Math 263 Section 005: Class 2 : Normal Distribution and *z*-scores

Deborah Hughes Hallett dhh@math.arizona.edu

Course Policies and Information at http://math.arizona.edu/~dhh/263-14.html. You need a WebAssign account for the course. Go to www.webassign.net. The class key is **Arizona 7989 8719**

Class 2 covers the normal distribution, including standard normal *z*-values, in Section 1.3 in Introduction to the Practice of Statistics, 7-th edn, by D. Moore, G. McCabe, B. Craig. (W.H. Freeman, 2012)

TYPES of VARIABLE
- **Discrete variables**: Age in years, number of days temperature goes over 100°F, class size
- **Continuous variables**: Exact temperature in °F, sea level, atmospheric pressure

What is a distribution? What does it tell you?
Shows the values that occur and how common each one is. Example: Distribution of ages, heights, scores.

## Histogram: Picturing the Distribution of Discrete Variable

What are the axes when a data set is represented graphically using a histogram? What would a distribution of income in a city look like?

Histogram of 2002 Wage Income, National Sample
Source: Current Population Survey(CPS)
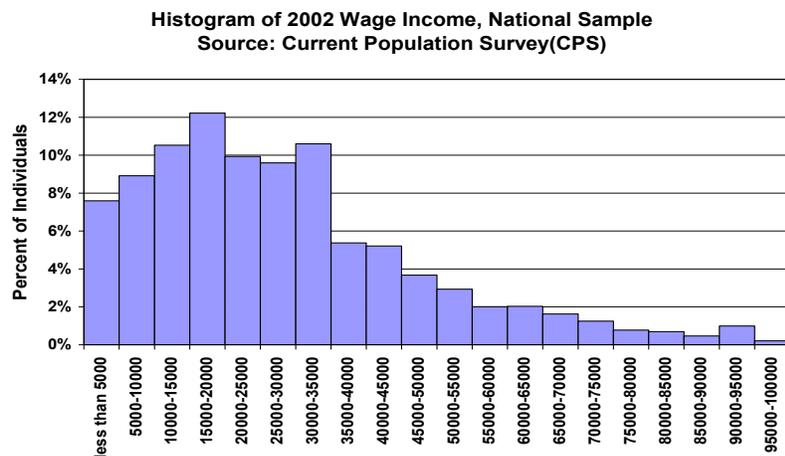
- **Horizontal axis**: Income

- **Vertical axis**: Number of people (frequency histogram) or percent of population (relative frequency histogram)

- **Shape:** Skewed right: More poor people than rich

Density Curve: Picturing the Distribution of a Continuous Variable
The histogram shows the actual vocabulary scores of a group of 7-th grade children. The smooth curve is the idealized curve what we imagine we would get if we took the population of all such children and made the bin widths very small. The smooth curve is called a ***density curve***.
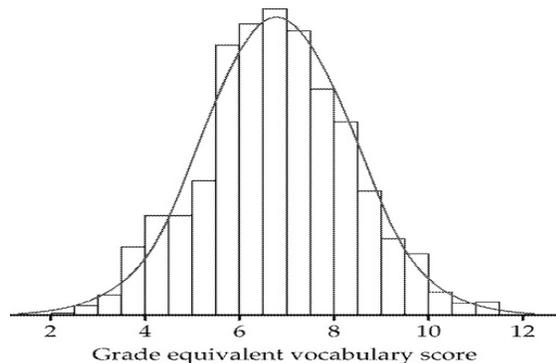

What do you notice about the smooth curve?
- Symmetric
- Bunched around a central value, 7
- Mean is 7
- Standard deviation is perhaps around 2

Using a distribution to measure proportions and probabilities

For a discrete variable, such as age in years, we can calculate the proportion of the population that has an exact value—for example, the proportion of 10-year olds in a school.

