Expanding the Use of Hierarchical Bayes Models in Marketing Research

Michael Conklin

56Stats LLC

Agenda

Brief History of HB in Marketing Research

Rethinking Statistics and Stan

Examples and Benefits of HB Models

History of HB Models in Marketing Research

BAMMCONF 1999

Bayesian Applications of Marketing Models

Greg Allenby (Ohio State) and Peter Lenk (U Michigan) provided MATLAB code for a generic Hierarchical Bayes Discrete Choice Model.

- Beginning in the late 1990’s there was a lot of academic development of hierarchical bayes models to assess the effect of marketing efforts.
  - These tended to be custom designed and coded models and were mostly used in research papers in academic journals.
- Market research practitioners could not easily access these models and generally used regular conjoint or discrete choice models with software provided by Sawtooth Software.
- Then Rich Johnson (founder of Sawtooth Software) attended BAMMCONF

History of HB Models in Marketing Research

Rich Johnson converts the MATLAB code provided at BAMMCONF to C and incorporates into Sawtooth Software's Conjoint Analysis suite of programs.

- HB Choice Models quickly become the standard in marketing research.
  - The models were applicable to a wide variety of research questions e.g. why do people choose the brands and products that they buy.
  - The set up was fairly easy
  - HB models produce coefficient estimates for each respondent. This adds to the flexibility of the analysis.
  - Even when the underlying models are main effects only, the heterogeneity of estimates across the population allows for interactions to become apparent.

History of HB Models in Marketing Research

When Maximum Difference (MaxDiff) models were developed it was trivial to treat them as a special case of the HB Choice Model.

- MaxDiff or Best-Worst Scaling was described by Louviere and Marley in 2005.
- In this approach respondents are shown sets of four items and asked which is the “best” item in the set and which is the “worst”
- This is transformed into a series of paired comparisons which can then be fit with the standard HB Choice Model software.
- This has become a common technique for eliciting an ordering of a number of items and because of HB fitting we get estimates for each person in the study.

Hierarchical Bayes does not mean just choice models

HB Modeling strategies allow us to build models that represent how we think the world works.

The problem for the practitioner is how to easily construct such a model

- Hierarchical Bayes simply means that there are multiple levels of aggregation in a model and that model is estimated using Bayesian inference techniques.
- In general, we are estimating parameters for individuals in our data, which are informed by an overall set of population level parameters, which are informed by a prior distribution on those parameters.
- This allows us to flexibly build models that reflect the way we think the world works.

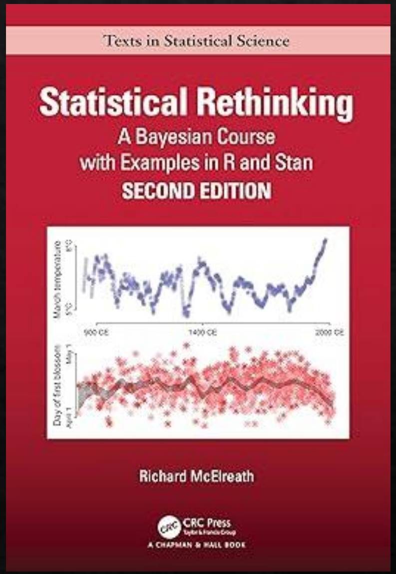
How to build different models

General purpose Bayesian Model software exists but each one requires learning program specific terminology and coding practices.

The biggest issue to me is the fragility of the software. It is easy to create a model that throws an error that is meaningless to a new user.

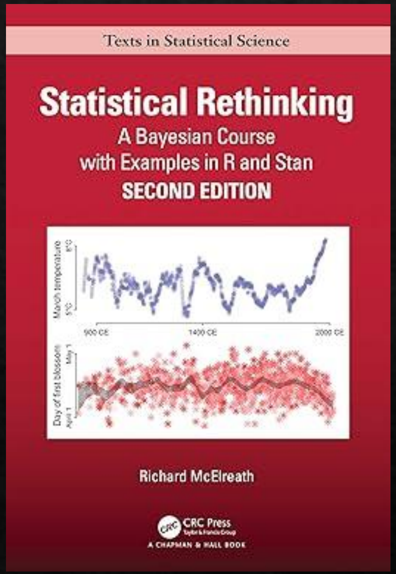
- First there were stand alone programs like BUGS and it's Windows version WinBugs.
- Bayesian Data Analysis (Gelman, Carlin, Stern and Rubin 2nd Edition (2004)) makes extensive use of R to feed data into BUGS.
- In 2012 the first version of Stan, a probabilistic programming language written in C++ specifically to specify Bayesian statistical models, was released by the Stan Development Team led by Andrew Gelman.
- Stan utilizes some creative methods that speed up Bayesian models substantially.

Enter "Statistical Rethinking" by Richard McElreath



- **Richard McElreath** is an American professor of anthropology and a director of the Max Planck Institute for Evolutionary Anthropology in Leipzig, Germany.
- Information on the full course that follows the book can be found at GitHub - rmcelreath/stat_rethinking_2024
- This takes you through the entire book with 10 video lectures (on YouTube) where you step by step learn to build a wide variety of Bayesian models
- The computational engine is the program Stan (by Andrew Gelman and associates) but McElreath provides R code and functions that sit on top of Stan making for ease of use without having to learn another single purpose modeling language.

