Chapter Six

Normal Distributions

The Normal Curve

Normal Probabilities

Percentiles and the Normal Curve

The Central Limit Theorem

The Normal Curve

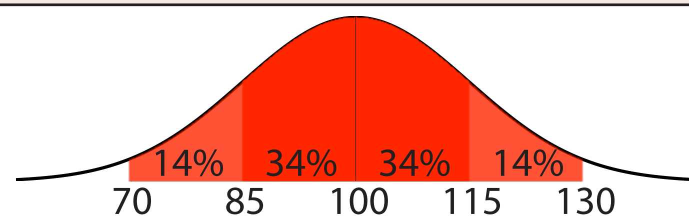
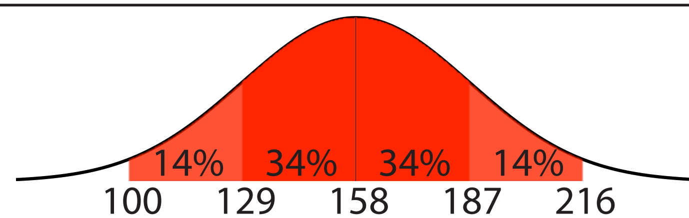
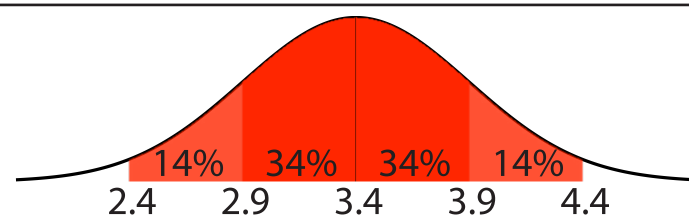
In a **normal distribution**, data values are more frequent the closer they are to the mean.

Exact frequencies of data values in a normal distribution can be calculated based on number of standard deviations above the mean, called ***z***.

68% of a normal curve is within one standard deviation of the mean ($-1 < z < 1$).

95% of a normal curve is within two standard deviations of the mean ($-2 < z < 2$).

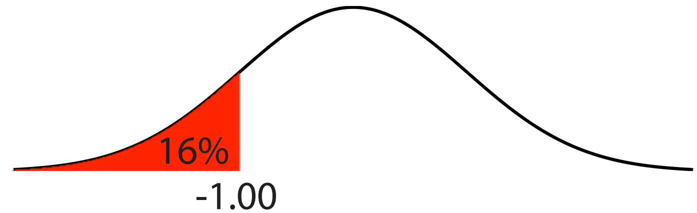
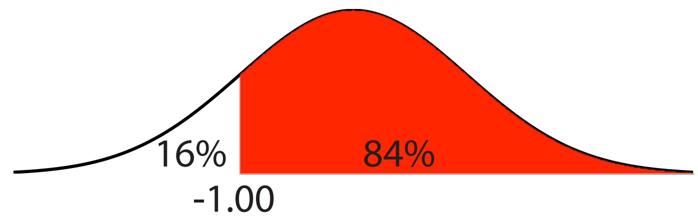
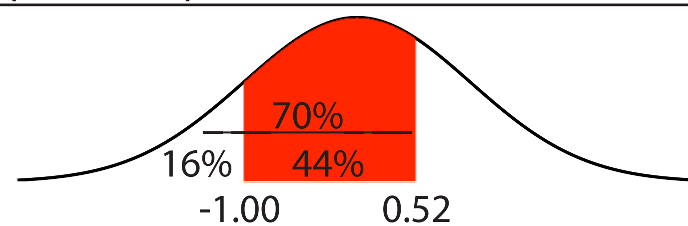
Real-world data sets typically are approximately normal, allowing for calculations of frequency or probability if the mean and standard deviation are known. Some examples are shown below.

Variable	Parameters	68% of data	Curve
IQ scores	$\mu = 100, \sigma = 15$	$85 < x < 115$	
PreCalc final	$\mu = 158, \sigma = 29$	$129 < x < 187$	
Newborn girls (kg)	$\mu = 3.4, \sigma = 0.5$	$2.9 < x < 3.9$	

Z Scores

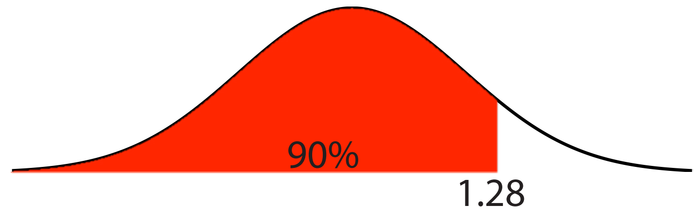
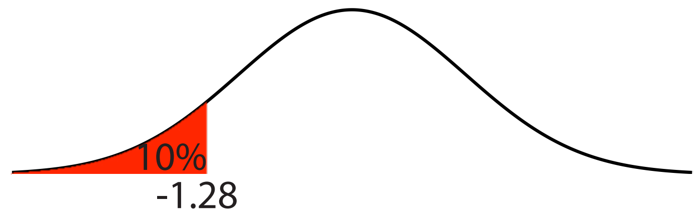
The ***z* score** of a data value is the number of standard deviations it is above the mean: $z = \frac{x-\mu}{\sigma}$.

