DOCUMENT RESUME

ED 468 463 TM 034 392

AUTHOR	De Ayala, R. J.
TITLE	The Effect of Missing Data on Estimating a Respondent's Location Using Ratings Data.
PUB DATE	2002-04-00
NOTE	21p.; Paper presented at the Annual Meeting of the National Council on Measurement in Education (New Orleans, LA, April 2-4, 2002).
PUB TYPE	Reports - Research (143) -- Speeches/Meeting Papers (150)
EDRS PRICE	EDRS Price MF01/PC01 Plus Postage.
DESCRIPTORS	*Measurement Techniques; Monte Carlo Methods; *Rating Scales; *Responses; Simulation
IDENTIFIERS	*Missing Data

ABSTRACT

In social science research there are a number of instruments that use a rating scale such as a Likert response scale. For a number of reasons, a respondent's response vector may not contain responses to each item. This study investigated the effect on a respondent's location estimate when a respondent is presented an item, has ample time to answer the item, but decides not to respond to the item. For these situations, different strategies have been developed for handling missing data. In this study, four different approaches for handling missing data were investigated for their capability to mitigate the effect of omitted responses on person location estimation. These methods included ignoring the omitted response, selecting the "midpoint" response category, Hot-decking, and a likelihood-based approach. A Monte Carlo study was performed and the effect of different levels of omission on the simulee's location estimates was determined. Results show that the Hot-decking procedure performed the best of the methods examined. Implications for practitioners were discussed. (Contains 6 figures and 10 references.) (Author/SLD)

TM

The effect of missing data on estimating a respondent's location using ratings data

R.J. De Ayala

University of Nebraska-Lincoln

R.J. De Ayala

Department of Educational Psychology

Teachers College Hall 114

University of Nebraska-Lincoln

Lincoln, NE 68588-0345

(402) 472-2736

email: rdeayala2@unl.edu

Paper presented at the annual meeting of the National Council on Measurement in Education, New Orleans, LA, April 2002.

Abstract

In social science research there are a number of instruments that utilize a rating scale such as a Likert response scale. For a number of reasons a respondent's response vector may not contain responses to each item. This study investigated the effect on a respondent's location estimate when a respondent is presented an item, has ample time to answer the item, but decides to not respond to the item. For these situations different strategies have been developed for handling missing data. In this study, four different approaches for handling missing data were investigated for their capability to mitigate against the effect of omitted responses on person location estimation. These methods included Ignoring the omitted response, selecting the "midpoint" response category, Hot-decking, and a Likelihood-based approach. A Monte Carlo study was performed and the effect of different levels of omissions on the simulees' location estimates was determined. Results showed that the Hot-decking procedure performed the best of methods examined. Implications for practitioners were discussed.

The effect of missing data on estimating a respondent's location using ratings data

In social science research there are a number of instruments that utilize a rating scale such as a Likert response scale. For a number of reasons a respondent's response vector may not contain responses to each item. Using Little and Rubin's (1987) terminology nonresponses that arise from an a priori decision to not administer certain (e.g., in the adaptive administration of an instrument or when respondents are directed to answer only relevant items (see Schulz & Sun, 2001)) represent conditions in which the missingness process may be ignored for purposes of estimating the person's location on the latent continuum of interest (Mislevy & Wu, 1988; Mislevy & Wu, 1996). In contrast, nonresponses for "not-reached" item(s) occur because an respondent has insufficient time to even consider responding to the item(s). Assuming the subject responds to the items in serial order these not-reached items can be identified as collectively occurring at the end of an instrument. Another source of missing data occurs because respondents have the capability of choosing not to respond to certain items on an instrument. These (intentionally) omitted responses represent nonignorable missing data (Lord, 1980; Mislevy & Wu, 1988; Mislevy and Wu, 1996). This latter condition is referred to as missing not at random (MNAR). This study investigated the effect on a respondent's location estimate when a respondent is presented an item, has ample time to answer the item, but decides to not respond to the item (i.e., the MNAR case).

Different strategies have been developed for handling missing data. For example, respondents with missing data may be dropped so that one performs a complete-case analysis (Groves, Dillman, Eltinge, & Little, 2002). Alternatively, one may replace the missing values by 'estimates' to produce 'complete data' and these are then analyzed by standard methods. The replacement of the missing values by estimates is known as imputation. A commonly used approach replaces the missing value with the mean of the variable (i.e., mean substitution). A second strategy is hot-deck imputation (Hanson, 1978, cited in Groves, Dillman, Eltinge, & Little, 2002). Hot-decking is based on matching the respondent with the omitted response(s) to another individual based on variables (e.g., items) that are observed for both persons (if there are multiple matching candidates, then an individual is selected at random). The omitted responses are replaced with the responses from the matched individual. Both of these strategies are considered to be single imputation methods. A single imputation method is an approach where each missing value is replaced by a plausible value and then the 'complete' data are analyzed (Sinharay, Stern, & Russell, 2001). While specific single imputation methods may have specific disadvantage(s), a general disadvantage of single imputation methods is that they cannot represent all of the uncertainty about which value to impute (Groves, Dillman, Eltinge, & Little, 2002). To address this disadvantage of single imputation methods multiple imputation has been developed. In multiple imputation a set of M datasets are created, each containing different sets of imputation of missing values (Groves, Dillman, Eltinge, & Little, 2002). Each of these M datasets is analyzed and the results across the M analyses are combined to produce an estimate plus an assessment of its variability. In contrast to imputation methods, maximum likelihood (ML) utilizes a stochastic model and makes inferences based on the likelihood function of the incomplete data (Groves, Dillman, Eltinge, & Little, 2002). When data are missing at random the likelihood approach yields valid inferences about the relevant parameters (Groves, Dillman, Eltinge, & Little, 2002; Sinharay, Stern, & Russell, 2001).