Enter "Statistical Rethinking" by Richard McElreath



- Each chapter and gives detailed examples of different models and includes R code to easily structure and adapt those models to your own situation.
- The R code provided automatically calls the Stan computational engine to run the Bayesian models.
- R code for extracting posterior results make it easy to not only look at the average results but also examine the variation in results to determine if differences really matter.

Why do we want to build HB Models?

An Example

A Taste Test

Client wishes to test 9 variations of a product

Products include competitive offerings

Want to determine which of the client offerings has the best/most liked taste profile

How do the offerings compare to competitive offerings, including the current market leader.

Nine items are too many for a single consumer to taste and compare – so each respondent will taste 4 of the nine items and rate each item as they taste it.

Photo by Nate Johnston on Unsplash

Avoiding Bias

- It is well known in market research (and in Olympic Skating Judging) that the order in which a stimulus is presented will likely have an effect on the judgement of that stimulus.
- So – we create a balanced test plan. Each product is tasted in each of the 4 positions an equal number of times (in our case 55)

	Position			
	1	2	3	4
Product 1	55	55	55	55
Product 2	55	55	55	55
Product 3	55	55	55	55
Product 4	55	55	55	55
Product 5	55	55	55	55
Product 6	55	55	55	55
Product 7	55	55	55	55
Product 8	55	55	55	55
Product 9	55	55	55	55

Standard analysis vs Bayes

Standard Analysis

- Assume balanced design means that position bias can be ignored.
- Calculate criterion variable for each product by simple averaging.
- If comparison to other studies is required, calculate criterion variable only when the product is shown in the first position

Bayesian Model

- Observed criterion variable is modeled by
  - An overall intercept +
  - A product effect +
  - A position effect +
  - A respondent effect +
  - A product x position interaction effect

But, how do we build these models?

Rethinking Statistics

- The Statistical Rethinking book takes you through increasingly complex Bayesian models for a number of types of data.
- All of them can be fit with a single function in R – ulam which sets the model up for fitting with Stan and then runs it, and returns an R model object.
- Once the model is fit and converged (about 1 minute on my machine) we can get the posterior estimates with one simple command.
- Estimates<-extract.samples(mod1)
- This object provides all of the simulated posterior estimates that I used in this presentation.

The model used in the examples

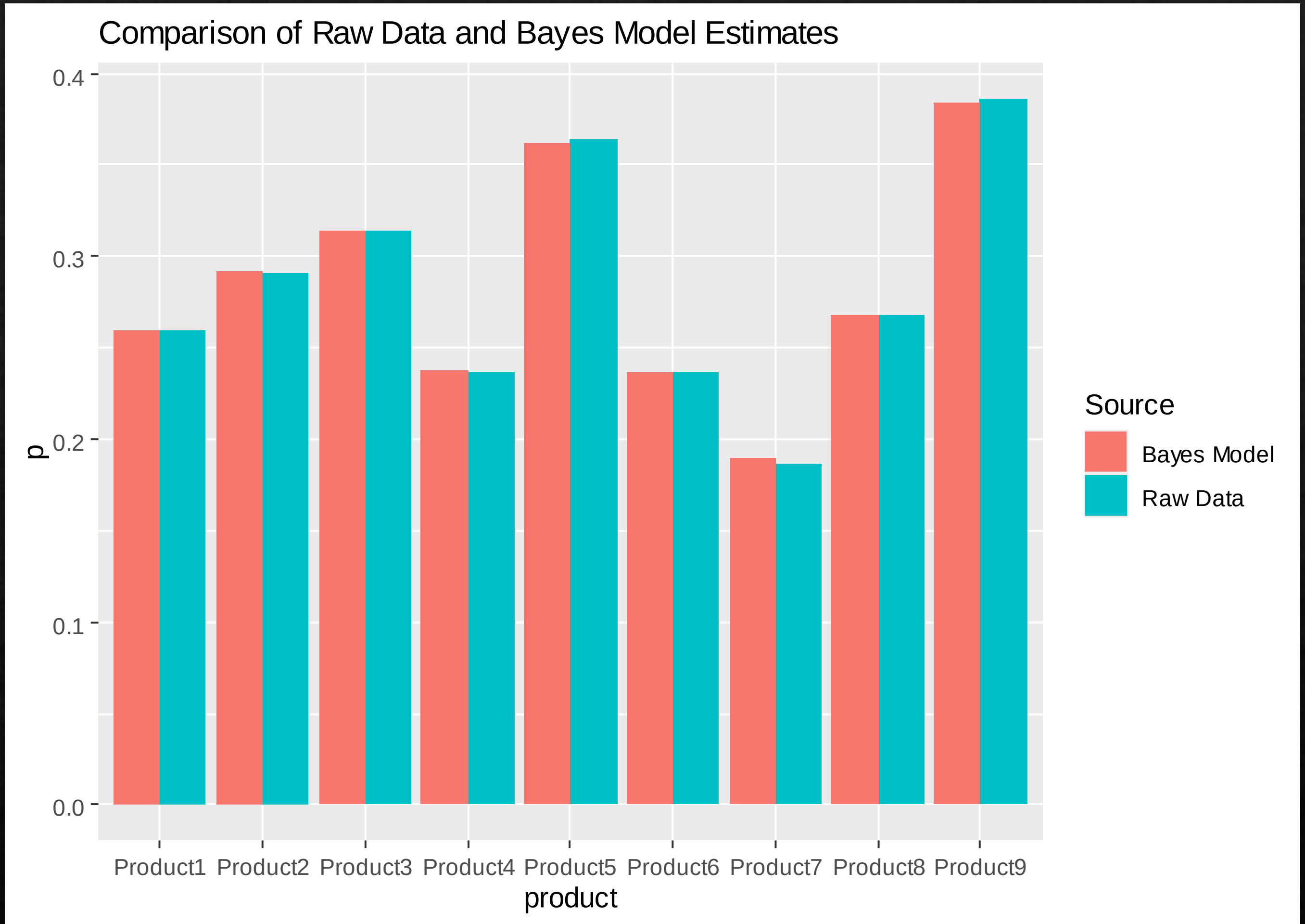
```
dat <- list(  S = d$T3Hedonic1,  product = d$product,  position = d$pos,
resp = d$resp

mod1 <- ulam(  alist(

S ~ dbinom( 1 , p ) ,

logit(p) <-oint+g[product]+a[resp]+b[position,product] ,

#adaptive priors - non centered

transpars> matrix[position,9]:b <-compose_noncentered( sigma_pos ,
L_Rho_pos , z_pos ),matrix[9,position]:z_pos ~ normal( 0 , 1 ),

# fixed priors

oint~dnorm(0,1.5),    g[product]~dnorm(0,1),    sigma_resp~dexp(1),
a[resp]~dnorm(0,1),    vector[9]:sigma_pos~dexp(1),

cholesky_factor_corr[9]:L_Rho_pos~lkj_corr_cholesky(2),

### compute ordinary corr matrix

gq> matrix[9,9]:Rho_pos <<- Chol_to_Corr(L_Rho_pos)       ), data=dat ,
chains=4 , log_lik=TRUE )
```

