# Introduction to Statistics Tutorial: Probability and Probability Distributions Part 2

INCOGEN, Inc.

*2008*

# Outline

- Binomial Distribution
- Normal Distribution
- Case Study

# Binomial Probability Distribution

A binomial experiment has the following properties:

1. The experiment must have a *fixed number of trials.*
2. The trials must be *independent*. (The outcome of any individual trial doesn't affect the probabilities in the other trials.)
3. Each trial results in one of two outcomes (Typically, the outcomes are classified as a success or failure.)
4. The probabilities must remain *constant* for each trial.
5. We are interested in the number of successes during the trials.

INCOGEN

# Notation for the Binomial Probability Distribution

*n* = fixed number of trials

*x* = specific number of successes in *n* trials

*p* = probability of *success* in *a single* trial

*q* = probability of *failure* in *a single* trial

($q = 1 - p$)

p(*x*) = probability of getting exactly *x* successes among *n* trials

Be sure that *x* and *p* both refer to the same category being called a success.

# Formula for the Binomial Probability Distribution

$$P(x) = \frac{n!}{(n-x)!\,x!} \cdot p^{x} \cdot q^{n-x}$$

**Number of outcomes with exactly *x* successes among *n* trials**

**Probability of *x* successes among *n* trials for any one particular order**

# Binomial Distribution Example

Find the probability of having five left-handed students in a class of twenty-five. Assume that the probability of being left-handed is 10% or 0.1

Solution:

$n = 25$, $x = 5$, $p = 0.1$

$$P(5) = \frac{25!}{(25-5)! \cdot 5!} \cdot (0.1)^5 \cdot (0.9)^{20} = 0.064593$$

INCOGEN

# Mean and Variance of the Binomial Distribution

**Mean** $\mu_X = np$

**Variance** $\sigma_X^2 = np(1-p)$

Example: Flip a coin 8 times. If the coin is fair, then *p* = 0.5.

The mean number of heads = *np* = 8*0.5 = 4.

The variance is equal to *np(1-p)* = 8*0.5*0.5 = 2.

INCOGEN

# Continuous Random Variables

- Assume the infinitely many values corresponding to points on a line interval
- Can not assign a positive probability to each of these uncountable values
- Use probability curves or distributions – the probability equals the area under the curve

INCOGEN

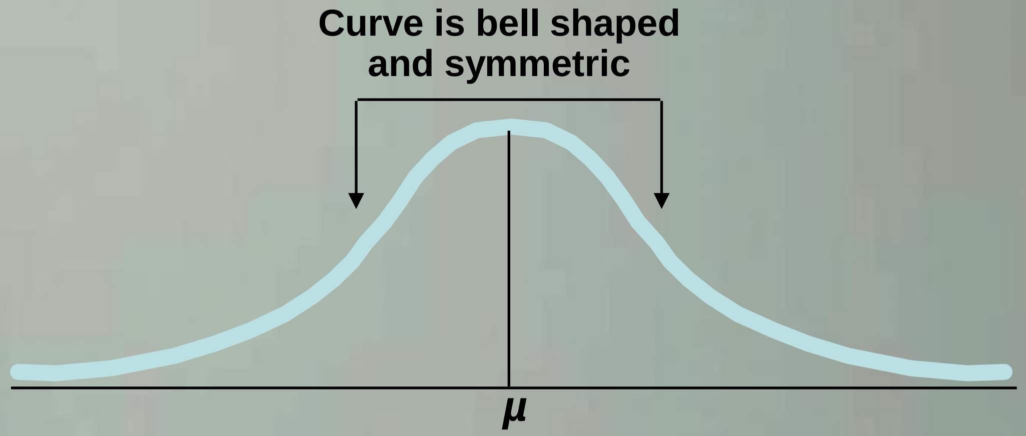
# Normal Probability Distribution

Many times our data has a mound-shaped frequency distribution. When the data is also continuous, we use a normal probability distribution to approximate it.

INCOGEN

# Normal Probability Distribution

**The total area under the normal probability curve is equal to 1.**

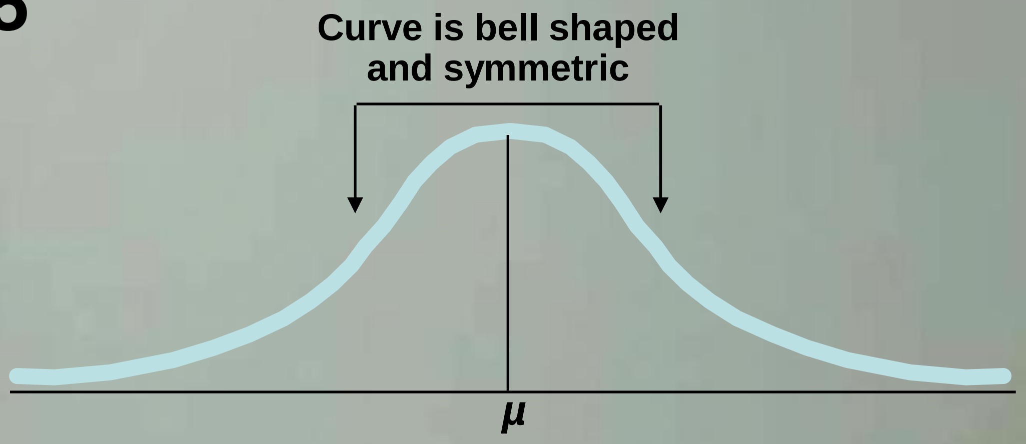



**There is a correspondence between *area* and *probability*.**

# Normal Probability Distribution

**The mean $\mu$ locates the center of the distribution.**

**Area to the right of the mean equals 0.5**




**Area to left of the mean equals 0.5**

# Normal Probability Distribution

We use a table to calculate the area under the curve in which we may be interested.

