12-2013

A Guide to Testing a Proportion when there may be Misclassifications

David L. Farnsworth
Rochester Institute of Technology

Jonathan R. Bradley
Rochester Institute of Technology

Math. Scientist **38,** 135–144 (2013)
Printed in England

A GUIDE TO TESTING A PROPORTION WHEN THERE MAY BE MISCLASSIFICATIONS

JONATHAN R. BRADLEY,* *The Ohio State University*
DAVID L. FARNSWORTH,** *Rochester Institute of Technology*

Abstract

Ignoring possible misclassifications when testing for a proportion can lead to erroneous decisions. A statistical test is described that incorporates misclassification rates into the analysis. Easily checked safeguards that ensure that the test is appropriate are given. Additionally, the test provides a procedure when the hypothesis stipulates that the proportion is zero. Applications of the test are illustrated with examples which show that it is practical. Comprehensive guidance is supplied for the practitioner.

Keywords: Errors in measurement; false-positive rate; false-negative rate; hypothesis test; misclassification rate; misdiagnosis rate; survey question

2010 Mathematics Subject Classification: Primary 62F03
Secondary 62D05; 62P25

1. Introduction

We are interested in performing a hypothesis test about a population's proportion, π, which has a certain characteristic. For example, in a survey or questionnaire, π may be the proportion that would answer *yes* to a certain yes–no question if the individuals were truthful. In a medical setting, the fraction of the population with a certain disease or condition may be of interest. The relevant statistic is the reported proportion with the characteristic in a sample of fixed size.

Without misclassifications, this situation is very simple and is presented in undergraduate textbooks. It is a binomial experiment for a large homogeneous population and random sampling. The random variable that is the sample proportion is a complete sufficient statistic, the maximum likelihood estimator, and the minimum variance unbiased estimator of π [10, pp. 323, 387, 394], [16, pp. 642–643, 657–659, 671–684].

However, in the real world, errors are made. People respond incorrectly to questions. There may be imperfections in the question causing wrong answers, giving a truthful answer may cause embarrassment, or incorrect answers can be caused by forgetfulness of the individuals sampled. There could be transcription errors. In medical testing, there are false-positive and false-negative diagnoses. There will always be some mistakes. The implications of the misuse of survey data or medical diagnoses, for example, when there are potential mistakes or errors, can be detrimental.

Because of concerns for errors in measurement, each discipline's literature concerning ways to reduce errors is vast. Most of it is concerned with research designs for reducing errors, measuring the effects of errors, and estimating errors. Those designs include using callbacks or

Received 27 February 2013; revision received 29 March 2013.
* Current address: Department of Statistics, University of Missouri, Columbia, MO 65211, USA.
Email address: bradleyjr@missouri.edu
** Postal address: School of Mathematical Sciences, Rochester Institute of Technology, Rochester, NY 14623, USA.
Email address: dlfsma@rit.edu

repeated measurements, a gold standard, a panel of experts, randomized responses for sensitive questions [21], and unrelated questions [11]. For surveys see [2], [3, pp. 359–399], [14], [17], and [20].

We address the error rates and accommodate them in the hypothesis test for a single proportion. The sample's proportion estimates the expected proportion when there are errors. The goal is to incorporate the errors into the analysis in order to test for the true value, π. This is a binomial experiment that is contaminated by misclassifications. We present an introduction to testing for a proportion in an environment of misclassifications and offer guidance for performing tests. The frequentist perspective is used. Extensive guidance from the Bayesian point of view, in which all proportions have prior distributions, can be found in [6] and [9]. For the most part, we speak in terms of a survey question, but there are other applications, such as in medical testing.

2. Error rates

Let T be 0 or 1 to denote the true answer to the question for a randomly selected individual. That answer, which is not observed, is either *yes* ($T = 1$) or *no* ($T = 0$). The probability $\pi = \mathbb{P}(T = 1)$ is the proportion of individuals whose correct answer would be *yes*. Let R be 0 or 1 to denote an individual's response to the question with $R = 1$ indicating *yes* and $R = 0$ indicating *no*. The probability that the individual will respond *yes* is $p = \mathbb{P}(R = 1)$.

