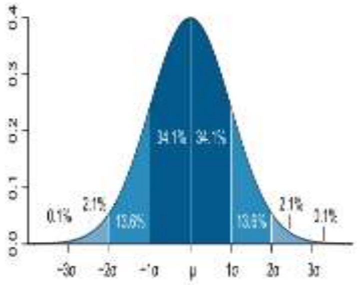
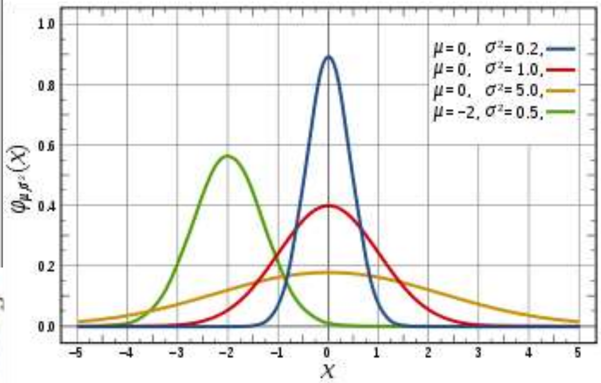
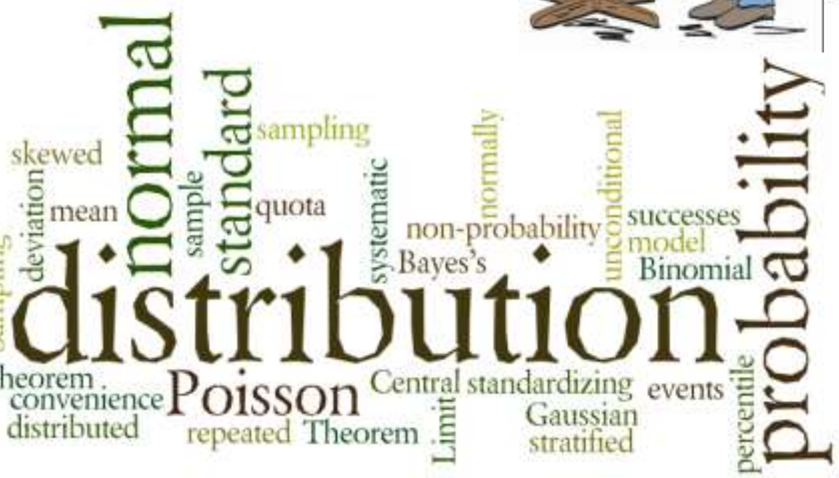
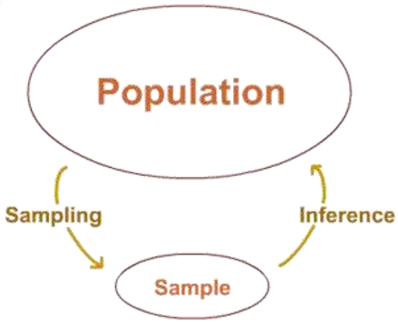
Continuous Probability

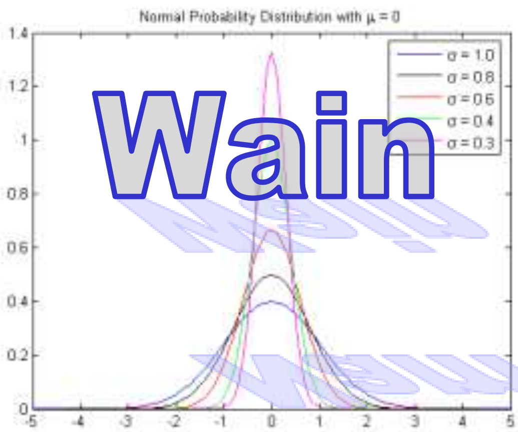


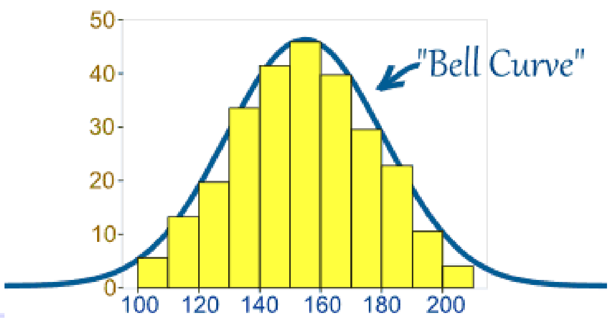




Name: Mr. Wain







Continuous Random Variables and their Probability Distributions

Continuous Random Variables

- A continuous random variable (CRV) is one that can take any value in an interval on the real number line.
- No limit on the accuracy, for example if someone's weight (kg) is given as 83, implies the weight is between 82.5 and 83.5, if the weight is given as 83.3, it implies the weight is between 83.25 and 83.35, etc..
- Therefore $\Pr(W = w) = 0$
- from text:

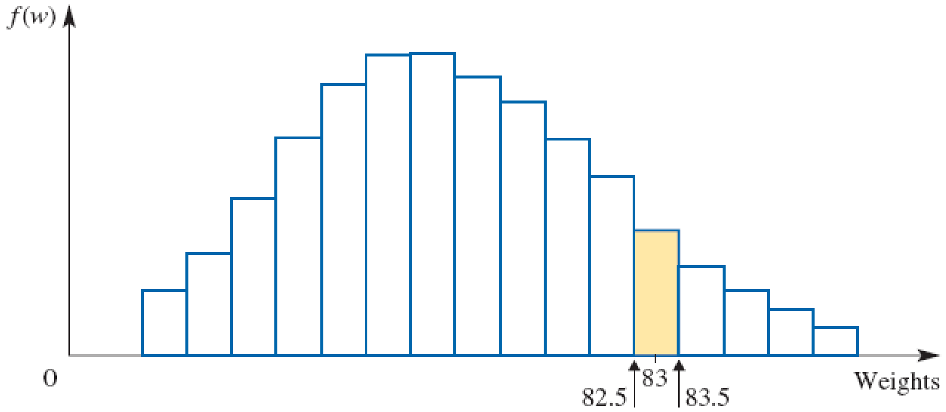
In practice, considering W taking a particular value is equivalent to W taking a value in an appropriate interval.

Thus, from above:

$$\Pr(W_0 = 83) = \Pr(82.5 \le W < 83.5)$$

To determine the value of this probability you could begin by measuring the weight of a large number of randomly chosen people, and determine the proportion of the people in the group who have weights in this interval.

Suppose after doing this a histogram of weights was obtained as shown.



-

From this histogram:

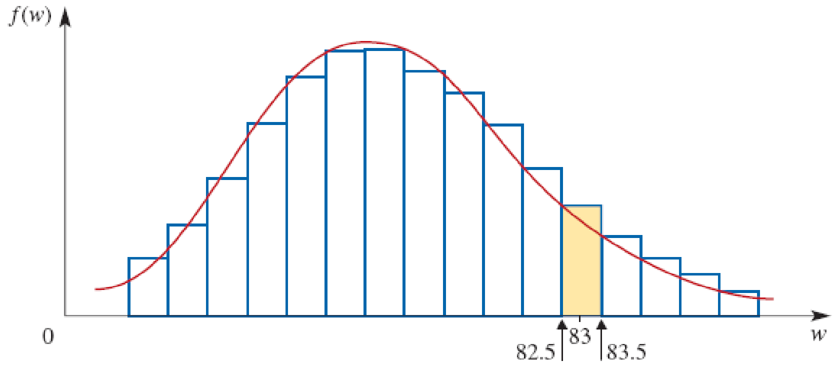
$$\Pr(W_0 = 83) = \Pr(82.5 \le W < 83.5) = \frac{\text{shaded area from 82.5 to 83.5}}{\text{total area}}$$

If the histogram is scaled so that the total area under the blocks is one, then:

$$\text{P}(W_0 = 83) = \Pr(82.5 \le W < 83.5) = \text{area under block from 82.5 to 83.5}$$

-

Suppose that the sample size gets larger and the class interval width gets smaller. If theoretically this process is continued so that the intervals are arbitrarily small, the histogram can be modelled by a smooth curve, as shown in the following diagram.



- The curve obtained here is of great importance for a continuous random variable.

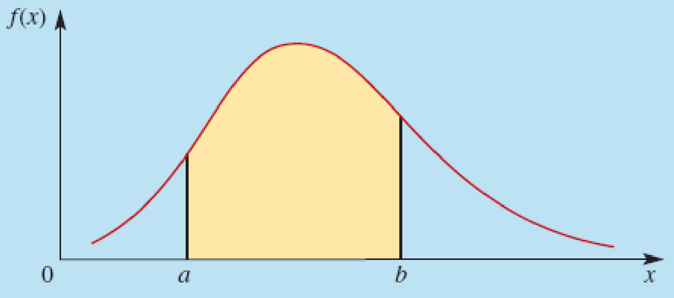
The function f whose graph models the histogram as the number of intervals is increased is called the **probability density function**. The probability density function f is used to describe the probability distribution of a continuous random variable, X.

Now, the probability of interest is no longer represented by the area under the histogram, but the area under the curve. That is:

$$\begin{aligned}\Pr(W_0 = 83) &= \Pr(82.5 \leq W < 83.5) \\ &= \text{area under the graph of the function with rule } f(w) \text{ from 82.5 to 83.5} \\ &= \int_{82.5}^{83.5} f(w)dw\end{aligned}$$

In general, for the continuous random variable X with density function f:

$\Pr(a < X < b) = \int_a^b f(x)dx$ is given by the area of the shaded region.



-
- In order to be a probability density function a function must satisfy certain conditions.

If the range of the continuous random variable X is $[a, b]$ then the domain of its probability density function f is $[a, b]$. The probability density function will satisfy two properties:

1 $f(x) \geq 0$ for all $x \in [a, b]$ and **2** $\int_a^b f(x)dx = 1$

Note that $\Pr(X < c) = \Pr(X \leq c)$

$$= \int_a^c f(x)dx$$

-

- A probability density function (or its natural extension) satisfies the following two properties:
  1 $f(x) \geq 0$ for all x **2** $\int_{-\infty}^{\infty} f(x)dx = 1$

- Note that, since the probability of X taking any exact value is zero, then:

  $$\Pr(a < X < b) = \Pr(a \leq X < b) = \Pr(a < X \leq b) = \Pr(a \leq X \leq b)$$

  That is, there is no difference between the numerical values of all of these expressions.

