STAT220: WINTER 2006: QUIZ 1: JAN 19
SOLUTIONS for 20-minute quiz.

There is not a unique correct explanation, but the points in **bold** are the ones you needed to make.
The comments in *italic* give a little more detail.

1. (**Explain your answers**) A small undergraduate college has 4000 students, evenly distributed among the four classes: freshman, sophomore. junior, senior. A survey is designed to investigate the types of employment undertaken by students.

(i) (2 points). A complete alphabetical list of the students is available; it is proposed to take every 20 th. name on the list (that is, number 20, 40, 60, 80, ..., number 4000). Is this an acceptable sample? Is it a simple random sample?

Yes it is an acceptable sample. It would be hard to think how any factor associated with employment types could also be associated with being number 20, 40, ...
In particular, these names will be spread through the alphabet in approx the same way that they are on the total list, Also the four year classes will be approx equally represented.

No it is not a simple random sample because:
EITHER Aardvark,Abe will not be on it, but Zymmermann,Zach will.
That is, there is a bias for/against the really extreme names, alphabetically
OR Smith, Kylie and Smith, Larry will not both be on it.
That is, people with very similar names will not be independently sampled/not.
OR:... *There are a few other possible correct explanations here.*

(ii) (2 points). Another proposal is that four interviewers should be chosen, one from each class, and that each should take the first 50 people they meet from their class as their sample. Is this an acceptable sample?

No this is not an acceptable sample. The first people each interviewer meets is not going to be a random sample of the class. There are many (confounding) factors that would be related both to getting sampled under this scheme AND also likely associated with the student's type of employment.
For example: they will meet their friends. Likely their friends will have similar employment types to the interviewer. They will meet people on campus. Being on campus will be associated with on-campus/off-campus type employment.

(iii) (2 points) Propose another scheme that combines good features of the two proposals (i) and (ii).

Stratifying by the class year (as in (ii)) is a good idea. Taking every 20 th. name of a class list (as in (i)) of each year-class would be one way to do it.
(OR, one could start counting at a random point in the first 20 ...)
Stratifying is good here, because year-class is likely associated with employment type. But note it is not NECESSARY – the classes will be roughly evenly balanced taking every 20 th. name, or taking a simple random sample.

2.(**One-word answers are sufficient**) The following three histograms show the distribution of annual household incomes (in units of $1000) in three suburbs Avondale (A), Burnaby (B) and Chester (C) of a large US city.

Figure from FPP P.34, Question 4

(i) (2 points) City Councillors are interested in what proportion of households have income <u>more than</u> $50,000 per year. For each of the 3 suburbs, is this proportion about <u>equal</u> half, well <u>over</u> half, or well <u>under</u> half?

| Avondale: **over** | Burnaby: **under** | Chester: **equal** |

(ii) (1 point) In one suburb, there are two quite different groups of households, one quite well off and the other much poorer. Which suburb?
Burnaby (or "**B**")

(iii) (1 point) In Burnaby, are there more households with incomes in the range $40,000 to $50,000 than in the range $90,000 to $100,000? (Yes or no.)
No