Since the area to the right of the mean is 0.5, $P(X > \mu) = 0.5$

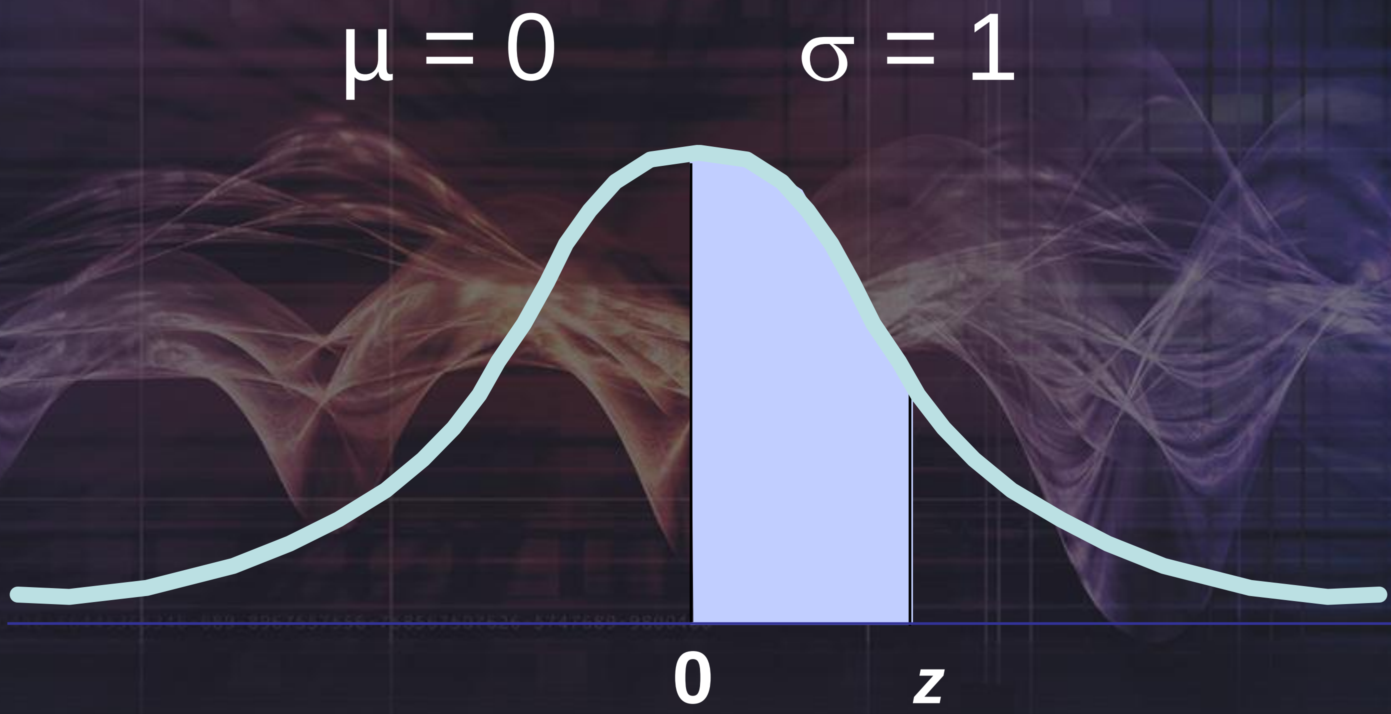
Since probability = area, the $P(X = a) = 0$ where a is a point. (There is no area to a point.)

INCOGEN

# Standard Normal Probability Distribution

A normal probability distribution that has a mean of 0 ($\mu=0$) and a standard deviation of 1 ($\sigma=1$)

