Accuracy in Medical Claims: Estimation of Overpayment Amounts and Accuracy Rates. An Example of Use of the Ratio Estimator to Improve Precision of Estimators.

Information taken from a Report to the Dept of Health and Human Services, State of North Carolina.

Kenneth H. Pollock, North Carolina State University.

1. The Estimation Method

1.1 Background

The federal and state government agencies responsible for reimbursing health care providers (Doctors, Hospitals, Home Care Providers, Nursing Homes and many others) are concerned about the potential for the providers to over bill for their services. This over billing could be due to lack of awareness of government regulations or in a few cases due to deliberate attempts to defraud the agencies for profit. Sometimes millions of dollars of over billing are involved and criminal proceeding may be initiated if this can be proved beyond a reasonable doubt.

The agencies have designed sampling plans to verify that the providers have not over billed the agencies. This involves drawing a probability sample from the complete computer list of all possible claims. Then the sample claims are "verified". In some cases the verification involves just checking in detail all the paperwork submitted to support the bill presented for that claim whereas in other cases a medical specialist is sent out on site to do a very detailed check of each claim in the sample. The verification can be very expensive so statisticians are involved to see the sampling plans are valid and that the most efficient estimators are used.

1.2 Definitions and Guidelines

Sampling Unit

The sampling unit will be an individual medical claim for the survey period under review.

Sample

The sample will be the set of individual claims for billed services for the survey period under review that are actually checked for overpayment or under payment.

Universe (Population)

The universe or population under study consists of the total set of individual claims for billed services of the class of services being reviewed for a period of time. The universe is often divided into **strata** based on different types of claims. (We do not get into stratification in this simple presentation). The universe or frame of claims is a computer file so that it is very easy to draw probability samples from the frame.

Measured Variable (Payment Error of a Claim)

The measured variable is the amount of payment error on a sample claim and that is given by
y_i = Amount billed and actually paid – amount that should have been billed and paid

1.3 Estimation Equations for Simple Random Sampling

These are the standard estimation equations for a quantitative variable using simple random sampling where the sampling is without replacement. The source is a good sampling theory book such as Cochran (1978) or Thompson (2002). I have basically followed the Cochran (1978) notation which is slightly different from that in your textbook Thompson (2002) because I adapted this document from one I prepared for them..

Some Definitions

N is the total number of units in the population.
n is the total number of units in the sample.
y_i is the error in payment amount of unit i. It is calculated as the difference between what was paid on a claim minus what should have been paid on a claim. It is assumed measured without error.
x_i is the actual paid amount of unit i. (We would suspect that y is linearly related to x with no intercept, that is regression through the origin. If the amount paid approaches zero then the amount over billed must also approach zero. This is not rocket science here!).
$\bar{y}$ is the mean error in payment amount calculated for the sample.
s^2 is the variance in the error in payment amount variable calculated for the sample.
Y is the population total error in payment amount. This is unknown and one parameter we are trying to estimate.
X is the population total payment amount. This is a known value.
$E = \frac{Y}{X}$ is the error in payment amount expressed as a rate and another parameter we are trying to estimate. (This is the ratio R in Thompson’s notation). This is sometimes of more interest than Y. Note that the related accuracy rate is
A = 1-E.

c is the number of units(claims) in error in the sample
C is the number of units in error in the Universe
E_c is the error rate based on number of units in error divided by the number of units in the Universe.

$$E_c = \frac{C}{N}$$

1.3.1 Difference or Simple Expansion Estimator

Estimate of Absolute Error in Payment Amount in the Population

A simple unbiased but inefficient estimate of Y (the total amount overpaid for the universe of claims) which does not use the relationship between y and x is:

$$\hat{Y}_d = N\bar{y}$$

The estimated variance and standard error of the point estimate of error in payment amount in the population is

$$\mathrm{Var}(\hat{Y}_d) = N(N-n)s^2/n$$

$$\mathrm{SE}(\hat{Y}_d) = \sqrt{\{N(N-n)s^2/n\}}\,.$$

The confidence interval estimates of the error in payment amount in the population is

$$\hat{Y}_d \ \pm\ t\ \ x\ \ \mathrm{SE}(\hat{Y}_d),$$

where t is the value from the students t table (with (n-1) degrees of freedom) that gives the appropriate confidence level. If n is large normal distribution values can be used.

Estimate of Rate of Error in Payment Amount in the Population

This estimator is

$$\hat{E}_d = \frac{\hat{Y}_d}{X} = \frac{N\bar{y}}{X}$$

Standard error and confidence interval results follow easily because X is known and because

$$\mathrm{Var}(\hat{E}_d) = \mathrm{Var}(\hat{Y}_d)/X^2\,.$$

$$\hat{\mathrm{Var}}(\hat{E}_d) = N(N-n)s^2/X^2 n$$

$$\hat{SE}(\hat{E}_d) = \sqrt{\mathrm{Var}(\hat{E}_d)} \qquad (1)$$

1.3.2 Ratio Estimator

The corresponding ratio estimate for the absolute error in payment amount which does use the relationship between y and x is

$$\hat{Y}_r = \frac{\bar{y}}{\bar{x}} X \ ,$$

and for the error rate in payment amount is

$$\hat{E}_r = \frac{\bar{y}}{\bar{x}} \, .$$

These are the standard ratio estimators and are approximately unbiased in large samples. The variance and standard error estimates and confidence interval estimates are given in standard sampling theory books.

$$Var(\hat{Y}_r) = N(N-n)\frac{s_r^{\ 2}}{n}$$

$$S\hat{E}(\hat{Y}_r) = \sqrt{V\hat{a}r(\hat{Y}_r)}$$

$$V\hat{a}r(\hat{E}_r) = \frac{N(N-n)}{X^2}\frac{s_r^{\ 2}}{n} \qquad (2)$$

$$S\hat{E}(\hat{E}_r) = \sqrt{V\hat{a}r(\hat{E}_r)}$$

with

$$s_r^{\ 2} = \sum_1^n (y_i - \hat{E}_r x_i)^2 /(n-1)$$

Clearly these estimators will be more efficient than the difference estimators presented earlier because they exploit the relationship between y and x.

As noted earlier, in practice stratified random sampling is often used and there are extensions of the ratio estimators to this sampling design.

1.3.3 Error Rate Estimator based on Claims

$$\hat{E}_c = \frac{c}{n}$$

This is the standard estimation problem for a population proportion and therefore

$$S\hat{E}(\hat{E}_c) = \sqrt{(\frac{N-n}{N})[\frac{\hat{E}_c(1-\hat{E}_c)}{n}]} \qquad (3)$$

$$\hat{\mathrm{E}}_\mathrm{c} \ \pm \ \mathrm{z} \ \ \mathrm{x} \ \ \mathrm{SE}(\hat{\mathrm{E}}_\mathrm{c})$$