Let the false-positive rate be $p_1 = \mathbb{P}(R = 1 \mid T = 0)$ and the false-negative rate be $p_2 = \mathbb{P}(R = 0 \mid T = 1)$. These rates may have single point estimates, imprecise estimates, or probability distributions. All three situations are addressed in Sections 5 and 6, below.

Since

$$\begin{aligned} p &= \mathbb{P}(R = 1) \\ &= \mathbb{P}(R = 1 \text{ and } T = 1) + \mathbb{P}(R = 1 \text{ and } T = 0) \\ &= \mathbb{P}(R = 1 \mid T = 1)\mathbb{P}(T = 1) + \mathbb{P}(R = 1 \mid T = 0)\mathbb{P}(T = 0) \\ &= (1 - p_2)\pi + p_1(1 - \pi) \\ &= (1 - p_1 - p_2)\pi + p_1, \end{aligned} \tag{1}$$

the probabilities π and p are linearly related for each p_1 and p_2. The inequality

$$1 - p_1 - p_2 > 0 \tag{2}$$

ensures that p increases with increasing π, that p decreases with decreasing π, and that $p > 0$. Because p_1 and p_2 are error rates, this inequality is a mild demand. If $1 - p_1 - p_2 < 0$, then the identifications of *yes* and *no* can be interchanged so that the new complementary error rates satisfy (2), see [8, p. 707]. We assume that (2) is true.

The probability of a *yes* response, p, can be less than, equal to, or greater than the probability, π, that a correct or truthful answer would be *yes*, that is $p < \pi$, $p = \pi$, or $p > \pi$, depending upon the error rates and value of π. For example, using (1) and holding π at 0.20: $p_1 = 0.01$ and $p_2 = 0.10$ give $p = 0.188 < \pi$, $p_1 = 0.02$ and $p_2 = 0.08$ give $p = 0.20 = \pi$, and $p_1 = 0.10$ and $p_2 = 0.01$ give $p = 0.278 > \pi$.

3. Random variables

There are two Bernoulli random variables, T and R. The variable T has parameter π. The variable R has parameter p. Simple random sampling is performed from a large and homogeneous population. The predetermined sample size is n, and X is the count with the characteristic in the sample.

The expected value $\mathbb{E}(T)$ is π and $\mathrm{var}(T) = \pi(1-\pi)$, since T is a Bernoulli variable. Similarly, $\mathbb{E}(R) = p$ and $\mathrm{var}(R) = p(1-p)$. The covariance is

$$\begin{aligned}\mathrm{cov}(T, R) &= \mathbb{E}((T - \mathbb{E}(T))(R - \mathbb{E}(R))) \\ &= \mathbb{E}(TR) - \mathbb{E}(T)\mathbb{E}(R) \\ &= 1 \cdot 1 \cdot \mathbb{P}(T = 1 \text{ and } R = 1) - \pi p \\ &= \mathbb{P}(R = 1 \mid T = 1)\mathbb{P}(T = 1) - \pi p \\ &= (1 - p_1 - p_2)\pi(1-\pi),\end{aligned}$$

and the correlation coefficient is

$$\rho = \mathrm{corr}(T, R) = \frac{\mathrm{cov}(T, R)}{\sqrt{\mathrm{var}(T)\,\mathrm{var}(R)}} = (1 - p_1 - p_2)\sqrt{\frac{\pi(1-\pi)}{p(1-p)}}.$$

As the error rates p_1 and p_2 approach zero, the correlation coefficient approaches 1. Often, ρ^2 is called the coefficient of reliability [18, p. 1356].

It is sensible to require that there be a positive correlation between the correct response and the offered or recorded response [22, p. 858]. Since $\mathrm{corr}(T, R) > 0$ only if $1 - p_1 - p_2 > 0$, (2) appears again.

