

Examination Guidelines HEVAL5140 –
Methods for effectiveness evaluations
in health care
Fall 2021

Henning Øien

Instructions

This open book home-exam is to be solved independently. An important part of the assessment is to evaluate the students' ability to answer the questions by applying and discussing the material using their *own* words. Students who are well prepared, and understand the material, should be able to write good answers to many of the questions without consulting the lecture notes or other course material.

The students may use inline references (author and year of publication/lecture or slide number) when referring to sources from the course material. If they refer to sources not included in the curriculum, a reference list that includes the author name, year of publication, and name of journal should be included. When students copy formulas or graphs from sources, they need to refer to author, page/slide number and equation/figure number. They should only use the course material and published articles in peer-reviewed journals as references. The students were told in general to *not* use direct quotations. The assignment should be in their own words. If the students found it essential to use direct quotations, quotation marks should be used, as well as a reference to page and paragraph number of the source.

Answers that are simply reproductions of the course material, without any interpretation or explanation, receive zero points. Answers that use direct quotations without proper referencing receive zero points.

The answer guidelines (this document) are written in *R*, but the students can use the statistical software and text editor of their choice. The only requirements are that the code should be included as an appendix, and that the assignment should be handed in as a single pdf. Failure to adhere to these requirements can result in loss of points.

Please, see below a non-exhaustive list of points that should be addressed in good answers.

Exercise 1: Potential outcomes (20%)

Consider the simple hypothetical example in Table 1. This example involves eleven patients each of whom is infected with coronavirus. There are two treatments: ventilators and bedrest. Table 1 displays each patient's potential outcomes in terms of years of post-treatment survival under each treatment. Larger outcome values correspond to better health outcomes.

Table 1: Potential outcomes for eleven patients receiving ventilators Y^1 or bedrest Y^0

Patient	Y^1	Y^0	Age
1	1.5	13.0	29
2	1.5	6.5	35
3	1.5	5.2	19
4	7.5	7.8	45
5	7.5	1.3	65
6	9.0	9.1	50
7	10.5	10.4	77
8	10.5	13.0	18
9	12.0	2.6	85
10	13.5	7.8	96
11	15.0	9.1	77

1.0. Calculate each patient's treatment effect.

Guide to answer:

```
library(dplyr)
# Make the perfect doctor data:
df <- data.frame("patient_i" = 1:11,
                 "y1i"       = c(1,1,1,5,5,6,7,7,8,9,10)*1.5,
                 "y0i"       = c(10,5,4,6,1,7,8,10,2,6,7)*1.3,
                 "age"       = c(29, 35, 19, 45, 65, 50, 77, 18, 85, 96, 77))
df <- df %>% mutate(delta_i = y1i - y0i)
df
```

```
#>   patient_i y1i y0i age delta_i
#> 1         1  1.5 13.0 29   -11.5
#> 2         2  1.5  6.5 35    -5.0
#> 3         3  1.5  5.2 19    -3.7
#> 4         4  7.5  7.8 45    -0.3
#> 5         5  7.5  1.3 65     6.2
#> 6         6  9.0  9.1 50    -0.1
#> 7         7 10.5 10.4 77     0.1
#> 8         8 10.5 13.0 18    -2.5
#> 9         9 12.0  2.6 85     9.4
#> 10        10 13.5  7.8 96     5.7
#> 11        11 15.0  9.1 77     5.9
```

The individual treatment effect is found by subtracting $Y_i^1 - Y_i^0$ for each patient.

1.1. What is the average treatment effect for ventilators compared to bedrest? Which type of intervention is more effective on average?

Guide to answer:

```
library(dplyr)
# Calculate ate
ate <- mean(df$y1i) - mean(df$y0i)
ate
```

```
#> [1] 0.3818182
```

The *ate* is positive. Ventilators are more effective on average.

1.2. Suppose the "perfect doctor" knows each patient's potential outcomes and as a result chooses the best treatment for each patient. Fill in the treatment assignment and the observed outcome based on what the perfect doctor chooses.

