Probability: Random Variables

A Normal Distribution Problem

In these notes we show how to use Normal probability table to solve problems involving normal distributions with any mean (μ) and any standard deviation (σ). Consider the next problem:

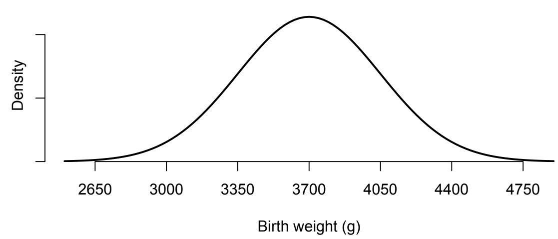
Suppose the birth weights of the infants born at a hospital follow a Normal distribution with mean of 3700 g and a standard deviation of 350 g.

We will work out the answers to these questions:
1. What percent of infants born at the hospital weigh less than 3000 grams?
2. What percent of infants weigh between 4000 and 4500 grams?
3. What birth weight is the first quartile?

Let X be the birth weight of a randomly chosen baby, in statistical notation we can write

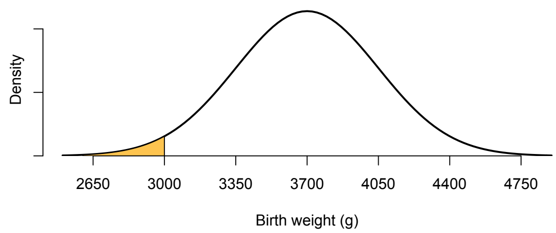
$$X \sim N(3700, 350)$$

The normal distribution is our model for the birth weights of babies at the hospital, so we calculate probabilities as areas under this normal density curve:



The peak is at our mean of 3700, and one standard deviation is 350 grams. So for example from the empirical rule, we expect 68% of the babies born at the hospital to be in range from 3350 ($3700 - 350$) to 4050 ($3700 + 350$) grams.
Let's go to our first question

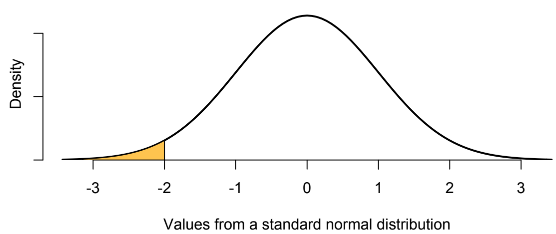
1. What percent of infants born at the hospital weigh less than 3000 grams?
This could also be worded as: What is the probability that an infant will weigh less than 3000 g? Therefore we need this area under our normal density, corresponding to the $P(X < 3000)$:



If we have a computer, we can get probabilities for any normal distribution. If we need to rely on values from the standard normal distribution, we just have to know that any normal random variable can be transformed to a standard normal random variable by subtracting its mean and dividing the result by its standard deviation. That is, for $X \sim N(\mu, \sigma)$ we get:

$$Z = \frac{X - \mu}{\sigma} \sim N(0, 1)$$

This is a standard normal distribution. Hence the area to the left of 3000 in our problem correspond to the area below -2 in standard normal distribution (since $\frac{3000-3700}{350} = -2$):

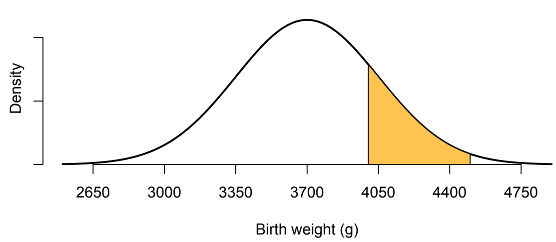


Now we can easily find the desired probability from standard normal probability table:

$$P(X < 3000) = P(Z < \frac{3000 - 3700}{350}) = P(Z < -2) = 0.0228$$

So only about 2% of babies weigh less than 3000 grams.

2. What percent of infants weigh between 4000 and 4500 grams?
We need this area under normal density curve:



Note that normal probabilities are typically given as the area below a value, so to calculate this area we find the area under the density below 4500 g and subtract the area below 4000 g. So we calculate:

$$P(4000 \leq X \leq 4500) = P(X \leq 4500) - P(X < 4000) =$$

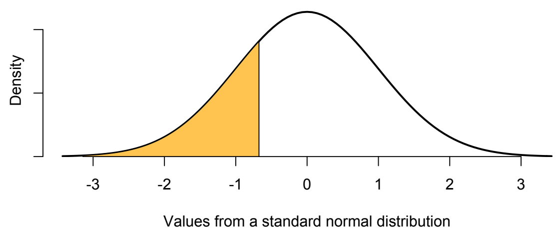
$$P(Z \leq \frac{4500 - 3700}{350}) - P(Z < \frac{4000 - 3700}{350}) =$$

$$P(Z \leq 2.29) - P(Z < 0.86) = 0.9890 - 0.8051 = 0.1839$$

So approximately 18% of babies have birth weights between 4000 and 4500 grams.

3. What birth weight is the first quartile?

The first quartile is the value such that the probability of being less than it is 0.25.



Here the shaded area is exactly 25%. For a standard Normal distribution this value is half way between -0.67 and -0.68, we approximate it as -0.675. This is the first quartile for a standard normal, but we need this for our birth weights with a mean of 3700 and standard deviation of 350 grams. To convert a value from any normal distribution to a standard normal distribution, we subtract μ and divide by σ. To go in the opposite direction, we do the opposite calculation. That is if $Z \sim N(0, 1)$ then $Z * \sigma + \mu$ has a $N(\mu, \sigma)$ distribution. Using the above equation we can easily get the first quartile for the Birthweight from the first quartile for a standard normal:

$$-0.675 * 350 + 3700 = 3463.75$$

Hence the first quartile is about 3464 grams.