

## Continuous Probability Distributions

Recall that we have contrasted continuous variables from discrete variables by noting that while there are gaps between possible values of a discrete variable, a continuous variable can take any value in some interval of the number line.

Suppose  $X$  is a **discrete** random variable taking on the values  $1, 2, \dots, n$  with **probability distribution function**  $f_X(x)$ . Then

$$P[2 \leq X \leq 4] = f_X(2) + f_X(3) + f_X(4)$$

Suppose  $X$  is a **continuous** random variable with **probability density function**  $f_X(x)$ . Then

$$P[a < X < b] = \int_a^b f_X(x) dx \quad (1)$$

Suppose  $X$  is a continuous random variable taking on any value  $x \in [c, d]$ . Then the probability density function satisfies

$$\begin{aligned} f_X(x) &> 0 && \text{if } x \in [c, d] \\ &= 0 && \text{otherwise} \end{aligned}$$

$$\int_c^d f_X(u) du = 1,$$

The probability  $P(a < X < b)$  equals

$$\int_a^b f_X(u) du$$

A random variable  $X$  is said to follow a Gaussian (Normal) distribution with population parameters  $\mu$  and  $\sigma$  if its probability density function is given by

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

A random variable  $X$  is said to follow the standard Normal distribution if its population parameters  $\mu$  and  $\sigma$  equal 0 and 1, respectively. In this case,

$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

There are a number of reasons why the normal distribution is one of the most useful of all probability densities:

- It is simple. It is completely characterized by two population parameters,  $\mu$  and  $\sigma$ .
- It reasonably approximates many distributions found in nature (“it works”).
- It is a physical law that random variables generated from sums of other random variables tend to become normally distributed as the number of variables summed gets larger (the famous Central Limit Theorem).

Example ( Normal distribution)

Suppose  $X$  represents blood pressure and suppose that the population mean  $\mu = 129$  and the population standard deviation  $\sigma = 19.8$ .

$$\begin{aligned} P[X > 150] &= \int_{150}^{\infty} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \\ &= \int_{150}^{\infty} \frac{1}{\sqrt{2\pi(19.8)^2}} e^{-\frac{(x-(129))^2}{2(19.8)^2}} \end{aligned}$$

The probability that a Normal random variable takes on a value within an interval is equal to the area under the part of the normal density which lies above the interval.

Unfortunately, there is no simple formula for calculating this area, so we need to use a table.

Fortunately, we only need a table for the **standard normal density** with mean  $\mu = 0$  and standard deviation  $\sigma = 1$ .

If we want to calculate probabilities for a general normal random variable  $X$  with mean  $\mu$  and SD  $\sigma$ , we need to construct a new random variable, called the **standardized score** of  $X$ ,

$$Z = \frac{X - \mu}{\sigma}.$$

Given values for  $\mu$  and  $\sigma$ , we can actually go back and forth between the “ $X$  scale” and the “ $Z$  scale:”

$$X = \mu + \sigma Z.$$

This is similar to how we go back and forth between Celsius ( $C$ ) and Fahrenheit ( $F$ ):

$$C = \frac{F - 32}{1.8}, \text{ and}$$

$$F = 32 + 1.8 \times C.$$

Two simple rules can be very helpful in calculating normal probabilities:

- Since the total area under any density is 1,

$$P[Z > z] = 1 - P[Z \leq z].$$

- Since the normal density is symmetric about 0,

$$P[Z < z] = P[Z > -z] = 1 - P[Z \leq -z];$$

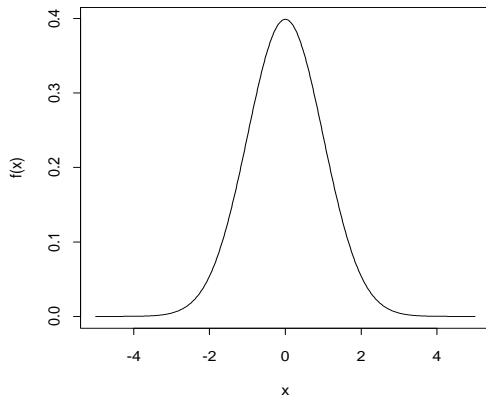
## Applications of the Normal distribution

Table 3 in the Appendix gives  $P[Z > z]$  for specific values of  $z$ .

