5.7. (a) The response variable is whether the student has developed symptoms of ADHD after two years (yes/no). The explanatory variable described in the problem is the frequency of digital media use. There are probably many other explanatory variables (associated with ADHD development) that were not described in the problem.
(b) Researchers did not "assign" the students to any type of a treatment or intervention group. They simply observed behavior over the two-year period. This is an observational study.
(c) An observational study may reveal an association exists between ADHD and digital media use, but this should not be interpreted as causation. There are many possible lurking variables too, for example,

- injuries
- environmental factors

Both of these could be associated with ADHD development. There is also the possibility ADHD simply develops later on in a child's life (i.e., after the 10th grade year). Therefore, although the child did not have symptoms when the two-year period started, symptoms could have developed later as part of the child's genetic predisposition to the condition.

5.16. (a) The response variable is whether a subject developed colon cancer (yes/no). The explanatory variable is the antioxidant. There are probably many other explanatory variables (associated with colon cancer development) that were not described in the problem. This is an experiment.
(b) The researchers would use randomization to assign subjects to one of the four antioxidant (treatment) groups:

- Group 1: Daily beta-carotene
- Group 2: Daily vitamins C and E
- Group 3: Daily beta-carotene + vitamins C and E
- Group 4: Placebo.

The subjects would be followed over a four-year period and then tested for colon cancer at the end of the trial.
(c) Give each subject a three digit code such as 000, 001, 002, 003, ..., 864. You could use Table of Random Digits in Moore and Notz, or you could use R:

```
> colon = seq(1,864,1)
> sample(colon,5,replace=FALSE)
[1] 610  41 368 741 367
```

Subjects labeled 610, 41, 368, 741, and 367 would be the first five subjects assigned to the beta-carotene group.

If the researcher wanted to balance the number of subjects for each treatment group, s/he would use randomization to assign the first 214 subjects to Group 1, the next 214 subjects to Group 2, and the next 214 subjects to Group 3. The remaining 214 subjects would be assigned to Group 4.
(d) This means the differences in the colon cancer percentages among the four treatment groups were small and could have arisen by random chance (i.e., by natural sampling variability).
(e) People whose diets include fruits and vegetables may have better diets overall, better overall health, better exercise habits, better lifestyle choices, etc. In other words, all of the people in this trial may have been very low risk to begin with. Performing this trial targeting a different population of subjects may give different results.

5.19. This is an experiment. Students were randomly assigned to the two groups:

- Group 1: Ads spoken in high pitch
- Group 2: Ads spoken in low pitch.

(a) The explanatory variable is the ad pitch (high or low). There might be other explanatory variables (associated with perceived sandwich size rating) that were not described in the problem.
(b) The response variable is the perceived size of the sandwich, which is given as a rating on the numerical scale: -3, -2, -1, 0, 1, 2, and 3.
(c) Yes, they could have included a third group where the ads are spoken in normal pitch. Therefore, if there is a placebo effect in this experiment, its effect could be assessed directly by including the third group.

5.24. This means that the differences between students who used software and those who did not were small and could have arisen by random chance.

5.25. (a) In this problem, the authors want you to design an experiment. There are two groups of subjects:

- Group 1: Cocoa pill
- Group 2: Placebo.

Baseline heart health measurements will be made on all 50 subjects. We will then use randomization to assign subjects to one of the two groups. The subjects will be followed over a two-year period and their heart health will be assessed again.

(b) Give each subject a code such as 1, 2, 3, ..., 49, 50. We can use R to randomly select 25 subjects:

```
> subjects = seq(1,50,1)
> sample(subjects,25,replace=F)
 [1]  5 14 15  7 29 32 37 44 39  9 34  8 13 23 43 50 10 31 16 42 38 46 22 36 47
```

Subjects whose codes match these random numbers will be assigned to Group 1 (the cocoa pill group). Reading down each column (i.e., the first column subjects are coded 1, 2, ..., 10), the subjects assigned to Group 1 are:

Campanella (5)	Herman (14)	Hornsby (15)	Dean (7)	Newhouser (29)
Palmer (32)	Sisler (37)	Vance (44)	Spahn (39)	Feller (9)
Ruffing (34)	Duffy (8)	Greenberg (13)	Lemon (23)	Traynor (43)
Young (50)	Foxx (10)	Paige (31)	Hubbell (16)	Tinker (42)
Snider (38)	Waner (46)	Lajoie (22)	Seaver (36)	Williams (47)

5.28. There are two studies being described.

- The first study is an observational study with matching. Subjects who exercise regularly are matched with subjects who don't.
- The second study is a comparative experiment where subjects will be randomized to the two groups:
  - Group 1: Regular exercise
  - Group 2: Usual habits.

Randomization is our best weapon to create two groups of subjects who are "similar on average." This gives the experiment an advantage. However, in the first study, if researchers "match" each subject who exercises with "a similar" subject who doesn't, this is also an attempt to maintain balance between the two groups. I guess it depends how good the matching is. Both studies have their own merit.

If the researchers performed the first study without matching, then clearly the second study (an experiment) would be preferred. The first study could have subjects in the two groups who are highly dissimilar on average. The differences between the groups would then be confounded with the effect of regular exercise when assessing bone loss reduction.