Example 1: Suppose that the random variable X has the density function with the rule:

$$f(x) = \begin{cases} cx & 0 \leq x \leq 2 \\ 0 & if \ x > 2 \ or \ x < 0 \end{cases}$$

a. Find the value of c that makes f a probability function;

b. Find $\Pr(X > 1.5)$

Solution: Calculator: solve(∫(cx,x,0,2)=1,c)

a.

$$\int_{-\infty}^{0} 0\,dx + \int_{0}^{2} cx\,dx + \int_{2}^{\infty} 0\,dx = 1$$
$$\left[\frac{cx^2}{2}\right]_0^2 = 1$$
$$\left[\frac{4c}{2}\right] - \left[\frac{0}{2}\right] = 1$$
$$2c = 1$$
$$c = 0.5$$

b.

$$\Pr(X > 1.5) = \int_{1.5}^{2} 0.5x\,dx$$
$$= 0.5\left[\frac{x^2}{2}\right]_{1.5}^{2}$$
$$= 0.5\left(\frac{4}{2} - \frac{2.25}{2}\right)$$
$$= 0.5 \times 0.875$$
$$= 0.4375$$

Example 2: Consider the function f with the rule:

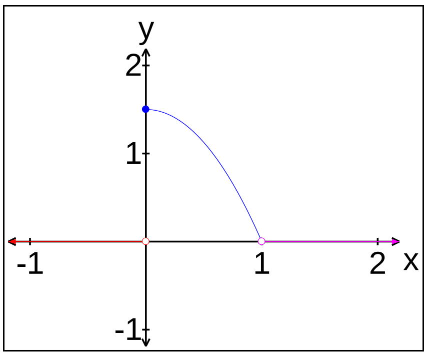
$$f(x) = \begin{cases} 1.5(1 - x^2) & 0 \leq x \leq 1 \\ 0 & if \ x > 1 \ or \ x < 0 \end{cases}$$

a. Sketch the graph of f;

b. Show that f is a probability density function;

c. Find $\Pr(X > 0.5)$.

Solution:

a. 

b. $\int_{-\infty}^{\infty} f(x)\,dx = \int_0^1 0.5(1-x^2)\,dx = 1.5\left[x - \frac{x^3}{3}\right]_{0.5}^{1} = 1.5\left[\left(1-\frac{1}{3}\right)-0\right] = 1.5\times\frac{2}{3} = \frac{3}{2}\times\frac{2}{3} = 1$

c. $\Pr(X > 0.5) = \int_{0.5}^{1} 0.5(1-x^2)\,dx = 1.5\left[x - \frac{x^3}{3}\right]_{0.5}^{1} = 1.5\left[\left(1-\frac{1}{3}\right)-\left(0.5-\frac{0.125}{3}\right)\right] = 0.3125$

Example 3: Consider the exponential probability density function f with the rule:

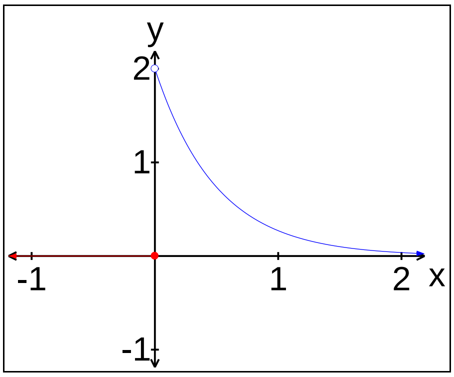
$$f(x) = \begin{cases} 2e^{-2x} & x > 0 \\ 0 & x \le 0 \end{cases}$$

a. Sketch the graph of f ;

b. Show that f is a probability density function;

c. Find $\Pr(X > 1)$.

Solution:

a. 

b. $\int_{-\infty}^{\infty} f(x)\,dx = \lim_{k\to\infty}\int_0^k 2e^{-2x}\,dx = \lim_{k\to\infty}\left[\frac{2e^{-2x}}{-2}\right]_0^k = \lim_{k\to\infty}\left(-e^{-2k}\right)-\left(-e^{0}\right) = 0 - -1 = 1$

Calculator: limit(∫($2e^{-2x}$,*x*,*o*,*k*),*k*,∞) or ∫($2e^{-2x}$,*x*,*o*,∞)

c. $\Pr(X > 1) = \lim_{k\to\infty}\int_1^k 2e^{-2x}\,dx = \lim_{k\to\infty}\left[-e^{-2x}\right]_1^k = \lim_{k\to\infty}(-e^{-2k}) - (-e^{-2}) = 0 + e^{-2} = \frac{1}{e^2} = 0.1353\ (4d.p.)$

- **Ex15A** 1, 2, 3, 4, 5, 6, 7, 9, 11, 15

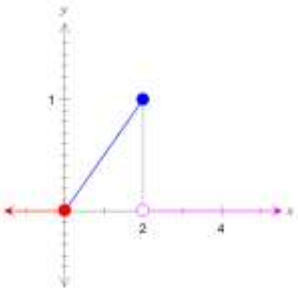
Mean, median and mode for a continuous random variable.

Mean (Expected Value)

$$E(X) = \int_{-\infty}^{\infty} xf(x)\,dx$$

Example: Find the expected value of the random variable *I* which has probability density function with rule:

$$f(x) = \begin{cases} 0.5x & 0 \le x \le 2 \\ 0 & x < 0 \text{ or } x > 2 \end{cases}$$



Solution: $E(X) = \int_{-\infty}^{\infty} xf(x)\,dx = \int_0^2 x(0.5x)\,dx = \int_0^2 0.5x^2\,dx = \left[\frac{x^3}{6}\right]_0^2 = \frac{8}{6} - 0 = \frac{4}{3}$

Example: If X is the random variable with probability density function f :

$$f(x) = \begin{cases} 0.5x & 0 \le x \le 2 \\ 0 & x < 0 \text{ or } x > 2 \end{cases}$$

Find:

a. the expected value of X^2;
b. the expected value of e^x (4 d.p).

Solution:

a. $E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx = \int_0^2 x^2(0.5x)\,dx = \int_0^2 0.5x^3\,dx = \left[\frac{x^4}{8}\right]_0^2 = \frac{16}{8} - 0 = 2$

b. $E(e^x) = \int_{-\infty}^{\infty} e^x f(x)\,dx = \int_0^2 e^x(0.5x)\,dx = \int_0^2 0.5e^x x\,dx = 4.1945$

Percentiles and the Median

50th percentile = $\int_{-\infty}^{p} f(x)\,dx = 0.5$

in general: the "nth" percentile, $\int_{-\infty}^{p} f(x)\,dx = \frac{n}{100}$

Example: The duration of telephone calls to the order department of a large company is a random variable *X* minutes with probability density function:

$$f(x) = \begin{cases} \frac{1}{3}e^{\left(\frac{-x}{3}\right)} & x > 0 \\ 0 & x \le 0 \end{cases}$$

Find the value of a such that 90% of phone calls last less than a minutes.

Solution:

$$\int_0^a \frac{1}{3}e^{\left(\frac{-x}{3}\right)}\,dx = 0.9$$

$$\left[-e^{\left(\frac{-x}{3}\right)}\right]_0^a = 0.9$$

$$-e^{\left(\frac{-a}{3}\right)} + 1 = 0.9$$

$$-e^{\left(\frac{-x}{3}\right)} = -0.1$$

solve(∫(1/2*e*^-*x*/3, *x*, 0, *a*)=0.9, *a*)

$$e^{\left(\frac{-x}{3}\right)} = 0.1$$

$$\frac{-x}{3} = \log_e 0.1$$

$$x = -3\log_e 0.1 \approx 6.908$$

therefore 90% of the calls to this company last less than 6.908 minutes.

Example: Suppose the probability density function of weekly sales of topsoil, *X* (in tonnes), is given by the rule:

$$f(x) = \begin{cases} 2(1-x) & 0 \le x \le 1 \\ 0 & x < 0 \quad or \quad x > 1 \end{cases}$$

Find the median value of *X* and interpret the result.

Solution:

$$\int_0^a (2-2x)\,dx = 0.5$$

$$\left[2x - x^2\right]_0^a = 0.5$$

$$2m - m^2 = 0.5$$

solve(∫(2-2*x*, *x*, 0, *a*)=0.5, *a*)

$$2m^2 - 4m + 1 = 0$$

$$m = 0.293 \; or \; m = 1.707$$

Because $0 \le x \le 1$ the median = 0.293 tonnes. This means 50% of the weekly sales will be less than 0.293 tonnes.

Mode

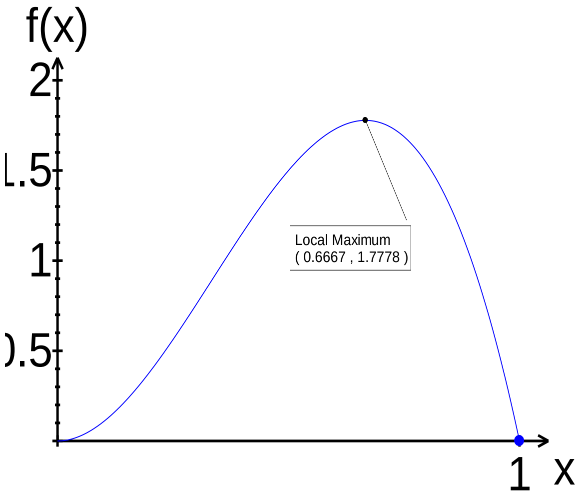
- the most common value of the variable *X*.
- for the probability density function this will be the local maximum.

Example: Find the mode of the CRV, *X* with the probability density function:

$$f(x)=\begin{cases}12x^2(1-x)\,dx & 0\le x\le 1\\ 0 & x<0 \text{ or } x>1\end{cases}$$

Solution:

$$f(x)=12x^2-12x^3, \quad x\in[0,1]$$
$$f'(x)=24x-36x^2=0$$
$$12x(2-3x)=0$$
$$x=0 \text{ or } x=\frac{2}{3}$$
$$\text{Maximum at } x=\frac{2}{3}$$



fmax(*f*(*x*),*x*) |0≤ *x* ≤1

- **Ex15B** 1, 2, 3, 4, 5, 7, 8, 10, 12, 13, 14, 15

Measures of Spread

$$Var(X) = \sigma^2 = E\left[(X-\mu)^2\right] = \int_{-\infty}^{\infty} (x-\mu)^2 f(x)\,dx = E(X^2) - \mu^2$$

$$sd(X) = \sqrt{Var(X)}$$

Example: Find the variance and standard deviation of the random variable *X* which has the probability density function with rule:

$$f(x) = \begin{cases} 0.5x & 0 \le x \le 2 \\ 0 & x < 0 \text{ or } x > 2 \end{cases}$$

Solution:

$$Var(X) = \sigma^2 = E(X^2) - \mu^2$$

$$E(X^2) = \int_0^2 x^2(0.5x)\,dx = \int_0^2 0.5x^3\,dx = \left[\frac{x^4}{8}\right]_0^2 = \frac{16}{8} - 0 = 2$$

$$E(X) = \mu = \int_0^2 x(0.5x)\,dx = \int_0^2 0.5x^2\,dx = \left[\frac{x^3}{6}\right]_0^2 = \frac{8}{6} - 0 = \frac{4}{3}$$

$$\therefore Var(X) = \sigma^2 = E(X^2) - \mu^2 = 2 - \left(\frac{4}{3}\right)^2 = \frac{2}{9}$$

$$sd(X) = \sqrt{\frac{2}{9}} = \frac{\sqrt{2}}{3}$$

Range

The range of a random variable is the difference between the smallest and largest value (of *X*). i.e calculate from the domain of the probability density function.

