On Stratification in Medicare Investigations

Don Edwards

Dept. of Statistics
University of South Carolina
edwards@stat.sc.edu

October 14, 2009
13th Annual Medicare / Medicaid Statistics and Data Analysis Conference
Omaha, NE

Acknowledgments

This research would not have been possible without the support of Palmetto GBA, Inc. and Tricenturion, Inc. I am especially grateful for advice and feedback received from:

- Gail Ward-Besser, Palmetto GBA
- Petko Kostadinov, TriCenturion
- Jennifer Lasecki, Palmetto GBA

Outline

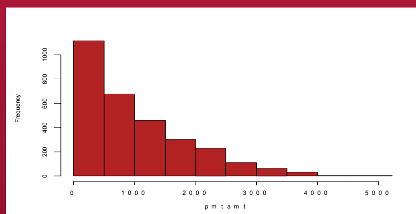
- Three Payment Populations
- What Are We Doing?
- In God We Trust – the rest of us need to have data
- Stratification by Payment Amount
- Summary
- Appendix I: *samptest*
- Appendix II: the Minimum Sum Method

Three Payment Populations

1. Pediatric Services

Shape: Right skew, no separation from 0

N = 3000, total pmt = $3.1M

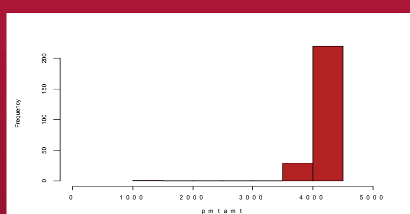


Three Payment Populations

2. Power Wheelchairs

Shape: Similar pmt amts, separation from 0

N = 250, total pmt = $1.0M

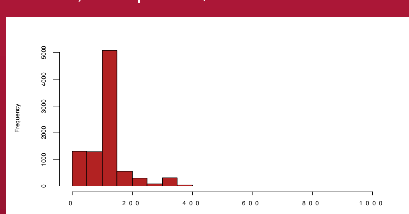


Three Payment Populations

3. Home Health

Shape: in-between examples 1 and 2

N = 9000, total pmt = $1.1M



What Are We Doing?

We are sampling overpayments - the sampled objects are payments, but the observation in each case is an overpayment.

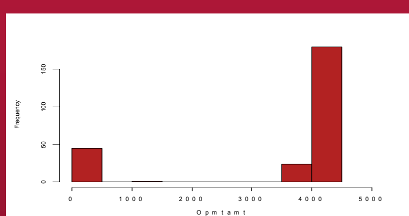
(a) The operating characteristics of the sampling and extrapolation plan depends on the distribution of the overpayments.

(b) Any planning calculations made using payment amounts are potentially very misleading unless the denial rate is 100%

(c) We do not know the overpayment distribution, but we can infer a lot about its plausible shape from the pmt distribution.

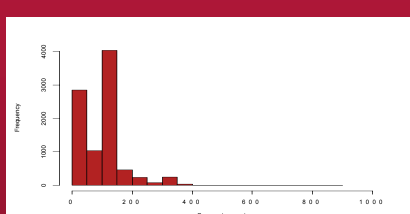
Wheelchair example: one Plausible Overpayment Population

Compare to slide 5. This assumes an 80% denial rate, denials evenly distributed among payments



Home Health example: one Plausible Overpayment Population

Compare to slide 6. This assumes an 80% denial rate, denials evenly distributed among payments



What Are We Doing?

We are not *estimating* the total universe overpayment. We are finding a 90% lower confidence bound for the total overpayment.

(a) Regardless of the true nature of the overpayment population, the sampling and extrapolation plan should underrecoup 90% of the time.

(b) Subject to (a) and available resources, we seek to recover as much of the universe overpayment (small or large) as possible.

(c) All else equal, it's better to recoup 90% from each of 4 providers than 95% from 1 provider.