This study was concerned with the accuracy of person location estimates when respondents choose not to answer one or more items on an instrument that uses rating scales. The responses to affective or attitudinal instruments that use a rating scale, such as, a Likert response scale, may be modeled using Andrich's (1978a, 1978b) rating scale model (RSM). The RSM states that the probability of responding in category x_i of an (m+1)-category item i can be obtained by

$$p(x_{ik} \mid \theta) = \frac{\exp \sum_{j=0}^{x} \exp(\theta - (b_i + \tau_j))}{\sum_{k=0}^{m} \exp \sum_{j=0}^{k} (\theta - (b_i + \tau_j))} , \qquad (1)$$

where θ is the person location on the latent continuum being measured by the instrument and b_i is the item's location on the same continuum. τ_j represents the j^{th} threshold or the location of the transition from one response category to the next; $\tau_0 \equiv 0$. Therefore, there are m τs estimated for the m+1 response categories across all items.

The RSM is a member of the Rasch family and, as such, the RSM assumes items are equally effective at discriminating among examinees. Moreover, the unweighted sum of the respondent's responses (scale score) is a sufficient statistic for estimating the respondent's location. Therefore, it might be expected that because omitting responses affects the individuals' scale score that estimation of the person's location will be adversely affected.

Four different approaches for handling missing data were investigated for their capability to mitigate against the effect of omitted responses on person location estimation. These methods included Ignoring the omitted response, selecting the "midpoint" response category, Hot-decking, and a Likelihood-based approach.

Ignoring the omitted response had the effect of reducing the number of items used for estimating the person's location and thereby affecting the respondent's sufficient statistic for location estimation. This strategy of ignoring nonignorable missing data assumes that the omissions do not contain any useful information for estimating the respondent's location.

Replacing the omitted response with the "midpoint" response category (in effect, assuming the response is neutral-like) does not reduce the number of items used in calculating the sufficient statistic. However, to the extent that this 'neutral' response is not reflective of the respondent's true response (e.g., strongly disagreeing with an item) this approach may introduce additional measurement error.

The Hot-decking strategy selects a respondent (say, B) who is most similar to the respondent with the missing response(s) (say, A) in terms of the respondent's string , but who has also answered the item that respondent A did not respond to. Respondent B's response to the item in question is used for respondent A's response to the item.

In the Likelihood approach the various possible responses are substituted for each omitted response and the likelihood of that response pattern is calculated conditional on the location estimate, $\hat{\theta}$, corresponding to the response vector's sufficient statistic. For instance, let us say that the respondent has omitted one item and there are four possible response options (1, 2, 3, 4). In this approach the omitted response would be replaced a response of 1 and the likelihood based the corresponding sufficient statistic's $\hat{\theta}$ calculated. Then the omitted response would be replace by a response of 2 and the likelihood recalculated and so forth for responses of 3 and 4. The θ associated with the largest of the four likelihoods was taken as the $\hat{\theta}$. Obviously, as the number of omissions increases the number of combinations of potential responses also increases. This strategy attempts to determine what the most likely responses should be. It is assumed that if a respondent does not feel that an item is not applicable to him or her, that he or she is presented the opportunity to select a 'not applicable' category and, therefore, omissions are a function of a desire not to answer a particular question (e.g., the question is of a sensitive nature).

Method

Data Generation:

The simulation data were modeled on a empirical data set. This empirical data set consisted of 4282 respondents. These data consisted of responses to a questionnaire concerning sexual behavior and was administered as part of an HIV Counseling and Testing program. Fifteen four-point Likert scale (1=strongly disagree to 4=strongly agree) questions concerning opinions about condom use formed the scale of interest. Because a respondent may omit an item as a function of many different factors (e.g., uncomfortableness with the question, etc.) and there were no explicit measures of these factors it was decided to not use a parametric approach for modeling the empirical data. Because the omission pattern across the scale scores differed for persons who responded in one category (e.g., strongly agree) versus another response category on an item, four contingency tables were created for each item using the 4282 respondents. Each contingency table consisted of a two-level response type variable versus the scale score variable. The two-level response type variable reflected omission and one of the response alternatives. For example, for one table the response type variable consisted of response omission and responding strongly disagree, for a second table the response type variable consisted of response omission and responding disagree, etc. The scale score was transformed into deciles. Based on these tables the proportion of individuals omitting a response to an item conditional on the fractile were calculated. Some tables had cells with zero frequencies. In these cases, a value of 0.5 was substituted for the zero frequency before calculating the proportions (i.e., resulting 'frequency' was 0.005).

The generation of the simulation data required item parameter estimates. Using only individuals that had complete response strings (N = 3473), BIGSTEPS (Linacre & Wright, 2001) was used to obtain item parameter estimates for the RSM.