4. Estimators

From (1), $p_1 < p < 1 - p_2$, since $0 < \pi < 1$. The maximum likelihood estimator (MLE) of p is

$$\hat{p} = \begin{cases} p_1, & \text{if } \frac{X}{n} \le p_1, \\ \frac{X}{n}, & \text{if } p_1 < \frac{X}{n} < 1 - p_2, \\ 1 - p_2, & \text{if } 1 - p_2 \le \frac{X}{n}, \end{cases} \tag{3}$$

[10, p. 316], [16, pp. 720–722]. Since the MLEs $\hat{p}$ and $\hat{\pi}$ are also related by (1) (see [10, p. 316] and [16, p. 680]), $\hat{p} = (1 - p_1 - p_2)\hat{\pi} + p_1$, where $\hat{\pi}$ is the MLE of π. So,

$$\hat{\pi} = \frac{\hat{p} - p_1}{1 - p_1 - p_2} \tag{4}$$

and

$$\hat{\pi} = \begin{cases} 0, & \text{if } \frac{X}{n} \le p_1, \\ \frac{X/n - p_1}{1 - p_1 - p_2}, & \text{if } p_1 < \frac{X}{n} < 1 - p_2, \\ 1, & \text{if } 1 - p_2 \le \frac{X}{n}. \end{cases} \tag{5}$$

Since we assume (2), $\hat{\pi}$ is well defined.

The estimators of p and π in (3) and (5) on a boundary would indicate that the analysis has broken down. Assuming that the model is correct, we should check that the probabilities are very small and that the statistic X/n will have a value off-scale, which would cause the estimates to be on a boundary. The probability that $\hat{\pi} = 0$ and $\hat{\pi} = 1$ for (5) can be computed, given numerical values for π, p_1, p_2, and n. Assuming that n is sufficiently large, so that X/n is approximately normally distributed with mean p and variance $p(1-p)/n$, we have

$$\mathbb{P}(\hat{\pi} = 0) = \mathbb{P}(\hat{p} = p_1) = \mathbb{P}\left(\frac{X}{n} \le p_1\right) \approx \mathbb{P}\left(Z \le \frac{p_1 - p}{\sqrt{p(1-p)/n}}\right) \tag{6}$$

and

$$\mathbb{P}(\hat{\pi} = 1) = \mathbb{P}(\hat{p} = 1 - p_2) = \mathbb{P}\left(\frac{X}{n} \ge 1 - p_2\right) \approx \mathbb{P}\left(Z \ge \frac{(1-p_2) - p}{\sqrt{p(1-p)/n}}\right), \tag{7}$$

with p given by (1) and Z a standard normal variable. For example, for $\pi = 0.06$, $p_1 = p_2 = 0.15$, and $n = 75$, we obtain $p = 0.192$ and

$$\mathbb{P}(\hat{\pi} = 0) = \mathbb{P}(\hat{p} = p_1) = \mathbb{P}\left(\frac{X}{n} \le 0.15\right) \approx \mathbb{P}(Z \le -0.92) \approx 0.18.$$

However, for $n = 600$ and the same values for π, p_1, and p_2, $\mathbb{P}(\hat{\pi} = 0) = \mathbb{P}(\hat{p} = p_1)$ is only 0.004. Subsequently, we assume that the sample size is sufficiently large, to ensure that $p_1 < X/n < 1 - p_2$, so that the estimators are not on the boundaries.

5. The test

The null hypothesis is $\mathrm{H}_0\colon \pi = \pi_0$, and the alternative hypothesis can be one-sided or two-sided. For specificity, we take $\mathrm{H}_\mathrm{A}\colon \pi > \pi_0$. The test statistic is X/n. For large sample sizes, so that the normal distribution approximation is appropriate and the probabilities in (6) and (7) are negligible,

$$\mathbb{E}(\hat{\pi}) = \frac{\mathbb{E}(X/n) - p_1}{1 - p_1 - p_2} = \frac{p - p_1}{1 - p_1 - p_2} = \pi, \tag{8}$$

from (1) and (5). Also,

$$\operatorname{var}(\hat{\pi}) = \frac{\operatorname{var}(X/n)}{(1 - p_1 - p_2)^2} = \frac{p(1-p)/n}{(1 - p_1 - p_2)^2}. \tag{9}$$