Guide to answer:

```
library(dplyr)
# Calculate perfect doctor treatment, d_i = 1 if treatment is positive, zero otherwise
df <- df %>% mutate(d_i = if_else(delta_i > 0, 1, 0))
# Fill in observed y
df <- df %>% mutate(y = if_else(d_i == 1, y1i, y0i))
df

#>   patient_i y1i y0i age delta_i d_i y
#> 1         1  1.5 13.0 29  -11.5  0 13.0
#> 2         2  1.5  6.5 35   -5.0  0  6.5
#> 3         3  1.5  5.2 19   -3.7  0  5.2
#> 4         4  7.5  7.8 45   -0.3  0  7.8
#> 5         5  7.5  1.3 65    6.2  1  7.5
#> 6         6  9.0  9.1 50   -0.1  0  9.1
#> 7         7 10.5 10.4 77    0.1  1 10.5
#> 8         8 10.5 13.0 18   -2.5  0 13.0
#> 9         9 12.0  2.6 85    9.4  1 12.0
#> 10        10 13.5  7.8 96    5.7  1 13.5
#> 11        11 15.0  9.1 77    5.9  1 15.0
```

1.3. Calculate the simple difference in outcomes (SDO). How similar is it to the ATE?

Guide to answer:

```
sdo <- mean(df$y[df$d_i == 1]) - mean(df$y[df$d_i == 0])
sdo
```

```
#> [1] 2.6
```

The SDO is positive, but larger than the ATE. The treated live on average longer than the untreated.

1.4. Calculate the average treatment effect on the treated (ATT) and untreated (ATU). How similar are each of these to the SDO? How similar are each of these to the ATE?

Guide to answer:

```
# average treatment effect on the treated.
att <- mean(df$y1i[df$d_i == 1]) - mean(df$y0i[df$d_i == 1])
att
```

```
#> [1] 5.46
```

The treated gain roughly 5.5 years.

```
# average treatment effect on the untreated.
atu <- mean(df$y1i[df$d_i == 0]) - mean(df$y0i[df$d_i == 0])
atu
```

```
#> [1] -3.85
```

The untreated lose 3.85 years if they had received the treatment.

1.5. Show that the SDO is numerically equal to the sum of ATE, selection bias and heterogeneous treatment effects bias. You will need to calculate the ATE, selection bias and heterogeneous treatment effects bias, combine them in the appropriate way, and show that their sum is equivalent to the SDO.

Guide to answer:

```
# calculate selection bias
sb <- mean(df$y0i[df$d_i == 1]) - mean(df$y0i[df$d_i == 0])
sb
```

```
#> [1] -2.86
```

```
# Calculate the hteb
hteb <- (1- mean(df$d_i))*(att - atu)
hteb
```

```
#> [1] 5.078182
```

```
ate + sb + hteb
```

```
#> [1] 2.6
```

1.6. Create a dataset based on the perfect doctor treatment assignment from part (1). This dataset should only contain the treatment assignment (D), age (Age) and the observed outcome (Y). Then estimate the following equation:

$$Y_i = \alpha + \delta D_i + \epsilon_i$$

What is the estimated δ identical to?

Guide to answer:

```
df.obs <- df %>% select(patient_i, d_i, y, age)
lm(y ~ d_i, data = df.obs)
```

```
#>
#> Call:
#> lm(formula = y ~ d_i, data = df.obs)
#>
#> Coefficients:
#> (Intercept)          d_i
#>          9.1           2.6
```

- The students should recognize that this is simply a comparison of means in treatment and control group, and therefore simply the SDO.

1.7. Now run the following multivariate regression controlling for age.

$$y_i = \alpha + \delta D_i + \beta \text{age}_i + \epsilon_i$$

Report the coefficient on δ . Is it equal to ATE, SDO, ATT or ATU? Did controlling for age recover the ATE?

Guided answer:

```
lm(y ~ d_i + age, data = df.obs)

#>
#> Call:
#> lm(formula = y ~ d_i + age, data = df.obs)
#>
#> Coefficients:
#> (Intercept)          d_i          age
#>      8.00825       1.01808       0.03342
```

- The estimate is closer to the ATE than the SDO. This is because age is positively correlated with treatment and negatively correlated with the outcome.
- Bonus if students discuss age as a confounding variable.

1.8. Estimate two regression equations: The first equation is an auxiliary regression of D onto Age . The second equation regresses Y onto $\hat{D}$ which is the residual from the first auxiliary equation. Compare the coefficient on D from the multivariate regression in 1.9, to the coefficient on $\hat{D}$ in the second equation. What does this tell you about how to interpret multivariate regressions? Which theorem have you illustrated with this exercise?

Answer:

```
m1 <- lm(d_i ~ age, data = df.obs)
res_m1 <- residuals(m1)

m2 <- lm(y ~ res_m1, data = df.obs)
m2

#>
#> Call:
#> lm(formula = y ~ res_m1, data = df.obs)
#>
#> Coefficients:
#> (Intercept)      res_m1
#>      10.282       1.018
```

This illustrates the "regression anatomy theorem" section 2.24 in the textbook "Causal inference. The Mixtape". It says that a specific coefficient in a multivariate regression can be interpreted as the association between the specific covariate and the outcome after the correlation between the specific covariate and the other explanatory variables have been taken out.

