The Normal Distribution

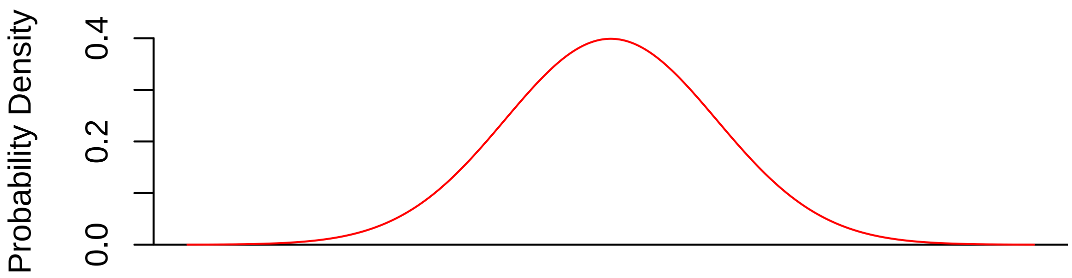
Mary Lindstrom
(Adapted from notes provided by Professor Bret Larget)

February 10, 2004

The Normal Distribution

- The Normal Distribution (AKA Gaussian Distribution) is our first distribution for continuous variables and is the most commonly used distribution.

- The normal density curve is the famous symmetric, bell-shaped curve.



Central Limit Theorem

Why is it such an important distribution? Two related reasons:

1. The central limit theorem states that many statistics we calculate from large random samples will have approximate normal distributions (or distributions derived from normal distributions), even if the distributions of the underlying variables are not normally distributed.

2. Many measured variables nave approximately normal distributions. This is true because most things we measure are the sum of many smaller units and the central limit theorem applies.

Central Limit Theorem

These facts are the basis for most of the methods of statistical inference we will study in the last half of the course.

- Chapter 4 introduces the normal distribution as a probability distribution.

- Chapter 5 culminates in the central limit theorem, the primary theoretical justification for most of the methods of statistical inference in the remainder of the textbook.

The Normal Density

Normal curves have the following bell-shaped, symmetric density.

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{y-\mu}{\sigma}\right)^2}$$

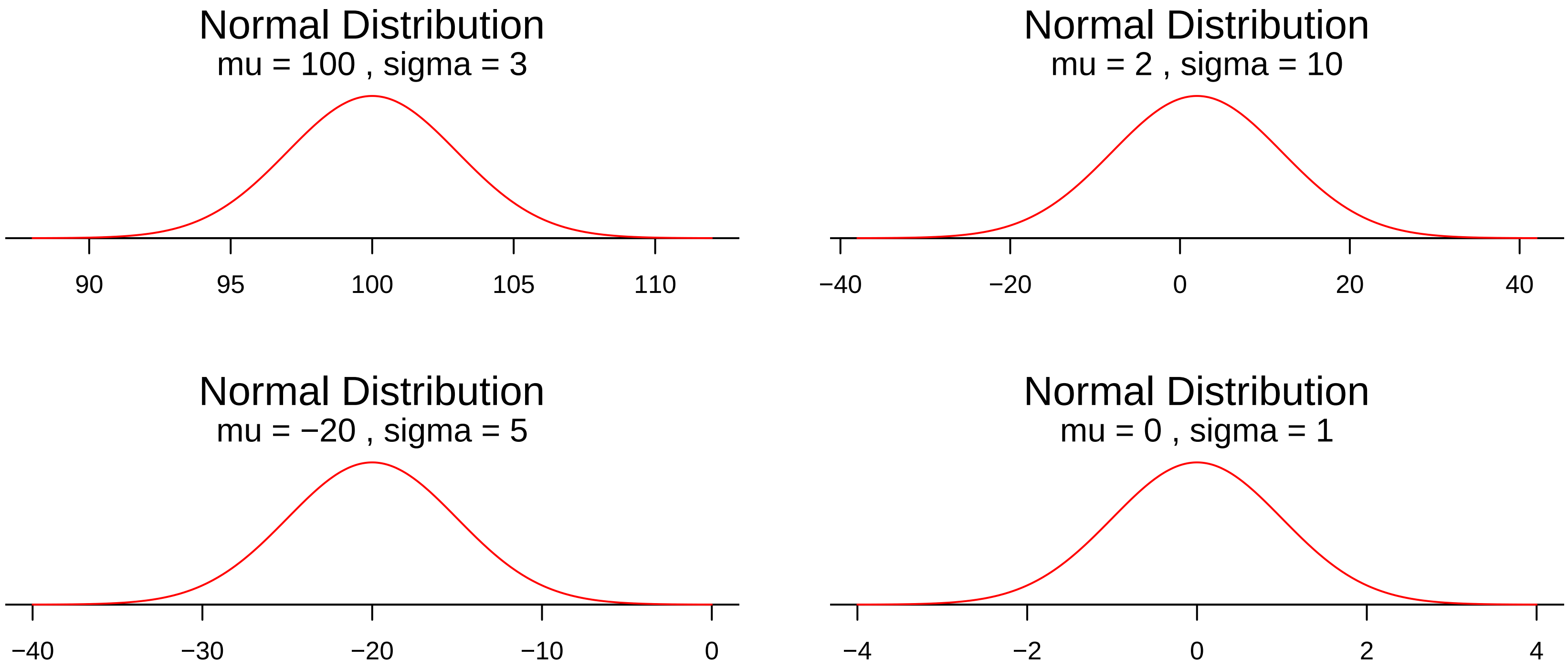
Parameters: The parameters of a normal curve are the mean μ and the standard deviation σ.

If Y is a normally distributed Random Variable with mean μ and standard deviation σ, then we write:

$$Y \sim \mathcal{N}(\mu, \sigma^2)$$

Standard Shape

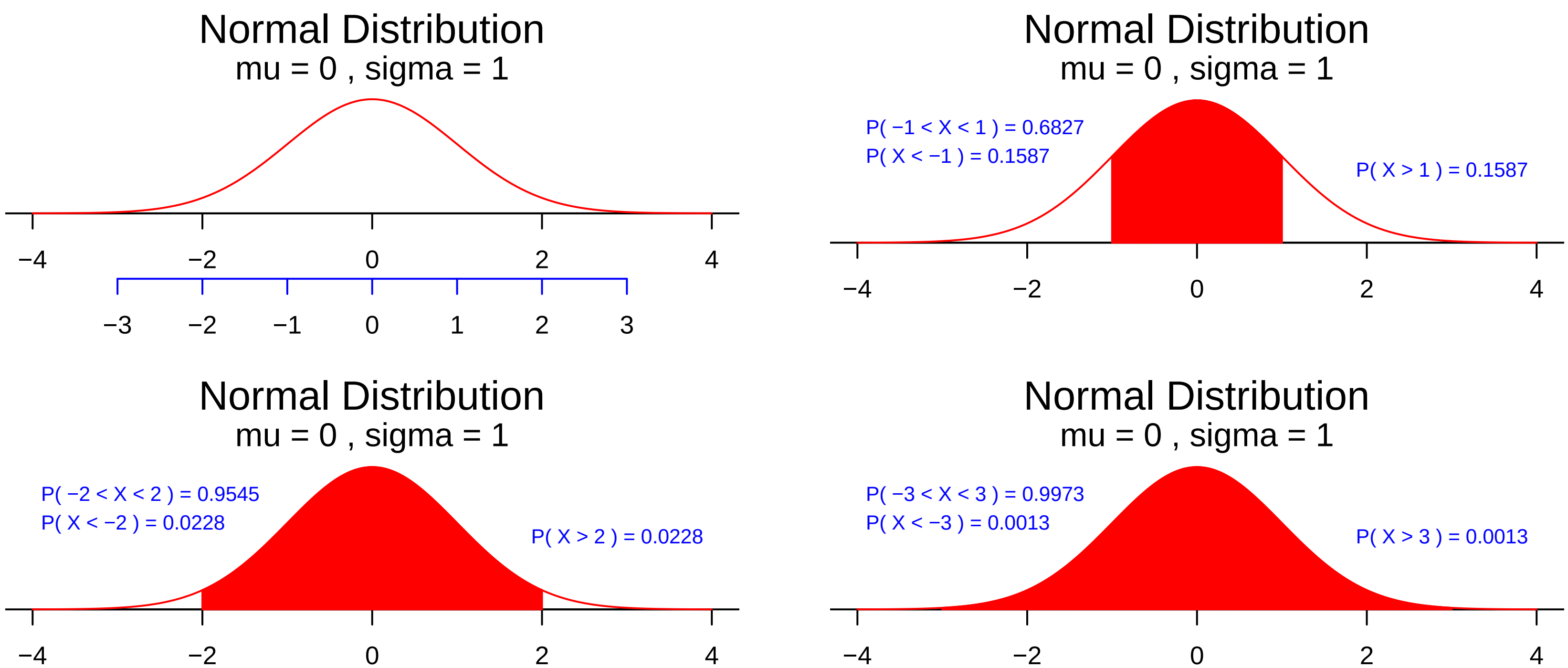
All normal curves have the same shape so that every normal curve can be drawn in exactly the same manner, just by changing labels on the axis (this is not true for the Binomial or Poisson distributions).



