

# Final

Stat 451

April 14, 2022

This is a take-home exam. The final will be due at midnight on Tuesday, April 19. When you finish, email me your exam with the subject ‘Stat 451 Final’. Please do your work in a .Rmd file so that I can see your code. Please hand in both your .Rmd file and either an .html file, a .pdf file, or a .doc file. Please name your files with your last name first, so my files would be named: ‘FellinghamFinal.Rmd’ and ‘FellinghamFinal.pdf’, ‘FellinghamFinal.html’ or ‘FellinghamFinal.doc’.

As always, I expect your final to be your own work. You may use any notes and any help files available for the programs, but searching for similar code on the web would not be appropriate. You should not ask for help from any person currently alive in the mortal state.

The first question is worth 75 points and should be the focus until you believe you have completed it. The second question is worth 25 points.

1. On an annual basis, each county Assessor is required by Utah law to list and value on an assessment roll all property subject to *ad valorem* taxation. Iron County is located in southwest Utah approximately 265 miles south of Salt Lake City, UT and 170 miles north of Las Vegas, NV on the I-15 corridor. The Iron County Assessor’s office assesses values on approximately 35,000 parcels of property on approximately 620,000 acres.

The data file *ironco.txt*, will be mailed to you, contains data on selling price for various properties, as well as information of covariates that may be related to selling price. The columns are described below:

- obs - observation number (in quotes)
- price - selling price of the property
- lot - lot acreage
- floors - number of floors (not including basement)
- const - assessed construction quality on a scale of 1 (poor) to 4 (excellent)
- roof - assessed roof condition on a scale of 1 (poor) to 4 (excellent)
- build - assessed home condition on a scale of 1 (poor) to 4 (excellent)
- area - square footage of home
- yr.built - year the home was built
- eff.age - evaluation by the assessor of the home’s equivalent market age
- baths - number of full bathrooms
- gar - indicator for presence of a garage
- basmt - indicator of presence of a basement

Make the response the square root of price and write a model to estimate the response based on an intercept, the area of the home, the assessed roof condition, and the indicator for the presence of a basement. The likelihood is normal. Make your priors for the  $\beta$ ’s fairly flat and use a gamma prior for the error variance with parameters shape=3 and rate=.1. Run the model in JAGS or Stan. Make sure the chains have converged using the methods we have covered in class. Show the 95% equal tail interval estimates for  $\beta_{\text{area}}$ . Plot the posterior density of  $\beta_{\text{area}}$ .

2. The data we are considering are from the Body Project, a dissonance based eating disorder prevention intervention. Female adolescents were randomized to one of four conditions: the dissonance intervention (DI; n=114), a healthy-weight management program (HW; n=117), an expressive writing control (EW,

n=123), and an assessment-only control (AO; n=126). For the purposes of this exam, we are only considering two treatments, DI and AO. The DI treatment was delivered to groups of participants, (average group size=6.7). The AO treatment was delivered to individuals. We would expect the responses in the DI groups to show some within group correlation. The researchers' question of interest was whether the DI treatment lowered the response (called TII2 in the data file).

The data file (called *bp.dat*) will be emailed to you. There are seven columns in the data file. The first column is a subject id that has issues because we are only looking at two treatments. The second column is a groupid that also had some issues. Ignore the first two columns. The next column is the treatment condition. It is 0 for the DI treatment and 3 for the AO treatment, the only two you care about. The next column is labeled DIS and it also is a treatment indicator, 1 for the DI treatment and 0 for the AO treatment. These two columns have identical information. The next column is TII2, which is the response of interest. The final two columns are subject id and group id that have been fixed up for this data set.

There are many ways to think about a model in this case. I think the most reasonable one for us is a hierarchical model, as there are multiple sources of variability. Subject variability in both the AO and DI treatments, as well as group variability for the DI treatment. For picking priors you should be aware that all the variance components for this data are fairly small (all less than 1). You should write code to evaluate the conjecture that DI is a better treatment than AO.

The most important item you will hand in is your code, either JAGS or Stan. If you get your code to run I would also like to see convergence diagnostics, a summary output table, an equal tail interval for the parameter used to indicate the possible presence of a treatment effect, and a plot of the posterior density of this same parameter.