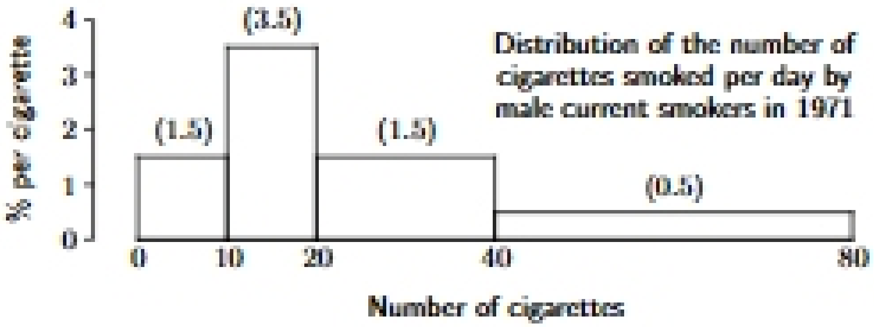
HISTOGRAMS AND PERCENTILES

- What is the 25^{th} percentile of a histogram?

The point on the horizontal axis such that ______ of the area under the histogram lies to the left of that point (and ______ to the right).

- What is the 25^{th} percentile in this case?



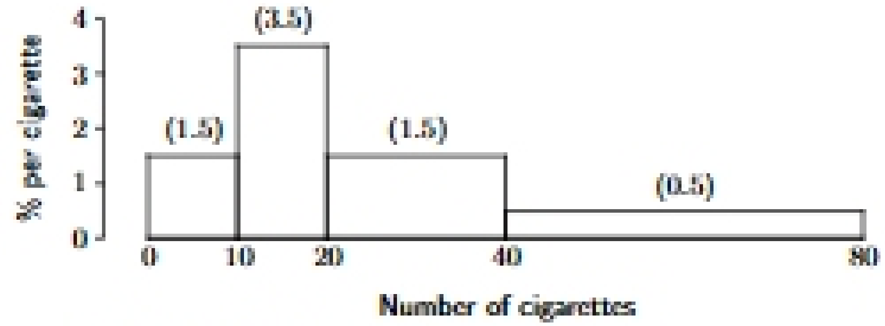
Point on horizontal axis	Area to the left of it
10	15%
11	15% + 3.5%
12	15% + 7%
13	15% + 10.5%

So the 25^{th} percentile is about 13. What does that say about the people in the study?

1 out of 4 of them smoked 13 or fewer cigarettes per day.

- Other percentiles are defined in a similar way. E.g,, the 95^{th} percentile is the point on the horizontal axis such that 95% of the area under the histogram lies to the left of it.

  - What is the 50^{th} percentile for the cigarette histogram?



  - 50^{th} percentile = 20.
  - ______ out of ______ of these men smoked ______ or fewer cigarettes per day.

- The 25^{th}, 50^{th}, and 75^{th} percentiles are called quartiles:

  25^{th} percentile = first quartile (1Q)
  50^{th} percentile = second quartile (2Q) = the median
  75^{th} percentile = third quartile (3Q)

- The interquartile range (IQR) is the distance between the first and third quartiles; this is measure of spread that is not sensitive to outlying values.

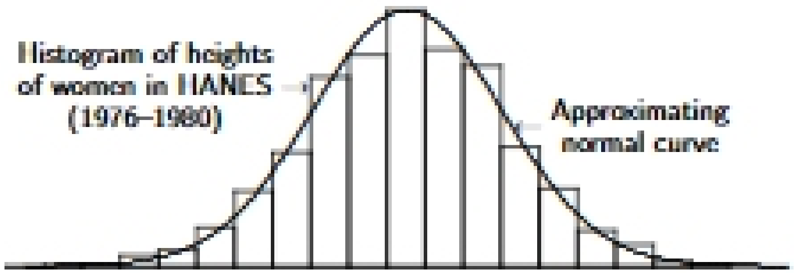
  - In the cigarette example,

$$\begin{aligned} 1Q &\approx 13 \\ 3Q &\approx 37 \\ IQR &= 3Q - 1Q \approx 37 - 13 = 24 \end{aligned}$$