The 68–95–99.7 Rule

For every normal curve,

- 68% of the area is within one SD of the mean,
- 95% of the area is within two SDs of the mean,
- 99.7% of the area is within three SDs of the mean.



Standardization

If Y is a normally distributed Random Variable $\quad Y \sim \mathcal{N}(\mu, \sigma^2)$

And we define

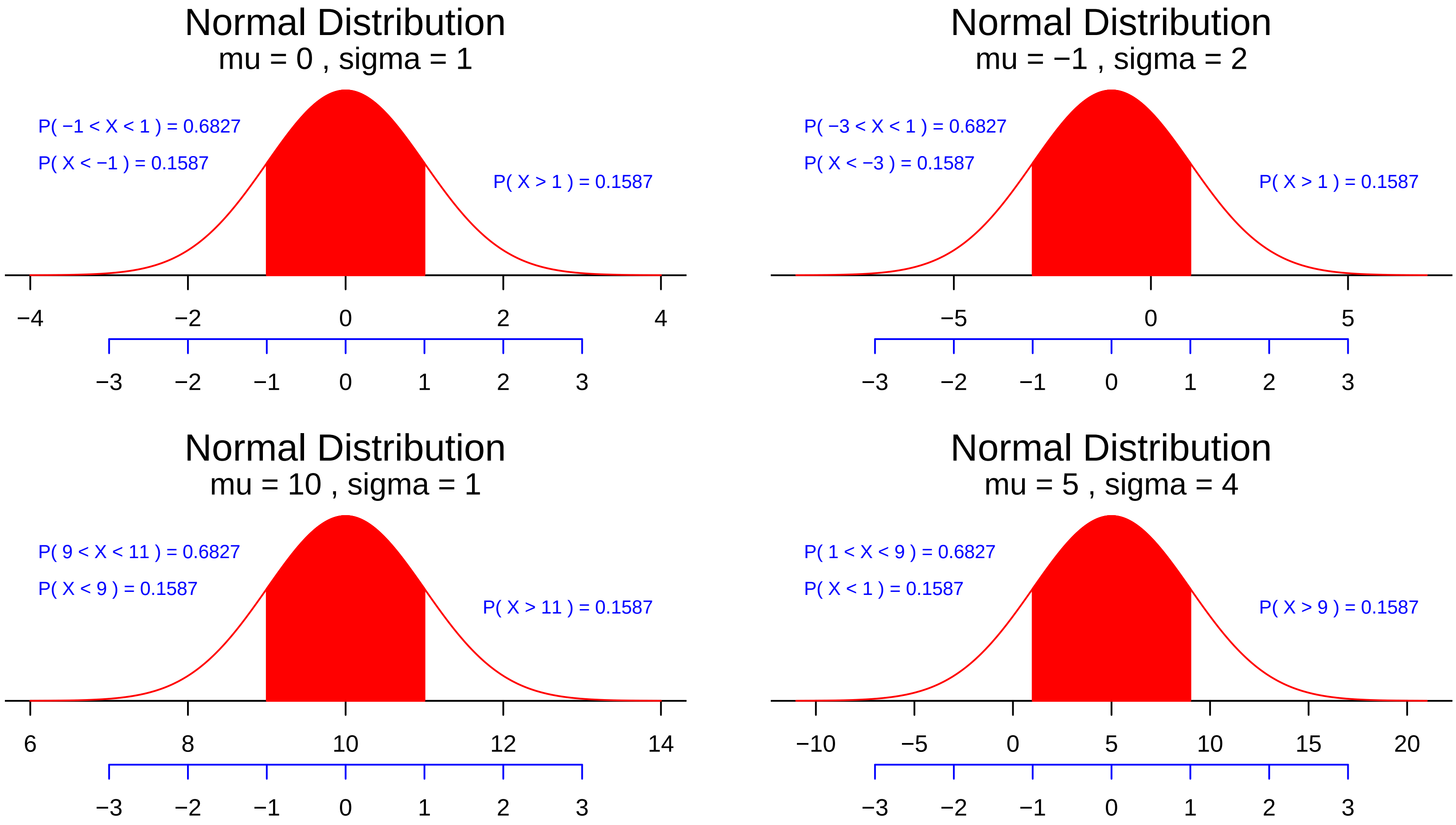
$$Z = \frac{Y - \mu}{\sigma}$$

Then Z is a standard normal random variable

$$Z \sim \mathcal{N}(0, 1)$$

Every problem that asks for an area under a normal curve is solved by first finding an equivalent problem for the standard normal curve.

Standardization



Example Calculation

Suppose that egg shell thickness is normally distributed with a mean of 0.381 mm and a standard deviation of 0.031 mm.

Find the proportion of eggs with shell thickness less than 0.34 mm.