Standard vs Bayes

The standard approach – The overall criterion variable average for each product is recreated almost exactly by the Bayesian Model.

BUT

With the Bayesian model we can explore all of the individual independent effects built into the model.

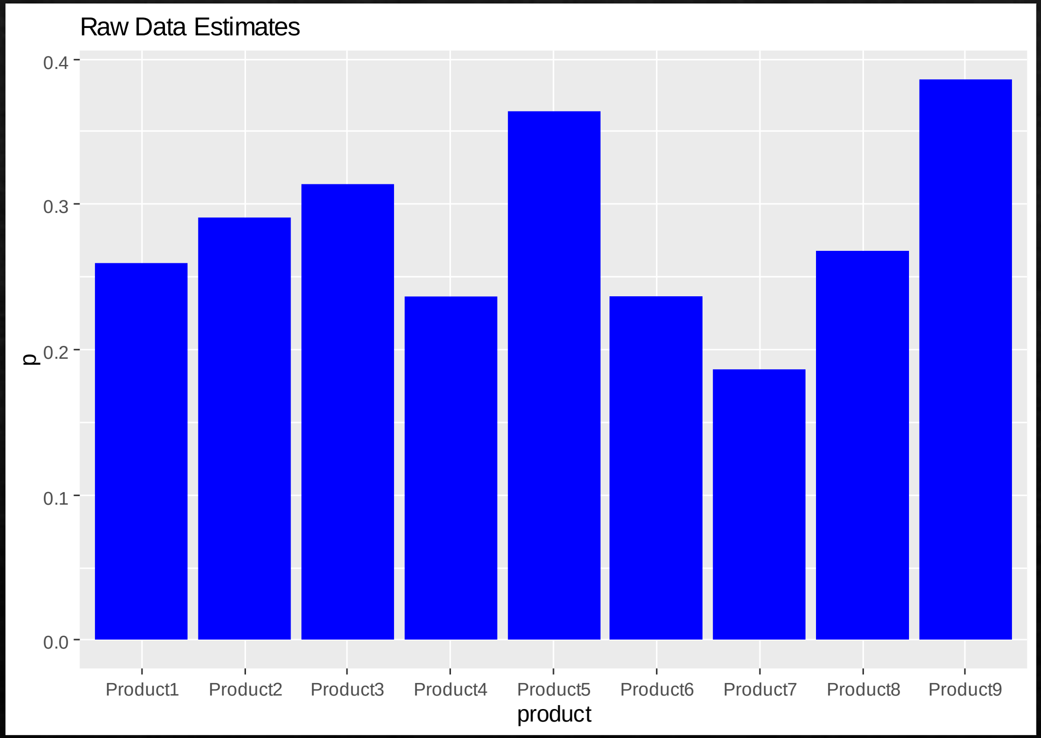


Product 9 and Product 5 appear to be the leaders in performance

Standard Model

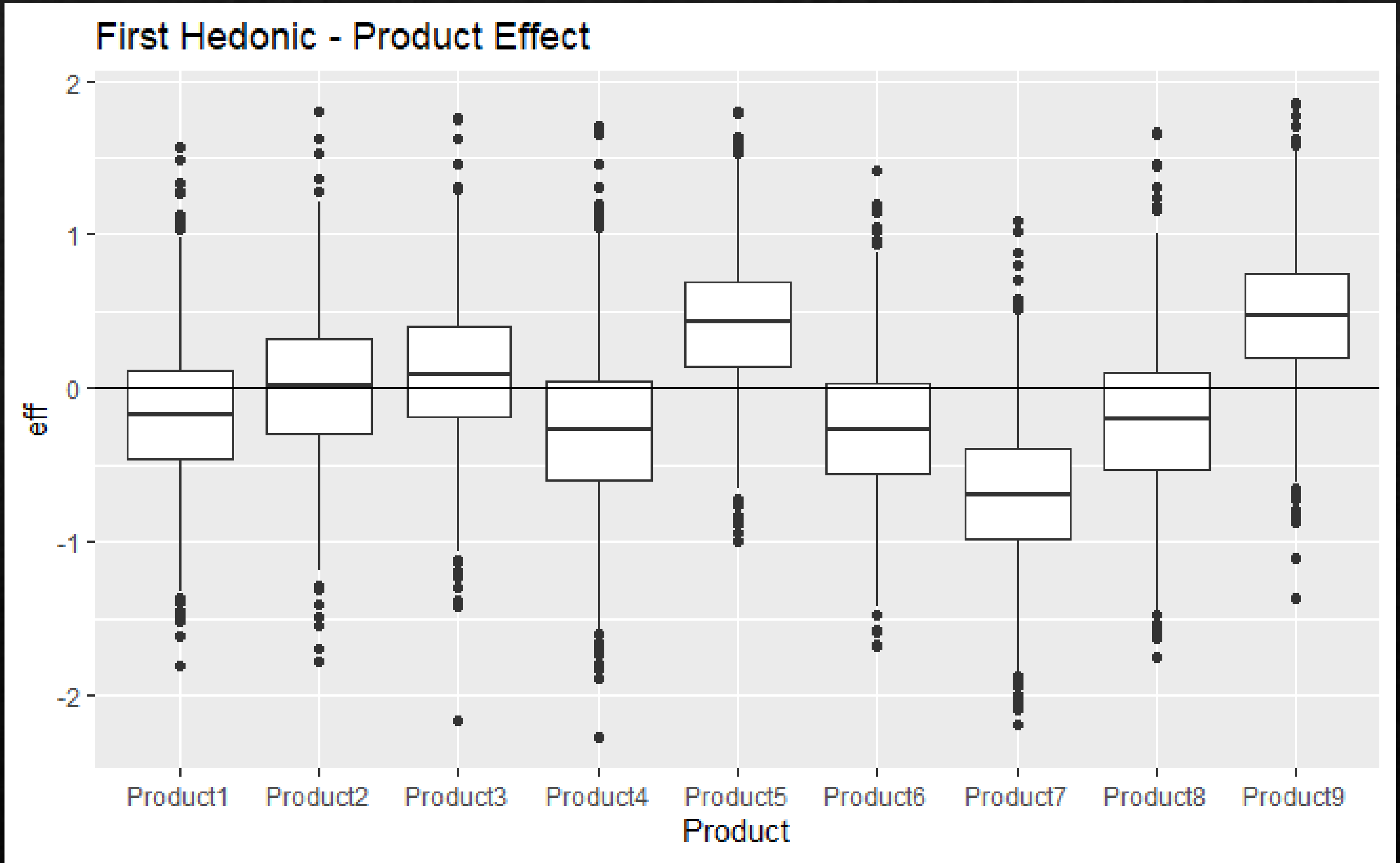
Significance of Multiple Comparison Differences (Tukey HSD)

Product	1	2	3	4	5	6	7	8	9
1	1.00	1.00	0.94	1.00	0.26	1.00	0.74	1.00	0.07
2	1.00	1.00	1.00	0.94	0.74	0.94	0.26	1.00	0.38
3	0.94	1.00	1.00	0.67	0.96	0.67	0.07	0.98	0.74
4	1.00	0.94	0.67	1.00	0.07	1.00	0.96	1.00	0.01
5	0.26	0.74	0.96	0.07	1.00	0.07	0.00	0.38	1.00
6	1.00	0.94	0.67	1.00	0.07	1.00	0.96	1.00	0.01
7	0.74	0.26	0.07	0.96	0.00	0.96	1.00	0.60	0.00
8	1.00	1.00	0.98	1.00	0.38	1.00	0.60	1.00	0.12
9	0.07	0.38	0.74	0.01	1.00	0.01	0.00	0.12	1.00



Here we see 8 significant differences in the raw data between pairs of products

With Bayes we can look at the posterior distribution of Product Effects independent of the position and respondent effects.