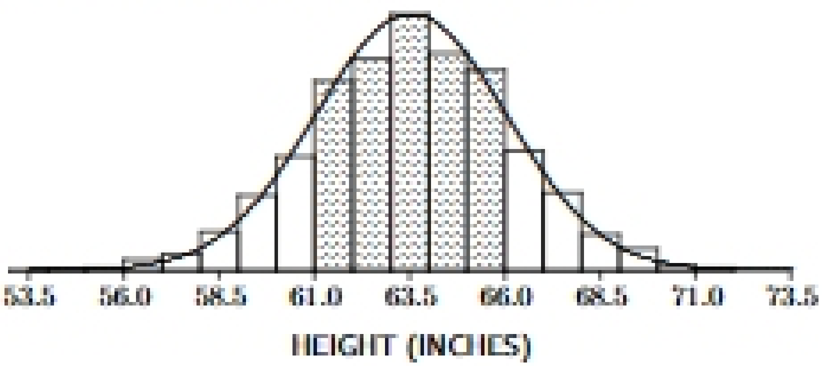
WHY ARE NORMAL CURVES OF INTEREST?

• Normal curves often provide a simple, compact way of describing how some variable is distributed.

  • Many variables (e.g., height, blood pressure, ..., but not years of education, ...) have histograms which follow (match up well with) a normal curve:



• For such variables, areas under the histogram — that is, population percentages — can be approximated by the corresponding areas under the normal curve:



• Areas under the normal curve can be computed easily knowing only the average and the SD.

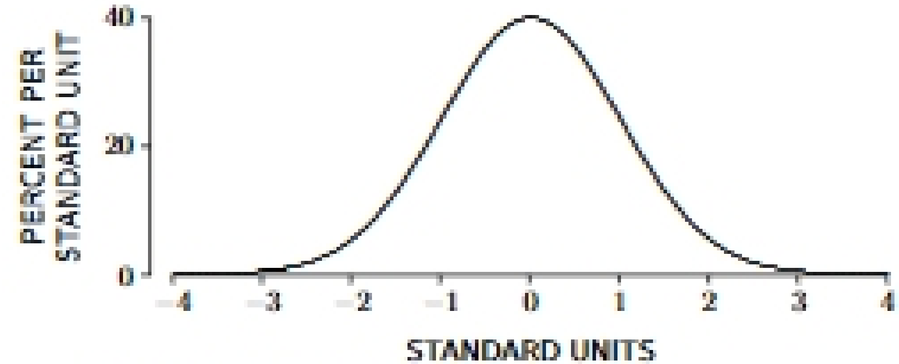
• Normal curves are well known and well understood.

  • A convenient means of communication.

• As Chapter 18 explains, the sampling distribution of sample averages tends to follow the normal curve.

  • This is the cornerstone of statistical inference!

THE STANDARD NORMAL CURVE



• The equation of the curve is

$$\text{ordinate} = \frac{100\%}{\sqrt{2\pi}} e^{-(\text{abscissa})^2/2}.$$

• Two very important properties are:

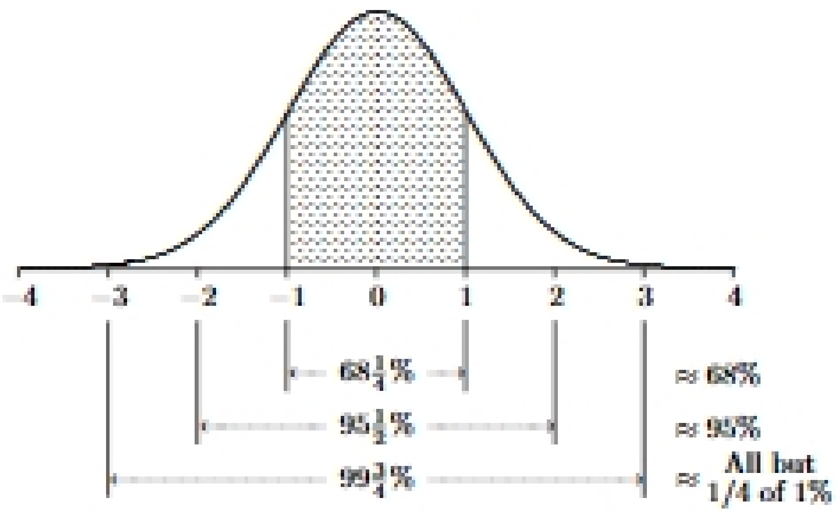
  • The total area under the curve is __100%__.