In God We Trust

THE REST OF US NEED TO HAVE DATA

We do not know the overpayment distribution, but we can infer a lot about its plausible shape from the payment distribution.

- If the payment distribution is right-skewed and not separated from 0, the overpayment distribution will almost certainly be right-skewed.
- If the payments are similar in size and separated from 0, under a high denial rate the overpayment distribution will almost certainly be bimodal and left-skewed.

In God We Trust

THE REST OF US NEED TO HAVE DATA

Cochran (≈ God) 3rd ed. p.41:

"when we sample from positively skew populations…the frequency with which $\bar{y} - (1.96)s_{\bar{y}}$ *is greater than* μ *is* less *than 2.5%…"*

Attempted translation by Edwards (*not* God):

When we sample from a right-skewed overpayment population, the actual underrecoupment rate of a Central Limit Theorem-based "90% lower confidence bound" is greater than 90%.

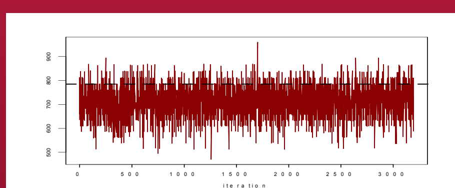
In God We Trust
THE REST OF US NEED TO HAVE DATA

The *bad news* is that vice-versa holds:

When we sample from a left-*skewed overpayment population, the actual underrecoupment rate of a Central Limit Theorem-Based "90% lower confidence bound" is* less *than 90%.*

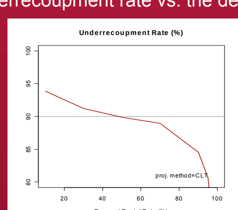
In God We Trust
THE REST OF US NEED TO HAVE DATA

Don't take my word for it. We can easily sample repeatedly from the overpayment population shown on slide 7, and discover that the "90% lower bound" only works 86.5% of the time when n = 45.



In God We Trust
THE REST OF US NEED TO HAVE DATA

Often this problem gets worse with higher denial rates. Repeat the process just described for denial rates between 0.1 and 1.0 (wheelchair pop). A plot of the underrecoupment rate vs. the denial rate:



In God We Trust

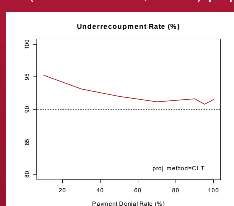
THE REST OF US NEED TO HAVE DATA

- An even more serious problem: the CLT lower bound can be greater than the total amount paid, rendering it completely indefensible as a recoupment demand. The chance of this happening approaches 40% for the wheelchair population when the denial rate is close to 1.
- Using a lower bound based on the ratio estimator fixes the above problem, but not the problem with the underrecoupment rate being << 90%.
- For any population with payment amounts similar and separated from 0, use the minimum-sum (Edwards et al 2005) extrapolation method.

In God We Trust

THE REST OF US NEED TO HAVE DATA

Don't abandon CLT methods for all populations. It is conservative when the overpayment population is right-skewed, e.g. for the pediatric services and home health (shown below, n=45) populations.



In God We Trust

THE REST OF US NEED TO HAVE DATA

BOTTOM LINE: In this day and age, there is no reason to use a sampling-and-extrapolation plan without testing it thoroughly.

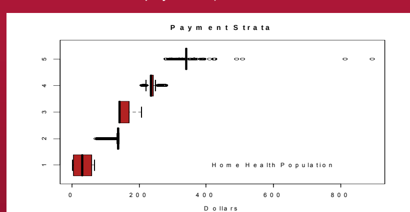
TEST YOUR SAMPLING PLAN BEFORE YOU USE IT

Stratified Sampling by Payment Amounts

Scheaffer et al (≈ God), 4th ed., pp 98-99:
"Stratification may produce a smaller bound on the error of estimation than would be produced by a simple random sample of the same size. This result is particularly true if the measurements within strata are homogeneous."

