

## STAB22 section 3.1

**3.1** To do this, you have to be able to say that all students at your college, as a group, are “like you and your friends” in the sense that all students (as a group) will have similar attitudes towards this TV show. This is hard to justify; as we’ll see later in Chapter 3, the best way to get a sense of how “all students” feel is to take a random sample of them. As it is, the opinions of you and your friends are like anecdotal evidence, in that you cannot know how well they will generalize, and you might suspect that they may not very well.

**3.3** Not all people are like Jamie (not everyone is a “hard-core computer programmer”), so not everybody is likely to share Jamie’s preferences for Jolt Cola. (The nature of computer programming is that people doing it like to work at night, or stay awake for long periods of time, and a lot of programmers have a preference for some kind of caffeinated drink. Coke and Pepsi have some caffeine in them, but not nearly as much as Jolt, and not everyone is looking for caffeine in their favourite drink.)

**3.4** This is only one kind of accident, with a combination of circumstances that happen to lead to “not wearing a seat belt” being a better choice. Not all accidents are like this; there are many kinds of accident where a seat belt would stop you going through the windshield or otherwise prevent you from a worse fate. Looking at all accidents, or a random sample of them, would give you a better sense of how likely a driver is to get into different kinds of accident, and how helpful a seat belt, air bag etc. is in reducing the severity of the outcome.

If you want an “explain briefly” answer: this is anecdotal evidence (that Herman may remember precisely because it is unusual) that doesn’t reflect the whole phenomenon of car crashes.

**3.5** In Chapter 2, exercises 2.139 and 2.140, we used available data on basketball scores. These were collected because people wanted to

know how the basketball teams were doing (presumably those web pages get enough hits to make them worth while maintaining). Maybe not many of the web page visitors were using the data to make scatterplots, though!

Of course, there are many other possible answers, and other actual studies that available data were used for, but this is a simple example.

**3.6** The link <http://www.thestar.com/article/445835> is to a Toronto Star article about a large number of Canadians who don’t have a family doctor. The figure of 4.1 million came from the Canadian Community Health Survey, in which 65,000 Canadians were chosen (presumably at random) and asked some questions about their health. (The figure wasn’t arrived at by asking every single Canadian “do you have a family doctor?”, but by figuring out how many of the 65,000 sampled didn’t have one, and scaling upwards to the number of adults in the country.) Some of the other conclusions described in the article were that recent immigrants were less likely to have a family doctor, and that men were less likely than women to have a regular family doctor. Related to all of this is the difficulty immigrant doctors face in being able to practice in Canada, availability of doctors in rural areas, and so on, but that’s outside the scope of this survey.

(I found this article by going to the Toronto Star web site, and searching there for “survey”. This was the first hit when I looked, and there were other hits about different surveys if you want one that interests you more. The Canadian Community Health Survey can be searched for too: I found links to it from the Statistics Canada and Health Canada websites, from which it seems clear that the survey covered a lot of things apart from being able to find a doctor.)

**3.7** This is not an experiment, because the people involved were not asked to use (or not use) a cell phone. It is therefore an observational study, but one of the best kinds (called a case-control

study), where you see if people who have brain cancer are higher cell-phone users (after allowing for age, sex, race etc.) than people who do not.

If you wanted to do an experiment here, you would *control* cell phone usage: you would have a number of subjects, and you would (randomly) assign some of them to be cell phone users for the duration of the experiment, and the rest to be non-users. At the end of the experiment, you would see whether the people you assigned to be users have a higher incidence of brain cancer than those who were not. This is a difficult experiment to run because you would have to follow these people for a (large) number of years, and even then there might be residual effects from pre-experiment cell phone use. Also, brain cancer is rare, and you might not see many cases at all in your experiment. This is why people use case-control studies, particularly for rare diseases: there is no need to impose a treatment, and because you use people who are known to have the disease, the rarity of the disease doesn't matter.

