 Teacher notes – sample proportion

**Note:** This resource is to be used with the Microsoft Excel resource sample-proportions.XLSM

Sample introduction to the task for students

Suppose we had population data of the heights of individuals. From this data a sample of 20 people is chosen and the proportion of people whose height is under 150cm is recorded. The value of this sample proportion could be anywhere from $\frac{0}{20}$ to $\frac{20}{20}$.

Imagine we repeatedly draw samples and recorded the values for this sample proportion each time.

The aim of this task is to explore how these values are distributed.

Teacher notes

If $n=sample size$ and $p=population$ , when $np>10$ and $n(1-p)>10$, the distribution of the sample proportion values will be approximately normal. This is independent of the shape of the original population. When these inequalities do not hold, the distribution will be skewed.

When the distribution is approximately normal, the mean of the sample proportion will approximate the actual proportion in the population.

The standard deviation of the sample proportion can be approximated by the formula: $\sqrt{\frac{p(1-p)}{n}}$

Using sample-proportions.XLSM

In the spreadsheet, the user can:

1. Set a mean and standard deviation for the population. This is normally distributed. Alternatively a user's population can be pasted into column N. It is not necessary for the population to be normally distributed for this investigation.
2. Select a range of values. When a sample is selected, we will consider the proportion in the sample which meets this criteria.
3. Select a sample size (10, 20, 50 or 100)

After setting the previous parameters the user can:

1. Add a histogram .
2. Select a sample. This will generate a sample from the population data which will be displayed in the table in columns H and I.
3. Update the frequency table. This will update the frequency table in columns K and L with the latest sample. This will also update the histogram.

**Note:** Steps 2 and 3 can be repeated to increase the number of samples. The teacher should consider modelling this first so that students have the opportunity to observe the changing values in the tables and the changing display of the histogram. After repeating steps 2 and 3, you may wish to click ‘Run 20 samples’ to quickly increase the quantity of data. This process can also be repeated several times.

1. Clear the simulation data and delete the histogram before setting up a new scenario. You can use the green buttons to do this.

Considerations when using the resource:

* What values can $\hat{p}$ (the sample proportion) take in this specific experiment?
* Are the values of $\hat{p}$ different between samples? If so, why?
* What is the relationship between$ p$, the proportion in the population, and $\hat{p}$,
	+ Consider $p$ (C7) and the mean of the sample proportion (C29)
	+ Consider the approximate standard deviation of the sample proportion (F8) and the standard deviation of the experimental sample proportion (C30)
* Consider the shape of the distribution (see the histogram) of sample proportions where:
	+ The sample size is large
	+ The sample size is small
	+ Where $np>10$ and $n(1-p)>10$ compared to when they are less than 10.
* What are the limitations of the sample proportion that may hinder its use in solving problems ?

Refer to sample size and the value of $np$ and $n(1-p)$.

Why is this method important?

If a binomial distribution can be approximated by a normal distribution, solutions to complex problems can be approximated using what we know about a normal distribution.

Consider the following scenario: Suppose 54% of students are shorter than 150cm. If a random sample of 100 students was chosen, what is the probability that, at most, 49% of the students (**i.e.** 49 students or less) will be shorter than 150cm?

If this was to be completed using binomial probability we would need to consider $P(0 students) + P(1 student) + …+ P(49 students)$, which would be time consuming. Let’s consider this in terms of the sample proportion.

Firstly, consider if the distribution of the sample proportion would be normally distributed:

$$n=100, p=0.54$$

$$np=54$$

$$n\left(1-p\right)=46$$

**i.e.** The distribution of the sample proportions can be approximated using the normal distribution with:

$$μ\_{\hat{p}}=p=0.54$$

$$σ\_{\hat{p}}=\sqrt{\frac{p(1-p)}{n}}=\sqrt{\frac{0.54(1-0.54)}{100}}=0.0498…≈0.05$$



How can this be interpreted? Consider the standard normal results. From this information, the probability of the sample proportion being between 0.49 and 0.59 would be 68%. This means if we took a sample of 100 students, 68% of the time between 49 and 59 students would be shorter than 150cm. That is, if we took a sample of 100 students, the probability that 49 to 59 students would be shorter than 150cm is 68%.

In our example, we want to consider the probability that the sample proportion will be less than 0.49, which is 1 standard deviation below the mean.

i.e. $z\_{0.49} =\frac{0.49-0.54}{0.05}=-1$

This is the section of the curve less than 0.49

$P\left(<49\right)=0.5-0.34=0.16$ or 16%.

The probability of at most 49 students being shorter than 150cm is approximately 16%.