

AN ITEM RESPONSE ANALYSIS OF STATISTICS USE IN THE WORKPLACE

John A Harraway

University of Otago, New Zealand

Dalton F. Andrade

Federal University of Santa Catarina, Brazil

jharraway@maths.otago.ac.nz

Training in statistics at university should be informed at least in part by what graduates will have to do with acquired statistical knowledge after graduation. A sample of 977 employed graduates with PhD and Masters degrees in seven specialties with statistics pre-requisites at university identifies which of 46 statistics based techniques (the items) they use in their work. A two parameter item response model uses 32 of the 46 items to build a scale measuring the extent of statistics use in the workplace and creates a value for each graduate which is used to summarize differences between the use of statistics in the seven specialties. Implications for syllabus construction to better prepare graduates for the workplace are discussed.

INTRODUCTION

Statistical activities and research methods which graduates encounter in the workplace can be useful for planning academic curricula that are aligned with primary workplace needs. This study quantifies the extent of use of statistics in the workplace. Harraway and Barker (2005) collated information from 977 New Zealand university PhD and Masters graduates in statistics, economics, marketing, the biological sciences, nutrition, food science and psychology. Of the 977 respondents, 98% were employed or recently employed indicating opinions expressed represented the views of employed graduates with higher degrees.

The 46 statistical techniques and research methods nominated in the survey and listed in Harraway and Barker (2005) are the 46 items for an item response analysis. For each respondent the value 1 indicates an item is *used* by that respondent in the workplace while the value 0 indicates the item is *not used*. The data matrix has binary entries with 977 rows and 46 columns. The item response analysis creates a scale for the *extent of statistics use in the workplace*, places each item in a position on this scale indicating how much that item is used in the workplace, and compares statistics use for all graduates in each of the seven specialties.

METHOD

Items for building the scale of *extent of statistics use in the workplace* are identified using the two parameter item response function

$$\text{Prob}(U_i = 1 | \theta_j) = \frac{1}{1 + \exp[-a_i(\theta_j - b_i)]}$$

where U_i is a binary random variable with value 1 when item i is used in the workplace, 0 otherwise and θ_j represents the level of the *extent of statistics use in the workplace* of individual respondent j . The two parameters a_i and b_i are the “discrimination” and “difficulty” parameters of item i respectively (Lord, 1980 or Andrade *et al.*, 2000). All calculations are carried out using the software BILOG-MG (Zimowski, *et al.*, 1996), based on the scale with mean 0 and standard deviation 1.

A high value for b_i suggests item i is difficult, which in the context of this study indicates that the particular item is not widely used in the workplace. A high value for a_i suggests item i will discriminate well between *use* and *no use* in the workplace and therefore will be important in the development of the *extent of statistics use in the workplace* scale.

The item parameters are estimated by the marginal maximum likelihood method. Items that present a value for the discrimination parameter greater than or equal to 0.65 are selected for construction of the scale. This identifies 32 items (statistical topics or research methods). Using

the estimates of the item parameters so obtained, empirical Bayesian estimates are found for the *extent of use* value for each respondent (Baker and Kim, 2004). These estimates are rescaled to have mean 50 and standard deviation 15 with most of the values therefore varying from 0 to 100. The results ordered by item use are shown in Table 1.

In order to interpret the *extent of statistics use* scale we define anchoring levels (Beaton and Allen, 1992) of 5, 20, 35, 50, 65, 80, 95 and 110 on the scale and identify an anchoring item by the following definition: if X and Y are two consecutive anchoring levels define item i to be an anchoring item for level Y if and only if

- (i) $\text{Prob}(U_i = 1 | \theta = Y) \geq 0.65$
- (ii) $\text{Prob}(U_i = 1 | \theta = X) \leq 0.50$
- (iii) $\text{Prob}(U_i = 1 | \theta = Y) - \text{Prob}(U_i = 1 | \theta = X) \geq 0.30$

That is, respondents at level Y are *using* the item and respondents at level immediately before X are *not using* the item.