The $\mathbb{P}$-value is

$$\begin{aligned}\mathbb{P}\left(Z \ge \frac{\pi_{\text{data}} - \mathbb{E}(\hat{\pi}_0)}{\sqrt{\operatorname{var}(\hat{\pi}_0)}}\right) &= \mathbb{P}\left(Z \ge \left(\frac{X/n - p_1}{1 - p_1 - p_2} - \frac{p_0 - p_1}{1 - p_1 - p_2}\right) \Big/ \sqrt{\frac{p_0(1-p_0)/n}{(1 - p_1 - p_2)^2}}\right) \\ &= \mathbb{P}\left(Z \ge \frac{X/n - p_0}{\sqrt{p_0(1-p_0)/n}}\right),\end{aligned} \tag{10}$$

where

$$p_0 = (1 - p_1 - p_2)\pi_0 + p_1. \tag{11}$$

Equation (8) implies that $\hat{\pi}$ is an unbiased estimator of the true proportion π. Generally, X/n is a biased estimator of π, since its bias is $\mathbb{E}(X/n) - \pi = p - \pi = p_1(1-\pi) - p_2\pi$ from (1), see [1, p. 481]. The bias is zero only in the unlikely event that the probability that an element will be incorrectly classified with the feature equals the probability that an element will be incorrectly classified without the feature, that is

$$p_1(1-\pi) = \mathbb{P}(R=1 \mid T=0)\mathbb{P}(T=0) = \mathbb{P}(T=0 \text{ and } R=1)$$

must be equal to

$$p_2\pi = \mathbb{P}(R=0 \mid T=1)\mathbb{P}(T=1) = \mathbb{P}(T=1 \text{ and } R=0).$$

Ignoring the error rates would give a test based on the biased estimator X/n for π. This bias can be very large, even for reasonable error rates. For example, for $\pi = 0.15$ and $p_1 = p_2 = 0.08$, the bias is $0.08 \times 0.85 - 0.08 \times 0.15 = 0.056$ and the relative bias is

$$\frac{0.15 - 0.056}{0.15} 100\% = 62.7\%.$$

Since $\mathbb{E}(X/n) = p = (1 - p_1 - p_2)\pi + p_1$ from (1), X/n is an unbiased estimator of the probability p that an individual will respond *yes*.

Equation (9) implies that $\hat{\pi}$ is consistent, since the variance approaches zero as n increases [10, pp. 203–207], [16, pp. 195–197]. Using some algebra, we can show that

$$\operatorname{var}(\hat{\pi}) = \frac{\pi(1-\pi)}{n} + \frac{1}{(1-p_1-p_2)^2}\left(\frac{p_2(1-p_2)}{n}\pi + \frac{p_1(1-p_1)}{n}(1-\pi)\right).$$

The first term is the pure sampling variance, and the second term contains a weighted average of the variances from the two types of measurement errors. Since $\operatorname{var}(\hat{p}) = (1-p_1-p_2)^2 \operatorname{var}(\hat{\pi})$ from (4), there is a similar expression for $\operatorname{var}(\hat{p})$. These relationships can be used for addressing the relative costs of the two errors.

Since the usual binomial test for a proportion is employed, the normal approximation and formulas for power and sample size are known [10, pp. 265–267], [16, pp. 216–220, 609–611].

Example 1. As an illustration, consider a test in which the true proportion of *yes* responses in a survey is $\pi = 0.25$ against the one-sided alternative $\pi > 0.25$, and take the misclassification rates to be $p_1 = p_2 = 0.05$. The expected proportion of *yes* responses in the sample is $p_0 = (1 - 0.05 - 0.05)0.25 + 0.05 = 0.275$. In a sample of size $n = 400$, there were 123 respondents who answered *yes*. The normal distribution approximation is appropriate, since $np_0 > 10$ and $n(1 - p_0) > 10$. The probabilities that $\hat{\pi} = 0$ and $\hat{\pi} = 1$ in (6) and (7) are negligible, since

$$\begin{aligned}\mathbb{P}(\hat{\pi}=0) &= \mathbb{P}\left(\frac{X}{n} \le 0.05\right)\\ &= \mathbb{P}\left(Z \le \frac{0.05-0.275}{\sqrt{0.275(1-0.275)/400}}\right)\\ &\approx \mathbb{P}(Z \le -10.08)\\ &\approx 0\end{aligned}$$

and

$$\mathbb{P}(\hat{\pi}=1) = \mathbb{P}\left(\frac{X}{n} \ge 1 - 0.05\right) \approx \mathbb{P}(Z \ge 30.23) \approx 0.$$

From (10), we obtain

$$\mathbb{P}\text{-value} = \mathbb{P}\left(Z \geq \frac{\frac{123}{400} - 0.275}{\sqrt{0.275(1-0.275)/400}}\right) \approx \mathbb{P}(Z \geq 1.46) \approx 0.072.$$

The exact $\mathbb{P}$-value, obtained without the normal distribution approximation, has the same value since $\sum_{i=123}^{400} \binom{400}{i} 0.275^i (1-0.275)^{400-i} = 0.072$.

