

Notes: Questions 1 and 2 will only be checked for completion.

Sampling from data with replacement in SPSS has proven to not be such a trivial task; I have not been able to find how to do it. The bootstrap is very feasible in other statistical software programs, but since you most likely will not be using these programs on a regular basis in the future, not to mention that most of these programs can be extremely difficult to learn, I cannot assign a good bootstrap problem. However, anyone who can give me a procedure for sampling from data with replacement in SPSS will get **80** (yes, eighty) extra points on this homework assignment.

1. (a) In your own words, describe (in detail) the entire process of hypothesis testing. Leave nothing to the imagination (I am aware we went over this in class, but it is important to fully grasp the concept of hypothesis testing.)
(b) Discuss when we should use hypothesis testing versus confidence intervals.
(c) Discuss how confidence intervals can be used to test hypotheses.
(d) Explain what type I and type II errors are. Give examples of both.
2. Describe in detail the process of bootstrap resampling. Be sure to cover the following points:
 - Why bootstrap resampling is valid
 - When the bootstrap is useful/should be used
 - How to calculate 95% confidence intervals using bootstrap samples
3. Suppose you want to investigate the attitudes of students at this school toward the statistics department faculty's commitment to teaching. You have been provided enough funds to contact about 200 students.

Give a detailed description of your sampling design. Be sure to cover the following:

- Your population. Are you going to sample from the entire student body, from just undergrads, or from some other subset of students?
- What type of sampling technique (simple random sampling, stratified random sampling, etc) will you be using?
- How are you going to collect information from the students?
- Finally, what could go wrong with your study?

4. An experiment was performed to determine the effectiveness of a new drug in preventing anxiety. The results are below:

Group	n	Mean Time Until Recurrence (Days)	s
Experimental	210	147.21	16.97
Control	188	141.03	24.47

- Calculate 95% confidence intervals for the true mean recurrence time for the experimental and control groups.
 - Can you conclude the new drug actually helps? Why or why not?
 - Previous studies claim the mean recurrence time for people using the drug is 150 days. Construct a hypothesis test to test the validity of this claim. State your null and alternative hypotheses, your p-value, and whether or not you reject the null hypothesis.
 - For each group, calculate the number of observations needed to make the 95% confidence intervals *at most* 0.01 units wide.
5. One study compared the hemoglobin levels of breast-fed and formula-fed infants. The blood hemoglobin levels for the infants at 12 months of age are summarized in the table below:

Group	n	\bar{x}	s
Breast-fed	23	13.3	1.7
Formula-fed	19	12.4	1.8

- Find 99% confidence intervals for the true mean hemoglobin level for the breast-fed and formula-fed groups.
 - Can you conclude that breast-fed babies on the whole have higher hemoglobin levels?
 - Test the claim that the actual mean hemoglobin level in breast-fed babies is 15. State the null and alternative hypotheses, report your p-value, and state whether or not you reject the null hypothesis.
 - What assumptions are we required to make about the data in order for your answers to parts (a) through (c) to be valid? Explain.
6. Suppose we have a random number generator that randomly generates any integer between 0 and some upper limit with equal probability. The upper limit is supposed to be 20, but we believe it to be larger.
- State the null and alternative hypotheses. How can we test the null hypothesis in a way such that the probability of a type I error is zero?
 - Suppose we do not reject the null hypothesis. Why can we still not say that we “accept” the null hypothesis in this case?