August 4, 1994

TWO-SAMPLE MCNEMAR TESTS FOR COMPLEX SURVEY DATA

Jenny Thompson and Robin Fisher
Jenny Thompson, Demographic Statistical Methods Division, Bureau of the Census,
Washington, DC 20233

KEY WORDS: Current Population Survey, Parallel Survey, Nonparametric Statistics.

Feuer and Kessler (1989) generalized the McNemar test (1947) to a two sample situation where the hypothesis of interest is that the marginal changes in each of two independent sample's tables are equal. We show two refinements of this test for complex survey data, which require different estimates of variance. In particular, we present these tests along with applications from the Current Population Survey's Parallel Survey split panel data and from the Current Population Survey's CATI Phase-in data.

1. Introduction

Feuer and Kessler (1989) generalized the McNemar test (1947) to a two-sample situation where the hypothesis of interest is that the marginal changes in each of two independent samples' 2 x 2 tables are equal. The application presented was for a two sample cohort analysis and assumed simple random sampling.

Further modifications of the test statistic are necessary for a complex survey data application of the two-sample McNemar test. First, because the data are not obtained through a simple random sample, a different estimate of the variance is required. Second, unless the survey has a longitudinal design, a separate link of individuals in two consecutive months' of data must be performed. In general, such a link will use a set of demographic variables and will include some false matches. This induces another variance component to the model, the error due to false matches.

We show two refinements of this test for complex survey data, which require separate estimates of variance. In particular, we present these tests along with applications to the Current Population Survey's Parallel Survey split panel study and from the Current Population Survey's CATI Phase-in Project. In Section 2 we describe these test modifications including background on the one and two-sample McNemar tests (Section 2.1), modifications for complex survey data (Section 2.2), and some remarks on applications to several months' data (Section 2.3). Section 3 presents the results from the application of these tests specifically to CPS Parallel Survey Data and to CPS CATI Phase-in data. We make some concluding remarks in Section 4. Details on the estimation of variances and covariances are included in the appendices.

2. Test and Modifications

2.1 General

A sample is randomly split into two independent representative samples (split panels). After a baseline measurement is taken, a new technique is administered in one panel, the treatment panel. The other panel serves as a control.

The responses are matched longitudinally after the second measurement is taken. A response can be +, -, or * (missing). Since this is matched data, the "**" cell will be empty.

This scenario is represented pictorially as

Treatment Panel

Month 2 Treatment

Month 1 No Treatment		+	-	*	
	+	x_{++}	x_{+-}	x_{+*}	$x_{+.}$
	-	x_{-+}	x_{--}	x_{-*}	$x_{-.}$
	*	x_{*+}	x_{*-}		$x_{*.}$
		$x_{.+}$	$x_{.-}$	$x_{.*}$	n

Control Panel

Month 2 No Treatment

Month 1 No Treatment		+	-	*	
	+	x_{++}'	x_{+-}'	x_{+*}'	$x_{+.}'$
	-	x_{-+}'	x_{--}'	x_{-*}'	$x_{-.}'$
	*	x_{*+}'	x_{*-}'		$x_{*.}'$
		$x_{.+}'$	$x\ '$	$x_{.*}'$	n'

where n is not necessarily equal to n′.

For each panel, define

$M_{(12)}$ as the set of cases which have month 1 and month 2 responses (matched cases). This set contains $n_{(12)} = (x_{++} + x_{+-} + x_{-+} + x_{--})$ elements;
$M_{(10)}$ as the set of cases which have month 1 responses, but no month 2 response. This set contains $n_{(10)} = (x_{+*} + x_{-*})$ elements;
$M_{(02)}$ as the set of cases which have month 2 responses, but no month 1 response. This set contains $n_{(02)} = (x_{*+} + x_{*-})$ elements.

First, consider the one-sample case. Traditionally, the one-sample McNemar test statistic is constructed from the $n_{(12)}$ and $n_{(12)}'$ matched responses. In the one-sample scenario, we test the hypothesis

H_0: $\text{Prob}(x_{+-}) = \text{Prob}(x_{-+})$
H_1: Not H_0

i.e., the hypothesis that the movement from one state to the other (+ to -, or - to +) is zero. We also refer to this movement as the flux.

The one-sample test can be a useful diagnostic in the two-sample situation. We examine the Control panel estimates to see if there is zero movement. Any significant movement in the Treatment panel can be measured as a deviation from zero flux or as a change in the probability of a "+."