However, a much different $\mathbb{P}$-value would be obtained if we had mistakenly ignored the possibility of misclassifications. The $\mathbb{P}$-value would then be only

$$\mathbb{P}\left(Z \geq \frac{\frac{123}{400} - 0.25}{\sqrt{0.25(1-0.25)/400}}\right) \approx \mathbb{P}(Z \geq 2.66) \approx 0.004.$$

6. Estimating and using the error rates

Since $\partial p/\partial p_1 = 1 - \pi$ and $\partial p/\partial p_2 = -\pi$ from (1) for fixed π, the probability, p, of a *yes* response is especially sensitive to the false-positive rate, p_1, when π is small and to the false-negative rate, p_2, when π is large. This sensitivity influences the hypothesis test. The decision in the test can be greatly influenced by the error rates that are used in the analysis, especially if π_0 is near 0 or 1. Determining p_1 and p_2 carefully is critical for a meaningful test.

6.1. Point estimates of the error rates

There are many methods for assigning values to p_1 and p_2. These include past experience with the question, information from the provider of the question, and various reliability exercises.

A range of values can be used for the error rates. Values for the error rates within the suggested intervals are tried and the combination that gives the largest $\mathbb{P}$-value is used, in order to be statistically conservative.

Another method is to carry out two separate simple experiments [20, pp. 924–925]. A random sample of elements whose true state is $T = 0$ is taken and the fraction that responds $R = 1$ is used as an estimator of $p_1 = \mathbb{P}(R = 1 \mid T = 0)$. Similarly, a sample of elements whose true state is $T = 1$ is taken and the fraction that responds $R = 0$ is the estimate of $p_2 = \mathbb{P}(R = 0 \mid T = 1)$. This has the advantages of simplicity and of being binomial experiments with well-known statistical properties. The estimators are random variables, which should be part of the analysis. Knowing T for some individuals is called having a gold standard. This might be almost impossible to obtain, too expensive, or may not exist. Tenenbein [18], [19] proposed a somewhat more complex scheme of double sampling part of a sample, based on a gold standard. There are more complicated procedures, such as the use of a panel of experts [3], [14].

6.2. Using a probability distribution of the error rates

Another method employs a compounding or mixing distribution [7], [10, p. 191], [15]. This might be appropriate in cases in which the error rates have distributions, perhaps reflecting an inhomogeneity of the testing process or instrument. When the error rates p_1 and p_2 are independent, the $\mathbb{P}$-value is

$$\int_{x/n}^{1} \left(\int_{\mathrm{D}(p_2)} \int_{\mathrm{D}(p_1)} \frac{1}{\sqrt{2\pi p_0(1-p_0)/n}} \exp\left(-\frac{1}{2}\frac{(t/n - p_0)^2}{p_0(1-p_0)/n}\right) f_1(p_1) f_2(p_2)\, \mathrm{d}p_1\, \mathrm{d}p_2 \right) \mathrm{d}t. \tag{12}$$

In the outer integration's limit, x is the number of *yes* responses in the sample of size n. From (11), $p_0 = (1 - p_1 - p_2)\pi_0 + p_1$. The approximating normal density that appears in the integral is considered a conditional distribution of X (or of the dummy variable t in the integral), depending on p_1 and p_2, which have distributions of their own. The mixing or weighing distributions of p_1 and p_2 are $f_1(p_1)$ and $f_2(p_2)$ with domains $\mathrm{D}(p_1)$ and $\mathrm{D}(p_2)$. The result of the integration is independent of the order of integration with respect to p_1 and p_2, when their probability density functions are continuous [4, p. 239]. Of course, care must be taken to ensure that integrations are performed only where $0 < p_0 < 1$. Generally, the integration requires numerical methods or a numerical computing software package such as MATHEMATICA®, MATLAB®, or MAPLE®.

One choice of distribution for the error rates is the uniform distribution on subintervals of the unit interval. The probability density function is zero outside the subintervals.

