620-270 Applied Statistics

Problem Set 2 (tutorial 2: s2/02)

In Tutorial 2, you should do the questions on Power (Questions 5 to 8) from Problem Sheet 1.

In addition, the following four questions revise the basic principles of study design.

2.1 For each of the experimental situations described below,

i. identify the experimental units, the explanatory variable(s) (factors), and the response variable;
ii. determine whether each variable is numerical, categorical, continuous, discrete, nominal, ordinal;
iii. determine whether the study is designed or observational.

(a) Can aspirin help heart attacks? The Physician’s Health Study, a large medical experiment involving 22000 male physicians, attempted to answer this question. One group of 11000 physicians took an aspirin every second day, while the rest took a placebo. After several years it was found that the subjects in the aspirin group has significantly fewer heart attacks than subjects in the placebo group.

(b) New varieties of corn with altered amino acid patterns may have higher nutritive value than standard corn, which is low in the amino acid lysine. An experiment compares two new varieties, called opaque-2 and floury-2, with normal corn. Corn-soybean meal diets using each variety of corn are prepared at three different protein levels: 12%, 16% and 20%. There are thus nine diets in all. Researchers assign 10 one-day-old male chicks to each diet and record their weight gain after 21 days. The weight gain of the chicks is a measure of the nutritive value of their diet.

2.2 (From Devore and Peck, 2.25) An article in the San Luis Obispo Tribune described an experiment designed to investigate the effect of creatine supplements on the development of muscle fibers. The article states that the researchers

... looked at 19 men, all about 25 years of age and similar in weight, lean body mass and capacity to lift weights. Ten were given creatine — 25 grams a day for the first week, followed by 5 grams a day for the rest of the study. The rest were given a fake preparation. No one was told what they were getting. All the men worked out under the guidance of the same trainer.

The response variable measured was gain in fat–free mass (in percent of body mass).

(a) What confounding variables are identified in the statement above, and what strategy did the researchers use to deal with them?
(b) Was this a balanced design? Explain.
(c) This was a blind study, i.e. the men participating in the experiment were not told whether they were receiving creatine or the placebo? Do you think this was important? Explain.
(d) How could this experiment have been run as a double-blind study? Do you think it would have been a good idea to have made this a double-blind experiment? Explain.
2.3 As part of a study investigating the effect of smoking on infant birthweight a physician examines the records of 40 nonsmoking mothers, 40 light-smoking, and 40 heavy-smoking mothers. The mean birthweights (in kg) for the three groups are respectively 3.43, 3.29 and 3.21.

(a) What are the independent and dependent variables? What are the experimental units?

(b) Is this a designed experiment or an observational study? Explain your choice.

(c) What are the potential confounding variables in this case? Explain how you would eliminate the effect of at least some of the variables.

(d) Has replication been used in this study? Explain, stating its purpose.

2.4 (From Devore and Peck, 2.32) The level of lead in children’s blood should be kept low, and one way to reduce a child’s exposure to lead is to control dust in the home.

Researchers wished to study the effects of families’ dust control practices in the home on children with low to mild elevations in blood lead levels. They identified 100 families whose children had elevated lead levels and randomly assigned them to two groups. The families in one group were given a pamphlet on how to avoid lead exposure, cleaning supplies and a demonstration of proper cleaning techniques. The families in the other group were given only the pamphlet on how to avoid lead exposure. The researchers measured the lead levels in the children’s blood at the beginning of the study and again seven months later.

(a) What is the population for this study?

(b) What are the experimental units?

(c) Is this an observational study or a designed experiment? Explain briefly.

(d) Give the explanatory variable and the response variable and state whether each variable is categorical or numerical.

(e) Discuss the role and importance of randomisation in this study.

(f) Was a control group used in this study? Discuss the purpose of using a control group.

(g) Discuss possible confounding variables and explain how they could be controlled.

(h) Discuss whether and how blocking should be used.