The two-sample hypothesis is

H_0: $\text{Prob}(x_{-+}) - \text{Prob}(x_{+-}) = \text{Prob}(x_{-+}') - \text{Prob}(x_{+-}')$
H_1: Not H_0

In other words, the difference in the probabilities of switching in the two directions is the same, regardless of the treatment, or equivalently, the difference in panel fluxes is zero.

The Feuer and Kessler generalization (1989) to a two-sample McNemar test (described in 2.2.1. below) is confined to the $M_{(12)}$ and $M_{(12)}'$ sets. With an additional assumption, however, the unmatched responses can be included in computation of the test statistics. This assumption and resultant modification are described in section 2.2.2.

2.2 Complex Survey Modifications

2.2.1. Modification One: Restrict Analysis to Longitudinally Linked Data

This method is a straightforward application of the two-sample McNemar test, using longitudinally linked data from a complex survey. The domain for both months of data is given by $M_{(12)}$.

To construct the test statistic, consider one panel. Feuer and Kessler (1989) noted that

$$\begin{aligned} Prob(x_{-+}) - Prob(x_{+-}) &= [(Prob(x_{++}) + Prob(x_{-+})) - (Prob(x_{++}) +(Prob(x_{+-}))] \\ &= [Prob(x_{.+})^{\circ} - (Prob(x_{+.})^{\circ}] \\ &= p_2^{\circ} - p_1^{\circ} \end{aligned}$$

where p_2° is the marginal probability of a + response month 2, given that the respondent responded both months; and
where p_1° is the marginal probability of a + response month 1, given that the respondent responded both months.

The one-sample test statistic constructed from this panel's data is

$$Z_{\bar{1}} = \frac{p_{2}^{-} - p_{1}^{-}}{\sqrt{Var(p_{2}^{-} - p_{1}^{-})}}, \ where \ p_{1}^{-} = \frac{x_{++} + x_{+-}}{n_{(12)}}, \ p_{2}^{-} = \frac{x_{++} + x_{-+}}{n_{(12)}}$$

Given two independent panels, the two-sample test statistic is

Install Equation Editor and double-click here to view equation. (1a)

If the survey is designed to collect longitudinal data, then this modification is a natural extension of the method described by Feuer and Kessler. The extension is the use of weighted estimates and complex survey variances and covariances in place of simple random sample variances. For this type of survey design, an effective mechanism to link individuals from month to month is presumably in place. Often, however, this is not the case, and one data set must be physically linked to another. Consequently, the $n_{(12)}$ elements in the domain will contain some false matches, and some actual matches may be inadvertently excluded.

2.2.2. Modification Two: Use Each Month of Data to Construct Estimates of Marginal Probabilities

This method omits the longitudinal linkage step altogether, noting that the construction of the test statistic relies on estimates of marginal probabilities. To do this, assume that under the null hypothesis, the expected value of (Prob($x_{*.}$) - Prob ($x_{.*}$)) is zero. This is described for a simple random sampling application in Marascuilo et al (1988).

The domain for the first month of data is given by Install Equation Editor and double-click here to view equation. which contains $n_{(12)} + n_{(10)} = n_1$ elements. The domain for the second month of data is given by Install Equation Editor and double-click here to view equation. which contains $n_{(12)} + n_{(02)} = n_2$ elements.

The one-sample test statistic constructed from the unlinked data is given by Install Equation Editor and double-click here to view equation. .

Given two independent panels, the two-sample test statistic is

Install Equation Editor and double-click here to view equation. (1b)

As with the application described in 2.2.1, all estimates are weighted estimates, and variances are complex survey variances (as opposed to simple random sampling variances).

2.3 Linear Combinations

We can use our estimated covariance matrix to test linear combinations of Install Equation Editor and double-click here to view equation. over time, where Install Equation Editor and double-click here to view equation. and $\underline{p}_1$, $\underline{p}_2$, $\underline{p}'_1$, and $\underline{p}'_2$ are vectors containing the marginal probabilities for the time period under consideration.

Perhaps the most interesting (to our applications) of these tests is of the hypothesis H_0: $\underline{1}'\underline{\mu} = 0$,where $\underline{\mu}$ is the expected value of one of the vectors described above. Other general linear hypotheses of this form could be equally interesting. One might wish to test for contrast by time period, for example testing the average difference from January through June against the remainder of the year's data.

Another test of particular interest is the "omnibus hypothesis," where we test H_0: $\underline{\mu} = \underline{0}$.

The test statistics for this hypothesis are Install Equation Editor and double-click here to view equation. , and Install Equation Editor and double-click here to view equation. each of which has an approximate chi-squared distribution with r degrees of freedom, where r is the dimension of the vector of interest.

