Biemer (2011) Latent Class Analysis of Survey Error. Wiley

R_1 will be overestimated. In other words, positive correlations between the errors in the two measurements will make the measurements appear to be more consistent than they are, thus negatively biasing the index of inconsistency. Stated another way, positively correlated errors will result in erroneously attributing greater reliability to the measurements.

To summarize, some violations of the parallel assumptions will result in negative bias while others will result in positive bias in the estimates of R. In general, the bias in $\hat{R}$ is unpredictable. Examples of this unpredictability are provided in the following illustration.

2.3.3 Example: Reliability of Marijuana Use Questions

To illustrate the estimation methodology, we use data from a large, national survey on drug use. In this survey, data on past-year marijuana use were collected in the same interview using three different methods. Responses were recoded to produce three dichotomous measures of past-year marijuana use denoted by y_1, y_2, and y_3, where 1 denotes use and 0 denotes no use. The frequencies for all possible response patterns are shown in Table 2.5. Although the sample was drawn by a complex multistage unequal probability design, simple random sampling will be assumed for the purposes of the illustration. Cell counts have been weighted and rescaled (see Section 5.3.4 for more detail on the approach) to the overall sample size, n = 18,271.

The parallel measures assumption can be tested by comparing the three proportions, p_{1++}, p_{+1+}, and p_{++1}. (Note that we have extended the "+" notation for two measurements in Section 2.2.2 to three measurements.) If a test of equality is rejected, the assumption of parallel measures is not supported by the data; however, failure to reject does not suggest that y_1, y_2, and y_3 are parallel. We compute the marginal proportion for y_1 as p_{1++} as (1181 + 96 + 17 + 15)/18,269 = 0.072. In the same way, the marginal proportion for y_2 is $p_{+1+} = 0.084$ and for y_3, it is $p_{++1} = 0.084$. Although the proportions are all close, p_{1++} differs significantly from the other two proportions, and thus the test is rejected,

Table 2.5 Past-Year Marijuana Use for Three Measures

y_1	y_2	y_3	Count
1	1	1	1,181
1	1	0	96
1	0	1	17
1	0	0	15
0	1	1	113
0	1	0	150
0	0	1	229
0	0	0	16,470

indicating that y_1 is no
are parallel cannot be
estimates of reliability
biased. Nevertheless,
measurements.
The estimation of
as in (2.34). Note tha
(when only one of y_1,
1 (when $y_1 = y_2 = y_3 =$
latter occurring with f

To compute $\hat{\sigma}_y^2$ from
$\bar{p}$ as

$\bar{p} = \frac{1\times1181}{}$

and thus

$s_1^2 = \frac{1181\times(1-0.08)^2}{}$

$= 0.06612$

The estimate of σ_y^2
Combining these resu

and the index of inco
Using the US Cen
tude of $\hat{I}$ (i.e., $0 \le \hat{I} \le$
this result suggests th
year marijuana use is
We can use the fir
tions for the case whe
Table 2.4 over the y
$p_{11} = 0.0699$, $p_{10} = 0.0$
$p_{1+} = 0.072$ and $p_{+1} =$

›sitive correlations between the
he measurements appear to be
y biasing the index of inconsis-
ed errors will result in errone-
ısurements.
allel assumptions will result in
ve bias in the estimates of R. In
ples of this unpredictability are

Questions

use data from a large, national
t-year marijuana use were col-
rent methods. Responses were
res of past-year marijuana use
and 0 denotes no use. The fre-
shown in Table 2.5. Although
ge unequal probability design,
he purposes of the illustration.
e Section 5.3.4 for more detail
= 18,271.
ested by comparing the three
ave extended the "+" notation
e measurements.) If a test of
measures is not supported by
gest that y_1, y_2, and y_3 are paral-
s p_{1++} as (1181 + 96 + 17 + 15)/
roportion for y_2 is $p_{+1+} = 0.084$
rtions are all close, p_{1++} differs
and thus the test is rejected,

or

Count
1,181
96
17
15
113
150
229
16,470

indicating that y_1 is not parallel compared to y_2 or y_3. However, that y_2 and y_3 are parallel cannot be rejected on the basis of this test. Thus, we expect our estimates of reliability on the basis of all three measurements to be somewhat biased. Nevertheless, we proceed with the estimation of R using all three measurements.

The estimation of R using (2.27) will be demonstrated with s_2^2 computed as in (2.34). Note that p_i is either 0 (which occurs when $y_1 = y_2 = y_3 = 0$), $\frac{1}{3}$ (when only one of y_1, y_2, or y_3 is 1), $\frac{2}{3}$ (when only one of y_1, y_2, or y_3 is 0), or 1 (when $y_1 = y_2 = y_3 = 1$). Thus, $p_i q_i$ can only take on the values 0 or $\frac{2}{9}$, the latter occurring with frequency 620 (= 96 + 17 + 15 + 113 + 150 + 229). Thus

$$s_2^2 = \frac{3}{18{,}271(3-1)}\left(\frac{2}{9}\times 620\right) = 0.01131$$

To compute $\hat{\sigma}_y^2$ from (2.28), we first compute s_1^2 using (2.34). We compute $\bar{p}$ as

$$\bar{p} = \frac{1\times 1181 + \frac{2}{3}\times(96+17+113) + \frac{1}{3}\times(15+150+229)}{18{,}271} = 0.08$$

and thus

$$s_1^2 = \frac{1181\times(1-0.08)^2 + 226\times(\frac{2}{3}-0.08)^2 + 394\times(\frac{1}{3}-0.08)^2 + 16{,}470\times(0-0.08)^2}{18{,}271-1}$$
$$= 0.06612$$

The estimate of σ_y^2 from (2.28) is therefore $0.06612 + \frac{2}{3}\times 0.01131 = 0.07366$. Combining these results, we find that the estimator of R is

$$\hat{R} = 1 - \frac{0.01131}{0.07364} = 0.8465$$

and the index of inconsistency $\hat{I} = 0.15$.

