

The Central Limit Theorem for proportions

There is a companion version of the CLT for the sampling distribution of “proportion” type statistics that arise from categorical variables. Here is what it says:

When the conditions are met, the sampling distribution of sample proportions follows the normal distribution $N(p, \sqrt{\frac{p(1-p)}{n}})$, where n =sample size, p =true value of the proportion in the population of interest.

The conditions that must be met:

1. The sample size must be large enough.
2. The sample must be made up of independent observations (or cases).

A warmup

To help us understand inferential stats at a fundamental, conceptual level, let us first work backwards. That is, we will assume the true value of the parameter of interest is known, and we will look at the probability of getting the statistic we found in our sample.

1. In order to gauge public opinion about a referendum on the ballot in the upcoming elections in Indiana, a news organization conducted telephone interviews with 602 voters registered in the state. They found 74% of those interviewed supported the referendum.

Let p denote the true proportion of supporters of the referendum. For each of the following possible values of p estimate the probability of getting our sampled value, $\hat{p} = 0.74$.

(a) $p = 0.7$.

(b) $p = 0.75$.

(c) $p = 0.78$.

2. A children's research organization studied the television viewing habits of 822 children between the ages of 8 and 14 years. They found the average time these children spent in front of a television per year was about 1054 hours.

Let μ, σ respectively denote the true mean and standard deviation of children's TV watching time. For each of the following scenarios, estimate the probability of getting our sampled value, $\hat{\mu} = 1054$ hours.

(a) $\mu = 1020, \sigma = 400$ hours.

(b) $\mu = 1040, \sigma = 400$ hours.

(c) $\mu = 1080, \sigma = 400$ hours.

Is there a moral to these stories?

Conjecture some thoughts!