

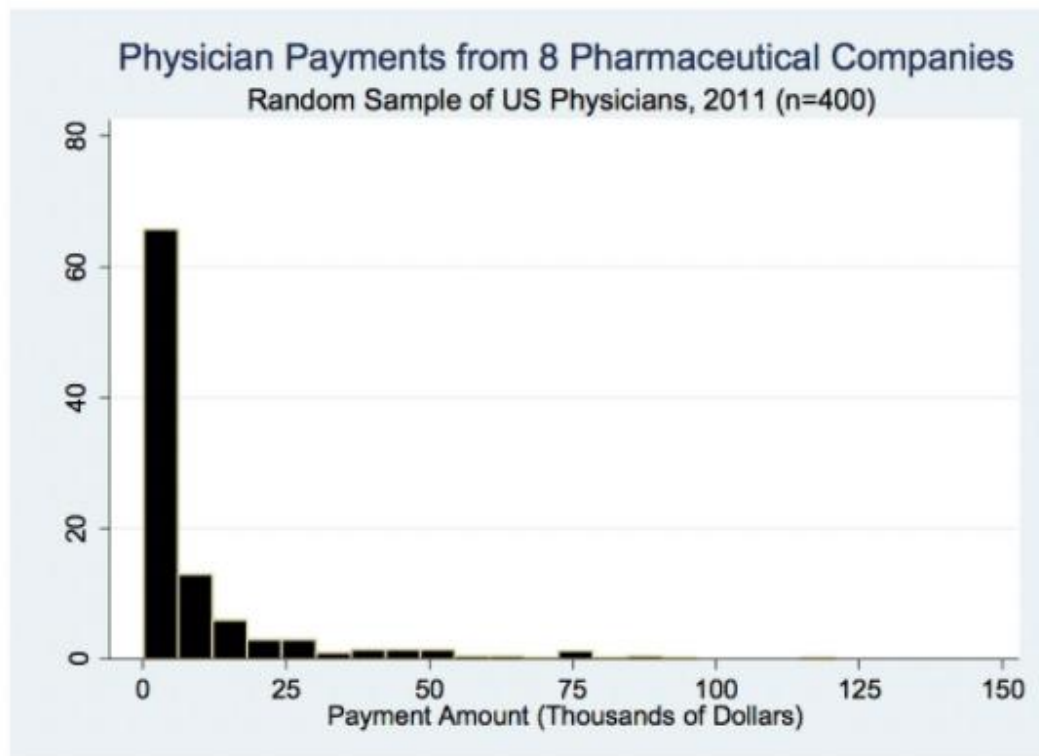
1. Which of the following is true about an observational prospective cohort study being designed to study the association between cigarette smoking (smoking ≥ 1 pack of cigarettes per day, smoking < 1 pack of cigarettes per day, or no smoking) to be assessed on 3/1/19, and catching a cold between 3/2/19 and 3/31/19?
- ☐ This design minimizes the potential for systematic differences between the exposure groups.
 - ☐ Both the exposure and outcome are measured at the same point in time.
 - ☒ Subjects self-select to be in the exposure groups, so potential confounders will need to be considered when assessing the outcome/exposure relationship.
 - ☐ This design ensures that that smoking increases the chances of getting a cold in March of 2019.
2. Which of the following statements defines the 97.5th percentile value for a data sample, regardless of the sample data distribution?
- ☐ The value such that 97.5% of the sample observations are larger than this value, and 2.5% of the sample observations are smaller than this value.
 - ☒ The value such that 2.5% of the sample observations are larger than this value, and 97.5% of the sample observations are smaller than this value.
 - ☐ The value that is 97.5% larger than the sample mean.
 - ☐ This value always corresponds to the sample mean plus 2 standard deviations.

3. 500 randomly selected university students were selected for a study on smoking status and stress levels. At the time of the study, students were asked to rate their stress levels on a scale of 0 – 100 with 0 indicating no stress. Students were also asked to disclose whether they were a non-smoker, a light smoker, or a heavy smoker.

Of the 500 students, 75 identified themselves as heavy smokers, 150 identified themselves as light smokers, and 275 identified themselves as non-smokers. The average stress level reported for heavy smokers was 60.4. The average stress level reported for light smokers was 45.7. The average stress level reported by non-smokers was 56.6. Based on these results, which of the following statements is true?