Exercise 2 Instrumental Variables (20%)

2.0. Suppose we are interested in estimating the causal returns to years of schooling, and that the causal relationship can be described by the following model:

$$Y_i = \alpha + \delta S_i + \eta_i,$$

where Y is the log of earnings and S_i is years of schooling. What is the interpretation of δ ?

Guide to answer:

The students should recognize that this is a causal model, and make interpretations that are similar to these:

- δ is the causal effect of years of education on earnings, and is interpreted as the percentage change in earnings caused by one more year of schooling.
- In the model δ is specified to be the same for everyone (i.e. no subscript i) and therefore $\delta = ATE = ATT$ of years of schooling on earnings.

2.1. Assume that you have access to observational data on years of schooling and earnings. Would a regression of log earnings on years of schooling recover δ ? Why or why not?

Guide to answer:

- The students should mention selection bias.
- It will not recover the causal effect of schooling, because educated and less educated people are different in many respects, which are highly likely to be related to both schooling and earnings. Confounding variables.

2.2. Consider using instrumental variables to recover δ . What are the requirements for a valid instrument?

Guide to answer:

- The students should mention the relevance, independence and exclusion requirements.
- They can also mention SUTVA, but SUTVA was not discussed in detail in lectures.

2.3. In Section 7.7.1., page 352, in "Causal Inference: The Mixture" by Cunningham (2021) a study that uses geographic proximity to colleges as an instrument for S_i is discussed. Specifically, the instrument is a dummy that is equal to 1 if you live in a county with a college ("college-in-the-county dummy variable"). Discuss whether distance to college is a valid instrument for identifying δ .

Guide to answer:

- A valid instrument should be as-good-as randomly assigned, and therefore independently distributed of confounders.
- Whether you live close to a college is not randomly assigned, and therefore highly likely to be correlated with confounding variables. Proximity to college is not a valid instrument.

Exercise 3: The effect of sickness insurance reform in Norway (60%)

In this exercise, you are going to discuss and replicate some of the analysis in "Activation against absenteeism - Evidence from a sickness insurance reform in Norway" by Hernæs (2018). The data `sickness_absence_county.csv` is posted here:

Data:

3.0. Write a short report of the attached paper "Sickness insurance reform in Norway" by Hernæs (2018). The report should include a summary of the paper, and a critical discussion of the empirical approach. The summary should identify the research questions that the paper tries to answer, how the paper answers the questions, and the results (about 1 page). The discussion of the empirical approach should give a description and critical assessment of the applied methods and its identifying assumptions. Focus on the following questions: What are the coefficient(s) of interest(s)? What is (are) the key identifying assumption(s)? Are the identifying assumptions likely to hold? Are there data limitations, and do you have any suggestions for alternative analyses and sensitivity checks? The report can be less, but should be no longer than 3 pages (12pt, double spaced).

Guide to answer:

- This should be cohesive text written in the students' own words and arguments i.e. the text should not be a copy of the text and arguments from the article.
- In the first page, it is important that the student is able to effectively identify the papers' core argument: What is the research question and statement?
- It is important that the students are able to critically discuss the empirical approach. They should identify the key identifying assumptions, coefficients of interest, threats to the identifying assumption, data limitations, suggestions for alternative analyses.
- Especially the students should recognize the problem of estimating standard errors in a classical DiD regression in cases where there are only one treated unit, and few control units. And how Synthetic control provides a potential solution to this problem.

3.1. Compute the "before-after" estimator for the causal effect of the sickness insurance reform: Compute the average work hours lost due to sickness absence in percent (`sick_pct`) in Hedmark after and including the second quarter in 2013, before the second quarter in 2013, and take the difference. Do you think this gives a credible estimate of the causal effect of the sickness insurance reform?

Guide to answer:

```
hed.mean.before <- mean(d %>% filter(post == 0 & fnr == 4) %>% pull(sick_pct))
hed.mean.after  <- mean(d %>% filter(post == 1 & fnr == 4) %>% pull(sick_pct))
hed_after_before <- hed.mean.after-hed.mean.before
hed_after_before
```

```
#> [1] -1.168831
```

We find that average work hours lost due to sickness absence drops by roughly 1.17 percentage points.

It is not credible, because there could be other factors that have changed over time that affects sickness absence.