# Use the
# Standard Normal Distribution Table



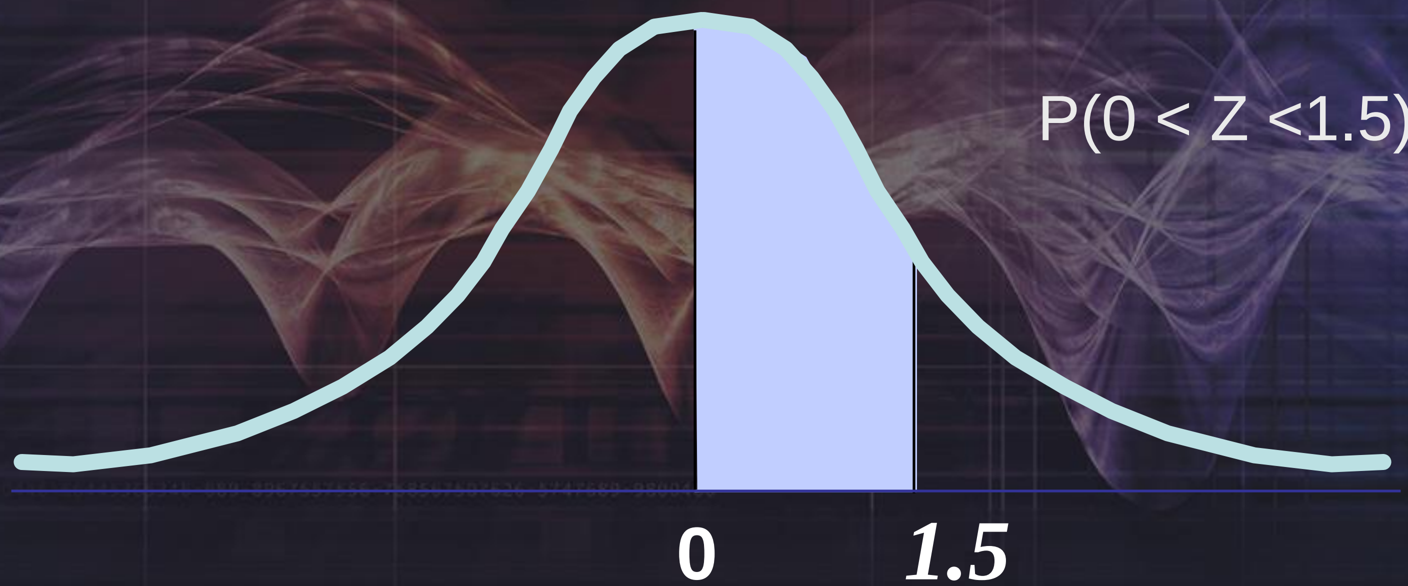

INCOGEN

# Standard Normal Distribution Table

| z | 0 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0.004 | 0.008 | 0.012 | 0.016 | 0.0199 | 0.0239 | 0.0279 | 0.0319 | 0.0359 |
| 0.1 | 0.0398 | 0.0438 | 0.0478 | 0.0517 | 0.0557 | 0.0596 | 0.0636 | 0.0675 | 0.0714 | 0.0753 |
| 0.2 | 0.0793 | 0.0832 | 0.0871 | 0.091 | 0.0948 | 0.0987 | 0.1026 | 0.1064 | 0.1103 | 0.1141 |
| 0.3 | 0.1179 | 0.1217 | 0.1255 | 0.1293 | 0.1331 | 0.1368 | 0.1406 | 0.1443 | 0.148 | 0.1517 |
| 0.4 | 0.1554 | 0.1591 | 0.1628 | 0.1664 | 0.17 | 0.1736 | 0.1772 | 0.1808 | 0.1844 | 0.1879 |
| 0.5 | 0.1915 | 0.195 | 0.1985 | 0.2019 | 0.2054 | 0.2088 | 0.2123 | 0.2157 | 0.219 | 0.2224 |
| 0.6 | 0.2257 | 0.2291 | 0.2324 | 0.2357 | 0.2389 | 0.2422 | 0.2454 | 0.2486 | 0.2517 | 0.2549 |
| 0.7 | 0.258 | 0.2611 | 0.2642 | 0.2673 | 0.2704 | 0.2734 | 0.2764 | 0.2794 | 0.2823 | 0.2852 |
| 0.8 | 0.2881 | 0.291 | 0.2939 | 0.2967 | 0.2995 | 0.3023 | 0.3051 | 0.3078 | 0.3106 | 0.3133 |
| 0.9 | 0.3159 | 0.3186 | 0.3212 | 0.3238 | 0.3264 | 0.3289 | 0.3315 | 0.334 | 0.3365 | 0.3389 |
| 1 | 0.3413 | 0.3438 | 0.3461 | 0.3485 | 0.3508 | 0.3531 | 0.3554 | 0.3577 | 0.3599 | 0.3621 |
| 1.1 | 0.3643 | 0.3665 | 0.3686 | 0.3708 | 0.3729 | 0.3749 | 0.377 | 0.379 | 0.381 | 0.383 |
| 1.2 | 0.3849 | 0.3869 | 0.3888 | 0.3907 | 0.3925 | 0.3944 | 0.3962 | 0.398 | 0.3997 | 0.4015 |
| 1.3 | 0.4032 | 0.4049 | 0.4066 | 0.4082 | 0.4099 | 0.4115 | 0.4131 | 0.4147 | 0.4162 | 0.4177 |
| 1.4 | 0.4192 | 0.4207 | 0.4222 | 0.4236 | 0.4251 | 0.4265 | 0.4279 | 0.4292 | 0.4306 | 0.4319 |
| 1.5 | 0.4332 | 0.4345 | 0.4357 | 0.437 | 0.4382 | 0.4394 | 0.4406 | 0.4418 | 0.4429 | 0.4441 |
| 1.6 | 0.4452 | 0.4463 | 0.4474 | 0.4484 | 0.4495 | 0.4505 | 0.4515 | 0.4525 | 0.4535 | 0.4545 |

# Use the
# Standard Normal Distribution Table

$\mu = 0$ $\sigma = 1$




P(0 < Z <1.5) = ?

