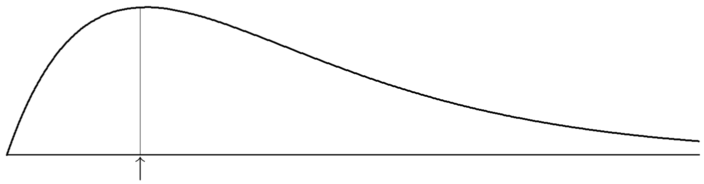
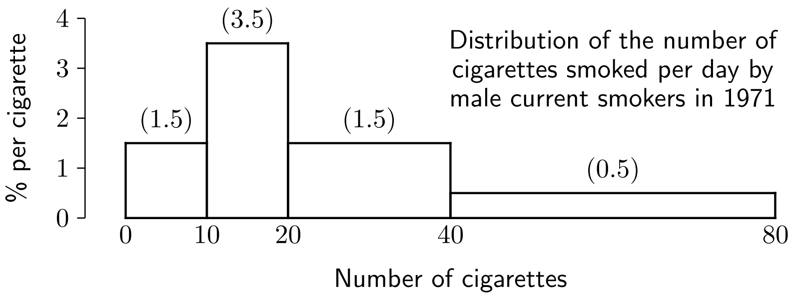
HISTOGRAMS AND PERCENTILES

- 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?



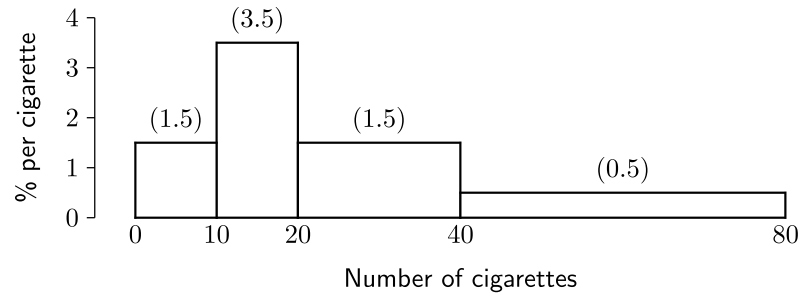
Point on horizontal axis	Area to the left of it
10	15%
11	15% + 3.5%
12	15% + 7%
13	15% + 10.5%

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.

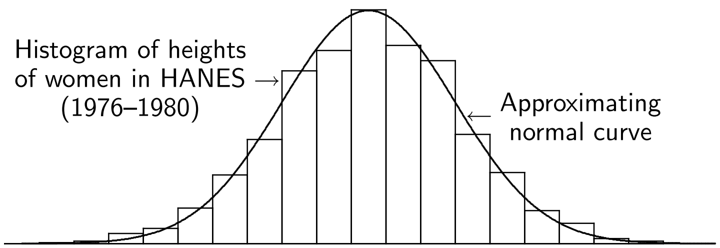
  - In the cigarette example,

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

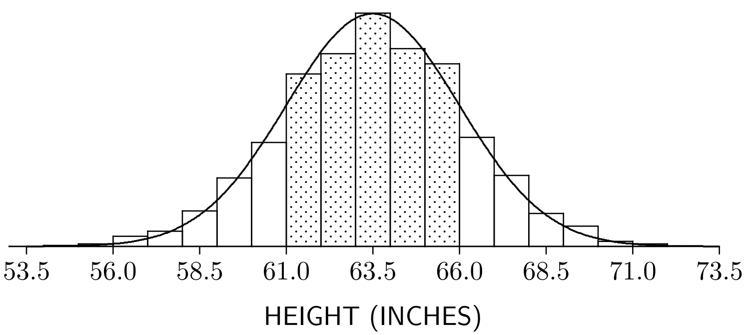
WHY ARE NORMAL CURVES OF INTEREST?

• 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.

