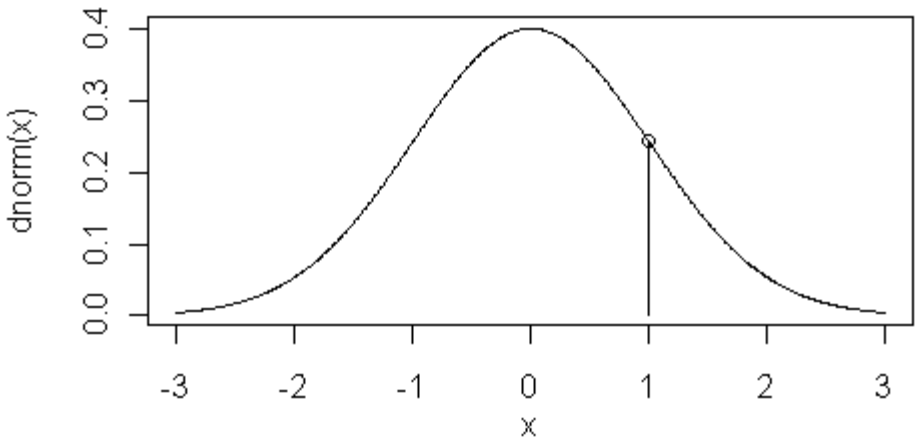
PROBABILITY DISTRIBUTIONS EXERCISES: CRITICAL VALUES, AREAS UNDER CURVES, QUANTILES, AND CHECKS FOR NORMALITY

Probability Distributions

R has density and distribution functions built-in for a number of probability distributions, including those in the following table:

distribution	function	type
binomial	binom	discrete
chi-squared	chisq	continuous
F	f	continuous
hypergeometric	hyper	discrete
normal	norm	continuous
Poisson	pois	discrete
Student's t	t	continuous
uniform	unif	continuous

By prefixing a "d" to the function name in the table above, you can get probability density values (pdf). By prefixing a "p", you can get cumulative probabilities (cdf). By prefixing a "q", you can get quantile values. By prefixing an "r", you can get random numbers from the distribution. We demonstrate using the normal distribution.

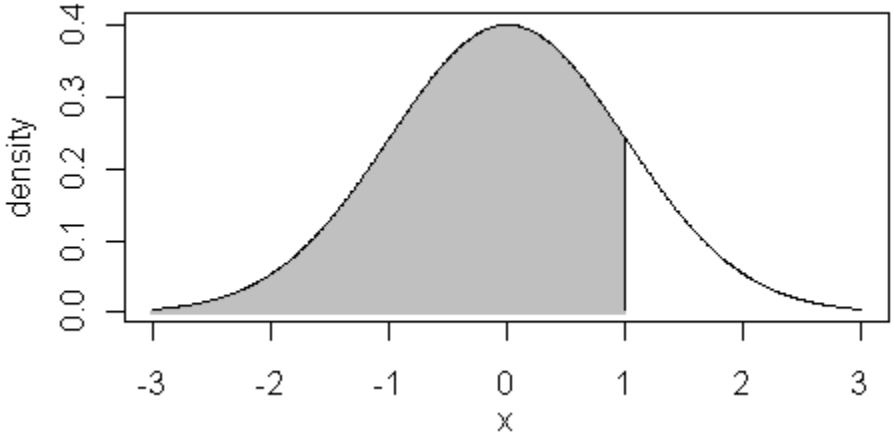


The **dnorm()** function returns the height of the normal curve at some value along the x-axis. This is illustrated in the figure at left. Here the value of **dnorm(1)** is shown by the vertical line at x=1.

```
> dnorm(1)
[1] 0.2419707
```

With no options specified, the value of "x" is treated as a standard score or z-score. To change this, you can specify "mean=" and "sd=" options. In other words, **dnorm()** returns the probability density function or pdf.

The **`pnorm()`** function is the cumulative density function or cdf. It returns the area below the given value of "x", or for x=1, the shaded region in the figure at right.



Please answer the following questions:

1) What percentage of the total area under the curve above is in the shaded region?
2) What percentage of the total area under the curve is bounded by $-2 < X < 0$?
3) What percentage is included in the area $X < -2$ while including $X > 1$?

To get quantiles or "critical values", you can use the **`qnorm()`** function. Please answer these questions:

4) What is the critical value for $p < 0.90$?
5) What are the critical values for two-tailed confidence interval at $p < 0.01$?
6) What are the deciles (or decile boundaries) from the unit normal distribution?
7) What are the quartiles (or quartile boundaries) from the unit normal distribution?

To use these functions with other distributions, more parameters may be needed:

8) What is the area under the t probability density function below $t = 2.101$, $df = 8$?
9) What are the lower and upper bounds of the 95% confidence interval for the chi-square distribution, $df = 3$ (lower) and 12 (upper)?

Random numbers are generated from a given distribution:

10) Randomly draw 30 numbers from a uniform distribution and plot a frequency histogram of their values.
11) Randomly draw 100 numbers from a normal distribution with a mean of 5 and a standard deviation of 7 and plot the resulting density function over the range {-20 to +30}

Empirical Quantiles

12) Calculate the quantiles for the built-in data set "**`rivers`**" and compare these results to what you get from running the **`summary()`** function on "**`rivers`**".
13) Use the **`quantile()`** function to separately calculate the: a) quintiles of rivers; b) deciles of rivers; and c) 55% percentile of rivers.

Checks For Normality

14) Generate a Normal Q-Q Plot for the **`rivers`** data set.
15) Use two different R normality-checking functions to assess whether rivers is normally distributed. Do the two functions return consistent evaluations of the normality of the rivers data set?