INCOGEN

# Standard Normal Distribution Table

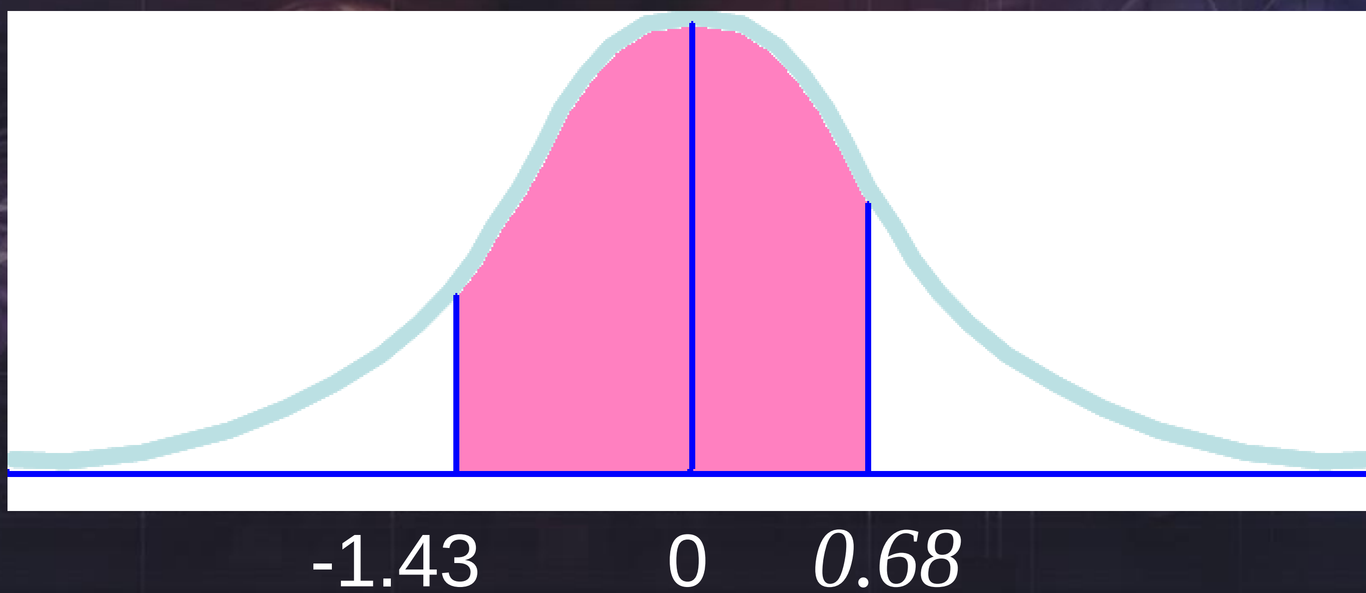
| z | 0 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0.004 | 0.008 | 0.012 | 0.016 | 0.0199 | 0.0239 | 0.0279 | 0.0319 | 0.0359 |
| 0.1 | 0.0398 | 0.0438 | 0.0478 | 0.0517 | 0.0557 | 0.0596 | 0.0636 | 0.0675 | 0.0714 | 0.0753 |
| 0.2 | 0.0793 | 0.0832 | 0.0871 | 0.091 | 0.0948 | 0.0987 | 0.1026 | 0.1064 | 0.1103 | 0.1141 |
| 0.3 | 0.1179 | 0.1217 | 0.1255 | 0.1293 | 0.1331 | 0.1368 | 0.1406 | 0.1443 | 0.148 | 0.1517 |
| 0.4 | 0.1554 | 0.1591 | 0.1628 | 0.1664 | 0.17 | 0.1736 | 0.1772 | 0.1808 | 0.1844 | 0.1879 |
| 0.5 | 0.1915 | 0.195 | 0.1985 | 0.2019 | 0.2054 | 0.2088 | 0.2123 | 0.2157 | 0.219 | 0.2224 |
| 0.6 | 0.2257 | 0.2291 | 0.2324 | 0.2357 | 0.2389 | 0.2422 | 0.2454 | 0.2486 | 0.2517 | 0.2549 |
| 0.7 | 0.258 | 0.2611 | 0.2642 | 0.2673 | 0.2704 | 0.2734 | 0.2764 | 0.2794 | 0.2823 | 0.2852 |
| 0.8 | 0.2881 | 0.291 | 0.2939 | 0.2967 | 0.2995 | 0.3023 | 0.3051 | 0.3078 | 0.3106 | 0.3133 |
| 0.9 | 0.3159 | 0.3186 | 0.3212 | 0.3238 | 0.3264 | 0.3289 | 0.3315 | 0.334 | 0.3365 | 0.3389 |
| 1 | 0.3413 | 0.3438 | 0.3461 | 0.3485 | 0.3508 | 0.3531 | 0.3554 | 0.3577 | 0.3599 | 0.3621 |
| 1.1 | 0.3643 | 0.3665 | 0.3686 | 0.3708 | 0.3729 | 0.3749 | 0.377 | 0.379 | 0.381 | 0.383 |
| 1.2 | 0.3849 | 0.3869 | 0.3888 | 0.3907 | 0.3925 | 0.3944 | 0.3962 | 0.398 | 0.3997 | 0.4015 |
| 1.3 | 0.4032 | 0.4049 | 0.4066 | 0.4082 | 0.4099 | 0.4115 | 0.4131 | 0.4147 | 0.4162 | 0.4177 |
| 1.4 | 0.4192 | 0.4207 | 0.4222 | 0.4236 | 0.4251 | 0.4265 | 0.4279 | 0.4292 | 0.4306 | 0.4319 |
| 1.5 | 0.4332 | 0.4345 | 0.4357 | 0.437 | 0.4382 | 0.4394 | 0.4406 | 0.4418 | 0.4429 | 0.4441 |
| 1.6 | 0.4452 | 0.4463 | 0.4474 | 0.4484 | 0.4495 | 0.4505 | 0.4515 | 0.4525 | 0.4535 | 0.4545 |

So the $P(0 < Z < 1.5) = 0.4332$

INCOGEN

# Example:  Standard Normal

Using the standard normal table, calculate the area under the normal curve between z = -1.43 and z= 0.68; that is, find P( -1.43 < z < 0.68)



