

This work is licensed under a [Creative Commons Attribution-NonCommercial-ShareAlike License](https://creativecommons.org/licenses/by-nc-sa/4.0/). Your use of this material constitutes acceptance of that license and the conditions of use of materials on this site.



Copyright 2009, The Johns Hopkins University and John McGready. All rights reserved. Use of these materials permitted only in accordance with license rights granted. Materials provided "AS IS"; no representations or warranties provided. User assumes all responsibility for use, and all liability related thereto, and must independently review all materials for accuracy and efficacy. May contain materials owned by others. User is responsible for obtaining permissions for use from third parties as needed.



JOHNS HOPKINS  
BLOOMBERG  
SCHOOL *of* PUBLIC HEALTH

## Lecture 3d: Practice Problem Solutions

---

John McGready  
Johns Hopkins University

# Estimating a 95% Confidence Interval

1. A healthcare information company is interested in estimating the average charge for a standard patient visit to a chiropractor in Maryland, after applying the discount negotiated with a large HMO plan. Data is collected from 16 randomly selected chiropractic practices in Maryland, and the following are some summary statistics:
  - Mean charge: 25.50 USD
  - SD of charges: 2.10 USD

# Estimating a 95% Confidence Interval

1. Assuming the charge data is normally distributed for all chiropractic practices in Maryland, estimate a range of amounts that most (95%) of the chiropractic practices in Maryland charge for a standard patient visit
  - If the sample data comes from a distribution of normally distributed values, then we can estimate this interval from the sample results by using  $\bar{x} \pm 2 \times s$ .
  - In this example, applying this logic gives an interval of
  - $25.50 \pm 2 \times 2.10 \rightarrow 25.50 \pm 4.20 \rightarrow (\$21.30, \$29.70)$ .

## Estimating a 95% Confidence Interval

2. Without assuming normality, estimate a range of amounts that most (95%) of the chiropractic practices in Maryland charge for a standard patient visit
  - This is a trick question: given only the sample mean and standard deviation, it is not possible to answer this question

## Estimating a 95% Confidence Interval

3. Assuming the charge data is normally distributed for all chiropractic practices in Maryland, estimate a 95% confidence interval for the mean amount charged by Maryland chiropractors
- We can go ahead and estimate a 95% CI for the true average charge among all Maryland chiropractors using the formula,
  - $\bar{x} \pm t_{0.95,16-1} \times SE(\bar{x})$
  - (i.e.,  $\bar{x} \pm 2.13 \times \frac{s}{\sqrt{n}}$  )
  - For these sample results, this gives:
  - $25.50 \pm 2.13 \times \frac{2.10}{\sqrt{16}} \rightarrow 25.50 \pm 1.12 \rightarrow$   
 $(\$24.38, \$26.62)$

## Estimating a 95% Confidence Interval

4. Without assuming normality, estimate a 95% confidence interval for the mean amount charged by Maryland chiropractors
  - Actually, the normality assumption about the original individual data is NOT needed to use our results to estimate 95% CI for the true underlying population mean of the population from which the sample was taken
  - Hence the answer to this question is EXACTLY the same as the answer to the previous question (3)
  - This highlights one of the powers of the CLT related results: that we can estimate a valid confidence interval for a population mean using only a single random sample of data, regardless of what the population distribution (and random sample distribution) of the individual level data is

## Estimating a 95% Confidence Interval

5. What is the difference in the interpretation of the intervals created in questions 1 and 2 and questions 3 and 4?
  - Same answer as problem 5 from lecture 3C practice problems