3. Applications

In this section, we apply the one and two-sample McNemar techniques for unlinked data outlined in 2.2.2 and 2.3 to two separate sets of data: PS split panel data and CPS CATI Phase-in data. Section 3.1 provides some background on the two studies' designs. Section 3.2 describes the panel estimates and variance estimates. Section 3.3 outlines our diagnostics. Section 3.4 provides our results, for both one and two-sample applications. Tables One and Two (section 3.4.1) provide the results for PS split panel data. Tables Three and Four (section 3.4.2) provide the results for the CPS CATI Phase-in data.

3.1 Background

The official monthly civilian labor force estimates from January 1994 onward are based on data from a comprehensively redesigned Current Population Survey (CPS). The redesign included implementation of a new, fully computerized questionnaire, and an increase in centralized computer-assisted telephone interviewing (CATI). To gauge the effect of the CPS redesign on published estimates, a Parallel Survey (PS) was conducted using the new questionnaire and data collection procedures from July 1992 through December 1993[1]. Special studies were embedded in both the PS and the CPS during the same time period to provide data for testing hypotheses about the effects of the new methodological differences on labor force estimates: the PS split panel study and the CPS CATI Phase-in Project (a continuation of the study presented in Shoemaker, 1993).

---

1 This survey is also referred to as the CPS CATI/CAPI Overlap Sample Survey (CCO).

The effect of increased centralized computer-assisted telephone interviewing was of particular interest. Findings from the study described in Shoemaker (1993) had shown that including centralized telephone interviews tended to yield a larger unemployment rate. The two-sample McNemar test appeared to be a good vehicle for examining this phenomenon. In both the CPS and the PS, households are interviewed for 4 consecutive months, not interviewed for the next 8 consecutive months, and then interviewed for another 4 consecutive months. The first and fifth interviews are conducted by a personal visit, and the subsequent interviews are conducted by telephone whenever possible. Thus the first and fifth interviews provide a baseline measurement of labor force status; the second and sixth interviews provide a "post-treatment" measurement of labor force status.

To create the panels for both studies, sample within selected sample areas was randomly divided into two representative panels using systematic sampling methods. The treatment panel was designated as CATI eligible. This meant that the sample households in the panel were eligible for interview at a centralized facility after the initial (first and fifth) interviews. To be interviewed by CATI, a respondent must have a telephone and speak English or Spanish, and must agree to be interviewed in subsequent months by telephone. Not all households in this panel were interviewed by CATI. The other panel served as a control.

The monthly unemployment rate is the primary statistic of interest published from Current Population Survey data. This rate is defined as the estimated number of unemployed persons divided by the estimated number of persons in the civilian labor force (the denominator does not include military personnel, persons under sixteen years old, or people who are no longer looking for work, or retired persons). Our primary goal was to understand how including CATI interviews influenced the probability of changing labor force status, in this case from unemployed to not unemployed (or vice versa). The statistics for the one and two-sample McNemar tests use unemployment to population ratios, rather than unemployment rates. This allows for a slightly more precise estimate of the proportion by decreasing the variability of the test statistic.

3.2 Estimates

Each month/panel estimate is an unbiased estimate. That is, the weights used to produce the estimates were strictly a function of the probability of selection: each weight is the product of the baseweight (the inverse probability of selection for a PSU), the weighting control factor (an adjustment for field subsampling), and an adjustment factor for the probability of inclusion in a split panel.

We performed both of the described modifications of the two-sample McNemar test on five sets of PS split panel data. In this case, the final adjustment factor was a constant because of the split panel design. Thus, the final adjustment canceled out in the estimates, although it was included in the variance estimates.

Additionally, we performed the second modification (unlinked data) on thirteen sets of CPS CATI Phase-in data. We did not attempt the first modification for two reasons. First, a longitudinal link between two consecutive months of data proved quite difficult. We were required to match on a set of eight demographic variables and were rarely able to manage more than a seventy percent match rate. Second, the final adjustment factor for the probability of inclusion in a split panel was not necessarily constant, making individual cell estimation complicated.

Variances of levels were computed with generalized variance functions (GVFs). For more details, see Fisher et al (1993). Robert Fay used his VPLX software to calculate replicate estimates of correlation between rotation groups for unemployed and for civilian labor force using September 1992 through December 1993 data from the CPS. We used these correlations for the test statistics based on unlinked data, assuming that they would not differ by survey (CPS versus PS) or by geography (national versus subnational). We derived an expression for the within-panel correlation for civilian population by relating previously calculated autocorrelations (Fisher and McGuinness, 1993) and variance estimates to the individual rotation group estimates. See Appendix B for more details.

