Draw a sketch of the appropriate Normal before looking up the Normal table or using R.

For a continuous random variable

probability density function (pdf) $f(x) \; \forall x$ (often represented as graph)

$$f(x) = \frac{dF}{dx}$$

cumulative distribution function (CDF) $F(x) = P(X \le x) = \int_{-\infty}^{x} f(y)dy$

$$P(X \le x) = P(X < x) \quad \text{as } P(X = x) = 0 \;\; \forall x$$

expected value or mean $E(X) = \int_{x} x f(x)dx$

expected value of function $E(g(X)) = \int_{x} g(x)f(x)dx$

Eg: $E(X^2) = \int_{x} x^2 f(x)dx$

variance $Var(X) = E(X^2) - (E(X))^2$

General Normal Distribution

A General Normal random variable $X \sim N(\mu, \sigma^2)$ has mean μ and variance σ^2.

The pdf is $f(x) = \dfrac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$

Standardising a Normal

If $X \sim N(\mu, \sigma^2)$ and $Z \sim N(0,1)$, then

$$P(X \le x) = P\Big(\frac{X-\mu}{\sigma} \le \frac{x-\mu}{\sigma}\Big) = P\Big(Z \le \frac{x-\mu}{\sigma}\Big)$$

1. Data collected by Adelaide sports doctor Geoffrey Verrall and AFL sports scientist Jamie Hepner suggest genetics play a bigger part in achieving a top-level football career than skill or desire ... They found those shorter than 180cm could just about forget trying for a place in the elite league - unless they happened to be an indigenous or Pacific Islander player, who thrive on leg speed and lightning reflexes ... Generally, white Australian males must be tall (about 188cm) to make it in the AFL.

   http://www.heraldsun.com.au/sport/afl/size-matters-at-afl-level/story-e6frf9jf-1226650225771

   (a) What is the average height of Australian males?

   http://www.abs.gov.au/ausstats/abs@.nsf/lookup/4338.0main+features212011-13

   (b) Assuming that the population standard deviation is 7 and that heights are Normally distributed, draw a model for the heights of Australian men.

   ```
   curve(dnorm(x,175.6,7),xlim = c(140, 200), xlab="Height", ylab="Density",
         main="Australian Men Heights",col="blue")
   ```

   (c) AFL's shortest player is Caleb Daniel. How tall is Caleb? How likely is it to find an Australian man of Caleb's height or shorter? http://www.aflplayers.com.au/article/big-dreams-for-afls-smallest-player/

(d) Adam Goodes is 191cm. How likely is it to find an Australian man of Adam's height or taller?

(e) If the average height of Australian women is 161.8cm, what proportion of Australian men are below this average?

Solution

(a) 175.6cm.

(b) $X \sim N(\mu = 75.6, \sigma^2 = 7^2)$.

(c) Caleb is 167cm.

Height of Australian men in cms (x)

$$P(X \le 167) = P(\frac{X - 175.6}{7} \le \frac{167 - 175.6}{7}) = P(Z \le -1.228571) \approx 0.11$$

```
pnorm(167,175.6,7)
```

```
## [1] 0.1096163
```

```
pnorm(-1.228571)
```

```
## [1] 0.1096163
```

(d)

Height of Australian men in cms (x)

$$P(X \ge 191) = P(\frac{X - 175.6}{7} \ge \frac{191 - 175.6}{7}) = P(Z \ge 2.2) \approx 0.01$$

```
1-pnorm(191,175.6,7)
```

```
## [1] 0.01390345
```

```
1-pnorm(2.2)
```

```
## [1] 0.01390345
```

(e)

Height of Australian men in cms (x)

$$P(X \leq 161.8) = P(\frac{X - 175.6}{7} \leq \frac{161.8 - 175.6}{7}) = P(Z \leq -1.971429) \approx 0.02$$

```
pnorm(161.8,175.6,7)
```

```
## [1] 0.02433744
```

```
pnorm(-1.971429)
```

```
## [1] 0.02433741
```

2. Standardising Normal Distribution

In 2011-12, the Australian Bureau of Statistics reported a tendency for high levels of sedentary be-
haviour across the Australian adult population, with Australian adults spending an average of 13 hours
a week watching TV.

http://www.abs.gov.au/ausstats/abs@.nsf/lookup/4364.0.55.004Media%20Release22011-12

Let X be the number of hours per week spent watching TV by an Australian adult. Model X by a
Normal with standard deviation $\sigma = 2$ hours.