The simulated data were generated on the basis of the RSM and the item parameter estimates of the empirical data were treated as known. For each 0.1 of logit from -2.0 to 2.0 (inclusive) 1000 θs were generated for a total of 41,000 simulees. For each simulee the probability of a response in each category was calculated according to the RSM. These probabilities were then accumulated across response categories and compared to a uniform random number [0,1]. If the random number was less than or equal to probability of a category's cumulative probability, then that category's ordinal position was the response for the item. To generate the omission data, the scale score for each simulee's response vector was determined and the simulee assigned to one of the ten fractiles. For each item the simulee's response was used to determine which of the four contingency tables for the item should be used. Based on the simulee's fractile assignment the appropriate relative frequency of omission was compared to a uniform random number [0,1]. If the uniform random number was less than or equal to the relative frequency for omission, conditional on the simulee's fractile, then the response was changed to be an omission, otherwise the simulee's response to the item was not changed. For example, for an item the relative frequency of omission for an respondent in the third fractile might be 0.40, 0.30, 0.20, 0.10 for the strongly disagree, disagree, agree, strongly agree categories, respectively. If the simulee's generated response to the item was strongly disagree, then a uniform random number would be generated and compared to 0.40. If this random number was, for instance 0.3, then the simulee's response to this item would be changed to reflect that it had been omitted. This process was repeated for each of the 15 items and for all simulees. Therefore, each simulee had a complete response vector and a response vector containing omitted responses (a.k.a., the omission vector).

Ability Estimation: For each simulee, an $\hat{\theta}$ based on the complete response vector and another based on the omission vector was obtained using maximum likelihood estimation (MLE) with the RSM. For the omission vector the various imputation methods described above were used to impute the missing response and then MLE was used for location estimation. Therefore, each simulee had a θ, an $\hat{\theta}$ based on the complete response vector, an $\hat{\theta}$ based on Ignoring the omitted response(s), an $\hat{\theta}$ based on using the Midpoint category for the omitted response(s), an $\hat{\theta}$ based on Hot-decking, and an $\hat{\theta}$ based on using the Likelihood strategy.

Analysis: Independent and Dependent Variables:

Each level of imputation method factor was crossed by the number of items omitted in the response vector (Nomitted). The Nomitted factor consisted of three levels: 1, 2, and 3 omitted responses. These three levels of Nomitted, 1, 2, and 3, represent approximately 7%, 13%, and 20% of the test length, respectively. The dependent variables were the various location estimates.

To assess the effect of omission on the accuracy of the person location the Root Mean Square Error of the estimate (RMSE) and bias were calculated with respect to the simulee's known location. In addition, RMSE and bias were calculated for the location estimate obtained using the complete response data. Because of the way the locations were generated it was possible to investigate the effect of omitted responses as a function of location as well as across the ability scale. RMSE was calculated according to:

$$\mathrm{RMSE}(\theta_k) = \sqrt{\frac{\sum (\hat{\theta} - \theta_k)^2}{n_k}} \tag{2}$$

$$\mathrm{Bias}(\theta_k) = \frac{\sum (\hat{\theta} - \theta_k)}{n_k} \tag{3}$$

where $\hat{\theta}$: location estimate based on one of the estimation methods using either the complete data ($\hat{\theta}_c$) or missing data ($\hat{\theta}_o$)

θ_k: simulee's location at logit k (-2.0, -1.9, -1.8, ..., 2.0)

n: the number of simulees at logit k

RMSE and Bias were calculated separately for the complete vectors and omission vectors. Because RMSEs for the complete vectors represented how well the simulees' locations could be estimated on the basis of complete response data, the RMSEs for the omission vectors were compared to the corresponding RMSEs for the complete vectors; this was also true for Bias. These absolute differences between the RMSE for the omission and complete vectors as well as the difference between Bias based on the omission and complete vectors were examined graphically for each condition. Only θ points with at least 10 observations were plotted. All statistics were calculated using convergent cases with listwise deletion of missing data.

Results

The item locations were distributed between -0.98 and 0.86 with τ_1 = -0.30, τ_2 = -0.02, and τ_3 = 0.31. The number of simulees that omitted 1, 2, and 3 items were 7040, 3224, and 2994, respectively. For the one omit level the average trait values was $\bar{\theta}$ = 1.0421 (SD = 0.9721; Skew = -0.4402), for the two-omit level $\bar{\theta}$ = -0.1911 (SD = 1.2181; Skew = -0.2872), and for the three-omit level $\bar{\theta}$ = -1.1971 (SD = 1.0550; Skew = 0.1027). Table 1 contains descriptive statistics, the fidelity coefficients based on complete data ($r_{\theta\hat{\theta}_c}$) and that based on missing data ($r_{\theta\hat{\theta}_o}$), as well as the correlation between the location estimate based on complete data and that based on missing data ($r_{\hat{\theta}_c\hat{\theta}_o}$). As would be expected, $r_{\theta\hat{\theta}_o}$s for a given level of Nomitted were always less than the corresponding $r_{\theta\hat{\theta}_c}$ for that level of Nomitted. With respect to missing data strategy, the $\hat{\theta}_o$s had the strongest linear relationship with the θs under the Hot-decking approach for the 1 and 2 omit levels. For three omits the Midpoint strategy yielded the largest fidelity coefficient. However, the difference between the largest and smallest fidelity coefficients (i.e., $r_{\theta\hat{\theta}_o}$) across levels was always less than 0.02.