We used the unlinked data correlations as a poor approximation for the linked data. Regrettably, we did not have direct replicate estimates of linked data correlation, which we would intuitively expect to be higher than the unlinked. We were also unable to determine a unique relationship between our autocorrelations and our monthly estimates of variance which we could use to obtain linked data correlations. The consequence of this approximation (using unlinked data correlations to approximate linked data) was artificially similar results for the two modifications' applications. In fact, the results from the two modifications of the test using PS data were virtually identical. The estimated standard errors for the linked data are larger in part because we had a smaller sample size for the linked data.

Moreover, there were some unresolved problems with the variance estimator for the linked data. Because we did not have a separate estimator for within-panel correlations for the linked data, we used the estimator for the unlinked data. We suspect this approach tends to underestimate these correlations: the people are matched and don't change. Second, we were unable to estimate the component of variance due to matching error. Since we did not know the relative sizes of these effects, we could not tell the overall effect on the variance estimator for linked data.

3.3 Diagnostics

Small expected sample sizes in individual cells will result in highly variable and consequently unreliable tests. We are not aware of a general method of calculating adequate sample sizes for this type of analysis using complex survey data. Instead, as a naive approach, we used a slightly modified version of the traditional Pearson chi-squared test diagnostic to form a cut-off value as follows:

As defined in Section 2.2.2, let
$x_{+.}$ = unweighted unemployed persons in month 1;
$x_{-.}$ = unweighted not-unemployed persons in month 1;
$x_{.+}$ = unweighted unemployed persons in month 2;
$x_{.-}$ = unweighted not-unemployed persons in month 2.
Recall that in the case of the usual contingency table,under the assumption of independence (and ignoring missing values). In our estimates of expected cell size, we used unlinked marginal data. The sample sizes corresponding to the two months of data are different; we used a geometric mean for our denominator.

A commonly used rule in contingency table analysis is that expected cell sizes should be at least five. However, both the CPS and PS designs are highly clustered, and we felt that the cut-off value should be adjusted upwards accordingly, multiplying by a design effect. Furthermore, we knew that the rows and columns are correlated, so we felt that the minimum cut-off value for expected cell sizes in both panels needed to be further increased. Our final cut-off expected cell size value was ten.

3.4 Results

3.4.1 Parallel Survey Split Panel Study

Parallel Survey data was collected monthly. Unique identifiers were assigned to each sample household. Theoretically, members of the same household could be matched longitudinally from month to month with small error. In general, our match rate from the first month to the second month was close to ninety percent, although we were unable to determine the extent of the false matches included in the set of linked data. This property, along with our inability to directly estimate within panel covariances for the linked data as described in 3.2., led us to omit the linked data results, focusing on the results from the unlinked data. Additionally, small expected cell sizes in the Control panel led us to omit data from several sets of adjacent months from this analysis.

For the reader's convenience, Table One provides summary statistics for the one-sample "monthly" tests for each panel which were based on unlinked data from the PS's split panels. Table Two provides summary statistics for the two-sample tests based on unlinked data.

The reported values of p_1, p_2, p_1', and p_2' are percentages of estimated unemployed to estimated total population for the panel. Recall that p_1 and p_1' are the panel ratio of estimated unemployed from the first and fifth interviews to the estimated panel population from the first and fifth interviews; p_2 and p_2' are the panel ratio of estimated unemployed from the second and sixth interviews to the estimated panel population from the second and sixth interviews Data from the time frame of February 1993 - March 1993 are omitted: a CATI facility was closed during the March interview week because of a blizzard.

Table One: One-Sample McNemar Tests for Individual PS Panels -- Unlinked Data

	Treatment Panel				Control Panel			
Time Frame	p_2-p_1	se(p_2-p_1)	Z-Statistic	P-Value	p_2'-p_1'	se(p_2'-p_1')	Z-Statistic	P-Value
10/92 - 11/92	-0.62	0.29	-2.18	0.03	2.44	0.81	3.02	0.00
11/92 - 12/92	-0.47	0.28	-1.68	0.09	0.11	0.83	0.14	0.89
04/93 - 05/93	-0.76	0.27	-2.84	0.00	0.20	0.72	0.27	0.78
06/93 - 07/93	-0.04	0.27	-0.16	0.88	0.97	0.71	1.38	0.17
08/93 - 09/93	-0.66	0.27	-2.42	0.02	-1.73	0.68	-2.54	0.01

The one-sample McNemar tests above test the probability that the proportion unemployed does not change between the initial and the subsequent interview within the same panel. We use the Control panel to examine the unemployment flux from one month to the next in the absence of CATI. Note that the two significant point estimates are in the opposite direction.

