

GR5065 Homework 6 Answer Key

Ben Goodrich

Due April 25, 2022 at 4PM

```
set.seed(20220425)
```

1 Social Surveys

```
if (!file.exists("GSS_2020_panel_stata_1a.zip")) {  
  download.file("https://gss.norc.umd.edu/Documents/stata/GSS_2020_panel_stata_1a.zip",  
               destfile = "GSS_2020_panel_stata_1a.zip")  
  unzip("GSS_2020_panel_stata_1a.zip")  
}  
library(haven)  
GSS <- as_factor(read_dta(file.path("GSS_2020_panel_stata_1a", "gss2020panel_r1a.dta")))  
ANES <- as_factor(read_dta(file.path("anes_timeseries_2020_gss_stata_20220408",  
                                   "anes_timeseries_2020_gss_stata_20220408.dta")))  
  
library(dplyr)  
ANES_GSS <- inner_join(ANES, GSS, by = c(YEARID = "yearid"))
```

Here I am going to model a question about support for gay marriage

```
ANES_GSS$y <- factor(ANES_GSS$marhomo_2, ordered = TRUE)  
ANES_GSS$degree_2 <- factor(ANES_GSS$degree_2, ordered = TRUE)  
ANES_GSS$age_2 <- as.numeric(as.character(ANES_GSS$age_2))  
ANES_GSS$relig16_2 <- with(ANES_GSS, ifelse(is.na(relig16_1a),  
                                           as.character(relig16_1b),  
                                           as.character(relig16_1a)))  
ANES_GSS <- filter(ANES_GSS, relig16_2 != "native american") # too few
```

1.1 Prior Predictive Distribution

First, we take a look at the default priors

```
library(brms)  
options(mc.cores = parallel::detectCores())  
get_prior(y ~ mo(degree_2) + sex_2 + s(age_2) + relig16_2, data = ANES_GSS, family = cumulative)
```

##	prior	class	coef group resp dpar
##	(flat)	b	
##	(flat)	b	modegree_2
##	(flat)	b	relig16_2catholic
##	(flat)	b	relig16_2christian
##	(flat)	b	relig16_2hinduism
##	(flat)	b	relig16_2jewish
##	(flat)	b	relig16_2muslimDislam

[illegible]

which need to be made proper in order to draw from the prior predictive distribution

```
prior <- brm(y ~ mo(degree_2) + sex_2 + s(age_2) + relig16_2,
  data = ANES_GSS, family = cumulative, sample_prior = "only",
  prior = prior(normal(0, 1), class = "b"), seed = 20220425)
```

In this case, just putting a standard normal prior on the coefficients happens to achieve what we want in the prior predictive distribution, namely putting about equal chances on all possible outcomes

```
prop.table(table(posterior_predict(prior)))
```

```
##
##           1           2           3           4           5
## 0.2200820 0.1916614 0.1894764 0.1894767 0.2093035
```

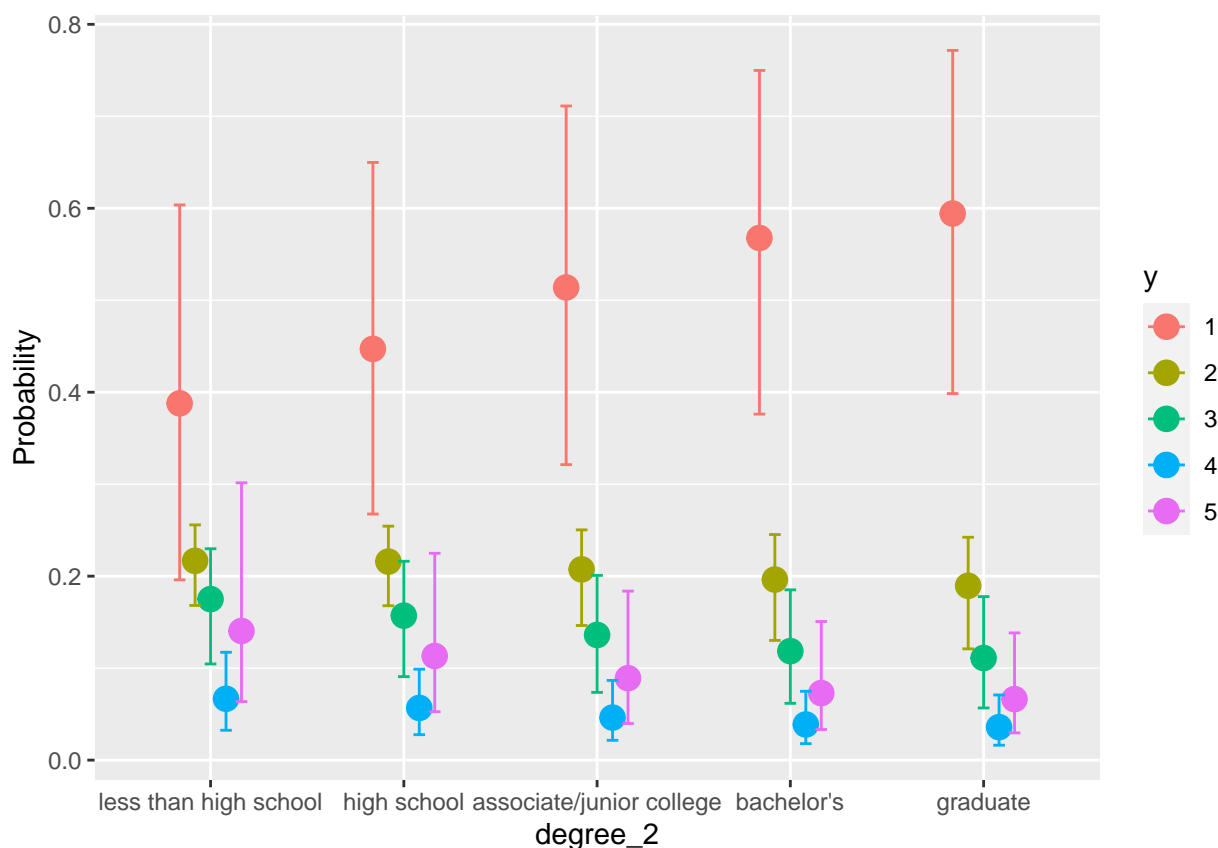
1.2 Posterior Distribution

```
post <- update(prior, sample_prior = "no", control = list(adapt_delta = 0.96))
```