Insert Table 1 about here

While the fidelity coefficients indicate the degree of linear agreement between two scales, they do not indicate the accuracy of estimation. To assess the accuracy of location estimation RMSE($\hat{\theta}$) and Bias($\hat{\theta}$) were calculated. Because the RMSE($\hat{\theta}$) based on complete response data indicates how well one can expect to do with this item pool, the RMSE plots represent the difference between the RMSE($\hat{\theta}$) based on the complete response data value and that based on the response vectors with missing data. These RMSE($\hat{\theta}$) differences as a function of θ for the various missing data strategies are presented in Figures 1 - 3 for the one-, two-, and three-omit levels, respectively. From Figure 1 one sees that there appears to be little difference between Ignoring omits, the Midpoint, and the Hot-decking methods for the upper half of the θ continuum. Moreover, for this portion of the θ continuum the Likelihood method did not function as well as the other strategies. In the lower half of the θ continuum the Midpoint, the

Hot-decking, and the Likelihood methods were very similar to one another. It appears that Hot-decking performed the best for the greatest range θ for the one-omit level.

Insert Figure 1 about here

Figures 2 and 3 show that despite the large fidelity coefficients for the Ignore Omit(s) missing data strategy, that this method does not yield accurate $\hat{\theta}$s for most of the θ scale for both the two- and three-omit conditions. As was the case for the one-omit level, the Hot-decking procedure had, in general, RMSE($\hat{\theta}$)s that agreed better with those based on complete data than did the other methods for θs above -0.9 for both the two- and three-omit conditions. The Midpoint missing data strategy performed almost as well as the Hot-decking procedure, with the Likelihood strategy performing worse than both the Midpoint and Likelihood procedures.

Insert Figures 2 and 3 about here

Similar to RMSE($\hat{\theta}$), the Bias($\hat{\theta}$) based on the complete response data indicates how well one can expect to do with this item pool, therefore the plots represent the difference between this value and the Bias($\hat{\theta}$) based on the missing data response vectors. The Baseline represents perfect agreement between the Bias($\hat{\theta}$) based on complete data and that using a missing data procedure for location estimation. With respect to Bias($\hat{\theta}$) and the one-omit level (Figure 4) one finds that the Hot-decking missing data procedure exhibited greater agreement with the bias based on complete data than did the other methods. The Midpoint strategy also showed similar bias to that of the complete data results between -1.0 and 1.0. Ignoring the missing data introduced substantially more overestimation bias than was found with the complete data. The Likelihood approach tended to yield $\hat{\theta}$s that were larger than those based on complete data throughout the θ continuum.

Insert Figure 4 about here

For the two-omits level (Figure 5) the Hot-decking performed similar to that observed with the complete data and better than the other missing data procedures between approximately -1.0 to 1.5. As was the case with the one-omit level, the Midpoint strategy performed almost as well as Hot-decking in terms of agreeing with the Bias($\hat{\theta}$) based on complete response vectors. In general, the Likelihood procedure performed similar to the Midpoint and Hot-decking procedures between -0.5 and 0.2 and better than Ignoring the omitted responses. However, across the θ continuum the Likelihood strategy did not do as well as Hot-decking or using the midpoint value. As was the case with the one-omit level, Ignoring the omits resulted in $\hat{\theta}$s that were smaller than those based on complete data.

Insert Figure 5 about here

The results for the three-level omits (Figure 6) followed that of the two-level omits. Above $\theta \cong -1.0$ Bias($\hat{\theta}$) based on Hot-decking the missing data vectors agreed well with that from the complete response vectors. The Midpoint strategy also performed reasonably well above $\theta \cong -0.3$ with the Likelihood approach performing similarly or slightly better as θ increased above -1.0. Compared to the Bias($\hat{\theta}$) based on complete data, Ignoring the omitted responses performed substantially worse than the other strategies for most of the continuum.

Insert Figure 6 about here

Discussion

Respondents choose to not answer certain questions for a variety of reasons. This study used a nonparametric approach with empirical data in order to provide realistic guidance for generating simulation data to ascertain the effect of omission on location estimation with the RSM.

As would be expected, as the number of omissions increased the accuracy of $\hat{\theta}$ decreased for a given missing data strategy. The above results seem to indicate that with the RSM omits should not be ignored. This was particularly true for the two and three omit conditions. While Ignoring omits yielded reasonably accurate $\hat{\theta}$s in terms of RMSE (see Figure 1), Figure 4 showed that these estimates exhibited underestimation bias throughout the θ continuum. Of the imputation methods, Hot-decking appeared to be, overall, the best strategy to use in terms of producing RMSE($\hat{\theta}$) and Bias($\hat{\theta}$) that agreed with that obtained from complete data. It should be noted that given the logit range represented in the item pool, the RMSE($\hat{\theta}$) and Bias($\hat{\theta}$) below -1.0 and above 1.0 may only be indicative, but not very stable.

Although the Midpoint imputation method introduced measurement error, from an estimation perspective using the Midpoint strategy produced reasonably good $\hat{\theta}$s for one- and two-omit conditions. This strategy did not performed as well in the three omit condition as it had in the one- and two-omit conditions. This may be due the cumulative effect of the introduction of measurement error with each imputation. For instance, comparatively speaking, imputing a neutral response for the first omit may not be that deleterious in terms of the amount of measurement error introduced (i.e., whenever the actual response was not neutral then imputing a neutral response introduced error). However, as one imputes more neutral responses then the sufficient statistic becomes more distorted due to the increased measurement error. As a result, the accuracy of the corresponding $\hat{\theta}$ is degraded.

