Psychometric data printed for examination results from Computing and Communication Services

"CCS_exam_results.doc" prepared by Dwayne Schindler (extension 4205) March 29, 2009

```
   Question   2                                                                Key 4, Index 4
*********************************************************************************************
*                                                                                           *
* Group     Number           Other       1          2          3          4          5      *
*                           n     p    n     p    n     p    n     p    n     p    n     p  *
* Low         18            0  0.00    0  0.00   10  0.56    0  0.00    8  0.44    0  0.00  *
* Middle      27            0  0.00    0  0.00    7  0.26    1  0.04   19  0.70    0  0.00  *
* High        18            0  0.00    0  0.00    4  0.22    0  0.00   14  0.78    0  0.00  *
* Total       63            0  0.00    0  0.00   21  0.33    1  0.02   41  0.65    0  0.00  *
*                                                                                           *
* Z-Score                     N/A        N/A       -0.47      -0.13       0.25       N/A    *
* Mean Test Score             N/A        N/A       31.86      35.00      38.41       N/A    *
*                                                                                           *
*********************************************************************************************

           Difficulty:  0.65,           Bi-Serial:  0.41,           Point-Bi-Serial:  0.34
 Corrected Difficulty:  0.65, Corrected Bi-Serial:  0.40, Corrected Point-Bi-Serial:  0.32
       Discrimination:  0.33
```

Groups are split into three groups: the top 27 percent, the bottom 27 percent, and the middle 46 percent. In the current example, the 63 subjects were split into 18 students having high scores (28.57% of the group), 18 students having low scores (28.57% of the group), and 27 students between those extremes (42.86% of the group).

Difficulty is equal to the probability of obtaining the correct response. High values indicate easy items whereas low values indicate difficult items.

$$\begin{aligned} Difficulty &= \frac{Correct\ responses}{Total\ number\ of\ responses} \\ &= \frac{41}{63} \\ &= 0.65 \end{aligned}$$

Discrimination is the difference between the proportions of individuals responding correctly in extreme groups; it ranges in value from +1.0 to -1.0. (Sax, 1989). Values that are positive indicate items that discrimination appropriately between the high group and the low group. Values that are negative indicate items that the low group responded to correctly but for some reason the high group did not. These would be considered suspect items. Values close to zero indicate items for which both groups did equally well and, thus, provide no discrimination value.

$$\begin{aligned} Discrimination &= \frac{Correct\ responses\ in\ high\ group}{Total\ number\ of\ responses\ in\ high\ group} - \frac{Correct\ responses\ in\ low\ group}{Total\ number\ of\ responses\ in\ low\ group} \\ &= \frac{14}{18} - \frac{8}{18} \\ &= 0.7778 - 0.4444 \\ &= 0.33 \end{aligned}$$

Thus 77.78% of the students in the upper group but only 44.44% of the students in the lower group responded correctly to this item.

Point-biserial correlation

One important characteristic of the item discrimination index is the extent to which it differentiates between the high and low performers on the test. An index with this discrimination can be obtained by correlating the performance on the item with the total test or subtest score. Thus, one type of item discrimination index consists of a correlation coefficient. Assuming that an item can be scored right or wrong, and that the total test or subtest score assumes at least interval scale measurement, we have the conditions for applying the point-biserial correlation coefficient. The point-biserial correlation-coefficient is the product-moment correlation when one variable is dichotomous, and the other variable is continuous and measured on at least an interval scale.

$$r_{pb} = \frac{\overline{X}_H - \overline{X}_L}{s}\sqrt{pq}$$

where $\overline{X}_H$ = mean of the scores on the continuous variable of the individuals passing the item

$\overline{X}_L$ = mean of the scores on the continuous variable of the individuals failing the item

s = the standard deviation of all scores on the continuous variable

p = the proportion of individuals responding correctly to the item

q = the proportion of individuals responding incorrectly to the item

Items with high point-biserial correlations are usually retained, and those with low or negative correlations are rejected. An item with a high correlation is retained because the high value indicates similarity (over the group of examinees) between performance on the item and performance on the total test or subtest (Lemke & Wiersma, 1976; Klugh, 1974).

Biserial correlation

This correlation is similar to the point-biserial correlation except that it assumes that the dichotomous split is the result of recoding a continuous measure. Thus, there is an underlying continuous measure (distribution) with the success/failure simply representing the two extremes of the continuum.

$$r_{pb} = \left(\frac{\overline{X}_H - \overline{X}_L}{s}\right)\left(\frac{pq}{y}\right)$$

where $\overline{X}_H =$ mean of the scores on the continuous variable of the individuals passing the item

$\overline{X}_L =$ mean of the scores on the continuous variable of the individuals failing the item

$s =$ the standard deviation of all scores on the continuous variable

$p =$ the proportion of individuals responding correctly to the item

$q =$ the proportion of individuals responding incorrectly to the item

$y =$ ordinate (height) of the unit normal curve at the point of the division between the p and q proportions under the curve