One of the most commonly used distributions for rates is the standard beta distribution with

$$f_i(p_i) = \frac{\Gamma(\alpha_i + \beta_i)}{\Gamma(\alpha_i)\Gamma(\beta_i)} p_i^{\alpha_i - 1}(1 - p_i)^{\beta_i - 1}, \qquad 0 < p_i < 1,\ \alpha_i > 0,\ \beta_i > 0,\ \text{for } i = 1, 2,$$

[10, p. 192], [15, p. 161]. One way to assign values to the parameters α_i and β_i is the method of moments [5, pp. 243–245], [16, pp. 664–669] in which the sample mean, $\bar{p}_i$, and the sample variance, $s^2_{p_i}$, of historical values of p_i are set to be equal to the corresponding moments of the beta distributions,

$$\mu_{p_i} = \frac{\alpha_i}{\alpha_i + \beta_i} \quad \text{and} \quad \sigma^2_{p_i} = \frac{\alpha_i \beta_i}{(\alpha_i + \beta_i)(\alpha_i + \beta_i - 1)}.$$

Solving for α_i and β_i gives

$$\bar{p}_i + \frac{\bar{p}_i^2(1 - \bar{p}_i)}{s^2_{p_i}} \quad \text{and} \quad 1 - \bar{p}_i + \frac{\bar{p}_i(1 - \bar{p}_i)^2}{s^2_{p_i}},$$

respectively. Hogg *et al.* [10, p. 588] discussed other criteria for selecting the values of the parameters. As in Example 1, the probabilities in (6) and (7) should be checked. If the support of p_1 or p_2 is a proper subset of the unit interval, then nonstandard beta distributions might be required [5, pp. 167–168], [12, pp. 210–275].

When the joint distribution $f_{12}(p_1, p_2)$ of p_1 and p_2 over the domain $\mathrm{D}(p_1, p_2)$ is used, the $\mathbb{P}$-value can be expressed as

$$\int_{x/n}^{1} \left(\iint_{D(p_1,p_2)} \frac{1}{\sqrt{2\pi p_0(1 - p_0)/n}} \exp\left(-\frac{1}{2} \frac{(t/n - p_0)^2}{p_0(1 - p_0)/n} \right) f_{12}(p_1, p_2)\, \mathrm{d}p_1\, \mathrm{d}p_2 \right) \mathrm{d}t. \tag{13}$$

The Dirichlet distribution is often used as a joint distribution for two proportions [8, p. 707], [13, pp. 485–491, 504–511]. Its probability density function is

$$f_{12}(p_1, p_2) = \frac{\Gamma(\alpha_1 + \alpha_2 + \alpha_3)}{\Gamma(\alpha_1)\Gamma(\alpha_2)\Gamma(\alpha_3)} p_1^{\alpha_1 - 1} p_2^{\alpha_2 - 1} (1 - p_1 - p_2)^{\alpha_3 - 1},$$

with $\alpha_i > 0$ (for $i = 1, 2, 3$), $0 < p_1 < 1$, $0 < p_2 < 1$, and $1 - p_1 - p_2 > 0$. The last inequality is (2), which is automatically part of this distribution. This is a bivariate generalization of the univariate beta distribution. Its mean and variance have relatively simple formulas, and the marginal probability density functions are beta distributions [13, pp. 489, 504]. If a discrete distribution of p_1 and p_2 is available, then the inner double integral in (13) becomes a sum.

We close this section with a numerical example.

Example 2. We test $H_0 : \pi = 0.25$ that the proportion of the population who represent a truthful *yes* is 0.25 against the one-sided alterative hypothesis $H_A : \pi > 0.25$. We use a sample of 600 respondents. The sample's count of those who say *yes* is $x = 185$, so $x/n = \frac{185}{600} = 0.308$.