IQR

The Interquartile Range is the range of the middle 50% of the distribution. It is the difference between the 75th percentile and the 25th percentile.

Example: Find the IQR of the random variable *X* which has the probability density function:

$$f(x) = \begin{cases} 2x & 0 \le x \le 1 \\ 0 & x < 0 \text{ or } x > 1 \end{cases}$$

Solution:

25th percentile:

$$\int_0^a 2x\,dx = 0.25$$

$$\left[x^2\right]_0^a = 0.25$$

$$a^2 = 0.25$$

$$a = \pm 0.5 \qquad a = 0.5$$

75th percentile:

$$\int_0^a 2x\,dx = 0.75$$

$$\left[x^2\right]_0^a = 0.75$$

$$a^2 = 0.75$$

$$a = \pm\sqrt{0.75} \qquad a = \frac{\sqrt{3}}{2}$$

IQR= $\frac{\sqrt{3}}{2} - \frac{1}{2} = \frac{\sqrt{3}-1}{2} \approx 0.366$

- **Ex15C** 1, 2, 3, 5, 6, 9, 11

The Normal Distribution

Normal distributions are described by symmetric, bell-shaped curves. Area under the curve shows what decimal fraction of the observations lie in any region. The total area enclosed by the graph is equal to 1.

Standardised Normal Distribution
Example: There are two national (USA) college entrance exams, the Scholastic Aptitude Test (SAT) and the American College Testing program (ACT). Scores on the SAT's are approximately normal with a mean of 500 and a standard deviation of 100. Scores on the ACT are approximately normal with mean of 18 and a standard deviation of 6. Use this information to answer the following: Julie scores 630 on the mathematics part of the SAT. John takes the ACT mathematics test and scores 22. Assuming both tests measure the same kind of ability, who has the better result?

Solution:

Julie: Score is 1.3 standard deviations above the mean: $\frac{630-500}{100}=\frac{130}{100}=1.3$

John: Score is 0.67 standard deviations above the mean: $\frac{22-18}{6}=\frac{4}{6}=0.67$

Therefore Julie had the better result!

So to standardise a normal result:

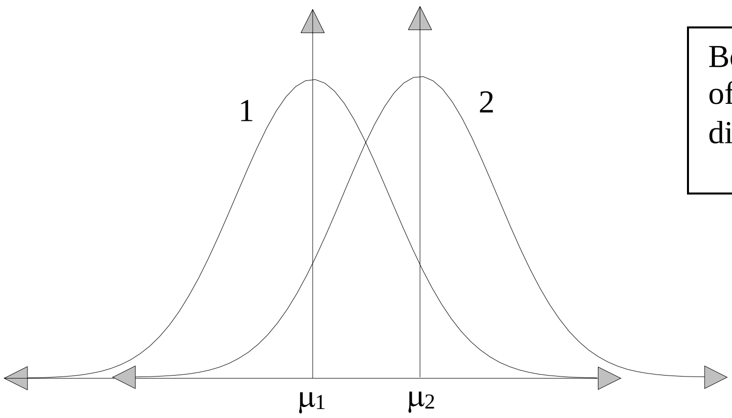
$$Z=\frac{x-\mu}{\sigma}$$ where Z is the standardised normal value.

Notation: The random variable X has a normal distribution with mean μ and variance σ^2 is written as: $X \sim N(\mu,\sigma^2)$.
In the above example for SAT : $X \sim N(500{,}10000)$

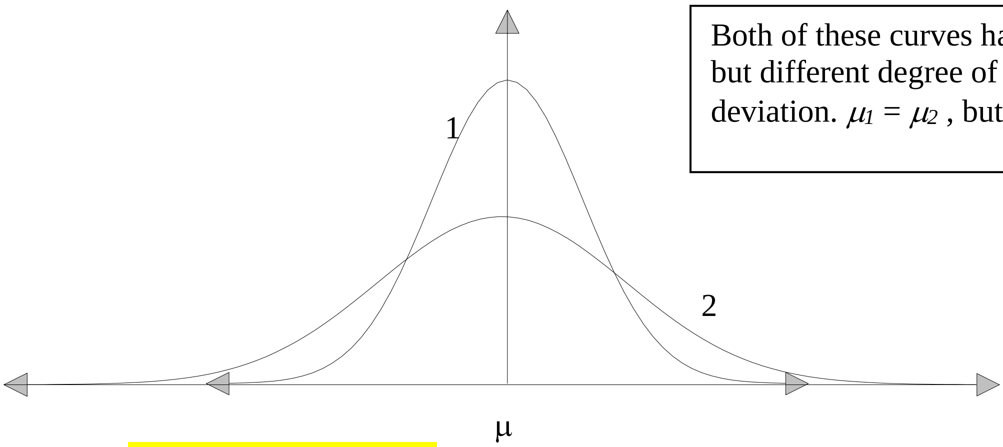
Whereas the Standard normal variable, Z , has a mean of 0 and a standard deviation of 1.
$Z \sim N(0,1)$

The Normal Probability Function is:

$$f(z)=\frac{1}{\sqrt{2\pi}}e^{\left(\frac{-1}{2}(z)^2\right)} \quad \text{or} \quad f(x)=\frac{1}{\sigma\sqrt{2\pi}}e^{\left(\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right)}$$



Both of these curves have the same degree of spread (i.e. standard deviation), but have different means. . $\mu_1 < \mu_2$, but $\sigma_1 = \sigma_2$



Both of these curves have the same mean but different degree of spread (i.e. standard deviation. $\mu_1 = \mu_2$, but $\sigma_1 < \sigma_2$

- **Ex16A** 1, 2, 3, 4, 6, 8, 9

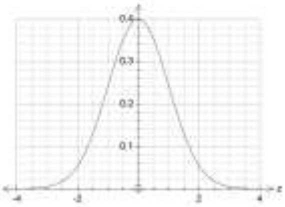
Standardised Values (Z values)

Example: Sketch the Normal Curve and shade the area for the following:

(a) (i) $\Pr(Z \le 2.5)$ (ii) $\Pr(Z < 0)$ (iii) $\Pr(Z \le 1.234)$

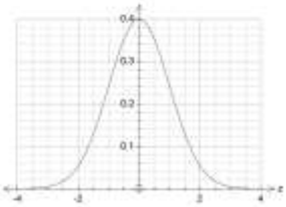
Solution:

(i)



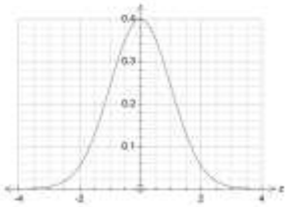
$\Pr(Z \le 2.5) = 0.9938$

(ii)



$\Pr(Z \le 0) = 0.5$

(iii)



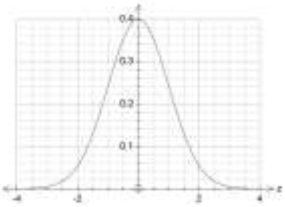
$\Pr(Z \le 1.234) = 0.8914$

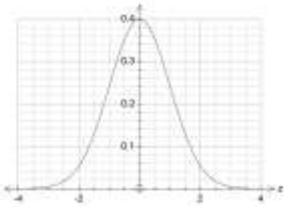
(b) (*Tech Free)

Given $\Pr(Z > 0.78) = 0.2177$ and $\Pr(Z \le 1.234) = 0.8914$ find

(i) $\Pr(Z \le -0.78) =$ $\Pr(Z \le -0.78) = \Pr(Z > 0.78) = 1 - \Pr(Z < 0.78) = 1 - 0.7823 = 0.2177$

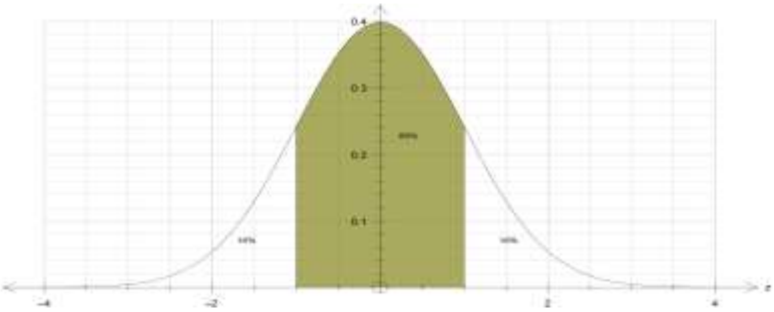
(ii) $\Pr(-0.78 \le Z \le 1.234) = \Pr(Z \le 1.234) - \Pr(Z \le -0.78) = 0.8914 - 0.2177 = 0.6737$



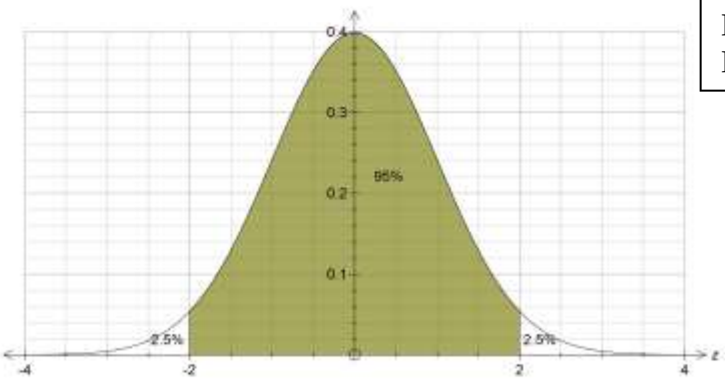


68 – 95 – 99.7 % rule

68%: $\mu - \sigma < X < \mu + \sigma$

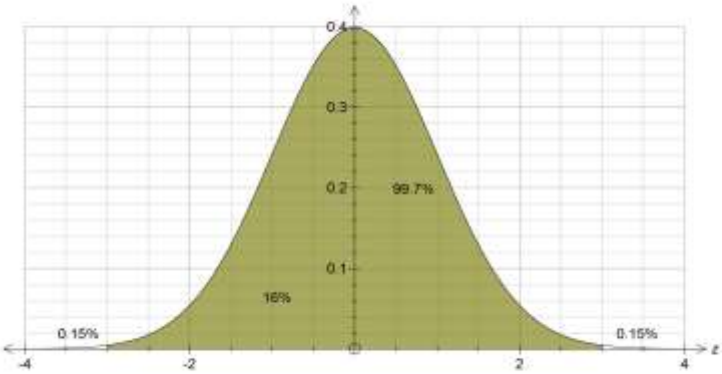


95%: $\mu - 2\sigma < X < \mu + 2\sigma$



Ti-nSpire:
Menu – Probability – Distributions
Normal Cdf

99.7%: $\mu - 3\sigma < X < \mu + 3\sigma$

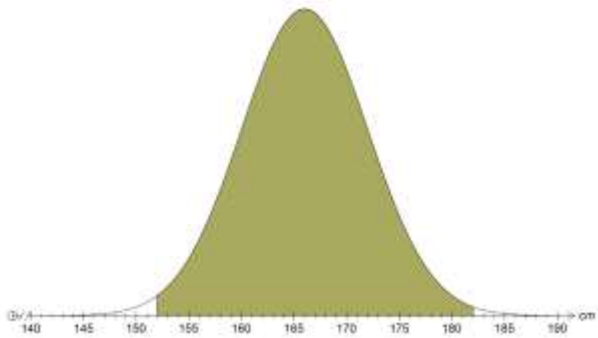


Example: The distribution of heights of young women is approximately normal with mean, $\mu = 166$ cm and standard deviation, $\sigma = 6$.