- ☒ On average, heavy smokers have stress values of 3.8 greater than non-smokers
 - ☐ On average, heavy smokers have stress values of 3.8 times less than the stress levels of non-smokers
 - ☐ On average, heavy smokers have stress values of 3.8 times the stress levels of non-smokers
 - ☐ On average, heavy smokers have stress values of 3.8 less than non-smokers
4. Which of the following sample statistics is least sensitive to the influence of outliers for a sample continuous data measures?
- ☐ The sample standard deviation.
 - ☐ The sample variance.
 - ☒ The sample median (50th percentile)

5. This graph shows the total payment amounts in 2011 (in thousands of US\$) for a **random sample** of 400 US physicians who received payments from any of 8 major pharmaceutical companies in 2011. The vertical axis represents the percent of physicians. (Source: Propublica "Dollars for Docs" Online Database)



Sample Summary Statistics (n=400)
(estimates in thousands of dollars)

Mean (\bar{x}): 10

Standard Deviation (s): 17

Range: 0.1 to 121

Percentiles:

2.5th: 0.3

25th: 1.5

50th: 3.6

75th: 10

97.5th: 74

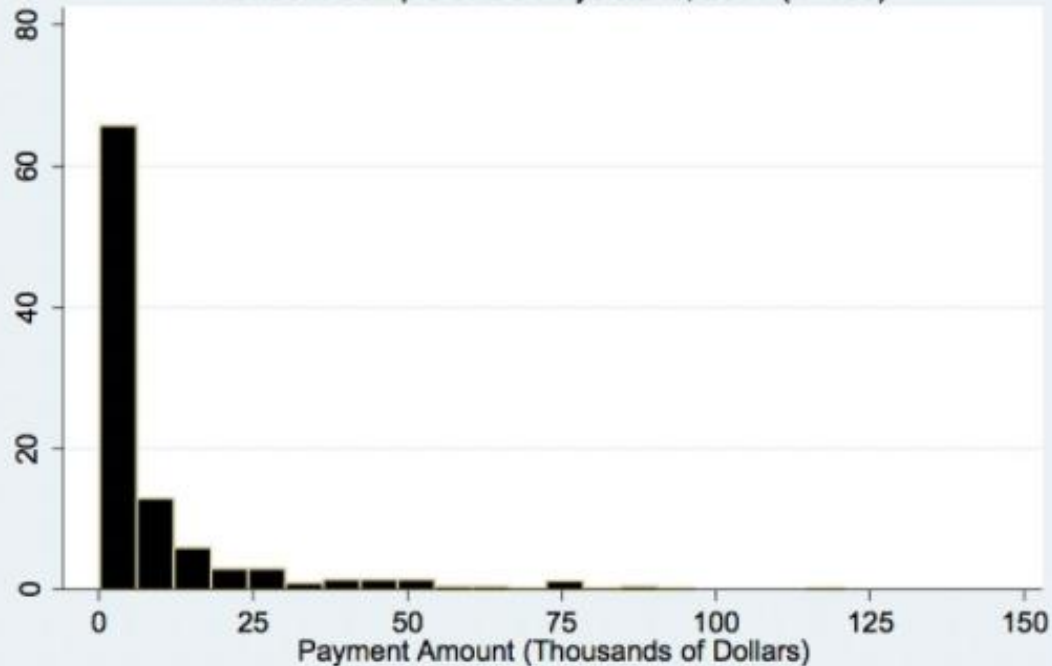
Which of the following statements best tells the story of these data?

Which of the following statements best tells the story of these data?

- ☐ The percentage of physicians receiving a given payment amount was relatively similar for all values across the range of payment values.
- ☐ Physicians did not receive adequate payments from the pharmaceutical companies.
- ☐ The majority of physicians received relatively “large” payments ($> \$25,000$) and a smaller percentage received payments less than this majority.
- ☒ The majority of physicians received relatively “small” payments ($< \$25,000$) and a smaller percentage received payments greater than this majority.

