A1. EXERCISES FROM CHAPTER 1

1. For each of the following variables, determine whether the variable is nominal, ordinal, interval/ratio, or Likert scale::

    a. Admitting diagnosis of patients admitted to a mental health clinic.

        NOMINAL. Population: all mental health patients at clinics (or perhaps just this particular clinic, depending on what you are interested in)

    b. Weights of babies born in a hospital during a given year.

        INTERVAL-RATIO. Population: All babies.

    c. Pain scores provided by patients on a scale of 0 to 10.

        ORDINAL. (it doesn't really matter whether fractional responses are allowed here – distance on the scale cannot be assumed to have meaning). Population: All patients.

    d. Gender of patients visiting an eye clinic.

        NOMINAL. Population: All patients visiting eye clinics.

    e. Score on the *Psychology Today* depression test (see website).
        LIKERT SCALE (almost certainly they are summing/averaging ordinal responses). Population: All people.

    f. Range of motion for the elbow joints of students enrolled in a university health services curriculum.

        INTERVAL-RATIO. Population: All people of the typical ages of students in this curriculum.

    g. Temperatures for day-old infants who remain at least 24 hours in a hospital.

        INTERVAL-RATIO. Population: All day-old infants who remain hospitalized at least 24 hours.

    h. Disease-stage for patients having Parkinson's disease as defined using a 4-point scale.
        ORDINAL. Population: All people having Parkinson's disease.

    i. Satisfaction scores given by patients upon release from a hospital stay.

        LIKERT SCALE. Population: All patients in hospitals.

    j. Amounts of blood transfusions given to patients who have experienced trauma.

        INTERVAL-RATIO. Population: All patients who experience trauma.

2. For each item in question #1, identify the population of interest to the researcher.

See populations identified in #1 above.

3. Suppose that the hospital at which you work wishes to estimate the average age of their patients. For each of the following sampling schemes, assess the described sample for bias and confounding, explaining why certain samples are poor and ultimately making some argument for which sampling scheme you believe is best.

   It is important to keep in mind the population here: All patients at this hospital.

   a. Data are collected for every 25th patient admitted to the hospital during the next three weeks.

      This is a SYSTEMATIC sampling scheme (a type of convenience sample that incorporates an element of randomization). It is likely to be nearly as good as a RANDOM sampling scheme, and since random sampling is generally impossible, this would be our sampling scheme of choice.

   b. Dr. Adams specializes in treating cancer. Data are collected for all of her patients from the next two months.

      This is a CONVENIENCE sample and a poor choice at that. Cancer patients tend to be older and therefore this is likely to overestimate the average age for patients at this hospital (an example of sampling bias).

   c. Data are collected for all patients who are referred (for any reason) to Children's Hospital during the next year.

      This is another CONVENIENCE sample and another poor choice. Since we are likely only looking at children in this sample, the average age of all patients at the hospital would be grossly underestimated (sampling bias).

   d. Data are collected for all patients who come through the entrance and stop at a marked table to fill out the form, after which they receive a coupon for $5 off a meal in the hospital cafeteria.

      This is a CONVENIENCE sample (though it might seem self-selected, the experimenter did choose the group to be sampled) but one that is prone to substantial non-response bias. It seems likely to underestimate the average age of people visiting the hospital. The reason for this is that elderly patients likely have greater degree of illness and therefore are more likely to ignore the table in favor of more quickly pursuing treatment.

   e. Data are collected for all patients who visit the emergency room between 1:00 p.m. December 31 and 11:00 a.m. January 1.

      Another CONVENIENCE sample, and one that is likely to again underestimate age since younger people are more "active" during that time-frame and therefore more likely to need hospital care as a result of their activities.

4. You have three available potential treatments for a newly identified infection that is prevalent only in women (the presence of infection is easily identified based on a somewhat painful rash). For a variety of reasons it is not feasible to treat patients with more than one of the three treatments. It is also of note that, typically, the body would eventually fight off the infection on its own. Identify the research questions of interest here and design an experiment that should help to answer them.

The research questions would include (1) are any of these treatments better than doing nothing and (2) if so, which treatment is best. Since it is known that typically the body will fight off the infection on its own, a placebo should be used in this study (unless this is deemed to be unethical). There will be four groups (one for each treatment and the placebo group) and patients should be randomized into those groups. Data would be collected on the time it takes to fight off the infection for each patient on a given treatment.

5. Consider the Hrisanfow and Hagglund article published in the Journal of Clinical Nursing (you'll need to use a library to obtain the article).

Article Reference: Hrisanfow, E. and Hagglund, D. (2013). Impact of cough and urinary incontinence on quality of life in women and men with chronic obstructive pulmonary disease. *Journal of Clinical Nursing*, 22(2), 97-105.

a. Identify the population of interest to the researchers.

The population of interest to the researchers would seem to be anyone with COPD. However based on the description of participants

b. Identify the primary response variable(s) and associated data type(s).

Their primary variables of interest seem to be quality of life scores (probably Likert scale) and individual symptoms (ordinal – Likert items)

c. Only 66% of study participants completed the survey. Discuss whether this is likely to result in bias.

Answers may vary. It seems likely that patients for whom COPD is less of a burden might be less likely to complete the survey (perhaps feeling it doesn't really apply to them). This might result in negative response bias (i.e. generally worse symptoms and scores.

d. Would results of the study apply to people who are 35 years of age?

No, probably not. The sample did not include anyone under 50 years of age. Health is likely better at age 35 and this would likely affect quality of life scores and symptoms (in a positive manner for the 35-year-old).

e. Consider all variables listed in the first column of Table 1 (page 100 of the manuscript). Identify the level of measurement for each.

Interval-ratio variables: Age, BMI
Ordinal variables: COPD severity, Cough, Phlegm
Nominal variables: Marital status, Smoking history

f. The researchers intend to evaluate the impact of urinary incontinence (UI) on quality of life. It is noted that the average BMI for male study participants with UI is nearly two units higher than for those without UI. Explain why this could be a problem.

The issue here is that BMI might impact quality of life in some way. Therefore BMI could be confounded with UI. That is to say that if differences in quality of life are found, it would be impossible to know whether they were a result of UI, BMI, or some combination of the two.

6. Find another research article in your discipline that involves sampling. Assess the sampling design used by the researchers for potential effects of bias and confounding. Make suggestions for the improvement of their design.

Answers would vary widely.