For a continuous variable, such as the quantity of oil discovered in a region. We expect no discoveries to lead to *exactly* 2 trillion cubic feet of oil, for example, so the proportion having any particular value is zero. We can only calculate the proportion *between* two values. Thus we have:

- **Histogram: Proportion, or probability, is height of bars**
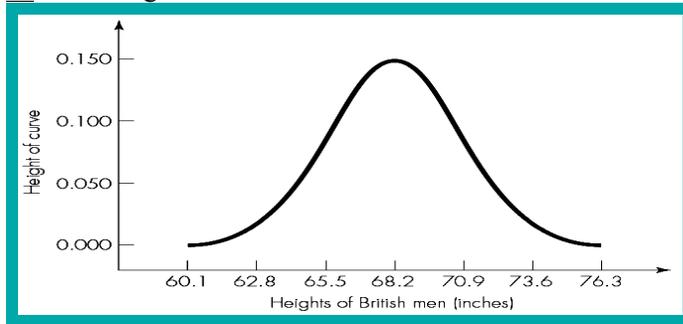- **Density Curve: Proportion, or probability, is area under curve**

Proportion given by
area

a   *b*   *x*
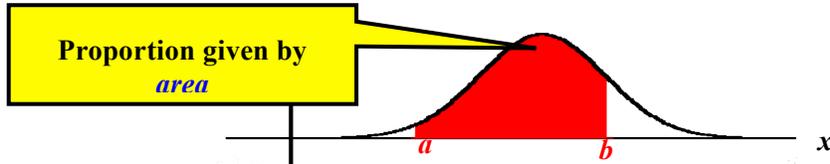
The proportion of data values lying between two values is the *area* under the density curve between these values.

The NORMAL DISTRIBUTION
Represented by a bell shaped curve, bunched around the mean.

ex   Heights of men[1]. Estimate the mean and standard deviation.

Mean: 68.2 inches
Standard Deviation: 2.7 inches

ex   How do the mean and standard deviation of a normal distribution affect the graph?

Normal Distribution
Mean = 0, Standard deviation = 1

Normal Distribution
Mean = 1, Standard deviation = 0.8
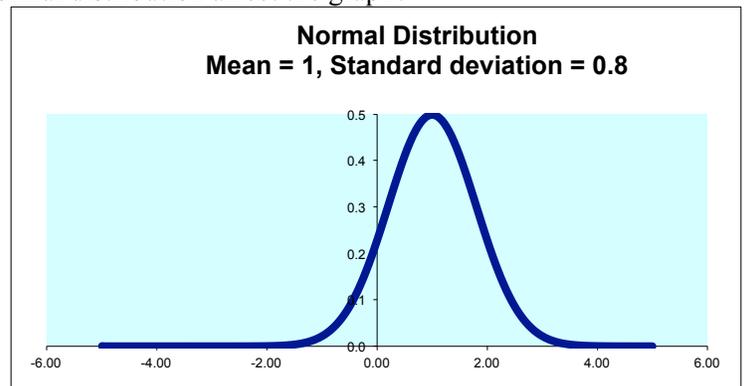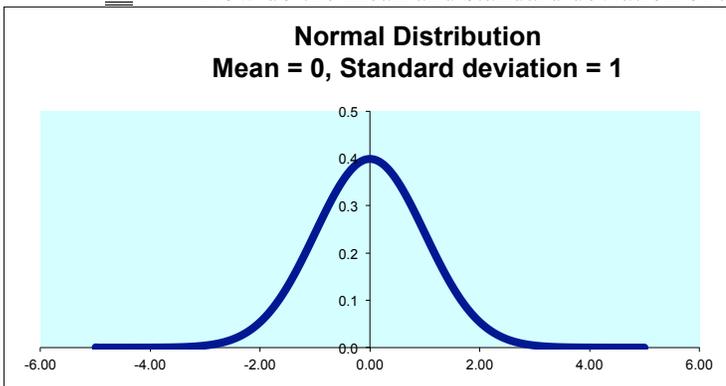
- Mean fixes center of graph;
- Standard deviation measures spread

Notice that each graph spreads roughly three times the standard deviation on either side of the mean. Thus the standard deviation is approximately 1/6 of the range.

---

[1] From Jessica Utts, *Seeing Through Statistics* 3rd edition. Thomson-Brooks Cole.

ex: How does a normal distribution look with mean 3 and standard deviation 1? With mean 0 and standard deviation 0.5? Sketch these distributions.

