This assignment includes problems related to paired designs and analysis of categorical data.

1. Exercises 8.2 and 8.3 (page 319).
**Women with silicone breast implants are more likely than those without to smoke, be heavy drinkers, use hair dye, and have had an abortion. Use the language of statistics to explain why this casts doubt on a hypothesis that silicone breast implants cause illness.**

   Solution: Factors such as smoking, drinking, using hair dye, and having an abortion are confounded with the having breast implants as an explanation for illness. Women with implants may be different than women without implants with respect to illness, but the confounding variables make it difficult to claim that the implants are the cause of the increased illness.

   In the previous problem the explanatory variable is whether or not a woman has had breast implants, the response variable is a measure of illness, and the observational units are women. Extraneous variables include smoking status, drinking status, whether or not a woman dyes her her, and whether or not a woman has had an abortion.

2. Exercise 8.5 (page 319).
**U.S.-born white women and African-born black women now living in the United States have similar percentages of low birthweight babies, but U.S.-born black women have a higher percentage of low birthweight babies. Use the language of statistics to discuss the relationship among low birthweight, race, and mother's birthplace.**

   Solution: The association between low birthweight and race for is apparently spurious — if race were a causal factor, we would expect to see elevated percentages of low birthweight babies for African-born black women as well. Other confounding factors (such as socioeconomic level, access to prenatal care, etc.) may be different between U.S.-born white and black women and might be the true cause of the difference in percentage of low birthweight.

3. Exercise 8.7 (page 328).
**A 1976 research paper showed that terminal cancer patients given vitamin C survived longer than did historical controls, but surgeons selected which patients were given vitamin C. Randomized, controlled experiments have shown that vitamin C is not effective. Explain how the 1976 study was biased.**

   Solution: because the groups were not randomized, surgeons may have tended to give vitamin C to healthier patients and not to patients for which they thought no treatment could be effective.

4. Write functions in R (perhaps by editing the example in file `allocate.R`) that will allocate subjects for an experiment as in Exercise 8.8 and 8.10 (page 328) and 8.18 (page 337).

   Solution: Here are possible solutions:

```
> allocate8.8 <- function() {
+     groups <- c(rep("Standard", 10), rep("New", 20))
+     group <- sample(groups)
+     subject <- 1:30
+     alloc <- data.frame(subject, group)
+     print("Standard:")
+     print(alloc$subject[alloc$group == "Standard"])
+     print("New:")
+     print(alloc$subject[alloc$group == "New"])
+     return(invisible(alloc))
+ }
> allocate8.8()

[1] "Standard:"
 [1]  1  4  7 10 18 19 20 24 28 29
[1] "New:"
 [1]  2  3  5  6  8  9 11 12 13 14 15 16 17 21 22 23 25 26 27 30

> allocate8.10 <- function() {
+     groups <- c(rep("A", 3), rep("B", 3), rep("C", 3), rep("D",
```

```
+         3), rep("E", 3))
+     group <- sample(groups)
+     subject <- 1:15
+     alloc <- data.frame(subject, group)
+     print("Group A:")
+     print(alloc$subject[alloc$group == "A"])
+     print("Group B:")
+     print(alloc$subject[alloc$group == "B"])
+     print("Group C:")
+     print(alloc$subject[alloc$group == "C"])
+     print("Group D:")
+     print(alloc$subject[alloc$group == "D"])
+     print("Group E:")
+     print(alloc$subject[alloc$group == "E"])
+     return(invisible(alloc))
+ }
> allocate8.10()

[1] "Group A:"
[1]  1  5 15
[1] "Group B:"
[1]  2  9 14
[1] "Group C:"
[1]  6  8 13
[1] "Group D:"
[1]  4 11 12
[1] "Group E:"
[1]  3  7 10

> allocate8.18 <- function() {
+     heifers <- c(rep("Method 1", 4), rep("Method 2", 4))
+     young <- c(rep("Method 1", 4), rep("Method 2", 4))
+     mature <- c(rep("Method 1", 5), rep("Method 2", 5))
+     group <- c(sample(heifers), sample(young), sample(mature))
+     subject <- c(paste("heifer", 1:8), paste("young", 1:8), paste("mature",
+         1:10))
+     alloc <- data.frame(subject, group)
+     print(alloc)
+     return(invisible(alloc))
+ }
> allocate8.18()

     subject    group
1   heifer 1 Method 1
2   heifer 2 Method 1
3   heifer 3 Method 2
4   heifer 4 Method 2
5   heifer 5 Method 2
6   heifer 6 Method 2
7   heifer 7 Method 1
8   heifer 8 Method 1
9    young 1 Method 2
10   young 2 Method 2
11   young 3 Method 1
12   young 4 Method 2
```

```
13   young 5 Method 2
14   young 6 Method 1
15   young 7 Method 1
16   young 8 Method 1
17  mature 1 Method 2
18  mature 2 Method 2
19  mature 3 Method 1
20  mature 4 Method 1
21  mature 5 Method 2
22  mature 6 Method 1
23  mature 7 Method 1
24  mature 8 Method 2
25  mature 9 Method 1
26 mature 10 Method 2
```

