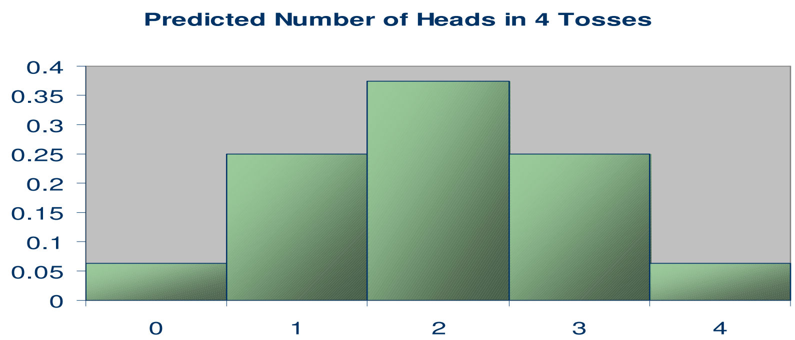
Chapter 6 Normal Probability Distributions

6-1 Overview

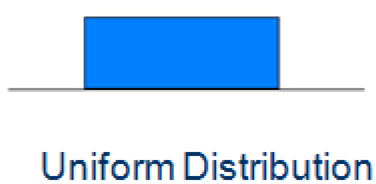
Review:

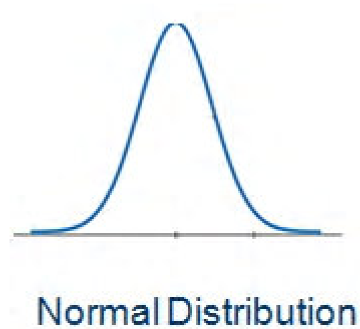
- Discrete random variable—a random variable whose values are countable or finite; a random variable whose possible values are isolated points along a number line.
- Discrete probability distribution—describes how likely each value of the random variable is to occur; can be modeled by a probability histogram.



- Continuous random variable—a random variable with an infinite number of possible values; a random variable whose possible values form an interval along the number line.
- Continuous probability distribution –a smooth curve that serves as a model for the population distribution of a continuous variable. The total area under the curve equals 1

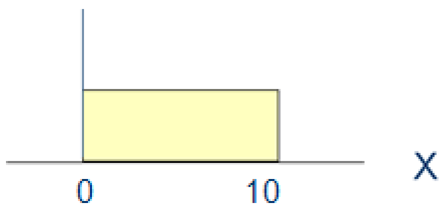
Examples of Continuous Probability Distributions





Uniform Distribution
A continuous random variable has a uniform distribution if its values spread evenly over the range of possibilities. The graph of a uniform distribution results in a rectangular shape.

Example: Suppose a morning commuter train never leaves before its scheduled departure time. The length of time that elapses between the scheduled departure time and the actual departure time is uniformly distributed between 0 and 10. Let X = length of time.

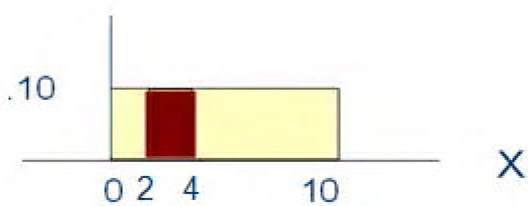


What must the height of the curve be in order for the probability to be equal to 1?

Area of a rectangle = base * height = 10 * height = 1

Therefore, the height must be 1/10 or 0.10.

Because X is a continuous random variable, it doesn't make sense to talk about the probability that X assumes an exact value, say, P(X = 3). It does make sense to talk about the probability of X being in some interval, say, $P(2 < X < 4)$.