(a) Using the curve below, mark its mean and standard deviations.

(b) Between what two values do the heights of the central 95% of young women lie? Shade this area on the curve.

(b) 154 - 178



- **Ex16B** 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15

Inverse Problems

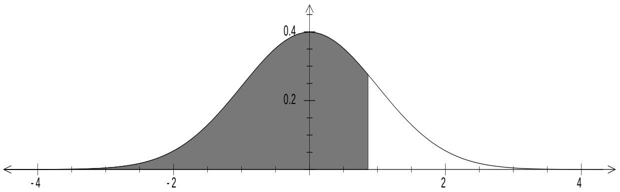
Example: Find by first drawing a diagram.

(a) $\Pr(Z < k) = 0.8$ (b) $\Pr(Z < k) = 0.6$ (c) $\Pr(Z < k) = 0.95$

(d) $\Pr(Z < k) = 0.3$ (e) $\Pr(Z > k) = 0.15$ (f) $\Pr(Z > k) = 0.65$

Solution:

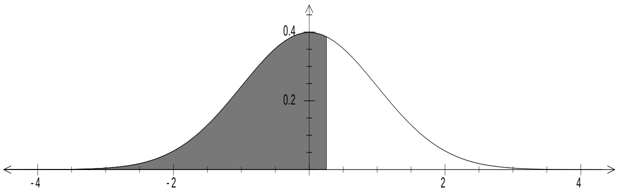
(a)



$\Pr(Z < k) = 0.8$

$\therefore k = 0.842$

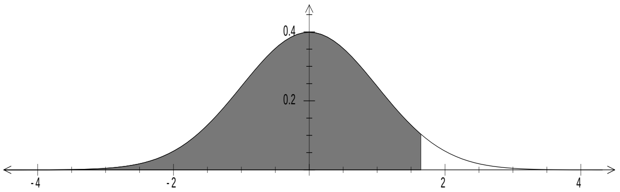
(b)



$\Pr(Z < k) = 0.6$

$\therefore k = 0.253$

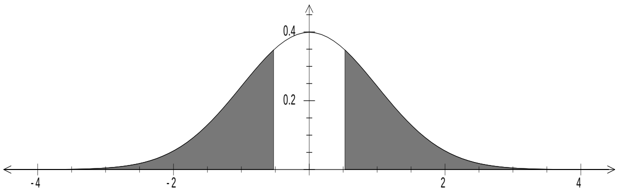
(c)



$\Pr(Z < k) = 0.95$

$\therefore k = 1.645$

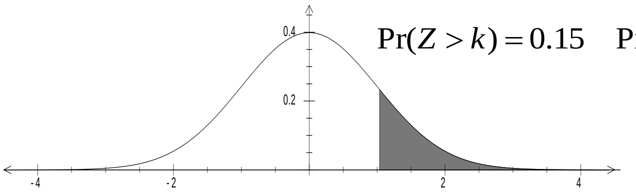
(d)



$\Pr(Z < k) = 0.3$ $\Pr(Z < m) = 0.7$

$\therefore k = -0.524$

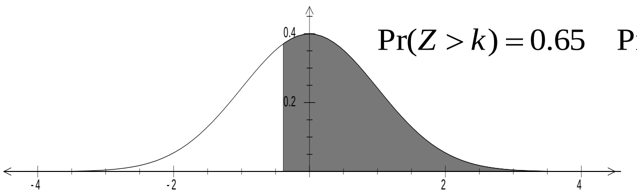
(e)



$\Pr(Z > k) = 0.15$ $\Pr(Z < k) = 0.85$

$\therefore k = 1.036$

(f)



$\Pr(Z > k) = 0.65$ $\Pr(Z < m) = 0.65$

$\therefore k = -0.385$

- **Ex16C** 1 adeh, 2 abfh, 4, 5, 6, 7, 8, 9, 10 , 11

Normal & Inverse Applications

Example

A manufacturer of electric light globes finds that his articles have an average life of 1200 burning hours with a standard deviation of 200 hours. Assuming that the distribution of life-times is normal,
(i) what is the probability of a globe selected at random having a life between 1240 and 1320 hours?
(ii) out of a batch of 200 globes, how many would be expected to fail in the first 880 burning hours?
(iii) what proportion of globes manufactured would be expected to have a life less that 1100 hours or more than 1460 hours?

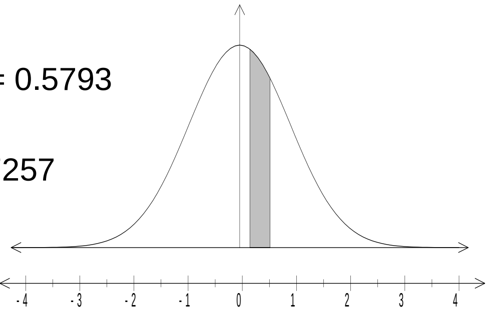
Solution:

Let X = the number of burning hours

(a) $\Pr(X < 1240) = \Pr\left(Z < \frac{1240-1200}{200}\right) = \Pr(Z < 0.2) = 0.5793$

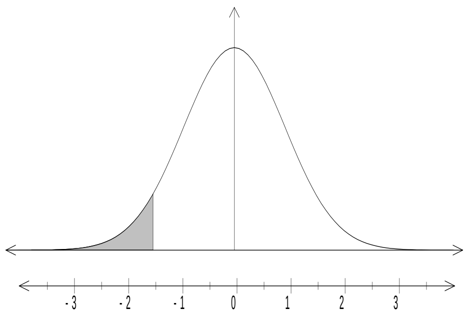
$\Pr(X<1320) = \Pr\left(Z < \frac{1320-1200}{200}\right) = \Pr(Z < 0.6) = 0.7257$

$\Pr(1240< X < 1320)= 0.7257 - 0.5793 = 0.1464$



(b) $\Pr(X < 880) = \Pr(Z< -1.6) = \Pr(Z > 1.6) = 1 - \Pr(Z < 1.6)= 1 - 0.9452 = 0.0548$

Expected number = 200 x 0.0548 = 11

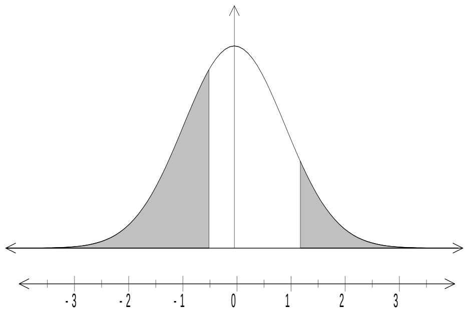


(c) $\Pr(X < 1100) = \Pr\left(Z < \frac{1120-1200}{200}\right)$

$= \Pr(Z < -0.5) = 0.3085$

$\Pr(X> 1460) = \Pr\left(Z > \frac{1460-1200}{200}\right) = \Pr(Z > 1.3)$

$= 1 - \Pr(Z < 1.3)= 1 - 0.9032 = 0.0968$



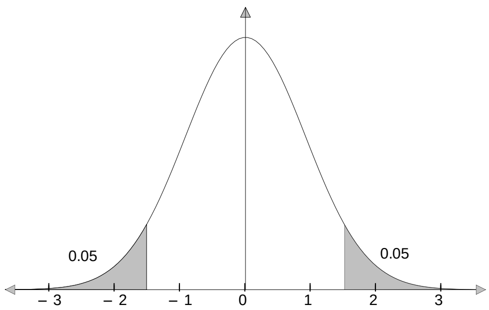
$\Pr(X< 1100 \text{ or } X > 1460) = 0.3085 + 0.0968 =$
0.4053 (40.53 %)

- **Ex16C** 12, 13, 17 abc, 18 ab; **Ex16D** 1a, 2a, 3a ,4

Example

A lathe turns out brass cylinders with a mean diameter of 2.16 cm and a standard deviation of 0.08 cm. Assuming that the distribution of diameters is normal, find the limits to the acceptable diameters if, on checking, it is found that 5 per cent in the long run rejected because they are oversize and 5 per cent are rejected because they are undersize.

Solution



$$Z = \frac{x-\mu}{\sigma}$$

$$\pm 1.6449 = \frac{x-2.16}{0.08}$$

$$x = 2.16 \pm 0.08 \times 1.6449$$

$$x = 2.16 \pm 0.13$$

$$x = 2.29 \ or \ 2.03 \ cm$$

- **Ex16C** 14, 15, 17 def, 18cde; **Ex16D** 1b, 2 bc, 3 b, 5, 6

Example: The lengths of newborn girls are normally distributed with a mean of 49.50cm and a variance of 3.24 cm.

(i) Girls A's length at birth was 53.3cm. What percentile is she in for height (length)?

(ii) Her sister, Girl B, was in the 97th percentile for length. How long was she at birth?

(iii) Girl C was in the 25th percentile for length at birth. How long was she?

Solution:

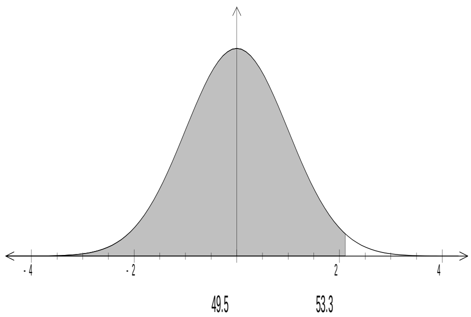
(i) Let X = birth length in cm.

$$\Pr(X < 53.3)$$

$$z = \frac{x-\mu}{\sigma} \qquad \mu = 49.5 \quad \sigma = \sqrt{3.24} = 1.8$$

$$z = \frac{53.3-49.5}{1.8} = 2.111$$

$$\Pr(Z < 2.111) = 0.9826$$



Therefore the girl is in the 98th percentile. This means 98% of all girls are shorter than girl A.

