36-303 Homework 4

Reading for next week: Review Lohr 2.1, 2.5-2.6. Read Lohr Ch 8

The following questions are due Thursday, 2/19/2015, by 3:00PM

1) [adapted from Groves (2009), p255 #1] In order to gauge general public support for energy conservation, the following (fictional) question was included in a survey of American adults:

Do you agree or disagree with the following statement: I would support President Obama’s decision to use the U.S. military to help local cities achieve energy independence by installing more energy efficient public lighting.

a) What construct is this question trying to measure?
b) Do you expect this question to give valid, reliable and unbiased measurements of that construct? Why or why not?
c) Suggest some possible ways the question could be improved

2) [adapted from Groves (2009), p255-6 #5]

a) Thinking about the cognitive processing models of the response task, describe potential problems you see in the wording and response options for this question:

Many people who own vehicles have regular service work done on them, such as having the oil changed. What kind of service work do you usually have done on your vehicle? Specify one or two.

• Oil changes
• Fluid replacement
• Tune-up
• Body repair
• Warranty-related services
• Tire care
• Transmission overhaul
• Air conditioner treatment

b) Suggest some possible ways the question could be improved.

3) [adapted from Groves (2009), p256 #6] Consider the following survey questions:

i) During the past four weeks, beginning [DATE FOUR WEEKS AGO] and ending today, have you done any housework, including cleaning, cooking, yard work, and household repairs, but not including any activities carried out as part of your job?

ii) In the past week, how many times did you drink alcoholic beverages?

iii) Living where you do now and meeting the expenses you consider necessary, what would be the smallest income (before any deductions) you and your family would need to make ends meet each month?

For each question i-iii,

a) Write 2-3 potential probes for use in a cognitive interview. At least one should be specific to the question. Explain why you chose each probe. (For example, what part of the response process are you targeting? Is there a specific aspect of the question that seems likely to cause problems?)

b) Based the readings and lectures from last week, diagnose the problem(s) with each question and propose new question wording that solves (or mitigates) those problems.
4) (You should work on most of this problem individually, but it’s OK to discuss parts f & g with classmates.) Revisit http://www.stat.cmu.edu:3838/jsmurray/SamplingDistMean/, the app from HW 2. In this problem you will inspect the properties of confidence intervals obtained using the Central Limit Theorem (CLT) approximation to the sampling distribution of the mean (Section 2.5).

a) Change $N$ to 500 and $n$ to 20, and sampling design to SRS without replacement.
b) Inspect the “Population info” tab. Describe the population distribution, using the histogram, boxplot, and summary statistics.
c) Switch to the “Samples” tab, and take 100 samples using the appropriate button. Check the “Show approximate 95% CI” checkbox. The bottom plot shows the 100 samples of $\bar{y}$, plus their approximate 95% confidence intervals according to the Central Limit Theorem (CLT). Below that, the number of confidence intervals containing the true value is reported. Record this value. What would we expect this number to be if the CLT approximation is good?
d) Take another 1,000 samples. Check the “Show CLT approx” box. This shows the CLT normal approximation to the sampling distribution (solid line, in the first plot). Does it seem accurate? (Justify your answer) (Check the “Zoom sample histogram” box to get a closer look at the distribution)
e) Change the population to “Type 2” and inspect the “Population info” tab. Describe the population distribution, using the histogram, boxplot, and summary statistics.
f) Repeat steps c) and d) for this new population. If the results are different, explain why.
g) Increase $n$ to 100 and repeat steps c and d. Is the CLT approximation better or worse than in part f)? Why?