front-door

yousri elfattah

2024-07-29

## R Markdown

This is an R Markdown document. Markdown is a simple formatting syntax for authoring HTML, PDF, and MS Word documents. For more details on using R Markdown see <http://rmarkdown.rstudio.com>.

When you click the **Knit** button a document will be generated that includes both content as well as the output of any embedded R code chunks within the document.

library(dagitty)  
library(ggdag)

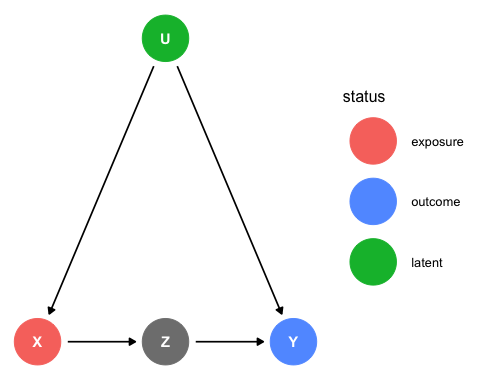
##   
## Attaching package: 'ggdag'

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

# Chapter 9

## Front-Door Criterion

library(dagitty)  
library(ggdag)  
g <- dagitty("dag{  
 X <- U -> Y   
 X -> Z -> Y   
 X [exposure]  
 Y [outcome]  
 U [unobserved]  
}")   
# set coordinates for graph layout  
coords <- list(X=c(X=1, Z=2, Y=3, U=2),  
 Y=c(X=1, Z=1, Y=1, U=2))  
coords\_df<-coords2df(coords)  
coordinates(g)<-coords2list(coords\_df)  
#Plot causal graph  
ggdag::ggdag\_status(g) + ggdag::theme\_dag()



Simulate observational data set

library(mvtnorm)  
# simulate data (linear Gaussain model)  
n <- 1e5  
# we will generate correlated, normal random variables with the program rmmvnorm in the mvtnorm library  
out <- rmvnorm(n, mean = c(0, 0),   
 sigma = matrix(c(1,.6, .6, 1), 2, 2))  
Ux <- out[,1]  
Uy <- out[,2]  
Uz <- rnorm(n)  
X <- 0.1 \* Ux  
Z <- 1.2 \* X + 0.3 \* Uz  
Y <- 2.0 \* Z + 0.2 \* Uy  
model.data <- data.frame (  
 X <- X,  
 Z <- Z,  
 Y <- Y  
)

## Linear regression

Exstimate path coefficient X -> Z

model\_1 <- lm(Z ~ X, model.data)  
model\_1.coefs <- coef(model\_1)  
print(model\_1.coefs)

## (Intercept) X   
## -0.001310523 1.194918974

Estimate path coefficient Z -> Y

model\_2 <- lm(Y ~ Z, model.data)  
model\_2.coefs <- coef(model\_2)  
print(model\_2.coefs)

## (Intercept) Z   
## 0.0009189206 2.1383890094

model\_3 <- lm(Y ~ X, model.data)  
model\_3.coefs <- coef(model\_3)  
print(model\_3.coefs)

## (Intercept) X   
## -0.002639476 3.585885865

## Drawing a scatter plot

library(ggplot2)  
# Plot data points  
p <- ggplot(model.data, aes(x = X, y = Y)) +  
 geom\_point(alpha = 0.1, size = 0.1)   
  
# Creating the lines  
line1\_data <- data.frame(  
 X = range(model.data$X),  
 Y = model\_1.coefs["X"] \* model\_2.coefs["Z"] \* range(model.data$X)  
)  
  
line2\_data <- data.frame(  
 X = range(model.data$X),  
 Y = model\_3.coefs["X"] \* range(model.data$X)  
)  
  
# Add informative text box  
textbox <- paste0(  
 "Average Treatment Effect (ATE)\n",  
 "True value = 2.4\n",  
 "Estime value (Red regression slope) = ", round(model\_1.coefs["X"] \* model\_2.coefs["Z"],4), "\n",  
 "Estimate value (Blue regression slope) = ", round(model\_3.coefs["X"],4)  
)  
  
# Plotting the lines - Increased size to 2 for thicker lines  
p +  
 geom\_line(data = line1\_data, aes(X, Y), colour = "red", size = 2) +  
 geom\_line(data = line2\_data, aes(X, Y), linetype = "dashed", colour = "blue", size = 2) +  
 annotate(  
 "text",  
 x = min(model.data$X),  
 y = max(model.data$Y),  
 label = textbox,  
 hjust = 0,  
 vjust = 1,  
 size = 4,  
 colour = "black"  
 )