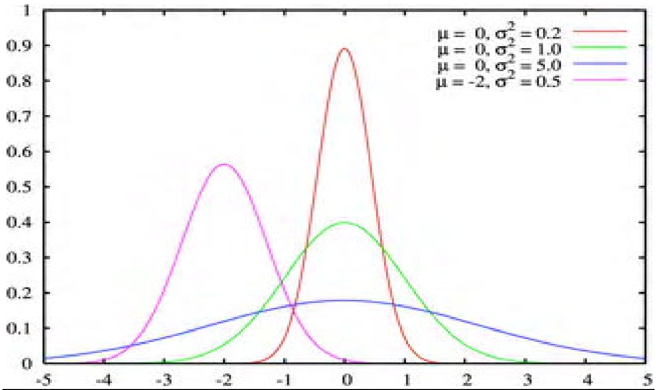
$P(2 < X < 4) = 4*0.10 = 0.4$

Normal Distribution

normal distribution—a bell-shaped, symmetric probability distribution; it is centered at the mean, μ, and its shape is determined by σ, the standard deviation.

human blood pressures, IQ scores, birth weights are normally distributed

Examples of Normal Distributions



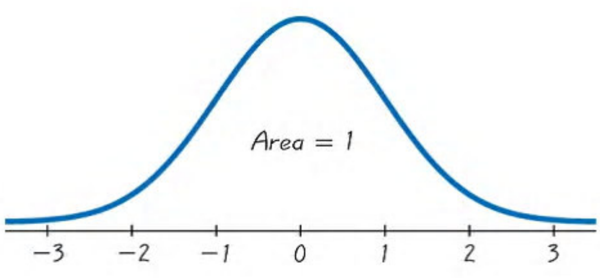
The formula for the normal distribution is given by

$$f(x) = \frac{e^{-1/2(\frac{x-\mu}{\sigma})}}{\sigma\sqrt{2\pi}}$$

6-2: The Standard Normal Distribution

standard normal distribution – a normal probability distribution that has a mean of 0 and a standard deviation of 1. The total area under the curve is equal to 1.

A standard normal random variable is denoted by z. We will be interested in finding the probability that z falls in some interval.



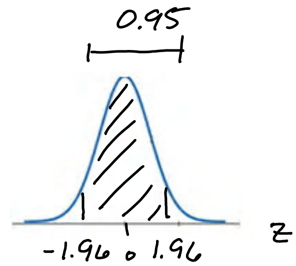
You should always draw a picture of the normal distribution when finding probabilities. Use your calculator to find the areas under the curve.

2nd → VARS → normalcdf (left bound, right bound, mean, standard deviation)

Examples:

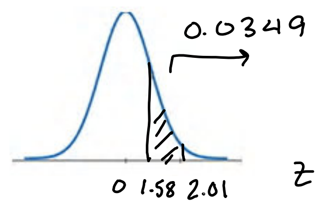
P(-1.96 < Z < 1.96)

normalcdf(-1.96, 1.96, 0, 1) = 0.95



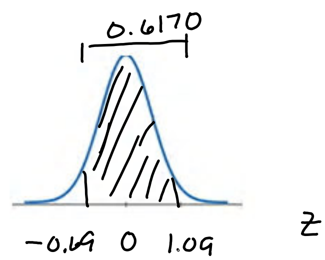
P(1.58 < Z < 2.01)

normalcdf(1.58, 2.01, 0, 1) = 0.0349



P(-0.69 < Z < 1.09)

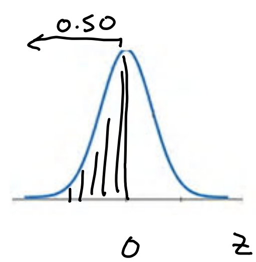
normalcdf(-0.69, 1.09, 0, 1) = 0.6170



P(Z < 0)

normalcdf(-9999, 0, 0, 1) = 0.50

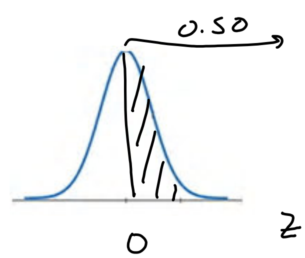
(Note: The left bound is negative infinity; just use a large negative number instead.)



P(Z > 0) = 0.50

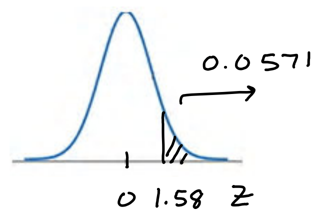
normalcdf(0, 9999, 0, 1) = 0.50

(Note: The right bound is positive infinity; just use a large positive number instead.)