Recall that  $F_Z(z) = P[Z \leq z] = 1 - P[Z > z]$ .

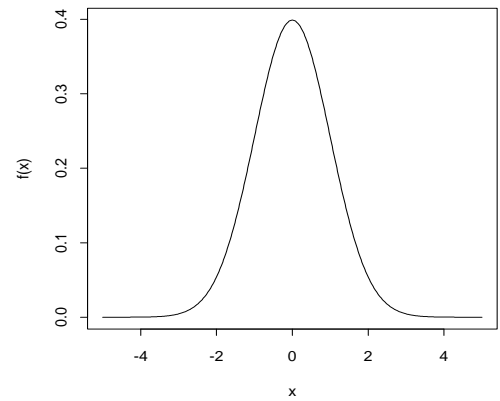
Note that  $F_Z(z) = P(Z \leq z)$  is often denoted by  $\Phi(z)$

and  $1 - \Phi(z) = P(Z > z)$ .



If  $Z$  is a standard normal random variable, find  $P[-1 < z < 1]$ .

From Table 3 in the Appendix of your book,  $P(Z > 1) = 0.159$



If  $X$  is a normally distributed random variable representing blood pressure (BP) with mean  $\mu = 129$  and  $\sigma = 19.8$ .

$z$	$P[Z > z]$
0.51	0.305
1.06	0.145
1.96	0.025

1. What proportion of men have BP greater than 150 mmHg?
2. What value of BP cuts off the upper 2.5
3. What value of BP cuts off the lower 2.5

Suppose serum cholesterol levels  $X$  for children in Wisconsin have mean 175mg/100ml and SD 30mg/100ml. If we assume “normal levels” are within  $\pm 2$  standard deviations of the mean, what proportion of children have “normal levels?”

The scale below corresponds to standardized cholesterol level:

$$Z = \frac{X - 175}{30}.$$

Suppose we want to know the limits within which 95% of the population lies. From Table A.3, we get that  $P[Z > 1.96] = 0.025$

so that  $P[-1.96 \leq Z \leq 1.96] = 0.95$ . Now,

$$\begin{aligned} & P[-1.96 \leq Z \leq 1.96] \\ &= P[30 \times (-1.96) + 175 \leq 30 \times Z + 175 \leq 30 \times 1.96 + 175] \\ &\doteq P[116 \leq X \leq 234]. \end{aligned}$$

We usually write these limits as (116,234) or [116,234].

Forced vital capacity (FVC), a standard measure of pulmonary function, is the volume of air a person can expel in 6 seconds. Current research looks at potential risk factors, such as smoking, air pollution, indoor allergies, that may affect FVC in grade-school children. One problem is that sex, age, and height affect pulmonary function, and these variables must be corrected for before looking at other risk factors. One way to make these adjustments for a particular child is to find the mean  $\mu$  and standard deviation  $\sigma$  for children of the same age (in 1-year groups), sex, and height (in 2-in. height groups) from large surveys and compute a standardized FVC. The standardized FVC then approximately follows a  $N(0,1)$  distribution, provided the original FVC values were approximately normal. Suppose a child is in poor pulmonary health if his or her standardized FVC is less than -1.5. What percentage of children are in poor pulmonary health ?

Suppose a child is considered to have normal lung growth if his or her standardized FVC is within 1.5 standard deviations of the mean. What proportion of children are within this normal range ? What proportion of children are within 2 standard deviations of the mean (how does this match up with the empirical rule) ?

What proportion of children have standardized FVC's between -1.2 and 0.8 ?