

## CES-D IRT Item Parameter Estimates, Scores and Standard Errors with Custom Weighted Z-Scores and Percentile Ranks

### Parameter Estimates and Created Variables

CES-D ITEM PARAMETER ESTIMATES. These four variables, shown below, represent item response theory (IRT) item parameter estimates for each Center for Epidemiologic Studies-Depression Scale (CES-D) item, including measures of discrimination ( $\hat{a}$ ) and severity ( $\hat{b}_1, \hat{b}_2, \hat{b}_3$ ), which were calibrated using a graded response model in Multilog.

CES-D IRT SCORE and STANDARD ERROR. These two variables represent the CES-D IRT scores ( $\hat{\theta}$ ) and their standard errors of measurement (SE), which are presented in standardized metric and were calculated using the CES-D IRT parameter estimates.

CES-D CUSTOM WEIGHTED Z-SCORE and PERCENTILE RANK. These two variables were calculated using the cross-sectional custom weights for each survey wave within each cohort, which correct the raw data for the effects of over-sampling, differential base year participation and differential wave and item non-response.

### Data collected

The CES-D IRT item parameter estimates, scores and standard errors were calculated using data from following cohorts and surveys:

Mature Women – 1995, 1997, 1999, 2001, 2003

Young Women – 1995, 1997, 1999, 2001, 2003

NLSY79 – 1992, 1994, Age 40

Young Adult – 1994, 1996, 1998, 2000, 2002, 2004, 2006, 2008

### Mature Women and Young Women

Seven of the CES-D scale items, which appear in Table 1, were presented to respondents in the 1995 (R33678.00 – R33684.00), 1997 (R41361.00 – R41367.00), 1999 (R50935.00 – R50941.00) and 2001 (R61869.00 – R61875.00) surveys. The 2003 surveys included the full CES-D scale (R72915.00 – R72934.00), as presented in Table 2.

**Table 1. Subset of CES-D Scale Items Presented in the 1995, 1997, 1999, and 2001 Mature Women and Young Women Surveys**

---

I felt that I could not shake off the blues even with help from my family or friends.

I had trouble keeping my mind on what I was doing.

I felt that everything I did was an effort.

My sleep was restless.

I felt lonely.

I felt sad.

I could not get “going.”

---

**Table 2. Full CES-D Scale Presented in the 2003 Mature Women and Young Women Surveys**

- 
1. I was bothered by things that usually don't bother me.
  2. I did not feel like eating; my appetite was poor.
  3. I felt that I could not shake off the blues even with help from my family or friends.
  4. I felt I was just as good as other people.
  5. I had trouble keeping my mind on what I was doing.
  6. I felt depressed.
  7. I felt that everything I did was an effort.
  8. I felt hopeful about the future.
  9. I thought my life had been a failure.
  10. I felt fearful.
  11. My sleep was restless.
  12. I was happy.
  13. I talked less than usual.
-

- 
14. I felt lonely.
  15. People were unfriendly.
  16. I enjoyed life.
  17. I had crying spells.
  18. I felt sad.
  19. I felt that people dislike me.
  20. I could not get “going.”
- 

### **NLSY79 and Young Adult**

For the NLSY79 1992 survey, respondents were presented with the full CES-D scale (R38949.00 – R38968.00), as it appears in Table 2. Respondents from the NLSY79 1994 (R49783.00 – R49789) survey and all of the Young Adult (1994: Y03360.00 - Y3366.00; 1996: Y06364.00 – Y06370.00; 1998: Y09309.00 – Y09315.00; 2000: Y11615.00 – Y11621.00; 2002: Y13966.00 – Y13972.00; 2004: Y16481.00 – Y16487.00; 2006: Y19199.00 – Y19205.00; 2008: Y22334.00 – Y22340.00) surveys were presented with seven of the CES-D scale items, which appear in Table 3. When NLSY79 respondents were surveyed at age 40, they were presented with nine of the CES-D items (H00003.00 – H00011.00), which are displayed in Table 4.

### **Table 3. Subset of CES-D Scale Items Presented in the NLSY79 1994 Survey and All of the Young Adult Surveys**

---

- I did not feel like eating; my appetite was poor.
- I had trouble keeping my mind on what I was doing.
- I felt depressed.
- I felt that everything I did was an effort.
- My sleep was restless.
- I felt sad.
- I could not get “going”.
-

**Table 4. Subset of CES-D Scale Items Presented to NLSY79 Respondents at Age 40**

---

I did not feel like eating; my appetite was poor.