RESULTS

The 32 item parameters and scale probabilities for each item are shown in Table 1. The levels where an item i can be considered as an anchoring item are highlighted using brackets. Since a reasonable number of items do not satisfy this definition we further define items for quasi anchoring if they are close to these defined levels. We consider these to be anchoring. They are also in brackets.

Table 1: Item parameters and levels on extent of use scale.

Item	Item Parameters		Levels of the scale (50,15)							
	a_i	b_i	5	20	35	50	65	80	95	110
GRAPHS	0.0825	39.21	0.06	0.17	0.41	(0.71)	0.89	0.97	0.99	1.00
TESTS	0.1156	49.72	0.01	0.03	0.15	0.51	(0.85)	0.97	0.99	1.00
SLREG	0.1164	50.26	0.01	0.03	0.14	0.49	(0.85)	0.97	0.99	1.00
ANOVA	0.1375	54.92	0.00	0.01	0.06	0.34	(0.80)	0.97	1.00	1.00
POSTHOC	0.1114	61.53	0.00	0.01	0.05	0.22	(0.60)	0.89	0.98	1.00
MULTREG	0.1135	61.78	0.00	0.01	0.05	0.21	0.59	(0.89)	0.98	1.00
FACTDES	0.0854	69.31	0.00	0.01	0.05	0.16	0.41	(0.71)	0.90	0.97
PCA	0.1301	69.48	0.00	0.00	0.01	0.07	0.36	(0.80)	0.97	0.99
MANOVA	0.0959	69.81	0.00	0.01	0.03	0.13	0.39	(0.73)	0.92	0.98
NONLREG	0.0939	70.00	0.00	0.01	0.04	0.13	0.38	(0.72)	0.91	0.98
REPMEAS	0.0917	70.22	0.00	0.01	0.04	0.14	0.38	(0.71)	0.91	0.97
LOGREG	0.0830	71.84	0.00	0.01	0.04	0.14	0.36	(0.66)	0.87	0.96
POWER	0.0800	73.84	0.00	0.01	0.04	0.13	0.33	(0.62)	0.84	0.95
THEORY	0.0458	74.15	0.04	0.08	0.14	0.25	0.40	(0.57)	0.72	0.84
NONPREG	0.0954	74.21	0.00	0.01	0.02	0.09	0.29	(0.63)	0.88	0.97
RANDEFTS	0.1111	74.39	0.00	0.00	0.01	0.06	0.26	(0.65)	0.91	0.98
CLUSTER	0.0780	74.48	0.00	0.01	0.04	0.13	0.32	(0.61)	0.83	0.94
BLOCKING	0.0782	75.30	0.00	0.01	0.04	0.12	0.31	(0.59)	0.82	0.94
DISCRIM	0.1093	76.79	0.00	0.00	0.01	0.05	0.22	(0.59)	0.88	0.97
LGLINMOD	0.0805	77.30	0.00	0.01	0.03	0.10	0.27	0.55	0.81	0.93
COMPINT	0.0597	79.27	0.01	0.03	0.07	0.15	0.30	0.51	0.72	0.86
EXPLORFA	0.0734	79.84	0.00	0.01	0.04	0.10	0.25	0.50	(0.75)	0.90
CANCORR	0.1076	80.96	0.00	0.00	0.01	0.03	0.15	0.47	(0.82)	0.96
MDS	0.0876	81.02	0.00	0.00	0.02	0.06	0.20	0.48	(0.77)	0.93
CORRESP	0.1082	81.85	0.00	0.00	0.01	0.03	0.14	0.45	(0.81)	0.95
PATHANAL	0.0715	81.89	0.00	0.01	0.03	0.09	0.23	0.47	(0.72)	0.88
CROSSOV	0.0608	84.96	0.01	0.02	0.05	0.11	0.23	0.43	(0.65)	0.82
MRKRECAP	0.0574	87.09	0.01	0.02	0.05	0.11	0.22	0.40	0.61	0.79
SURVANAL	0.0648	87.65	0.00	0.01	0.03	0.08	0.19	0.38	(0.62)	0.81
META	0.0584	88.91	0.01	0.02	0.04	0.09	0.20	0.37	0.59	0.77
BAYESIAN	0.0634	90.43	0.00	0.01	0.03	0.07	0.17	0.34	0.57	0.78
STOCHAST	0.0532	94.56	0.01	0.02	0.04	0.09	0.17	0.32	0.51	0.69

