PH251D Fall 2019 - Project 2 (v. 2019-12-01-A)

FirstName MI LastName

2019-MM-DD

Download this Rmarkdown template (`PH251D2019_LastName_Project2.Rmd`) and edit.

Create a project folder called `project2` on your computer. You will put all your Project 2 files in this folder.

Use this template and R Markdown to demonstrate the skills below. Be sure to compile this document to .DOC, .HTML, or PDF file for submission. Be sure to change `output` option in YAML.

You will need to install `bnlearn`, `tableone`, `survey`, and `sandwich` packages.

1. Reading in CSV file

The CAGE is a four-item questionnaire for screening for alcohol use disorder. A patient can answer "yes" to zero, one, two, three, or four questions. Mayfield conducted a CAGE study in 1974 with 366 patients, recorded the number of CAGE questions that were answered "Yes" (range was 0 to 4). A psychiatric social worker independently conducted in-depth interviews and each patient was classified by whether he or she had alcohol use disorder or not (AUD: "Y", "N"). The AUD status was considered the gold standard. The study results are available here: https://raw.githubusercontent.com/taragonmd/data/master/cage.csv

Read in the data set and display a two-way continency table using the `xtabs` function. The column headings should be the `AUD` status.

```
## R code here to display two-way table similar to above
```

2. Recode integer variable into a factor with two levels

Recode the `num_yes` variable into a new factor variable called `CAGE`:

Factor	Level	`num_yes` values
`CAGE`	`pos`	>= 3
	`neg`	<= 2

Use the `xtabs` function to display 2 by 2 table of CAGE by AUD.

```
## display 2 x 2 table
```

3. Calculate the sensitivity and specificity of the CAGE questionnaire

The sensitivity is P(CAGE = pos | AUD = Y), and the specificity is P(CAGE = neg | AUD = N).

```
## calculate sensitivity and specificity
```

4. Posterior probabilities as a function of prior probabilities

You have learned that the posterior (post-test) probability is a function of the prior (pre-test) probability, and sensitivity and specificity of the diagnostic test (in this case, the CAGE questionnaire). For alcohol use disorder (AUD), the positive predictive value (PPV) is the probability that a patient has AUD given that they answered "yes" to 3 or 4 items on the CAGE questionnaire.

Using the sensivity and specificity you calculated above, plot a graph displaying on how PPV changes as a function of the prior probability. Some of the R code is provided, including Bayes' theorem for calculating PPV.

```
## plot of PPV vs prior probability (prior)
## use `sens` and `spec` from previous results
prior <- seq(from = 0, to = 1, by 0.05)
ppv <- (sens * prior) / (sens * prior + (1 - spec) *(1 - prior))
```

5. Bayesian network for probabilistic reasoning

Install the `bnlearn` package as described here: http://www.bnlearn.com/. The causal model is AUD -> CAGE. Create a Bayesian network (BN) model by fitting the causal model to data (I have done this for you).

```
library(bnlearn)
```

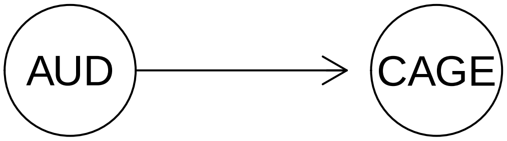
```

Attaching package: 'bnlearn'

The following object is masked from 'package:stats':

    sigma
```

```
curl <- 'https://raw.githubusercontent.com/taragonmd/data/master/cage2.csv'
aud <- read.csv(curl, header = TRUE)
summary(aud)
```

```
 AUD      CAGE
 N:224   neg:265
 Y:142   pos:101
```

```
dag <- empty.graph(nodes=c("AUD", "CAGE")) # Build the Bayesian network
dag <- set.arc(dag, from = "AUD", to = "CAGE")
graphviz.plot(dag, layout = "circo") # Plot the BN
```

```
Loading required namespace: Rgraphviz
```



```
bn <- bn.fit(dag, data = aud)        # Fit the BN to the data
# cpquery
```

Assuming the prior probability of AUD is 5%, use the `cpquery` function from `bnlearn` package to calculate the PPV; that is P(AUD = "Y" | CAGE = "pos").

6. Deconfounding: IPW with stratified contingency table

In 1986, Charig published a retrospective observational study comparing open surgery (Treatment A) to percutaneous nephrolithotomy (Treatment B) for the treatment of kidney stones. Treatment B is noninvasive. Success was defined as no stones at three months. The data are presented in Table below.

Table 2: Comparison of Treatment A to B for kidney stones, by size

Stone size	Treatment A	Success (%)	Treatment B	Success (%)
Small stones	81/87	**93**	234/270	87
Large stones	192/263	**73**	55/80	69

Stone size	Treatment A	Success (%)	Treatment B	Success (%)
Combined	273/350	78	289/350	**83**

Urologists select open surgery (Treatment A) for more severe cases, and larger stones are considered more severe. In other words, severity causes treatment selection, and severity lowers success. Stone size (severity) is a common cause or fork that may be confounding the causal relationship between treatment and success (outcome).

Review Section 4.3 from https://escholarship.org/uc/item/8000r5m5

The study results are available here: https://github.com/taragonmd/data/blob/master/kidney.csv

Read in the data set. Using the inverse probability weighting (IPW) method (see Section 4.3.1.1) and report on the average causal effect (ACE) in terms of the following:

(a) causal risk difference
(b) causal risk ratio
(c) causal odds ratio

```
## put R code here
```

7. Deconfounding: IPW with marginal structural model (MSM)

Repeat the analysis above using IPW/marginal structural model (MSM) that uses the propensity scores and the generalized linear model (`glm` function in R). Edit the R code from Section 4.3.1.2 and report on the average causal effect (ACE) in terms of the following:

(a) causal risk difference
(b) causal risk ratio
(c) causal odds ratio

You will need to install the `tableone`, `survey`, and `sandwich` package.

```
## Put R code here
```