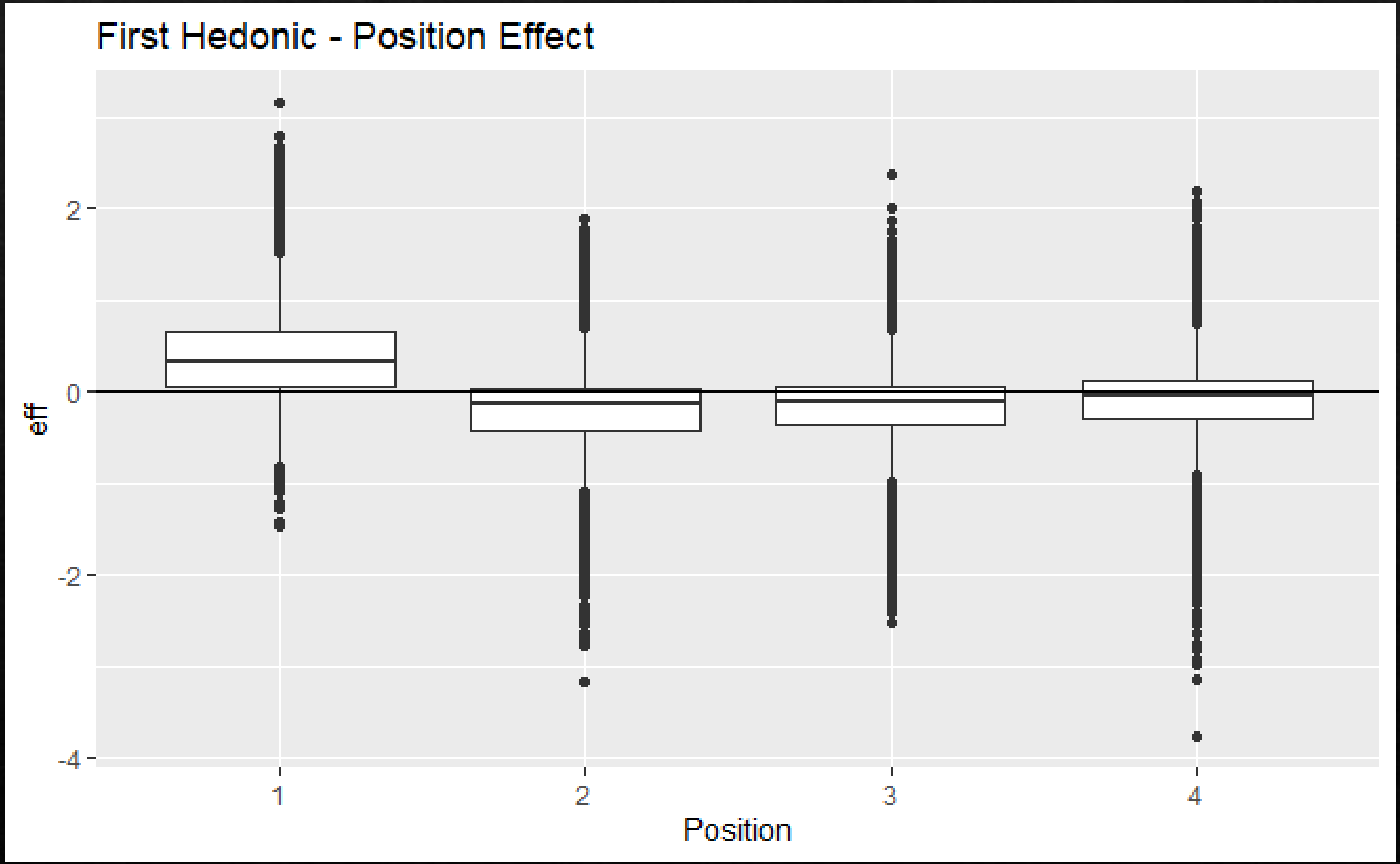


Here we can do multiple comparisons by looking at 2000 simulations of the posterior estimates and calculating how often each product is greater than each other product. By looking at just the independent product effects we can see we get a better estimate of the differences between the products. (12 significant differences)