The explanatory variable is cell-phone usage (amount, or a yes/no), and the response is brain cancer (yes/no).

**3.8** There was no attempt to control the amount of time watching TV, so it's an observational study. The explanatory variable is the amount of TV watched, and the response is "later aggressive behaviour" (however you measure that).

A child's TV watching might be associated with how much TV his/her parents watch, or what chores the child is asked to do around the house, or how much homework the child has, or how much time the child spends with friends (and I'm sure you can come up with more). It can be difficult to figure out which way the cause and effect goes: it could be that TV watching influences something else that itself is the real cause of aggressive behaviour.

**3.9** This one is an experiment because the software company chose whether each student got the software or just read the text. This

explanatory variable (software/text) was imposed on the students, and the response is "increase in understanding" as measured by difference in test scores (I would imagine).

**3.11** In this experiment, the experimental units are the material on which the experiment was conducted, here food samples. The treatments are the radiation levels, and the response is the oxidation of the lipid. A factor is a (usually categorical) variable that may affect the response; its levels are the different categories or values that were used in the experiment: here the 9 different radiation levels.

If other lipids are "like" the one studied, we can have some confidence in generalizing the results of this experiment to other lipids, but if this is not known, we should be very hesitant, and wait for similar experiments to be done on the other relevant types of lipid.

**3.12** This is an experiment because the treatment (instruction method) is imposed on the students (experimental units) in the course. The response is the score on the standardized drawing test. The factor is (as usual) the same as the treatment: it is instruction method, and its levels are the different methods, web-based and paper-based. The students in this course may not be like students in other courses who might be learning the same material, but to the extent that they are similar, the results might be (cautiously) generalized to other settings. (It would be better if the experiment were done on a random sample of all computer graphics technology students.)

**3.17** In (a), what could happen is that the students who sit towards the front of the class are different in terms of how well they do the lab exercise, compared to the students who sit at the back. If the students sitting at the front get one form of the exercise and the students sitting at the back get the other, we won't know whether any differences in results are because of the different forms of the lab exercise, or because of differences in the students who got that exercise.

In short, the way in which the treatments (lab exercises) are assigned may be mixed up with the results that are observed. An improved experiment would be to randomize which rows (or, better, which students) get which form of the exercise (with half the class getting each).

In (b), there is no control group. Maybe the leadership program appears to work because all students would improve on this standardized leadership test over time. (Think of increasing maturity, say.) So we have to form a control group: half the students go through this new leadership program, and the other half go through the previous leadership program (if there is one), or meet every so often to discuss leadership, or do nothing apart from take this test (twice), or something like that.

In (c), there could be a difference between fall and spring. (I taught calculus once, in the winter semester, and my class consisted largely of people who had also taken calculus in the fall, and didn't make it through.) So I would expect the students in the second-semester course to be weaker than those in the first-semester course. Also, the second-semester course is different from the first-semester course (botany vs. zoology), so we are not comparing like with like.

The experiment could be improved by only looking at (say) zoology, and conducting the experiment over several years, sometimes with the new method used in the fall and sometimes in the spring. Or a mixture of different courses could be used (at the cost of a more complicated analysis), making sure that the new method is used with each course sometimes in the fall and sometimes in the spring (again requiring several years to complete the experiment).

**3.18** In (a), sorting by last name isn't randomizing (for instance, some ethnic groups have last names predominantly starting with certain letters). The 50 names could be shuffled in Minitab, or you could use Table B to pick them out 10 at a time, or you could use the procedure described in Example 3.12.

In (b), what could happen is that you get 6 heads and 2 tails (say), so you end up with 6 subjects getting the first treatment and only 2 getting the second. Randomly shuffling the list of subjects, and then picking out the first four on the shuffled list to get the first treatment, is better. (The proposed method is not *so* bad, especially when you have a lot of subjects, because you will get *about* the same number in each group, but having *exactly* the same number in each group is better, and not that hard to do.)