1.3 Interpretation

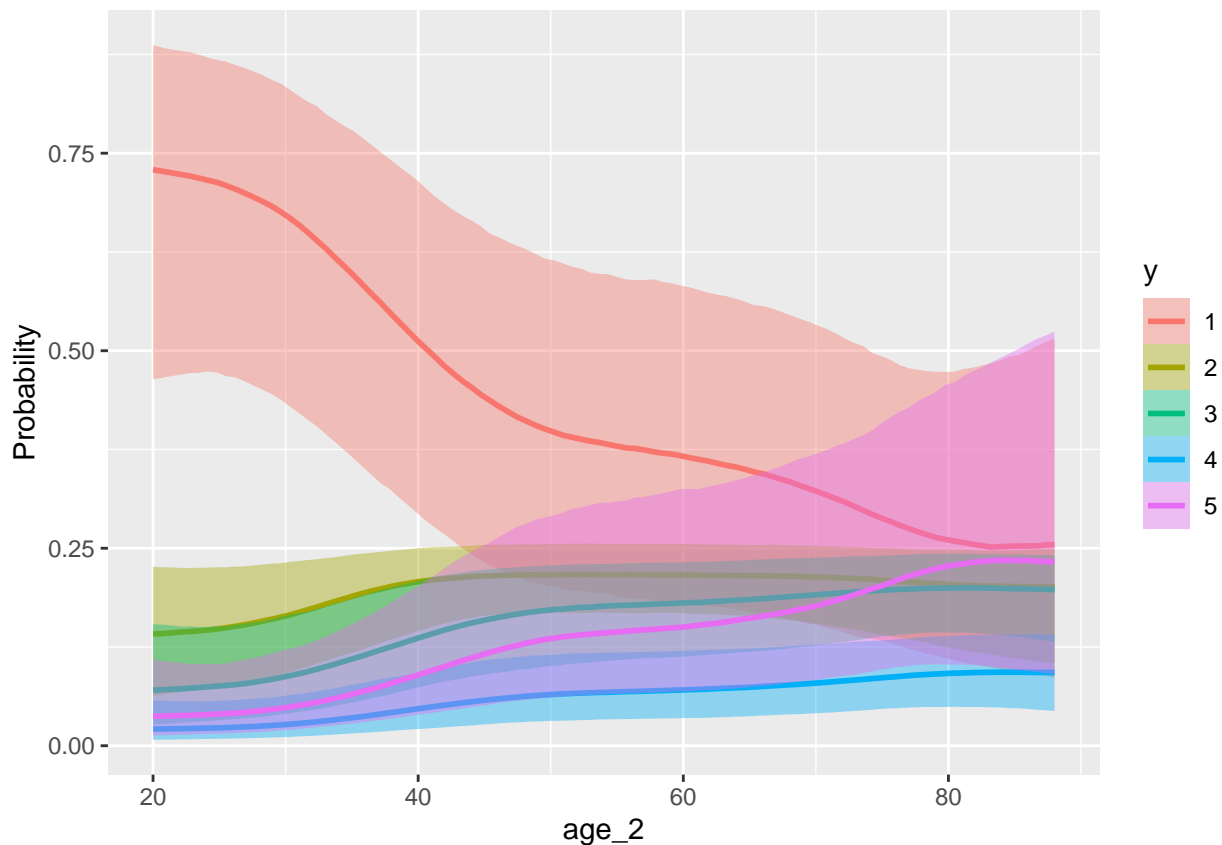
The probability of strongly agreeing with same-sex marriage increases with the educational degree obtained, although there is little difference between having only a bachelor's degree and having a graduate degree. The probability of every other outcome category decreases with degree obtained.

```
plot(conditional_effects(post, effects = "degree_2", categorical = TRUE))
```



