These problems concern a recent research article* that appeared in a journal published by the CDC, regarding full adherence to national guidelines for physical activity, fruit and vegetable consumption, and nonsmoking among individuals with and without coronary heart disease (CHD). Please read the full article (although there may be terms and concepts which we have not yet discussed in the course), located at the link:

http://www.stat.wisc.edu/~ifischer/Intro_Stat/SIBS/CDC/Article.pdf

and answer the following questions.

1. Table 1 (page 10) displays the variables used in the study; note that they are all categorical. Classify each one according to whether it is ordinal or nominal; if the latter, classify further as either binary or not binary. (10 pts)

2. Table 3 (page 14) displays some summary statistics of the sample individuals associated with full adherence, both without and with CHD. (35 pts)

   (a) For the “Age” category, sketch density histograms for both groups; use R if you wish. Compare and make some general comments suggested by their appearance. (15 pts)

   Note: For simplicity, assume the endpoint convention that the age interval written 18-30 is actually [18, 31), the interval written 31-39 is actually [31, 40), etc. Note also that these intervals are quintiles. (Odds Ratios are not involved in this!)

   (b) For the “Without CHD” group, calculate the group mean and median age. Which of the two would you consider to be a more representative “measure of center” for this group, and why? (10 pts)

   (c) For the “Without CHD” group, calculate the IQR. (10 pts)

3. Consider Fig. 1a, the Venn diagram for the “Without CHD” group, and assume that the values indicated are accurate estimates of the corresponding population values. (That they sum to 101% is due to roundoff error, and is not relevant.) Answer the following. (3 pts ea = 27 pts)

   (a) What is the probability that a randomly selected individual fully adheres to the national recommendations for smoking abstinence, or fruit and vegetable intake, regardless of physical activity?

   (b) What is the probability that a randomly selected individual fully adheres to the national recommendations for smoking abstinence, as well as fruit and vegetable intake, regardless of physical activity?

   (c) What is the probability that a randomly selected individual fully adheres to the national recommendations for smoking abstinence, from among those individuals who fully adhere to the national recommendations for fruit and vegetable intake, regardless of physical activity?

   CONTINUED...

* According to the “Methods / Design” section (page 2), the data for this article were obtained through the BRFSS survey system; see http://www.cdc.gov/BRFSS/ for more information.
(d) What is the probability that a randomly selected individual *smokes*, from among those individuals who fully adhere to the national recommendations for fruit and vegetable intake, regardless of physical activity?

(e) What is the probability that a randomly selected individual *smokes*, from among those individuals who *do not* fully adhere to the national recommendations for fruit and vegetable intake, regardless of physical activity?

(f) Are full adherence to smoking abstinence and physical activity statistically independent (approximately)? If so, explain what this means in non-statistical terms.

(g) What is the probability that a randomly selected individual fully adheres to exactly one of these three national recommendations?

(h) What is the probability that a randomly selected individual fully adheres to exactly two of these three national recommendations?

(i) What is the probability that a randomly selected individual fully adheres to all three of these national recommendations, *given* that they fully adhere to at least two of them?

4. **Note:** In order to answer this question, it is necessary to understand a concept from Chapter 6. Namely, a 95% confidence interval (CI) for the true value of a population parameter is a sample-generated interval estimate that presumably contains this unknown value with 95% “confidence” (a notion related to “probability”). Recall the introductory “age at first birth” example presented in this early PowerPoint document. (8 pts)

In large public health surveys such as this – which, recall, are cross-sectional studies – the definition of Odds Ratio (OR) of an outcome event $A$, with respect to event $B$, must be slightly modified from longitudinal studies, such as cohort (where $A = D^+$ and $B = E^+$), and case-control (where $A = E^+$ and $B = D^+$), but its basic structure $OR = \frac{Odds(A | B)}{Odds(A | B^c)}$ remains the same.†

(a) It can be observed from the Sex category of Table 3 that among those *with* CHD, $OR = 1.23$ for females, with a 95% CI of 0.63-2.41. (Note: In both CHD categories, males are treated as the reference group, with $OR = 1$, in order to make comparison between the sexes easier.) Interpret the highlighted statement above, in the context of the study. (4 pts)

(b) The last paragraph of the “Results” section of this paper (page 5) includes the following excerpt regarding Table 3: “Among individuals *without* a history of CHD,… female sex ($OR = 1.47$; 95% CI, 1.23-1.76) [was] associated with full adherence.” Why is this conclusion formally justified, and why can it not be formally claimed for those *with* CHD in part (a)? (4 pts)

---

† Likewise, the numerator odds of $OR$ are defined as $\frac{P(A | B)}{P(A^c | B)}$, the denominator odds are defined as $\frac{P(A | B^c)}{P(A^c | B^c)}$. 

---
5. Also according to the “Results” section of this study, 5.6% of the individuals in the final sample reported CHD. Suppose the study had also included a fourth national recommendation of a daily regimen of low-dose aspirin, with the following information. From among those individuals who reported CHD, 21% fully adhered to the aspirin regimen; from among those individuals who reported no CHD, 83.5% fully adhered to the aspirin regimen. 

(a) What is the probability that CHD was reported, from among those individuals who fully adhered to the aspirin regimen? (2 pts)

What is the probability that no CHD was reported, from among those individuals who fully adhered to the aspirin regimen? (2 pts)

(b) What is the probability that CHD was reported, from among those individuals who did not fully adhere to the aspirin regimen? (2 pts)

What is the probability that no CHD was reported, from among those individuals who did not fully adhere to the aspirin regimen? (2 pts)

(c) Calculate the odds of CHD, among those individuals who fully adhered to the aspirin regimen. Interpret this quantity in context. (4 pts)

Calculate the odds of CHD, among those individuals who did not fully adhere to the aspirin regimen. Interpret this quantity in context. (4 pts)

(d) Compute the odds ratio of CHD, relative to adherence to the aspirin regimen. Interpret this quantity in context. (4 pts)