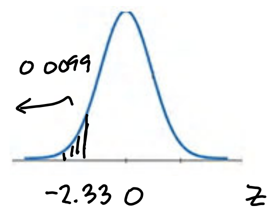
P(Z > 1.58)

normalcdf(1.58, 9999, 0, 1) = 0.0571



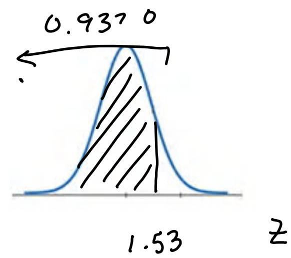
P(Z < -2.33)

normalcdf(-9999, -2.33, 0, 1) = 0.0099



P(Z < 1.53)

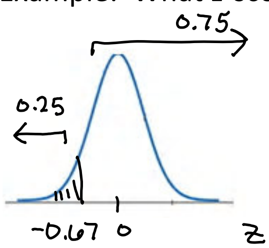
normalcdf(-9999, 1.53, 0, 1) = 0.9370



Finding Z-scores from probabilities

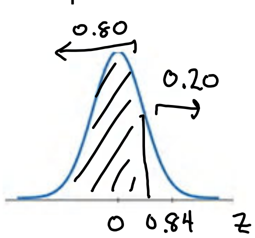
Use 2^{nd} → VARS → invnorm (area to the left, mean, standard deviation)

Example: What z-score separates the bottom 25% from the top 75%?



invnorm (0.25, 0, 1) = -0.67

Example: What z-score separates the bottom 80% from the top 20%?



invnorm (0.80, 0, 1) = 0.84

6-3 Applications of Normal Distributions

Most normal distributions are not standard normal; in other words, $\mu \neq 0$ and/or $\sigma \neq 1$.

Example: IQ scores are normally distributed with a mean of 100 and a standard deviation of 15.

Let X = IQ score. What is P(X < 130)?

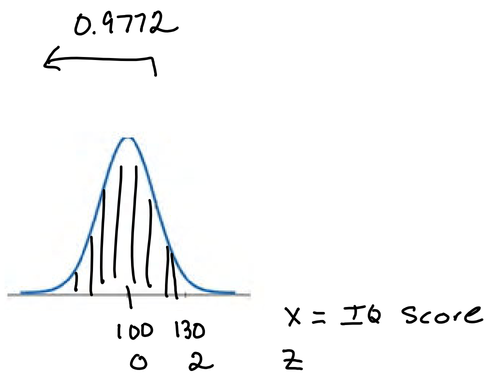
We can standardize the scores using the following formula and rounding to two decimal places:

$$z = \frac{x-\mu}{\sigma} = \frac{130-100}{15} = 2.00$$

X = 130 is 2 standard deviations above the mean.

Therefore, Z = 2.00 . P(X < 130) = P(Z < 2) = normalcdf(-9999, 2, 0, 1) = 0.9772 (from calculator)

Therefore, approximately 97.72% of individuals have IQs less than 130.

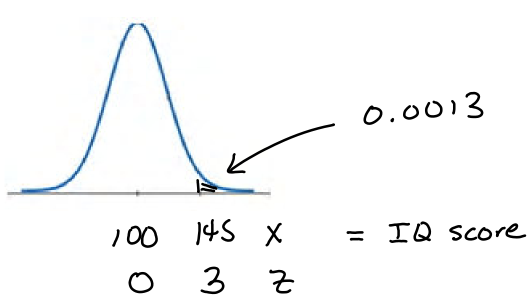


Example: P(X > 145) (the borderline for genius)

$$z = \frac{145 - 100}{15} = 3.00$$

P(X > 145) = P(Z > 3.00) = normalcdf(3, 9999, 0, 1) = 0.0013

Therefore, approximately 0.13% of the population has IQs at the genius level.