Normal Distribution: Mean = 3, Standard Deviation = 1

Normal Distribution: Mean = 0, Standard Deviation = 0.5
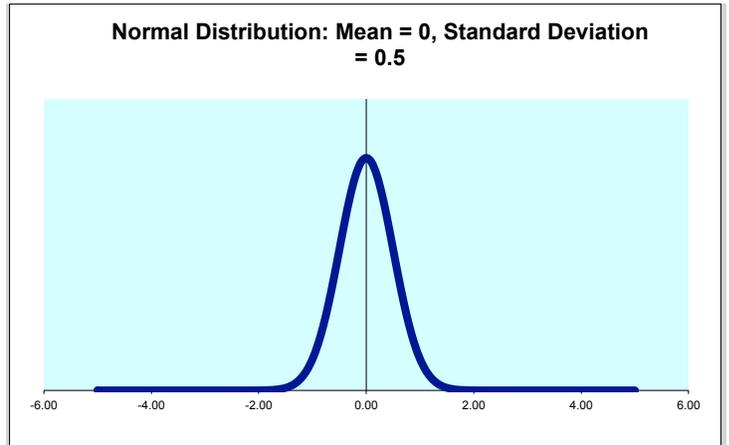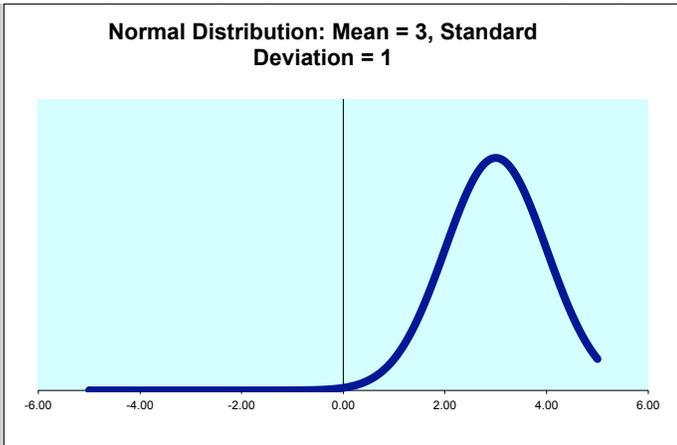
Ex. Which of the following quantities are likely to be normally distributed? Make rough sketch.
*   Incomes in a city Not normal—because not symmetric. There are more low incomes than high; a few very high incomes makes the distribution skewed right.

*   Scores on a standardized test (for example ACT, SAT) Approximately normal. Math SAT-2008: Mean = 515, SD= 116. The maximum and minimum scores possible on the text are 800 and 200. Since $3 \cdot 116 = 348$ and $515 + 348 = 863$ and $515 - 348 = 167$, the distribution cannot be exactly normal.

Distribution of SAT Math Scores
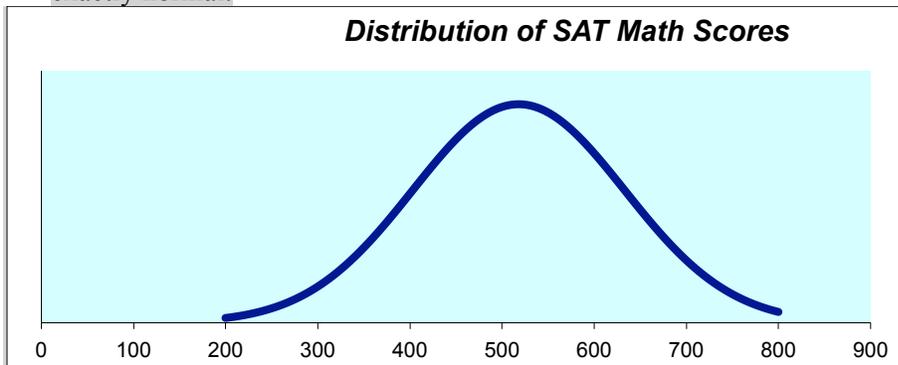
"Rule of Thumb": for the Normal Distribution
For all normal curves, it can be shown that, approximately

