Estimating confidence intervals for the median hospital charge

Cody Olsen, Amy Donaldson, Andrea Thomas, Larry Cook

University of Utah School of Medicine, Intermountain Injury Control Research Center

Presenter Disclosures

Cody Olsen

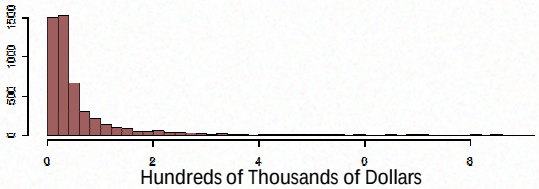
(1) The following personal financial relationships with commercial interests relevant to this presentation existed during the past 12 months:

No relationships to disclose

INTRODUCTION TO THE PROBLEM

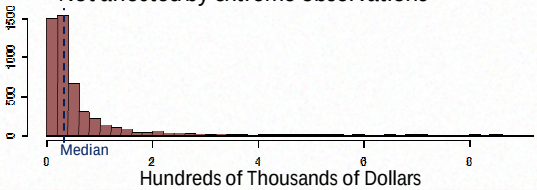
Hospital charges

- Charges are a useful outcome in injury research
- Charges are typically skewed



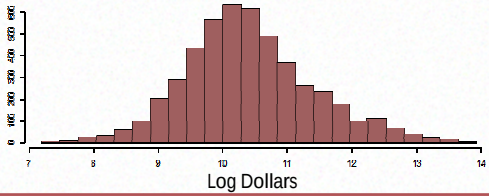
Median Charges

- The sample median is often used to describe charges
  - 50th percentile, or middle observation
  - Not affected by extreme observations



Hospital charges

- Transformed charges are often analyzed



Complication: Missing Data

- Administrative hospital datasets often have missing data
- This can be handled using multiple imputation
  - Fill-in missing data with plausible values determined by a statistical model
  - Results in multiple imputed datasets, each with different sets of imputed values
  - Combine results from multiple datasets for final result (Rubin, D.B. 1987)

A confidence interval for the Median

A confidence interval for the median can be obtained by:

- Log transformation of the data
- Quantile regression: A non-parametric method combined with bootstrapping

Methodological Assumptions

- Transformation Method
  - Assumes that the distribution of charges is normal when transformed
  - Uses the mean and a 95% confidence interval for the mean of the transformed distribution, then back-transform
- Quantile regression
  - No distributional assumptions
  - Uses the sample median and bootstrapping to obtain 95% confidence interval

Study Objectives

- Using multiple imputed charges, compare confidence intervals from:
  - Transformation method
  - Quantile regression
- Simulate data from a known distribution
- Apply the methods to an administrative dataset: motor vehicle crash related injuries

SIMULATIONS

Simulation Methods

Using a known population:

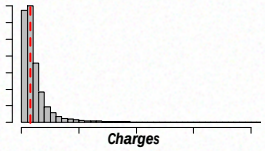
- Sample n=2000
- Simulate 10% missing data
- Impute missing data: 5 imputed datasets
- Estimate confidence intervals for the median
- Repeat 100 times
- Compare Mean Square Error, Bias, Coverage, Mean Width of the Confidence Interval

Simulated Populations

1. Population which is normal when log transformed

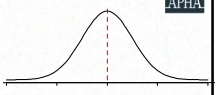
Median = $15,636, Variance of log(charges)= 0.88

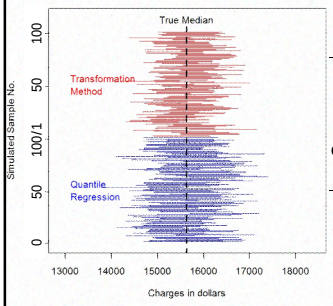
2. Empirical distribution based on observed injury-related charges
   - Utah hospital discharge database (n=112k)
   - Adjusted for inflation



Median = $15,636, Variance of log(charges) = 0.88

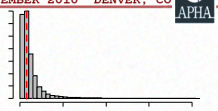
Simulation 1: Normal

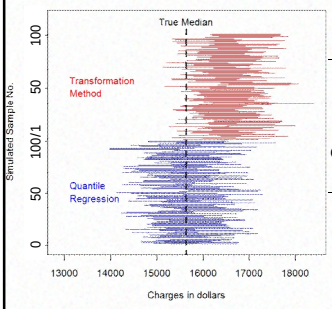




	Transformation Method	Quantile Regression
MSE	105,990	120,401
Bias	$24	$32
Coverage	95%	96%
Width	$1,268	$1,656