The values of b_i place the 32 items i in order of use. The numbers of respondents reaching the levels on the *extent of statistics use* scale are calculated. Table 2 reports the frequency and percentage for each one of the seven specialties considered and also the total.

Table 2: Frequency (percentage) of respondents by specialty

Specialty		Level of Statistics Use					Total
		20	35	50	65	80	
Stat	Count(%)	30(25)	33(27)	35(29)	19(16)	4(3)	121(100)
Econ	Count(%)	10(12)	32(38)	32(38)	11(13)	0(0)	85(100)
Mark	Count(%)	28(34)	33(40)	15(18)	4(5)	2(2)	82(100)
Biosci	Count(%)	71(21)	80(23)	101(29)	73(21)	20(6)	345(100)
Foodsc	Count(%)	7(23)	14(47)	4(13)	4(13)	1(3)	30(100)
Nutrit	Count(%)	7(23)	4(13)	13(42)	6(19)	1(3)	31(100)
Psych	Count(%)	70(25)	80(28)	77(27)	54(19)	2(1)	283(100)
Total	Count(%)	223(23)	276(28)	277(28)	171(18)	30(3)	977(100)

It follows that values over 80 for b_i suggest that the matching item is not widely used while low values indicate extensive use in the workplace of the item. For example the values for Level 50 and Level 65 show that GRAPHS, TESTS, SLREG, ANOVA and POSTHOC are relatively widely used while the values of Level 95 indicate EXPLORFA, CANCEORR, MDS, CORRESP, PATHANAL, CROSSOV, SURVIVAL are not used much relatively.

We can also conclude that only 3% (30 out of 977) reached Level 80 indicating that only 3% of the respondents use the items MULTREG, FACTDES, PCA, MANOVA, NONREG, REPMEAS, LOGREG, POWER, NONPREG, RANDEFTS, CLUSTER, BLOCKING and DISCRIM. As none of the respondents reached Level 95, we can further conclude that the items EXPLORFA, CANCEORR, MDS, CORRESP, PATHANAL, CROSSOV and SURVANAL are not used by the respondents.

DISCUSSION

The scale reported in Table 1 and the percentages reported in Table 2 show that graphs, introductory testing procedures, simple linear regression, analysis of variance and posthoc tests are widely used. These are the main topics in introductory courses on statistical methods and this study confirms a need for students in the specialties reviewed to be taught these items to prepare them for the workplace. On the other hand survival analysis, meta analysis, Bayesian statistics and stochastic processes are not used. These topics are rarely taught in service statistics courses and since the specialties with the majority of respondents have only service statistics needs these conclusions may not be accurate for the graduates in statistics.

Table 3: Percentage of respondents that reached the level

Specialty	Level of Statistics use					
	20	35	50	65	80	95
Stat	100	75.21	47.93	19.01	3.31	0.00
Econ	100	88.24	50.59	12.94	0.00	0.00
Mark	100	65.85	25.61	7.32	2.44	0.00
Biosci	100	79.42	56.23	29.96	5.80	0.00
Foodsc	100	76.67	30.00	16.67	3.33	0.00
Nutrit	100	77.42	64.52	22.58	3.23	0.00
Psych	100	75.27	47.00	19.79	0.71	0.00

In Table 3. the percentages of the respondents in each of the seven specialties who reached the designated levels of *statistics use* are reported. The clear downward trend for each specialty is shown in Figure 1. Inspection of the values for Level 65 shows there is greater use of statistics in the sciences than in the business areas and this should be reflected in the amount of statistics included in the different specialties at university. There are significant differences between the specialties at Level 50.

