to the true measure in the same way as the measure of the same construct by the dependent variable.
2) The dependent variable may result, in part, from variables not measured by the covariates. For these data it could be that women or men perform better in the classroom and this fact is reflected in the salary variable but not in the covariates.
3) The covariates are fallible measure of their target constructs. As previously indicated number of publications is not likely to reflect completely an individuals research productivity. Research Productivity maybe more completely reflected in salary because administrators or committees may actually scrutinize and evaluate publications.
4) Additivity, particularly linearity, can not be assumed for variables of the kind used in social science investigations. Linear models in the social sciences are used to roughly fit monotonic relationships. However the ad hoc measurement procedures do not allow one to put much faith in a linear model really fitting data. When two groups differ on a covariate and the relationship between the variate and covariate is non-linear, some degree of bias in the results of the covariance adjustment is bound to occur.

What Bayer and Astin have shown from their regression analysis is that one can predict salary with sex as a variable better than if one ignores sex. However it is not clear that this increased prediction is due to sex per se or because these other variables which also reflect sex differences do so fallibly. Since number of publications etc. fallibly reflect research productivity etc., it can not be clear from these data that women are paid less because they are women or because they publish less.

---

[1]Supported by National Institute of Education Contract Number NIE-C-74-0115.

REFERENCES

Bayer, Alan E. and Helen S. Astin. (1975). "Sex differentials in the academic reward system". Science 188: 796-802.

Cochran, W. G. (1970). "Some effects of errors of measurement on multiple correlation". Journal of the American Statistical Association 65: 22-34.

DeGracie, J. S. and W. A. Fuller. (1972). "Estimation of the slope and analysis of covariance when the concomitant variable is measured with error". Journal of the American Statistical Association 67: 930-937.