I felt that I could not shake off the blues even with help from my family or friends.

I had trouble keeping my mind on what I was doing.

I felt depressed.

I felt that everything I did was an effort.

My sleep was restless.

I felt lonely.

I felt sad.

I could not get “going.”

---

### **Depressive Symptoms**

The Center for Epidemiological Studies – Depression Scale (CES-D) measures an individual’s current level of depressive symptoms and is intended for use in general population surveys (Radloff, 1977). The 20-item scale (see Table 2) comprises 16 negative symptoms and four positive symptoms, representing a single continuum from depression to happiness (Wood, Taylor, & Joseph, 2010). Symptom severity is measured by asking the frequency of occurrence for each item during the preceding week. Response options range from 0 (rarely or none of the time/1 day) to 3 (most or all of the time/5-7 days).

### **Why These Variables May be Helpful**

Depression is a mental health condition with both a substantial prevalence and impact. Some have conjectured that it is the single largest cause of lost time on the job worldwide. The original method of scoring these questions was to use a 0-3 scale for the responses (reverse coding where appropriate) and then sum. This is the approach of “Classical Test Theory” (CTT), which used to dominate the psychometrics of scaling. There are several problems with CTT that are relevant for users of the depression data. First, if the questions asked from year-to-year or from cohort to cohort differ, the scores are not comparable. Second, CTT imposes the very strong restriction that a “1” means the same thing for all questions, and third, that going from a “1” to a “2” likewise is equally informative about depression for all questions. Item Response Theory does not impose these restrictions.

Our scoring of the CES-D using IRT generates a  $\theta$  value that is comparable across rounds, across cohorts and summarizes the information contained in all the responses to the CES-D questions for a particular round. We also provide an estimated standard error for  $\theta$  to provide the user with guidance on the precision of the estimated value of  $\theta$ .

Because  $\theta$  measures depressivity with error (this being unavoidable given the data resources at hand), we suggest that when  $\theta$  is being used as a right-hand side variable in a regression, users consider taking advantage of the repeated measures on  $\theta$  available in the various rounds, by using these other measured values as “instruments” for the observation of  $\theta$  being used as a regressor. The method of instrumental variables (IV) is due to Geary and Reiersol, dating back to 1945, and is discussed in many advanced texts on statistical methods for the social sciences. By using IV, the well-known problem of attenuation bias due to measurement error can be overcome. The substantial stability of CES-D scores over the life course makes the use of IV especially efficacious. Whether IV is appropriate depends on the model specification and the assumed error structure.

Moreover, because the subset of CES-D questions asked changes and IRT scoring accounts for such changes, using IRT simplifies comparisons over the life course and across cohorts, which are staples of longitudinal analysis.

### **Item Response Theory**

Within the item response theory (IRT) framework, the latent construct ( $\theta$ ) being measured (i.e., depressive symptoms) is assumed to follow a standard normal distribution. Given the meaningfully ordered (with respect to  $\theta$ ), multiple response options appearing with the CES-D items, the IRT analyses were conducted using a graded response model (GRM; Samejima, 1969), in Multilog (Thissen, Chen, & Bock, 2003). For a GRM, the  $a$  parameter or slope estimate ( $\hat{a}$ ) represents item discrimination, which indicates how well an item differentiates between individuals with varying levels of  $\theta$ . Items with low slopes (i.e., close to zero) are problematic because they do not distinguish between individuals with varying levels of depressive symptoms. Therefore, items with higher  $\hat{a}$  values are generally more desirable than those with lower  $\hat{a}$  values. Each  $b$  parameter or severity estimate ( $\hat{b}_1, \hat{b}_2, \hat{b}_3$ ) identifies the point along  $\theta$  where one response category becomes more likely to be endorsed than any other option, given the respondent's level of depressive symptoms. Items with equally distributed  $\hat{b}$  values, across the range of  $\theta$ , identify clear distinctions between individuals with varying levels of depressive symptoms, according to the response options that they choose. Items with  $\hat{b}$  values that are extreme (greater than 4.5 standard deviations in either direction) or too close together are less desirable because knowledge of the selected response option does not provide clear information regarding an individual's level of  $\theta$ . The unidimensionality (i.e., the items measure a single latent construct) of the CES-D scale (Wood et al., 2010), ensures that any subset of CES-D items will also provide a unidimensional measure of depressive symptoms. The presentation of the CES-D IRT scores in standardized metric, with mean of zero and standard deviation of one, allows for easy and meaningful comparison of scores and standard errors with standardized scores from other scales.