The omnibus hypothesis test (i.e., that $[(p_2'-p_1')_i = 0, i=1...5)$ was significant (p-value=0.00), so we tested the mean of these points. Because we were unable to reject this test (p-value=0.24), we felt that the Control panel one-sample tests did not provide any evidence of a distinct switching trend and did not test any further linear combinations.

We expected a certain amount of rotation group bias to be present in these estimates. In Adams (Bureau of the Census, 1991), the estimates of p_1 constructed from the first and fifth months in sample of the full CPS were roughly six percent larger than their respective second and sixth month in sample analogues (p_2). Consequently, estimates of (p_2 - p_1) calculated from the full CPS data were generally negative. As seen in Table One, this was not the case with the PS Control panel's estimates: counter to our intuition, the estimated difference (p_2'-p_1') is generally positive. This could be a function of the time difference, a geographic difference, or a design difference. Adams used 1987 data from the CPS to calculate national estimates of biases associated with rotation groups. Thus in each of these one-sample tests, the net movements are intertwined with an unmeasured effect from rotation group bias.

Note the negative unemployment flux in the Treatment panel. This observation is substantiated by the significant result from the formal test of the omnibus test (p-value=0.00), and the significant result for the hypothesis $\underline{1}'\underline{\mu}=0$ (p-value=0.00).

The two-sample McNemar test results are presented below.

Table Two: Two-Sample McNemar Tests -- Unlinked PS Data

Time Frame	$(p_2-p_1)-(p_2'-p_1')$	$se[(p_2-p_1)-(p_2'-p_1')]$	Z-Statistic	P-Value
10/92 - 11/92	-3.06	0.86	-3.58	0.00
11/92 - 12/92	-0.58	0.88	-0.66	0.51
04/93 - 05/93	-0.95	0.77	-1.24	0.22
06/93 - 07/93	-1.02	0.76	-1.34	0.18
08/93 - 09/93	1.08	0.74	1.47	0.14

Individually, the monthly results do not demonstrate a clear difference in the unemployment flux between the two panels. On the other hand, the omnibus test (i.e. the test of the hypothesis $((p_2-p_1)-(p_2'-p_1'))_i$, i=1...5) is significant (p-value=0.00). The mean unemployment flux seems to be lower in the treatment panel as evidenced by the significant test results of the hypothesis $\underline{1}'\underline{\mu} = 0$, where $\underline{\mu}$ is the vector of $(p_2-p_1)-(p_2'-p_1'))_i$'s, with each element corresponding to a month's estimate (p-value=0.01).

In these tests, we make statements about contrasts in a table of probabilities, looking for indicators of the effect of a treatment on unemployment movement. As mentioned earlier, some rotation group bias is present in the one-sample tests. The tested hypotheses examine combinations of the net movement within a panel and rotation group bias. This problem is somewhat mitigated in the two-sample tests. Indeed, if rotation group bias is an additive term which affects both panels equally, it will cancel out of the test statistic. Moreover, this effect will be alleviated somewhat in the two-sample test even if it is not the same between the two panels or is multiplicative. A preliminary sensitivity analysis bears this out: we found sensitive one-sample test results and insensitive two-sample tests results.

The two-sample t-tests presented in Thompson (1994) failed to detect a difference by panel in mean unemployment rate using the PS split panel data. This contrasts with the CPS CATI Phase-in results: over two years, the CATI (Treatment) panel had consistently significantly higher unemployment rates than the non-CATI (Control) panel. See Shoemaker (1993). In this analysis of PS split panel data, we have evidence that unemployment is lower in the presence of CATI. There are, however, some problems with the data. First, as previously mentioned, there is some confounding in the Treatment (CATI) panel, since not all respondents in this panel have their second interview conducted from a centralized telephone facility. Second, in each month the expected sample size in the pertinent Control panel cells was near ten, which could be small enough to make the distribution behave unpredictably. This latter problem is not an issue with the CPS CATI Phase-in study analysis presented in 3.4.2.

3.4.2 CPS CATI Phase-in Project Results

The CPS CATI Phase-in project was a continuation of the study presented in Shoemaker (1993). The primary purpose of this study was to measure the effect of including CATI interviewing on the unemployment rate. CATI interviewers in this study used an automated version of the old CPS pencil-and-paper questionnaire, which had a slightly modified version of the lead-in labor force question. More details are provided in Thompson (1994). The data considered in this paper are from the same time period as the PS split panel data examined in 3.4.1: October 1992 through December 1993, again omitting the February 1993 - March 1993 time frame. Expected cell sizes in both the Treatment (CATI) and Control (non-CATI) panels were well over one hundred, and so all other contiguous months of data are included.