In (c), the batches of rats might be differently good (in terms of what the experiment is measuring), so that the treatment that happens to get the first batch of rats (say) may end up looking good without actually *being* good. This can be fixed by randomizing the individual rats: rat #1 gets treatment 2, rat #2 gets treatment 3, rat #3 gets treatment 3, etc. (or whatever the randomization produces). The randomization can be done ahead of time, because the experimenter knows that 80 rats will be used, 20 in each treatment. A potential difficulty is that the experiment may be set up so that it is easier to run one treatment on all 20 rats, rather than a bunch of different treatments on smaller numbers of rats (as the design I'm proposing requires), but no-one said that randomization would make running the experiment easy!

**3.21** First obtain some seeds (or maybe some young plants) for the plant of interest. These will be the experimental units. The treatment is going to be "compost tea", and it will be good to have a control group, say having the plants grow in regular compost. Randomly split the available seeds/plants into two groups. The seeds/plants in the first group will be grown in compost tea and the ones in the second group will be grown in regular compost (or however your control group is going to work). Decide on how long the experiment will run (this may depend on how long the plants take to grow), and, at the end, measure something that describes the effectiveness of growth: how tall the plants are, what fraction of seeds germinated (if you started from seeds), that kind of thing.

Because of the randomization and the structure of the experiment, if you find that the compost-tea plants did grow better, you have good evidence that it was the compost tea that was the reason.

**3.23** For these next few problems, you can start by identifying the explanatory variable(s) and the response. The factor is the explanatory variable, which, in an experiment, is controlled by the experimenters. The levels are the particular values or categories of the explanatory variable that were chosen for the experiment; the treatment is the combination of levels used. The experimental units or subjects are the “material” on which the experiment was done: things or people.

Thus the explanatory variable here is amount of light, and the response is dry weight. The factor is also amount of light, and the treatment is full light, or light shaded to 5%. The experimental units are the pine seedlings.

**3.24** There are two explanatory variables (and hence factors) here: physical activity intervention and nutrition intervention. The responses are also two: amount of physical activity and lunchtime consumption of fat (both things are observed). The treatments are four: each factor is either a “yes” or a “no” for each school, so there are  $2 \times 2 = 4$  treatments (factor level combinations). The experimental units are middle schools. (Careful here: students at each school undergo the interventions, but schools in this experiment are randomized and measured as a whole.)

**3.25** The purpose of this study is to see how the introduction method of the interviewer, and the offer to send the survey results later, affects whether the interview is completed. Thus there are two explanatory variables (factors): the introduction method, and the offer to send the results. There is one response, whether the interview is completed. There are three choices for the method of introduction, name, university or both; there are two choices for the offer to send survey results (offered or not offered). So there

are  $3 \times 2 = 6$  treatments (explanatory variable combinations). The subjects are people (strictly, those people appearing in the telephone directory, or however the selection of people to call was made).

**3.26** The explanatory variable/factor is taking aspirin, the levels being “take an aspirin every second day” and “take a placebo”. The response variable is “having a heart attack”. The experimental subjects are the male physicians.

Without a diagram: divide the physicians into 2 groups of 11,000. Give the first group the aspirin treatment, and give the second a placebo. Observe the number of heart attacks in each group.

The logic here seems backwards, but this is correct: *if* the aspirin has no effect above placebo, we could work out what differences in number of heart attacks between the two groups is plausible — that is, what difference you might see by chance. (In Chapter 5 and beyond, we’ll see how this is done.) To say that the aspirin group had “significantly fewer” heart attacks means not only that the aspirin group had fewer heart attacks, but also that the difference is bigger than you would expect by chance.

**3.27** There are 36 subjects altogether. Number the subjects 1–36 in some way (along the rows seems to make most sense, since the subjects are in alphabetical order if you read along the rows). There are four treatments, so we want to divide the subjects into 4 groups of 9 subjects each.