1. We define our random variable Y = egg shell thickness.
2. We know that

$$Y \sim \mathcal{N}(0.381, 0.031^2)$$

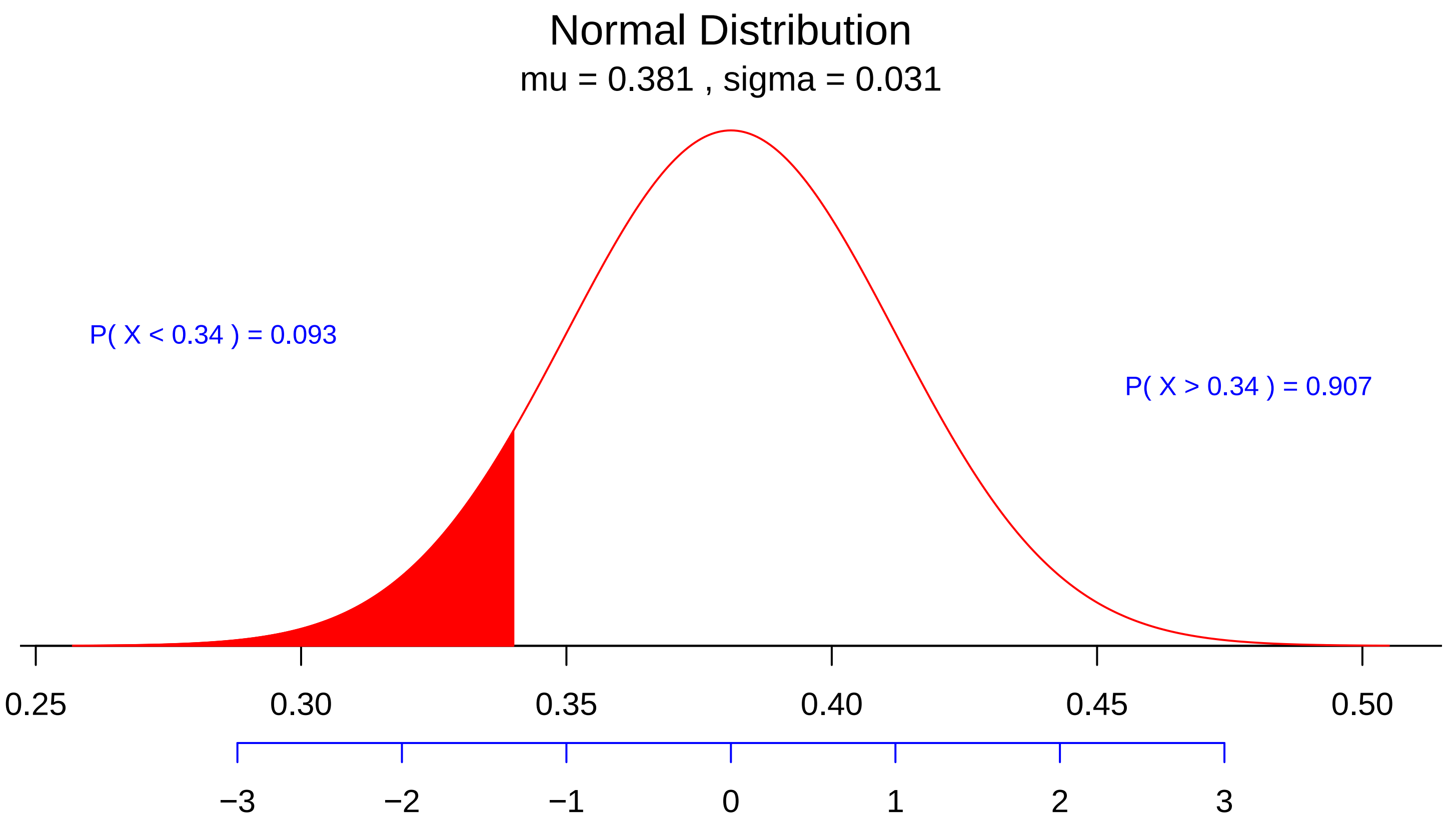
3. We wish to know

$$\Pr\{Y < 0.34\}$$

Example Calculation

Here is a way to do it in R.

```
> source("prob.R")
> gnorm(0.381, 0.031, b = 0.34)
```



Example Calculation

If you don't have R handy, the way to do this is to transform it into a question about a standard normal distribution.

$$\begin{aligned}\Pr\{Y < 0.34\} &= \Pr\{Y - 0.381 < 0.34 - 0.381\} \\ &= \Pr\{[Y - 0.381]/0.031 < [0.34 - 0.381]/0.031\} \\ &= \Pr\{Z < -1.322\}\end{aligned}$$

Where Z is a standard normal random variable: $Z \sim \mathcal{N}(0, 1)$.

This is exactly the type probability tabulated in a standard normal table

Standard Normal table

- The standard normal table lists the area to the left of z under the standard normal curve for each value from -3.49 to 3.49 by 0.01 increments.

- The normal table is on the inside front cover of your textbook.

- Numbers in the margins represent z.

- Numbers in the middle of the table are areas to the left of z.

Standard Normal table

Here is a portion of the table on the inside cover of your book:

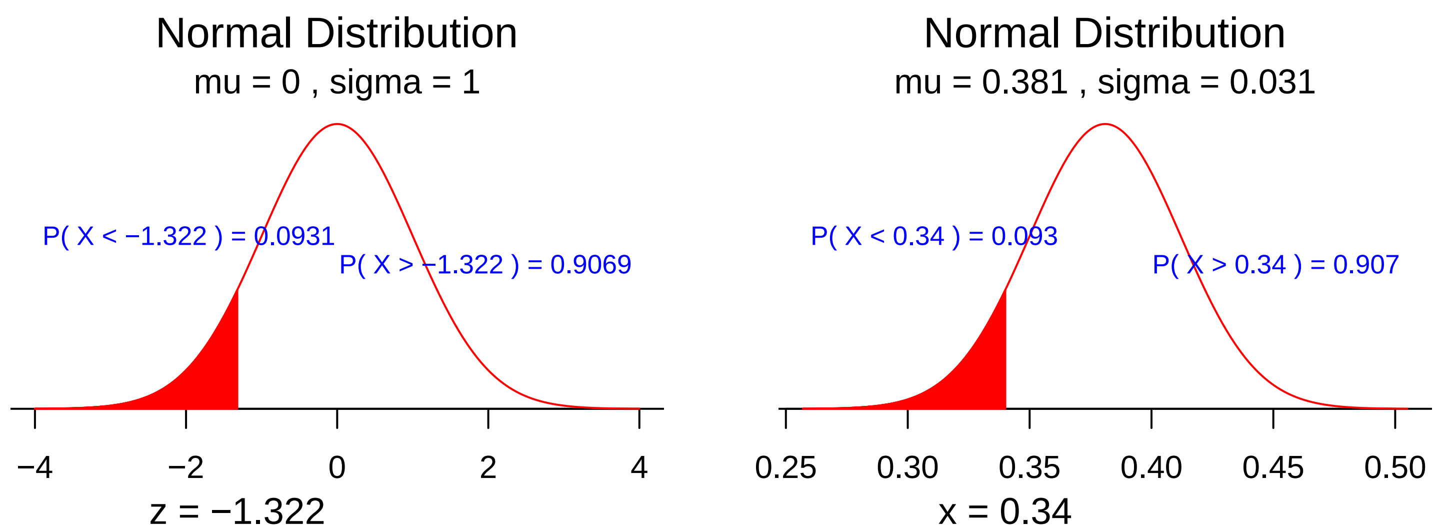
	0	.01	.02	.03
-1.4	0.0808	0.0793	0.0778	0.0764
-1.3	0.0968	0.0951	0.0934	0.0918
-1.2	0.1151	0.1131	0.1112	0.1093
-1.1	0.1357	0.1335	0.1314	0.1292
-1.0	0.1587	0.1562	0.1539	0.1515

We want the area to the left of -1.322

- we round to -1.32
- look up -1.3 in the row labels on the left
- look up .02 in the column labels
- to find $P(Z < -1.32) = 0.0934$.

Standard Normal table

- R can do this for general values of z
- and R can do the standardization for you.
- Recall that our original question was given $Y \sim \mathcal{N}(0.381, 0.031^2)$ what is $\Pr\{Y < 0.34\}$.

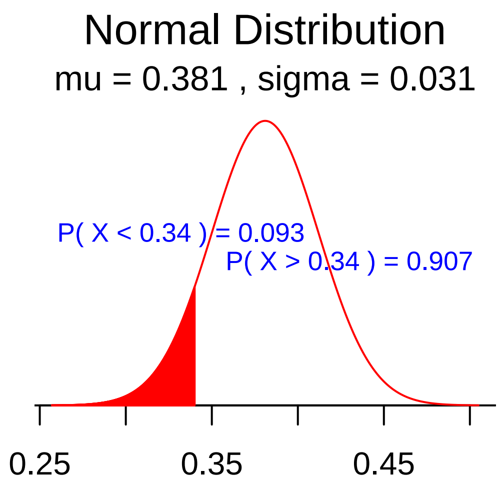


Example Area Calculations: Area to the left

Find the proportion of eggs with shell thickness less than 0.34 mm.

```
> pnorm(q = 0.34, mean = 0.381, sd = 0.031)
[1] 0.09298744
```

```
> gnorm(0.381, 0.031, b = 0.34, sigma.axis = F)
```



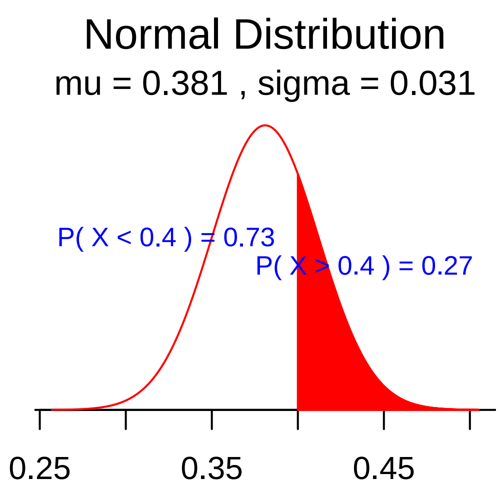
$$\begin{aligned} \Pr\{Y < 0.34\} &= \Pr\{[Y - 0.381]/0.031 < [0.34 - 0.381]/0.031\} \\ &= \Pr\{Z < -1.322\} = 0.0934 \end{aligned}$$