(a) Plot the curve.

```
curve(dnorm(x,13,2),10,20, main="Weekly hours of TV watching by Australian adults")
```

(b) For an Australian adult chosen at random, find the probability that the person watches between
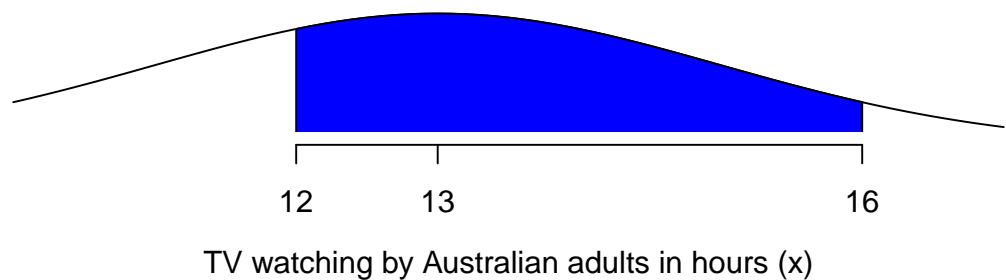12 and 16 hours of TV per week.

(c) What is the 90% percentile for weekly TV watching by Australian adults?

```
qnorm(0.9,13,2)
```

```
## [1] 15.5631
```

Solution

$X \sim N(\mu = 13, \sigma^2 = 4)$.

TV watching by Australian adults in hours (x)

(b)

$$P(12 < X < 16) = P\left(\frac{12 - 13}{2} < \frac{X - 13}{2} < \frac{16 - 13}{2}\right) = P(-0.5 < Z < 1.5) = \Phi(1.5) - \Phi(-0.5) = 0.6246553 \approx 0.6.$$

```
pnorm(16,13,2) - pnorm(12,13,2)

## [1] 0.6246553

pnorm(1.5)-pnorm(-0.5)

## [1] 0.6246553
```

(c) We want q such that $P(X < q) = 0.9$.

So $0.9 = P(X < q) = P\left(\frac{X - 13}{2} < \frac{q - 13}{2}\right) = P\left(Z < \frac{q - 13}{2}\right)$.

As $0.9 = P(Z < 1.28)$ (from table or R), we solve $1.28 = \frac{q - 13}{2}$, which gives $q = 2 * 1.28 + 13 = 15.56$.

```
qnorm(0.9,13,2)

## [1] 15.5631
```

Alternatively, we can use R directly:

```
qnorm(0.9)

## [1] 1.281552
```

3. Expectation of Normal Distribution

Suppose that X is the breaking strength of a rope in pounds, with distribution $N(100, 16)$. Each 30 metre coil of rope brings a return of \$25 provided $X > 95$. If $X \leq 95$ then the rope is used for a different purpose and a return of only \$10 per coil is made. Find the expected return per coil.

Solution

$X =$ the breaking strength of a rope in pounds $\sim N(100, 16)$.

(1) Good quality rope : If $P(X > 95)$ then \$25 return.

$$P(X > 95) = P\left(\frac{X - 100}{4} > \frac{95 - 100}{4}\right) = P(Z > -5/4) = P(Z < 5/4) = \Phi(5/4) = 0.8943.$$

(2) Poor quality rope : If $P(X < 95)$ then \$10 return.

$P(X < 95) = 1 - P(X > 95) = 0.1057.$

Hence expected return per coil of rope is:
$E(X) = P(X > 95) \times \$25 + P(X < 95) \times \$10 = 0.8943 \times \$25 + 0.1057 \times \$10 = \$23.4145.$

Extra Questions

4. Standard Normal Probabilities

   Given $Z \sim N(0,1)$.

   (a) Plot Z.

   ```
   curve(dnorm(x),-3,3)    #Plots Standard Normal from -3 to 3.
   ```

   (b) From looking at the plot, estimate $P(Z \leq 1)$ and $P(Z \leq -1)$. Check by looking up the Normal table, and using R.

   ```
   pnorm(1)    #Calculates P(Z < 1)

   ## [1] 0.8413447
   ```

   (c) Fill out the following table. What do these probabilities represent?

$P(-1 \leq Z \leq 1)$	$P(-2 \leq Z \leq 2)$	$P(-3 \leq Z \leq 3)$
0.68		

   ```
   pnorm(1)-pnorm(-1)
   ```

   (d) Using the Normal table, find the values of a, b and c such that $P(Z \leq a) = 0.7291$, $P(Z > b) = 0.10$ and $P(|Z| < c) = 0.90$. Check your answers in R.