The one-sample McNemar test results for both panels are presented below. Test statistics are constructed with unlinked data. The reported values of p_1, p_2, p_1', and p_2' are percentages of estimated unemployed to estimated total population for the panel.

Table Three: One-Sample McNemar Tests for Individual CPS Panels -- Unlinked Data

	Treatment Panel				Control Panel			
Time Frame	p_2-p_1	se(p_2-p_1)	Z-Statistic	P-Value	p_2'-p_1'	se(p_2'-p_1')	Z-Statistic	P-Value
10/92 - 11/92	1.13	0.16	7.63	0.00	0.05	0.47	0.11	0.92
11/92 - 12/92	0.07	0.17	0.44	0.66	-0.14	0.47	-0.30	0.76
12/92 - 01/93	0.43	0.13	3.46	0.00	0.72	0.43	1.68	0.09
01/93 - 02/93	0.00	0.14	0.03	0.97	-0.91	0.43	-2.11	0.03
03/93 - 04/93	-0.25	0.14	-1.81	0.07	-0.16	0.39	-0.40	0.69
04/93 - 05/93	0.63	0.13	4.99	0.00	-0.18	0.43	-0.42	0.67
05/93 - 06/93	0.88	0.13	6.56	0.00	0.47	0.38	1.22	0.22
06/93 - 07/93	0.84	0.13	6.49	0.00	-0.32	0.46	-0.68	0.49
07/93 - 08/93	-0.07	0.14	-0.51	0.61	-0.52	0.39	-1.32	0.19
08/93 - 09/93	0.42	0.13	3.17	0.00	-0.54	0.44	-1.21	0.23
09/93 - 10/93	0.06	0.12	0.52	0.60	-0.08	0.37	-0.22	0.83
10/93 - 11/93	1.05	0.12	8.45	0.00	-0.63	0.42	-1.50	0.13
11/93 - 12/93	0.18	0.14	1.27	0.20	-0.09	0.37	-0.23	0.82

As with the PS split panel data, the one-sample McNemar tests using the CATI Phase-in data test the probability that the proportion unemployed does not change between the initial and the subsequent interview within the same panel. As in 3.4.1, we use the Control panel to estimate the unemployment flux from one month to the next in the absence of CATI. The monthly tests for the Control panel do not appear to exhibit any particular movement. Furthermore, the omnibus hypothesis test was not significant (p-value=0.29), so we did not test any further linear combinations.

Again basing our expectations on the effects of rotation group bias presented in Adams (1991), we believed that the Control panel estimate of p_1' (from the first and fifth months in sample) would be larger than its respective second and sixth month in sample analog, p_2'. On the average, this was the case: although quite variable, the estimates of p_1' are on the average about 4 percent larger than the estimates of p_2'. Because both panels are representative samples from the same parent sample, we assume that the rotation group bias behaves similarly in both panels. The Treatment (CATI) panel estimates of p_2 are **larger** on the average than the estimates of p_1. Given the Control panel's estimates behavior, this phenomenon provides some evidence of a CATI effect.

Note the movement in the Treatment panel from **not** unemployed to unemployed. This observation is substantiated by the significant result from the formal test of the omnibus test (p-value=0.00), and the significant result for the hypothesis $\underline{1}'\underline{\mu}$=0 (p-value=0.00). In contrast to the PS results provided in 3.4.1., this data provides some evidence that unemployment rate is higher in the presence of CATI. This evidence is further substantiated by the two sample McNemar test results provided Table Four below.

Table Four: Two-Sample McNemar Tests -- Unlinked CPS Data

Time Frame	$(p_2-p_1)-(p_2'-p_1')$	$se[(p_2-p_1)-(p_2'-p_1')]$	Z-Statistic	P-Value
10/92 - 11/92	1.18	0.50	2.38	0.02
11/92 - 12/92	0.22	0.50	0.43	0.67
12/92 - 01/93	-0.29	0.45	-0.64	0.52
01/93 - 02/93	0.92	0.45	2.03	0.04
03/93 - 04/93	-0.10	0.42	-0.23	0.81
04/93 - 05/93	0.81	0.45	1.81	0.07
05/93 - 06/93	0.41	0.41	1.01	0.31
06/93 - 07/93	1.16	0.48	2.41	0.02
07/93 - 08/93	0.45	0.42	1.07	0.28
08/93 - 09/93	0.95	0.46	2.06	0.04
09/93 - 10/93	0.14	0.39	0.37	0.71
10/93 - 11/93	1.69	0.44	3.83	0.00
11/93 - 12/93	0.26	0.40	0.66	0.51

The individual monthly results in Table Four provide some evidence of difference in the unemployment flux between two panels. Furthermore, the omnibus test is significant (p-value=0.00). The mean unemployment flux in the Treatment panel seems to be higher as evidenced by the significant test results of the hypothesis $\underline{1}'\underline{\mu}$ = 0.