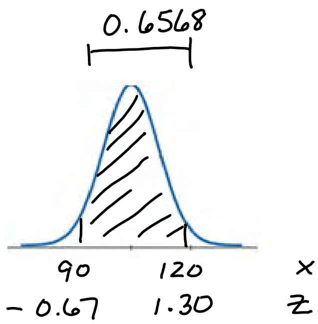


Example: P(90 < X < 110)

$$z = \frac{90-100}{15} = -0.67$$

$$z = \frac{120-100}{15} = 1.33$$

P(90 < X < 12) = P(-.67 < Z < 1.33) = normalcdf(-0.67, 1.33, 0, 1) = 0.6568



Example: The lengths of pregnancies are normally distributed with a mean of 268 days and a standard deviation of 15 days.

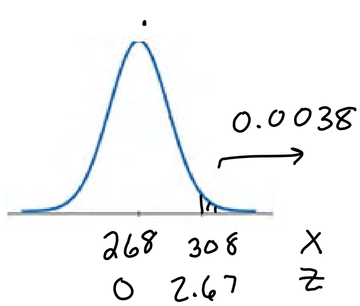
Find the probability of a pregnancy lasting 308 days or longer.

X = length of a pregnancy
μ = 268 and σ = 15

Convert X = 308 to a z score

$$z = \frac{308-268}{15} = 2.67$$

P(X > 308) = P(Z > 2.67) = normalcdf(2.67, 9999, 0, 1) = 0.0038



Example: A baby is premature if the length of pregnancy is in the lowest 4%. Find the length that separates premature babies from those who are not premature.

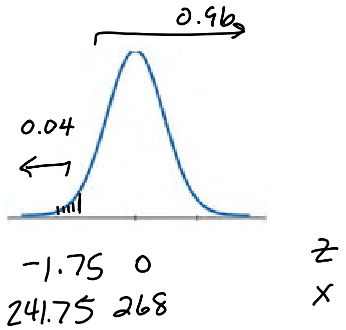
Find the z score that separates the bottom 4% from the top 96%.

z = invnorm(0.04, 0, 1) = -1.75

We need to find the x associated with a z score of -1.75.

x = $\mu + z\sigma$ = 268 + (-1.75)(15) = 241.75

Babies born before 241.75 days are considered premature



.

Example: Suppose that women's heights have a mean of 63.6 inches and a standard deviation of 2.5 inches. Find P_{10}, the height that separates the bottom 90% of women from the top (tallest) 10%.

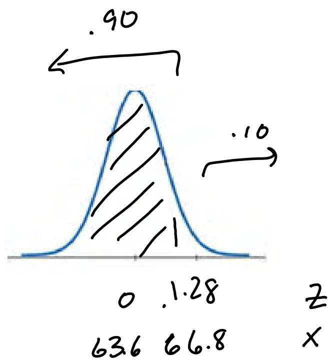
X = women's heights
μ = 63.6 and σ = 2.5

Find the z score that separates the bottom 90% from the top 10%.

z = invnorm(0.90, 0, 1) = 1.28

x = $\mu + z\sigma$ = 63.6 + (1.28)(2.5) = 66.8

10% of women are above 66.8 inches tall.



.90
.10
0
1.28
z
63.6
66.8
x

6-4 Sampling Distributions and Estimators

When we have a large population, we take a sample to estimate some population parameter, such as a population mean μ, or a population proportion, p.

Example: Suppose we're interested in the average salary of people on Cape Cod. We take a random sample of 100 individuals and compute $\overline{X}$, the average salary for the sample.

We could do this repeatedly to get an idea of which values of $\overline{X}$ are more likely and which are less likely. The average salary for the sample will <u>depend</u> on who is in the sample.

We can think of the sample mean, $\overline{X}$, as a random variable whose value changes depending on who is in the sample. Some values of $\overline{X}$ are more likely, some are less likely. Therefore *we can create a probability distribution for* $\overline{X}$ *, called a sampling distribution. It is a probability distribution of sample means.*

***The mean of the sample means is equal to the population mean, μ ***

(We are sampling with replacement, since for large populations, sampling with replacement is not significantly different than sampling without replacement. However, for this analysis, it's easier to think about these as independent events.)

