3. A medical study investigates the relationship between smoking and blood pressure. Participants are smokers and are separated into two groups. Group A gets help quitting and does not smoke for a month. Group B continues to smoke. At the end of the month, the systolic blood pressure (the top number in a blood pressure reading) of all participants is measured. The study records the following data:

<table>
<thead>
<tr>
<th>sex</th>
<th>age</th>
<th>blood pressure</th>
<th>group</th>
</tr>
</thead>
<tbody>
<tr>
<td>male</td>
<td>42</td>
<td>143</td>
<td>A</td>
</tr>
<tr>
<td>female</td>
<td>54</td>
<td>135</td>
<td>B</td>
</tr>
</tbody>
</table>

(a) How many participants does the study have?

(b) How many variables are in this data set?

(c) How many observations are in this data set?

(d) Is this an observational study or an experiment?

   ○ observational study   ○ experiment

(e) Can this study provide evidence that stopping smoking will have an effect on blood pressure?
(f) The boxplot to the right shows the distribution of the blood pressure variable within each group.

Below are histograms that also show the distribution of the blood pressure variable within each group. Which histogram belongs to group A and which belongs to group B?
4. The temperature each day in August follows a normal distribution with mean 85 degrees and standard deviation 10 degrees.

(a) What percentile does a 92 degree day fall into?

(b) How hot must a day be to be in the hottest 2% of August days?

(c) What is the probability that a day in August will be hotter than 100 degrees?

(d) What is the probability that a day in August will be within 13 degrees of the mean August temperature?
6. The following questions refer to the distribution depicted in this histogram:

(a) Approximately what is the median of this distribution?

(b) This distribution is
   - symmetric   - left-skew   - right-skew

(c) How does the mean compare to the median? The mean is:
   - about the same   - smaller   - larger   - not enough info to tell

(d) How close is this distribution to the normal distribution?
   - very close   - somewhat close   - far   - not enough info to tell

(e) Suppose a few observations around 100 are added to the distribution. Which of the following quantities will be substantially affected, and which will stay almost the same?

   - median:   - affected   - stay the same
   - mean:   - affected   - stay the same
   - interquartile range:   - affected   - stay the same
   - standard deviation:   - affected   - stay the same
   - maximum:   - affected   - stay the same
   - minimum:   - affected   - stay the same
5. Complete the following sentences:

(a) When conducting a hypothesis test, if the ________ is smaller than the significance level, we ________.
   Choose one answer:
   - p-value; reject the alternative hypothesis
   - p-value; reject the null hypothesis
   - test statistic; accept the null hypothesis
   - confidence interval; reject the null hypothesis.

(b) Rejecting the null hypothesis when it’s true is a _________. The frequency that this occurs is given by the _________ of the test.
   Choose one answer:
   - type-1 error; significance level
   - type-1 error; power
   - type-2 error; significance level
   - type-2 error; power

(c) The idea of statistical inference is to estimate information about ________ of a population using ________ of a sample.
   - parameters; statistics
   - the mean; the standard deviation
   - the level of significance; an observation
   - a confidence interval; a hypothesis test

(d) The smaller the ________, the stronger the evidence for the ________.
   - level of significance; null hypothesis
   - test statistic; null hypothesis
   - p-value; alternative hypothesis
   - p-value; null hypothesis
6. Suppose that a anti-cholesterol drug is being tested, and that $\mu$ is the average change in blood cholesterol levels after administering the drug to a patient. The testers wish to carry out a hypothesis test to determine if the drug has any effect on cholesterol.

(a) Write hypotheses for this test:

$$H_0:$$

$$H_A:$$

(b) The testers carry out the test with significance level 1%. Suppose that in truth, the drug has no effect on cholesterol levels. Answer the following true/false questions:

The p-value will be greater than .01.
$$\bigcirc$$ True   $$\bigcirc$$ False

In about 99 out of 100 tests carried out under these conditions, the test will fail to reject the null hypothesis.
$$\bigcirc$$ True   $$\bigcirc$$ False

Only 1 out of every 100 patients will have a change in blood cholesterol levels after taking the drug.
$$\bigcirc$$ True   $$\bigcirc$$ False
6. The box plot on the right shows the finishing times in hours of two groups of runners in the 2017 Boston marathon. Say whether following statements are true, false, or impossible to determine based on the plot:

- The median runner is faster in group 2 than in group 1.  
  True  False  Impossible
- The fastest runner is faster in group 2 than in group 1.  
  True  False  Impossible
- The interquartile range is larger in group 2 than in group 1.  
  True  False  Impossible
- All runners in the middle 50% of group 1 finished in four hours or less.  
  True  False  Impossible
- All runners in the middle 50% of group 2 finished in four hours or less.  
  True  False  Impossible
- There were more runners in group 1 than group 2.  
  True  False  Impossible
- The slowest runner was in group 2.  
  True  False  Impossible
- The fastest 25% of runners in group 1 finished in 3.5 hours or less.  
  True  False  Impossible
- The fastest 25% of runners in group 2 finished in 3.5 hours or less.  
  True  False  Impossible
7. A public health researcher does a study. She interviews 250 people and records the following information about each one: name, gender, resting heart rate (number of beats per minute), age, and hours of exercise per week.

(a) How many observations are there?

(b) How many cases are there?

(c) How many variables are there?

(d) List all categorical variables:

(e) List all quantitative variables:

8. The following are summary statistics for the income variable, giving the annual income of individuals in a dataset of full-time employed adults living in Nashville.

\[
\begin{array}{cccccccc}
\text{min} & Q1 & \text{median} & Q3 & \text{max} & \text{mean} & \text{sd} \\
$18,279 & $33,001 & $45,114 & $81,238 & $923,321 & $61,923 & $59,012 \\
\end{array}
\]

(a) Compute the interquartile range (IQR) for this variable.

(b) Is this distribution left-skewed, symmetric, or right-skewed? Explain briefly (one sentence is enough) how you know.