We believe that p_1 is equally likely to be anywhere between 0.03 and 0.08, so that $f_1(p_1) = 1/(0.08 - 0.03) = 20$ over the domain $D(p_1)$, which is $0.03 < p_1 < 0.08$, and $f_1(p_1) = 0$ otherwise, that is, p_1 has a uniform distribution. Similarly, p_2 is equally likely to be anywhere between 0.02 and 0.12, so that $f_2(p_2) = 1/(0.12 - 0.02) = 10$ over the domain $D(p_2)$, which is $0.02 < p_2 < 0.12$, and $f_2(p_2) = 0$ otherwise. From (12), the $\mathbb{P}$-value is

$$\int_{185/600}^{1} \left(\int_{p_2=0.02}^{p_2=0.12} \int_{p_1=0.03}^{p_1=0.08} \frac{1}{\sqrt{2\pi p_0(1-p_0)/n}} \times \exp\left(-\frac{1}{2} \frac{(t/n - p_0)^2}{p_0(1-p_0)/n} \right) 20 \cdot 10 \, \mathrm{d}p_1 \, \mathrm{d}p_2 \right) \mathrm{d}t. \tag{14}$$

Before evaluating the $\mathbb{P}$-value in (14), we should check that the normal distribution approximation is appropriate and that the probabilities in (6) and (7) are negligible. Since $p_0 > (1 - 0.08 - 0.12)0.25 + 0.03 = 0.23$, $np_0 > 600 \times 0.23 = 138 > 10$, while $p_0 < (1 - 0.02 - 0.03)0.25 + 0.08 = 0.3175$, $n(1 - p_0) > 600(1 - 0.3175) = 409.5 > 10$, the normal approximation may be used.

In the denominators of (6) and (7), use the identity $p(1-p) = (1-p_1-p_2)^2\pi(1-\pi) + p_2(1-p_2)\pi + p_1(1-p_1)(1-\pi)$, which is derived using (1). Then,

$$\begin{aligned}
&\mathbb{P}(\hat{\pi} = 0) \\
&\quad = \mathbb{P}\left(Z \le \frac{-(1-p_1-p_2)0.25}{\sqrt{\frac{1}{600}((1-p_1-p_2)^2 0.25 \cdot 0.75 + p_2(1-p_2)0.25 + p_1(1-p_1)0.75)}} \right) \\
&\quad = \mathbb{P}\left(Z \le \frac{-0.25\sqrt{600}}{\sqrt{0.25 \cdot 0.75 + (p_2(1-p_2)0.25 + p_1(1-p_1)0.75)/(1-p_1-p_2)^2}} \right) \\
&\quad \le \mathbb{P}\left(Z \le \frac{-0.25\sqrt{600}}{\sqrt{0.25 \cdot 0.75 + 0.0904}} \right) \\
&\quad \approx \mathbb{P}(Z \le -11.62) \\
&\quad \approx 0,
\end{aligned}$$

since

$$\begin{aligned}
\frac{p_2(1-p_2)0.25 + p_1(1-p_1)0.75}{(1-p_1-p_2)^2} &< \frac{0.12(1-0.12)0.25 + 0.08(1-0.08)0.75}{(1-0.02-0.03)^2} \\
&\approx 0.0904.
\end{aligned}$$

Similarly, $\mathbb{P}(\hat{\pi} = 1) \approx \mathbb{P}(Z \ge 34.85) \approx 0$.

Using MATHEMATICA, the $\mathbb{P}$-value is

$$\int_{0.308333}^{1}\int_{0.02}^{0.12}\int_{0.03}^{0.08}\frac{1}{\sqrt{2\pi}}\frac{\sqrt{600}}{\sqrt{((1-y-z)0.25+y)(1-(1-y-z)0.25-y)}}$$
$$\times\exp\left(-\frac{1}{2}\frac{(x-(1-y-z)0.25-y)^2}{((1-y-z)0.25+y)(1-(1-y-z)0.25-y)/600}\right)$$
$$\times 20\cdot 10\,\mathrm{d}y\,\mathrm{d}z\,\mathrm{d}x$$
$$\approx 0.260.$$

Ignoring the possibility of misclassifications, the $\mathbb{P}$-value would be

$$\mathbb{P}\left(Z \geq \frac{0.308-0.25}{\sqrt{(0.25\cdot 0.75)/600}}\right)=\mathbb{P}(Z\geq 3.28)=0.0005.$$

As in Example 1, failing to recognize and accommodate the misclassification rates gives a smaller $\mathbb{P}$-value and a subsequent false determination of a significant difference from the hypothesized value.

7. Hypothesized zero proportion

A particularly intriguing feature of this test is that it can be used for testing the null hypothesis that a probability of a *yes* is zero, that is, $\pi_0 = 0$. If the false-positive error rate, p_1, were zero, then $p_0 = 0$ under the null hypothesis and the binomial test and its normal distribution approximation fail to be applicable, because the test statistics would have a zero in their denominators. Indeed, if there is no chance of an error, this would be deterministic, not statistical, and one *yes* response would lead to rejection of the hypothesis.

