Homework 3A

Sam Bashevkin

2022-08-05

```
library(rethinking)
```

```
## Loading required package: rstan
```

```
## Loading required package: StanHeaders
```

```
##
## rstan version 2.26.11 (Stan version 2.26.1)
```

```
## For execution on a local, multicore CPU with excess RAM we recommend calling
## options(mc.cores = parallel::detectCores()).
## To avoid recompilation of unchanged Stan programs, we recommend calling
## rstan_options(auto_write = TRUE)
## For within-chain threading using `reduce_sum()` or `map_rect()` Stan functions,
## change `threads_per_chain` option:
## rstan_options(threads_per_chain = 1)
```

```
## Do not specify '-march=native' in 'LOCAL_CPPFLAGS' or a Makevars file
```

```
## Loading required package: cmdstanr
```

```
## This is cmdstanr version 0.5.2
```

```
## - CmdStanR documentation and vignettes: mc-stan.org/cmdstanr
```

```
## - CmdStan path: C:/Users/sbashevkin/Documents/.cmdstan/cmdstan-2.29.2
```

```
## - CmdStan version: 2.29.2
```

```
##
## A newer version of CmdStan is available. See ?install_cmdstan() to install it.
## To disable this check set option or environment variable CMDSTANR_NO_VER_CHECK=TRUE.
```

```
## Loading required package: parallel
```

```
## rethinking (Version 2.21)
```

```
##
## Attaching package: 'rethinking'
```

```
## The following object is masked from 'package:rstan':
##
##     stan
```

```
## The following object is masked from 'package:stats':
##
##     rstudent
```

```
library(dagitty)
```

5E1. Which of the linear models below are multiple linear regressions?

2, 3, and 4

5E2. Write down a multiple regression to evaluate the claim: Animal diversity is linearly related to latitude, but only after controlling for plant diversity. You just need to write down the model definition.

$$
\begin{aligned}
adiverity_i &\sim Normal(\mu_i, \sigma) \\
\mu_i &= \alpha + \beta_l latitude_i + \beta_p pdiversity_i \\
\alpha &\sim Normal(167, 40) \\
\beta_l &\sim Log - Normal(0, 5) \\
\beta_p &\sim Log - Normal(0, 5) \\
\sigma &\sim Exponential(1)
\end{aligned}
$$

5E3. Write down a multiple regression to evaluate the claim: Neither amount of funding nor size of laboratory is by itself a good predictor of time to PhD degree; but together these variables are both positively associated with time to degree. Write down the model definition and indicate which side of zero each slope parameter should be on.

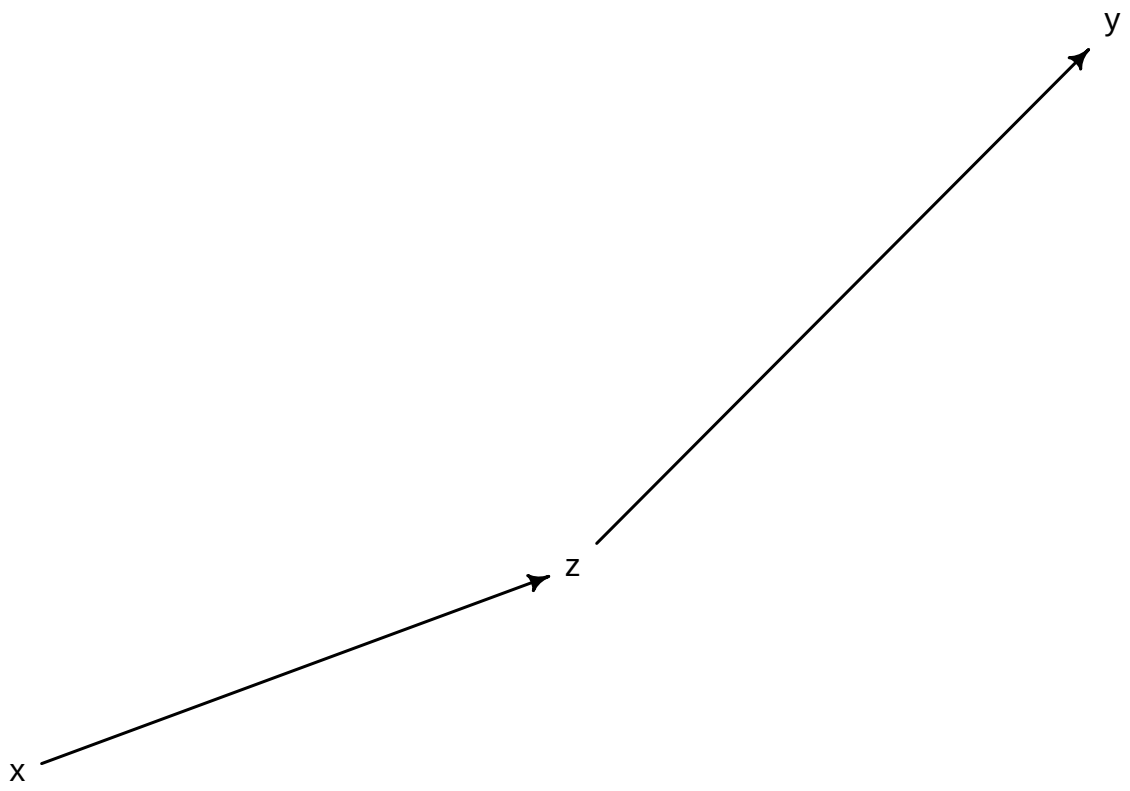
? I can't imagine the DAG for this

5E4.

1, 3, and 4

5M1. Invent your own example of a spurious correlation. An outcome variable should be correlated with both predictor variables. But when both predictors are entered in the same model, the correlation between the outcome and one of the predictors should mostly vanish (or at least be greatly reduced).

```
dag <- dagitty("dag {
    x -> z -> y
}")
drawdag(dag)
```



The pipe. Once stratified by z (e.g., z and x are included in model) there is no longer a relationship between x and y

x=air temp, z = snowmelt, y = inflow (in undammed river during spring)

5M2. Invent your own example of a masked relationship. An outcome variable should be correlated with both predictor variables, but in opposite directions. And the two predictor variables should be correlated with one another.