COMPUTATIONAL STATISTICS & PROBABILITY

FALL 2021
PROBLEM SET 3

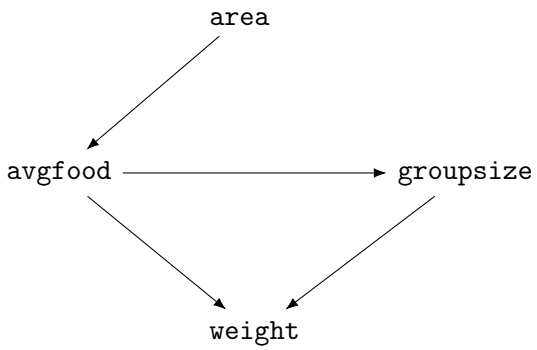
DUE DATE:	23:59:59 **November 24**, 2021
FORMATTING:	Acceptable formats are `.pdf`, `.R`, or `.Rmd` only.
	For `.Rmd` files, submit also a knitted `pdf`.
	For any format, each answer should be clear to a human without running code.
SUBMITTING:	Upload your completed assignment to Canvas.
	Submissions must comply with the homework policy on the course page.

---

1. Urban foxes are like street gangs. Groups vary from 2 to 8 individual foxes, and each group maintains its own (almost exclusive) urban territory. Some territories are larger than others. The data set `foxes` in the *rethinking* package consists of data for 116 foxes from 30 different urban groups in England. You can load and inspect the data in the usual way, namely by

```
library(rethinking)
data(foxes)
d <- foxes
precis( d )
```

   The variable `area` encodes that some territories are larger than others; the variable `avgfood` encodes that some territories have more average food than others. Suppose we want to model the `weight` of each fox. For the questions in this section, assume this causal DAG:



   Questions:

   a) Does territory size have a causal influence the weight of foxes? Construct a *quap model* to infer the total causal influence of `area` on `weight`. Does increasing the area available to each fox make it healthier (i.e., heavier)? I recommend that you standardize your variables and use prior predictive simulation to show that your model's predictions stay within the possible outcome range.

b) Now infer the causal impact of adding food (`avgfood`) to a territory. Would this make foxes heavier? Which covariates do you need to adjust to estimate the total causal influence of food?

c) Now infer the causal impact of group size (`groupsize`). Which covariates do you need to adjust to make this estimate? Inspect the posterior distribution of the resulting model. What do you think explains these data? Specifically, explain the estimates of the effects of `area`, `avgfood`, and `groupsize` on `weight`. How do they make sense together? (Hint: we covered an example in class which exhibited a similar relationship between predictors and outcome variable.)

2. Explain the difference between model *selection* and model *comparison*. What information is lost under model selection?

3. Use WAIC or LOO based model comparison on five different models, each using `weight` as the outcome, and containing the follow sets of predictor variables:

   (1) `avgfood + groupsize + area`

   (2) `avgfood + groupsize`

   (3) `avgfood + area`

   (4) `avgfood`

   (5) `area`

   Can you explain the relative differences in WAIC scores, using the fox DAG from above? Be sure to pay attention to the standard error of the score differences (dSE).