Theoretically, it was expected that the Likelihood approach would have performed better than it did because it attempted to determine which was the most likely response pattern based on the current complete information. However, this expectation was not realized. While under some conditions the Likelihood strategy performed comparable to the Midpoint strategy (e.g., the two-omit level between -0.7 and 0.2 as well as the three-omit condition between -0.25 and 0.25), the Likelihood never performed as well as Hot-decking across the θ continuum. In the Likelihood approach each omit was replaced by one of the possible responses and the likelihood of the response vector recalculated based on the response vector's $\hat{\theta}$. As is well-known certain response patterns (e.g., Guttman patterns) have a higher likelihood of occurrence than other response patterns. It was this property that was expected to be exploited by the Likelihood approach. These results seem to indicate that the most likely response pattern did not correspond to the observed complete response pattern. This may be due to the stochastic nature of the data.

The results indicate that when using the RSM and MLE for location estimation, Hot-decking may be the preferred approach to use with missing data. It should be noted that for this study the data were generated, in part, according to the RSM. To the extent that a nonRasch family model more accurately describes the data, then one may not observe results comparable to those seen here.

References

Andrich, D. (1978a). A rating formulation for ordered response categories. *Psychometrika, 43,* 561-573.

Andrich, D. (1978b). Application of a psychometric rating model to ordered categories which are scored with successive integers. *Applied Psychological Measurement, 2,* 581-594.

Groves, R.M., Dillman, D.A., Eltinge, J.L., & Little, R.J.A. (2002). *Survey nonresponse.* New York: Wiley.

Linacre, J.M., & Wright, B.D. (2001). *A user's guide to BIGSTEPS.* Chicago: MESA Press.

Little, R., & Rubin, D.B. (1987). *Statistical analysis with missing data.* New York: Wiley.

Lord, F. M. (1980). *Applications of item response theory to practical testing problems.* Hillsdale, NJ: Erlbaum.

Mislevy, R.J., & Wu, P. (1988). *Inferring respondent ability when some item responses are missing* (RR 88-48-ONR). Princeton, NJ: Educational Testing Service.

Mislevy, R.J., & Wu, P. (1996). *Missing responses and IRT ability estimation: Omits, choice, time limits, and adaptive testing* (RR 96-30-ONR). Princeton, NJ: Educational Testing Service.

Shulz, E.M., & Sun, A. (2001). Controlling for rater effects when comparing survey items with incomplete Likert data. *Journal of Applied Measurement, 2,* 337-355.

Sinharay, S., Stern, H.S., & Russell, D. (2001). The use of multiple imputation for the analysis of missing data. *Psychological Methods, 6,* 317-329.

Table 1: Descriptive Statistics and Fidelity Coefficients

Omits	Method	$r_{\theta\hat{\theta}_c}$	$r_{\theta\hat{\theta}_o}$	$r_{\hat{\theta}_c\hat{\theta}_o}$	$\bar{\hat{\theta}}_c$[a]	$\bar{\hat{\theta}}_o$	s_o	$Skew_o$
1 omit	Ignore	0.9264	0.9116	0.9820	1.0192	0.7703	0.7742	-1.4528
	Midpoint		0.9175	0.9867		0.9641	0.8151	-0.6997
	Hotdeck		0.9225	0.9894		0.9939	0.8453	-0.6257
	Likelihood		0.9095	0.9820		1.1709	1.0153	-0.4355
2 omits	Ignore	0.9586	0.9458	0.9846	-0.2240	-0.4905	1.1498	-0.8751
	Midpoint		0.9555	0.9807		-0.1155	0.9452	-0.1271
	Hotdeck		0.9563	0.9777		-0.1323	1.0069	-0.1116
	Likelihood		0.9469	0.9826		-0.1197	1.3632	-0.2977
3 omits	Ignore	0.9237	0.9083	0.9697	-1.2526	-1.5731	1.0808	-0.1191
	Midpoint		0.9174	0.9553		-0.8162	0.6512	0.5453
	Hotdeck		0.9014	0.9313		-0.9557	0.7470	0.4638
	Likelihood		0.8976	0.9608		-1.3282	1.1907	0.0067

[a]1 omit: $s_c = 0.9187$, $Skew_c = -0.7492$; 2 omits: $s_c = 1.2282$, $Skew_c = -0.5589$;
3 omits: $s_c = 1.0888$, $Skew_c = -0.1077$

