Quiz Week 6 - CLT and confidence intervals - Questions

FIT5197 teaching team

Question 1

The relationship between the central limit theorem and confidence intervals for the true mean of the sample mean distribution. (These means of means are making me confused, I don't know what it means.

(a) Let $X_1, \ldots, X_n$ be i.i.d. Random Variables (RVs) with $\mathbb{E}\left[X_i\right] = \mu, \mathbb{V}\left[X_i\right] = \sigma^2$. From the sample mean version of the Central Limit Theorem (CLT) we know that as $n \rightarrow \infty$

$$\overline{X} \xrightarrow{d} \mathcal{N}\left(\mu, \sigma^2/n\right)$$

where $\bar{X}$ is the RV for the sample mean. Also note that when dealing with normal distributions (as above) we transform the normal distribution we are working with into the standard normal distribution in order to compute probabilities. I.e. in this case we define

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \sim N(0, 1), z \in \mathbb{R}$$

Assuming the situation above with large n so the CLT can be applied, show why the two-sided confidence interval for the true population parameter for the sample mean, μ, provides bounds on this parameter that guarantees this parameter falls within these bounds with probability $1 - \alpha$. Moreover, specifically show that when we draw a specific sample $\mathbf{x} = \mathbf{x_1, \ldots, x_n}$ from the RVs X_i and obtain an observed sample mean $\bar{x}$, that the two-sided confidence interval can be expressed as:

$$\mathrm{CI}_\mu(1 - \alpha) = \left(\bar{x} - \frac{z_{\alpha/2}}{\sqrt{n}} \cdot \sigma, \bar{x} + \frac{z_{\alpha/2}}{\sqrt{n}} \cdot \sigma\right).$$

Hint: recall that

$$\mathbb{P}\left(-z_{\alpha/2} \leqslant Z \leqslant z_{\alpha/2}\right) \approx 1 - \alpha.$$

where $z_{\alpha/2}$ is the value of z that satisfies this probability formula.

(b) Practical example of the above: Consider the following sample of weights of individuals,

$$\mathbf{w} = \{61, 55, 57, 70, 59, 65, 66, 58\}$$

and given that the variance of weight for an individual is known to be $\sigma^2 = 16$. Find the two-sided 95% confidence interval for the mean of the sample mean distribution for weight. Hint: You will probably need the z-table in the [Formula Sheet (https://lms.monash.edu/mod/resource/view.php?id=7439150)](https://lms.monash.edu/mod/resource/view.php?id=7439150) for the unit.

(c) Why for the love of the universe do we have to calculate confidence intervals for the mean of a population?

Question 2

A small micro-loan bank has 500 loan customers. If the total annual loan repayments made by an individual is a random variable with mean $750 and standard deviation $900, approximate the probability that the average total annual repayments made across all customers is greater than $755.

Question 3

There are 10000 students participate in an exam and the exam score approximately follow a normal distribution. Given that 359 students get scores above 90 in the exam and 1151 students get lower than 60. If there are 2500 students pass the exam, find out the lowest score to pass.

Question 4

The light bulb in Monash university has an average lifetime of 1000 hours with a standard deviation of 50 hours. How many of these light bulbs should Monash stock up so that it can guarrantee that the light will be on for at least 7200 hours with a probability of at least 98%?

Question 5

During the European football championships in 2008, and the football World Cup in 2010, **an octopus called Paul** living at an aquarium in Oberhausen, Germany, was used to predict the outcome of football matches, mostly involving the German national football team. To obtain Pauls' predictions, his keepers at the aquarium would present him with two boxes of food before each match. Each box was covered in the flag of the two nations that were participating, and the box that Paul chose to feed from first determined which nation he predicted would win.

Paul was asked to predict the outcome of $\mathbf{14}$ matches, $\mathbf{12}$ of which involved Germany. He correctly predicted the outcomes of $\mathbf{12}$ matches, only incorrectly guessing that Germany would beat Croatia in the Euro 2008 group stage, and that Germany would beat Spain in the Euro 2008 final. Some people claimed he was an **"animal oracle"**:

Calculate an estimate of Paul's success rate at predicting football matches. Calculate a $95\ \%$ confidence interval for this estimate, and summarise/describe your results appropriately. Show all workings as required.

R code hackers nail-biting challenge

Consider the standard normal distribution characterised by random variable:

$$Z \sim N(0,1), z \in \mathbb{R}$$

Using R, through simulation with $n = 200$ samples obtained with the R function rnorm and the calculation of normalised histograms show that

$$\mathbb{P}\left(-z_{\alpha/2} \leqslant Z \leqslant z_{\alpha/2}\right) \approx 1 - \alpha.$$

appears to be correct for $\alpha = 0.1$, 0.05 and 0.025. Note you will have to obtain the value of $z_{\alpha/2}$ using the R function qnorm. How do your simulated values for $\mathbb{P}\left(-z_{\alpha/2} \leqslant Z \leqslant z_{\alpha/2}\right)$ compare to values for the same quantity computed using the R function pnorm. Which of these approaches gives a better estimate of $1 - \alpha$? Why?