Analytical Techniques 2: Practical

Set up

For this practical you need to download one Stata dataset (**sbp.dta**) and two program files (**cidiffni.ado** and **cishowi.ado**) from the AT Moodle page. Create a folder called **AT** on your computer and save the three files in this folder.

Exercise 2.1 Confidence Intervals for a Population Proportion

In a publication reporting results from a randomised controlled trial comparing the effect of a single dose of the drug prednisolone with placebo for treating children in hospital with acute asthma, the authors began their results section with the following statement:

> 3 out of 51 patients in the placebo group (5.9%, 95% CI –0.6% to 12.3%) and 20 out of 57 in the prednisolone group (35.1%, 95% CI 22.7% to 47.5%) were discharged at first examination.

(a) By hand, construct an ***approximate*** 95% CI for the percentage of children discharged at first examination in the ***prednisolone*** group.

(b) What is wrong with the published confidence interval for the percentage in the placebo group? Why do you think has this happened?

(c) Use the immediate Stata command **`cii`** to calculate 95% confidence intervals for the placebo and prednisolone groups separately using exact methods.

```
cii proportions 51 3
cii proportions 57 20
```

You can type **`help cii`** to get more information of the syntax for **`cii`**.

(d) Compare these to those presented above. Comment on your findings.

Exercise 2.2 Confidence Intervals for a Population Mean

In a study comparing two treatments for mildly elevated blood pressure, 17 patients were given treatment A for two weeks and had their systolic blood pressure (SBP) measured at the end of this period. For the following two weeks the patients were given no treatment. Finally, the patients were given treatment B for two weeks and had their SBP measured at the end of that period. The data are stored in the Stata dataset **`sbp.dta`**. Change the working directory to your AT folder and load the dataset.

(a) Use the **`summarize`** command to obtain the sample mean and standard deviation of SBP under treatment A (sbp_a). **By hand**, calculate a 95% CI for the population mean μ_A from which this sample is drawn. [Note: you can obtain the critical value from statistical tables or using **`display invt(16, 0.975)`**]. What are the assumptions that you are implicitly making to construct this confidence interval?

Use the command **`cii`** to confirm the accuracy of your calculation.

```
cii means 17 144.06 12.44
```

(b) Assuming that the population variance is known to be 100 $mmHg^2$, recalculate ***by hand*** the 95% CI for μ_A.

(c) Which of the two confidence intervals calculated in (a) and (b) above is wider? Explain why.

Exercise 2.3 Developing Understanding of Confidence Intervals

In the following exercises we will use the two user written Stata commands downloaded from Moodle to develop our understanding of confidence intervals.

First, we will use the command **`cidiffni`** to explore how the confidence interval changes with changes in the sample size. We will use the mean (144.06) and SD (12.44) from the previous exercise and then pretend they were derived from samples of different sizes: 10, 30, 50, 70, 90 and 110. Before running the command, how do you expect the sample size to affect the 95% CI for the population mean μ_A? Type the following command to explore this:

```
cidiffni 144.06 12.44, n(10,110,20) unknown level(95) graph
      name(g1, replace)
```

Note: the option **`n(10,110,20)`** means from 10 to 110 in steps of 20, **`level(95)`** means calculate 95% CIs, **`unknown`** means assume that the population variance is unknown, **`graph`** requests the CIs are shown in a graph as well as the table, and **`name(g1, replace)`** will allow you to keep the graph window open for comparison when you execute the next command.

(a) What happens to the intervals and why?

(b) In the above command the population variance was assumed unknown. Retype the above command removing the **`unknown`** option and change **`g1`** to **`g2`** in the **`name`** option. How do the intervals change?

Second, we will use the command **`cishowi`** to emphasise the general interpretation of confidence intervals. We will start by simulating drawing 50 samples of size 10 from a normal distribution with true population mean 150 and true population variance 100. The command calculates the confidence interval for each sample and plots them.

```
cishowi 10 150 100, repeat(50) level(95) graph name(g3, replace)
```

Note: **`10`** is the sample size, **`150`** is the population mean and **`100`** is the population variance. The option **`repeat(50)`** means to draw 50 different samples from the population.

(c) What proportion of the intervals include the population mean of 150? Do these intervals estimate the variance or do they assume that it is known? Try adding the option **`unknown`** and change the graph name to **`g4`** so that you can compare them.

If you have time repeat the calculation with more simulations (100, 500 or 1000) and try different confidence levels e.g. 90 or 99.

Exercise 2.4 Confidence Interval for a Population Variance

Continuing with the SBP dataset we will now calculate a 95% CI for the population variance (see case study 3 from lecture) of SBP under treatment A (sbp_a).

(a) Use the sample standard deviation to calculate ***by hand***, a 95% CI for the population variance σ_A^2 from which this sample is drawn. Find the relevant chi-square critical values using Neave's tables. Once you have found them you can check them in Stata, e.g:

```
display invchi2(16,0.025)
display invchi2(16,0.975)
```

(b) Use the Stata command **cii** to confirm the accuracy of your calculations.

```
cii variance 17 154.68
```

Exercise 2.5 Confidence Interval for a Population Rate

See case study 5 from lecture. During a 12-month survey 19 deaths due to sepsis were observed at a district general hospital.

(a) Use the normal approximation to calculate ***by hand***, a 95% CI for the monthly ***rate*** of death from sepsis. What assumptions are you making in your construction of the CI?

(b) Use the Stata command **cii** to calculate an exact 95% CI for the rate.

```
cii means 12 19, poisson
```

How does this compare to your answer to part (a)?

The survey was repeated over a 2-year period at 5 randomly selected district general hospitals. A total of 178 deaths due to sepsis were observed.

(c) Use the normal approximation to calculate ***by hand***, a 95% CI for the monthly ***rate*** of death from sepsis i.e. rate per hospital per month. What additional assumptions are you making here?

(d) Use the Stata command **`cii`** to calculate an exact 95% CI for the rate. How does this compare to your answer to part (c)?