Physician Payments from 8 Pharmaceutical Companies

Random Sample of US Physicians, 2011 (n=400)



Sample Summary Statistics (n=400) (estimates in thousands of dollars)

Mean (\bar{x}): 10

Standard Deviation (s): 17

Range: 0.1 to 121

Percentiles:

2.5th: 0.3

25th: 1.5

50th: 3.6

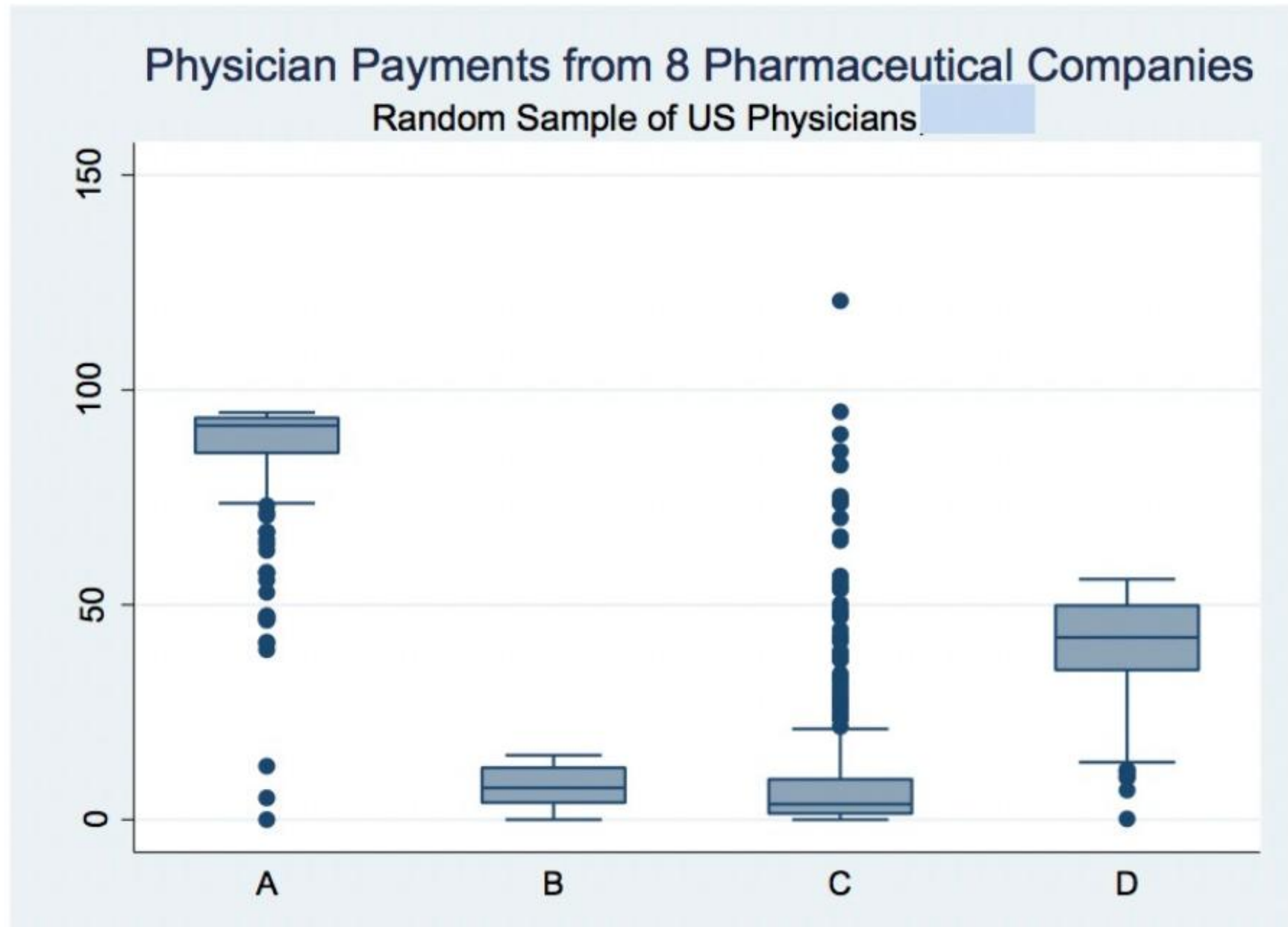
75th: 10

97.5th: 74

Given the distribution of the sample data, what is the most likely distribution of the payment data in the population of all US physicians in 2011?