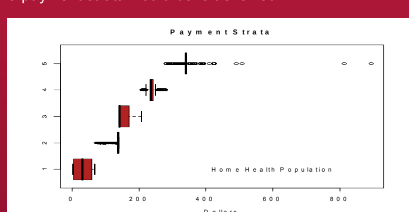
Stratified Sampling by Payment Amounts

If we stratify the payment population by payment amounts, will we achieve homogeneous strata, which will lead to a smaller bound on the error of estimation (i.e. tighter lower bounds for total overpayment?)



Stratified Sampling by Payment Amounts

NOT NECESSARILY for the overpayments! Imagine each of these strata with a 90% denial rate – each would extend from 0 to their largest payment amounts. Moreover, several of these overpayment strata would be left-skewed.



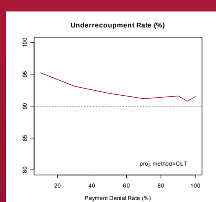
Stratified Sampling by Payment Amounts

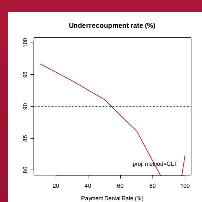
In short, if you stratify by payment amount,

YOU'RE STRATIFYING THE WRONG POPULATION

Stratified Sampling by Payment Amounts

A comparison for the Home Health population (#3): at left are the results for a simple random sample of size 45, at right for a stratified sample (5 strata, n_i=9 each).

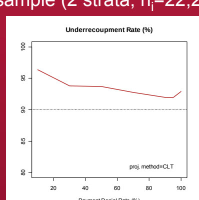


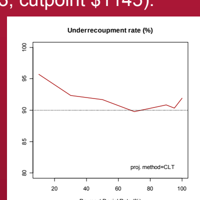


Stratified Sampling by Payment Amounts

In the right conditions, careful stratification can help.

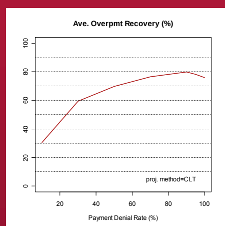
For the pediatric services population (#1; pmts right skewed, no separation from 0): at left are results for a simple random sample n=45, at right a stratified sample (2 strata, n_i=22,23, cutpoint $1145).

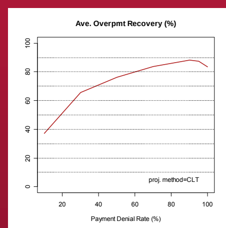




Stratified Sampling by Payment Amounts

Pediatric services population: Average overpayment recovery. At left: simple random sample of n=45, at right: stratified sample (2 strata,n=22,23 cutpt $1145).





Summary

- Lower Confidence Bounds using the Central Limit Theorem are usually conservative if the overpayment population is right-skewed, liberal if it is left-skewed.
- Careless stratification can destroy the confidence level of an otherwise perfectly adequate simple random sample, rendering it invalid for use.
- *Careful* stratification can sometimes yield a valid procedure with higher overpayment recovery than a simple random sample of the same size – the best candidates are payment populations which are right-skewed and not separated from 0.

Summary

UNLESS YOU ARE
DIVINELY OMNISCIENT,
TEST YOUR
SAMPLING PLAN
BEFORE YOU USE IT

REFERENCES

1. Cochran, W.G. (1977). *Sampling Techniques*, 3rd edition. New York: John Wiley and Sons.

2. Edwards, Don; Ward-Besser, Gail; Lasecki, Jennifer; Parker, Brenda; Wieduwilt, Kristin; Wu, Fuming; and Moorhead, Philip (2003). "The Minimum Sum Method: a Distribution-Free Sampling Procedure for Medicare Fraud Investigations." *Health Services and Outcomes Research Methodology* **4**: 241-263. (published 2005)

3. Scheaffer, R.E., Mendenhall, W., and Ott, L. (1990). *Elementary Survey Sampling*, 4th edition. Boston: PWS-Kent.

Appendix I: *samptest*

R programs for Medicare studies

functions included (* = used in this talk):