Example Area Calculations: Area to the right

Find the proportion of eggs with shell thickness more than 0.4 mm.

```
> 1 - pnorm(q = 0.4, mean = 0.381, sd = 0.031)
[1] 0.2699702
```

```
> gnorm(0.381, 0.031, a = 0.4, sigma.axis = F)
```

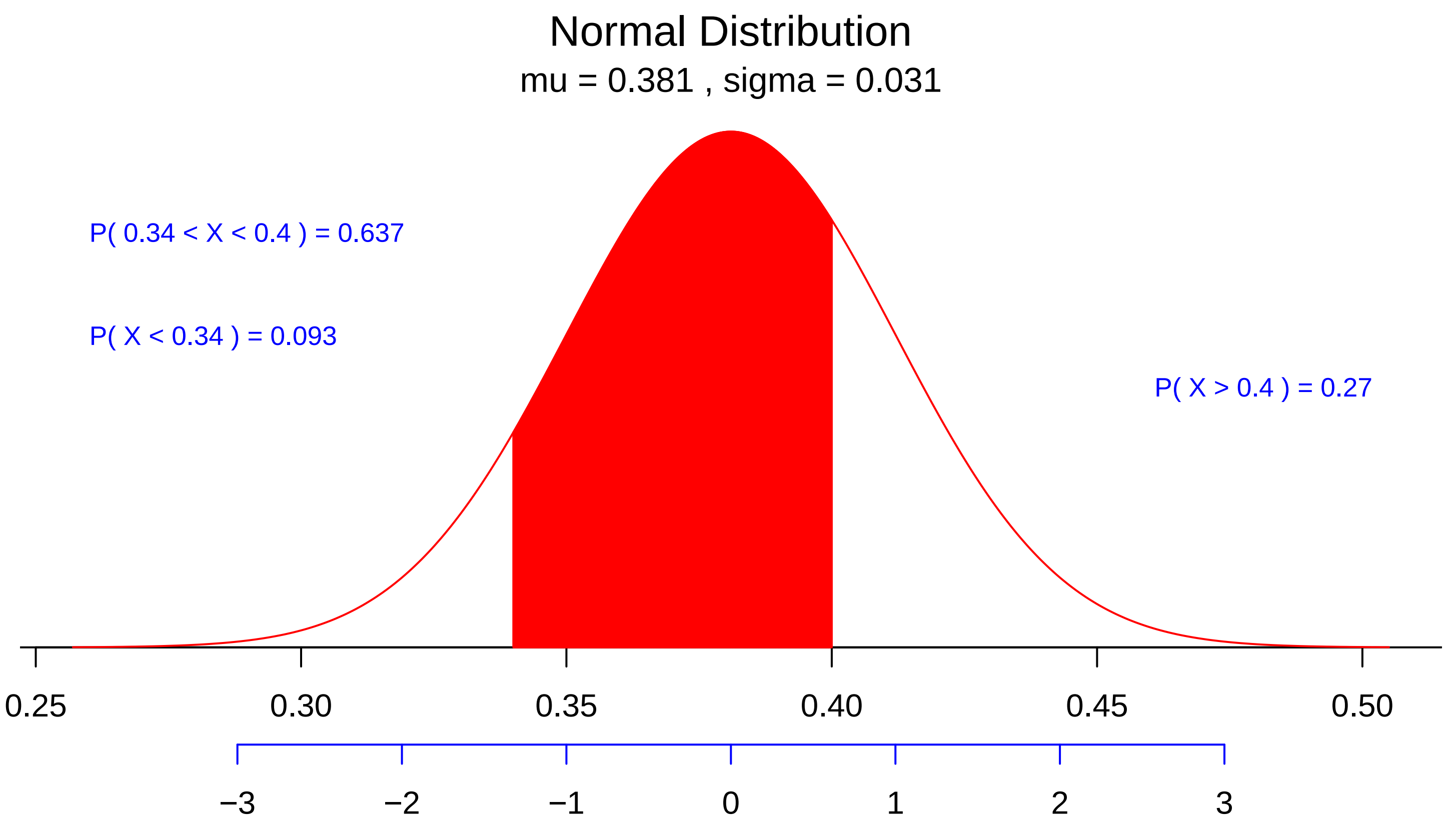


$$
\begin{aligned}
\Pr\{Y > 0.4\} &= \Pr\{[Y - 0.381]/0.031 > [0.4 - 0.381]/0.031\} \\
&= \Pr\{Z > 0.613\} = 1 - \Pr\{Z < 0.613\} = 1 - 0.7291 = 0.2709
\end{aligned}
$$

Example Area Calculations: Area between two values.

Find the proportion of eggs with shell thickness between 0.34 and 0.4 mm.

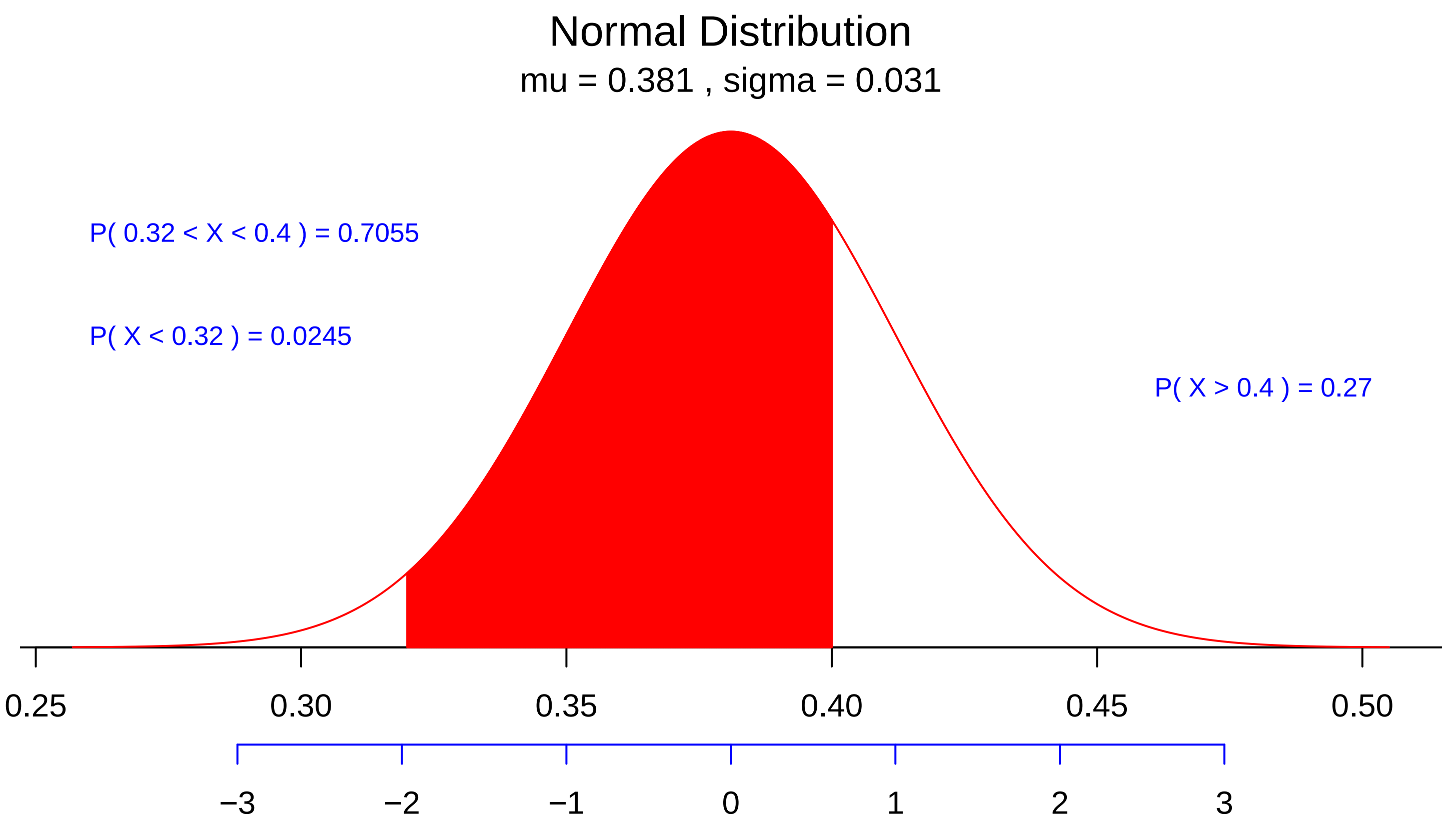
```
> gnorm(0.381, 0.031, a = 0.34, b = 0.4)
```



Example Area Calculations: Area outside two values.

