

Calibrating Personality Self-Report Scores to Acquaintance Ratings

John A. Johnson

Pennsylvania State University, DuBois



Abstract

By convention in individual personality assessment, scores on self-report questionnaires within ± 5 standard deviation of the mean score for that trait are considered "average," whereas scores outside that range are reported as "high" or "low" levels of the trait. To date, no one has examined how well this convention corresponds to perceptions of low, average, or high trait levels by acquaintances. The present study compares the accuracy of the conventional ± 5 SD cutoffs for low, average, and high trait levels with cutoffs derived from Optimal Data Analysis (ODA; Yarnold & Seltysik, 2005) for 160 participants who completed the IPIP-NEO (Goldberg, 1999) and were rated by acquaintances on the 30 facets and 5 domains of that instrument. In 32 of 35 comparisons, the ± 5 convention was found to be less accurate than the ODA method for deriving cutoffs.

The Cutoff Problem in Computer Narrative Reports

A perennial practical problem when providing feedback on computer scored personality inventories has been the how to choose definitive descriptions of personality from scores that are merely probabilistic signs of personality. Hofstee (1994) has argued that, given the imperfect reliabilities of personality measures, we should not pretend that our feedback to inventory respondents can be more precise than to tell them that they possess "low," "average," or "high" levels of a personality trait. This is essentially what the one-page "Your NEO Summary" feedback sheet provides (Costa & McCrae, 1992), although Costa and McCrae's NEO PI-R Profile Form uses two additional categories, "very low" and "very high."

On the NEO PI-R Profile Form, scores within ± 5 standard deviation of the mean score of a standardization sample are described as "average." Describing T-scores between 45 and 55 as "average" is a common practice for interpreting multi-scale inventories (e.g., Hogan & Hogan's 1992 HPI). For normally distributed traits, these cutoffs correspond to the 30th and 70th percentiles, which means that about 40% of the population would be described as "average." For normally distributed traits, about 15% of the population will score between .5 and 1 SD above the mean and approximately 15% will score higher than 1 SD above the mean. These scores might be described as "high" and "very high," and the lowest 30% of scores "low" and "very low."

What About Acquaintance Perceptions?

Although defining self-report inventory scores as "low," "average," or "high" according to whether they fall within ± 5 SD of the mean makes good statistical sense for labeling a score with respect to others' self-report scores, this does not mean that someone receiving an "average" score on a self-report measure will be perceived as average for that trait by acquaintances who know the person well. It is an empirical question how accurately the ± 5 SD cutoffs for a particular self-report measure will correspond to judgments of personality by acquaintances. The current study tested the accuracy of the ± 5 SD convention against a method called Optimal Data Analysis (ODA; Yarnold & Soltysik, 2005) for deriving cutoffs for maximizing valid classification.

METHOD

196 undergraduate students agreed to complete Goldberg's (1999) 300-item International Personality Item Pool (IPIP) representation of Costa and McCrae's (1992) NEO PI-R (hereafter, IPIP-NEO). Participants provided the names of three knowledgeable acquaintances, who were sent to a Web site containing descriptions of the five domains and 30 facets measured by the IPIP-NEO. Acquaintances were asked to rate participants on 35 scales with the following percentile anchor points: 1, 10, 20, 30, 40, 50, 60, 70, 80, 90, 99. From the 196 persons recruited, 160 (59 males, 101 females) actually completed the IPIP-NEO and had at least one acquaintance complete the personality ratings (116 had three raters, 39 had two raters, and 5 had one rater). Acquaintance ratings were averaged for participants with two or more raters to increase reliability. While future analyses may examine five categories of ratings, the current student classified averaged ratings between 30-70 percentile as "average" and ratings outside that range as "low" or "high." Below is a sample portion of the Web rating form.

Extraversion
Extraversion is marked by pronounced enjoyment with the external world. Extraverts enjoy being with people, are full of energy, and often experience positive emotions. They tend to be optimistic, active, outgoing, individuals who are likely to say "Yes!" or "Let's go!" in response to the invitation. Extrapens like to be talkative, assertive, and are often leaders in their groups.

Introversion
Introverts lack the exuberance, energy, and active levels of extraverts. They tend to gain less stimulation and enjoyment from the social world. Their lack of social involvement doesn't get interpreted as shyness or depression, the former being similar to a person who is not interested in the world. They are often reserved and reserved in their interactions. In socially, extraverts who are high on the extraversion dimension will seek out others and will be perceived as being approachable.

Friendliness
Friendly people greet the other people and greet themselves in a positive and warm way. They tend to be quick and to react to them to form close, warm relationships. Low levels of friendliness do not mean that individuals are hostile, but people with low friendliness do not reach out to others and are perceived as distant and reserved.

Openness
Openness is the degree to which people are curious, imaginative, and open-minded. Low openness individuals tend to be more traditional, and therefore active and, large crowds. They do not mind to do things with people, but they are not the person and they are usually in a much greater than the individuals who are high on openness.

Assertiveness
Assertiveness is the degree to which people are confident and assertive. High assertiveness individuals tend to be more confident and assertive. Low assertiveness individuals tend to be more reserved and less assertive.