The easiest method to determine the pq/y is to obtain the value from the table below that was published in Guilford (1954).

p(or q)	pq/y	p(or q)	pq/y	p(or q)	pq/y	p(or q)	pq/y
.99	.3715	.86	.5409	.73	.5961	.60	.6212
.98	.4048	.85	.5468	.72	.5989	.59	.6223
.97	.4277	.84	.5524	.71	.6015	.58	.6232
.96	.4456	.83	.5576	.70	.6040	.57	.6240
.95	.4605	.82	.5625	.69	.6063	.56	.6247
.94	.4735	.81	.5671	.68	.6085	.55	.6253
.93	.4848	.80	.5715	.67	.6106	.54	.6258
.92	.4951	.79	.5756	.66	.6124	.53	.6262
.91	.5043	.78	.5796	.65	.6142	.52	.6264
.90	.5128	.77	.5832	.64	.6158	.51	.6266
.89	.5206	.76	.5867	.63	.6174	.50	.6267
.88	.5279	.75	.5900	.62	.6188	.	
.87	.5346	.74	.5931	.61	.6200		

Caution Index

$$CI = \frac{\sum_{j=1}^{y_{i.}} (1 - y_{ij})(y_{.j}) - \sum_{j=y_i+1}^{J} (y_{ij})(y_{.j})}{\sum_{j=1}^{y_{i.}} y_{.j} - (y_{i.})(u')}$$

$$\begin{aligned}
&\textit{where } i = 1,\ 2, \cdots I \ \textit{ indexes the students} \\
&\qquad j = 1,\ 2, \cdots J \ \textit{ indexes the questions} \\
&\qquad y_{ij} = \textit{the response of student } i \textit{ on question } j \\
&\qquad y_{i.} = \textit{the number of correct responses of student } i \textit{ on all test questions} \\
&\qquad y_{.j} = \textit{the number of correct responses of all students on question } j \\
&\qquad u' = \textit{the average number of correct responses on all questions}
\end{aligned}$$

While there are a number of variations in the formula, the one indicated above is probably the easiest to apply using a simple program such as Microsoft Excel.

Sato suggests that a standard Caution Index value is 0.5. If CI is higher than 0.5, the administrator of the test ought to pay attention to the corresponding student.

	s5	s7	s9	s4	s10	s2	s13	s14	s6	s1	s15	s11	s3	s8	s12	Total	Caution Index a	b	c	d
q3	1	1	1	1	1	1	1	1	0	1	1	0	0	0	1	11	0		11	
q2	1	0	1	1	1	1	1	1	1	1	0	1	1	0	0	11	0		11	
q7	1	1	1	1	1	0	0	1	0	1	1	0	1	1	0	10	0		10	
q4	1	1	1	1	0	1	1	1	1	0	1	0	0	0	0	9	9		9	
q6	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	8	0		8	
q9	1	1	1	1	1	1	0	0	0	1	0	0	1	0	0	8	0		8	
q1	1	1	0	1	0	0	0	1	1	0	1	1	0	1	0	8		0		
q5	1	1	1	0	0	1	0	0	1	0	0	1	0	0	0	6		0		
q10	1	1	0	0	1	0	0	0	1	1	0	0	0	0	0	5		5		
q8	1	0	1	0	0	0	1	0	0	0	0	1	0	0	0	4		0		
Score	**10**	**8**	**8**	**7**	**6**	**6**	**5**	**5**	**5**	**5**	**4**	**4**	**3**	**2**	**2**	**80**	**9**	**5**	**57**	**48**

Average number of correct responses per question = 8

Let us look at the results for subject 6.

$$a = \sum_{j=1}^{y_{i.}} (1 - y_{ij})(y_{.j})$$
$$= 9$$

$$b = \sum_{j=y_i+1}^{J} (y_{ij})(y_{.j})$$
$$= 5$$

$$c = \sum_{j=1}^{y_{i.}} y_{.j}$$
$$= 57$$

$$d = (y_{i.})(u')$$
$$= 6 * 8$$
$$= 48$$

$$CI = \frac{\sum_{j=1}^{y_{i.}} (1 - y_{ij})(y_{.j}) - \sum_{j=y_i+1}^{J} (y_{ij})(y_{.j})}{\sum_{j=1}^{y_{i.}} y_{.j} - (y_{i.})(u')}$$
$$= \frac{9 - 5}{57 - 48}$$
$$= 0.44$$

Modified Caution Index

The Modified Caution Index (or MCI) provides an index for detecting students, or items which have produced unusual response patterns on multiple choice exams. Student MCI is sensitive to students who score low but get an inordinate number of difficult items correct and students who score high but get an inordinate number of easy items wrong. Similarly, Item MCI can point out items that are easy but high scoring students miss, and difficult items that low scoring students get.