Simulation 2: Empirical

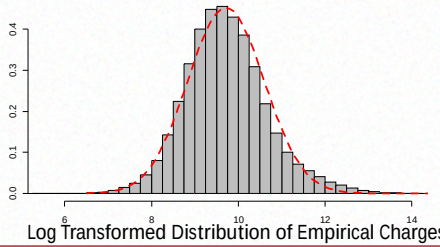




	Transformation Method	Quantile Regression
MSE	944,115	149,574
Bias	$900	$54
Coverage	30%	98%
Width	$1,431	$1,649

Simulation 2: Empirical hospital charges

Skewness persists after log transformation:

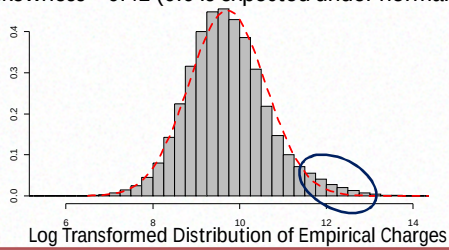


Log Transformed Distribution of Empirical Charges

Simulation 2: Empirical hospital charges

Skewness persists after log transformation:

- Skewness = 0.41 (0.0 is expected under normality)



Log Transformed Distribution of Empirical Charges

APPLICATION:

MOTOR VEHICLE CRASH RELATED INJURIES

CODES

Crash Outcome Data Evaluation System

- Uses medical outcomes related to motor vehicle crashes (MVC) for highway safety and injury research
- Emergency Department Charges
- Inpatient Hospital Charges
- Multiple imputation accounts for missing data

Motor Vehicle Crash related injuries

- 3 years of inflation-adjusted charges
- Utah dataset
  - Emergency Department (n = 53,950 × 5 imputations)
  - Hospital Inpatient (n = 4,827 × 5 imputations)
- Large multi-state dataset
  - 18 States included
  - Emergency Department (n = 1.37M × 5 imputations)
  - Hospital Inpatient (n = 182,398 × 5 imputations)

Computational Efficiency

	Time to compute	
Dataset Size	Transformation Method	Quantile Regression
n = 4,827 x 5	1 second	4 seconds
n = 1.37M x 5	17 seconds	1.74 hours

- Intel® Core™2 CPU 6600 @ 2.40 GHz, 2.0 GB of RAM
- SAS Procedures: Means (transformation method); Quantreg, MIAnalyze (quantile regression)

Motor Vehicle Crash related injuries

Emergency Department Charges: Median (95% CI)

	Transformation Method	Quantile Regression
Utah	$763 (757-770)	$705 (698-712)
Multi-State	$859 (857-860)	$826 (825-828)

Inpatient Charges: Median (95% CI)

	Transformation Method	Quantile Regression
Utah	$32,785 ($31,873-33,723)	$30,138 ($29,267-31,009)
Multi-State	$20,859 ($20,724-20,996)	$19,450 ($19,283-19,617)

DISCUSSION

Transformation Method

- Pros
  - Performed well under normality: narrow intervals with good coverage
  - Computationally efficient
- Cons
  - Provided poor coverage in empirical situation
  - Biased: over-estimated the true median
  - Interval estimate of a mean, which may not be equal to the median

Quantile Regression Method

- Pros
  - Unbiased estimate with good coverage in both simulations
  - Estimates the median regardless of the shape of the distribution
- Cons
  - Required more computing time
  - Wider confidence intervals under normality

Conclusions

- Quantile regression is an unbiased method to obtain confidence intervals for median hospital charges
- Carefully check normality assumptions when analyzing log transformed charges with standard methods
- Quantile regression can be used with multiply imputed charges

Future directions

Compare transformation methods to quantile regression for:

- Estimating differences between group medians
- Estimating regression coefficients in multivariable models of the median

References

- Rubin, D.B. (1987), *Multiple Imputation for Nonresponse in Surveys*, New York: John Wiley & Sons, Inc.
- Koenker, R., Hallock, K.F. (2001), Quantile regression. The Journal of Economic Perspectives 15 (4), 143-156
- Laurent AG (1963), Journal of the American Statistical Association 58, 231-235

Acknowledgments

- Thanks to National Highway Traffic Safety Administration (NHTSA)
  - The CODES Technical Resource Center at the University of Utah operates as a cooperative agreement with NHTSA to assist states with probabilistic linkage, analytical issues, and assist NHTSA with multistate studies

Cody Olsen
cody.olsen@hsc.utah.edu
801 – 581 – 5755
295 Chipeta Way
PO Box 581289
Salt Lake City, UT 84158-0289

QUESTIONS