Figure 1. Plot of $\bar{X}$ and S for the original scale, Dollars

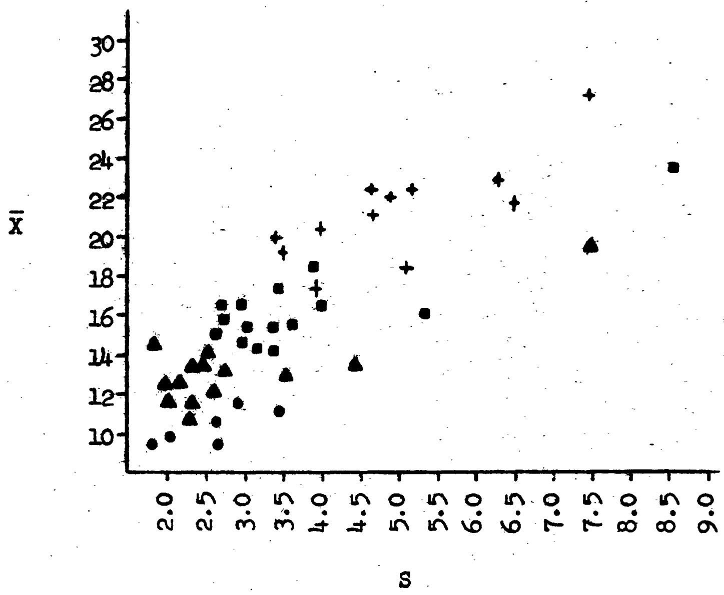

Figure 2. Plot of $\bar{X}$ and S for the transformed scale, Log Dollars

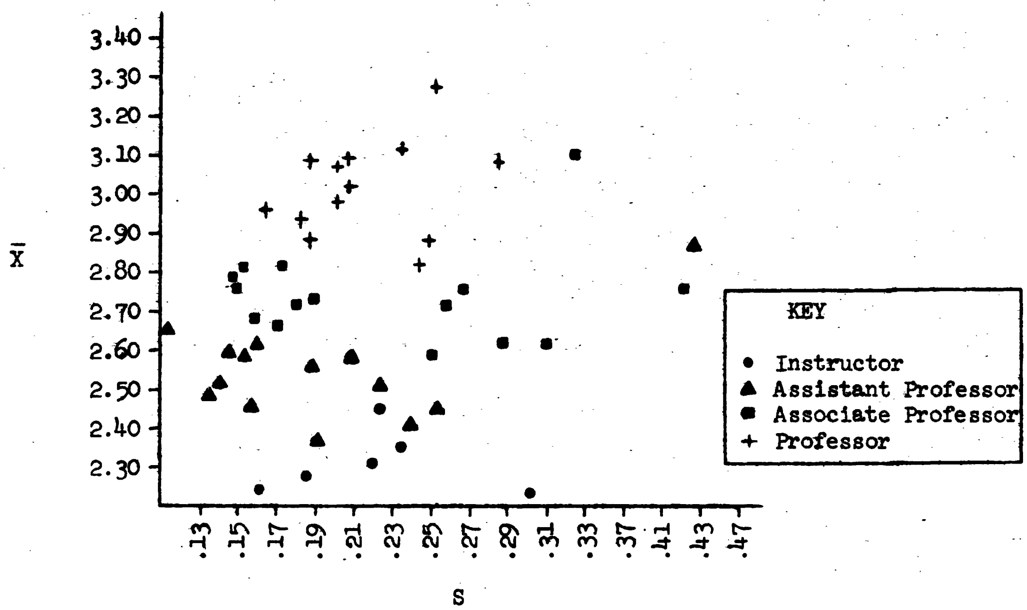

Figure 3. Mean salary (original scale) of Males (M) and Females (F) in the same field and holding the same academic rank

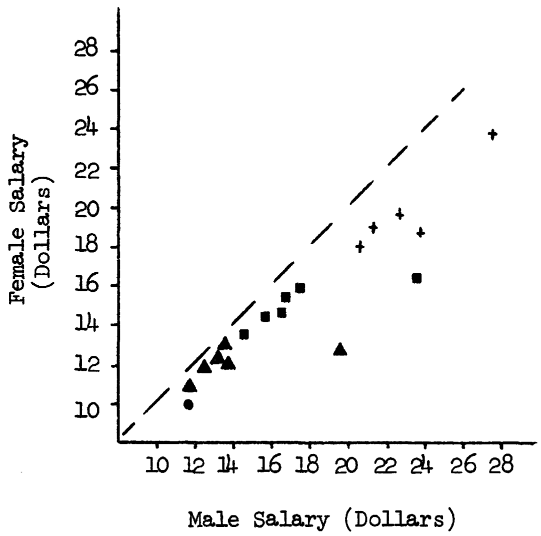

Figure 4. Mean salary (log scale) of Males (M) and Females (F) in the same field and holding the same academic rank

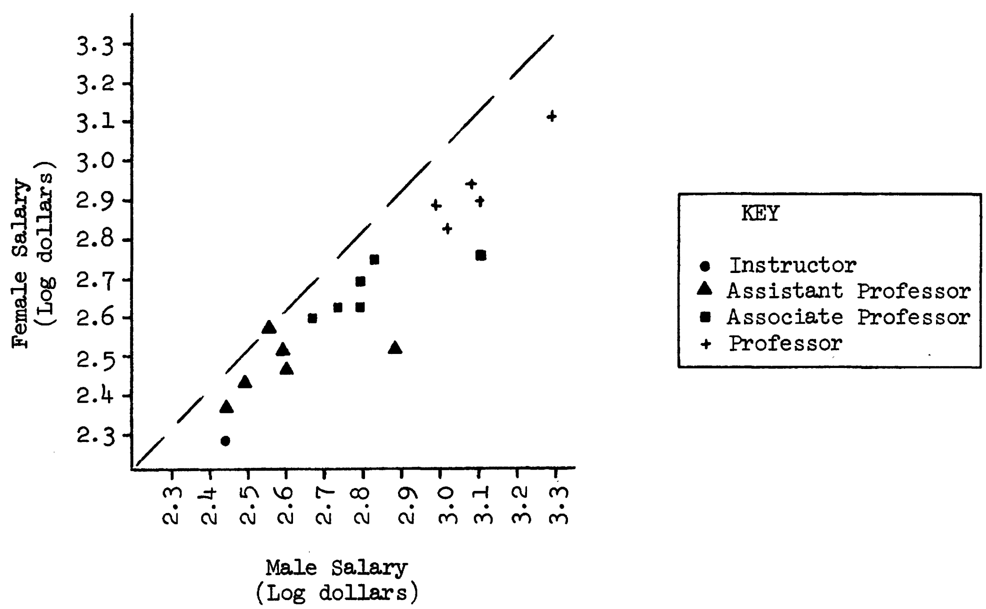

Figure 5. Number of Articles Published (# A) Plotted Against Log Salary (L $)