- *simpic**
- *SRStest* (CLT and Minimum Sum)*
- *get.LD* (formerly get.NeL)*
- *StRStest**
- *StRS.cutpts*; StRS.allocate*
- *Dblesamptest*
- *project*

- These programs are free – send me an email
- Developed under support of Palmetto GBA and TriCenturion
- You need to get *R*: *www.R-project.org* (also free)

Appendix II:
The Minimum Sum Method

New notation:

N = # of population pmts

D = unknown # of denied pop. Pmts

n = sample size (Simple RS)

d = # of denied sample payments

The Minimum Sum Method

Ref: Edwards *et al*, 2003 (appeared 2005).

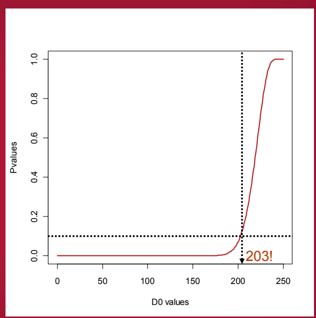
Step 1: Calculate L_D, a 90% lower confidence bound for D by inverting the level-0.10 test for H_0: $D \leq D_0$ vs. H_A: $D > D_0$ based on the hypergeometric distribution.

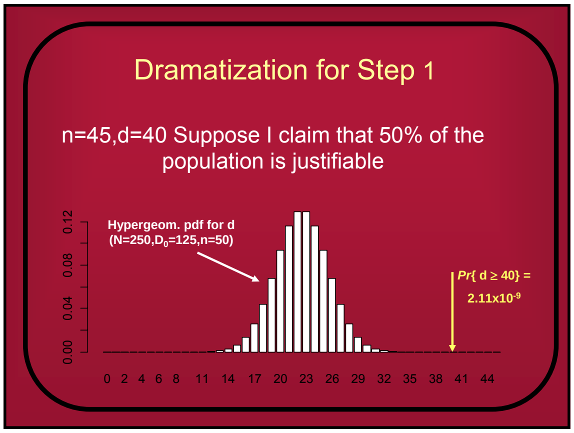
The Minimum Sum Method

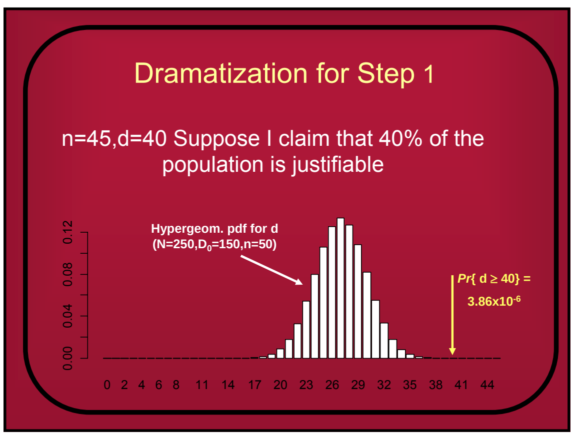
Example: Suppose a simple random sample of n=45 payments from the wheelchair population shows d=40 denied payments. Use *get.LD* in *samptest* or do the following directly in *R*:

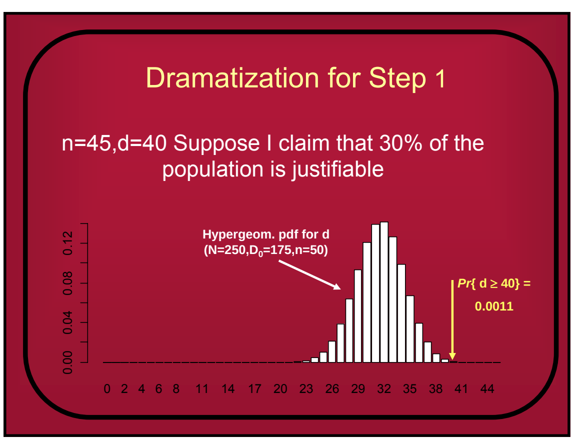
```
> N = 250
> n = 45
> d = 40
> Pvalues = 1-phyper( d-1, 0:N, N:0, n)
> min((0:N)[ Pvalues >= 0.10 ])
# [1] 203
```