(ii)

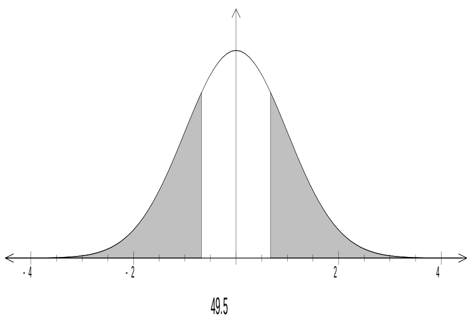
$$\Pr(X < ?) = 0.97$$

$$\therefore \Pr(Z < 1.881) = 0.97$$

$$1.881 = \frac{x-49.5}{1.8}$$

$$1.881(1.8) + 49.5 = x$$

$$x = 52.9cm$$

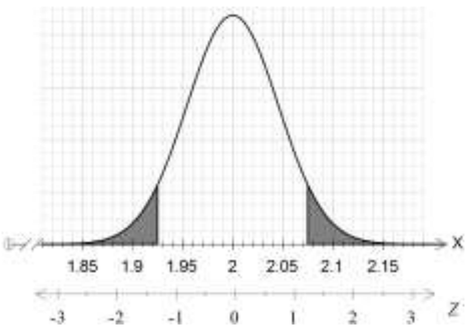


(iii)

$$\Pr(X < ??) = 0.25$$
$$\Pr(Z < -0.674) = 0.25$$
$$\therefore -0.674 = \frac{x - 49.5}{1.8}$$
$$-0.674(1.8) + 49.5 = x$$
$$x = 48.3cm$$

Example: Limits of acceptability imposed on the lengths of a certain batch of metal rods are 1.925 cm and 2.075 cm. It is observed that, on average, 5% are rejected as undersized and 5% are rejected as oversized.
Assuming that the lengths are normally distributed, find the mean and standard deviation of the distribution.

Solution:
Let X = the length of the metal rods (cm)
$X \sim N(\mu, \sigma^2)$



$$\Pr(X > 2.075) = 0.05 \qquad \therefore \Pr(Z > 1.6448) = 0.05$$
$$\Pr(X < 1.925) = 0.05 \qquad \therefore \Pr(Z < -1.6448) = 0.05$$

$$\therefore 1.6448 = \frac{2.075 - \mu}{\sigma} \quad \& \quad -1.6448 = \frac{1.925 - \mu}{\sigma}$$

Solve simultaneously

$$\mu = 2cm \quad \& \quad \sigma = 0.0456cm$$

- **Ex16D** 7, 10, 11; **Ch16** Review **TF** 1, 2, 3, 4, 5, 6, 7, 8, 9, 10 **ER** 7, 8, 9, 10

The Normal Approximation to the Binomial Distribution

- We know the shape of the binomial distribution depends on n and p.
- If n is large and p is not too close to 0 or 1 then the distribution is close to symmetrical and starts to look like a Normal distribution
- When to use the normal approximation?
  - If n is sufficiently large (~25* or more) *depends on p also
  - P is not too close to 0 or 1
  - **General rule: np and $n(1 - p)$ is greater than 5**
- If satisfied, then the Binomial random variable, X, will be approximately normally distributed with:
  - a mean of $\mu = np$;
  - a standard deviation of $\sigma = \sqrt{np(1-p)}$

Example: A sample of 1000 people from a certain city were asked to indicate whether or not they were in favour of the construction of a new freeway. It is known that 30% of people in this city are in favour of the new freeway. Find the approximate probability that between 270 and 330 people in the sample were in favour of the new freeway.

Solution:

Let X = the number of people in the sample who are in favour of the freeway.
X~Bi(n, p) $n = 1000$ & $p = 0.3$

$$\mu = n \times p = 1000 \times 0.3 = 300$$

$$\sigma = \sqrt{np(1-p)} = \sqrt{1000 \times 0.7 \times 0.3} = \sqrt{210}$$

$$\Pr(270 < X < 330) \approx \Pr\left(\frac{270-300}{\sqrt{210}} < Z < \frac{330-300}{\sqrt{210}}\right) \approx \Pr(-2.070 < Z < 2.070) \approx 0.9616$$

Note: if we used the Binomial distribution to calculate this we obtain:

$$\Pr(270 < X < 330) = 0.9583 \quad \text{and} \quad \Pr(270 \le X \le 330) = 0.9648$$

- **Ex16E** 1, 2, 3, 4, 5, 6

Sampling & Estimation

Populations & Samples

- A **population** is the set of all eligible members of a group which we intend to study.
- A **sample** is a subset of the population.
- A sample is used to make inferences about the population.
- We use sampling when:
    - The population is too large – e.g. all AFL supporters in Australia.
    - The population may be hard to access – e.g. all Blue Whales in the Pacific Ocean.
    - The data collection process may be destructive – e.g. testing every battery to see how long it lasts, mean that there would be no batteries left to sell.

Random Sample

- When we select our sample, we don’t want it to be biased towards a sub-group. For example, surveying people as they enter the MCG and asking “do enjoy Australian Rules football.
- A sample size then has to be selected in a way that every subset of the population has an equal chance of being selected.
- We call this the **Simple Random Sample**.
- Many methods can be used. A common method is to assign every member of the population a number and then use a **random number generator** to select the sample.

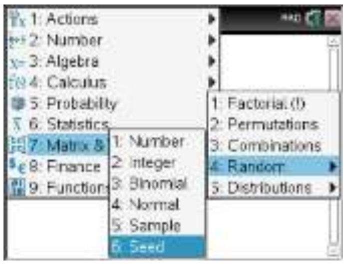
Example

A researcher wishes to evaluate how well the local library is catering to the needs of a town’s residents. To do this she hands out a questionnaire to each person entering the library over the course of a week. Will this method result in a random sample?

Probably not, since the members of the sample are already users of the Library. Would be better sampling from the whole of the town’s population.

Random Number Generator on the Ti-NSpire

- **Menu – Probability – Random – Seed** then enter 4 random numbers, e.g. the last 4 numbers of your phone number. (e.g. RandSeed 8595)
- To get a random number between 0 and 1 use **Menu – Probability – Random – Number.** (e.g. rand())
- For a random integer - **Menu – Probability – Random – Integer** .
    - To get 5 random integers between 1 and 30: randInt(1,30,5)



Example: Use a random number generator to select 6 students from the following class:

John	Mary	Bill	Jane	Fred	Lily
Shay	Rose	Lucy	Declan	Patrick	Louis
Harry	Ruby	Georgia	Tom	Darren	Alice
Kristin	Lee-Anne	Glen	Tim	Steve	Lachlan

Solution:
First assign numbers to each person in the class

John (1)	Mary (2)	Bill (3)	Jane (4)	Fred (5)	Lily (6)
Shay (7)	Rose (8)	Lucy (9)	Declan (10)	Patrick (11)	Louis (12)
Harry (13)	Ruby (14)	Georgia (15)	Tom (16)	Darren (17)	Alice (18)
Kristin (19)	Lee-Anne (20)	Glen (21)	Tim (22)	Steve (23)	Lachlan (24)

Then generate 6 random integers without repeats from 1 to 24.

e.g. 4, 15, 9, 7, 22, 13

which gives: Jane, Georgia, Lucy, Shay, Tim & Harry

The sample proportion as a random variable

- $\text{Population proportion } p = \dfrac{\text{number in population with attribute}}{\text{population size}}$
- In the above example there are 11 females and 13 males. So the proportion of female students in the class is $p = \dfrac{11}{24}$
- The **Sample Proportion,** $\hat{p}$, $\hat{p} = \dfrac{\text{number in sample with attribute}}{\text{sample size}}$ $\hat{p} = \dfrac{x}{n}$
- The proportion of females in our sample of 6 is $\hat{p} = \dfrac{3}{6} = \dfrac{1}{2}$
- The population proportion, p, is a **population parameter**; its value is constant.
- The sample proportion, $\hat{p}$, is a **sample statistic**; its value is not constant.

Example: Use the same data from the previous example, select another random group of six. Determine the proportion of females in the sample.

Solution:

e.g. Chosen 6 are: 5, 6, 11, 19, 24, 10: Fred, Lily, Patrick, Kristin, Lachlan, Declan

$$\hat{p} = \frac{2}{6} = \frac{1}{3}$$

- **Ex17A** 1, 2, 3, 5, 6, 8, 12, 13, 14

The exact distribution of the sample proportion

Sampling from a small population

Example: A bag contains six blue and four red balls. If we take a random sample of size 4, what is the probability that there is one blue ball in the sample ($\hat{p} = \frac{1}{4}$) ?

Solution:
Method 1:
This can be done in 4 different ways: RRRB or RRBR or RBRR or BRRR

$$\Pr(RRRB, RRBR, RBRR, BRRR) =$$
$$\left(\frac{4}{10}\times\frac{3}{9}\times\frac{2}{8}\times\frac{6}{7}\right)+\left(\frac{4}{10}\times\frac{3}{9}\times\frac{6}{8}\times\frac{2}{7}\right)+\left(\frac{4}{10}\times\frac{6}{9}\times\frac{3}{8}\times\frac{2}{7}\right)+\left(\frac{6}{10}\times\frac{4}{9}\times\frac{3}{8}\times\frac{2}{7}\right)$$
$$= 4\times\frac{2}{70} = \frac{8}{70} = \frac{4}{35}$$

Method 2:

There are $\binom{10}{4} = 210$ ways to select 4 balls from 10 balls.

There are $\binom{4}{3} = 4$ ways to select 3 red balls from 4 red balls and there are $\binom{6}{1} = 6$ ways to select 1 blue ball from 4 blue balls.

Therefore the probability of obtaining 3 reds and 1 blue is $\dfrac{\binom{4}{3}\times\binom{6}{1}}{\binom{10}{4}} = \frac{4\times 6}{210} = \frac{24}{210} = \frac{4}{35}$

The following is an extension of the previous example. This table gives the probability of obtaining each possible sample proportion $\hat{p}$ when selecting a sample size of 4.

Number of blue balls in the sample (x)	0	1	2	3	4
Proportion of blue balls in the sample, $\hat{p}$	0	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{3}{4}$	1
Probability	$\frac{1}{210}$	$\frac{24}{210}$	$\frac{90}{210}$	$\frac{80}{210}$	$\frac{15}{210}$

The last 2 rows form a probability distribution for the random variable, $\hat{p}$.

The distribution of a statistic which is calculated from a sample is called a **sampling distribution.**

Example: A bag contains six blue balls and four red balls. Use the sampling distribution in the previous table to determine the probability that the proportion of blue balls in a sample of size 4 is more than $\frac{1}{4}$.

Solution:

$$\Pr\left(\hat{P} > \frac{1}{4}\right) = \Pr\left(\hat{P} > \frac{1}{2}\right) + \Pr\left(\hat{P} > \frac{3}{4}\right) + \Pr\left(\hat{P} > 1\right)$$
$$= \frac{90}{210} + \frac{80}{210} + \frac{15}{210} = \frac{185}{210} = \frac{37}{42}$$

Sampling from a large population

This is the most common case.