The two-sample t-tests presented in Thompson (1994) also detected a **positive** difference by panel in mean unemployment rate using the CPS split panel data i.e, including CATI interviews resulted in a **higher** unemployment rate. These results were consistent with the CPS CATI Phase-in results presented in Shoemaker (1993). This analysis of CPS split panel data reinforces that conclusion. Again, it is impossible to attribute the positive net migration from not unemployed to unemployed entirely to the effect of CATI: the same confounding described in 3.4.1. is present in this Treatment (CATI) panel.

3.5 Discussion

The results in 3.4.1 and 3.4.2 yield opposite conclusions about the effect of CATI on unemployment flux (i.e., from not unemployed to unemployed or vice-versa). The CATI effect is not, however, the same in both tests. There are several differences between the two studies.

Perhaps the key difference is the questionnaire. The PS data was collected using the newly redesigned CPS questionnaire. The new questionnaire was designed as an automated instrument. In contrast, the old CPS questionnaire used for the CPS CATI Phase-in Project was designed as a pencil and paper instrument. Field interviewers were required to memorize complicated skip patterns. To minimize respondent burden, CPS interviews generally last about twenty minutes. Using an automated questionnaire, an interviewer can collect more (and more detailed) information in the same amount of time, since she no longer has to determine the path of the interview. The machine does it for her. Besides the automation difference, the wording of the labor force questions differs between the two questionnaires.

PS interviews were conducted using the same questionnaire both in the field interviews (using a laptop computer) or from the CATI facility. In contrast, the CPS CATI Phase-in interviews used two different versions of the old questionnaire: a paper version for the field interviews; and an automated version, with a slightly modified lead-in labor force question for the CATI interviews.

Given these questionnaire differences, and the caveats about the PS split panel data, it would be unwise to draw any clear conclusion about the effect of CATI alone from these two studies. Instead, we would recommend continuing to examine this effect by using two-sample McNemar techniques on the new CPS split panel data, which uses the old CATI Phase-in design and the redesigned, fully automated questionnaire.

4. Conclusion

We have presented two modifications of the two-sample McNemar test using complex survey data, with applications from the unlinked data modification. If the survey does not have a longitudinal design, then the application using the linked data will have an unknown variance/covariance structure and will include a variance component due to matching error. In this case, using the unlinked data makes sense with respect to the model's interpretation, although the statistic based on the (unlinked) estimates of marginal probabilities may be inferior to a well-developed linked model. If the survey has a longitudinal design, then the first method may be preferred, as it is a straight-forward extension of the traditional test, and consequently, the interpretation is equivalent to the textbook interpretation.

Areas for future research include investigations into the power of these tests in the context of complex sample data, variance/covariance estimation for linked data including matching error variance contributions, and the difference in efficiency in the two approaches. In data analytical applications, McNemar and two-sample McNemar tests seem to have uses in comparing aspects of different survey methods or effects on responses within a method over time. The approach is nonparametric in its conception; when the approximation is good, it avoids pitfalls that may be associated with model-based tests.

References

Adams, David E. (1991) "CPS Month in Sample (MIS) Bias Index Research," internal memorandum, Demographic Statistical Methods Division, Bureau of the Census, Washington, D.C.

Cochran, William G., *Sampling Techniques,* John Wiley and Sons, New York, 1977.

Feuer, Eric J. and Kessler, Larry G. (1989), Test Statistic and Sample Size for a Two-Sample McNemar Test, *Biometrics* ***45***, 629-636.

Fisher, Robin; Robison, Edwin; Thompson, Jenny; Welch, Michael (1993), Variance Estimation in the CPS Overlap Test, *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Fisher, Robin and McGuinness, Richard (1993), "Correlations and Adjustment Factors for CPS," internal memorandum, Demographic Statistical Methods Division, Bureau of the Census, Washington, D.C.

McNemar, Q. (1947), Note on the Sampling Error of the Differences Between Correlated Proportions of Percentages. *Psychometrika* ***12***, 153-157.

Marascuilo, Leonard A.; Omelich, Carol L.; Gokhale, D.V. (1988), Planned and Post Hoc Methods for Multiple-Sample McNemar (1947) Tests With Missing Data. *Psychological Bulletin* *103*, 238-245.

Rao, J.N.K. and Wu, C.C.J. (1988), Resampling Inference with Complex Survey Data, *Journal of the American Statistical Association 83*, 231-241.