Find the proportion of eggs with shell thickness smaller than 0.32 mm or greater than 0.40 mm.

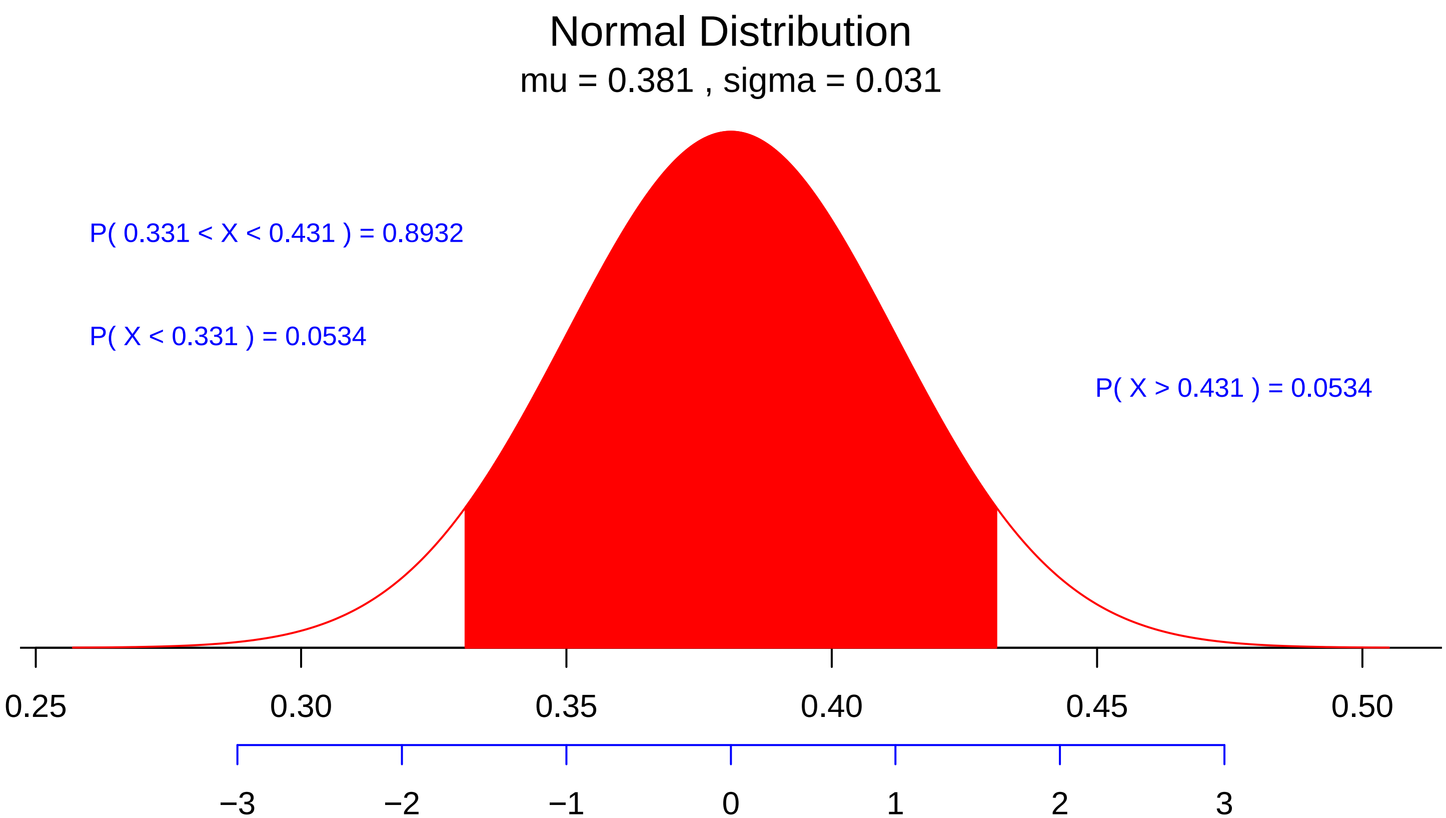
```
> gnorm(0.381, 0.031, a = 0.32, b = 0.4)
```



Example Area Calculations: Central area.

Find the proportion of eggs with shell thickness within 0.05 mm of the mean.

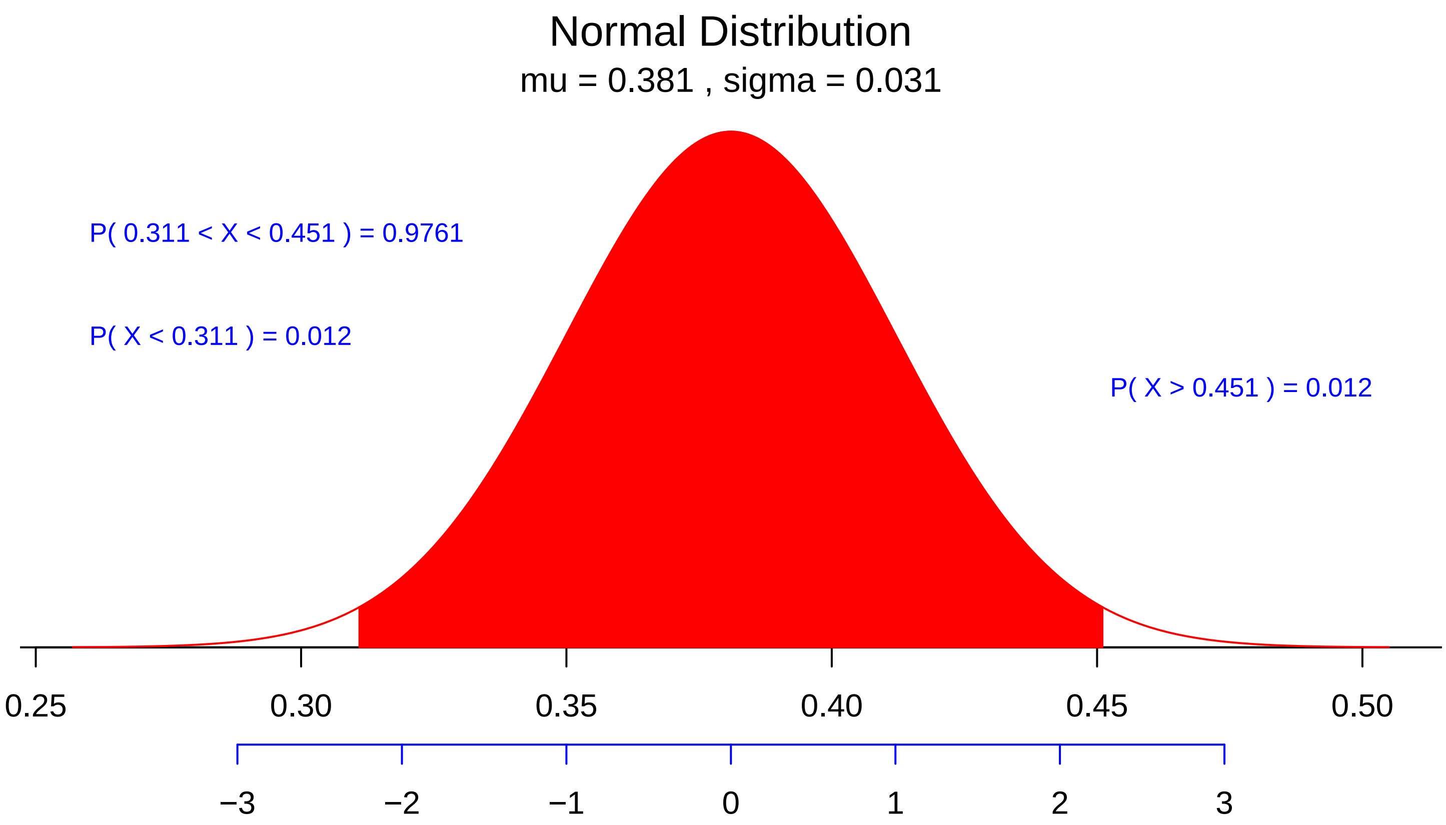
```
> gnorm(0.381, 0.031, a = 0.381 - 0.05, b = 0.381 + 0.05)
```



Example Area Calculations: Two-tail area.