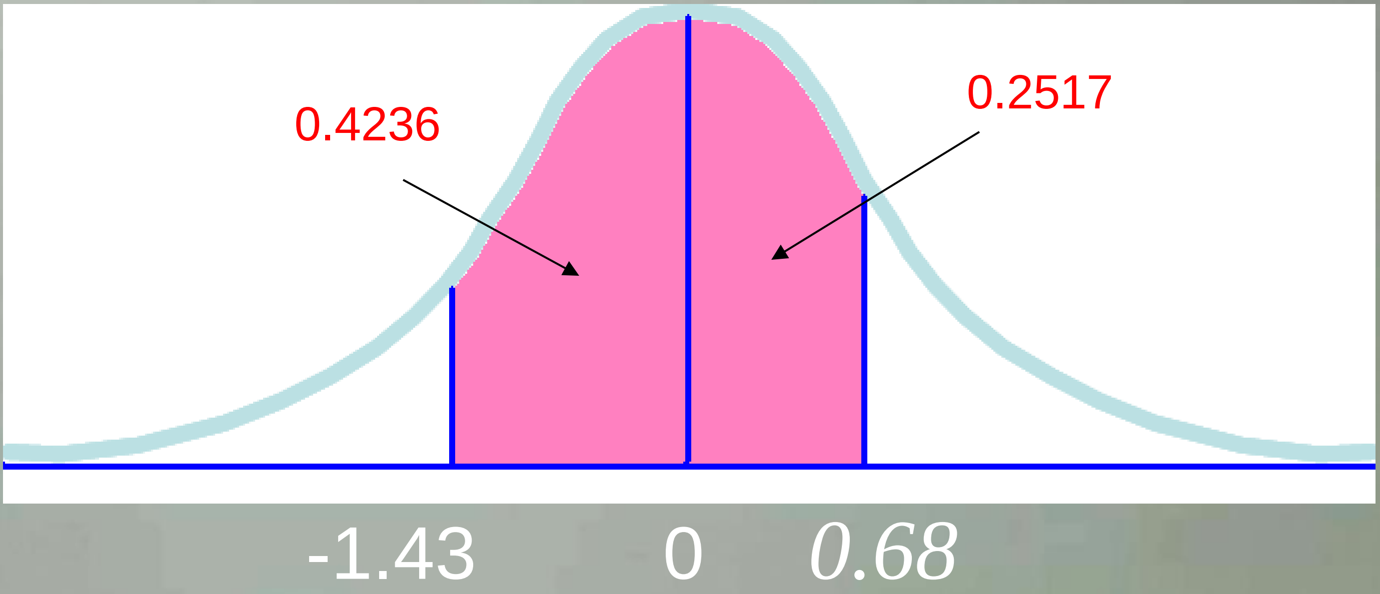
# Standard Normal Distribution Table

| z | 0 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0.004 | 0.008 | 0.012 | 0.016 | 0.0199 | 0.0239 | 0.0279 | 0.0319 | 0.0359 |
| 0.1 | 0.0398 | 0.0438 | 0.0478 | 0.0517 | 0.0557 | 0.0596 | 0.0636 | 0.0675 | 0.0714 | 0.0753 |
| 0.2 | 0.0793 | 0.0832 | 0.0871 | 0.091 | 0.0948 | 0.0987 | 0.1026 | 0.1064 | 0.1103 | 0.1141 |
| 0.3 | 0.1179 | 0.1217 | 0.1255 | 0.1293 | 0.1331 | 0.1368 | 0.1406 | 0.1443 | 0.148 | 0.1517 |
| 0.4 | 0.1554 | 0.1591 | 0.1628 | 0.1664 | 0.17 | 0.1736 | 0.1772 | 0.1808 | 0.1844 | 0.1879 |
| 0.5 | 0.1915 | 0.195 | 0.1985 | 0.2019 | 0.2054 | 0.2088 | 0.2123 | 0.2157 | 0.219 | 0.2224 |
| 0.6 | 0.2257 | 0.2291 | 0.2324 | 0.2357 | 0.2389 | 0.2422 | 0.2454 | 0.2486 | 0.2517 | 0.2549 |
| 0.7 | 0.258 | 0.2611 | 0.2642 | 0.2673 | 0.2704 | 0.2734 | 0.2764 | 0.2794 | 0.2823 | 0.2852 |
| 0.8 | 0.2881 | 0.291 | 0.2939 | 0.2967 | 0.2995 | 0.3023 | 0.3051 | 0.3078 | 0.3106 | 0.3133 |
| 0.9 | 0.3159 | 0.3186 | 0.3212 | 0.3238 | 0.3264 | 0.3289 | 0.3315 | 0.334 | 0.3365 | 0.3389 |
| 1 | 0.3413 | 0.3438 | 0.3461 | 0.3485 | 0.3508 | 0.3531 | 0.3554 | 0.3577 | 0.3599 | 0.3621 |
| 1.1 | 0.3643 | 0.3665 | 0.3686 | 0.3708 | 0.3729 | 0.3749 | 0.377 | 0.379 | 0.381 | 0.383 |
| 1.2 | 0.3849 | 0.3869 | 0.3888 | 0.3907 | 0.3925 | 0.3944 | 0.3962 | 0.398 | 0.3997 | 0.4015 |
| 1.3 | 0.4032 | 0.4049 | 0.4066 | 0.4082 | 0.4099 | 0.4115 | 0.4131 | 0.4147 | 0.4162 | 0.4177 |
| 1.4 | 0.4192 | 0.4207 | 0.4222 | 0.4236 | 0.4251 | 0.4265 | 0.4279 | 0.4292 | 0.4306 | 0.4319 |
| 1.5 | 0.4332 | 0.4345 | 0.4357 | 0.437 | 0.4382 | 0.4394 | 0.4406 | 0.4418 | 0.4429 | 0.4441 |
| 1.6 | 0.4452 | 0.4463 | 0.4474 | 0.4484 | 0.4495 | 0.4505 | 0.4515 | 0.4525 | 0.4535 | 0.4545 |