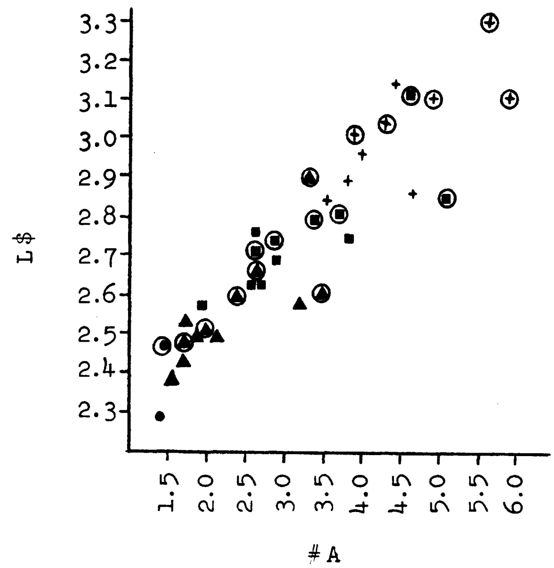

Figure 6. Number of Articles Published for the Male and Female Groups

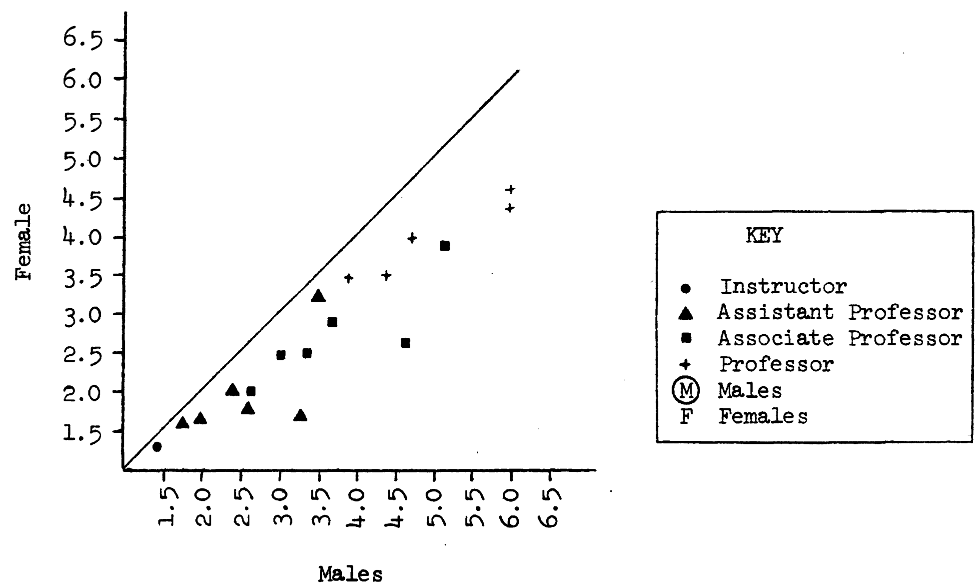

Figure 7. Number of Books Published (# B) Plotted Against Log Salary (L $)

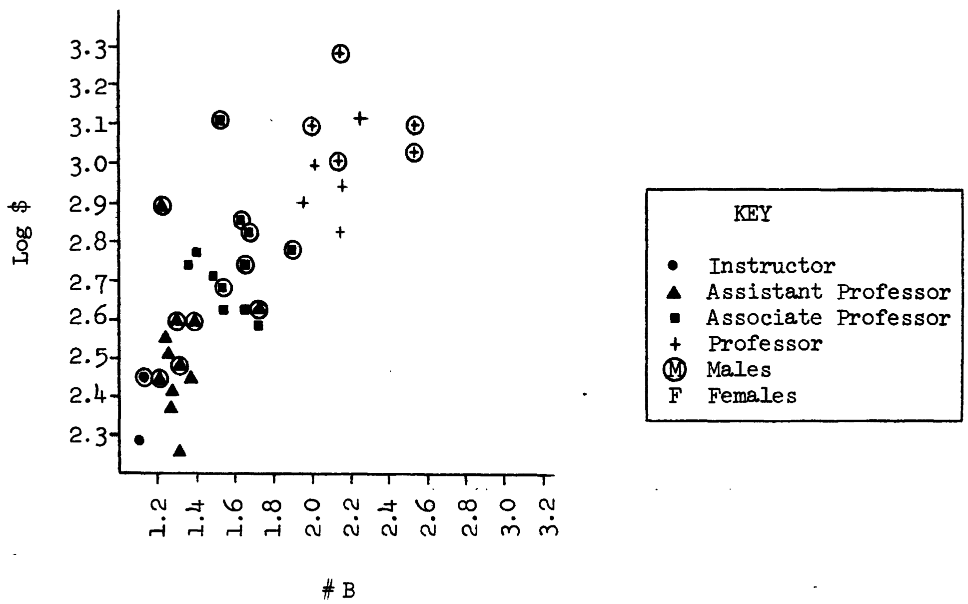

Table 1: Means and Standard Deviations of Faculty Salary (in units of $1000) by Academic Rank (R) and Departmental Affiliation (D)

