QSCI 381 HW 4

Simon-Hans Edasi

5/11/2023

1. (3 points) In your own words please summarize the meaning of:

    (a) (1 point) The Law of Large Numbers.

    If you do an experiment enough times, the average outcome of the experiment will tend towards an expected value.

    (b) (1 point) The Central Limit Theorem.

    Regardless of the sampled population distribution, the distribution of means of samples will approach a normal distribution with increasing samples.

    (c) (1 point) The theorem underlying the distribution of the sample mean and standard deviation.

    If you draw enough random samples, the uncertainty of those samples will approach a normal distribution.

2. (1 point) Why are estimates of population parameters (statistics) useless without a measure of their inherent error?

    Because without error estimates we have no confidence in the estimate because it could be based on anything or nothing.

3. (3 points) What does the sampling distribution of a statistic, e.g., the sample mean, represent?

    The sampling distribution is a distribution of the statistics being measured in the sample. Say we were studying jellybeans in a population of jars on a table. Each jar is a sample, and the variance of color of jellybeans would be our statistics. We can measure the mean and variance of each color within each sample of the population.

4. (6 points) With regard to experimental design:

    (a) (2 points) Define a treatment and give an example.

    A treatment a statistical method applied to data. An example would be finding the standard deviation

    (b) (2 points) Distinguish between systematic and random errors.

    systematic errors are consistent and proportional, and can be traced back to a source. Random errors are chance encounters with fate.

    (c) (2 points) Distinguish between an experimental and an observational study.

    Observational studies are based purely on observations. Experimental studies introduce variables and observe outcomes.

5. (1 point) Complete the following statement by John Tukey: "Numerical quantities focus on expected values, graphical summaries on ...."

    unexpected values

6. (1 point) In the phrase random sample, what does the word random mean?

Each sample has the same probability of being selected.

7. (4 points) Suppose we wish to investigate the effect that different foods have on a species of fish. We place the food in the tanks containing the fish. We record the weight increase of each fish.

   (a) (1 point) What is the response?

   weight change in the fish

   (b) (1 point) What is the experimental unit?

   The fish tanks

   (c) (1point) How could we redesign the experiment to make the fish the experimental unit?

   Feed the fish individually different foods

   (d) (1 point) Was an actual treatment applied?

   No, all that we have done so far is collect data. No treatment yet.

8. (4 points) Suppose we wish to conduct a gene expression experiment. We take the RNA from two groups of male subjects (those with a disease and those without it). We apply their RNA to the microarray chip (slide) that contains genetic material (genes). The RNA is hybridized to the microarray, and fluorescence is used to measure the gene expression. There are several types of arrays available and different manufacturers for each.

   (a) (2 points) What might be two possible sources of error from confounding if we are not careful?

   Different refractive indexes of the array could impede fluorescence to different degrees. Quality of array can also affect measurements and be a confounding variable.

   (b) (1 point) What is the experimental unit?

   The array to which the RNA is applied

   (c) (1 point) What is the treatment?

   No treatment was mentioned in the prompt. If this is asking for one, I suppose I would use a binomial test to determine probability of disease showing up on different arrays?

9. (5 points) In 2010, the Physicians Foundation conducted a survey of physician's attitudes about health care reform, calling the report "a survey of 100,000 physicians." The survey was sent to 100,000 randomly selected physicians practicing in the United States,40,000 via post office mail and 60,000 via email. A total of 2,379 completed surveys were received.

   (a) (3 points) State carefully what population is sampled in this survey and what is the sample size. Can you draw conclusions from this study about all physicians practicing in the United States?

   This survey is sampling the population of practicing physicians in the United States with a sample size of 2,379. I would say no, we can not draw conclusions from this study about all physicians practicing in the United States.

   (b) (1 point) Given that the nonresponse rate is defined to be (1 – the percentage of responses), what is the nonresponse rate?

   1 - (2379 / 100000) = 0.97621

   (c) (1 point) Why is it misleading to call the report "a survey of 100,000 physicians"?

   Because they didn't actually survey 100,000 physicians.

10. (12 points) You measure the lengths (in cm) of 14 specimens of a certain insect and find that sample mean to be 22 cm and the sample standard deviation to be 1.2 cm. You may assume that the population is approximately normally distributed.

(a) (10 points) Find a 95% confidence interval for the population mean. Show all of your work including clearly stating the distribution you assume and the numerical values of the corresponding critical values. Start by stating the confidence interval in general terms.

Assuming a t distribution due to the small sample size, we look up a value for t^* for our degrees of freedom $n-1$. We are given a sample mean $\hat{x}$ and a sample error s.

$$n = 14; \quad df = n - 1 = 13; \quad \hat{x} = 22; \quad s = 1.2; \quad t^* = 2.160;$$

$$\mu = \hat{x} \pm \frac{t^* \cdot s}{\sqrt{n}} = 22 \pm \frac{2.160 * 1.2}{\sqrt{14}} = 22 \pm 0.693 = [21.307, 23.693]$$

(b) (2 points) Provide a one sentence description of what this confidence interval means in words. Be explicit in your description.

This confidence interval is a range of values that will contain the true population mean 95% of the time.