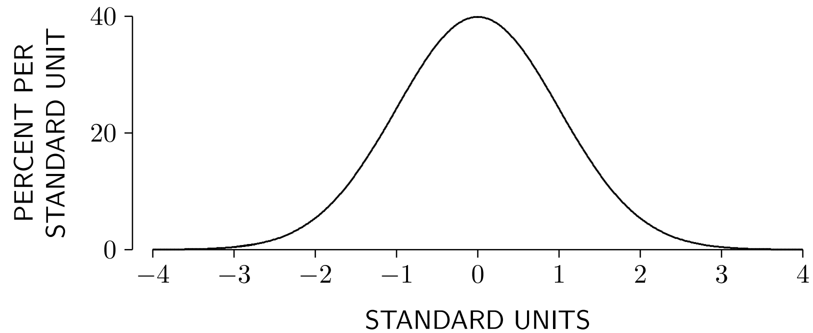
• Normal curves are well known and well understood.

  • 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:

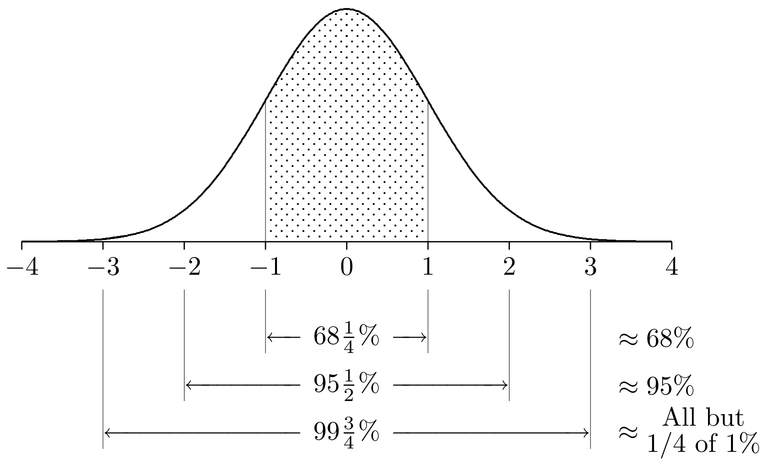
    • The total area under the curve is 100%.

      • Just like a histogram.

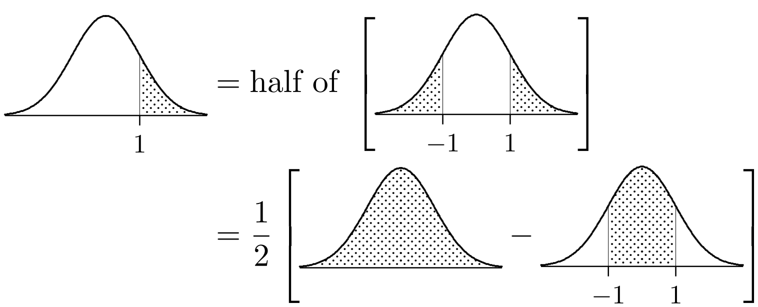
    • The curve is symmetric about 0.

A BRIEF TABLE OF AREAS UNDER THE STANDARD NORMAL CURVE

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

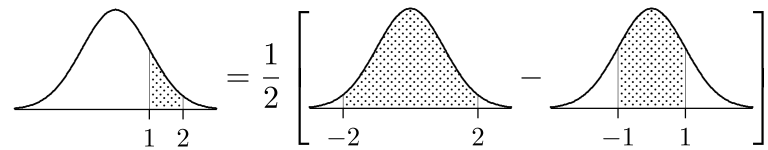


• What is the area under the standard normal curve to the right of 1?



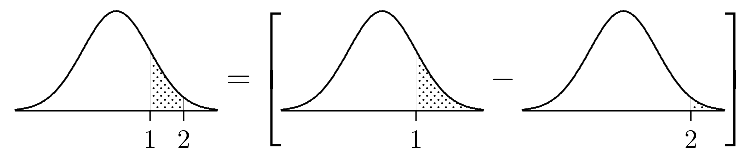
$$= \frac{1}{2}[100\% - 68\%] = \frac{1}{2}32\% = 16\%.$$

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



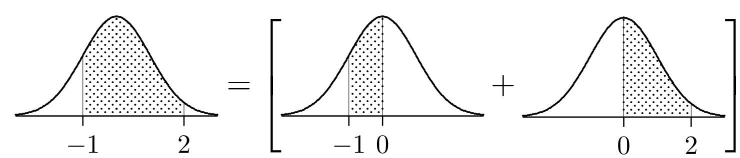
$$= \frac{1}{2}[95\% - 68\%] = \frac{1}{2}27\% = 13\tfrac{1}{2}\%.$$

Alternatively



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

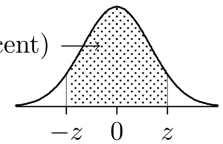
• What is the area under the standard normal curve between −1 and 2?