    • Just like a histogram.

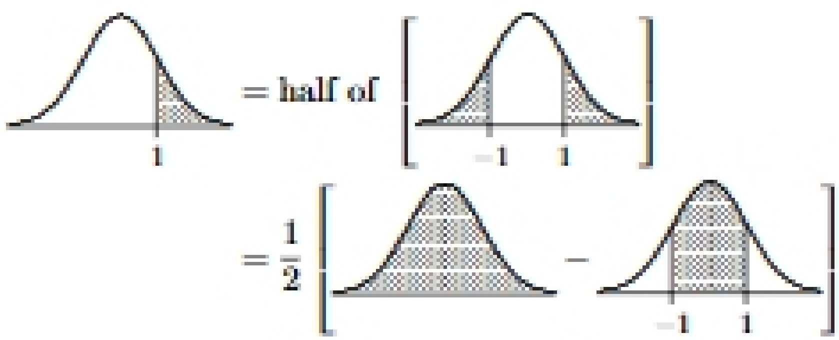
  • The curve is _symmetric_ about 0.

A BRIEF TABLE OF AREAS UNDER THE STANDARD NORMAL CURVE

• The following figure shows some "benchmark" areas under the standard normal curve:

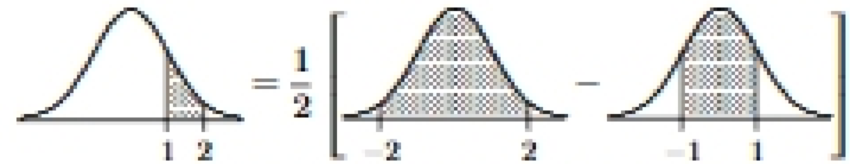


• What is the area under the standard normal curve to the right of 1?



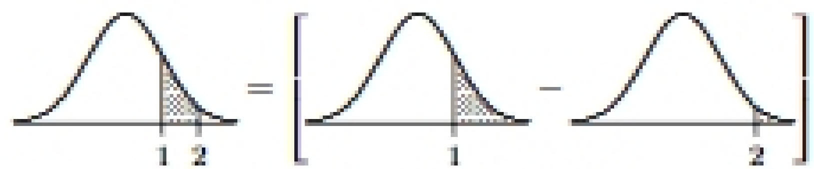
$$= \frac{1}{2}[100\% - 68\%] = \frac{1}{2}32\% = 16\%.$$

• What is the area under the standard normal curve between 1 and 2?



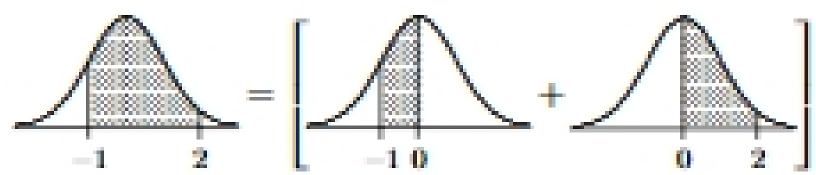
$$= \frac{1}{2}[95\% - 68\%] = \frac{1}{2}27\% = 13\tfrac{1}{2}\%.$$

Alternatively



$$= 16\% - 2\tfrac{1}{2}\% = 13\tfrac{1}{2}\%.$$

• What is the area under the standard normal curve between −1 and 2?



$$= \frac{1}{2}68\% + \frac{1}{2}95\% = 34\% + 47\tfrac{1}{2}\% = 81\tfrac{1}{2}\%.$$