Shoemaker, Harland H. (1993), Results from the Current Population Survey CATI Phase-in Project, *Proceedings of the Section on Survey Research Methods*, American Statistical Association.

Thompson, Jenny (1994), Mode Effects Analysis of Labor Force Estimates, *CPS Overlap Analysis Team Technical Report 3*, Bureau of the Census, Washington, D.C.

Wolter, Kirk M., *Introduction to Variance Estimation*, Springer-Verlag, New York, 1985.

Appendix A

General Construction of the Test Statistic

Let $a = x_{+.}$ = number of observed +'s in first month
b = sample size for first month
$c = x_{.+}$ = number of observed +'s in second month
d = sample size for second month

A zero subscript denotes the expected value for a, b, c, and d.

A = the set containing the a +'s
B = the set containing the b persons in the first month's population
C = the set containing the c +'s
D = the set containing the d persons in the second month's population

Then $$\hat{p}_1 = \frac{a}{b},\ \hat{p}_2 = \frac{c}{d},\ and\ Var(\hat{p}_2 - \hat{p}_1) = Var(\frac{a}{b}) + Var(\frac{c}{d}) - 2Cov(\frac{a}{b},\frac{c}{d}).$$

An approximate expression for the variance of a ratio of random variables is given by the Taylor Series expansion e.g., Cochran (1977).

Install Equation Editor and double-click here to view equation.

The same form of Taylor expansion can be used to obtain an approximate expression for the covariance of two ratios. We have

Install Equation Editor and double-click here to view equation.

Install Equation Editor and double-click here to view equation.

A common approximation which is exact for simple random sampling follows.

Install Equation Editor and double-click here to view equation.

The premise is approximately true in the unlinked situation (modification two) and is exactly true in the linked situation (modification one). If we conservatively assume Install Equation Editor and double-click here to view equation. the expression for the covariances becomes

Install Equation Editor and double-click here to view equation. (2)

Note that if B=D (as in the linked version of the test statistic), then the covariance expression becomes

Install Equation Editor and double-click here to view equation. (3)

For the linked data modification, $(p_2 - p_1)$ is estimated by Install Equation Editor and double-click here to view equation.
where X_{-+}, X_{+-}, and $N_{(12)}$ are weighted estimates, and Var$(p_2 - p_1)$ is estimated by

Install Equation Editor and double-click here to view equation. (4)

For the unlinked data modification $(p_2 - p_1)$ is estimated by Install Equation Editor and double-click here to view equation. where $X_{.+}$, $X_{+.}$, N_1, and N_2 are weighted estimates (note that the estimates of $X_{.+}$ and $X_{+.}$ are usually different than the estimates used for the first modification), and Var$(p_2 - p_1)$ is estimated by

Install Equation Editor and double-click here to view equation. (5)

If we accept the assumption that Install Equation Editor and double-click here to view equation. , then the covariance expression can be readily approximated by Install Equation Editor and double-click here to view equation.

Details about the calculation of the variances and covariances for our particular application are described in Section 4.2 and in Appendix B. We adapted generalized variance functions from Fisher et al (1993) for our variances for levels, but there are several other viable options for complex survey data such as balanced half-sample replication (e.g., Wolter, 1985) or some bootstrap estimators designed for complex surveys (e.g., Rao and Wu, 1988).

Appendix B

In this appendix we discuss the derivation of the covariance term in equation (5), Cov($X_{+\cdot}$, $X_{\cdot+}$). Consider the unlinked data.

Install Equation Editor and double-click here to view equation. (A1)

where

$X_{i,j}$ is a weighted sample level for month i, Month in sample (MIS) j

Note that $X_{1,j}$ and $X_{2,j+1}$ are from the same rotation group unless j=4 since a rotation group is out of sample for eight months after being in for four. We assumed that the correlations between $X_{i,j}$ and $X_{k,m}$ can be decomposed into three separate categories:

1) Install Equation Editor and double-click here to view equation.

This is a within rotation group correlation.

2) Install Equation Editor and double-click here to view equation.

This is a within month between rotation group correlation.

3) Install Equation Editor and double-click here to view equation.

This is a between rotation group between month correlation.

The covariance in (A1) becomes

Install Equation Editor and double-click here to view equation. (A2)

using the simplifying assumption that Var($X_{i,j}$) is constant for all i and j. The variance for a full month's estimate, Install Equation Editor and double-click here to view equation., is available in the form of a generalized variance function (GVF). We use this estimate to calculate Install Equation Editor and double-click here to view equation. by applying the following derivation:

Install Equation Editor and double-click here to view equation. (A3)