		Dollars (in thousands)				Log Dollars			
		Females		Males		Females		Males	
R[1]	D[2]	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s	$\bar{X}$	s
1	1								
	2	9.77	2.77			2.23	.328		
	3								
	4								
	5								
	6	10.48	2.87			2.32	.243		
	7	9.51	1.80			2.24	.182		
	8	10.02	2.19	11.96	2.97	2.28	.207	2.45	.245
	9	11.00	3.55			2.36	.256		
2	1								
	2	12.02	2.72	13.76	2.50	2.46	.276	2.61	.180
	3	13.27	2.76	13.65	4.47	2.56	.210	2.58	.230
	4			13.30	2.37			2.57	.167
	5			14.32	1.87			2.65	.132
	6	12.49	2.15	13.61	2.50	2.51	.163	2.59	.173
	7	10.92	2.40	11.82	2.09	2.37	.211	2.45	.178
	8	11.59	2.40	12.30	2.07	2.42	.259	2.49	.156
	9	12.90	3.65	19.72	7.53	2.52	.245	2.89	.476
3	1			16.72	4.11			2.75	.454
	2	14.22	3.40	16.47	2.78	2.61	.338	2.79	.168
	3	15.77	3.44	17.27	3.49	2.74	.210	2.83	.194
	4			15.48	3.16			2.71	.279
	5			16.00	2.79			2.76	.168
	6	15.02	2.66	16.72	3.10	2.69	.180	2.80	.171
	7	13.74	4.24	14.71	3.01	2.58	.271	2.67	.194
	8	14.04	3.19	15.65	3.67	2.61	.308	2.73	.205
	9	16.29	5.38	23.72	8.62	2.75	.289	3.10	.371
4	1								
	2	18.39	3.93	20.15	4.14	2.89	.209	2.98	.221
	3	18.93	5.12	22.88	6.30	2.90	.266	3.09	.306
	4			21.97	4.90			3.07	.222
	5			22.51	4.69			3.09	.209
	6	19.20	3.61	22.56	5.14	2.94	.203	3.09	.226
	7			19.46	3.53			2.95	.183
	8	17.47	3.97	20.97	4.67	2.83	.262	3.02	.227
	9	23.44	6.14	27.35	7.49	3.12	.255	3.27	.272

[1] 1 - Instructor
2 - Asst. Professor
3 - Assoc. Professor
4 - Professor

[2] 1 - Business
2 - Education
3 - Biology
4 - Physical Sciences
6 - Social Science
7 - Fine Arts
8 - Humanities
9 - Health

Table 2: Number of Articles (#A) and Books (#B) Published for Males (M) and Females (F) by Academic Rank (R) and Departmental Affiliation (D)

		#A		#B			
R	D	F	M	F	M	N_F	N_M
1	1					5	5
	2	1.31		1.26		105	13
	3					15	5
	4					8	5
	5					0	6
	6	1.57		1.28		47	13
	7	1.40		1.14		35	4
	8	1.30	1.42	1.10	1.12	122	26
	9	1.36		1.10		69	2
2	1					12	22
	2	1.78	2.62	1.35	1.71	184	55
	3	3.19	3.50	1.22	1.35	59	46
	4		3.54		1.14	10	69
	6		2.66		1.29	2	41
	6	2.02	2.32	1.34	1.33	93	100
	7	1.60	1.74	1.29	1.21	63	38
	8	1.67	1.99	1.30	1.34	178	120
	9	1.71	3.28	1.23	1.16	107	25
3	1		3.88		1.72	8	25
	2	2.53	3.32	1.55	1.87	143	68
	3	3.88	5.16	1.35	1.66	62	56
	4		4.28		1.45	15	92
	5		3.68		1.37	2	57
	6	2.93	3.52	1.53	1.69	102	95
	7	1.91	2.67	1.71	1.56	34	45
	8	2.52	2.96	1.68	1.66	124	103
	9	2.63	4.64	1.41	1.52	51	25
4	1					7	15
	2	3.56	3.88	1.95	2.14	113	94
	3	4.63	5.83	2.00	2.01	41	92
	4		5.27		1.87	12	137
	5		4.94		1.90	1	83
	6	4.00	4.65	2.14	2.58	69	140
	7		3.02		1.98	21	46
	8	3.47	4.30	2.13	2.53	76	148
	9	4.44	5.65	2.24	2.15	25	34