

Estimating Reliability from CAT

FAQ: How is reliability estimated in computer adaptive testing?

Answer: The answer is provided at: http://official-asvab.com/reliability_res.htm

As noted by Sireci, Thissen and Wainer (1991), marginal reliability is a function of the between person variance in theta and the average error variance:

$$\rho^2 = \frac{\sigma_{\theta}^2 - \overline{\sigma_E}^2}{\sigma_{\theta}^2}.$$

Assuming theta is estimated with mean of 0 and SD of 1, marginal reliability can be estimated as:

$$\rho^2 = \frac{1 - \overline{\sigma_E}^2}{1}.$$

If sample sizes are large, the error variance can be estimated by averaging the variance (square of the estimated posterior standard deviations, PSD) across individuals (Bock & Mislevy, 1982):

$$PSD_i = \left[\frac{\sum_{k=1}^q (\theta_k - \bar{\theta}_i)^2 L_i(\theta_k | u) \cdot f(\theta_k)}{\sum_{k=1}^q L_i(\theta_k | u) \cdot f(\theta_k)} \right]^{1/2}$$

Where θ_k is one of q equally-spaced quadrature points, $L(\theta_k | u)$ is the likelihood of θ_k given the response pattern (u_1, u_2, \dots, u_n) over n items. $f(\theta_k)$ represents the proportion of the population estimate θ_k and $\bar{\theta}_i$ is the expected a posteriori (EAP) estimate of the ability of person i , approximated by:

$$\bar{\theta}_i = \left[\frac{\sum_{k=1}^q \theta_k \cdot L_i(\theta_k | u) \cdot f(\theta_k)}{\sum_{k=1}^q L_i(\theta_k | u) \cdot f(\theta_k)} \right].$$

For an application, see Haley et al. (2010). For an extension to the ML estimator see Nicewander and Thomasson (1999).

References

Bock, R.D., & Mislevy, R.J. (1982). Adaptive EAP estimation of ability in a microcomputer environment. *Applied Psychological Measurement*, 6, 431-444.

Haley, S. M., Chafetz, R. S., Tian, F., Montpetit, K., Watson, K., Gorton, G., & Mulcahey, M. J. (2010). Validity and reliability of physical functioning computer-adaptive tests for children with cerebral palsy. *J Pediatr Orthop*, 30, 71-75.

Nicewander, W. A., & Thomasson, G. L. (1999). Some reliability estimates for computerized adaptive tests. *Applied Psychological Measurement*, 23, 239-247.

Sireci, S.G., Thissen, D., & Wainer, H. (1991). On the reliability of testlet-based tests. *Journal of Educational Measurement*, 28, 237-247.