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

A NORMAL TABLE

• The following table is like the one on page A86 of FPPA, except that it omits the columns of heights of the normal curve:



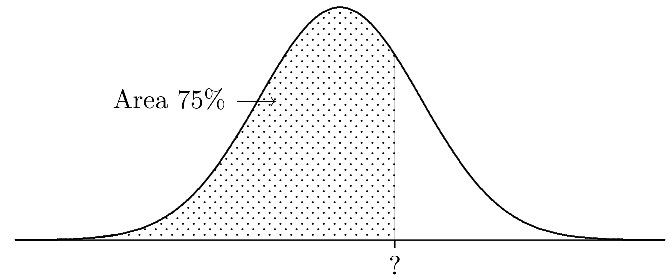
z	Area	z	Area	z	Area
0.00	0.	1.50	86.64	3.00	99.730
0.05	3.99	1.55	87.89	3.05	99.771
0.10	7.97	1.60	89.04	3.10	99.806
0.15	11.92	1.65	90.11	3.15	99.837
0.20	15.85	1.70	91.09	3.20	99.863
0.25	19.74	1.75	91.99	3.25	99.885
0.30	23.58	1.80	92.81	3.30	99.903
0.35	27.37	1.85	93.57	3.35	99.919
0.40	31.08	1.90	94.26	3.40	99.933
0.45	34.73	1.95	94.88	3.45	99.944
0.50	38.29	2.00	95.45	3.50	99.953
0.55	41.77	2.05	95.96	3.55	99.961
0.60	45.15	2.10	96.43	3.60	99.968
0.65	48.43	2.15	96.84	3.65	99.974
0.70	51.61	2.20	97.22	3.70	99.978
0.75	54.67	2.25	97.56	3.75	99.982
0.80	57.63	2.30	97.86	3.80	99.986
0.85	60.47	2.35	98.12	3.85	99.988
0.90	63.19	2.40	98.36	3.90	99.990
0.95	65.79	2.45	98.57	3.95	99.992
1.00	68.27	2.50	98.76	4.00	99.9937
1.05	70.63	2.55	98.92	4.05	99.9949
1.10	72.87	2.60	99.07	4.10	99.9959
1.15	74.99	2.65	99.20	4.15	99.9967
1.20	76.99	2.70	99.31	4.20	99.9973
1.25	78.87	2.75	99.40	4.25	99.9979
1.30	80.64	2.80	99.49	4.30	99.9983
1.35	82.30	2.85	99.56	4.35	99.9986
1.40	83.85	2.90	99.63	4.40	99.9989
1.45	85.29	2.95	99.68	4.45	99.9991

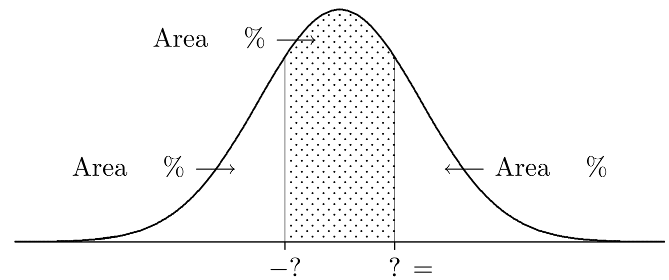
Benchmarks

z	Area
	50%
1.	
	90%
	95%
2.	
3.	

THE QUARTILES OF THE STANDARD NORMAL CURVE

• What is the first quartile of the standard normal curve?





