1. [20 pts] *Multiple Choice (4 parts). For each part, circle the roman numeral of the one best answer.*

   (a) [5 pts] When making a public report on a survey, which of the following is **not** required?

   [iii] The name of the statistical package used to do the analyses.

   (b) [5 pts] Which of the following statements is **most correct**:

   [iii] It is possible to construct a representative sample without random sampling, but it is more difficult to argue that it is really representative.

   (c) [5 pts] *Beneficence* is a basic ethical principle for research with human subjects. Which of the following is **not** an aspect of beneficence?

   [iii] Make sure the subjects get the benefit of some compensation for their participation.

   (d) [5 pts] Two important fractions in sample surveys are the *sampling fraction* n/N and the *response rate* r/n (where N is the population size, n is the intended sample size, and r is the number in the sample that actually responded). Which of the following is **not** true, for a simple random sample without replacement from the target population?

   [iv] You can get a more representative sample by increasing n, regardless of the response rate.

2. [20 pts]

   (a) [5 pts] During the past four weeks, beginning on [date 4 weeks ago] and ending today, have you done any exercise, including sports, physically active hobbies, and aerobic exercises, but not including any activities carried out as part of your job or in the course of ordinary housework?

   ___ *This item is OK as-is.*

   X *This item may not be OK. I want to ask the person who is trying these items the following question, to get information to improve this item:*

   Here are some possible probes:

   - Are there any kinds of exercise that you can think of, that are not included in the question yet?
   - How are you thinking about exercise involved in getting to and from work (e.g. biking)?
   - Do you know what aerobic means?
   - Do you have any trouble remembering your exercise habits for the past month?

   etc.

   (b) [5 pts] How many times a week do you have milk, butter or other dairy products?

   ___ *This item is OK as-is.*

   X *This item may not be OK. I want to ask the person who is trying these items the following question, to get information to improve this item:*

   Here are some possible probes:

   - Can you list other dairy products that the question may be referring to?
   - Dairy products are often ingredients in other foods. How would you answer the question if you consumed one of these foods, but no dairy products directly?
   - Do you think the amount dairy products consumed should affect your answer?
   - What period of time did you use to answer the question? This past week? An "average" week for you? How does an "average" week compare to this past week's consumption?

   etc.

   (c) [5 pts] Some people feel the US federal government should take action to reduce the national debt, even if it means unemployment would go up a lot. Others feel the government should take action to reduce the rate of unemployment even if it means the national debt would go up a lot. Where would you place yourself on the following 7-point scale?

1	*2*	*3*	*4*	*5*	*6*	*7*
Reduce Debt						*Reduce Unemployment*

___ *This item is OK as-is.*

__X__ *This item may not be OK. I want to ask the person who is trying these items the following question, to get information to improve this item:*

Here are some possible probes:

- Do you know what the national debt is?
- [If person marks middle category (4):] Did you mark that because you're truly neutral, because you don't know, don't want to reveal your opinion, or ...
- [If person marks any other category:] Did you find it difficult to place yourself on this scale? Are there too many/too few categories?

etc.

(d) [5 pts] During the past 12 months, since [date], about how many days did illness or injury keep you in bed more than half the day? Include days while you were an overnight patient in a hospital.

___ *This item is OK as-is.*

__X__ *This item may not be OK. I want to ask the person who is trying these items the following question, to get information to improve this item:*

Here are some possible probes:

- Is it difficult to remember as far back as a year? Let's try this month by month (or quarter by quarter etc.) and see if that changes your answer by much...
- If you stayed overnight in a hospital for one night, did you think of counting that as one day or two?

etc.

3. [20 pts] The American Hospital Association (AHA)...

(a) [2 pts] What is the target population *for this survey?*

Target population is apparently <u>All community hospitals</u> (or community hospital CEOs, or a similar answer).

(b) [2 pts] What are the units *in the target population?*

The units are <u>hospital CEO's</u> (or the hospitals themselves, or similar...).

(c) [2 pts] What is the sampling frame*?*

The sampling frame is <u>the list of fax and email addresses</u> that they used to distribute the survey.

(d) [2 pts] What sampling method *was used?*

No "sampling method" was used. Or you could say they tried to implement a census.

(e) [2 pts] What method of data collection *was used to contact units in the sample and collect data from them?*

This was a self-administered questionaire (SAQ), emailed or faxed to the respondents, that the respondents had to return (presumably by the same methods). [Any reasonable subset of this answer is OK.]