Two-tail area. Find the proportion of eggs with shell thickness more than 0.07 mm from the mean.

```
> gnorm(0.381, 0.031, a = 0.381 - 0.07, b = 0.381 + 0.07)
```



Using R

We have seen how to use R and the new function **gnorm** to graph normal distributions where the graphs include some probability calculations. You can also make the same calculations without graphs using the function **pnorm** (the "p" stands for probability). The following lists the commands for all of the previous computations.

Area to the left. Find the proportion of eggs with shell thickness less than 0.34 mm.

```
> pnorm(0.34, 0.381, 0.031)
[1] 0.09298744
```

Area to the right. Find the proportion of eggs with shell thickness more than 0.36 mm.

```
> 1 - pnorm(0.36, 0.381, 0.031)
[1] 0.75093
```

Using R

Area between two values. Find the proportion of eggs with shell thickness between 0.34 and 0.36 mm.

```
> pnorm(0.36, 0.381, 0.031) - pnorm(0.34, 0.381, 0.031)
[1] 0.1560825
```

Area outside two values. Find the proportion of eggs with shell thickness smaller than 0.32 mm or greater than 0.40 mm.

```
> pnorm(0.32, 0.381, 0.031) + 1 - pnorm(0.4, 0.381, 0.031)
[1] 0.2945190
```

Central area. Find the proportion of eggs with shell thickness within 0.05 mm of the mean.

```
> 1 - 2 * pnorm(0.381 - 0.05, 0.381, 0.031)
[1] 0.8932345
```

Two-tail area. Find the proportion of eggs with shell thickness more than 0.07 mm from the mean.

```
> 2 * pnorm(0.381 - 0.07, 0.381, 0.031)
[1] 0.02394164
```

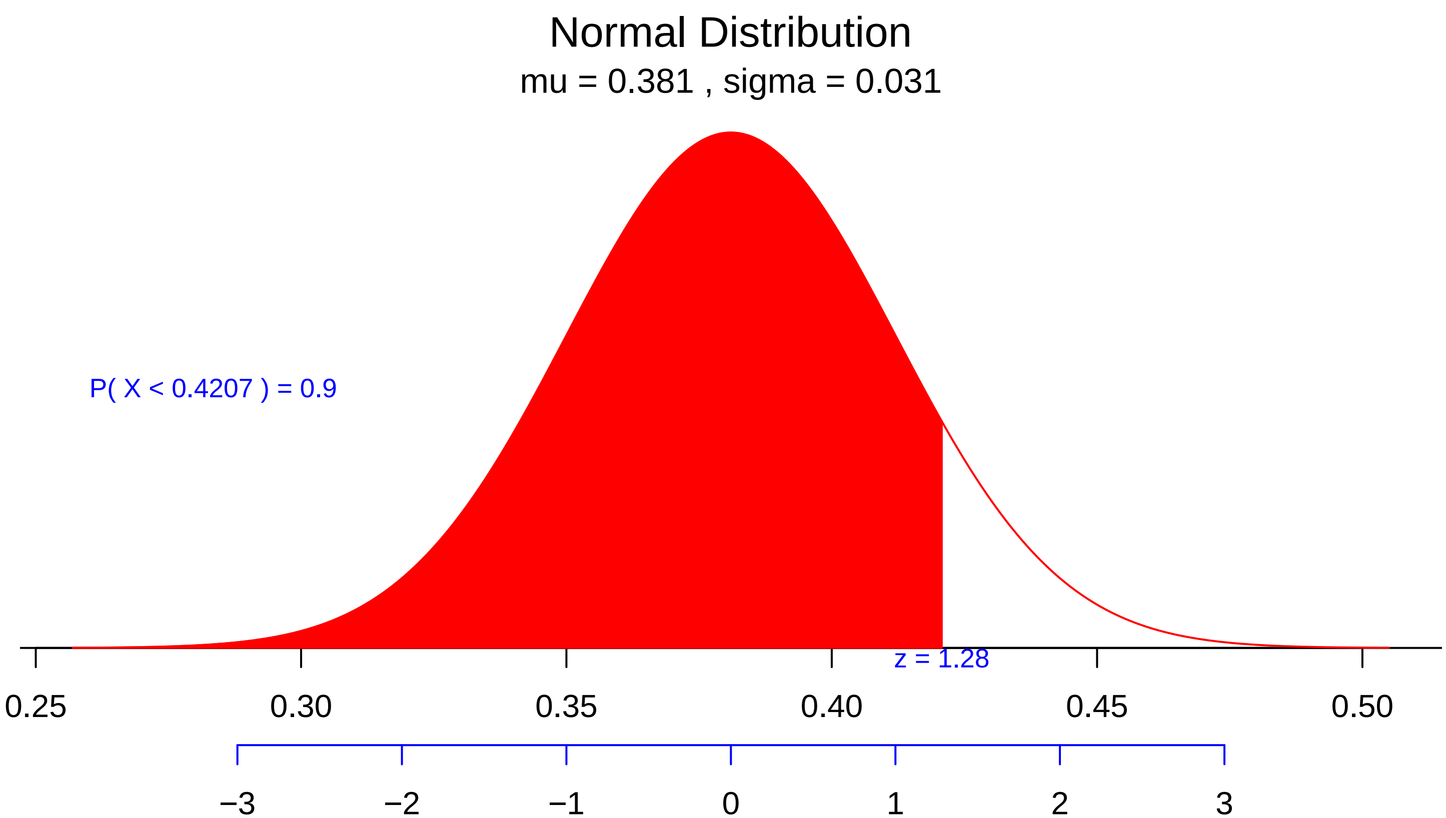
Quantiles

Quantile calculations ask you to use the normal table backwards. You know the area but need to find the point or points on the horizontal axis.

Example Quantile Calculations

Percentile. What is the 90th percentile of the egg shell thickness distribution?