# Example: Standard Normal

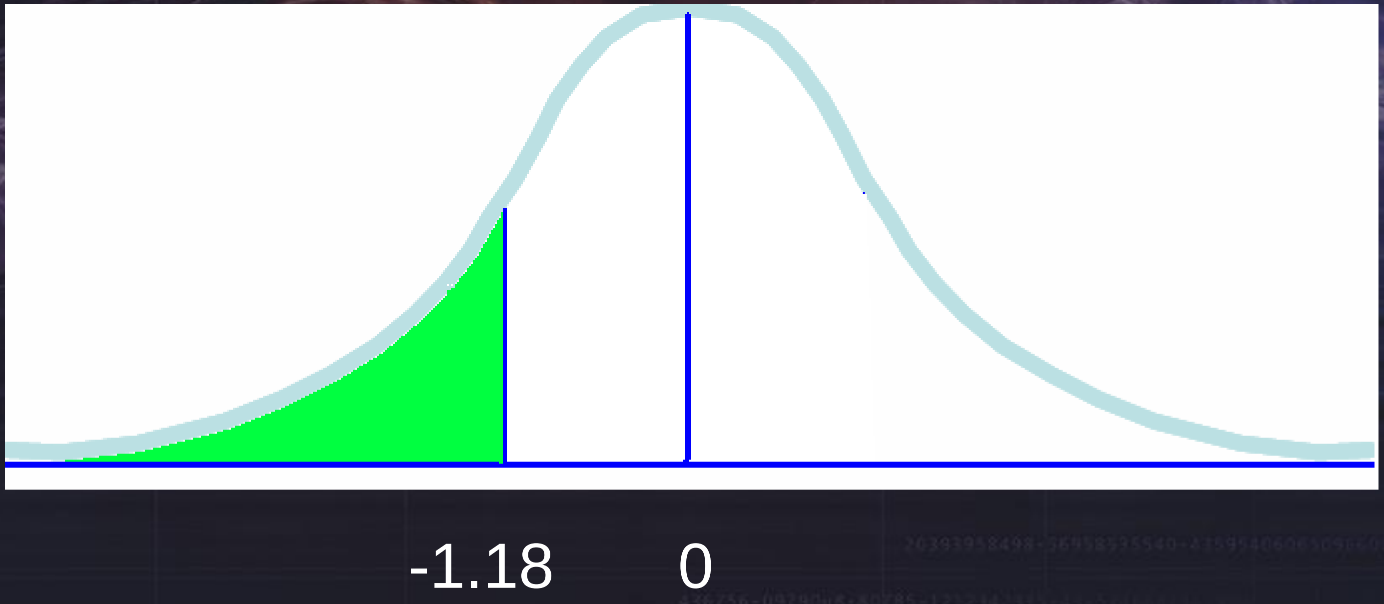
The area from -1.43 to 0 is 0.4236 and the area from 0 to 0.68 is 0.2517

So P( -1.43 < z < 0.68) = 0.4236 + 0.2517 =0.6753



# Example:  Standard Normal

Using the standard normal table, calculate the area under the normal curve that is less than -1.18; that is, find $P(z < -1.18)$




INCOGEN

# Standard Normal Distribution Table

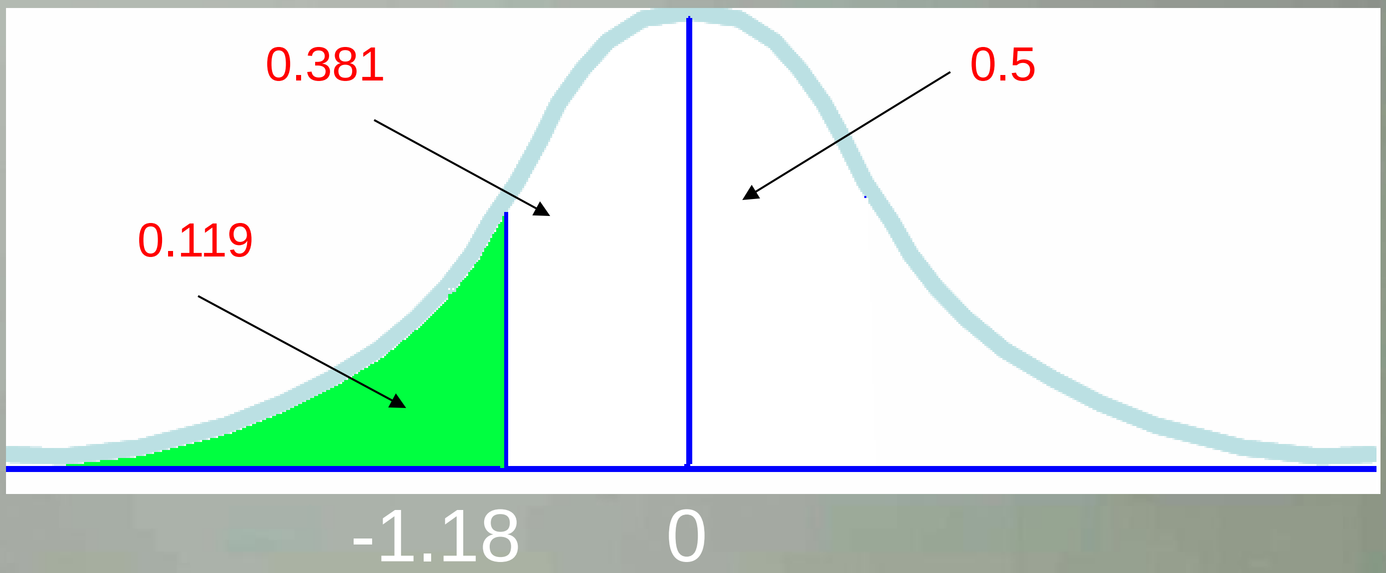
| z | 0 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---:|---:|---:|---:|---:|---:|---:|---:|---:|---:|---:|
| **0** | 0 | 0.004 | 0.008 | 0.012 | 0.016 | 0.0199 | 0.0239 | 0.0279 | 0.0319 | 0.0359 |
| **0.1** | 0.0398 | 0.0438 | 0.0478 | 0.0517 | 0.0557 | 0.0596 | 0.0636 | 0.0675 | 0.0714 | 0.0753 |
| **0.2** | 0.0793 | 0.0832 | 0.0871 | 0.091 | 0.0948 | 0.0987 | 0.1026 | 0.1064 | 0.1103 | 0.1141 |
| **0.3** | 0.1179 | 0.1217 | 0.1255 | 0.1293 | 0.1331 | 0.1368 | 0.1406 | 0.1443 | 0.148 | 0.1517 |
| **0.4** | 0.1554 | 0.1591 | 0.1628 | 0.1664 | 0.17 | 0.1736 | 0.1772 | 0.1808 | 0.1844 | 0.1879 |
| **0.5** | 0.1915 | 0.195 | 0.1985 | 0.2019 | 0.2054 | 0.2088 | 0.2123 | 0.2157 | 0.219 | 0.2224 |
| **0.6** | 0.2257 | 0.2291 | 0.2324 | 0.2357 | 0.2389 | 0.2422 | 0.2454 | 0.2486 | 0.2517 | 0.2549 |
| **0.7** | 0.258 | 0.2611 | 0.2642 | 0.2673 | 0.2704 | 0.2734 | 0.2764 | 0.2794 | 0.2823 | 0.2852 |
| **0.8** | 0.2881 | 0.291 | 0.2939 | 0.2967 | 0.2995 | 0.3023 | 0.3051 | 0.3078 | 0.3106 | 0.3133 |
| **0.9** | 0.3159 | 0.3186 | 0.3212 | 0.3238 | 0.3264 | 0.3289 | 0.3315 | 0.334 | 0.3365 | 0.3389 |
| **1** | 0.3413 | 0.3438 | 0.3461 | 0.3485 | 0.3508 | 0.3531 | 0.3554 | 0.3577 | 0.3599 | 0.3621 |
| **1.1** | 0.3643 | 0.3665 | 0.3686 | 0.3708 | 0.3729 | 0.3749 | 0.377 | 0.379 | 0.381 | 0.383 |
| **1.2** | 0.3849 | 0.3869 | 0.3888 | 0.3907 | 0.3925 | 0.3944 | 0.3962 | 0.398 | 0.3997 | 0.4015 |
| **1.3** | 0.4032 | 0.4049 | 0.4066 | 0.4082 | 0.4099 | 0.4115 | 0.4131 | 0.4147 | 0.4162 | 0.4177 |
| **1.4** | 0.4192 | 0.4207 | 0.4222 | 0.4236 | 0.4251 | 0.4265 | 0.4279 | 0.4292 | 0.4306 | 0.4319 |
| **1.5** | 0.4332 | 0.4345 | 0.4357 | 0.437 | 0.4382 | 0.4394 | 0.4406 | 0.4418 | 0.4429 | 0.4441 |
| **1.6** | 0.4452 | 0.4463 | 0.4474 | 0.4484 | 0.4495 | 0.4505 | 0.4515 | 0.4525 | 0.4535 | 0.4545 |

