Andrew Gard - equitable.equations@gmail.com

The log-normal distribution in *R*

The **log-normal distribution** models data for which the logarithm is normally distributed.

The **log-normal distribution** models data for which the logarithm is normally distributed. This means that all observations are positive and may differ by several orders of magnitude.

The **log-normal distribution** models data for which the logarithm is normally distributed. This means that all observations are positive and may differ by several orders of magnitude. Examples include household incomes and lengths of YouTube comments.

The **log-normal distribution** models data for which the logarithm is normally distributed. This means that all observations are positive and may differ by several orders of magnitude. Examples include household incomes and lengths of YouTube comments.

A log-normal distribution is completely determined by two parameters, μ and σ, the mean and standard deviation of the corresponding normal distribution. These are **not** the mean and standard deviation of the log-normal distribution itself!

There are 4 basic functions in R for calculating in the log-normal distribution.

There are 4 basic functions in *R* for calculating in the log-normal distribution.

- `rlnorm`$(\mathtt{n}, \mu, \sigma)$ generates *n* random values from the log-normal distribution with parameters μ and σ.

There are 4 basic functions in R for calculating in the log-normal distribution.

- `rlnorm`$(\mathtt{n}, \mu, \sigma)$ generates n random values from the log-normal distribution with parameters μ and σ.

  The following code gives 12 values from a distribution whose logarithm is normal with mean 6.1 and standard deviation 1.0.

```
rlnorm(12, 6.1, 1.0)
```

```
##  [1] 1576.48708  321.74736 1685.46920 1591.49534  674.95190   95.58826
##  [7]  176.16710  332.04801  443.29383 4937.69247  956.79912  200.53539
```

- `plnorm`$(\mathtt{q}, \mu, \sigma)$ returns the probability that a random observation X is less than or equal to q. It's the cumulative distribution function of the log-normal, $P(X \leq q)$.

- `plnorm`$(\mathtt{q}, \mu, \sigma)$ returns the probability that a random observation X is less than or equal to q. It's the cumulative distribution function of the log-normal, $P(X \leq q)$. As usual in R, the quantity q can be a vector, so you can compute several probabilities at once.

- `plnorm`(q, μ, σ) returns the probability that a random observation X is less than or equal to q. It's the cumulative distribution function of the log-normal, $P(X \leq q)$. As usual in R, the quantity q can be a vector, so you can compute several probabilities at once.

  The following code computes $P(X \leq 200)$, $P(X \leq 400)$, $P(X \leq 800)$, and $P(X \leq 1600)$ in the distribution whose logarithm is normal with mean 6.1 and standard deviation 1.0.

```r
plnorm(c(200, 400, 800, 1600), 6.1, 1.0)
```

```
## [1] 0.2113683 0.4567855 0.7205956 0.8993328
```

- `qlnorm`$(\mathtt{p}, \mu, \sigma)$ returns the quantity q such that `plnorm`$(\mathtt{q}, \mu, \sigma) = p$. It's the inverse cdf of the specified log-normal distribution. Again, p can be a vector.

- `qlnorm`$(\mathrm{p}, \mu, \sigma)$ returns the quantity q such that `plnorm`$(\mathrm{q}, \mu, \sigma) = p$. It's the inverse cdf of the specified log-normal distribution. Again, p can be a vector.

  The following code computes the 10^{th}, 30^{th}, 50^{th}, 70^{th}, and 90^{th} percentiles in the distribution whose logarithm is normal with mean 6.1 and standard deviation 1.0.

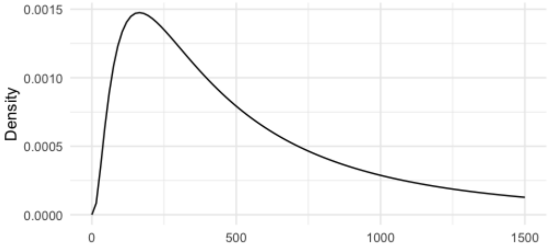
```
qlnorm(c(.1, .3, .5, .7, .9), 6.1, 1.0)
```

```
## [1]  123.7729  263.9077  445.8578  753.2525 1606.0798
```

- `dlnorm`$(\mathtt{x}, \mu, \sigma)$ is the probability density function (pdf) for the specified log-normal distribution.

- `dlnorm`$(\mathrm{x}, \mu, \sigma)$ is the probability density function (pdf) for the specified log-normal distribution. It's mostly used for graphing and for theoretical calculations.

```
ggplot() +
  geom_function(fun = dlnorm,
                args = list(mean = 6.1,
                            sd = 1.0)) +
  xlim(0, 1500) +
  labs(y = "Density")
```



Example. Freshman undergraduate enrollments at U.S. colleges have an approximate log-normal distribution with parameters $\mu = 6.1$ and $\sigma = 1.0$.

1. What is the probability that a randomly-selected college enrolls between 500 and 1000 freshmen in a year?

2. What is the 99^{th} percentile for U.S. college enrollment?

3. Simulate selecting 100 colleges at random. Plot a histogram of freshman enrollments at those colleges.