An MCI score for either an Item or a student that exceeds .30 indicates that this item or student has a pattern of responses that is unlike students or items with a similar number of correct responses. For students, an MCI exceeding .30 can mean that they guessed, were confused, careless, under stress, or perhaps they cheated on part of the test.

$$MCI = \frac{\sum_{j=1}^{y_{i.}} (1 - y_{ij})(y_{.j}) - \sum_{j=y_i+1}^{J} (y_{ij})(y_{.j})}{\sum_{j=1}^{y_{i.}} y_{.j} - \sum_{j=J+1-y_{i.}}^{J} y_{.j}}$$

$$\begin{aligned} &\textit{where } i = 1, 2, \cdots I \textit{ indexes the students} \\ &\quad j = 1, 2, \cdots J \textit{ indexes the questions} \\ &\quad y_{ij} = \textit{the response of student } i \textit{ on question } j \\ &\quad y_{i.} = \textit{the number of correct responses of student } i \textit{ on all test questions} \\ &\quad y_{.j} = \textit{the number of correct responses of all students on question } j \end{aligned}$$

While there are a number of variations in the formula, the one indicated above is probably the easiest to apply using a simple program such as Microsoft Excel.

	s5	s7	s9	s4	s10	s2	s13	s14	s6	s1	s15	s11	s3	s8	s12	Total	Modified Caution Index a	b	c	d
q3	1	1	1	1	1	1	1	1	0	1	1	0	0	0	1	11	0		11	
q2	1	0	1	1	1	1	1	1	1	1	0	1	1	0	0	11	0		11	
q7	1	1	1	1	1	0	0	1	0	1	1	0	1	1	0	10	0		10	
q4	1	1	1	1	0	1	1	1	1	0	1	0	0	0	0	9	9		9	
q6	1	1	1	1	1	1	1	0	0	0	0	0	0	0	1	8	0		8	8
q9	1	1	1	1	1	1	0	0	0	1	0	0	1	0	0	8	0		8	8
q1	1	1	0	1	0	0	0	1	1	0	1	1	0	1	0	8		0		8
q5	1	1	1	0	0	1	0	0	1	0	0	1	0	0	0	6		0		6
q10	1	1	0	0	1	0	0	0	1	1	0	0	0	0	0	5		5		5
q8	1	0	1	0	0	0	1	0	0	0	0	1	0	0	0	4		0		4
Score	**10**	**8**	**8**	**7**	**6**	**6**	**5**	**5**	**5**	**5**	**4**	**4**	**3**	**2**	**2**	**80**	**9**	**5**	**57**	**39**

Let us look at the results for subject 6.

$$a = \sum_{j=1}^{y_{i.}} (1 - y_{ij})(y_{.j})$$
$$= 9$$

$$b = \sum_{j=y_i+1}^{J} (y_{ij})(y_{.j})$$
$$= 5$$

$$c = \sum_{j=1}^{y_{i.}} y_{.j}$$
$$= 57$$

$$d = \sum_{j=J+1-y_{i.}}^{J} y_{.j}$$
$$= 39$$

$$MCI = \frac{\sum_{j=1}^{y_{i.}} (1 - y_{ij})(y_{.j}) - \sum_{j=y_i+1}^{J} (y_{ij})(y_{.j})}{\sum_{j=1}^{y_{i.}} y_{.j} - \sum_{j=J+1-y_{i.}}^{J_{i.}} y_{.j}}$$
$$= \frac{9 - 5}{57 - 39}$$
$$= 0.22$$

References

Chen, D.-J., Lai, A.-F., Liu, I-C., & Chen, D.T.K. The design and implementation of a diagnostic test system based on the enhanced S-P model. *International Conference on Computers in Education 2004*.
http://plum.yuntech.edu.tw/icce2004/Theme4/Ch108_DengJyiChen.pdf#search='caution%20index%20chen'

Guilford, J.P. (1954). *Psychometric methods*. New York: McGraw-Hill Book Company.

Klugh, H.E. (1974). *Statistics: The essentials for research (2nd ed.)*. New York: John Wiley & Sons, Inc.

Lemke, E., & Wiersma, W. (1976). *Principles of psychological measurement*. Chicago: Rand McNally College Publishing Company.

Oltman, P.K. (1985). Background characteristics of examinees showing unusual test behaviour on the Graduate Record Examinations. *GRE Board Professional Report GREB No. 82-8P ETS Research Report 85-39*. Princeton, N.J.: Educational Testing Service.
http://ftp.ets.org/pub/gre/gre-82-8p.pdf

Sax, G. (1989). *Principles of educational and psychological measurement and evaluation (3rd ed.)*. Belmont, California: Wadsworth Publishing Company.

Wu, H.Y. (1998). Software based on S-P chart analysis and its application. *Proceedings of the National Science Council ROC(D)*, *8*, 108-120.
http://nr.stic.gov.tw/ejournal/ProceedingD/v8n3/108-120.pdf