BEST COPY AVAILABLE

Figure 1: One Omit, RMSE

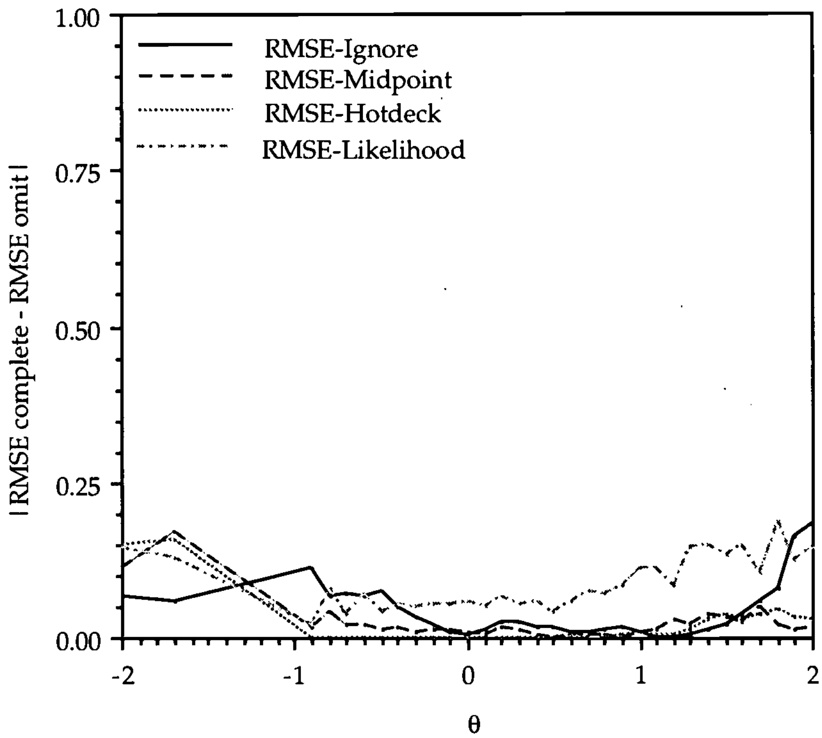


Figure 2: Two-Omits, RMSE

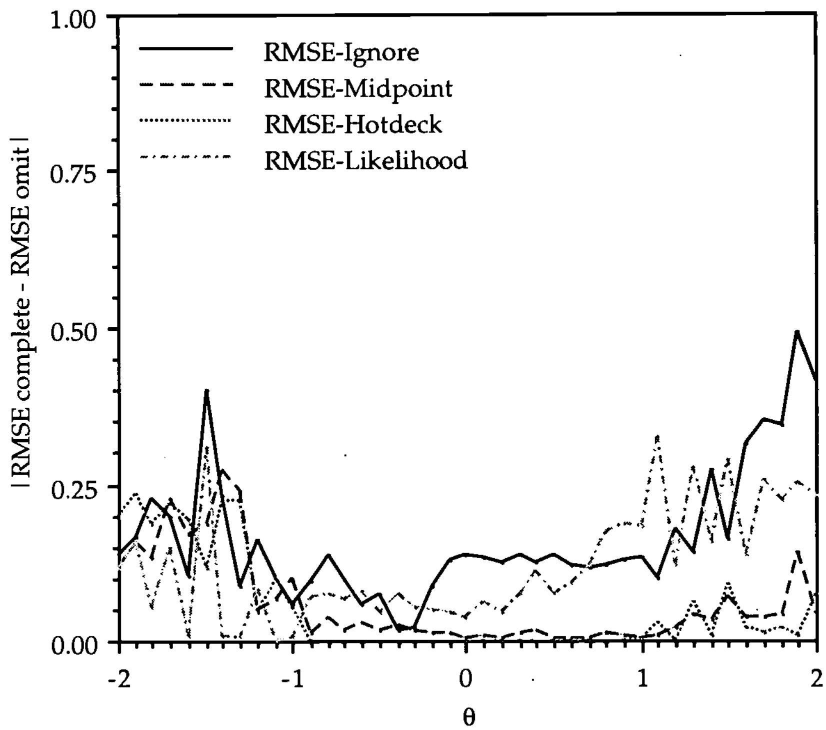


Figure 3: Three-Omits, RMSE