5. Exercise 8.12 (page 328).
**An acupuncturist evaluates the effectiveness of acupunture versus aspirin for treating headaches. Explain how lack of blinding biases the experiment.**

Solution: The acupuncturist believes that acupunture is effective, and this belief could cloud judgment while measuring the effectiveness in treating headaches.

6. Exercise 8.13 (page 335).
**Compare two different blocking schemes for a greenhouse experiment.**

Solution: The second scheme is better because locations within each block are similar to one another (in distance to the steam pipe heat source). Under the second blocking scheme, each fertilizer will be used for exactly two plants that are close to the steam pipe, two that are in the middle,and two that are farthest away. In the first scheme it is possible, just by chance, that one fertilizer will have all of its plants close to the heat source while another has all its plants far away.

Here is a possible random allocation in R.

```
> treatment <- matrix(0, 3, 12)
> treatment[1, 1:6] <- sample(6)
> treatment[2, 1:6] <- sample(6)
> treatment[3, 1:6] <- sample(6)
> treatment[1, 7:12] <- sample(6)
> treatment[2, 7:12] <- sample(6)
> treatment[3, 7:12] <- sample(6)
> treatment

     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12]
[1,]    5    6    2    1    3    4    4    2    5     6     1     3
[2,]    6    2    5    3    1    4    5    3    4     6     2     1
[3,]    4    6    5    2    3    1    5    4    6     1     3     2
```

7. Exercises 8.16 and 8.17 (pages 336–337).
**Exercise 8.16: In an experiment 18 rats are to be allocated into treatment groups $T_1$, $T_2$, and $T_3$. Each rat is in its own cage. There will be six cages on each of three tiers. Lighting is expected to be important, and the three tiers of cages receive different amounts of light. Discuss three allocation schemes for cage locations.**

Solution: From best to worst, the plans are:

(a) Plan III. By blocking for a suspected important extraneous factor, the groups will be similar in terms of lighting.

(b) Plan I. By complete randomization, any factor (known or unknown) is likely to be fairly balanced. However, the small sample size could have resulted in imbalance for the known important factor of lighting.

(c) Plan II. The worst design. Lighting effects would be completely confounded with treatment effects.

**Exercise 8.17: Investigator will measure yield in an agricultural field experiment. Harvesting will take time, and more growth could occur (especially after rain) in the last plots to be harvested. Pick the best harvesting plan.**

Solution:

(a) Plan II. By harvesting all plots of block one first, then block 2, and so on, the conditions for the plots within a block are kept as similar as possible.

(b) Plan I. Harvesting all of variety 1 first, then variety 2, and so on is a bad plan because conditions within a block may be different and the time of harvest would be confounded with variety of corn, the condition of most interest.

8. Exercise 8.31 (page 347).
**Researchers wish to compare two drugs and a placebo for their effect on forced expiratory volume among individuals who have exercise-induced asthma. There are 30 subjects available.**

Solution:

(a) **Should this be an experiment or an observational study?** This should be an experiment where treatment groups are randomly assigned.

(b) **What is the placebo effect in this context?** The placebo effect in this context is the effect that taking a nonactive ingredient in an inhaler has on the patients.

(c) **Briefly describe how to set up a randomized block design.** There are several possibilities. One is to take baseline FEV measurements and use these to partititon the individuals into ten groups of three. Then, each individual in a group of three would be randomly assigned to a treatment.