Multi Compare:Hedonic 1

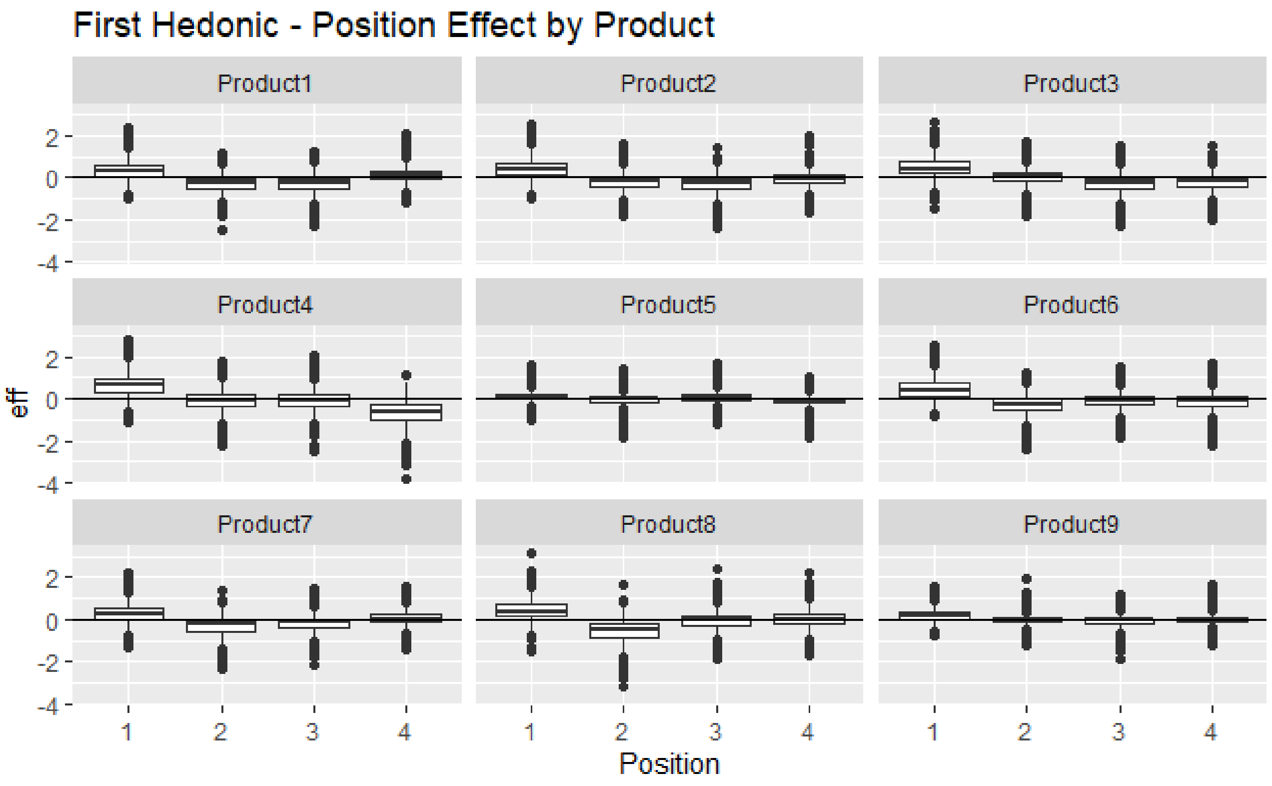
0.5	0.32	0.26	0.59	0.07	0.57	0.88	0.53	0.05
0.68	0.5	0.42	0.72	0.14	0.75	0.94	0.7	0.12
0.74	0.58	0.5	0.79	0.18	0.81	0.96	0.75	0.16
0.41	0.28	0.21	0.5	0.08	0.49	0.81	0.46	0.08
0.93	0.86	0.82	0.94	0.5	0.96	0.99	0.93	0.43
0.43	0.25	0.19	0.51	0.04	0.5	0.83	0.46	0.04
0.12	0.08	0.04	0.19	0.01	0.17	0.5	0.15	0.01
0.47	0.3	0.25	0.54	0.07	0.54	0.85	0.5	0.08
0.95	0.88	0.84	0.94	0.57	0.96	0.99	0.94	0.5
Product1	Product2	Product3	Product4	Product5	Product6	Product7	Product8	Product9