The area under a normal curve below a given *z* score z_1 can be found by looking up z_1 in a *z* table. Since this is a probability, it is expressed with probability notation $P(z < z_1)$.

Range	Area	Explanation	Example: $z_1 = -1.00$, $z_2 = 0.52$
Below z_1	$P(z < z_1)$	given in table	 $P(z < -1.00) = $ **16%**
Above z_1	$1 - P(z < z_1)$	complement of below z_1	 $P(z > -1.00) = 100\% - 16\% = $ **84%**
Between z_1 and z_2	$P(z < z_2) - P(z < z_1)$	everything below z_2 except what is also below z_1	 $P(-1.00 < z < 0.52) = 70\% - 16\% = $ **44%**

Percentiles

A percentile can be found in a normal distribution by identifying the *z* score needed for that percentile and converting the *z* score to a raw score: $x = \mu + z\sigma$. The examples below are for SAT math scores ($\mu \approx 500$, $\sigma \approx 100$).

Percentile	Z score	Raw score	Curve
90th	$z = 1.28$	$x = 500 + 1.28(100)$ $= 628$	
10th	$z = -1.28$	$x = 500 - 1.28(100)$ $= 372$	

The Law of Large Numbers

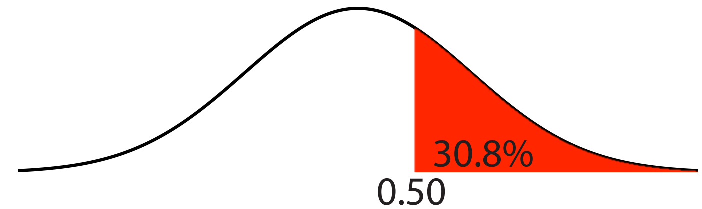
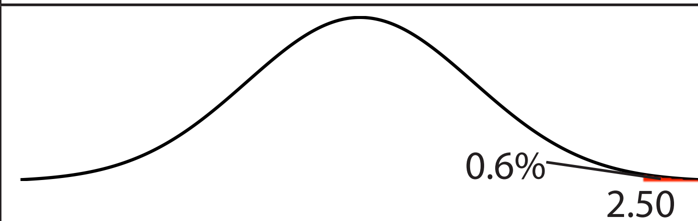
The **law of large numbers** states that the larger a sample is, the closer its statistics tend to be to the actual population parameters. As a result, as sample size increases, a sample statistic becomes more likely to fall within a given range that includes the population parameter.

Scenario example	Range Considered	Population Parameter	Effect of sample size
Percentage of heads out of *n* coins	$\hat{p} > .30$	$p = .50$	$\hat{p} > .30$ includes $p = .50$, so it becomes more likely as the sample size increases.
Average weight of freshmen boys	$80 < \bar{x} < 100$	$\mu = 112$ pounds	$80 < \bar{x} < 100$ does not include $\mu = 112$, so it becomes less likely as the sample size increases.

Sampling Distributions

A **sampling distribution** is a distribution of a statistic, such as $\bar{x}$, for a given sample size and population.

Based on the law of large numbers, the **central limit theorem** states that the larger the sample size is, the closer a sampling distribution will be to a normal distribution, and the less variation from the mean there will be in the samples. The standard deviation of the sampling distribution of means, called the **standard error** of the mean, can be estimated as $\sigma_{\bar{x}} \approx \frac{s}{\sqrt{n}}$.

Event	Z score	Probability	Curve
A random student has an SAT math score above 550.	$z = \frac{550 - 500}{100} = 0.5$	$P(x > 550)$ $= P(z > 0.5)$ $\approx 1 - .692$ $=$ **30.8%**	
The average of 25 random students' SAT math scores is above 550.	$z = \frac{550 - 500}{100 / \sqrt{25}} = 2.5$	$P(\bar{x} > 550)$ $= P(z > 2.5)$ $\approx 1 - .994$ $=$ **0.6%**	

Notice how the big difference between the two solutions demonstrates the law of large numbers.