(f) The press release reports that "99.5% of hospitals reported experiencing one or more drug shortage in the last six months and nearly half of the hospitals reported 21 or more drug shortages."

i. [5 pts] Do you think this is useful information for the AHA? Why or why not?

This is very useful information for the AHA to have. Here are some possible reasons:

- As a headline it is great for promoting the AHA, getting contributions, getting hospitals to join th AHA etc.

- If the survey results are valid (unbiased) and reliable (low uncertainty/variability in the results) then it reflects a serious problem in hospitals.

Another acceptable answer might be: "This is NOT useful information for the AHA because it is not a scientifically valid survey." Possible reasons could include

- No random sampling
- Only CEO's with an axe to grind would self-select to return the survey form
- No nonresponse followup
- Not clear their frame actually has good coverage for the target population (the hospitals in the frame are probably AHA members, e.g.).

etc.

ii. [5 pts] Do you think this is a scientifically valid inference about the target population? Explain carefully why or why not, using <u>at least two</u> *concepts from the lecture notes or textbook.*

Here are some reasons (need two):

- No random sampling
- Self-selection bias in returning the surveys
- No nonresponse followup
- Not clear their frame actually has good coverage for the target population (the hospitals in the frame are probably AHA members, e.g.).

etc.

4. [25 pts] . . . assume that the AHA survey is a *simple random sample (SRS) without replacement*. . .

(a) [10 pts] 17% of hospitals in the sample reported that patient treatment was delayed because of a drug shortage. Calculate a 95% confidence interval for the same percentage in the population, based on the 820 responses in the sample.

We want the confidence interval to be

$$(0.17 - 2\sqrt{1 - n/N} \cdot SD/\sqrt{n},\ 0.17 + 2\sqrt{1 - n/N} \cdot SD/\sqrt{n})$$

where

- $\sqrt{1 - n/N} = \sqrt{1 - 820/5100} = 0.92$
- $SD \approx \sqrt{0.17(1 - 0.17)} = 0.38$
- $\sqrt{n} = \sqrt{820} = 28.64$

So the interval is

$$(0.17 - 2(0.92)(0.38)/28.64,\ 0.17 + 2(0.92)(0.38)/28.64) = (0.15,\ 0.19)$$

[if you used 1.96 instead of 2, the interval would be the same].

(b) [5 pts] What is the margin of error (ME) in part (a)?

The ME is $2(0.92)(0.38)/28.64 = 0.02$ (or just the half-width of the interval in part (a).

(c) [10 pts] Suppose we only want a ME of 0.01. What should the sample size be?

Using the estimate 0.17 from above, the SRS with replacement sample size would be

$$n \geq n_0 = \frac{(2)^2(0.17 \cdot (1 - 0.17))}{(0.01)^2} = 5644$$

and then the sample size for SRS without replacement would be

$$n \geq \frac{(5100)(5644)}{5100 + 5644} = 2679.11\ (or\ 2680)$$

Note: If we do not believe 0.17 is a good estimate then we might use the worst-case value 0.5 instead. This leads to

$$n \geq n_0 = \frac{(2)^2(0.5 \cdot (1-0.5))}{(0.01)^2} = 10000$$

and

$$n \geq \frac{(5100)(5000)}{5100+5000} = 3377.48 \text{ (}or\text{ 3378)}$$

Note: *the ridiculously large sample sizes here are due to a typo of mine: I asked for a ME of 0.01 when I wanted to ask for 0.05!*

5. [15 pts] . . . informed-consent form. . .

   (a) [12 pts] Identify (using the paragraph numbers in the figure) any four of the six essential elements for an informed consent form (by filling out the table below):

Name of Element	**Number of Paragraph Where it Appears**
Purpose of the survey	(1)
Risks/Benefits for the Participant	(3)
Confidentiality	(4)
Voluntary; participant is allowed to "opt-out"	(5)

   (b) [3 pts] Identify an element of informed consent forms that is missing from the figure.

   There is no information about how to contact anyone if the participants have questions or concerns about the research.

   Note: the six components of informed consent are:

   - Purpose
   - Risks/Benefits
   - Confidentiality
   - Compensation for harm
   - Contact info for any questions
   - Participation is voluntary