Chapter 3

The Normal Distribution

In this chapter we will discuss the following topics:

- The normal density and the density curve with the R-function named **dnorm** (*d*ensity).
- The cumulative normal distribution function with the R-function named **pnorm** (*p*robability).
- The Quantiles with the R-function named **qnorm** (*q*uantile).

There are three arguments to the functions **dnorm**(x, μ, σ) and **pnorm**(x, μ, σ), where x is an observation from a normal distribution that has mean μ and standard deviation σ. The argument p to the function **qnorm**(p, μ, σ) is the proportion of observations in the normal distribution that are less than or equal to the corresponding quantile x.

- The density for a continuous distribution measures the probability of getting a value close to x. Continuous random variables have a density at a point since they have no probability at a single point, $P(X = x)$. For the normal distribution we compute the density using the function **dnorm**(x, μ, σ).
- The function **pnorm**(x, μ, σ) computes the proportion of observations in the normal distribution that are less than or equal to x; that is $P(X \leq x)$, where X is $N(\mu, \sigma)$.
- The function **pnorm**$(x, \mu, \sigma, lower.tail = FALSE)$ computes the proportion of observations in the normal distribution that are greater than or equal to x; that is $P(X \geq x)$, where X is $N(\mu, \sigma)$.
- The function **qnorm**(p, μ, σ) returns the *quantile* for which there is a probability of p of getting a value less than or equal to it. Thus, the *quantile* is the value x such that $P(X \leq x) = p$ for a given p. I other words, **qnorm** converts proportions to quantiles while **pnorm** converts quantiles to proportions which means that **qnorm** and **pnorm** are inverse functions of each other.

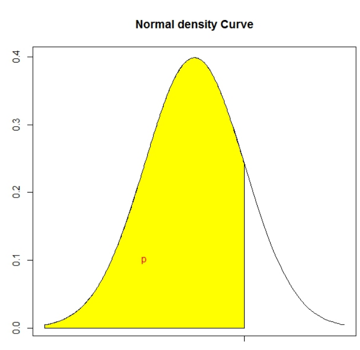


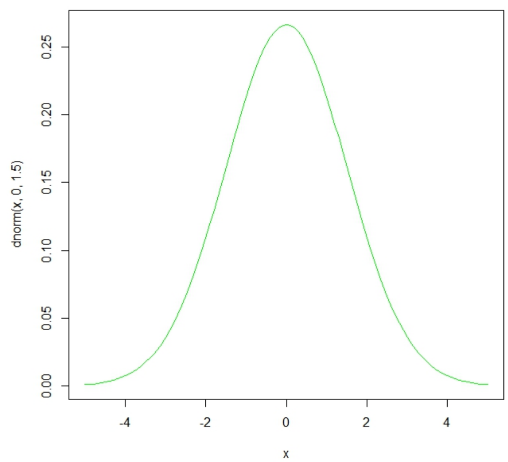
Figure 1: Normal density curve illustrating $P(X \leq x) = p$

The Normal Density Curve

Problem. Plot the bell curve of a normal distribution with mean 0 and standard deviation 1.5 over the domain $-5 \leq x \leq 5$.

Solution. Notice that the left and right tail of the distribution will be close to zero for this choice of domain.

```
> x=seq(-5,5,0.1)
> plot(x,dnorm(x,0,1.5),type="l",col="green")
```



Explanation. The code can be explained as follows:

- The function **x=seq(-5,5,0.1)** generates an array of equally spaced points from -5 to 5 in steps of 0.1 assigned x.
- The function **dnorm**(x,0,1.5) computes the density at x for a normal distribution with mean zero and standard deviation 1.5.
- The entry

  ```
  type="l" (Notice that this is the letter l)
  ```

  in the function **plot** connects the points by lines. By default, R plots the points.

The Normal Cumulative Distribution function

The probability that x is in the interval between a and b is the area under the density curve between a and b.

Problem. Suppose X is normal with mean 527 and standard deviation 105. Compute $P(X \leq 310)$.

Solution. We want to find the proportion of observations in the distribution that are less than or equal to 310. That is, the area under the curve to the left of the x-value 310. This can be done as:

```
> pnorm(310,527,105)
[1] 0.01938279
```

Thus, $P(X \leq 310) = 0.019$.

Problem. The amount of monsoon rain in Tucson is approximately Normally distributed with mean 5.89 inches and standard deviation 2.23 inches. (Data from 1895-2013) [1].
In what percents of all years is the monsoon rainfall in Tucson between 4 inches and 7 inches?

Solution. Let X be the amount of rain in inches. Here we want to find $P(4 \leq X \leq 7)$. This can be written as $P(X \leq 7) - P(X \leq 4)$, where X is $N(5.89, 2.23)$. In R this can be done as:

```
> pnorm(7,5.89,2.23)-pnorm(4,5.89,2.23)
[1] 0.4923238
```

Thus, in 49.23% of all years is the monsoon rainfall in Tucson between 4 inches and 7 inches.

Quantiles for the Normal distribution

Problem. Find the value of x such that the area to its right is 0.1 under the Normal curve with a mean of 400 and a standard deviation of 83.

Solution. Finding the value of x such that the area to its right is 0.1 is equivalent to finding the value of x such that the area to its left is 0.9. Thus, we wish to find x such that $P(X \leq x) = 0.9$, where X is from $N(400, 83)$. It can be done as:

```
> qnorm(0.9,400,83)
[1] 506.3688
```

Hence, $x = 506.37$ so $P(X \leq 506.37) = 0.9$ or $P(X \geq 506.37) = 0.10$.
Notice that this can also be done in R the following way:

```
> qnorm(0.1,400,83,lower.tail=FALSE)
[1] 506.3688
```

Explanation. The code can be explained as follows:

- The use of the option **lower.tail=FALSE** in the **qnorm** function returns the quantile for which the area $p = 0.1$ under the normal curve is to the right for the quantile.

If X is standard normal, we can use the functions, **dnorm(x)**, **pnorm(x)**, and **qnorm(x)**, where the default in R is $\mu = 0$ and $\sigma = 1$.

If we want to compute the quantiles given in the table of the standard normal distribution, we will use the function **qnorm(x)**.

Problem. Compute the 1. quartile of the standard normal distribution.

Solution. Here we want to find the value of z such that the area to the left of z is 0.25. That is, we wish to find z such that $P(Z \leq z) = 0.25$, where Z is standard normal:

```
> qnorm(0.25)
[1] -0.6744898
```

Thus, the 1. quartile is -0.674.

Problem. The math SAT scores among U.S. college students is approximately normally distributed with a mean of 500 and standard deviation of 100. Alf scored 600? What was his percentile? (The percentile is the value for which a specified proportion of observations given in percents fall below it.)

Solution. Let X be the SAT score for the student Alf. We want to find $P(X \leq 600)$. Thus, in R we obtain:

```
> pnorm(600,500,100)
[1] 0.8413447
```

Hence, Alf's SAT score is the 84.13 percentile.

Alternatively, we can first standardize the score such that $z = \frac{600-500}{100} = 1.00$ and then compute $P(Z \leq 1.00)$, where Z is standard normal. We obtain:

```
> pnorm(1)
[1] 0.8413447
```

References

[1] `National Weather Service Forecast for Tucson, AZ, at http://www.wrh.noaa.gov/twc/monsoon/monsoon.php`

00