Using Table B at line 151, read off the digits in twos (36 being a two-digit number), and discard 00 or anything bigger than 36. This gives: 03 (Bezawada), 80 (discard), 22 (Mi), 93 (discard), 41 (discard), 29 (Shu), 26 (Paul) and so on. If you get a number you’ve had before, discard it. Keep going until you have 9 names selected (we have 4 so far). These form the first group (to receive antidepressant alone). Select groups 2 and 3 (placebo alone, antidepressant plus stress management) in the same way,

taking care not to select names you’ve chosen before (a subject can’t be in two different groups). When you’ve done this, you’ll have nine names left; these go into group 4 (placebo plus stress management). There’s no point in continuing with table B to select these names one at a time, since you know which group they’ll end up in.

Instead of using Table B, you can use Minitab to do the randomization. There are many ways to do it; one way is this:

- type the 36 names into column 1 (or if you prefer, the numbers 1–36, but the names themselves will make your job easier later)
- Select Calc, Random Data and Sample from Columns. You want to sample 36 rows from C1 and store them in (say) C2; make sure to leave Sample with Replacement at the bottom unselected, or else you’ll get duplicate names. This produces a “random shuffle” of the 36 names, stored in column 2.
- click OK.

My results are shown in Figure 1. Yours will be different. Then select the groups by reading off groups of nine names along the rows (as far as Kim for group 1, as far as Nolan for group 2 and so on). Since randomness is randomness, if you want to read off the names going down the columns, that will work too, although of course you’ll get different groups that way. (The right way to do this is to specify ahead of time whether you’re going to read off the names along rows or columns; that way, nobody can accuse you of making the choice to suit yourself.)

In case you were wondering, making a random selection from 1–4 for each subject wouldn’t work, because you may not get 4 groups of the same size (you might get 11 1’s, for instance). But you would expect to get 4 groups of “about” the same size.

**3.28** The explanatory variable/factor is “strength of marijuana cigarettes”, the levels being “strong” and “weak”. The responses

Data Display				
C2				
xu	saygin	cetin	li	archberger
cheng	chronopoulou	daye	kim	wang
lu	watkins	lipka	hatfield	park
hua	guha	nolan	mehta	martin
kumar	tyner	daggy	towers	shu
rau	codrington	leaf	vassilev	mi
olbricht	bezawada	tang	engelbrecht	anderson
paul				

Figure 1: Random shuffle of names as done by Minitab

are work output and earnings. The experiment is designed by collecting 20 subjects (see (b)), divide them at random into 2 groups of 10, give one group strong cigarettes and the other group weak, and then observe the work output.

Number the subjects (listed on page 194) from 1 (actually 01) to 20 along the rows, and then go along the rows of Table B, selecting 2 digits at a time, starting at line 101, until you have 10 names. These form one group, and the other 10 names form the other group. This works the same way as 3.27: you discard any numbers you can’t use or have used before. Because there are only 20 subjects, you’ll end up doing a lot of discarding.

I get 19 (Wayman), 22, 39, 50, 34 (discard these), 05 (Cunningham), 75, 62, 87 (discard these), 13 (Mitchell), 96, 40, 91, 25, 31, 42, 54, 48, 28, 53, 73, 67, 64, 71, 50, 99, 40, 00 (discard all of these), 19 (already had this one, so discard), 27, 27, 75, 44, 26, 48, 82, 42, and so on. Keep going until you have 10 names, moving on to the next row of Table B when you run out of numbers in one row. These 10 names get the strong cigarettes, and the other 10 get the weak ones.

Using Minitab for this is a lot less work. Follow the same idea as 3.27: type the 20 names into a column, get a random shuffle of these names, and then pick out the first 10 for the “strong” group and the other 10 for the “weak” group. It’s almost more work to describe than it is to do!

**3.29** There are 3 treatments (black tea extract, green tea extract, placebo). The 18 rats are randomly assigned to one of these three groups. The growth of cataracts was observed.