Although the effect of age was not restricted to be monotonic, it turned out that way in the posterior

```
plot(conditional_effects(post, effects = "age_2", categorical = TRUE))
```



1.4 PSISLOOCV

```
loo(post)
```

```
##
## Computed from 4000 by 762 log-likelihood matrix
##
##           Estimate   SE
## elpd_loo   -975.3 22.0
## p_loo       16.9  1.2
## looic      1950.5 44.0
## -----
## Monte Carlo SE of elpd_loo is 0.1.
##
## All Pareto k estimates are good (k < 0.5).
## See help('pareto-k-diagnostic') for details.
```

This has no problems with high Pareto k values

```
post2 <- brm(y ~ degree_2 + sex_2 + s(age_2) + relig16_2, data = ANES_GSS,
  family = cumulative, prior = prior(normal(0, 1), class = "b"),
  seed = 20220425, control = list(adapt_delta = 0.95))
```

```
loo_compare(loo(post), loo(post2))
```

```
##           elpd_diff se_diff
## post      0.0         0.0
## post2    -1.1         1.3
```

But there is not much difference in the ELPD between the two models.

1.5 Posterior Predictive Distribution

The posterior predictive distribution is not uniform

```
prop.table(table(posterior_predict(post)))
```

```
##
##           1           2           3           4           5
## 0.50687566 0.19175033 0.13846719 0.05235696 0.11054987
```

It is similar to the marginal distribution of the *observed* outcomes

```
prop.table(table(ANES_GSS$y))
```

```
##
##           strongly agree           agree
##           0.5019506           0.1898570
## neither agree nor disagree           disagree
##           0.1417425           0.0520156
##           strongly disagree
##           0.1144343
```

However, there are many people who did not answer the question, and it is interesting to see that the model predicts them in the same way as for the people who did answer:

```
prop.table(table(posterior_predict(post,
                                newdata = filter(ANES_GSS, is.na(y),
                                                !is.na(age_2)))))
```

```
##
##           1           2           3           4           5
## 0.49328802 0.19212758 0.14173840 0.05445876 0.11838724
```

It is possible that not everyone was asked this question after the election.

2 The Replication Crisis in Psychology

The “Crisis in Psychology”, as stated in the *New York Times* article, is essentially that

Since 2011, a methodological reform movement has been rattling the field, raising the possibility that vast amounts of research, even entire subfields, might be unreliable.

This methodological reform movement has done more than raise the possibility that vast amounts of research might be unreliable; it has gone a long way toward demonstrating that vast amounts of research are unreliable. At a minimum, one cannot easily tell which articles that were published in various journals are reliable and which ones are not because journals engaged in many dubious practices, chief among them that the article had to obtain a p -value of less than 0.05 in order to be published (although that might not be sufficient). As Andrew Gelman has tried to point out, when journals filter by whether a p -value is less than 0.05, you cannot worry solely about Type I and Type II errors but also have to worry about Type S and Type M errors. A Type S error is when the point estimate has a different sign than the parameter it estimates, and a Type M error is when the point estimate has a substantially different magnitude than the parameter it estimates. Conditional on the p -value being less than 0.05, the probability of a Type S and / or a Type M error is larger than the unconditional probability, so journals may be filled with Type S and Type M errors.

Although the *New York Times* article does not say so, many articles in many fields do not have a fully specified generative model (or Directed Acyclic Graph) for the endogenous known they are studying. Rather, they hope to substantiate a hypothesis about the effect of one predictor (i.e., one arrow in a DAG) without

having much idea about the combined effects of the other predictors (i.e., the rest of the unwritten DAG). In situations like these, it is easy for the contribution to the signal by the predictor of interest to be much less than the apparent noise in the predictor driven by all of the other predictors that the model has not accounted for. Frequentist point estimation with finite samples is not particularly reliable in this situation, even on its own terms, which is to say that the distribution of point estimates across studies that differ only in which N observations are randomly sampled from the population is not well-described by a normal distribution. Thus, the p -values that were used to decide whether to consider an article for publication would not be distributed standard uniform, even if the null hypothesis of no effect were true.