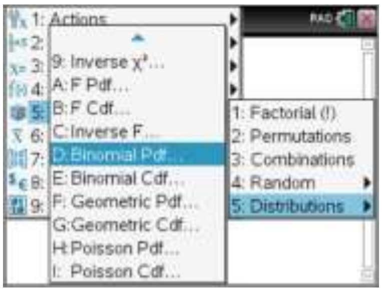
For example, suppose 70% of all 17 year olds in Australia attend school. i.e. $p = 0.7$

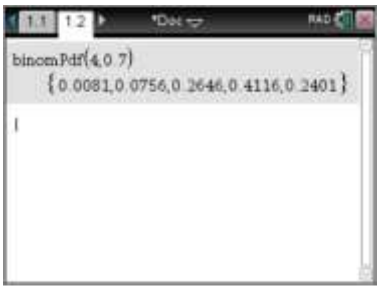
Now consider selecting a random sample of size 4 from the population of all 17 year olds in Australia. As this binomial then:

$$\Pr(X = x) = \binom{4}{x} 0.7^x 0.3^{4-x} \qquad x = 0,1,2,3,4$$

The sampling distribution will look like this:

Number at school in the sample (x)	0	1	2	3	4
Proportion at school in the sample, $\hat{p}$	0	0.25	0.5	0.75	1
Probability $\Pr(\hat{P} = \hat{p})$	0.0081	0.0756	0.2646	0.4116	0.2401





The population that the sample of size = 4 is being taken from is such that each item selected has a probability of $p = 0.7$ of success.

So we define the random variable, $\hat{p} = \frac{X}{4}$, $\hat{p} = \frac{x}{n}$ where X is a binomial random variable, $p = 0.7$ and $n = 4$. So the above table could be written as:

x	0	1	2	3	4
$\hat{p}$	0	0.25	0.5	0.75	1
$\Pr(\hat{P} = \hat{p}) = \Pr(X = x)$	0.0081	0.0756	0.2646	0.4116	0.2401

For example the probability, in the random sample of 4, the proportion attending school less than 50% is:

$$\Pr(\hat{P} < 0.5) = \Pr(\hat{P} = 0) + \Pr(\hat{P} = 0.25)$$
$$= 0.0801 + 0.0756$$
$$= 0.0837$$

The Mean and Standard Deviation of the sample proportion

Example: Use the following distribution to determine the mean and standard deviation of the sample proportion $\hat{P}$ from the previous example of 17 year olds.

x	0	1	2	3	4
$\hat{p}$	0	0.25	0.5	0.75	1
$\Pr\left(\hat{P}=\hat{p}\right)=\Pr(X=x)$	0.0081	0.0756	0.2646	0.4116	0.2401

Solution:

$$E\left(\hat{P}\right)=\sum \hat{p}\times\Pr\left(\hat{P}=\hat{p}\right)$$
$$=0\times0.0081+0.25\times0.0756+0.50\times0.2646+0.75\times0.4116+1\times0.2401$$
$$=0.7$$

$$sd\left(\hat{P}\right)=\sqrt{E(\hat{P}^2)-\left[\left(E(\hat{P}\right)\right]^2}$$
$$E\left(\hat{P}^2\right)=\sum \hat{p}^2\times\Pr\left(\hat{P}=\hat{p}\right)$$
$$=0^2\times0.0081+0.25^2\times0.0756+0.50^2\times0.2646+0.75^2\times0.4116+1^2\times0.2401$$
$$=0.5425$$

$$sd\left(\hat{P}\right)=\sqrt{0.5425-0.7^2}=0.2291$$

If we are selecting a random sample of size n from a large population, then the mean and standard deviation of the sample proportion $\hat{P}$ are given by:

$$E\left(\hat{P}\right)=p \qquad \text{and} \qquad sd\left(\hat{P}\right)=\sqrt{\frac{p(1-p)}{n}}$$

Note: the standard deviation of a sample statistic is called the **standard error**.

From our 17 year olds example, we can see that these are the same as calculated in the previous example:

p = 0.7

$$sd\left(\hat{P}\right)=\sqrt{\frac{p(1-p)}{n}}=\sqrt{\frac{0.7\times0.3}{4}}=0.2291$$

Example: Suppose that 70% of 17 year olds in Australia attend school. If a random sample of size 20 is chosen from this population, find the probability that the sample proportion:

a is equal to the population proportion (0.7);

b lies within one standard deviation of the population proportion;

c lies within two standard deviations of the population proportion.

Solution:

a

If the sample proportion is $\hat{p} = 0.7$ and the sample size is 20, then the number of students in the sample is 0.7 x 20 = 14.

$$\Pr\left(\hat{P} = 0.7\right) = \Pr(X = 14) = \binom{20}{14} \times 0.7^{14} \times 0.3^{6} = 0.1916$$

b

$$sd\left(\hat{P}\right) = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.7 \times 0.3}{20}} = 0.1025$$

$1\ s.d.\quad from\quad the\quad mean: 0.7 - 0.1025 = 0.5975\quad and\quad 0.7 + 0.1025 = 0.8025$

$0.5975 \times 20 = 11.95 \quad \& \quad 0.8025 \times 20 = 16.05$

$\Pr(0.5975 \le \hat{P} \le 0.8025) = \Pr(11.95 \le X \le 16.05) = \Pr(12 \le X \le 16) = 0.7796$

c

$2\ s.d.\quad from\quad the\quad mean: 0.7 - 2 \times 0.1025 = 0.495\quad and\quad 0.7 + 2 \times 0.1025 = 0.905$

$\Pr(0.495 \le \hat{P} \le 0.905) = \Pr(9.9 \le X \le 18.1) = \Pr(10 \le X \le 18) = 0.9752$

- **Ex17B** 1, 2, 4, 5, 8, 11, 12, 14

Approximating the distribution of the sample proportion

If the sample size, *n*, is large it is impractical to work out probabilities associated with the sample proportion.
We overcome this by approximating the distribution of the sample proportion.
As we have already seen a binomial distribution is well approximated by the normal distribution (**remember *np* and *np(1 – p)* should be greater than 10**).

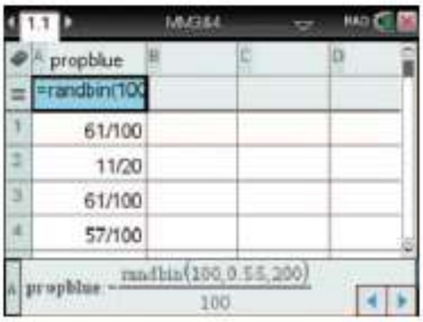
- $\mu = p$ $\qquad\qquad$ $\sigma = \sqrt{\dfrac{p(1-p)}{n}}$

Example: Assume 55% of people in Australia have blue eyes. Use your calculator to illustrate a possible distribution of sample proportions, $\hat{p}$, that may be obtained when 200 different samples (each of size 100) are selected from the population.

Solution:

Using the TI-Nspire

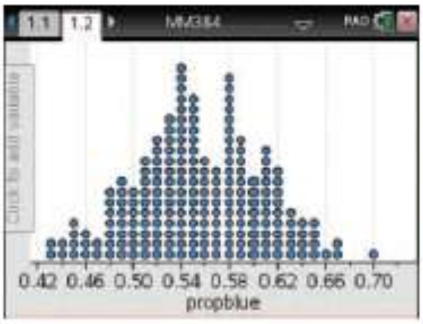
- To generate the sample proportions:
  - Start from a **Lists & Spreadsheet** page.
  - Name the list 'propblue' in Column A.
  - In the formula cell of Column A, enter the formula using (Menu) > **Data** > **Random** > **Binomial** and complete as: = randbin(100, 0.55, 200)/100



Note: The syntax is: randbin(*sample size*, *population proportion*, *number of samples*)
To calculate as a proportion, divide by the sample size.

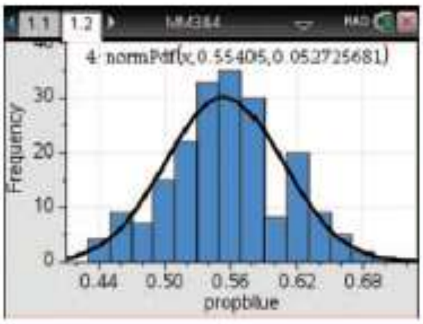
- To display the distribution of sample proportions:
  - Insert a **Data & Statistics** page ((ctrl)(I) or (ctrl)(doc ▾)).
  - Click on 'Click to add variable' on the *x*-axis and select 'propblue'. A dotplot is displayed.

Note: You can recalculate the random sample proportions by using (ctrl)(R) while in the **Lists & Spreadsheet** page.



- To fit a normal curve to the distribution:
  - (Menu) > **Plot Type** > **Histogram**
  - (Menu) > **Analyze** > **Show Normal PDF**

Note: The calculated Normal PDF, based on the data set, is superimposed on the plot, showing the mean and standard deviation of the sample proportion.



Example: Assume 60% of people have driver's licence. Using the normal approximation, find the approximate probability that, in a randomly selected sample size 200, more than 65% of people have a driver's licence.

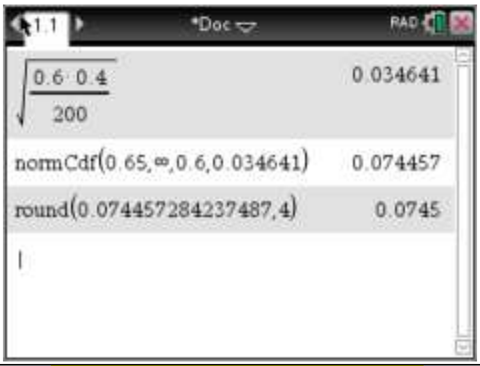
Solution:

$n = 200$ & $p = 0.6$ and since n is large the distribution of $\hat{P}$ is approximately normal,

so $\mu = p = 0.6$

$$\sigma = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.6 \times 0.4}{200}} = 0.0346 \quad (0.034641)$$

$$\Pr\left(\hat{P} > 0.65\right) = 0.0745$$



- **Ex17C** 1, 2, 4, 5, 7, 9, 10

Confidence intervals for the population proportion

Point estimates
The value of the sample proportion $\hat{p}$ can be used to estimate the population proportion *p*. since this is a single-valued estimate, it is called the **point estimate** of *p*.

Interval estimates
An **interval estimate** for the population proportion *p* is called a **confidence interval** for *p*.

Using the Normal approximation and standardising, remember $Z = \frac{x-\mu}{\sigma}$.

We know that $\Pr(-1.96 < Z < 1.96) = 0.95$, so

$$\Pr\left(-1.96 < \frac{\hat{P}-p}{\sqrt{\frac{p(1-p)}{n}}} < 1.96\right) \approx 0.95 \text{ which can be written as}$$

$$\Pr\left(-1.96\sqrt{\frac{p(1-p)}{n}} < \hat{P}-p < 1.96\sqrt{\frac{p(1-p)}{n}}\right) \approx 0.95 \text{ then}$$

$$\boxed{\Pr\left(\hat{P}-1.96\sqrt{\frac{p(1-p)}{n}} < p < \hat{P}+1.96\sqrt{\frac{p(1-p)}{n}}\right) \approx 0.95}$$

This is the **95% confidence interval** for *p*.

