Practical assignment trial-based cost-effectiveness analysis

X. Pouwels

2021-09-02

```
rm(list  = ls()) # clear environment
options(scipen = 999) # Disable scientific notations
library(tidyverse)
```

```
## -- Attaching packages --------------------------------------- tidyverse 1.3.0 --

## v ggplot2 3.3.3     v purrr   0.3.4
## v tibble  3.0.5     v dplyr   1.0.3
## v tidyr   1.1.2     v stringr 1.4.0
## v readr   1.4.0     v forcats 0.5.1

## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
```

```
library(knitr)
library(boot)
```

Aim

The aim of this practical assignment is to perform a health economic analysis using data collected during a clinical trials. The data used in this assignment was obtained during a randomised controlled trial (RCT) comparing two interventions for treating esophageal cancer. In total, 200 patients were randomized between usual care (open transthoracic esophagectomy) and the new intervention (robot-assisted thoraco-laparoscopic esophagectomy). For each patient, data was collected on the time required for the procedure in the operating room, the occurrence of complications, the length of stay in the hospital, and the quality of life after the procedure. In addition, data on the cost of resources was collected from the literature.

Instuctions

1. Download the folder `Practical_Trial_based_CEA` from the Canvas page and save it on your computer
2. Open the `Assingment_TB_CEA_start.R` file
3. Load the data for this assignment (`trial_based_CEA.RData`) using the `load()` function

4. When loaded, 8 objects (1 data frame, `df`, and 7 values) should appear in your environment
5. When performing the assignment, please document your code (using R markdown for instance)
6. Please keep your answers for the discussion

Required packages

- `tidyverse`: for data manipulation and plotting
- `knitr`: for reporting, especially when using R markdown
- `boot`: for bootstrapping

Objects

`df`

This dataframe contains 9 variables, each row of the dataframe contains the data from one participant:

- ID = Participant ID
- Procedure = Group to which the participant belongs; 1 = usual care (open transthoracic esophagectomy, 2 = the new intervention (robot-assisted thoraco-laparoscopic esophagectomy)
- OR_time = Operating time, in minutes
- Major_compl = 'Did the participant experience a major complication?'; 1 = yes, 0 = no
- Minor_compl = 'Did the participant experience a minor complication?'; 1 = yes, 0 = no
- Length_icu = Length of stay at the intensive care unit (ICU) in days
- Length_mcu = Length of stay at the intensive care unit (MCU) in days
- Death = 'Did the participant experience die at the hospital?'; 1 = yes, 0 = no
- Qol = Quality of life of the participant after the procedure

other objects

- c_Open = costs of the open surgery (usual care)
- c_Robot = costs of the new intervention (robot-assisted thoraco-laparoscopic esophagesctomy)
- c_ICU_day = costs of one day at the ICU
- c_MCU_day = costs of one day at the MCU

- c_Major_compl = costs of a major complication
- c_Minor_compl = costs of a minor complication
- c_OK_hour = costs of operating in the OK, per hour

Assignment and questions

1. Calculate the mean operating time, risk of complications, length of stay at ICU and MCU, risk of death, and quality of life after procedure, for both trial arms. *Hint: use the functions* `group_by` *and* `summarise` *from the* `tidyverse` *package to do so.*
   1.a. Which treatment provides the best health outcomes?
   1.b. Is robot-assisted surgery an improvement upon usual care with respect to all outcomes?

2. Create three new variables in `df` called Cost_procedure, Cost_compl, Cost_hosp, which respectively contain the costs related to the procedure (operatime time + procedure), the costs related to complications, and the costs related to hospitalization for each participant. To do so, use the objects beginning with `c_`. *Hint: Multiply the column of the dataframe* `df` *by the corresponding objects.*

3. Create a new variable in `df` called Total_costs, which contains the total costs per participant. Calculate the mean costs per group. Based on this information:
   3.a. Which treatment is cheaper on average, given the observed trial data?
   3.b. What do you think is the probability that the new intervention is cheaper?
   3.c. What do you think is the probability that the new intervention leads to health gain?

4. Calculate the difference in mean quality of life and mean total costs between the new intervention and usual care (i.e. the incremental effects and incremental costs). Calculate the ICER using this information. *Hint: use the results you obtained in previous steps.*

5. Save your dataframe `df` as a .csv file. Ensure that your total effects per participants are in the `Qol` variable and the total costs in the `Total_costs` variable.

6. Open the R shiny app using the following command in R and follow the instructions.

```
runGitHub("Teaching", "Xa4P", subdir = "Basics/shiny_app_cea/", ref = "main")
```

7. Run the bootstrap analysis by pushing the "Run bootstrap analysis" in the Shiny app, after having uploaded the file you just created (if this doesn't work, you can use the `df_trial_cea.csv` file that is provided in the map of this assignment. To do so, choose a number of bootstrap samples (at least 1,000 and up to 20,000) and a random seed number (between 1 and infinity). Wait a couple of seconds that the analysis runs. The results will automatically appear on the screen. The results that you see are related to the uncertainty that remains after observing this set of trial data. In addition to the bootstrap results, the mean results of the sample data (that you have calculated earlier), are also provided.
   7.a. Check the average estimates from the bootstrap samples. Do these results correspond to the point estimates from the observed data?
   7.b. Check the confidence intervals for the additional health effects and for the additional costs of the intervention. Can you explain why they are so wide?
   7.c. Check the probabilities of a cost-effectiveness outcome in each one of the four quadrants. How well did you guess the probability that the new intervention is cheaper, or improves health (questions 3b and c)?
   7.d. No confidence interval is given for the ICER (cell M64) can you explain why such a confidence interval is not informative by looking at the incremental CEA plane in column AA?
   7.e. Rerun the bootstrap procedure 3 times using different seed numbers. Do the results of the bootstrap analysis change substantially? What would we need to do to reduce this variation in outcomes?