INCOGEN

# Example: Standard Normal

The area from 0 to 1.18 is 0.381. Since the curve is symmetric, the area from -1.18 to 0 is also 0.381. The area under each half of the curve is 0.5; therefore,

P( z < -1.18) = 0.5 – 0.381 = 0.119



# Normal Probability Distribution

Any Normal distribution can be standardized to use the standard normal tables.

Let x be normally distributed with mean of 10 and a standard deviation of 2.

To find $P(X > 11)$, we convert it to a standard normal probability using:

$$z = \frac{x - \mu}{\sigma}$$

INCOGEN

# Normal Probability Distribution

Let x be normally distributed with mean of 10 and a standard deviation of 2.

$$z = \frac{x - \mu}{\sigma}$$

Z = (11 – 10)/2 = 0.5

so

P(X > 11) = P(Z > 0.5)

Then we use the standard normal table to calculate the probability.

P(Z > 0.5) = 0.3085

INCOGEN

# Standard Normal Distribution Table

| z | 0 | 0.01 | 0.02 | 0.03 | 0.04 | 0.05 | 0.06 | 0.07 | 0.08 | 0.09 |
|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 0 | 0.004 | 0.008 | 0.012 | 0.016 | 0.0199 | 0.0239 | 0.0279 | 0.0319 | 0.0359 |
| 0.1 | 0.0398 | 0.0438 | 0.0478 | 0.0517 | 0.0557 | 0.0596 | 0.0636 | 0.0675 | 0.0714 | 0.0753 |
| 0.2 | 0.0793 | 0.0832 | 0.0871 | 0.091 | 0.0948 | 0.0987 | 0.1026 | 0.1064 | 0.1103 | 0.1141 |
| 0.3 | 0.1179 | 0.1217 | 0.1255 | 0.1293 | 0.1331 | 0.1368 | 0.1406 | 0.1443 | 0.148 | 0.1517 |
| 0.4 | 0.1554 | 0.1591 | 0.1628 | 0.1664 | 0.17 | 0.1736 | 0.1772 | 0.1808 | 0.1844 | 0.1879 |
| 0.5 | 0.1915 | 0.195 | 0.1985 | 0.2019 | 0.2054 | 0.2088 | 0.2123 | 0.2157 | 0.219 | 0.2224 |
| 0.6 | 0.2257 | 0.2291 | 0.2324 | 0.2357 | 0.2389 | 0.2422 | 0.2454 | 0.2486 | 0.2517 | 0.2549 |
| 0.7 | 0.258 | 0.2611 | 0.2642 | 0.2673 | 0.2704 | 0.2734 | 0.2764 | 0.2794 | 0.2823 | 0.2852 |
| 0.8 | 0.2881 | 0.291 | 0.2939 | 0.2967 | 0.2995 | 0.3023 | 0.3051 | 0.3078 | 0.3106 | 0.3133 |
| 0.9 | 0.3159 | 0.3186 | 0.3212 | 0.3238 | 0.3264 | 0.3289 | 0.3315 | 0.334 | 0.3365 | 0.3389 |
| 1 | 0.3413 | 0.3438 | 0.3461 | 0.3485 | 0.3508 | 0.3531 | 0.3554 | 0.3577 | 0.3599 | 0.3621 |
| 1.1 | 0.3643 | 0.3665 | 0.3686 | 0.3708 | 0.3729 | 0.3749 | 0.377 | 0.379 | 0.381 | 0.383 |
| 1.2 | 0.3849 | 0.3869 | 0.3888 | 0.3907 | 0.3925 | 0.3944 | 0.3962 | 0.398 | 0.3997 | 0.4015 |
| 1.3 | 0.4032 | 0.4049 | 0.4066 | 0.4082 | 0.4099 | 0.4115 | 0.4131 | 0.4147 | 0.4162 | 0.4177 |
| 1.4 | 0.4192 | 0.4207 | 0.4222 | 0.4236 | 0.4251 | 0.4265 | 0.4279 | 0.4292 | 0.4306 | 0.4319 |
| 1.5 | 0.4332 | 0.4345 | 0.4357 | 0.437 | 0.4382 | 0.4394 | 0.4406 | 0.4418 | 0.4429 | 0.4441 |
| 1.6 | 0.4452 | 0.4463 | 0.4474 | 0.4484 | 0.4495 | 0.4505 | 0.4515 | 0.4525 | 0.4535 | 0.4545 |