6-5 The Central Limit Theorem

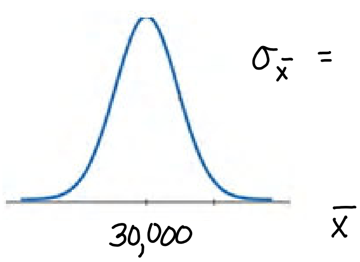
The central limit theorem involves **two** distributions—the original probability distribution of some random variable, X, and the probability distribution of sample means, $\overline{X}$.

<u>Central Limit Theorem</u>: Suppose you have a random variable X from a population with mean = μ and standard deviation σ. If you repeatedly choose simple random samples of size n and compute the sample mean, $\overline{X}$, for each sample, the distribution of $\overline{X}$ will approach a <u>normal distribution</u> with a mean $\mu_{\overline{X}}$ = μ and standard deviation $\sigma_{\overline{X}}$ = $\sigma / \sqrt{n}$

For the central limit to apply, the sample size must be greater than 30 or the original distribution must be normal (or both!)

Example: Suppose that the average salary in a certain area is $30,000 with a standard deviation of $7,000. What is the probability that you take a sample of 100 people and the average salary for the sample, the sample mean $\overline{X}$ is below $29,000.

By the central limit theorem, $\overline{X}$ is normally distributed with a mean of $\mu_{\overline{X}}$ = μ = 30,000 and a standard deviation of $\sigma_{\overline{X}}$ = $\sigma / \sqrt{n}$ = 7000/ $\sqrt{100}$ = 700

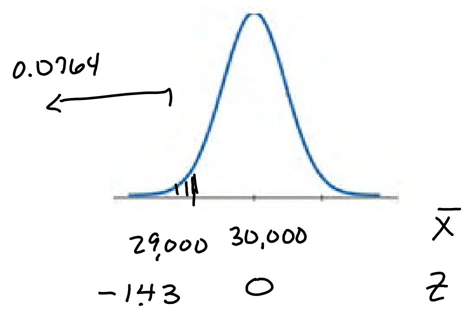


$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{7000}{\sqrt{100}} = 70$$

$$Z = \frac{29{,}000 - 30{,}000}{700} = -1.43$$

P($\overline{X}$ < 29,000) = P(Z < -1.43) = .0764

Therefore, the probability that you take a random sample of 100 people from this area and that the average salary for the sample is less than $29,000 is about 7.64%.

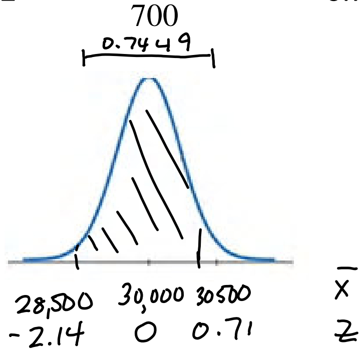


Example: What is the probability that you take a sample of 100 people and the average salary for the sample, the sample mean $\bar{X}$ is between \$28,500 and \$30,500?

P(28,500 < $\bar{X}$ < 30,500) = P(-2.14 < Z < 0.71) = 0.7449

Z = $\dfrac{28{,}500 - 30{,}000}{7{,}000/\sqrt{100}}$ = -2.14

Z = $\dfrac{30{,}500 - 30{,}000}{700}$ 0.71



Example: Suppose x = the fuel efficiency for a 2005 Toyota Camry, and x is distributed with μ = 24 mpg and σ = 3. What is the probability that we take a sample of 50 Camrys and the average fuel efficiency for the sample, $\bar{X}$, is less than 23 mpg? In other words, what is P($\bar{X}$ < 23)?

$\mu_{\bar{X}}$ = μ = 24 mpg

$\sigma_{\bar{X}}$ = $\sigma/\sqrt{n}$ = 3/$\sqrt{50}$ = 0.4243 mpg.

z = (23 – 24)/ 0.4243 = -2.36

P($\bar{X}$ < 23) = P(Z < -2.36) = 0.0091

Thus, there is a 0.91% probability that for a sample of 50 cars, the average fuel efficiency for that sample would be less than 23 mph.