   ```
   qnorm(0.7291)    #Calculates a st P(Z < a) = 0.7291
   ```

   Solution

   (a) From Normal tables, $P(Z \leq 1) = 0.8413$
   $P(Z \leq -1) = 1 - P(Z \leq 1) = 0.1587$.

   (b)

$P(-1 \leq Z \leq 1)$	$P(-2 \leq Z \leq 2)$	$P(-3 \leq Z \leq 3)$
0.6827 or 0.68	0.9545 or 0.95	0.9973 or 0.99

   Comment: This table represents the probability of being 1,2 and 3 standard deviations away from the mean. Eg The probability of being within 1 standard deviation of the mean is approximately 68%.

   (c) Using the Normal table,
   if $P(Z \leq a) = 0.7291$, then $a = 0.6101$
   if $P(Z > b) = 0.10$, then $P(Z < b) = 0.90$, so $b = 1.2816$
   if $P(|Z| < c) = 0.90$, then $P(Z < c) = 0.95$, hence $c = 1.645$

5. Normal Distribution

   Each day random samples of a city's water supply are analysed and the concentration of fluoride (parts per million, ppm) is measured. It is found that the distribution of this concentration is approximately normal with mean of 1.3 ppm and a standard deviation of 0.15 ppm.

   (a) On what proportion of days is the concentration between 1.45 and 1.53 ppm?

   (b) What is the 80th percentile of the distribution? What does this represent?

   Solution

X = the concentration of fluoride $\sim N(1.3, 0.15^2)$.

(a) $P(1.45 \le X \le 1.53) = P(\frac{1.45 - 1.3}{0.15} \le \frac{X - 1.3}{0.15} \le \frac{1.53 - 1.3}{0.15}) = P(1 < Z < 1.53) = \Phi(1.53) - \Phi(1) = 0.9370 - 0.8413 = 0.0957$

```
pnorm(1.53,1.3,0.15)-pnorm(1.45,1.3,0.15)

## [1] 0.09605838
```

(b) We want to find q such that $P(X < q) = 0.8$.

$P(X < q) = P(\frac{X - 1.3}{0.15} < \frac{q - 1.3}{0.15}) = P(Z < \frac{q - 1.3}{0.15}) = 0.8$.

From looking up Normal tables, we find that $\frac{q - 1.3}{0.15} = 0.8416$. So $q = 1.4262$.

Represents: We expect the level of fluoride to be less than 1.4262 80% of the time.

```
pnorm(1.4262,1.3,0.15)  # which is approximately 80%

## [1] 0.7999194
```

6. Continuous Distribution by hand

A random variable X has the following CDF:

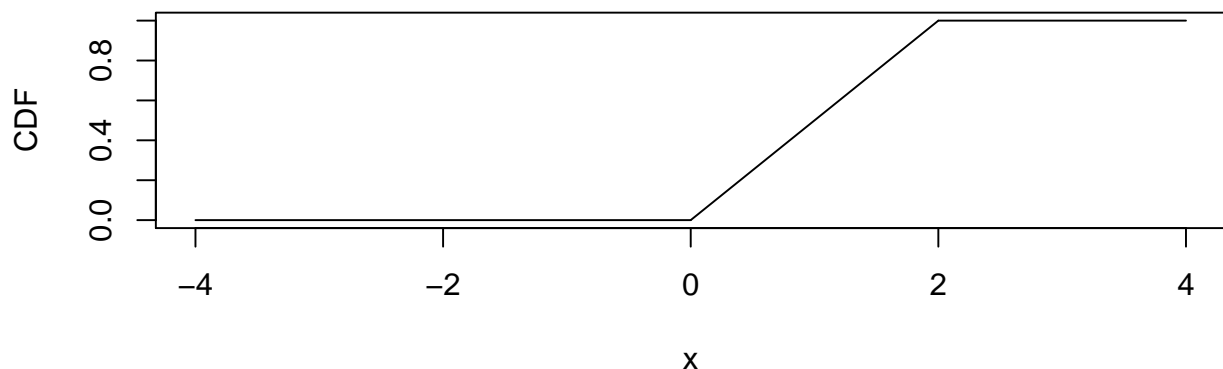$$F(x) = \begin{cases} 0 & x \le 0 \\ x/2 & x \in (0, 2) \\ 1 & x \ge 2 \end{cases}$$

(a) Sketch the CDF.