Using the US Census Bureau's rule of thumb for interpreting the magnitude of $\hat{I}$ (i.e., $0 \le \hat{I} \le 0.2$ is good, $0.2 < \hat{I} \le 0.5$ is moderate, and $\hat{I} > 0.5$ is poor), this result suggests that the average reliability of the three measures of past-year marijuana use is good.

We can use the first two columns of Table 2.5 to demonstrate the calculations for the case where there are only two measurements. Begin by collapsing Table 2.4 over the y_3 measurement to form a 2 × 2 crossover table having $p_{11} = 0.0699$, $p_{10} = 0.0018$, $p_{01} = 0.0144$, and $p_{00} = 0.9140$. Thus, $g = 0.01615$ and $p_{1+} = 0.072$ and $p_{+1} = 0.084$. The NDR is 0.072 – 0.084 = –0.012. To test for parallel measurements, we compute the standard error as $\sqrt{Var(\text{NDR})}$, where Var(NDR) is as given in (2.48). This yields $\sqrt{(0.01615+1/18{,}271)/18{,}271}$ $\doteq 0.001$, and NDR is highly significant. However, the size of the NDR is small relative to p_{1+} and p_{+1}, so the departure from parallel measures does not seem important for the purposes of reliability estimation.

The denominator of R is $0.072(1-0.084)+0.084(1-0.072)=0.1439$. It follows from (2.37) that $\hat{R}=1-(0.0161)/(0.144)=0.89$, which again indicates good reliability. To calculate κ from (2.42), we first compute $p_0=p_{11}+p_{00}=0.984$ and then $p_e=p_{1+}p_{+1}+p_{0+}p_{+0}=0.856$. Hence, $\kappa=(0.984-0.856)/(1-0.856)=0.89$, which, as expected, is the same number produced by the calculation of $\hat{R}$.

As previously noted, both estimates of reliability are biased to some extent because of the failure of the parallel assumptions to hold. The result in (2.58) suggests that SRV may be biased upward because $D_{12}^2>0$ and, consequently, estimates of R should be biased downward. This may not be true if $\rho_{12}\neq 0$, but no information on this correlation is available. Because the assumption that $D_{23}^2=0$ is plausible is based on the test of equality of p_{+1+} and p_{++1}, we can also compute $\hat{R}$ using y_2 and y_3 and compare this estimate with our previous results. This yields an estimate of $\hat{R}=0.81$, which is lower than the value computed for y_1 and y_2. One possible explanation is that $\rho_{12}\sqrt{\text{SRV}_1\text{SRV}_2}>D_{12}^2$ which, as can be seen from (2.58), results in a positive bias in $\hat{I}$ computed from y_1 and y_2. Unfortunately, the data in Table 2.5 are not sufficient for resolving this puzzling result.

2.3.4 Designs Based on a Subsample

Reinterviews will usually cost less than the original interviews because they often require less effort to contact and interview the respondent. However, most survey budgets cannot bear the costs of reinterviewing the entire sample, especially considering that the reinterview data are solely for evaluation purposes. Often it is sufficient to reinterview a subsample of between 10% and 50% of the survey respondents to achieve acceptable precision for the estimates of reliability. Let $n'<n$ denote the size of a simple random subsample of the original sample. Because this subsample constitutes a simple random sample from the entire population, the theory developed in the previous sections still holds after replacing n everywhere by n'. However, it is possible to obtain better precision for the estimates of R, I, and κ by computing the denominator of these ratios over the full sample.

An alternative estimator of σ_y^2 can be constructed by using only the main survey (i.e., assuming $m=1$) and noting from (2.23) that an estimator of the total variance σ_y^2 is $\hat{\sigma}_y^2=s^2$, where s^2 as defined in (2.14) is based on the n main survey observations ignoring the repeated measurements. Let $q_{1+}=1-p_{1+}$ and $q_{+1}=1-p_{+1}$. If $p_{1+}q_{1+}$ or $p_{+1}q_{+1}$ calculated from the interview–reinterview table is based on a small sample size, a more efficient estimator can be produced by

substituting $p_{1+}q_{1+}+p$
units in the sample ar
the reliability ratio be

A somewhat related s
random. For example
households and house
dormant for 8 month
surveys, only telephone
high costs of personal
problem is reinterview
survey nonresponse r
nism, the estimates of
biased if the proportio
very different error dis
still be computed and
it is important to reali
tation of the interview

2.4 RELIABILITY

Related to the concep
evaluation is the use
these are used extens
these types of measur
and to contrast them
However, the topic is
further beyond the pr

2.4.1 Scale Score Me

Scale score measuren
embedded in a quest
equivalent. The items
individual items, is th
psychological and soc
Checklist (CBCL) is
children (Achenbach
lems, each scored on
3 = often true of the