68% of the data lies within 1 std dev of mean
95% of the data lies within 2 std dev of mean
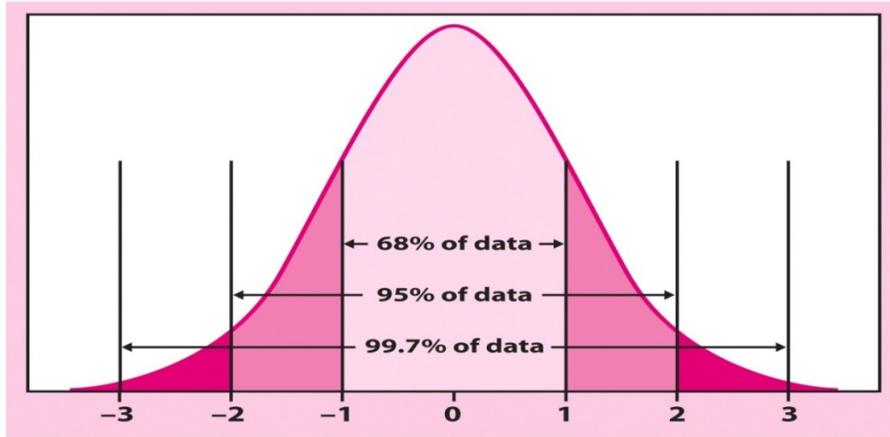99.7% of the data lies within 3 std dev of mean

Figure 13-8
Statistics: Concepts and Controversies, Sixth Edition
© 2006 W. H. Freeman and Company

ex: Women's heights are normally distributed with mean 65 inches (165 cm) and standard deviation 2.5 inches (6.4 cm).
(a) What proportion of women are less than 60 inches? (5ft)
(b) What proportion of women are less than 70 inches?
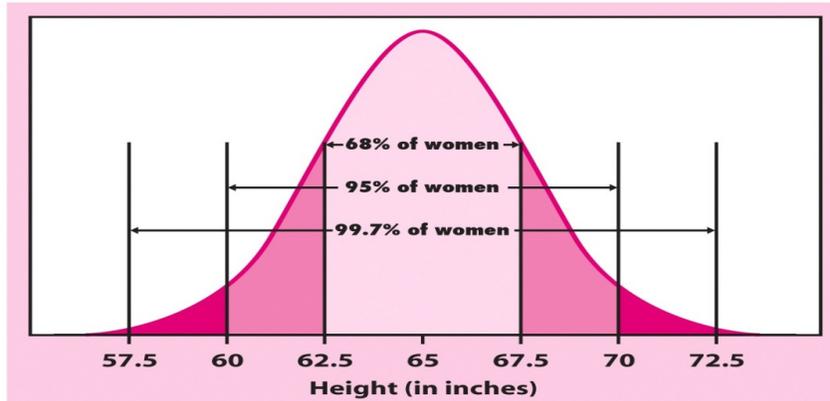(c) What proportion is more than 72 inches? (6ft)

Figure 13-9
Statistics: Concepts and Controversies, Sixth Edition
© 2006 W. H. Freeman and Company

(a) 60 inches is two standard deviations below the mean. About 2.5% of women are less than 60 inches. (2.5 = (100 – 95)/2)
(b) 70 inches is two standard deviations above the mean. About 97.5% of women are less than 70 inches. (97.5 = 100 – 2.5)
(c) 72 inches is close to 72.5 inches, which is three standard deviations above the mean. The proportion of women above 72.5 inches is 0.15%. (0.15 = (100 – 99.7)/2). The true answer is slightly larger than this.

How do we compute proportions if the value given is not an exact multiple of standard deviations away from the mean? (Like the 72)? **We use the z-score (or z-value)**, defined next.

THE Z-SCORE: Comparing Values in Different Normal Distributions

To compare values from different normal distributions (with different means and standard deviations), we find how many standard deviations each value is above or below its mean. This is the **z-score**:

$$z = \frac{\text{Value} - \text{Mean}}{\text{Standard deviation}}$$

The values of z have the **standard normal distribution**, with mean = 0 and standard deviation = 1.

Ex: On the 2008 SAT, which of the following scores represents the best performance:[2] 580 on reading, 595 on math, or 575 on writing?