The IRT item calibration of the CES-D scale items was conducted on a combined data set of responses from the four cohorts, across multiple years (see above;  $n = 20,225$ ; Table 5). Because previous research has identified the presence of differential item functioning (DIF) for some CES-D items across the life course (Cooksey, Eberwein, Gardecki, Ing, & Olsen, 2010), additional analyses were conducted to ensure the appropriateness of using a single set of item parameter estimates for calculating the CES-D IRT scores for respondents of different ages.

Separate item calibrations were conducted for eight age groups ranging from teens to those in their eighties and DIF analyses were conducted on the eight sets of parameter estimates. While differences between parameter estimates across age groups displayed statistical significance through chi-square difference tests in IRTLRDIF (Thissen, 2001), these differences are not substantively significant. The practical significance of the group differences was evaluated by comparing two sets of IRT scores for each age group; one based on their own item parameter estimates and one calculated from the combined data parameter estimates. The Spearman rank-order correlation coefficients for the two sets of IRT scores indicate a high degree of similarity in the percentile rankings of the scores ( $r \geq .97$ ); a person in the top 5% for depressive symptoms using an IRT score will most likely be in the top 5%, regardless of which set of parameter estimates are used to calculate the score. Given these results, the CES-D IRT item parameter estimates based on the combined data set were used to calculate all of the CES-D IRT scores and standard errors.

### Custom Weighted Scores and Percentile Ranks

Every NLS data release contains a set of cross-sectional weights. Using these weights provides a simple method for users to correct the raw data for the effects of over-sampling, clustering and differential base year participation. The custom weighted z-scores and percentile ranks were calculated using the CES-D IRT scores and the cross-sectional weights for each survey wave, within each NLS cohort.

**Table 5. IRT Item Parameter Estimates and Standard Errors for the CES-D Scale**

Item	Parameter	Estimate	Standard Error
1. I was bothered by things that usually don't bother me.	<i>a</i>	1.56	.03
	<i>b<sub>1</sub></i>	0.47	.02
	<i>b<sub>2</sub></i>	1.57	.03
	<i>b<sub>3</sub></i>	2.49	.05
2. I did not feel like eating; my appetite was poor.	<i>a</i>	1.37	.03
	<i>b<sub>1</sub></i>	1.16	.02
	<i>b<sub>2</sub></i>	2.04	.04
	<i>b<sub>3</sub></i>	2.93	.06
3. I felt that I could not shake off the blues even with help from my family or friends.	<i>a</i>	2.76	.06
	<i>b<sub>1</sub></i>	0.78	.01

	<i>b</i> <sub>2</sub>	1.49	.02
	<i>b</i> <sub>3</sub>	2.04	.03
4. I felt I was just as good as other people.	<i>a</i>	0.16	.03
	<i>b</i> <sub>1</sub>	-4.70	.65
	<i>b</i> <sub>2</sub>	-3.07	.39
	<i>b</i> <sub>3</sub>	-0.98	.19
5. I had trouble keeping my mind on what I was doing.	<i>a</i>	1.52	.03
	<i>b</i> <sub>1</sub>	0.39	.02
	<i>b</i> <sub>2</sub>	1.45	.02
	<i>b</i> <sub>3</sub>	2.44	.04
6. I felt depressed.	<i>a</i>	3.21	.05
	<i>b</i> <sub>1</sub>	0.66	.01
	<i>b</i> <sub>2</sub>	1.38	.01
	<i>b</i> <sub>3</sub>	1.90	.02
7. I felt that everything I did was an effort.	<i>a</i>	0.95	.02
	<i>b</i> <sub>1</sub>	-0.02	.02
	<i>b</i> <sub>2</sub>	1.11	.03
	<i>b</i> <sub>3</sub>	1.84	.05
8. I felt hopeful about the future.	<i>a</i>	0.10	.03
	<i>b</i> <sub>1</sub>	-8.51	1.89
	<i>b</i> <sub>2</sub>	-3.40	.90
	<i>b</i> <sub>3</sub>	2.88	.83
9. I thought my life had been a failure.	<i>a</i>	1.97	.05
	<i>b</i> <sub>1</sub>	1.48	.03
	<i>b</i> <sub>2</sub>	2.16	.04