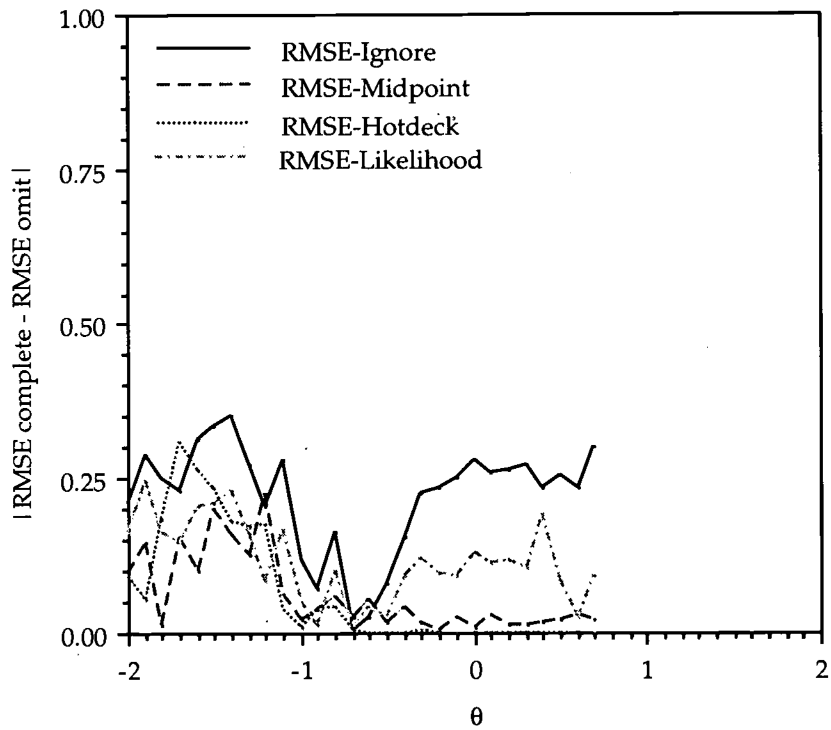


Figure 4: One-Omit, Bias

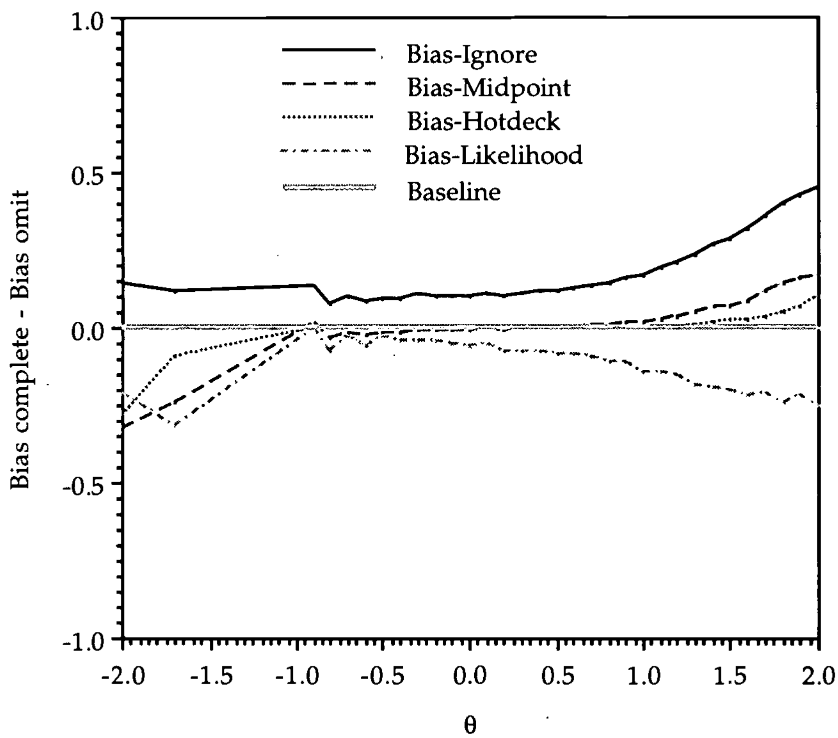


Figure 5: Two-Omits, Bias

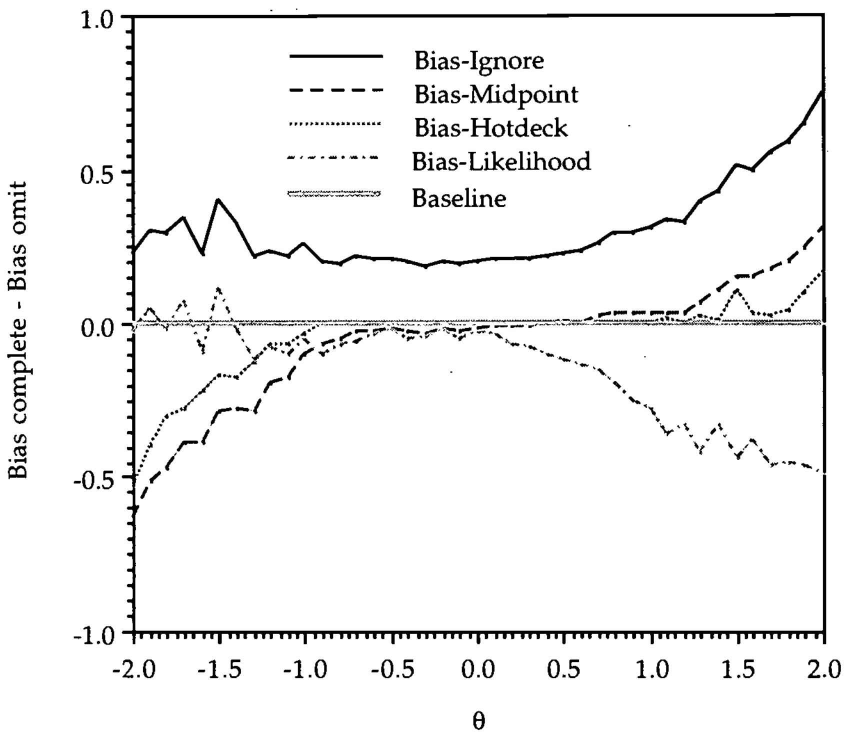
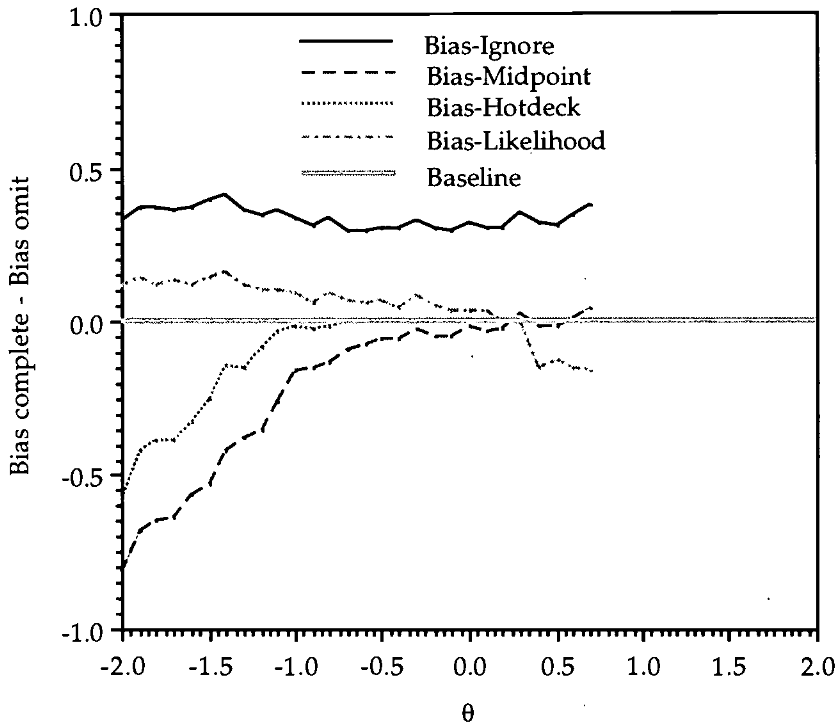