	Reading	Math	Writing
Mean	501	515	493
Standard Deviation	112	116	111

We find the z-scores for each test.

$$\text{Reading } z = \frac{580-501}{112} = 0.71$$
$$\text{Math } z = \frac{595-515}{116} = 0.69$$
$$\text{Writing } z = \frac{575-493}{111} = 0.74$$

Thus the writing score was the most impressive, as it was 0.74 standard deviations above the mean. The other two scores were 0.69 and 0.71 standard deviations above their respective means.

Z-TABLES

The normal table shows the z-values to the left (with the second decimal place across the top); the body of the table shows the proportion of values to the left of each z-value. (See picture at top of table.)

Finding the Proportion of Data in a Normal Distribution using the Table

ex   Find the proportion of data that has z-score less than 0.7.
From the table, the proportion is $0.7580 = 75.8\%$
ex   Find the proportion with z-score above 1.2.
The proportion below $z = 1.2$ is 0.8848. To get the proportion above $z = 1.2$, we subtract from the whole area, which is 1 (or 100%). Thus, we have $1 - 0.8848 = 0.1151$, or $100 - 88.49\% = 11.51\%$
ex Find the proportion with z-score between –0.20 and 1.4.
Looking up 1.4 gives 91.92% less than 1.4; similarly there are 42.07% less than –0.2. Thus
Proportion between is $91.92\% - 42.07\% = 49.85\%$
ex   What z-score is at the 70th percentile?
We want a z-score with 70% of the data to the left of it. From the table, this is between $z = 0.52$ and $z = 0.53$.
ex   What z-score has 75% of the data above it?
The score we want is at the 25th percentile, so about $z = -0.67$.

ex: Find the proportion of women who are shorter than 72 inches. (Heights are normally distributed with mean 65 inches and standard deviation 2.5 inches.)
Since
$$z = \frac{\text{Value} - \text{Mean}}{\text{Std Dev}},$$

---

[2] The College Board's *Total Group Profile Report: 2009 College-Bound Seniors*

we have
$$z = \frac{72 - 65}{2.5} = 2.8,$$
From the table, the proportion below $z = 2.8$ is $0.9974 = 99.74\%$; this is the proportion of women with height below 72 inches.

Ex: Find the proportion of women taller than 72 inches.
Since the 72 inches is the same, as in the previous example, the z-value is the same and so is the proportion from the table. To find the proportion above $z = 2.8$, we subtract from 1, so the proportion is $1 - 0.9974 = 0.0026 = 0.26\%$. This is the proportion taller than 72 ins.

Ex: How tall is a woman who is at the 30th percentile in height?
Look for 0.3 in the body of the table; it is between $z = -0.52$ and $z = -0.53$; it's closer to $z = -0.52$ so we pick that. Thus, if x is the value we are looking for, we have
$$-0.52 = \frac{x - 65}{2.5}.$$

Solving gives
$$x = 65 + 2.5(-0.52) = 63.7.$$
Thus, if a woman has height 63.7 inches, she is taller than 30% of women.
Ex: What percentile is a SAT reading score of 700? (Mean reading score is 503; standard deviation 113).
Since
$$z = \frac{\text{Value} - \text{Mean}}{\text{Std Dev}},$$
we have
$$z = \frac{700 - 503}{113} = 1.74,$$
so we look up $z = 1.74$ in the table and find that it has 95.91% of the data below it. So 95.91% is the percentile.
ex   What SAT math score is at the 90th percentile? (Mean math score is 518; standard deviation 115.)
We want a z-score with 90% of the data below it. From the table, this is $z = 1.28$. If the SAT score is x, we have
$$1.28 = \frac{x - 518}{115}$$
$$x - 518 = (1.28)115$$
$$x = 518 + (1.28)115 = 665.$$

Finding proportions using a TI-83/84

To find proportion between a and b, look under "Distr" menu and use
normalcdf $(a, b,$ mean, standard deviation)
To find the x value that corresponds to a proportion p, use
invNorm(p, mean, standard deviation)

Finding proportions using Excel

To find the proportion of the data to the left of x, use
= NORMDIST $(x,$ mean, standard deviation, true).
To find the x value that corresponds to a proportion p, use
= NORMINV $(p,$ mean, standard deviation)