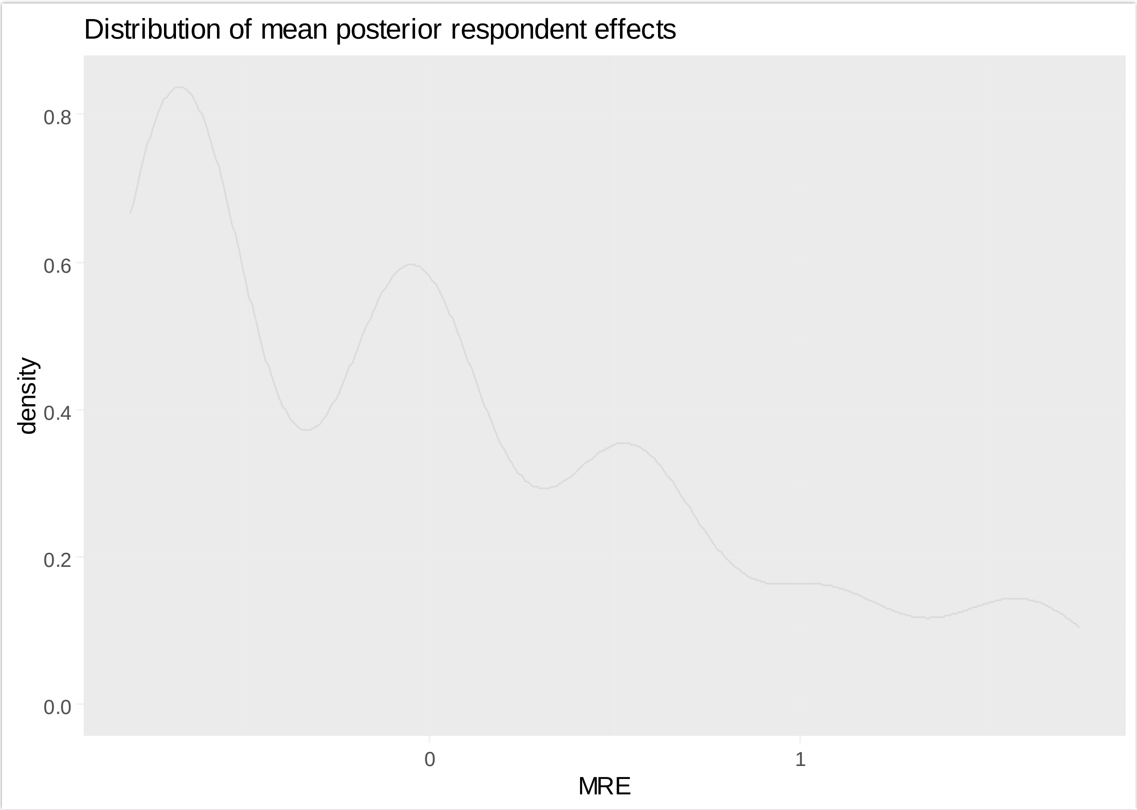
In addition, we can now also test the hypothesis that there is a position effect. Only the first position seems to matter



Is there a product x position interaction?

- Client was concerned that product 9 (which represented the most traditional option in the market) would have a bigger position effect than other products.
- A quick glance at the posterior distribution of the position by product interaction shows there is not a concern.





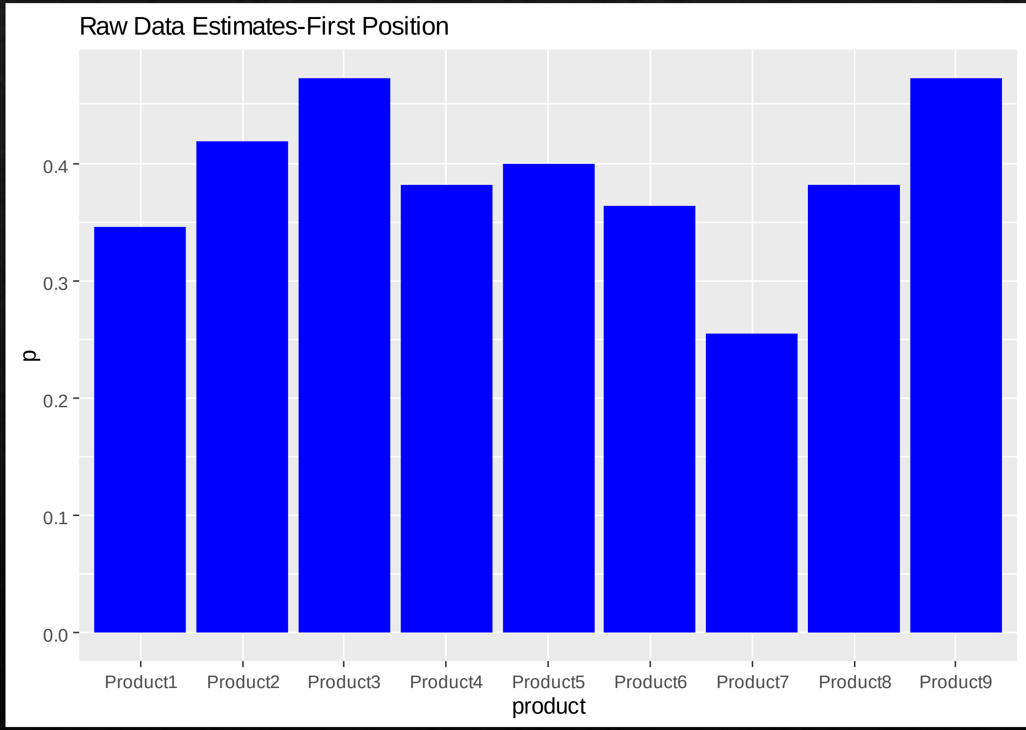
Respondent effects

- A large number of people with lower than average ratings
- A moderate amount with near zero differences
- A long tail of respondents with above average ratings.
- This shows heterogeneity in ratings across the population.

In the standard approach we frequently want to compare to other studies and in other studies single items may be evaluated.
So, we may want to look at the data limited to those products which were seen first.