Assign the rats numbers 01 to 18 (apparently lab rats don't have names!). Using Table B beginning on line 120, the six rats selected for group 1 are 16, 04, 07, 10, 13, 15; the six for group 2 are 05, 09, 08, 18, 03, 01 (which takes you right down to row 124 of Table B), and the other 6 rats go into group 3.

Or you can use Minitab, in the same way as 3.27. Put the numbers 1–18 into a column, and randomly sample them into another column. The first six numbers are the rats going into group 1, the next six go into group 2, and the last six go into group 3.

**3.30** The 90 subjects were randomly assigned into one of 3 groups (30 in each group). The subjects in each group received the treatment for that group shown in the question. Performance was observed after the break.

To use table B, take the digits in groups of 2, because the subjects are numbered up to 90, which has 2 digits. Starting at line 153, the first four chosen subjects are 07, 88, 65, 68. (These were chosen without any discarding, because 90 is close to 100.) Or you can use Minitab to randomly permute the numbers 1–90, then choose the first 30 numbers to form group 1, the next 30 to form group 2, and the last 30 to form group 3.

**3.31** There are 4 treatment groups in 3.24, so we need to divide the 24 schools into 4 groups of 6. In Minitab, put the numbers 1–24 into a column and randomly shuffle them into another column using Calc, Random Data and Sample from Columns. My results are shown in Figure 2. Thus I choose schools 8, 6, 18, 12, 21 and 14 for treatment 1, schools 22, 19, 11, 20, 7 and 13 for treatment 2, schools 2, 9, 4, 1, 16 and 3 for treatment 3, and schools 17, 10, 24, 5, 23 and 15 for treatment 4.

Data Display

C2	
8	6 18 12 21 14 22 19 11 20 7 13 2 9 4
1	16 3 17 10 24 5 23 15

Figure 2: Random shuffle of 24 schools

Or, as in the previous exercises, you can use Table B. Number the schools 01 to 24, and select digits in twos.

**3.32** The two factors are discount level and fraction of shoes on sale. There are 4 levels of each factor and thus  $4 \times 4 = 16$  combinations, which you can display in a table:

	Fraction of shoes on sale				
	25%	50%	75%	100%	
	20%	1	2	3	4
Discount level	40%	5	6	7	8
	60%	9	10	11	12
	80%	13	14	15	16

The 16 treatment combinations are shown. The response is the “sale attractiveness rating” given by each subject.

With 96 students, there will be  $96/16 = 6$  students getting each treatment combination. So number the students from 1 to 96 (01–96), and use Minitab or Table B to pick them in 6's; the first six get treatment combination 1, the next six get treatment combination 2, and so on, taking the treatment combinations from the list or sketch you made in (a). (In my case, that could mean reading along the rows of my table.)

**3.33** The factors are: type of roller, dyeing cycle time and temperature. There are 2 types of roller, 2 cycle times and 2 temperatures, so there are  $2 \times 2 \times 2 = 8$  treatments (factor level combinations). If you want to list the treatments, you can:

Treatment	Rollers	Cycle time	Temperature
1	Metal	30	150
2	Metal	30	175
3	Metal	40	150
4	Metal	40	175
5	Bristle	30	150
6	Bristle	30	175
7	Bristle	40	150
8	Bristle	40	175

You can check that every combination of roller type, cycle time and temperature is represented here. This is a common type of design, called a “ $2^3$  factorial design”, which means 3 factors, each at 2 levels. Since  $2^3 = 8$ , this tells you at once that there must be 8 treatments. If there were just one fabric specimen getting each treatment, you would need 8 specimens, but since 4 specimens will be given each treatment, you need  $8 \times 4 = 32$  specimens altogether. (When you have more than one experimental unit getting each treatment, the design is said to be “replicated”, which is a good thing because it gives you more information about how the factors affect the response. Specifically, you can directly compare the 4 specimens that get the same treatment. How variable those 4 measurements are gives you an idea of how much of the variability in the data is random and how much is because of the effects of the treatments.)