However, if $\pi_0 = 0$ and the false-positive rate is nonzero, then $p_0 = p_1 > 0$, and the test described in Section 5 can be employed to test for a zero probability in an environment of misclassifications. In order to reject the hypothesis that the probability π is zero, the count x in the sample would have to exceed a positive number that is determined by the false-positive rate, the sample size, and the level of significance. Although the true probability is assumed to be zero, the probability of obtaining some *yes* responses is not zero.

References

[1] Bross, I. (1954). Misclassification in 2 × 2 tables. *Biometrics* **10,** 478–486.

[2] Cochran, W. G. (1968). Errors in measurement in statistics. *Technometrics* **10,** 637–666.

[3] Cochran, W. G. (1977). *Sampling Techniques*, 3rd edn. John Wiley, New York.

[4] Courant, R. (1961). *Differential and Integral Calculus*, Vol. 2. John Wiley, New York.

[5] Devore, J. L. (2008). *Probability and Statistics for Engineering and the Sciences*, 7th edn. Duxbury, Belmont, CA.

[6] Evans, M., Guttman, I., Haitovsky, Y. and Swartz, T. (1996). Bayesian analysis of binary data subject to misclassification. In *Bayesian Analysis in Statistics and Economics: Essays in Honor of Arnold Zellner*, eds D. A. Berry *et al*., John Wiley, New York, pp. 67–77.

[7] Everitt, B. S. and Hand, D. J. (1981). *Finite Mixture Distributions*. Chapman & Hall, London.

[8] Fujisawa, H. and Izumi, S. (2000). Inference about misclassification probabilities from repeated binary responses. *Biometrics* **56,** 706–711.

[9] Gaba, A. and Winkler, R. L. (1992). Implications of errors in survey data: a Bayesian model. *Manag. Sci.* **18,** 913–925.

[10] Hogg, R. V., McKean, J. W. and Craig, A. T. (2005). *Introduction to Mathematical Statistics*, 6th edn. Prentice Hall, Upper Saddle River, NJ.

[11] Horvitz, D. G., Shah, B. V. and Simmons, W. R. (1967). The unrelated question randomized response model. In *Proceedings of the Social Statistics Section*, American Statistical Association, Alexandria, VA, pp. 65–72.
[12] Johnson, N. L., Kotz, S. and Balakrishnan, N. (1995). *Continuous Univariate Distributions*, Vol. 2, 2nd edn. John Wiley, New York.
[13] Kotz, S., Balakrishnan, N. and Johnson, N. L. (2000). *Continuous Multivariate Distributions*, Vol. 1, *Models and Applications*, 2nd edn. John Wiley, New York.
[14] Lyberg, L. *et al.* (eds) (1997). *Survey Measurement and Process Quality*. John Wiley, New York.
[15] McLachlan, G. and Peel, D. (2000). *Finite Mixture Models*. John Wiley, New York.
[16] Rohatgi, V. K. (2003). *Statistical Inference*. Dover, Mineola, NY.
[17] Scheaffer, R. I., Mendenhall, W. and Ott, R. L. (2005). *Elementary Survey Sampling*, 6th edn. Duxbury, Belmont, CA.
[18] Tenenbein, A. (1970). A double sampling scheme for estimating from binomial data with misclassifications. *J. Amer. Statist. Assoc.* **65,** 1350–1361.
[19] Tenenbein, A. (1971). A double sampling scheme for estimating from binomial data with misclassifications: sample size determination. *Biometrics* **27,** 935–944.
[20] Walter, S. D. and Irwig, L. M. (1988). Estimation of test error rates, disease prevalence and relative risk from misclassified data: a review. *J. Clin. Epidemiol.* **41,** 923–937.
[21] Warner, S. L. (1965). Randomized response: a survey technique for eliminating evasive answer bias. *J. Amer. Statist. Assoc.* **60,** 63–69.
[22] Zelen, M. and Haitovsky, Y. (1991). Testing hypotheses with binary data subject to misclassification errors: analysis and experimental design. *Biometrika* **78,** 857–865.