ST2132 Tutorial 5 2210

1. Let $X_1, \ldots, X_n$ be IID RV's with expectation μ and variance σ^2.

   (a) The Central Limit Theorem say that as $n \to \infty$, the distribution of

   $$\frac{X_1 + \cdots + X_n - n\mu}{\sqrt{n}\sigma}$$

   converges to N(0,1). Apply it to derive the first probability statement on slide 6 of Survey Sampling II.

   (b) Deduce the probability statement on slide 7 of Suvey Sampling II.

   (c) Rewrite the statement in (b), given that the distribution of X_i is Bernoulli(p).

2. A market survey organisation took a simple random sample of 500 households in a town with 25,000 households; 239 of the sample households had internet access.

   (a) Describe the population parameter that can be estimated from the data.

   (b) What is the exact SE of your estimate of the parameter? Find a 95%-CI for the parameter. Round your parameter and SE estimates to two decimal places.

   (c) The CI depends on a normal approximation to a binomial distribution. State its number of trials and success probability.

3. 5. Laser altimeters can measure elevation to within a few inches, without bias, and with no trend or pattern to the measurements. As part of an experiment, 25 readings were made on the elevation of a mountain peak. These averaged out to 81,411 inches, and their sample SD (with denominator 24) was 30 inches. For b–d, state if it is true or false, and explain your answers briefly.

   (a) A data analyst writes: "Assume the 25 readings are realisations of IID RV's $X_1, \ldots, X_{25}$ with expectation 81,411 inches and SD 30 inches." Is this correct? If not, amend the statement.

   (b) $81,411 \pm 12$ inches is a 95%-confidence interval for the elevation of the mountain peak.

   (c) $81,411 \pm 12$ inches is a 95%-confidence interval for the average of the 25 readings.

   (d) There is about a 95% chance that the elevation of the mount peak is in the range $81,411 \pm 12$ inches.

   (e) What is the significance of the phrase "no trend or pattern"?

4. A statistician issues a challenge to 10,000 friends. Upon receiving 100 realisations from $\mathrm{N}(\mu, \sigma^2)$, a friend will guess the value of μ. Different friends get different data. The friends reach an agreement to return an interval estimate in the form $(\bar{x} - \mathrm{SE}, \bar{x} + \mathrm{SE})$.

   (a) Can you predict the fraction of intervals containing μ, to two decimal places?

   (b) Your task is to simulate this game with R. Write a code to generate 100 realisations from N(0,1) and compute the interval estimate (for what parameter?). Repeat the process to get 10,000 intervals. Then calculate the fraction of intervals containing the parameter.

   [Hint: Instead of a for loop, you can store all the realisations in a $10,000 \times 100$ matrix, then use apply() to construct the intervals in one go.]