- ☐ uniform
- ☐ left (negatively) skewed
- ☐ approximately normal
- ☒ right (positively) skewed

Which of the following boxplots corresponds to the physician payment distribution in this sample?



☐ A

☐ B

☒ C

☐ D

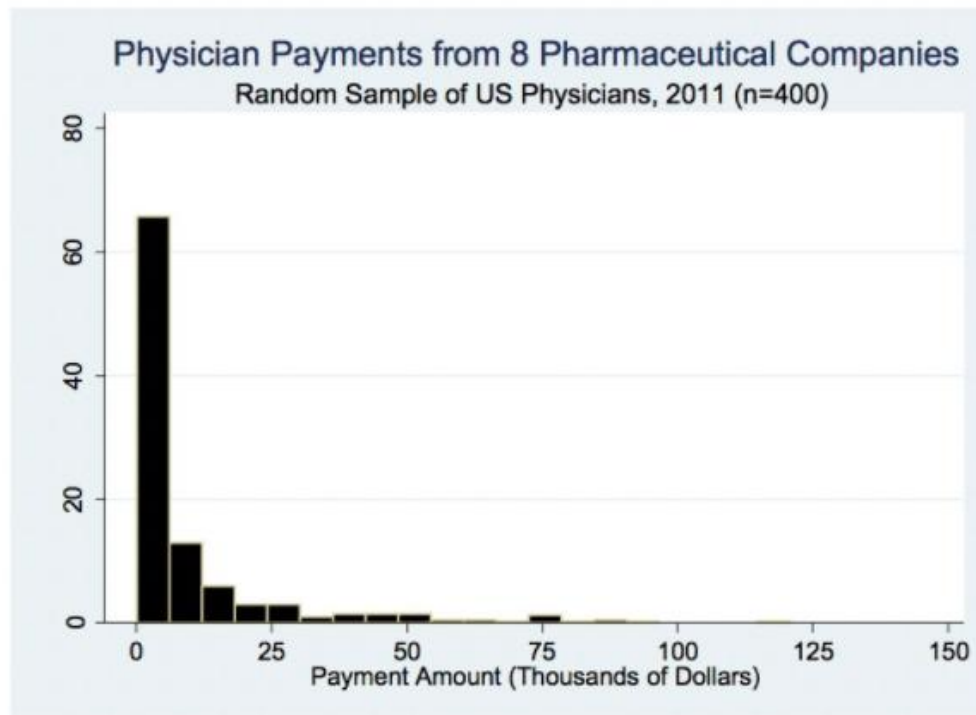
8. Suppose another researcher takes a random sample of 1000 subjects from the same population (US physicians who received payments from any of 8 major pharmaceutical companies in 2011). You have not yet seen these data. Likely, how will the sample standard deviation of these 1000 values (s_{1000}) compare to the sample standard deviation of individual physician payments in the sample of 400 ($s_{400}=17$)

☐ s_{1000} will likely be larger than 17.

☐ s_{1000} will be exactly equal to 17.

☐ s_{1000} will likely be smaller than 17.

☒ s will likely be similar in value to 17 (s), but the exact relationship between these two sample standard deviation values cannot be predicted based on the information given .



Sample Summary Statistics (n=400)
(estimates in thousands of dollars)

Mean (\bar{x}): 10

Standard Deviation (s): 17

Range: 0.1 to 121

Percentiles:

2.5th: 0.3

25th: 1.5

50th: 3.6

75th: 10

97.5th: 74

Using these data from this sample of 400 physicians, estimate the percentage of US physicians who received payments of greater than \$10,000 in 2011. This estimated percentage is :

☒ 25

☐ 50

☐ 75

☐ This cannot be estimated without having access to a standard normal table.

10. Which of the following is true about \bar{x} , the sample mean?

- ☐ The sample mean, \bar{x} , will always equal the underlying population mean
- ☒ The sample mean, \bar{x} , is the best estimate of the underlying population mean, based on the sample data.
- ☐ The sample mean, \bar{x} , tends to increase in value with increasing sample size.
- ☐ The sample mean, \bar{x} , tends to decrease in value with increasing sample size.