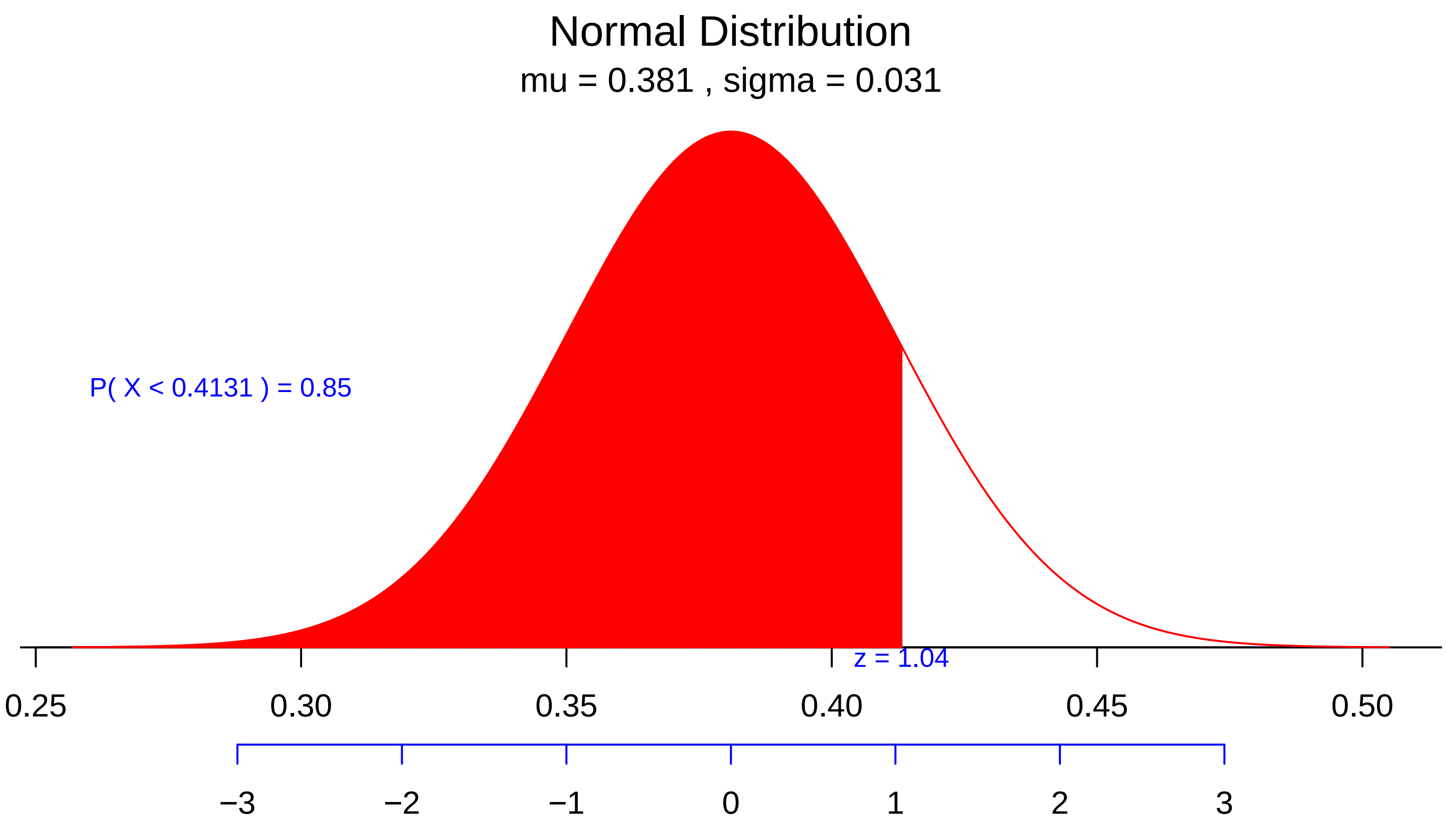
```
> source("prob.R")
> gnorm(0.381, 0.031, quantile = 0.9)
```



Example Quantile Calculations

Upper cut-off point. What value cuts off the top 15% of egg shell thicknesses?

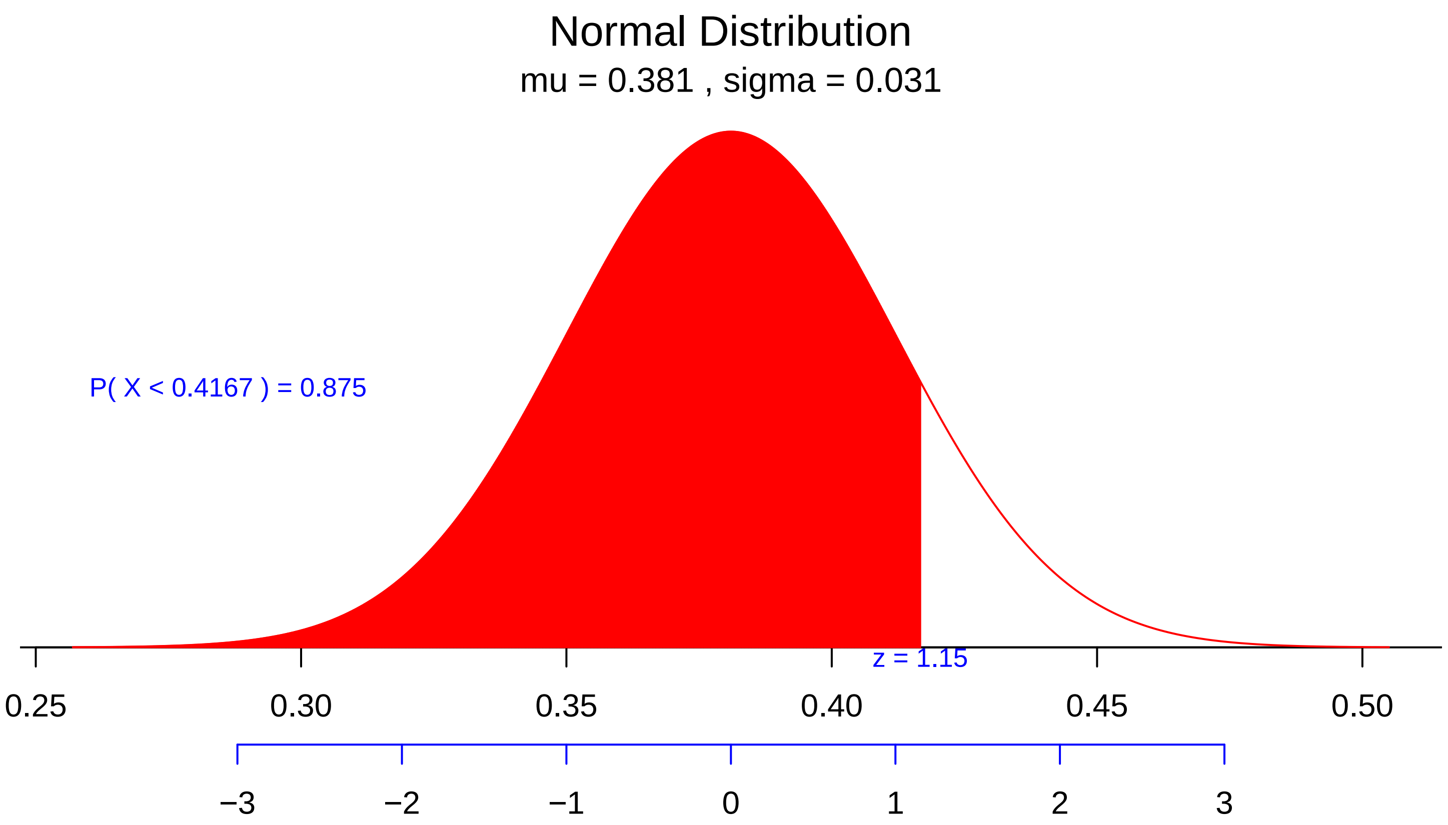
```
> gnorm(0.381, 0.031, quantile = 0.85)
```



Example Quantile Calculations

Central cut-off points. The middle 75% egg shells have thicknesses between which two values?

```
> gnorm(0.381, 0.031, quantile = 0.875)
```



Using R

You can also make the same calculations without graphs using the function `qnorm` (the "q" stands for quantile). The following lists the commands for all of the previous computations.

Percentile. What is the 90th percentile of the egg shell thickness distribution?

```
> qnorm(0.9, 0.381, 0.031)
[1] 0.4207281
```

Upper cut-off point. What value cuts off the top 15% of egg shell thicknesses?

```
> qnorm(0.85, 0.381, 0.031)
[1] 0.4131294
```

Central cut-off points. The middle 75% egg shells have thicknesses between which two values?

```
> qnorm(c(0.25/2, 1 - 0.25/2), 0.381, 0.031)
[1] 0.3453392 0.4166608
```

Approximation of Discrete distributions using the Normal distribution

- The normal distribution can also be used to compute approximate probabilities of discrete distributions.