Tukey HSD Multiple Comparisons

Product	1	2	3	4	5	6	7	8	9
1	1.00	1.00	0.91	1.00	1.00	1.00	0.99	1.00	0.91
2	1.00	1.00	1.00	1.00	1.00	1.00	0.71	1.00	1.00
3	0.91	1.00	1.00	0.99	1.00	0.96	0.32	0.99	1.00
4	1.00	1.00	0.99	1.00	1.00	1.00	0.91	1.00	0.99
5	1.00	1.00	1.00	1.00	1.00	1.00	0.82	1.00	1.00
6	1.00	1.00	0.96	1.00	1.00	1.00	0.96	1.00	0.96
7	0.99	0.71	0.32	0.91	0.82	0.96	1.00	0.91	0.32
8	1.00	1.00	0.99	1.00	1.00	1.00	0.91	1.00	0.99
9	0.91	1.00	1.00	0.99	1.00	0.96	0.32	0.99	1.00



- No significant differences – small sample size (55) seen first is the culprit
- Our sample size does not yield sufficient statistical power to identify any differences

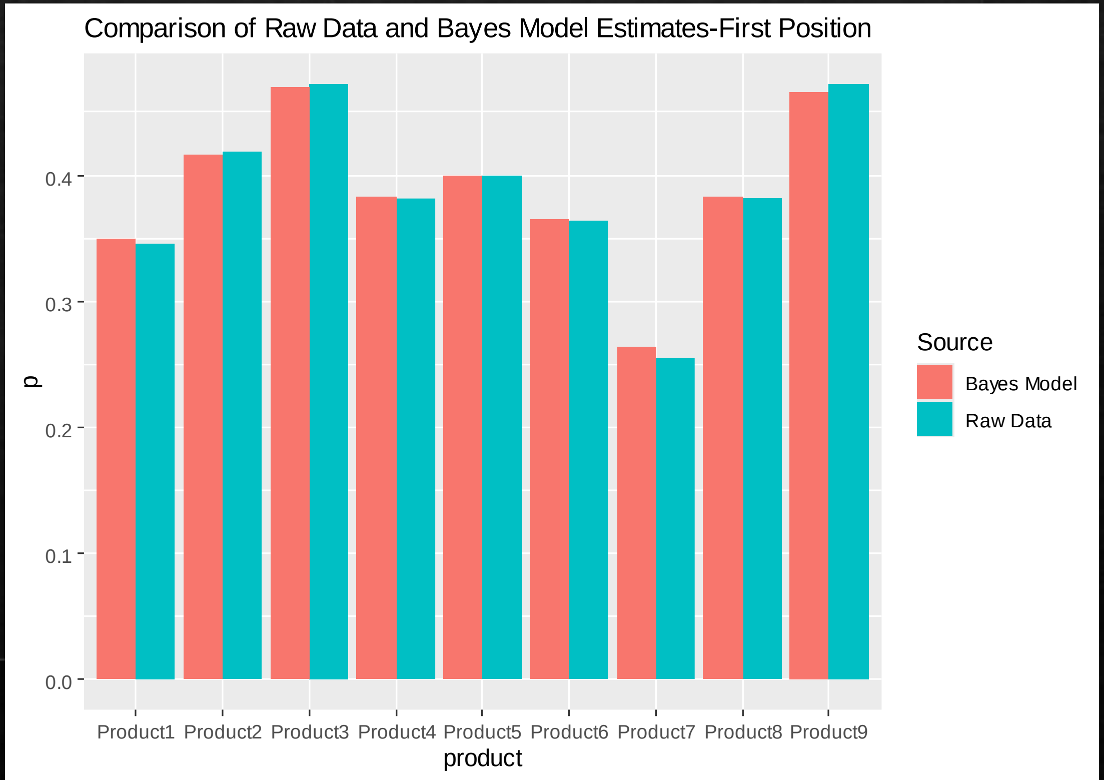
With the Bayesian model we can simulate posterior estimates for first position results

Bayesian model shows 7 significant differences between products (as opposed to 0). Most differences show product 7 to be much worse in performance

Multi Compare:Hedonic 1 - First Position

0.5	0.21	0.07	0.33	0.27	0.43	0.87	0.35	0.08
0.79	0.5	0.27	0.66	0.58	0.72	0.98	0.66	0.28
0.93	0.73	0.5	0.85	0.8	0.9	1	0.85	0.52
0.67	0.34	0.15	0.5	0.42	0.58	0.93	0.5	0.16
0.73	0.42	0.2	0.58	0.5	0.66	0.96	0.58	0.21
0.57	0.28	0.1	0.42	0.34	0.5	0.91	0.41	0.12
0.13	0.02	0	0.07	0.04	0.09	0.5	0.06	0.01
0.65	0.34	0.15	0.5	0.42	0.59	0.94	0.5	0.17
0.92	0.72	0.48	0.84	0.79	0.88	0.99	0.83	0.5
Product1	Product2	Product3	Product4	Product5	Product6	Product7	Product8	Product9

First position model estimates are consistent with observed raw data



Conclusions

- Bayesian models can be easily created using the code and concepts from the Statistical Rethinking book (unfortunately the book does not appear to be available as a free pdf anymore)
- However, the lectures that go along with the book and explain the concepts are available on YouTube.
- The world of Bayesian models lets you explore so much more in your data.

Questions?

If any occur to you later – feel free to email me at

Mike@56stats.com