Another aspect of this problem is that the Frequentist methods used to obtain these point estimates are anything but objective when used by actual humans with all sorts of professional incentives to conclude that their hypothesis is correct. Questions such as how to design the study and what data to collect are subjective, as is the now widely criticised practice of P-hacking: changing the estimation procedure if the previous estimates from the same data do not (sufficiently) provide the answer you are looking for, which also renders the distribution of the p -values non-uniform across datasets even if the null hypothesis were true.

To a large extent, the younger generation of psychology researchers — who had somewhat better methodological training and more inclination to discuss research findings on social media — objected to these well-entrenched but flawed practices. One question that is now asked is: Can a study (that often finds something implausible) be “replicated”? The p -value of the original study does not speak to the probability that another study will also obtain a small p -value because a p -value is conditional on the null hypothesis being true, which the original study has rejected.

But a lot of the problems with the use of p -values to make “either / or” decisions about whether a paper should be considered for publication are replicating themselves when making the “either / or” decision as to whether a study that originally obtained a small p -value has or has not been replicated by a subsequent study. On one hand, no two studies are identical, even if they are only intended to differ on which observations are randomly sampled from the population. So, it is not clear that the subsequent study should obtain the same results as the original. But worse, the consistency of findings between the original study and the subsequent ones is a matter of degree rather than an “either / or”. Thus, not everyone agrees with Amy Cuddy’s claim that some of the conclusions of the original power posing study have been replicated many times, apparently including at least one of the co-authors of the original power posing study.

Many of the people that criticize historical methodological practice in social psychology are not Bayesian but are often more open to Bayesian analysis than their predecessors. How can Bayesian analysis mitigate these problems? First, because a posterior distribution is conditional only on the data at hand (and the priors) it is “exact” for a given N and does not need to rely on asymptotic results that may not be applicable to a particular study. Moreover, we generally have no idea when an asymptotic result is going to hold reasonably well, so you would have to do the Bayesian analysis to see if the Frequentist analysis is similar.

Second, one of the issues with the original power pose study is that nearly everyone believed it was implausible that standing or sitting in a certain position for one minute could alter hormone levels enough to have a measureable effect on your well-being. In other words, the audience for the study had a prior that was near zero. By incorporating such a prior into the analysis, you reduce the opportunity to obtain a large estimate due to some accidental characteristic of the data that you happened to collect (which usually is not a sample from a well-defined population anyway) and if you do happen to obtain a posterior distribution that is largely bounded away from zero, even people who were initially skeptical should believe your results (but will often find some other aspect of the study to disagree with).

Third, with draws from the posterior distribution, you can evaluate a substantive utility function for different (even hypothetical) values of the predictors. Thus, you can move away from the typical Frequentist approach of establishing that the point estimate is “in the right direction” and is statistically significant toward the Bayesian approach of evaluating whether the effect is substantively important in light of our updated beliefs about it given the data. It seems quite likely that many of the statistically significant results in social psychology — even those that can be replicated reasonably well — are so small in substantive terms that they do not merit the hype that has been ascribed to them.

Fourth, the Bayesian approach leads naturally toward thinking about an evolution of beliefs about an effect as more studies are conducted. Perhaps the first study is suggestive but does not estimate an effect precisely enough to be sure that it is non-negligible. The next study can use the posterior distribution of the first study to form its prior and update that into the next posterior distribution. After enough studies have been conducted, scholars in this area should (although they may not) have a fairly precise set of beliefs about the effect. There is no coherent way of updating beliefs with Frequentist estimators, so what ends up happening is things like the first study cannot get published unless its p -value is less than 0.05 but if it does, then we are supposed to presume the effect has been found, unless the second study does not achieve a p -value less than 0.05 in which case we claim the first study “cannot be replicated” and revert back to the null hypothesis that the effect was never there at all. Binary thinking is terrible for science and yet many scientists are unable to conduct science without engaging in binary thinking.

The Bayesian approach provides an alternative to Frequentist notions of “replication”. Rather than redoing the original study with different data and hoping that the p -value is on the same side of 0.05 as in the original, Bayesians can use the posterior distribution of the original study as a prior distribution for a subsequent study and condition on the subsequent study’s data to obtain a second posterior distribution. This would move away from thinking about replication in a discretized way and instead continuously update our state of beliefs about the unknowns under consideration.

But it would take a monumental effort to get subfields with the social sciences to use Bayesian methods by default. Stan and related R packages like `rstanarm` and `brms` make it much easier to use Bayesian methods, almost as easy as using the corresponding Frequentist estimators. But plenty of social scientists do not even use R and among those that do, the vast majority still utilize maximum likelihood to estimate the model. You would need to have required courses that teach Bayesian methods and most faculty and most graduate students do not want that. They want Frequentist estimation to be what it is not. They want methods that ask as little as possible of the researchers using them. And that is not good for science.