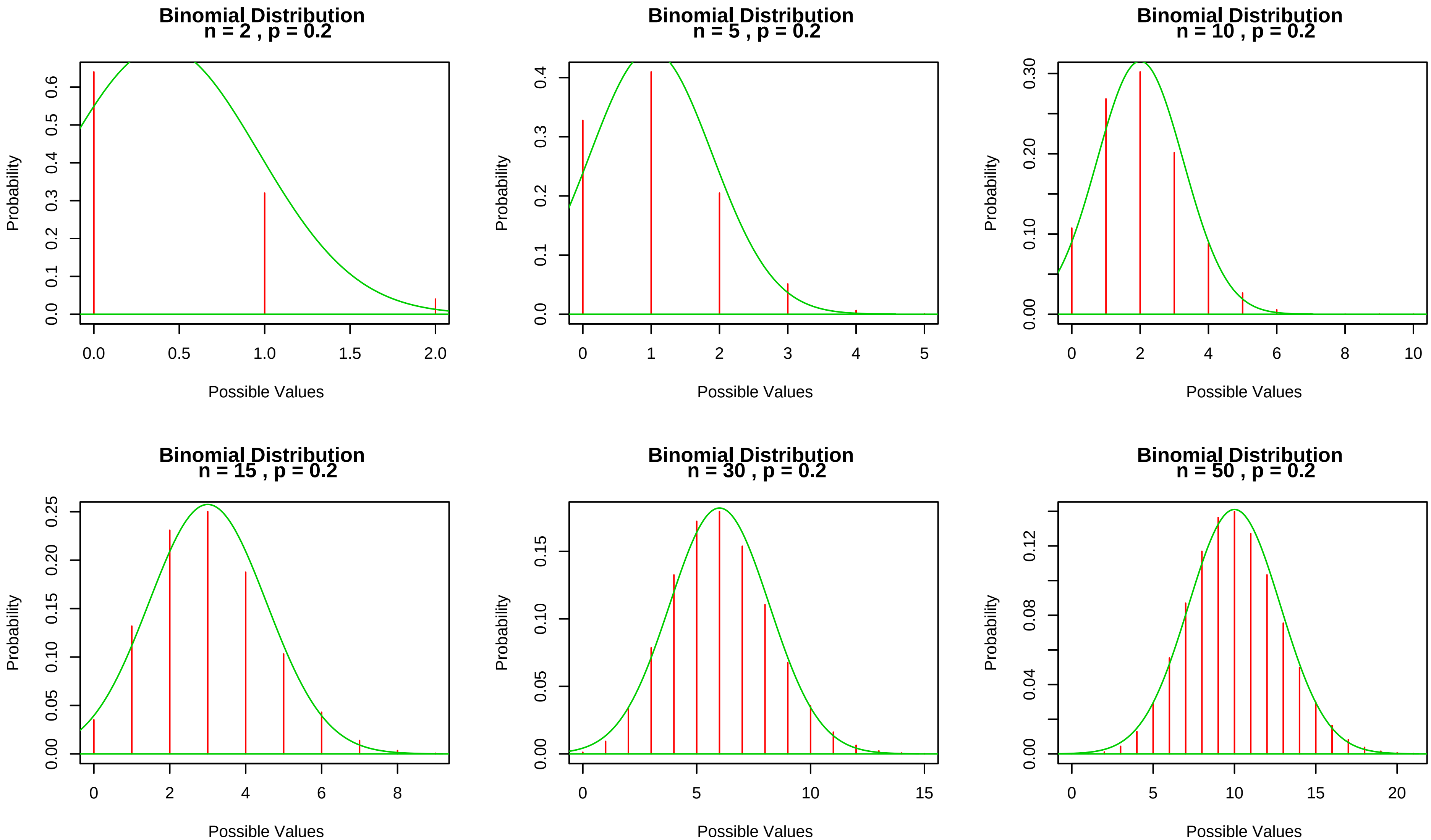
- Approximations to the binomial distribution use a normal curve with $\mu = np$ and $\sigma = \sqrt{np(1-p)}$.

- Approximations to the Poisson distribution use a normal curve with $\mu = \mu$ and $\sigma = \sqrt{\mu}$.

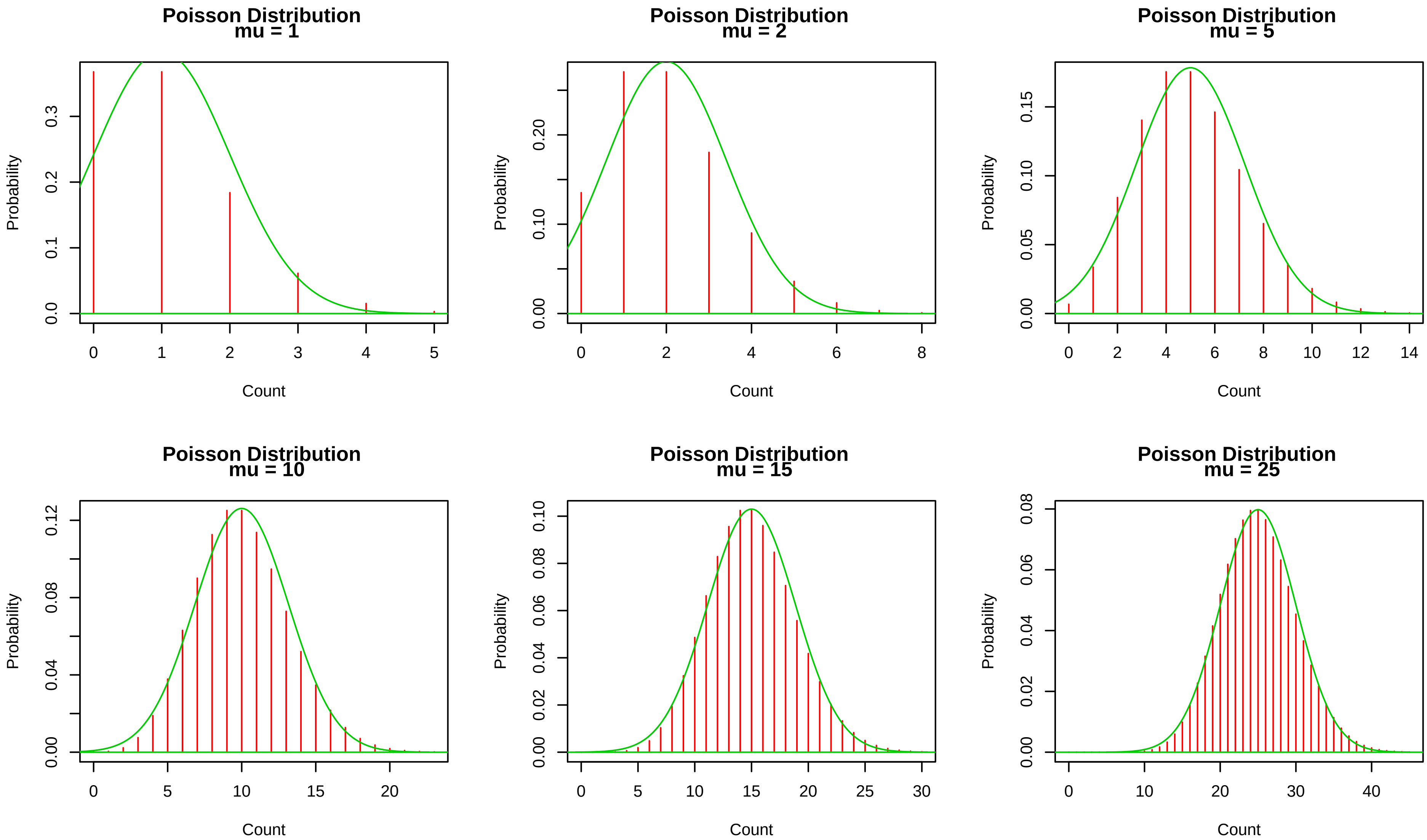
- When approximating a binomial probability, the approximation is usually pretty good when $np > 24$ and $n(1-p) > 24$.

- We will not be using the continuity correction described in your book.

Binomial Distributions: increasing n



Poisson Distributions: increasing n



Example: binomial

Assume Y has a binomial distribution with $n = 150$ and $p = 0.4$ and we want to compute $\Pr\{Y \leq 55\}$.

Exact:

```
> pbinom(55, 150, 0.4)
[1] 0.2274186
```

Normal approximation:

```
> mu = 150 * 0.4
> mu
[1] 60
> 150 * (1 - 0.4)
[1] 90
> sigma = sqrt(150 * 0.4 * 0.6)
> sigma
[1] 6
> pnorm(55, mu, sigma)
[1] 0.2023284
```

Example: Poisson

If Y has a Poisson distribution with $\mu = 25$ what is the probability that Y is greater than or equal to 16?

Exact:

```
> 1 - ppois(15, 25)
[1] 0.977707
```

Normal approximation:

```
> mu = 25
> sigma = sqrt(25)
> sigma
[1] 5
> 1 - pnorm(15, mu, sigma)
[1] 0.9772499
```

With R, do the exact calculation. With a calculator or normal table, use the normal approximation.