	<i>b</i> <sub>3</sub>	2.69	.06
10. I felt fearful.	<i>a</i>	1.88	.05
	<i>b</i> <sub>1</sub>	1.17	.02
	<i>b</i> <sub>2</sub>	2.08	.04
	<i>b</i> <sub>3</sub>	2.70	.06
11. My sleep was restless.	<i>a</i>	1.22	.02
	<i>b</i> <sub>1</sub>	0.23	.02
	<i>b</i> <sub>2</sub>	1.33	.03
	<i>b</i> <sub>3</sub>	2.18	.04
12. I was happy.	<i>a</i>	0.09	.04
	<i>b</i> <sub>1</sub>	-9.27	2.10
	<i>b</i> <sub>2</sub>	-4.05	.87
	<i>b</i> <sub>3</sub>	3.03	.74
13. I talked less than usual.	<i>a</i>	1.29	.03
	<i>b</i> <sub>1</sub>	0.91	.03
	<i>b</i> <sub>2</sub>	1.86	.04
	<i>b</i> <sub>3</sub>	2.74	.07
14. I felt lonely.	<i>a</i>	2.06	.04
	<i>b</i> <sub>1</sub>	0.81	.02
	<i>b</i> <sub>2</sub>	1.59	.03
	<i>b</i> <sub>3</sub>	2.18	.04
15. People were unfriendly.	<i>a</i>	0.99	.03
	<i>b</i> <sub>1</sub>	1.62	.05
	<i>b</i> <sub>2</sub>	2.88	.09
	<i>b</i> <sub>3</sub>	3.82	.13



16. I enjoyed life.	<i>a</i>	0.10	.02
	<i>b<sub>1</sub></i>	-7.44	1.52
	<i>b<sub>2</sub></i>	-4.14	1.08
	<i>b<sub>3</sub></i>	0.09	.23
17. I had crying spells.	<i>a</i>	2.20	.06
	<i>b<sub>1</sub></i>	1.35	.02
	<i>b<sub>2</sub></i>	2.02	.03
	<i>b<sub>3</sub></i>	2.66	.05
18. I felt sad.	<i>a</i>	2.83	.05
	<i>b<sub>1</sub></i>	0.59	.01
	<i>b<sub>2</sub></i>	1.49	.02
	<i>b<sub>3</sub></i>	2.15	.03
19. I felt that people dislike me.	<i>a</i>	1.56	.05
	<i>b<sub>1</sub></i>	1.64	.03
	<i>b<sub>2</sub></i>	2.58	.06
	<i>b<sub>3</sub></i>	3.26	.09
20. I could not get “going”.	<i>a</i>	1.52	.03
	<i>b<sub>1</sub></i>	0.57	.02
	<i>b<sub>2</sub></i>	1.76	.03
	<i>b<sub>3</sub></i>	2.64	.05

---

## References

Cooksey, E., Eberwein, C., Gardecki, R., Ing, P., & Olsen, R. J. (2010, September). Depressivity over the life course and across generations. Paper presented at the Society for Longitudinal and Life Course Studies Conference, Cambridge, UK.

Geary, R. (1949). Determination of Linear Relationships between Systematic Parts of Variables with Errors of Observation the Variances of which are Unknown. *Econometrica*, 17.

Radloff, L. S. (1977). The CES-D scale: A self-report depression scale for research in the general population. *Applied Psychological Measurement*, 1(3), 385-401.

Reiersol, O. (1945) Confluence Analysis by Means of Instrumental Sets of Variables. *Arkiv for Matematik, Astronomi och Fysik*, 32.

Samejima, F. (1969). Estimation of latent ability using a response pattern of graded scores. Psychometrika Monograph, No. 17.

Thissen, D. (2001). IRTLRDIF (Version 2.0b) [Computer software]. Chapel Hill, North Carolina: L. L. Thurstone Psychometric Laboratory.

Thissen, D., Chen, W., & Bock, R. D. (2003). Multilog (Version 7.03) [Computer software]. Lincolnwood, Illinois: Scientific Software International.

Wood, A. M., Taylor, P. J., & Joseph, S. (2010). Does the CES-D measure a continuum from depression to happiness? Comparing substantive and artifactual models. *Psychiatry Research*, 177, 120-123.