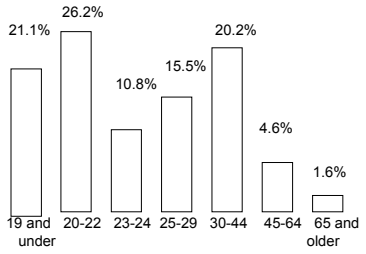
Review Question

The figure below shows the age distribution of students at the City College of New York. For example, 1.6% were 65 and older. The percentages start high, rise a little, then drop, climb, and finally drop again. How can this pattern be explained? Answer briefly.



The Normal Distribution

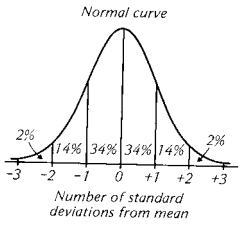
THE
NORMAL
LAW OF ERROR
STANDS OUT IN THE
EXPERIENCE OF MANKIND
AS ONE OF THE BROADEST
GENERALIZATIONS OF NATURAL
PHILOSOPHY ❖ IT SERVES AS THE
GUIDING INSTRUMENT IN RESEARCHES
IN THE PHYSICAL AND SOCIAL SCIENCES AND
IN MEDICINE AGRICULTURE AND ENGINEERING ❖
IT IS AN INDISPENSABLE TOOL FOR THE ANALYSIS AND THE
INTERPRETATION OF THE BASIC DATA OBTAINED BY OBSERVATION AND EXPERIMENT

Jack Youden

Where Are We Going?

- The normal curve and its properties
- Finding areas under the normal curve
- Approximating histograms by the normal curve

The Normal Curve



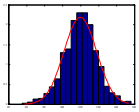
Some facts about the normal curve:

- It goes on forever in the horizontal direction
- The total area under the curve is 100%
- It is symmetric about 0
- There is hardly any area past +4 or -4.

The normal curve was discovered in 1720 by Abraham de Moivre, a French Huguenot who went to England to escape religious persecution.

$$\frac{100}{\sqrt{2\pi}} e^{-\frac{1}{2}x^2}$$

The normal curve was used to describe histograms of data around 1870 by Adolph Quetelet*, a Belgian astronomer.



*"If an individual at any given epoch in society possessed all the qualities of the average man, he would represent all that is great, good, or beautiful."

It is also called the "Gaussian density" in honor of Carl Friedrich Gauss, who was one of the first to use it.

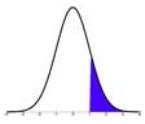
Tables

A NORMAL TABLE

z	Height	Area	z	Height	Area	z	Height	Area
0.00	39.89	0	1.50	12.95	86.64	3.00	0.443	99.730
0.05	39.84	3.99	1.55	12.00	87.89	3.05	0.381	99.771
0.10	39.69	7.97	1.60	11.09	89.04	3.10	0.327	99.806
0.15	39.45	11.92	1.65	10.23	90.11	3.15	0.279	99.837
0.20	39.10	15.85	1.70	9.40	91.09	3.20	0.238	99.863
0.25	38.67	19.74	1.75	8.63	91.99	3.25	0.203	99.885
0.30	38.14	23.58	1.80	7.90	92.81	3.30	0.172	99.903
0.35	37.52	27.37	1.85	7.21	93.57	3.35	0.146	99.919
0.40	36.83	31.08	1.90	6.56	94.26	3.40	0.123	99.933
0.45	36.05	34.73	1.95	5.96	94.88	3.45	0.104	99.944
0.50	35.21	38.29	2.00	5.40	95.45	3.50	0.087	99.953

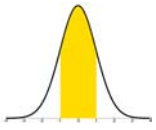
7

Using the Table to Find Areas under the Normal Curve



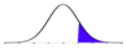
=

100% -



2

8



=

100% -

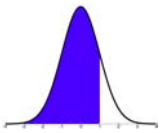
2

Example: What percentage if greater than z=1?

From the normal table, the area between –1 and +1 is 68.27%

So, the area outside –1 to +1 is 100 – 68.27 = 31.73% So, the area beyond +1 is 15.87% or round to 16%

9



= 100% -

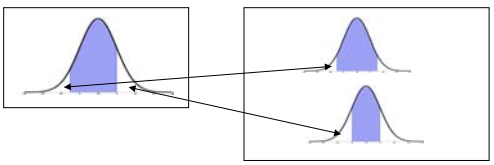
Continuing the previous example, the area less than z = 1 is 84%

Example: What is the area between -1.5 and 1?

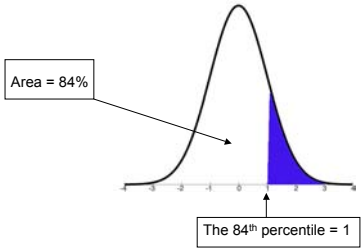
What's the picture of the area we want to find?

How does this relate to central areas?

Area between –1.5 and 1?

