JOHNS HOPKINS
BLOOMBERG
SCHOOL of PUBLIC HEALTH

Lecture 1c: Practice Problem Solutions

John McGready
Johns Hopkins University

Practice Problems

1. The following data is the annual income (in $1,000s of U.S. dollars) taken from nine students in the Hopkins internet-based MPH program:

37 102 34 12 111 56 72 17 33

Practice Problems

a) Calculate the sample mean income
   - Where n = 9 and x_1 through x_9 represent the nine observed values of income:
   - In this sample:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$\bar{x} = \frac{37+102+34+12+111+56+72+17+33}{9}$$

$$= \frac{474}{9} \approx 52.7$$ (the sample mean income is \$52,700)

Practice Problems

b) Calculate the sample median income
   - To calculate the sample median, we must first order our data from the lowest value to the highest value

12 17 33 34 37 56 72 102 111

   - Since there are nine observations, the median is the value directly in the center: the 5th ranked observation

12 17 33 34 37 56 72 102 111

   - The sample median income is $37,000

Practice Problems

c) Calculate the sample standard deviation of these incomes.

- In this sample of nine observations, the formula for the SD is:

$$s = \sqrt{\frac{\sum_{i=1}^{9}(x_i - \bar{x})^2}{8}}$$

- Now:

$$\frac{\sum_{i=1}^{9}(x_i - \bar{x})^2}{8} = \frac{(37-52.7)^2 + (102-52.7)^2 + ...(33-52.7)^2}{8}$$

$$\approx \frac{10,128}{8}$$

Practice Problems

c) So:

$$s = \sqrt{\frac{\sum_{i=1}^{9}(x_i - \bar{x})^2}{8}} = \sqrt{\frac{10{,}128}{8}} = \sqrt{1{,}267} \approx 35.6$$

- The sample standard deviation is \$35,600: on average, each of the nine observed incomes falls about \$35,600 away from the sample mean (above or below)

Practice Problems

d) What population could this sample represent?
   - It could be representative of the population of all Johns Hopkins Internet-based MPH students

Practice Problems

e) Which would change by a larger amount—the mean or median—if the 34 were replaced by 17, and the 12 replaced by a 31?
   - Notice that both changes do nothing to change the position of the median; therefore, the only statistic of the two that would change is the mean (the sample standard deviation would also change)

Practice Problems

2. The following data shows birthweights (oz) from seven consecutive deliveries at the Johns Hopkins Hospital

121 138 32 100 58 64 146

Practice Problems

a) Calculate the sample mean birthweight

b) Calculate the sample median birthweight

c) Calculate the sample standard deviation of these birthweight
   - I'm using Stata to do the first three questions!
   - Data in Stata:

```
. list

     +-----+
     |  bw |
     |-----|
  1. | 121 |
  2. | 138 |
  3. |  32 |
  4. | 100 |
  5. |  58 |
     |-----|
  6. |  64 |
  7. | 146 |
     +-----+
```

Practice Problems

a) Calculate the sample mean birthweight

b) Calculate the sample median birthweight

c) Calculate the sample standard deviation of these birthweight
   - *summarize* command with *detail* option

```
. summarize bw, detail

                            bw
-------------------------------------------------------------
      Percentiles      Smallest
 1%           32             32
 5%           32             58
10%           32             64       Obs                   7
25%           58            100       Sum of Wgt.           7

50%          100                      Mean           94.14286
                        Largest       Std. Dev.      43.67466
75%          138            100
90%          146            121       Variance       1907.476
95%          146            138       Skewness      -.1538512
99%          146            146       Kurtosis       1.544331
```

Practice Problems

d) What population could this sample represent?
   - This could possibly be representative of all births in a fixed time period (one month, one year?) at Johns Hopkins hospital, and possibly other U.S. urban teaching hospital
   - However, we may want to ask a few questions of the researcher who collected this data before making this conclusion

Practice Problems

e) Suppose this is a representative sample of births in a given year at Johns Hopkins
   - Suppose, instead of a sample of seven values, we have a sample of 100 birthweights
   - How should the mean, median and standard deviation of this sample compare to the same statistics for the sample of seven birthweights?

   - **Answer**: The mean, median and SD from the sample of 100 will likely not equal their counterparts from the sample of seven
   - However, all should be "similar" in value
   - There is no way to ascertain how each of these will differ across the samples (i.e., which sample would have the larger sample mean, etc.)