Figure 6: Three-Omits, Bias



U.S. Department of Education
Office of Educational Research and Improvement (OERI)
National Library of Education (NLE)
Educational Resources Information Center (ERIC)

ERIC®

TM034392

REPRODUCTION RELEASE

(Specific Document)

I. DOCUMENT IDENTIFICATION:

Title: THE EFFECT OF MISSING DATA ON ESTIMATING A RESPONDENT'S LOCATION USING RATINGS DATA	
Author(s): R J De Ayala	
Corporate Source: UNIV OF NEBRASKA	Publication Date:

II. REPRODUCTION RELEASE:

In order to disseminate as widely as possible timely and significant materials of interest to the educational community, documents announced in the monthly abstract journal of the ERIC system, *Resources in Education* (RIE), are usually made available to users in microfiche, reproduced paper copy, and electronic media, and sold through the ERIC Document Reproduction Service (EDRS). Credit is given to the source of each document, and, if reproduction release is granted, one of the following notices is affixed to the document.

If permission is granted to reproduce and disseminate the identified document, please CHECK ONE of the following three options and sign at the bottom of the page.

The sample sticker shown below will be affixed to all Level 1 documents	The sample sticker shown below will be affixed to all Level 2A documents	The sample sticker shown below will be affixed to all Level 2B documents
PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 1	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE, AND IN ELECTRONIC MEDIA FOR ERIC COLLECTION SUBSCRIBERS ONLY, HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2A	PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL IN MICROFICHE ONLY HAS BEEN GRANTED BY Sample TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC) 2B
Level 1 [X]	Level 2A [ ]	Level 2B [ ]
Check here for Level 1 release, permitting reproduction and dissemination in microfiche or other ERIC archival media (e.g., electronic) *and* paper copy.	Check here for Level 2A release, permitting reproduction and dissemination in microfiche and in electronic media for ERIC archival collection subscribers only	Check here for Level 2B release, permitting reproduction and dissemination in microfiche only

Documents will be processed as indicated provided reproduction quality permits.
If permission to reproduce is granted, but no box is checked, documents will be processed at Level 1.

I hereby grant to the Educational Resources Information Center (ERIC) nonexclusive permission to reproduce end disseminate this document es indicated above. Reproduction from the ERIC microfiche or electronic media by persons other than ERIC employees and its system contractors requires permission from the copyright holder. Exception is made for non-profit reproduction by libraries and other service agencies to satisfy information needs of educators in response to discrete inquiries.

Sign here,→ please

Signature: Rafael De Ayala	Printed Name/Position/Title: RAFAEL DE AYALA, PROFESSOR	
Organization/Address: UNIV OF NEBRASKA	Telephone: 402 472 2736	FAX: 402 472 8319
	E-Mail Address: rdeayala2@unl.edu	Date: 7/29/02

(over)

III. DOCUMENT AVAILABILITY INFORMATION (FROM NON-ERIC SOURCE):

If permission to reproduce is not granted to ERIC, *or*, if you wish ERIC to cite the availability of the document from another source, please provide the following information regarding the availability of the document. (ERIC will not announce a document unless it is publicly available, and a dependable source can be specified. Contributors should also be aware that ERIC selection criteria are significantly more stringent for documents that cannot be made available through EDRS.)

Publisher/Distributor:
Address:
Price:

IV. REFERRAL OF ERIC TO COPYRIGHT/REPRODUCTION RIGHTS HOLDER:

If the right to grant this reproduction release is held by someone other than the addressee, please provide the appropriate name and address:

Name:
Address:

V. WHERE TO SEND THIS FORM:

Send this form to the following ERIC Clearinghouse:

University of Maryland
ERIC Clearinghouse on Assessment and Evaluation
1129 Shriver Laboratory
College Park, MD 20742
Attn: Acquisitions

However, if solicited by the ERIC Facility, or if making an unsolicited contribution to ERIC, return this form (and the document being contributed) to:

ERIC Processing and Reference Facility
4483-A Forbes Boulevard
Lanham, Maryland 20706

Telephone: 301-552-4200
Toll Free: 800-799-3742
FAX: 301-552-4700
e-mail: ericfac@inet.ed.gov
WWW: http://ericfac.piccard.csc.com

EFF-088 (Rev. 2/2000)