

Part 0 – Open Stata, and make your own do-file

- Using windows explorer, make a new folder called `H:\rproject\clab\clab22`.
- Download the datafile `upop.dta` and do-file `progs.do` from **Bb** and put them in this folder.
- Start Stata through the start menu button
- In the white command window type `doedit` to start the do-file editor. Place the Stata screen on the left and the do file editor on the right such that you can easily switch between the two.
- In the first two lines of the do-file type


```
cls                                //this clears the screen
clear all                          //this clears the memory
cd "H:\rproject\clab\clab22"      //this is your path
set seed 123                       //random number init
```
- Save the do-file as `clab22.do`.
- In the following, as always, don't forget to apply Rules 1&2: paste the command in your `clab22.do` file, press `control-s` to save and `control-d` to run.

Part 1 – Causal effects in a sample

In part 1 of these exercises you are going to go back to the population data of unemployed people and their potential outcomes in terms of employment in 8 months depending on job seek assistance (JSA). In Clab 2.1 you showed by simulation that mean employment of a sample of (approx.) $n=500$ people is normally distributed. Also, you showed that the *t*-statistic of the mean was standard normally distributed, and tested two hypothesis ($\mathcal{H}_0: E[Y_i(0)] = 0$ and 0.30) with respect to employment for people without JSA.

Testing hypothesis about a population mean can be descriptively important. For example, policy makers are often interested in the average level of employment in the population. Most social science revolves around learning about causal effects. In our context, does JSA increase employment? In terms of potential outcomes this is the hypothesis $\mathcal{H}_0: E[Y_i(0)] = E[Y_i(1)]$. If JSA has no effect, that must mean that the average employment after 8 months is the same whether or not every unemployed person got JSA. A short way of writing this is $\mathcal{H}_0: \mu_0 = \mu_1$. In this part you will test this hypothesis using two samples of unemployed people, those with JSA ($x=1$) and those without ($x=0$).

1. Open the file `upop_rdraw_20k.dta` and then type

```
run progs.do
random_sample 1000
```

The first of these commands runs the do-file `progs.do` which contains some programs that you will use later (open it to check out the code). The second command draws a random sample of 1000 people from the population of unemployed, as you did last week.

2. Type

```
sum y if x==0
sum y if x==1
```

What are these summary statistics? Use them to test $\mathcal{H}_0: \mu_0 = \mu_1$ vs. $\mathcal{H}_1: \mu_0 \neq \mu_1$ using a t-test. Consult the slides if necessary. Use three digits in your calculations, and two

digits for the final answer. Throughout this course we always use a significance level $\alpha = 5\%$. What do you conclude?

3. The t-statistic that you have obtained can be put in a graph using a program defined in `progs.do`. Type

```
critvalplot, tstat(X) name(tstat)
```

where X is the t-statistic that you have obtained in question 2. Discuss what you see.

4. Of course Stata also has a direct command to do a mean comparison t-test:

```
ttest y, by(x) unequal
```

Your t-statistic may be slightly different, why? Look through the table to see if you understand everything.

5. The table also presents a so called “p-value”. Type

```
pvalplot, tstat(2.58) name(pval)
```

Use the graph to explain what a p-value is and how it can be used to reject \mathcal{H}_0 or not. The p-value in the graph is slightly smaller than in the table of question 4. Can you explain why? (Hint: look at the slide with remarks about the t-test)

Part 2 – Confidence intervals and precision

Another way of thinking about the t-test is to ask the question: “What hypotheses would not be rejected by the t-test?” This is the 95% confidence interval (CI). The data would not reject any of these values a null hypothesis. The formula for the 95% CI is of the estimated average treatment effect \widehat{ATE} is

$$95\% \text{ CI: } [\widehat{ATE} - 1.96 \cdot SE(\widehat{ATE}); \widehat{ATE} + 1.96 \cdot SE(\widehat{ATE})]$$

Confidence intervals are useful because they give you an idea how precise an estimate is. In the example above, for example, we have found $\widehat{ATE} = \bar{Y}_1 - \bar{Y}_0 = 0.077$ meaning employment is estimated to increase by 7.7%-points due to JSA. Still, if this estimate is very noisy, the true effect may be much higher(lower) than this. The CI gives you a range of values that you can expect the true parameter to be in. The narrower the CI is, the higher our confidence is that the population ATE is close to the \widehat{ATE} that you have found.¹

6. Re-type

```
ttest y, by(x) unequal
```

from question 4 to figure out what the 95% confidence interval is of the difference in means. Is it likely that the true effect is 0.20? Would an effect of 0.02 be rejected? If JSA is only worth the cost if it reduces unemployment by 4%-points or more, does this sample produce conclusive evidence?

7. To illustrate more what precision means, copy the following in your do-file editor and run it:

```
twayway (function y = normalden((x-0.303)/0.0220), range(0 1) lc(blue)) ///
  (function y = normalden((x-0.381)/0.0222), range(0 1) lc(red)), ///
  legend(order(1 "JSA=0" 2 "JSA=1")) xlabel(0 0.303 0.381 1) ///
  xline(0.303 0.381, lp(dash)) xtitle(mean employment) ///
  ytitle(density) name(diff)
```

Explain what you see.

¹ Strictly speaking this interpretations is only correct under a so called “Bayesian” approach, but in practice this is how CI’s are usually interpreted.



8. Now assume that your sample does not contain $n = 1000$, but just $n = 500$. The mean and standard deviations remain the same. Repeat the previous question with this smaller sample size. What will the standard errors now become? Use `name(diff2)` to name the graph. Is the estimated $\widehat{ATE} = 0.077$ still significant with this sample size? Compare both graphs.