Population Proportion

In statistics you usually wish to study a certain collection of individuals or items. This collection is known as the **population**

It would be useful to have information about every individual or item in the population. The gathering of this information is called a **census**

e.g. There are one million people in a particular city. We are interested in the proportion of people that would choose a prime number if they were asked to pick a number between 1 and 100.

The probability that a prime number was chosen is a **population proportion**

$$p = \frac{25}{100} = 0.25$$

Note: When a census of the population was held 249925 picked a prime number

As you cannot predict a response, the result of a census may not agree with the population proportion

Sample Proportion

It is often unrealistic to conduct a census, as it might;

- take to long to conduct the census
- cost too much to employ the people to do such a large census
- analysis of the results would also be time consuming

Therefore a more realistic aim would be to take a subset of the population, known as a **sample**, and gather the information via a **survey**

Ideally a sample should have all of the characteristics of the population

e.g. I surveyed 100 people from our city and asked them to choose number between 1 and 100 and 29 picked a prime number

The **sample proportion** would be $\hat{p} = \frac{29}{100} = 0.29$

If $X \sim \mathrm{Bin}(n,p)$

The **sample proportion** is the random variable

$$\boxed{\hat{p} = \frac{X}{n}}$$

$$P\left(\hat{p} = \frac{x}{n}\right) = P(X = x) = \binom{n}{x}(1-p)^{n-x}p^{x}$$

$$E\left(\hat{p}\right) = p \qquad Var\left(\hat{p}\right) = \frac{p(1-p)}{n}$$

$\hat{p}$ gives an estimate of the probability p

e.g. Prior to conducting our survey of 100 people, we wanted to see if 100 people would be sufficient to draw conclusions about the population.

So we ran 50 simulations using a spreadsheet and the number of primes produced in the simulations are listed in the following table

0.3	**0.25**	**0.24**	**0.29**	**0.31**	**0.27**	**0.28**	**0.26**	**0.2**	**0.3**
0.24	**0.3**	**0.26**	**0.27**	**0.21**	**0.32**	**0.17**	**0.22**	**0.26**	**0.32**
0.28	**0.3**	**0.27**	**0.29**	**0.23**	**0.21**	**0.28**	**0.18**	**0.22**	**0.19**
0.26	**0.26**	**0.31**	**0.22**	**0.26**	**0.3**	**0.22**	**0.23**	**0.23**	**0.26**
0.27	**0.35**	**0.27**	**0.27**	**0.22**	**0.22**	**0.21**	**0.2**	**0.24**	**0.3**

Let $\widehat{p}$ = proportion of people choosing a prime number

There are 25 prime numbers less than 100
$\therefore p = 0.25$

$$E\left(\hat{p}\right) = 0.25$$

$$Var\left(\hat{p}\right) = \frac{(0.25)(0.75)}{50}$$
$$= 0.00375$$

Let's say we would be happy if our survey produced an estimate that is within 5% of the actual value

$$p = 0.25 \quad \Rightarrow \; 0.2375 \le \hat{p} \le 0.2625$$

$$\hat{p} = \frac{x}{100} \quad \Rightarrow \; 23.75 \le x \le 26.25$$

$P(23 \le X \le 27)$, $X \sim \text{Bin}(100, 0.25)$

$$= \binom{100}{23}(0.75)^{77}(0.25)^{23} + \binom{100}{24}(0.75)^{76}(0.25)^{24} + \binom{100}{25}(0.75)^{75}(0.25)^{25}$$
$$+ \binom{100}{26}(0.75)^{74}(0.25)^{26} + \binom{100}{27}(0.75)^{73}(0.25)^{27}$$

$= 0.43601$

There is a 44% chance that our survey will produce an estimate within 5% of the actual results

Using a normal approximation

$$p = 0.25 \quad \Rightarrow \quad 0.2375 \le \hat{p} \le 0.2625$$

$$\begin{aligned} s &= \sqrt{0.00375} \\ &= 0.0612 \end{aligned}$$

$$\begin{aligned} P\left(|Z| \le \frac{0.2625 - 0.25}{0.0612} \right) &= 2\,\Phi(0.2042) - 1 \\ &= 0.1618 \end{aligned}$$

There is a 16% chance that our survey will produce an estimate within 5% of the actual results

