

STAT 261 Winter 2007

Problem Set 11

To be done on your own

1) The Gallup poll asks respondents how they would rate the honesty and ethical standards of people in different fields — very high, high, average, low, or very low. In 1992, 52% of the respondents gave medical doctors a rating of “very high” or “high”; only 50% rated college teachers that way. Assume the results are based on a single simple random sample of 1000 persons; each respondent rated clergymen, medical doctors, college teachers, and other professions.

Is the difference real, or a chance variation? Or do you need more information? Discuss briefly.

2) In the 1950s, the great statistician R. A. Fisher challenged every study suggesting that tobacco adversely affected the smoker’s health. Now, Fisher himself was a heavy smoker, and he was being paid by the tobacco industry, but he did point out some significant methodological weaknesses. He emphasized that the clear *association* between smoking and ill health did not imply *causation*. In particular, he advanced the “constitutional hypothesis”: Something in the genetic makeup of individuals predisposes them to certain diseases, and makes them more likely to take up smoking.

Researchers in Finland challenged this hypothesis by studying twins. They identified pairs of smoking-discordant monozygotic twins. (“Smoking-discordant” means that one of the twins smokes, and the other doesn’t; “monozygotic” means that they are genetically identical.) Then they looked to see which of the twins would die first.

During the 12-year course of the study, there were 22 pairs of which at least one twin died. The following table records the number of cases in which the first to die was the smoker or the nonsmoker, split by cause of death.

	Smokers	Nonsmokers
All causes	17	5
Coronary heart disease	9	0
Lung cancer	2	0

According to the first line, there were 17 pairs in which the smoking twin died first; 5 in which the nonsmoking twin died first. There were 9 pairs where the first twin to die died from CHD; all of these were smokers. There were 2 pairs where the first to die died of lung cancer. These were both smokers.

a) Analyse each row of the table. Frame a null hypothesis and an alternative, and test whether the difference between smokers and non-smokers could be due to chance.

b) Why is it significant that the twins were identified while still alive, and were then followed to see which ones died (a “prospective study”) rather than simply identifying pairs of smoking-discordant twins of which one member had already died (a “retrospective study”), which would have been easier and quicker?

c) If you were Fisher, what objections might you still raise?

(source: Freedman, Pisani, Purves, Statistics, p. 262–3.)