CS 598 – Advanced Bayesian Modeling – Spring 2025

Data Analysis Report

DUE: May 11, 2025

You will submit a PDF file containing your data analysis report, which must follow the format described below.

Important: You may not collaborate or discuss your analysis with anyone else. Plagiarism from *any* source is an academic integrity infraction.

Scenario: The data file `launches.csv` contains brief information about orbital launch attempts using various types of rocket launch vehicles in the small-lift payload category.[1] Each row represents a different launch attempt, and the columns are as follows:

`Date`	date of the launch attempt (in UTC, when available)
`Vehicle`	name of the type of launch vehicle used
`SinceFirst`	number of days after the first launch attempt for the type of vehicle used (zero for the first launch attempt)
`Success`	whether the attempt was fully successful (1 if so, 0 if not)

Use JAGS and R software, and use only the data in `launches.csv`. JAGS code should be included in the appropriate sections, but **all R code and any direct R text output listings you choose to include should be in the Appendix only**.

Your report must be neatly typed and can be at most **8 pages**, excluding the Appendix. It must follow this outline:

1. **Introduction** Provide brief background information about the space launch industry in recent years, and especially about small-lift launch vehicles and their use to place objects in Earth orbit. Use *at least two* sources, making sure to provide a *full* reference to each. (Footnote or endnote references are acceptable.) **Do not plagiarize!**

2. **Data** Briefly verbally describe and also *statistically* summarize each variable (except `Date`) in `launches.csv`. What overall proportion of launch attempts are successful? What proportion are successful among those that represent the *first attempt* for a type of launch vehicle? Which vehicle type has the most launches? Which type has the greatest apparent success rate?

3. **First Model** You will fit a Bayesian logistic regression model to explain launch success in terms of days since first launch attempt for the type of vehicle being used:

[1] Compiled from this source and pages linked from it: Wikipedia contributors. (2025, February 13). Small-lift launch vehicle. In *Wikipedia, The Free Encyclopedia.* Retrieved 15:57, March 19, 2025, from `https://en.wikipedia.org/w/index.php?title=Small-lift_launch_vehicle&oldid=1275597481`. Only vehicles listed as operational or having a launch attempt within the last decade are included. Only launch attempts initiated prior to March 19, 2025, and intended for orbital flight are included. A launch must be fully successful to qualify as a success. Launches to unintended, unrecoverable orbits generally count as failures.

- The response variable will be Bernoulli: 1 if the launch was successful, 0 if not.
- The model will be a logistic regression.
- The linear portion of the model will be like that of a simple linear regression: There will be an ordinary "intercept" term and a (single) coefficient multiplying the (centered and rescaled) days after first attempt. (Of course, there is no "error" term.) None of the parameters depends on vehicle type, i.e., the same "intercept" and the same "slope" coefficient apply to every vehicle type.
- The independent variable is a centered and rescaled version of `SinceFirst`: centered to have sample mean of zero, and rescaled to have sample standard deviation of 0.5 (*not* 1), as recommended in BDA3.
- As recommended in BDA3, the prior for the "intercept" should be $t_1(0, 10^2)$, the prior for the other coefficient should be $t_1(0, 2.5^2)$, and these should be independent. Note: These distributions are expressed in BDA3 notation. Be careful when converting to JAGS code.

This first model will *not* use any random effect terms. All launch outcomes will be regarded as (conditionally) independent under the sampling model, even among launches that use the same vehicle type.

- For the Bernoulli, use the `dbern` distribution specifier in JAGS.
- You may wish to consult the JAGS manual to make sure that you are correctly using the `dt` distribution specifier (for a t distribution).

Run your analysis (being careful to follow the usual procedures) and report as follows:

(a) List your JAGS code.

(b) Summarize the details of your computation, including number of chains, length of burn-in, number of iterations used per chain, any thinning (if used), and effective sample sizes of all parameters. Although you should use plots to check convergence, do *not* include them in your report.

(c) Approximate the posterior mean, posterior standard deviation, and 95% central posterior interval for each regression coefficient.

(d) Approximate the posterior probability that the "slope" coefficient related to days after first launch attempt is positive. Interpret this result. (In general, does it seem that a type of launch vehicle becomes more reliable over time?)

(e) Consider the probability of a successful launch on a first attempt (for the type of vehicle used). Based on this first model, approximate a 95% central posterior interval for this probability.

(f) Approximate the value of (Plummer's) DIC and its associated effective number of parameters. Compare the effective number of parameters with the actual number of parameters.

4. **Second Model** Now extend the first model by allowing each vehicle type to have a separate additive random effect:

- Starting with the first model (as described previously), add to the linear portion of the model a random effect term that varies by vehicle type, i.e., is the same for all launches of the same vehicle type but can be different for different vehicle types.
- For their prior, let these random effects be (conditionally) independent from a *normal* distribution with mean zero (since the model already has an intercept) and common *standard deviation* σ_{vehicle}.
- Let the hyperprior for σ_{vehicle} be approximately flat on $(0, \infty)$. (You need to determine how to implement this. It may require a preliminary run and some adjustment.)

Run your analysis (being careful to follow the usual procedures) and report as follows:

(a) List your JAGS code.

(b) Summarize the details of your computation, including number of chains, length of burn-in, number of iterations used per chain, any thinning (if used), and effective sample sizes of the top-level parameters. Although you should use plots to check convergence, do *not* include them in your report.

Note: Use overdispersed starting values, but try making them less extreme if you encounter convergence problems.

(c) Approximate the posterior mean, posterior standard deviation, and 95% central posterior interval for:

- the "intercept" term,
- the coefficient of the (centered and rescaled) number of days after first launch attempt, and
- the random effect standard deviation σ_{vehicle}.

(d) Display an approximate *density plot* for σ_{vehicle}. Based on inspection of the plot, would it be reasonable to assume that all launch vehicles are equally reliable (adjusting only for days after first launch attempt)?

(e) Approximate the value of (Plummer's) DIC and its associated effective number of parameters. Is this second model better than the first?

(f) Define the "most reliable" type of launch vehicle as the one *most likely* (under the posterior) to have the *largest* (most positive) random effect. *Name* the type of launch vehicle that is most reliable. Also, what is the approximate (posterior) probability that it has the largest random effect?

5. **Conclusions** Briefly summarize your results in a non-technical manner.

6. **Appendix** Provide the R code you used to conduct your analysis. Include comments that label the purpose of each block of code.

NOTES:

- Comma-separated variable (`.csv`) files can be read into R with `read.csv`.
- Effective sample sizes of at least 10000 are recommended for accuracy.
- If your computer runs out of memory, consider using thinning (e.g., the `thin` argument of `coda.samples`).

Point Allocations

Specifications		
	2	neatly typed
	2	no more than 8 pages (excluding Appendix)
Introduction		
	4	background given
	2	sources acknowledged
Data		
	2	description and summary of variables
	4	questions answered
First Model		
	5	(a)
	4	(b)
	2	(c)
	2	(d)
	1	(e)
	3	(f)
Second Model		
	4	(a)
	4	(b)
	3	(c)
	2	(d)
	3	(e)
	2	(f)
Conclusions		
	3	brief, clearly stated, appropriate summary of results
Appendix		
	2	all R code present
	2	comments for different blocks of code
Total:	58	