From this we could conclude that a sample size of 100 would be not be sufficient to model the entire population e.g. A recent census showed that 20% of the adults in a city eat out regularly

a) A survey of 100 adults in this city is to be conducted to find the proportion who eat out regularly. Show that the mean and standard deviation for the distribution of sample proportions of such surveys are 0.2 and 0.04 respectively

$$\begin{aligned}\bar{x} &= p \\ &= 0.2\end{aligned}$$

$$\begin{aligned}s &= \sqrt{\frac{p(1-p)}{n}} \\ &= \sqrt{\frac{(0.2)(0.8)}{100}} \\ &= 0.04\end{aligned}$$

b) Use the extract shown from a table giving values of $P(Z < z)$, where z has a standard normal distribution, to estimate the probability that a survey of 100 adults will find that at most 15 of those surveyed eat out regularly

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
1.1	0.8643	0.8665	0.8686	0.8708	0.8729	0.8749	0.8770	0.8790	0.8810	0.8830
1.2	0.8849	0.8869	0.8888	0.8907	0.8925	0.8944	0.8962	0.8980	0.8997	0.9015
1.3	0.9032	0.9049	0.9066	0.9082	0.9099	0.9115	0.9131	0.9147	0.9162	0.9177

$$P\left(Z < \frac{0.15 - 0.2}{0.04} \right) = \Phi(-1.25)$$

$$= 1 - \Phi(1.25)$$

$$= \underline{0.1056}$$

(c) An online retailer claims that 90% of all orders are shipped within 12 hours of being received. On a particular day, 121 orders were received and 102 orders were shipped within 12 hours.

(i) State the sample proportion of orders shipped within 12 hours.

$$\hat{p} = \frac{102}{121}$$

$$= 0.84 \quad \text{(to 2 dp)}$$

The distribution of the sample proportion of all orders that are shipped within 12 hours of being received on any day is approximately normal.

(ii) Assuming the online retailer's claim is true, find the probability that, in a random sample of 121, less than 85% of all orders are shipped within 12 hours.

$$E\left(\hat{p}\right) = 0.9 \qquad \mathrm{Var}\left(\hat{p}\right) = \frac{0.9 \times 0.1}{121}$$

$$= 7.438 \times 10^{-4}$$

$$\hat{p} \sim N(0.9, 7.438 \times 10^{-4})$$

$$P\left(\hat{p} < 0.85\right) = P\left(Z < \frac{0.85 - 0.9}{\sqrt{7.438 \times 10^{-4}}}\right)$$

$$= P(Z < -1.83)$$

$$= 1 - \Phi(1.83)$$

$$= 0.0336$$

(iii) Use the result from (ii) to evaluate the reasonableness of the online retailer's claim.

part (ii) tells me that there is a 3.4% chance of shipping less than 85% of orders within 12 hours

In part (i) our sample proportion was 84%, so this would have occurred by chance only 3.4% 0f the time

Therefore our sampling would suggest that the retailer's claim is dubious

2021 Extension 1 HSC Question 14 d)

At a certain factory, the proportion of faulty items produced by a machine is $p = \frac{3}{500}$, which is considered to be acceptable. To confirm that the machine is working to this standard, a sample of size n is taken and the sample proportion $\hat{p}$ is calculated.

It is assumed that $\hat{p}$ is approximately normally distributed with $\mu = p$ and $\sigma^2 = \frac{p(1-p)}{n}$

Production by this machine will be shut down if $\hat{p} \geq \frac{4}{500}$

The sample size is to be chosen so that the chance of shutting down the machine unnecessarily is less than 2.5%.

Find the approximate sample size required, giving your answer to the nearest thousand.

$$\hat{p} \sim N\left(0.006, \frac{(0.006)(0.994)}{n}\right) \Leftrightarrow Z \sim N(0,1) \text{ where } Z = \frac{\hat{p} - \mu}{\sigma}$$

$$P\left(\hat{p} \geq 0.008\right) < 0.025$$

$$P\left(Z \geq \frac{(0.008 - 0.006)\sqrt{n}}{\sqrt{(0.006)(0.994)}} \right) < 0.025$$

$$\frac{(0.008 - 0.006)\sqrt{n}}{\sqrt{(0.006)(0.994)}} > 2 \qquad \text{(emperical rule)}$$

$$\sqrt{n} > \frac{2\sqrt{(0.006)(0.994)}}{0.002}$$

$$n > \frac{4(0.006)(0.994)}{(0.002)^2}$$

$$n > 5964$$

A sample size of approximately 6000 will be required

Exercise 17D; 1 to 6, 8, 10, 12, 13, 15, 16, 18