DALITE Q3 - Parameters, Sampling Distributions and the Central Limit Theorem Solutions.

EPIB607 - Inferential Statistics[a]

[a]Fall 2018, McGill University

This version was compiled on September 21, 2018

This DALITE quiz will cover the building blocks of statistical inference.

Parameters and statistics | Sampling distributions | Central Limit Theorem (CLT)

1. Parameters Q1

Which of the following statements is false?

a) A parameter is a constant of unknown magnitude in a statistical model
b) The value of a parameter is in general unknown
c) A statistic is a number derived from a sample
d) **The population standard deviation can be estimated from the sample provided that we have a sample size greater than 30 (Correct)**

1.1. Correct rationales.

- No, you can have a standard deviation in a sample that has a 'n value' of smaller than 30 this would just change the estimated standard deviations accuracy.
- Population standard deviation can be estimated from a sample size regardless of the specific number

1.2. Incorrect rationales.

- Generally the sample standard deviation is estimated from the population standard deviation using the square root of the sample size.

2. Sampling Distributions Q1

A newborn baby has extermely low birth weight (ELBW) if it weighs less than 1000 grams. A study of the health of such children in later years examined a random sample of 219 children. Their mean weight at birth was $\bar{y} = 810$ grams. This sample mean $(\bar{y})$ is an unbiased estimator of the mean weight μ in the population of all ELBW babies. This means that

a) **In many samples from this population, the mean of many values of $\bar{y}$ will be equal to μ (Correct)**
b) As we take larger and larger samples from this population, the sample mean $\bar{y}$ will get closer and closer to μ
c) In many samples from this population, the many values of the sample mean $\bar{y}$ will have a distribution that is close to Normal

2.1. Correct rationales.

- As the sample mean $\bar{y}$ is an unbiased estimator of the mean weight μ in the population, the mean of many values of $\bar{y}$ will be equal to μ. Further, if we plot the sample means, the resulting sampling distribution of $\bar{y}$ will have the same mean value as the mean in the population distribution.
- Because the sample mean is an unbiased estimator of the mean weight. So the mean of many samples would be equal to μ.

2.2. Incorrect rationales.

- The larger the samples from the population, the smaller the standard deviations and the closer the mean values are to the average of the population.

3. CLT Q1

Cholesterol levels among fourteen-year-old boys are roughly Normal with mean 170 and standard deviation 30 milligrams per deciliter (mg/dl). You choose a simple random sample of 4 fourtheen-year-old boys and average their choleterol levels. If you do this many times, the mean of the average cholesterol levels you get will be close to

a) **170 (Correct)**
b) $170/4 = 42.5$
c) $170/\sqrt{4} = 85$

3.1. Correct rationales.

- The population has a normal distribution, which means that the sample mean of n independent observations will also have a normal distribution with a mean equal to mu.

3.2. Incorrect rationales.

4. CLT Q2

Cholesterol levels among fourteen-year-old boys are roughly Normal with mean 170 and standard deviation 30 milligrams per deciliter (mg/dl). You choose a simple random sample of 4 fourtheen-year-old boys and average their choleterol levels. If you do this many times, the standard deviation of the average cholesterol levels you get will be close to

a) 30
b) $4/\sqrt{30} = 0.73$
c) **$30/\sqrt{4} = 15$ (Correct)**

4.1. Correct rationales.

- By the CLT, the sample cholesterol levels will be normally distributed with a standard deviation equal to the population standard deviation divided by the square root of the sample size.

4.2. Incorrect rationales.

- This is standard deviation we are discussing, not standard error, standard error requires we divide by square root of N
- Multiple samples will approximate the population parameter
- Taking multiple sample, sample mean will be closer to the population mean and sample standard deviation will be closer to population standard deviation.

5. CLT Q3

The survival times of guinea pigs inoculated with an infections viral strain vary from animal to animal. The distribution of survival times is strongly skewed to the right. The central limit theorem says that

a) as we study more and more infected guinea pigs, their average survival time gets ever closer to the mean μ for all infected guinea pigs.
b) the average survival time of a large number of infected guinea pigs has a distribution of the same shape (strongly skewed) as the distribution for individual infected guinea pigs
c) **the average survival time of a large number of infected guinea pigs has a distribution that is close to Normal. (Correct)**

5.1. Correct rationales.

- CLT states that as a sample becomes large enough, the distribution takes on a normal distribution despite the the distribution shape of its population.

5.2. Incorrect rationales.