Red line:
P-values for
H_0: $D \leq D_0$ vs. H_A: $D > D_0$
for D_0 = 0,1,2,...,N = 250
given n=45,d=40
P = *Pr{d≥40}*



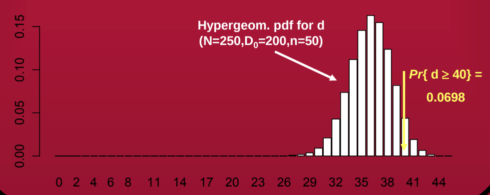

Dramatization for Step 1
n=45,d=40 Suppose I claim that 50% of the population is justifiable
Hypergeom. pdf for d
(N=250,D0=125,n=50)
Pr{ d ≥ 40} =
2.11x10-9
0.00 0.04 0.08 0.12
0 2 4 6 8 11 14 17 20 23 26 29 32 35 38 41 44


Dramatization for Step 1
n=45,d=40 Suppose I claim that 40% of the population is justifiable
Hypergeom. pdf for d
(N=250,D0=150,n=50)
Pr{ d ≥ 40} =
3.86x10-6
0.00 0.04 0.08 0.12
0 2 4 6 8 11 14 17 20 23 26 29 32 35 38 41 44


Dramatization for Step 1
n=45,d=40 Suppose I claim that 30% of the population is justifiable
Hypergeom. pdf for d
(N=250,D0=175,n=50)
Pr{ d ≥ 40} =
0.0011
0.00 0.04 0.08 0.12
0 2 4 6 8 11 14 17 20 23 26 29 32 35 38 41 44

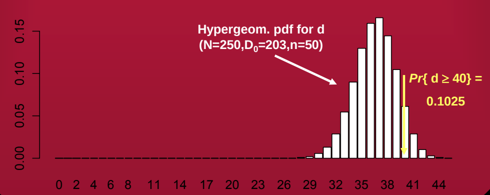
Dramatization for Step 1

n=45,d=40 Suppose I claim that 20% of the population is justifiable



Dramatization for Step 1

n=45,d=40 Suppose I claim that 18.8% of the population is justifiable



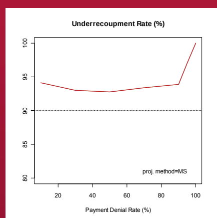
The Minimum Sum Method

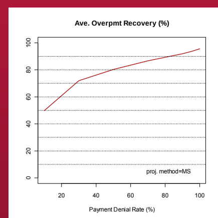
Step 2: Considering the results of Step 1, the MinSum recoupment demand is the sum of all sample overpayments plus the sum of the *smallest* remaining unsampled payments which would be in error.

Example: For the wheelchair population example, there would be at least 203 denied population payments. The minimum sum extrapolation is thus:
(40 sample overpayments) +
(the smallest 163 non-sampled payments)

The Minimum Sum Method

Operating Characteristics for the Wheelchair payment population, n=45





The Minimum Sum Method

The MinSum method has some good properties:

- Because Step 1 works for any population size and any sample size, the method is mathematically guaranteed to under-recoup in at least 90% of repeated samples.
- The minimum sum extrapolation *cannot* be greater than the total payment amount

The Minimum Sum Method

- The MinSum method recoups very well when payments are all nearly equal OR if the error rate is very high.
- However, because of the added conservatism of Step 2, for some populations the MinSum method will be *too* conservative to be useful...e.g. it would not recoup well for our example populations 1 or 3 unless the denial rate ≈ 95% or more.

The Minimum Sum Method

Operating Characteristics for the Pediatric Services payment population, n=45