$P(Z > 0.5) = 0.3085$

# Case Study

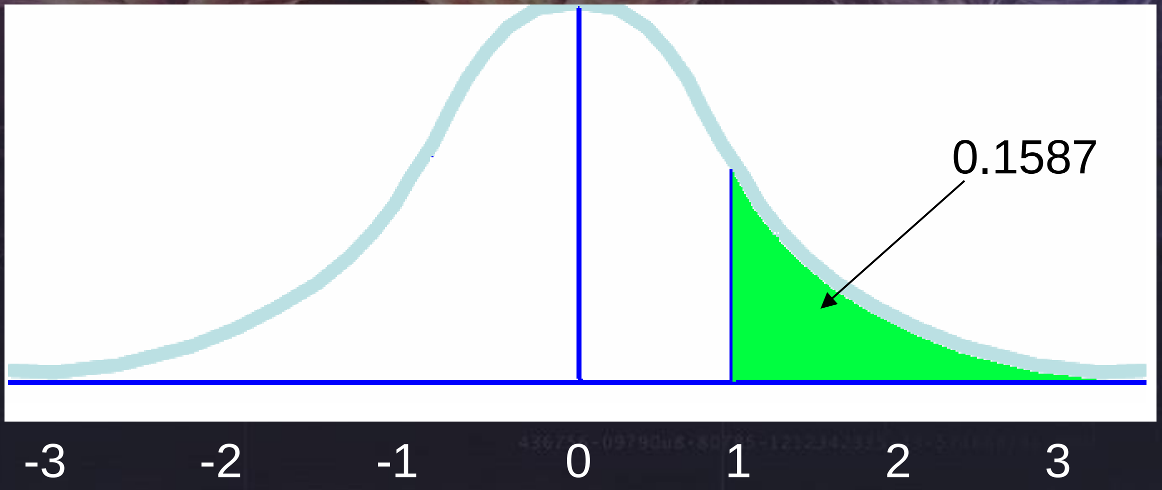
"An individual's BP (Blood Pressure) has important health implications; hypertension is among the most commonly treated chronic medical problems.  To examine variation in BP, Marczak and Paprocki (2001) found the mean and standard deviation in a group of healthy persons.  For men and women between the ages of 14 and 70, mean 24-h systolic pressure was 119.7 mm Hg, and the standard deviation was 10.9.  We use this information to calculate probabilities of any patient having a given BP." (Dawson and Trapp, 2004)

INCOGEN

# Case Study

Assuming systolic BP in normal healthy individuals is normally distributed with mean = 120 and sd = 10. (The actual study found these to be 119.7 and 10.7, respectively.) Let's find the area of the curve above 130 mm Hg?

$P(X > 130) = P(Z > (130-120)/10) = P(Z>1) = 0.1587$

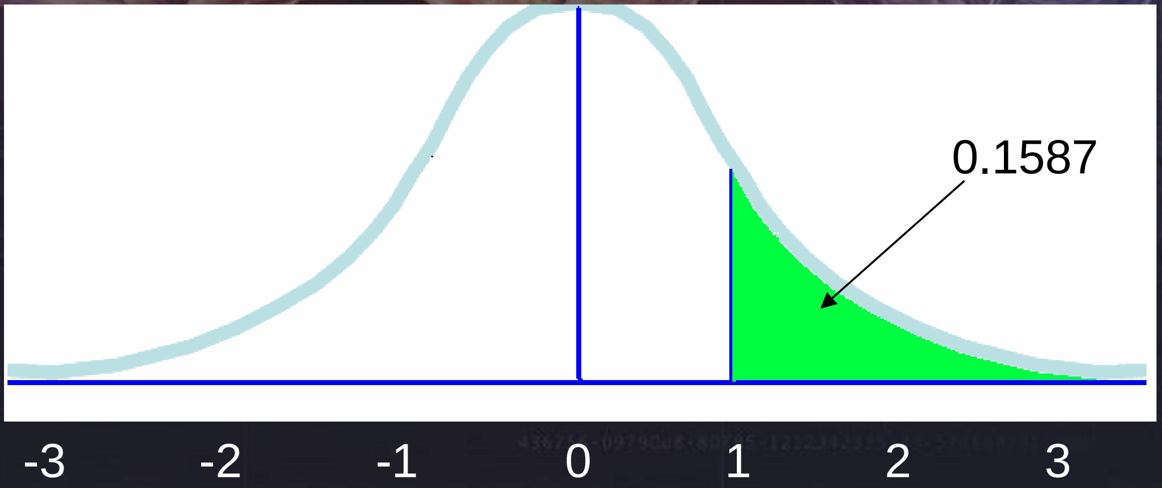


# Case Study

So what does this mean?

$P(Z>1) = 0.1587$

15.9% of normal healthy individuals have a systolic BP above 1 standard deviation; that is, above 130 mm Hg.



# References

This tutorial is comprised of materials from the following sources:

Introduction to Probability and Statistics by Mendenhall and Beaver. ITP/Duxbury.

The Cartoon Guide to Statistics by Gonick and Smith. HarperCollins.

Elementary Statistics by Triola. Addison-Wesley-Longman

Basic Statistics: an abbreviated overview by Ackerman, Bartz, and Deville. 2006 Accountability Conference

Basic and Clinical Biostatistics by Beth Dawson and Robert Trapp. McGraw-Hill, 2004.

INCOGEN