• The quartiles of the standard normal curve are:

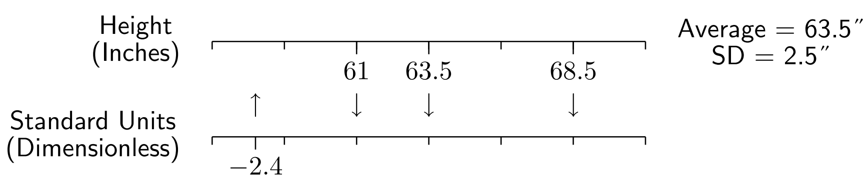
$$1Q = -0.675$$
$$2Q = 0$$
$$3Q = 0.675$$

• The interquartile range (IQR) for the standard normal curve is

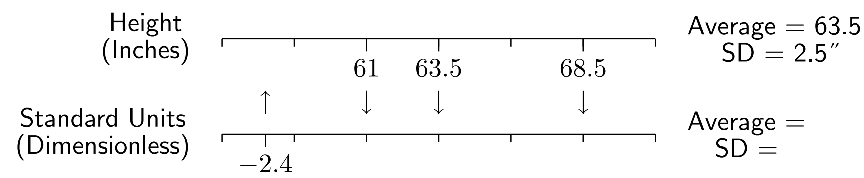
$$\text{IQR} = 3Q - 1Q = 0.675 - (-0.675) = 1.35 \approx 1.33 \approx 4/3.$$

STANDARD UNITS

- What are standard units?
  - Standard units say how many SDs a value is above (+ sign) or below (– sign) average.
- The women in the HANES study had heights averaging to 63.5 inches, with an SD of 2.5 inches.
  - What is 61″ in standard units?
    - 61″ is ________ inches ________ average.
    - That's ________ SD ________ average.
    - So 61″ is ________ in standard units.
  - What is 68.5″ in standard units?
    - 68.5″ is ________ inches ________ average.
    - That's ________ SDs ________ average
    - So 68.5″ is ________ in standard units.
  - What height is −2.4 in standard units?
    - The height is ________ SDs ________ average.
    - That's $\frac{12}{5} \times \frac{5}{2} = 6$ inches ________ average.
    - The height is ________.



- Reminder: standard units say how many SDs a value is above (+ sign) or below (− sign) average.



- Is there a formula for converting a value to standard units?
  - Yes, the formula is

$$\text{standard units} = \frac{\text{value} - \text{average}}{\text{SD}}$$

In our example, to express 68.5″ in standard units you compute

$$\frac{68.5'' - 63.5''}{2.5''} = \frac{5.0''}{2.5''} = 2$$

- Is there a formula for converting back from standard units to the original scale?
  - Yes, the formula is

$$\text{value} = \text{average} + (\text{standard units} \times \text{SD}).$$

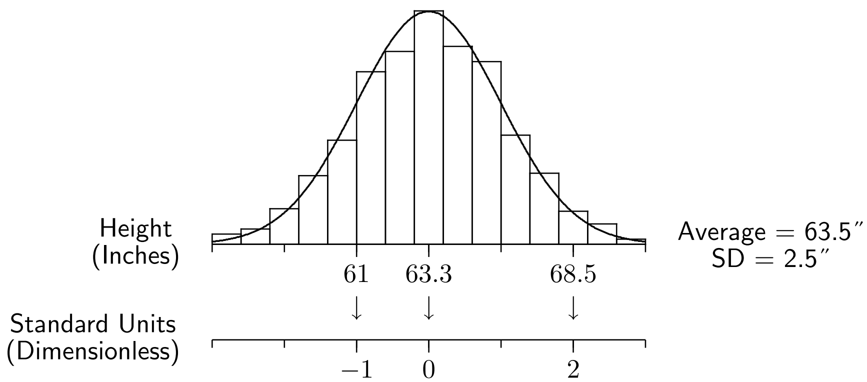
In our example, to find the height corresponding to −2.4 standard units, you compute

$$63.5'' + (-2.4 \times 2.5'') = 63.5'' - 6'' = 57.5''$$

THE NORMAL APPROXIMATION

• If a list of numbers follows the normal curve, the percentage of entries falling in a given interval can be estimated by first converting the interval to standard units and then finding the corresponding area under the standard normal curve. This procedure is called the normal approximation.