To do the randomization, you would number the specimens 1 (01) to 32, and randomly choose 4 (using Table B or Minitab) to get each treatment.

**3.38** The technical terms are “randomized”, “double-blind” and “comparative”. “Randomized” means that some fraction of the patients, probably half of them, chosen at random, received the CR morphine tablet and the rest received morphine in the standard way. “Double-blind” means that neither the patient nor the person administering the morphine knew which kind of morphine

it was (though you might imagine that patients receiving morphine in the standard way would know when they had received a “hit”). “Comparative” means that the study compared CR morphine with the standard delivery method, rather than just trying the new method and seeing how it worked. (There might be something about this hospital that makes morphine generally more or less effective; this way, the researchers are comparing the two delivery methods under the same conditions. In any case, the issue is not just whether CR morphine works; it’s about whether it works *better* than the standard delivery method.)

**3.39** “Randomized” means that the available subjects were assigned to receive St John’s wort or placebo (the subjects had no choice in the matter). If you read further on in the question, you’ll find that 98 participants ended up in the treatment group and 102 in the placebo group; this is probably because the two groups started out the same size, but some of the participants had to drop out of the study for reasons unrelated to the experiment (eg. they got a cold, or chose to withdraw from the study). “Double-blind” means that neither the participants nor the people giving the “drug” knew whether it was St John’s wort or the placebo. “Placebo-controlled” means that there was a control group, so that the effect of St John’s wort was being compared with something that was supposed to have no effect on major depression.

My diagram is in Figure 3.

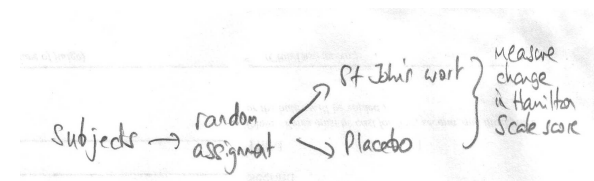


Figure 3: Experiment diagram for depression study

**3.43** A “matched-pair” design here would involve having each player do the trials twice: once with oxygen (before the final run, or

between each of the runs), and once without, randomizing as to whether the with-oxygen or without-oxygen trials are done first. Having three runs of 100 yards in one day is physically demanding enough (at least, it would be for me!), especially with the pressure of knowing that the final run will be timed, and so if all six 100-yard runs are done in one day, the results of the second trial, whether it is with oxygen or without, could be affected by exhaustion. This suggests that we will get more trustworthy results by doing the experiment over two days. (We are removing a possible factor, “tiredness”, from the experiment, which is always good to do.)

To do the randomization, number the players 01-20 (or use the players’ jersey numbers, if you prefer, moving on to the next two digits in Table B if you pick a number that isn’t on any of the players’ jerseys). Then use Table B in the usual way. I got: 12, 13, 04, 18, 19, 16, 02, 08, 17 and 10, skipping a 12 and a couple of 00s. These players get oxygen on their first-day trial (and not on their second), while the other 10 players get oxygen on their second-day trial and not on their first. It would probably also be smart to randomize the order in which the players run their trials on each day (say, using Minitab to produce two separate shuffled lists of the numbers 1–20, which are the “running order” for each day). This way, the running order isn’t mixed up with use of oxygen or not, which is sound statistical practice, even though it would be hard work for the coach to keep track of everything.

next digit could not be 0, which makes it (somewhat) predictable, and that’s no good.)

**3.46** Each digit in Table B is equally likely to be anything from 0 to 9, *independently* of any other digits you’ve looked at. So (a) is false, because you can’t say exactly how many digits will appear on each row. (It will be “about” 4, though, just not exactly 4). (b) is true, because that’s exactly how we select two-digit numbers (like the football players in 3.43). For (c), “random” just means “unpredictable”, in that we cannot say what is coming next; by that token, 0000 is as random as anything else. (If we knew we could not get 0000, as soon as we saw 000, we’d know that the