- *p* is the population proportion;
- $\hat{p}$ is the value of the sample proportion;
- *n* is the size of the sample from which $\hat{p}$ was calculated.

Example: Find an approximation 95% confidence interval for the proportion *p* of primary school children in Australia who regularly use social media, if we select a random sample of 20 children and find the sample population $\hat{p}$ to be 0.7.

Solution:

$\hat{p}$ = 0.7 and *n* = 20

$$\sigma = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.7\times 0.3}{20}} = 0.1025$$

so 95%:

$$(0.7-1.96\times 0.1025, 0.7+1.96\times 0.1025) = (0.499, 0.901)$$

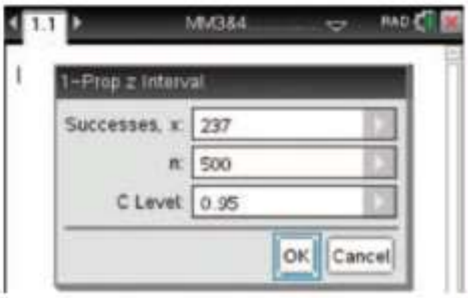
So based on a sample size of 20 there is a 95% confidence interval for the population proportion p is (0.499, 0.901).

Using a calculator to determine confidence intervals

Using the TI-Nspire

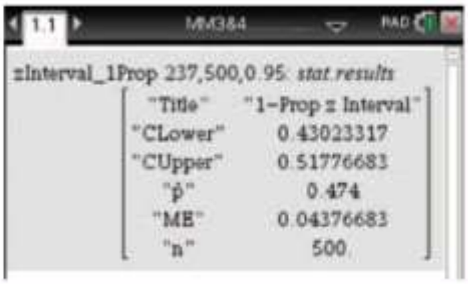
In a **Calculator** page:

- Use (Menu) > **Statistics** > **Confidence Intervals** > **1–Prop z Interval**.
- Enter the values $x = 237$ and $n = 500$ as shown.



- The 'CLower' and 'CUpper' values give the 95% confidence interval (0.43, 0.52).

Note: 'ME' stands for margin of error, which is covered in the next subsection.



Precision and margin of error

In the example with the school children, we found the 95% confidence interval (0.499, 0.901) for the proportion *p* of primary school children who use social media, based on a sample size of 20.
This means we are predicting the population proportion is somewhere in the range of approximately 50% to 90%. Not a good prediction.

If we change the random sample size to 200:

$P = 0.7$, $n = 200$ $\quad \Rightarrow \quad \sigma = \sqrt{\dfrac{p(1-p)}{n}} = \sqrt{\dfrac{0.7 \times 0.3}{200}} = 0.0324$

So the confidence interval is calculated as:

$$\boxed{\left(0.7 - 1.96\sqrt{\frac{0.7 \times 0.3}{200}},\quad 0.7 + 1.96\sqrt{\frac{0.7 \times 0.3}{200}}\right) = (0.636, 0.764)}$$

So based on a sample size of 200 and a sample estimate of 0.7, an approximate 95% confidence interval for the population proportion *p* is (0.636, 0.764) i.e. between 64% and 76%.

The difference between the sample estimate and the confidence interval endpoints is called the **Margin of Error, *M*.** For a 95% confidence interval it is given as:

$$\boxed{M = 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}}$$

Example: Determine the sample size required to achieve a margin of error of 2% in an approximate 95% confidence interval for the proportion p of primary school children in Australia who use social media, if the sample proportion $\hat{p}$ is found to be 0.7.

Solution:

$$M = 0.02 \ \& \ \hat{p} = 0.7 \quad M = 1.96\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$$

$$0.02 = 1.96\sqrt{\frac{0.7\times 0.3}{n}}$$

$$\left(\frac{0.02}{1.96}\right)^2 = \frac{0.7\times 0.3}{n} \quad \Rightarrow \quad n = \frac{0.7\times 0.3}{\left(\frac{0.02}{1.96}\right)^2} = 2016.84$$

So to achieve a margin of error of 2%, we need a sample size of 2017.

A 95% confidence interval for a population proportion p will have margin of error approximately equal to a specified value of M when the sample size is:

$$n = \left(\frac{1.96}{M}\right)^2 p^*(1-p^*)$$

p^* is an estimated value for the population proportion p.

In general, a C% confidence interval is given by:

$$\left(\hat{p} - k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + k\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$$

where k is such that: $\Pr(-k < Z < k) = \frac{C}{100}$

e.g. a 99% confidence will be $\left(\hat{p} - 2.58\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}, \hat{p} + 2.58\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right)$

because $\Pr(-2.58 < Z < 2.58) \approx 0.99$

Example: Calculate the 90%, 95% and 99% confidence interval for the primary school children in Australia who use social media, with a random sample size of 200 and $\hat{p}$ of 0.7.

Solution:

95%: (0.636, 0.764)

90%: $\left(0.7 - 1.65\sqrt{\frac{0.7\times 0.3}{200}}, \quad 0.7 + 1.65\sqrt{\frac{0.7\times 0.3}{200}}\right) = (0.647, 0.753)$

99%: $\left(0.7 - 2.58\sqrt{\frac{0.7\times 0.3}{200}}, \quad 0.7 + 2.58\sqrt{\frac{0.7\times 0.3}{200}}\right) = (0.616, 0.783)$

So obviously the greater the confidence then wider the interval is required.

- **Ex17D** 1, 2, 3, 4, 6, 7, 9, 11, 12

Past Exam Questions

2008 Exam 1

2008 Exam 2

Question 11

The probability density function for the continuous random variable X is given by

$$f(x)=\begin{cases}|1-x| & \text{if } 0\le x\le 2\\ 0 & \text{elsewhere}\end{cases}$$

The probability that $X<1.5$ is equal to

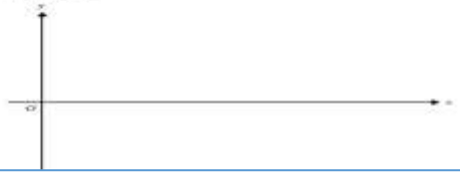
A. 0.125
B. 0.375
C. 0.5
D. 0.625
E. 0.75

Question 1

The time in hours that Sharelle spends training each day is a continuous random variable with probability density function given by

$$f(x)=\begin{cases}\frac{1}{64}(6-x)(x-2)(x+2) & \text{if } 2\le x\le 6\\ 0 & \text{elsewhere}\end{cases}$$

a. i. Sketch the probability density function, and label the local maximum with its coordinates, correct to two decimal places.



ii. What is the probability, correct to four decimal places, that Sharelle spends less than 3 hours training on a particular day?

iii. What is the mean time (in hours), correct to four decimal places, that she spends training each day?

2 + 2 + 2 = 6 marks

2009 Exam 1

2009 Exam 2

Question 6

The continuous random variable X has a normal distribution with mean 14 and standard deviation 2.

If the random variable Z has the standard normal distribution, then the probability that X is greater than 17 is equal to

A. $\Pr(Z>3)$
B. $\Pr(Z<2)$
C. $\Pr(Z<1.5)$
D. $\Pr(Z<-1.5)$
E. $\Pr(Z>2)$

Question 11

The continuous random variable X has a probability density function given by

$$f(x)=\begin{cases}\pi\sin(2\pi x) & \text{if } 0\le x\le \frac{1}{2}\\ 0 & \text{elsewhere}\end{cases}$$

The value of a such that $\Pr(X>a)=0.2$ is closest to

A. 0.26
B. 0.30
C. 0.32
D. 0.35
E. 0.40

Question 3

The Bouncy Ball Company (BBC) makes tennis balls whose diameters are normally distributed with mean 67 mm and standard deviation 1 mm. The tennis balls are packed and sold in cylindrical cans that each hold four balls. A tennis ball fits into such a can if the diameter of the ball is less than 68.5 mm.

a. What is the probability, correct to four decimal places, that a randomly selected tennis ball produced by BBC fits into a can?

2 marks

BBC management would like each ball produced to have diameter between 65.6 and 68.4 mm.

b. What is the probability, correct to four decimal places, that the diameter of a randomly selected tennis ball made by BBC is in this range?

2 marks

c. i. What is the probability, correct to four decimal places, that the diameter of a tennis ball which fits into a can is between 65.6 and 68.4 mm?

ii. A can of four balls is selected at random. What is the probability, correct to four decimal places, that at least one of these balls has diameter outside the desired range of 65.6 to 68.4 mm?

2 + 2 = 4 marks

BBC management wants engineers to change the manufacturing process so that 99% of all balls produced have diameter between 65.6 and 68.4 mm. The mean is to stay at 67 mm but the standard deviation is to be changed.

d. What should the new standard deviation be (correct to two decimal places)?

BBC sells tennis balls directly to tennis clubs once a year. If a tennis club buys its balls from BBC one year, there is an 80% chance it will buy its balls from BBC the next year. If a tennis club does not buy its balls from BBC one year, there is a 15% chance it will buy its balls from BBC the next year.

Suppose the Melbourne Tennis Club buys its tennis balls from BBC this year.

e. What is the exact probability that it will buy its tennis balls from BBC for the next three years?

2 marks

f. What is the exact probability that it will buy its tennis balls from BBC for exactly two of the next three years?

2 marks

Let p_n be the probability that the Melbourne Tennis Club will buy its tennis balls from BBC n years after 2009, given that it buys them from BBC in 2009.

g. Find the smallest value of n such that $p_n <$ [illegible]

2010 Exam 1

Question 5

Let X be a normally distributed random variable with mean 5 and variance 9 and let Z be the random variable with the standard normal distribution.

a. Find $\Pr(X>5)$.

1 mark

b. Find b such that $\Pr(X>7)=\Pr(Z<b)$.

2 marks

Question 7

The continuous random variable X has a distribution with probability density function given by

$$f(x)=\begin{cases}ax(5-x) & \text{if } 0\le x\le 5\\ 0 & \text{if } x<0 \text{ or if } x>5\end{cases}$$

where a is a positive constant.

a. Find the value of a.

3 marks

b. Express $\Pr(X<3)$ as a definite integral. (Do not evaluate the definite integral.)

1 mark

2010 Exam 2

Question 11

The continuous random variable X has a probability density function given by

$$f(x)=\begin{cases}\cos(2x) & \text{if } \frac{3\pi}{4}<x<\frac{5\pi}{4}\\ 0 & \text{elsewhere}\end{cases}$$

The value of a such that $\Pr(X<a)=0.25$ is closest to

A. 2.25
B. 2.75
C. 2.88
D. 3.06
E. 3.41

Question 13

The continuous random variable X has a normal distribution with mean 20 and standard deviation 6. The continuous random variable Z has the standard normal distribution.