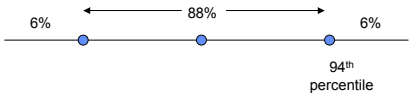


Percentiles of the Normal Curve

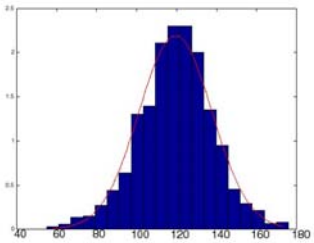


How would you find the 94th percentile?

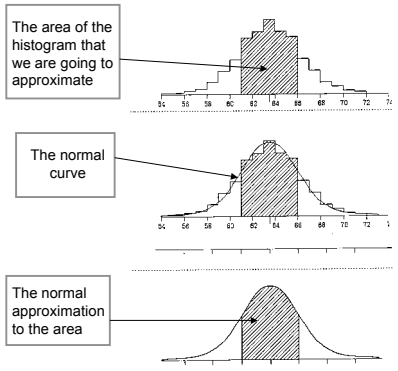
Finding the 94th percentile



Using the Normal Curve to Approximate Histograms



Histogram of All Birth Weights



16

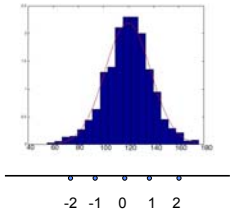
Conversion to Standard Units

Average = 119.6 SD = 18.2

Average=119.6
= 0 standard units

Average + 1 SD
= 119.6 + 18.2 = 137.8
= 1 standard units

Average – 2 SD
= 119.6 – 2 x 18.2
= 83.2
= -2 standard units



A value in standard units is how many SDs is that value from the average (including sign)

Standard Unit Conversion: (Value – Mean) / SD

17

Average = 119.6 SD = 18.2

Example 1: weight ⟶ standard units

Weight = 130 oz. Convert this weight to standard units.

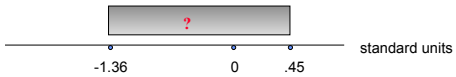
18

Average = 119.6 oz (7.5 lbs) SD = 18.2 oz (1.1 lbs)

Example 2:standard units ⟶ weight
What weight is -3 standard units?

Average = 119.6 oz (7.5 lbs) SD = 18.2 oz (1.1 lbs)

Example 3: By the normal curve approximation what percent of the babies weighed between 6 and 8 lbs?



Q: What is the area under a normal curve between -1.36 and .45?

Percentiles

There are 1236 birthweights

1236/4 = 309

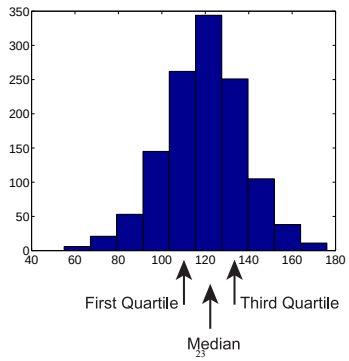
Ordered birthweights

309	**309**	**309**	**309**

The 309-th largest is called the 25-th percentile, or *first quartile.* 25% are less than it. It's value is **108.5**

The 618-th largest is called the 50-th percentile, or the *median.* 50% are less than it. It's value is **120**

The 927-th largest is called the 75-th percentile, or *the third quartile.* 75% are less than it. It's value is **131**.



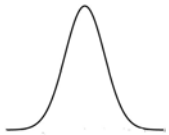
Percentiles from the normal curve approximation

From the normal curve approximation to the birthweight histogram, what is the first quartile?

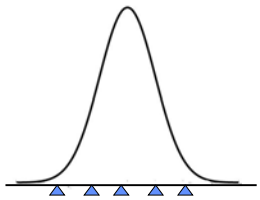
The Family of Normal Distributions

The normal curve we have been talking about is often called the "standard normal curve." Other normal curves have different averages and SDs.

Normal Applet



All Normal curves can be related to the Standard Normal by Relabeling the Axes



Group Differences Demonstration

Each applicant for a job is given a series of tests and interviews from which a numerical "suitability" score is calculated. This score is designed to predict success on the job and is used as the criterion for selecting applicants.

The applicants are from two demographic groups represented by the blue and red distributions. The same number of people are in each group. The blue distribution has a mean of 50; the red distribution has a mean of 60. Both distributions have standard deviations of 10. Applicants with scores greater than or equal to 50 are hired. The figure shows that among people hired, there are 1.683 people from the red distribution for each person from the blue distribution. This ratio increases rapidly as the cutoff increases.

The blue distribution has a mean of 50; the red distribution has a mean of 60. Both distributions have standard deviations of 10. Applicants with scores greater than or equal to 50 are hired. there are 1.683 people from the red distribution for each person from the blue distribution.

Blue: mean = 50 and SD = 10. So chance of being hired is 50%

Red: mean = 60 and SD =1 0. So 50 is –1.0 standard unit.

Area to right of 1.0 standard unit is 84% (check!)

For every 84 people hired from the Red group, 50 are hired from the Blue.

84/50 = 1.68