(b) Find the pdf and sketch it.

(c) Calculate $E(X)$ and $Var(X)$. Does this match up with your sketches?

Solution
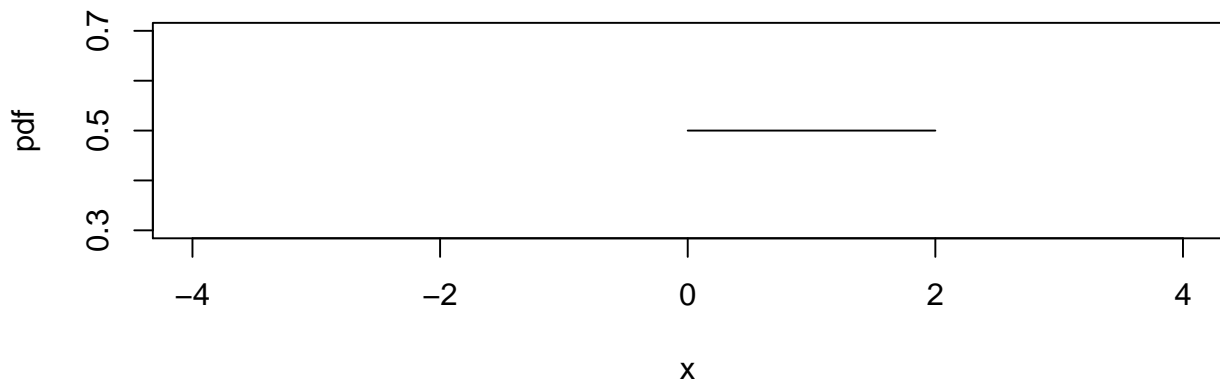
(a) Check your hand drawn sketch against this R output:

(b) To get the pdf, we differentiate the CDF. Hence,

$$f(x) = \begin{cases} 0 & x \le 0 \\ \dfrac{1}{2} & x \in (0,2) \\ 0 & x \ge 2 \end{cases}$$

giving

$$f(x) = \begin{cases} \dfrac{1}{2} & x \in (0,2) \\ 0 & \text{otherwise} \end{cases}$$

Again, check your hand drawn sketch against this R output:

(c) $E(X) = \displaystyle\int_{-\infty}^{\infty} x f(x)\,dx = \dfrac{1}{2}\int_0^2 x\,dx = \dfrac{1}{2}[\dfrac{x^2}{2}]_0^2 = 1$

Comment: As the pdf is symmetric, we expect the mean at $x = 1$.

$E(X^2) = \displaystyle\int_{-\infty}^{\infty} x^2 f(x)\,dx = \dfrac{1}{2}\int_0^2 x^2\,dx = \dfrac{1}{2}[\dfrac{x^3}{3}]_0^2 = 8/6$

Hence $Var(X) = 8/6 - (1)^2 = 1/3$, and $SD(X) \approx 0.58$.

Comment: Looking at the pdf, it seems reasonable that the standard deviation would be 0.58. (The mean is 1 and the full range of data is only 2).

7. Normal Probabilities

Given $X \sim N(5, 4)$.

(a) What is $P(X = 1)$?

(b) Calculate $P(X \le 1)$ and $P(-2 \le X \le 1)$.

(c) Use R to confirm your answers.

```
pnorm(1,5,2)      #Calculates P(X <=1)

## [1] 0.02275013
```

Solution

(a) For any continuous distribution, $P(X = x) = 0$ for all x.

(b) $P(X \le 1) = P\left(\frac{X-5}{2} \le \frac{1-5}{2}\right) = P(Z \le -2) = 1 - P(Z \le 2) = 1 - \Phi(2) = 0.0228$

$P(-2 \le X \le 1) = P\left(\frac{-2-5}{2} \le \frac{X-5}{2} \le \frac{1-5}{2}\right) = P(-3.5 \le Z \le -2) = P(2 \le Z \le 3.5) = \Phi(3.5) - \Phi(2) = 0.0226.$

(c)

```
pnorm(1,5,2)
```

```
## [1] 0.02275013
```

```
pnorm(-2)
```

```
## [1] 0.02275013
```

```
pnorm(3.5)-pnorm(2)
```

```
## [1] 0.0225175
```