The probability that Z is between -2 and 1 is equal to

A. $\Pr(18<X<21)$
B. $\Pr(14<X<32)$
C. $\Pr(14<X<26)$
D. $\Pr(8<X<32)$
E. $\Pr(X>14)+\Pr(X<26)$

2011 Exam 1

Question 5

The probability distribution function for the continuous random variable X is given by

$$f(x) = \begin{cases} |3 - x| & \text{if} \quad 2 \le x \le 4 \\ 0 & \text{elsewhere} \end{cases}$$

a. Find $\Pr(X < 3.5)$.

2 marks

b. Find $\Pr(X < 2.5 \mid X < 3.5)$.

2 marks

2011 Exam 2

Question 6

For the continuous random variable X with probability density function

$$f(x) = \begin{cases} \log_e(x) & 1 \le x \le e \\ 0 & \text{elsewhere} \end{cases}$$

the expected value of X, $E(X)$, is closest to

A. 0.358
B. 0.5
C. 1
D. 1.859
E. 2.097

Question 12

The continuous random variable X has a normal distribution with mean 30 and standard deviation 5. For a given number a, $\Pr(X > a) = 0.20$.

Correct to two decimal places, a is equal to

A. 23.59
B. 24.00
C. 25.79
D. 34.21
E. 36.41

Question 13

In an orchard of 2000 apple trees it is found that 1735 trees have a height greater than 2.8 metres. The heights are distributed normally with a mean μ and standard deviation 0.2 metres.

The value of μ is closest to

A. 3.023
B. 2.577
C. 2.230
D. 1.115
E. 0.223

Question 2

In a chocolate factory the material for making each chocolate is sent to one of two machines, machine A or machine B.

The time, X seconds, taken to produce a chocolate by machine A, is normally distributed with mean 3 and standard deviation 0.6.

The time, Y seconds, taken to produce a chocolate by machine B, has the following probability density function.

$$f(y) = \begin{cases} 0 & y < 0 \\ \dfrac{y}{16} & 0 \le y \le 4 \\ 0.25e^{-0.5(y-4)} & y > 4 \end{cases}$$

a. Find correct to four decimal places

i. $\Pr(3 \le X \le 5)$

ii. $\Pr(3 \le Y \le 5)$

1 + 3 = 4 marks

b. Find the mean of Y, correct to three decimal places.

3 marks

c. i. Find the median of Y.

ii. Find the value of a, correct to two decimal places, such that $\Pr(Y \le a) = 0.7$.

1 + 2 = 3 marks

d. It can be shown that $\Pr(Y \le 3) = \frac{9}{32}$. A random sample of 10 chocolates produced by machine B is chosen.

Find the probability, correct to four decimal places, that exactly 4 of these 10 chocolates took 3 or less seconds to produce.

3 marks

All of the chocolates produced by machine A and machine B are stored in a large bin. There is an equal number of chocolates from each machine in the bin.

It is found that if a chocolate, produced by either machine, takes longer than 3 seconds to produce then it can easily be identified by its darker colour.

e. A chocolate is selected at random from the bin. It is found to have taken longer than 3 seconds to produce. Find, correct to four decimal places, the probability that it was produced by machine A.

2012 Exam 1

Question 8

a. The random variable X is normally distributed with mean 100 and standard deviation 4.

If $\Pr(X < 106) = q$, find $\Pr(94 < X < 100)$ in terms of q.

2 marks

b. The probability density function f of a random variable X is given by

$$f(x) = \begin{cases} \dfrac{x+1}{12} & 0 \le x \le 4 \\ 0 & \text{otherwise} \end{cases}$$

Find the value of b such that $\Pr(X \le b) = \frac{5}{8}$.

3 marks

2012 Exam 2

Question 11

The weights of bags of flour are normally distributed with mean 252 g and standard deviation 12 g. The manufacturer says that 40% of bags weigh more than x g.

The maximum possible value of x is closest to

A. 249.0
B. 251.5
C. 253.5
D. 254.5
E. 255.0

2013 Exam 1

Question 9 (3 marks)

A continuous random variable, X, has a probability density function

$$f(x) = \begin{cases} \dfrac{\pi}{4}\cos\left(\dfrac{\pi x}{4}\right) & \text{if } x \in [0, 2] \\ 0 & \text{otherwise} \end{cases}$$

Given that $\dfrac{d}{dx}\left(x\sin\left(\dfrac{\pi x}{4}\right)\right) = \dfrac{\pi x}{4}\cos\left(\dfrac{\pi x}{4}\right) + \sin\left(\dfrac{\pi x}{4}\right)$, find $E(X)$.

2013 Exam 2

Question 22

Butterflies of a particular species die T days after hatching, where T is a normally distributed random variable with a mean of 120 days and a standard deviation of σ days.

If, from a population of 2000 newly hatched butterflies, 150 are expected to die in the first 90 days, then the value of σ is closest to

A. 7 days
B. 13 days
C. 17 days
D. 21 days
E. 37 days

c. When FullyFit surveyed all its gyms throughout the world, it was found that the time taken by members to complete S is a continuous random variable X, with a probability density function g, as defined below.

$$g(x) = \begin{cases} \frac{(x-3)^3+64}{256} & 1 \le x \le 5 \\ \frac{x+28}{120} & 5 < x \le 8 \\ 0 & \text{elsewhere} \end{cases}$$

i. Find $E(X)$, correct to four decimal places. 2 marks

ii. In a random sample of 200 FullyFit members, how many members would be expected to take more than four minutes to complete S? Give your answer to the nearest integer. 2 marks

2014 Exam 1

Question 8 (4 marks)

A continuous random variable, X, has a probability density function given by

$$f(x) = \begin{cases} \frac{1}{5}e^{-\frac{x}{5}} & x \ge 0 \\ 0 & x < 0 \end{cases}$$

The median of X is m.

a. Determine the value of m. 2 marks

b. The value of m is a number greater than 1.

Find $\Pr(X < 1 | X \le m)$. 2 marks

2014 Exam 2

Question 5

The random variable X has a normal distribution with mean 12 and standard deviation 0.5.

If Z has the standard normal distribution, then the probability that X is less than 11.5 is equal to

A. $\Pr(Z > -1)$
B. $\Pr(Z < -0.5)$
C. $\Pr(Z > 1)$
D. $\Pr(Z \ge 0.5)$
E. $\Pr(Z < 1)$

Question 16

The continuous random variable X, with probability density function $p(x)$, has mean 2 and variance 5.

The value of $\int_{-\infty}^{\infty} x^2 p(x)dx$ is

A. 1
B. 7
C. 9
D. 21
E. 29

Question 4 (14 marks)

Patricia is a gardener and she owns a garden nursery. She grows and sells basil plants and coriander plants.

The heights, in centimetres, of the basil plants that Patricia is selling are distributed normally with a mean of 14 cm and a standard deviation of 4 cm. There are 2000 basil plants in the nursery.

a. Patricia classifies the tallest 10 per cent of her basil plants as **super**.

What is the minimum height of a super basil plant, correct to the nearest millimetre? 1 mark

Patricia decides that some of her basil plants are not growing quickly enough, so she plans to move them to a special greenhouse. She will move the basil plants that are less than 9 cm in height.

b. How many basil plants will Patricia move to the greenhouse, correct to the nearest whole number? 2 marks

The heights of the coriander plants, x centimetres, follow the probability density function $h(x)$, where

$$h(x) = \begin{cases} \frac{\pi}{100}\sin\left(\frac{\pi x}{50}\right) & 0 < x < 50 \\ 0 & \text{otherwise} \end{cases}$$

c. State the mean height of the coriander plants. 1 mark

Patricia thinks that the smallest 15 per cent of her coriander plants should be given a new type of plant food.

d. Find the maximum height, correct to the nearest millimetre, of a coriander plant if it is to be given the new type of plant food. 2 marks

Patricia also grows and sells tomato plants that she classifies as either **tall** or **regular**. She finds that 20 per cent of her tomato plants are tall.

A customer, Jack, selects n tomato plants at random.

e. Let q be the probability that at least one of Jack's n tomato plants is tall.

Find the minimum value of n so that q is greater than 0.95. 2 marks

In another section of the nursery, a craftsman makes plant pots. The pots are classified as **smooth** or **rough**.

The craftsman finishes each pot before starting on the next. Over a period of time, it is found that if one plant pot is smooth, the probability that the next one is smooth is 0.7, while if one plant pot is rough, the probability that the next one is rough is p, where $0 < p < 1$. The value of p stays fixed for a week at a time, but can vary from week to week. The first pot made each week is always a smooth pot.

f. i. Find, in terms of p, the probability that the **third** pot made in a given week is smooth. 2 marks

ii. In one particular week, the probability that the **third** pot made is smooth is 0.61.

Calculate the value of p in this week. 2 marks

2015 Exam 1

Question 6 (3 marks)

Let the random variable X be normally distributed with mean 2.5 and standard deviation 0.3.

Let Z be the standard normal random variable, such that $Z \sim N(0, 1)$.

a. Find b such that $\Pr(X > 3.1) = \Pr(Z < b)$. 1 mark

b. Using the fact that, correct to two decimal places, $\Pr(Z < -1) = 0.16$, find $\Pr(X < 2.8 \mid X > 2.5)$.
Write the answer correct to two decimal places. 2 marks

Question 13

The function f is a probability density function with rule

$$f(x) = \begin{cases} ae^x & 0 \le x \le 1 \\ ae & 1 < x \le 2 \\ 0 & \text{otherwise} \end{cases}$$

The value of a is

A. 1

B. e

C. $\frac{1}{e}$

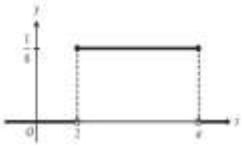
D. $\frac{1}{2e}$

E. $\frac{1}{2e-1}$

2015 Exam 2

Question 9

The graph of the probability density function of a continuous random variable, X, is shown below.



If $a > 2$, then $E(X)$ is equal to

A. 8

B. 5

C. 4

D. 3

E. 2

Question 3 (11 marks)

Mani is a fruit grower. After his oranges have been picked, they are sorted by a machine, according to size. Oranges classified as **medium** are sold to fruit shops and the remainder are made into orange juice.

The distribution of the diameter, in centimetres, of medium oranges is modelled by a continuous random variable, X, with probability density function

$$f(x) = \begin{cases} \frac{3}{4}(x-6)^2(8-x) & 6 \le x \le 8 \\ 0 & \text{otherwise} \end{cases}$$

a. i. Find the probability that a randomly selected medium orange has a diameter greater than 7 cm. 2 marks

ii. Mani randomly selects three medium oranges.

Find the probability that exactly one of the oranges has a diameter greater than 7 cm.

Express the answer in the form $\frac{a}{b}$, where a and b are positive integers. 2 marks

b. Find the mean diameter of medium oranges, in centimetres. 1 mark

ii. Suppose that instead of selecting only four lemons, n lemons are selected at random from a particular load.

Find the smallest integer value of n such that the probability of at least one lemon being underweight exceeds 0.5. 2 marks