M. Mohammadian

Statistical Data Analysis

Confidence Interval Estimation

1) Airport. The International Air Transport Association surveys business travellers to develop quality ratings for international airports. The maximum possible rating is ten. Suppose a simple random sample of business travellers is selected and each traveller is asked to provide a rating for Berlin-Brandenburg BER Airport. The ratings obtained from the sample of 50 business travellers follow. (Generate the dataset by:

```
> set.seed(56)
> BER <- rbinom(50,10,0.6)
> BER
[1] 6  5  7  5  7  8  7 …)
```

Construct a 95 per cent confidence interval estimate of the population mean rating for BER-Airport.

2) Alcoholic Beverages. Consumption of alcoholic beverages by young women of drinking age has been increasing in the UK, Europe and the US (The Wall Street journal, 15 February 2006). Data (annual consumption in litres) consistent with the findings reported in The Wall Street Journal article are shown for a sample of 20 European young women. (Generate the dataset by:

```
> set.seed(15)
> alc.bev <-  round(rnorm(20,130,65))
> alc.bev
 [1] 147  249  108  188  162  …)
```

Assuming the population is roughly symmetrically distributed, construct a 90 per cent confidence interval for the mean annual consumption of alcoholic beverages by young European women.

3) Medical Expenses. Public health agencies and insurance companies invest a great deal of time and money in developing models that accurately forecast medical expenses. Use the simulated dataset `insurance.csv` that contains hypothetical medical expenses for patients in the United States.

a) Form and plot the five-point summary and a histogram for the variable `charges`. (The distribution is not symmetrical!)
b) Provide a point estimate of the population mean of medical charges
c) What is the 95 per cent confidence interval estimate of the population mean?
d) At 90 per cent confidence, what is the margin of error?

4) Profit Estimates. According to a BusinessWeek report in early 2006, data from Thomson Financial showed that the majority of companies reporting profits had beaten estimates. A sample of 162 companies showed 104 beat estimates, 29 matched estimates, and 29 fell short.

a) What is the point estimate of the proportion that fell short of estimates?
b) Provide a 95 per cent confidence interval for the proportion that beat estimates.

5) Independent Palestinian State. In early December 2008, the Palestinian Centre for Policy and Survey Research carried out an opinion poll among adults in the West Bank and Gaza Strip. Respondents were asked their opinion about the chance of an independent Palestinian state being established alongside Israel in the next five years. Among the 1270 respondents, 34.6 per cent felt there was no chance of this happening.

a) Provide a 95 per cent confidence interval for the population proportion of adults who thought there was no chance of an independent Palestinian state being established alongside Israel in the next five years.
b) Provide a 99 per cent confidence interval for the population proportion of adults who thought there was no chance of an independent Palestinian state being established alongside Israel in the next five years.
c) What happens to the margin of error as the confidence is increased from 95 per cent to 99 per cent?

6) Euro. In a survey conducted by ICM Research in the UK towards the end of December 2008, 710 out of 1000 adults interviewed said that, if there were to be a referendum, they would vote for the UK not to join the European currency (the Euro €). What is the interval estimate of the population proportion of British adults who would vote for the UK not to join the European currency? Use 90 per cent and 95 per cent confidence.

7) Lifetime of Heat Pumps. A random sample of n = 20 heat pumps of a certain type yielded the following observations on lifetime in years: Generate sample data by

```
> set.seed(10)
> lifetime <-  round(rnorm(20,4,2),1)
> lifetime
[1] 4.0  3.6  1.3  2.8 ...
```

a) Assume that the lifetime distribution is normal and calculate a 95% CI for expected (true average) lifetime of all heat pumps of that type.
b) How should the interval of part (a) be altered to achieve a confidence level of 99%?