• Consider the heights of women in HANES:



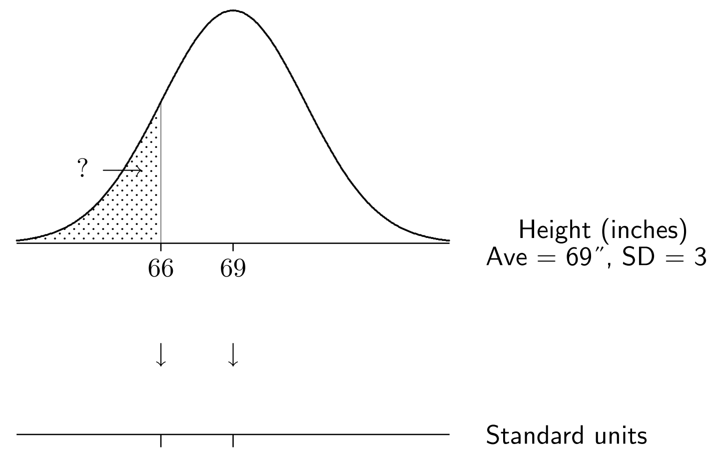
- The percentage of women with heights between 61″ and 68.5″ is exactly equal to the area under the ________ from ________ to ________, and approximately equal to the area under the ________ between ________ and ________, namely 81.5%.

• If a histogram follows the normal curve, about ________ percent of the area lies within one SD of the average, and about ________ percent within two SDs of the average.

• Warning: The normal approximation, especially for one-sided areas, is only valid if the histogram is approximately normal. Use your judgement.

USING THE NORMAL APPROXIMATION

• A group of people have heights that follow a normal curve with average 69″ and SD 3″.† About what percentage of these people have heights 66″ or under?

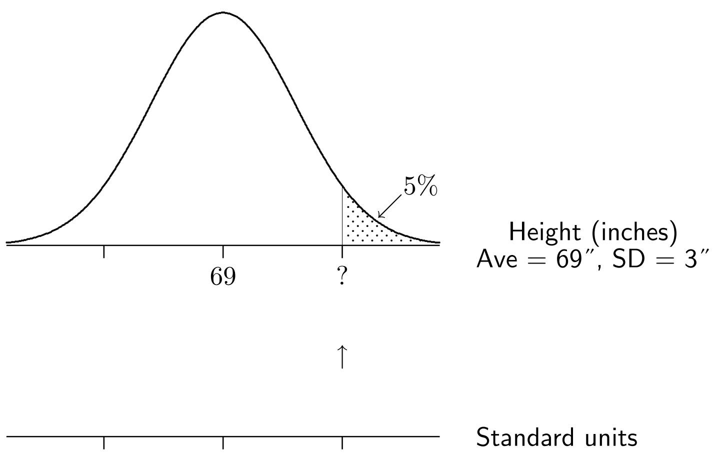


Answer = 16% (see page 5)

• The method: original units → standard units → standard normal curve.

† Are these men, or women? Ave height = 5 foot 9: they're men

• Same population (heights averaging to 69″ with an SD of 3″). What height is exceeded by 5% of the population?



Thus

$$? = \text{height} = \underline{\hspace{2cm}} \text{ SDs above average}$$
$$= 1.65 \times 3'' + 69''$$
$$= 5'' + 69'' = 74'' = 7' \, 2''$$

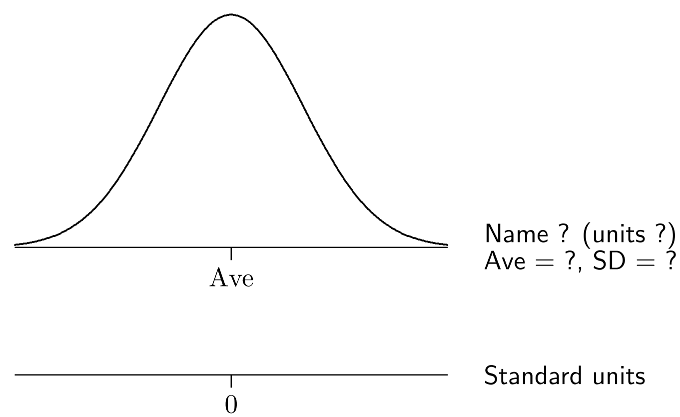
• The method: standard normal curve → standard units → original units.

SUMMARY

• What is the general procedure for working these kinds of problems?

DRAW THE PICTURE

- Sketch the normal curve
- Put in the axis for the original units
- Put in the axis for the standard units
- Shade the area of interest
- Proceed



• Be sure to follow this procedure on the homework, quizzes, and exams!