3.2. Do the calculation described in **3.1.** for the control counties (all counties, except Hedmark), and compute the 2×2 difference-in-differences estimator for the causal effect of the sickness insurance reform on `sick_pct`. When does the difference-in-differences estimator capture the causal effect of the sickness insurance reform?

Guide to answer:

```
con.mean.before <- mean(d %>% filter(post == 0 & fnr != 4) %>% pull(sick_pct))
con.mean.after  <- mean(d %>% filter(post == 1 & fnr != 4) %>% pull(sick_pct))
con_after_before <- con.mean.after - con.mean.before
con_after_before
```

```
#> [1] -0.5250573
```

```
dd_estimate <- hed_after_before - con_after_before
dd_estimate
```

```
#> [1] -0.6437739
```

- The students need to explain and discuss the parallel trend assumption.
- Good answers assess whether the parallel trend assumption is likely to hold in this case.

3.3. Use regression to compute the DD estimator. Is the DD estimate statistically significant? Discuss whether conventional standard errors are valid as an estimator of the uncertainty in the DD estimator.

Tip: Saturate your regression model by making a dummy for Hedmark, a dummy for the post period and an interaction between the Hedmark dummy and the post dummy.

Guide to answer:

```
d[, treat := (fnr == 4)]
d[, postXtreat := post*(fnr == 4)]
summary(lm(sick_pct ~ postXtreat + post + treat, data = d))

#>
#> Call:
#> lm(formula = sick_pct ~ postXtreat + post + treat, data = d)
#>
#> Residuals:
#>      Min       1Q   Median       3Q      Max
#> -1.91597 -0.45682 -0.01597  0.48403  2.18403
#>
#> Coefficients:
#>              Estimate Std. Error t value Pr(>|t|)
#> (Intercept)   6.11597     0.03763  162.526 < 2e-16 ***
#> postXtreat   -0.64377     0.27231  -2.364 0.018404 *
#> post         -0.52506     0.06418  -8.181 1.84e-15 ***
#> treatTRUE     0.59832     0.15965   3.748 0.000197 ***
#> ---
#> Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
#>
#> Residual standard error: 0.711 on 572 degrees of freedom
#> Multiple R-squared:  0.1423,    Adjusted R-squared:  0.1378
#> F-statistic: 31.64 on 3 and 572 DF,  p-value: < 2.2e-16
```

3.4. Plot the trend in average work hours lost due to sickness absence in Hedmark and the average work hours lost due to sickness absence in the other counties. Does it seem like the trends are similar in Hedmark and the other states prior to the intervention?

Tip: Make a new dataset with quarterly averages of `sick_pct` by the treatment dummy, and use this dataset to plot `sick_pct` by quarters.

Guide to answer:

```
plot.data <- d %>% group_by(treat, year_q) %>% summarize(sick_pct = mean(sick_pct))

ggplot(data = plot.data, aes(x = year_q, y = sick_pct, group = treat, color=treat)) +
  geom_vline(xintercept = "2013q1") +
  geom_line() +
  theme_bw()
```



There is a level difference, but the trends seem to be parallel, and, after the treatment quarter, the treatment and control group trends converge. Good answers include an explanation of why level differences do not invalidate a DD analysis.

3.5. Use the synthetic control method to estimate the effect of the reform on average work hours lost due to sickness absence (`sick_pct`). You need to decide which variables, and which pre-period, that should be used for matching. You can use exactly the same specification as is used in Hernæs (2018), but note that the results you get might be different because the optimization algorithm to find the weights is software dependent.

Make a plot, similar to Figure 1 in Hernæs, and assess whether the synthetic control region you found follows the Hedmark trend in the pre-treatment period. Comment on whether you think the synthetic control region produces a valid counterfactual trend in the post period for Hedmark.

Tip: If you use the `Synth` package in R, you should use the `quarters` variable to define the time variable in the data preparation function (`time.variable = "quarters"`), because the time variable has to be numeric.

Guide to answer:

- The students should produce a plot which is very similar to Plot 1 in Hernæs (2018).
- Should comment on the fit in the pre-treatment trend between the treated and synthetic treated unit.
- Good answers also show and discuss which regions receive a positive weight, and which variables are highly predictive of the treatment.
- Very good answers explain why or why not they use a validation period.

3.6. Explain and conduct at least two of the three sensitivity analyses described in sections 5.2-5.4 in Hernæs (2018) using the synthetic control model you estimated in **3.5**. If you are not able to conduct the analysis you should still try to describe and explain the purpose of the sensitivity analyses.

Guide to answer:

The student should conduct two sensitivity analyses, and show results that are similar to the results from sections 5.2-5.4 in Hernæs (2018).