Percent of correct classifications for the 35 ratings were compared for the two methods of deriving cutoffs (± 5 SD convention versus ODA). IPIP-NEO means and SDs from Johnson's (2005) sample of over 20,000 persons were used to describe whether scores were within ± 5 SD of the mean. ODA uses an iterative process to determine cutoffs that maximize correct classifications, employing a leave-one-out procedure (LOO) to confirm the stability of the predictive model. LOO is conceptually equivalent to cross-validation of the derived cutoff scores. For seven traits, insufficient cases in a category precluded the LOO procedure. Although ODA produces many statistics, including overall percentage accuracy in classification (PAC), sensitivity (PAC within each category), predictive value (percent of persons correctly classified), effect strength for sensitivity and predictive value (classification improvement over chance), and others (including model efficiency over base rates), current analyses were limited to overall PAC.

RESULTS

For 32 of 35 traits, the ODA method of selecting cutoffs produced a greater overall percentage accuracy of classification (PAC). The mean PAC for the ODA method was 62.2% (SD=12.9), whereas the mean PAC for the .5 SD method was 41% (SD=3.8); $t(34)=21.2, p < .0001$. The table below shows the accuracy of classification from the two methods for each of the 35 traits with the low and high cutoffs from the ODA method. Low cutoffs ranged from .17 SD above the mean to 2.4 SD below the mean, with an average of 1.21 SD below the mean. High cutoffs ranged from .03 SD below the mean to 2.31 SD above the mean, with an average of 1.27 SD above the mean.

Trait	Overall PAC	ODA Cutoffs		
		Low	High	
Friendliness	40	62	-1.35	1.01
Gregariousness	43	77	-1.98	1.78
Assertiveness	48	72	-1.59	0.91
Activity Level	41	64	-1.71	1.13
Excitement-Seeking	51	68	-1.88	1.38
Cheerfulness	39	67	-1.57	1.61
EXTRAVERSION	44	70	-2.40	0.88
Trust	41	52	-2.03	0.64
Morality	36	44	-0.14	1.55
Altruism	37	41	-0.20	1.41
Cooperation	35	68	-1.03	1.48
Modesty	40	78	-1.89	2.21
Sympathy	39	56	-1.30	0.69
AGREEABLENESS	41	33	0.17	0.78
Self-Efficacy	39	33	-0.55	-0.03
Orderliness	46	71	-1.79	0.94
Dutifulness	39	58	-1.05	0.13
Achievement-Striving	44	32	0.11	0.56
Self-Discipline	37	78	-0.59	0.88
Cautiousness	35	76	-1.84	2.03
CONSCIENTIOUSNESS	43	53	-0.57	1.32
Anxiety	47	66	-0.49	1.65
Anger	45	58	-1.41	1.63
Depression	41	67	0.01	2.11
Self-Consciousness	41	64	-0.11	1.84
Immoderation	41	68	-1.06	2.24
Vulnerability	36	66	-0.86	2.31
NEUROTICISM	43	71	-0.82	1.91
Imagination	35	77	-2.27	1.38
Artistic Interests	43	68	-2.37	0.89
Emotionality	37	63	-1.66	0.95
Adventurousness	41	65	-1.51	0.62
Intellect	40	57	-1.65	0.77
Liberalism	42	79	-1.84	1.88
OPENNESS	44	55	-1.23	0.98

DISCUSSION

Cutoff scores from Optimal Data Analysis (ODA; Yarnold & Soltysik, 2005) reliably classify acquaintance perceptions of low, average and high levels of traits more accurately than the convention of classifying scores within ± 5 SD of the mean as "average" and scores outside that range as "low" or "high." Optimal cutoffs varied considerably across scales, averaging more than 1 SD from the mean. Further analyses are needed to clarify the relative sensitivity, predictive value, effect strengths, and model efficiency for the two methods.

References

- Costa, P. T., Jr., & McCrae, R. R. (1992). *Revised NEO Personality Inventory (NEO PI-R™) and NEO Five-Factor Inventory (NEO-FFI) professional manual*. Odessa, FL: Psychological Assessment Resources.
- Goldberg, L. R. (1999). A broad-bandwidth, public-domain, personality inventory measuring the lower-level facets of several five-factor models. In I. Mervielde, I. Deary, F. De Fruyt, & F. Ostendorf (Eds.), *Personality Psychology in Europe, Vol. 7* (pp. 7-28). Tilburg, The Netherlands: Tilburg University Press.
- Hofstee, W. K. B. (1994). Who should own the definition of personality? *European Journal of Personality, 8*, 149-162.
- Hogan, R., & Hogan, J. (1992) *Hogan Personality Inventory manual*. Tulsa, OK: Hogan Assessment Systems.
- Johnson, J. A. (2005). Ascertain the validity of web-based personality inventories. *Journal of Research in Personality, 39*, 103-129.
- Yarnold, P. R., & Soltysik, R. C. (2005). *Optimal data analysis: A guidebook with software for windows*. Washington, DC: American Psychological Association.

