

An investigative agency is currently conducting a health survey for its clients. The focus of the investigation is to understand how the heights of young children are distributed in the small city of Galactica.

At the moment, the height of young children are thought to be normally distributed with a mean height of 132 cm with a standard deviation of 12.32 cm.

- (a) State the distribution and the parameter/s of the sample mean for samples of size 20. (2)

Another trainee states that the mode of the normal distribution is the same as the mean, he asserts that mode of a sample of heights will also be normally distributed with the same parameters as (a)

- (b) Perform a simulation of a sample $x_1, x_2, x_3, \dots, x_n$ of the heights with sample size $n = 20$ and a seed value of 1947. (2)

- (c) Calculate the empirical mode for the sample in (b) (2)

[Hint : `density()` outputs x and y , where y is the empirical pdf for each corresponding value of x]

- (d) Perform 10000 repetitions of parts (b) and (c) to obtain a bootstrapped sample of the mode using the same set seed as before. (9)

- (e) Plot a histogram showing the densities of the sample modes from part (d), with a y -axis that goes up to 0.15. Superimpose the density of the distribution discussed in part (a) on the histogram. (6)

- (f) Compare the distribution of the sample mode with that of the distribution of the sample mean given by the Central Limit Theorem, using the graph in part (e). (2)

Despite the differences observed in (e), the trainee still believes that empirical mode should be normally distributed.

- (g) Create a Q - Q Plot of the sample modes from part (d) and by adding a line to the Q - Q plot to show the expected result if the modes were normally distributed, comment whether there is any proof of trainee's assertion. (7)