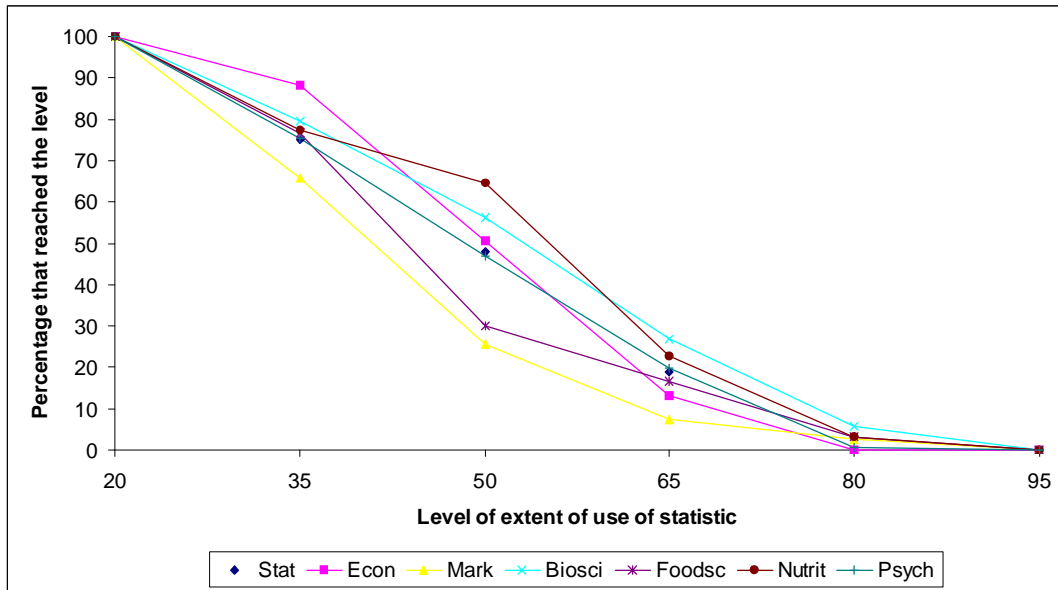


Figure 1: Specialty percentage by extent of use.

Further research could focus on statistics graduates in particular with an even more comprehensive list of statistics topics and research methods surveyed. Students who had completed both undergraduate and postgraduate statistics degrees could be included. This should provide insight into the statistical items we should be teaching in our courses for statistics specialists in order to prepare them for the workplace. An analysis of statistics software use in the workplace is also currently underway. Again this will suggest the most useful statistics software to include in our teaching.

REFERENCES

- Andrade, D. F., Tavares, H.R. and Valle, R.C. (2000). *Teoria da resposta ao item: Conceitos e aplicacoes*. Sao Paulo: Associacao Brasileira de Estatistica.
<http://www.inf.ufsc.br/~dandrade/tri>.
- Baker, F. B. and Kim, S-H. (2004). *Item Response Theory: Parameter Estimation Techniques*. Dekker Series of Textbooks and Monographs. New York: Marcel Dekker, Inc.
- Beaton, A. E. and Allen, N. L. (1992). Interpreting scales through scale anchoring. *Journal of Educational Statistics*, 17, 191-204.
- Harraway, J. A. and Barker, R. J. (2005). Statistics in the workplace: A survey of use by recent graduates with higher degrees. *Statistics Education Research Journal*, 4(2), 43-58,
<http://www.stat.auckland.ac.nz/serj>.
- Lord, F. M. (1980). *Applications of Item Response Theory to Practical Testing Problems*. Hillsdale: Lawrence Erlbaum Associates Inc.
- Zimowski, M. F., Muraki, E., Mislevy, R. J. and Bock, R. D. (1996). *BILOG-MG: Multiple-Group IRT analysis and Test Maintenance for Binary Items*. Chicago: Scientific Software Inc.