"Big Picture" Questions & Answers

1. What is the sampling distribution of the sample mean?

It is the probability distribution of the sample mean from a random sample of a particular size (n) taken from some population. It is best visualized by thinking of computing the sample mean for a sample of size n, and then repeating this step many times to produce many sample means, and now think of the density function which describes the relative frequencies of the various possible values for these sample means.

2. Why is it useful?

It is useful because it describes how far the sample mean might be from the population mean, and it is the population mean we are usually interested in.

3. What is a sampling distribution (more generally)?

In the description at 1. above, just replace "sample mean" by "statistic". Recall that a statistic, in general, is just a function of the all the sample values (in a random sample of size n). Usually we construct a statistic so that it will be an estimator of a parameter.

4. What are parameters?

Parameters are characteristics (usually numeric) of a population. They are usually the values that we try to estimate based on our sample data.

5. Why do we want estimates of parameters?

The "population" in these discussions is the group of numbers we want to know something about. e.g. the voting intentions of the electorate, the SFU student body, the stars in the sky over magnitude .1...

6. Why are point estimates not very useful?

An estimate of a parameter is not very reliable information unless it is reasonably precise – that is, unless the estimate has an acceptable standard deviation. So we need an estimate of precision to make point estimates more useful.

7. How do we provide more useful estimates?

...By using interval estimates, such as Confidence Intervals.

8. What determines the technique for CIs?

The question has an answer if we restrict our attention to CIs for the population mean. Then the determinants are the sample size, whether the population distribution is normal, and whether the population SD is known or not.

9. Why are proportions and averages treated similarly?

Because a proportion can be understood to be an average.

10. Why are proportions and averages treated differently?

Because 0-1 samples have less information per observation than samples of a truly quantitative measurement. For 0-1 samples, we usually need a large sample and then the CLT can be used to compute the CI.

11. Why does a small sample size cause problems?

1. For quantitative data: Because the standard deviation of the population cannot be reliably estimated from a small sample, and so we need to make allowance for this extra source of variation in constructing the CI (by using t distribution).

2. For categorical data: Because each observation has so little information, even 30 observations is quite small when estimating a proportion. While all proportions are in (0,1), the standard error for estimating a proportion is $\sqrt{p(1-p)/n}$, and if p=.5 and n=30, this is about 0.09, and ± 0.18 is quite a large part of (0,1).

12. When is the Central Limit Theorem needed?

When the population distribution is unknown, the CLT can give us a CI for a population mean as long as the sample is not too small.

13. How big does the sample have to be for the CLT?

In practice it depends on the degree of non-normality of the population distribution. But a rule of thumb is $n \geq 30$.

14. How is the CI related to the required sample size? The half-width of the confidence interval is equated to the allowable error of estimate, and this equation is solved for n.