USA pollster*

CHANGE SUBTITLE

Chris Yong Hong Sen

October 18, 2024

First sentence. Second sentence. Third sentence. Fourth sentence.

1 Introduction

Overview paragraph

Estimand paragraph

Results paragraph

Why it matters paragraph

Telegraphing paragraph: The remainder of this paper is structured as follows. Section 2….

2 Data

2.1 Overview

We use the statistical programming language R (R Core Team 2023)…. Our data (Toronto Shelter & Support Services 2024)…. Following Alexander (2023), we consider…

Overview text

2.2 Measurement

Some paragraphs about how we go from a phenomena in the world to an entry in the dataset.

*Code and data are available at: https://github.com/Monoji77/USA-pollster.

2.3 Outcome variables

Add graphs, tables and text. Use sub-sub-headings for each outcome variable or update the subheading to be singular.

Some of our data is of penguins (Figure 1), from Horst, Hill, and Gorman (2020).

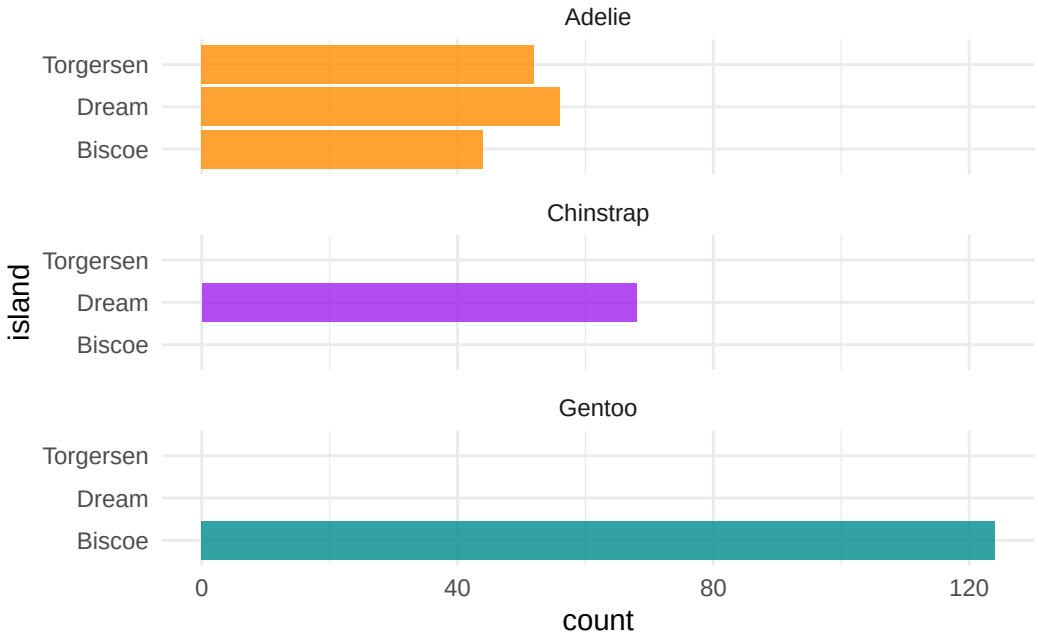


Figure 1: Bills of penguins

Talk more about it.

And also planes (Figure 2). (You can change the height and width, but don't worry about doing that until you have finished every other aspect of the paper - Quarto will try to make it look nice and the defaults usually work well once you have enough text.)

Talk way more about it.

2.4 Predictor variables

Add graphs, tables and text.

Use sub-sub-headings for each outcome variable and feel free to combine a few into one if they go together naturally.

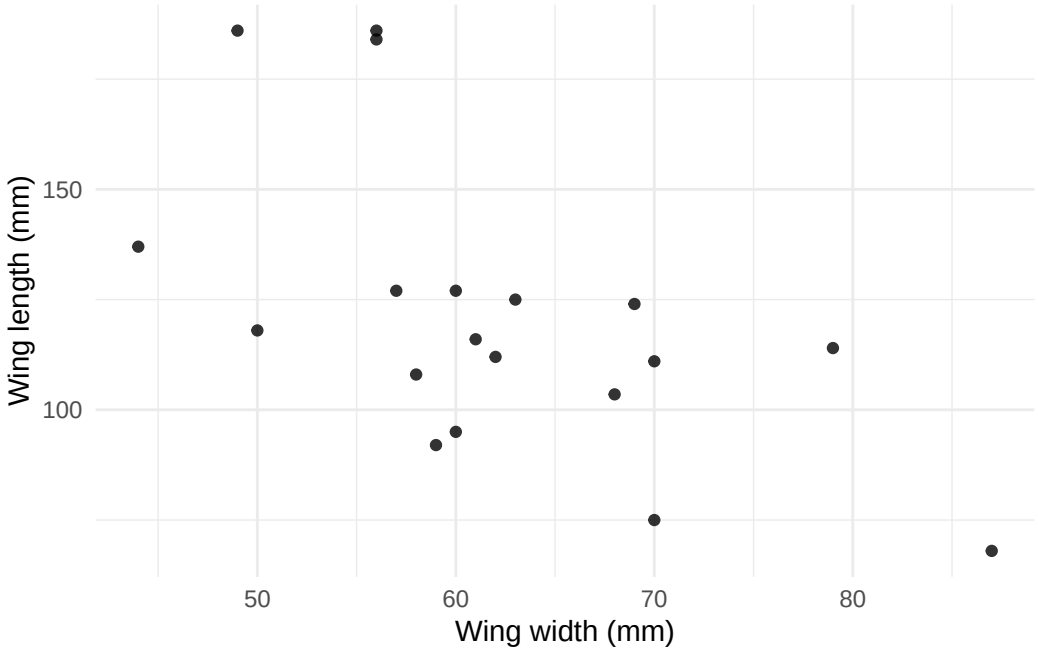


Figure 2: Relationship between wing length and width

3 Model

The goal of our modelling strategy is twofold. Firstly,…

Here we briefly describe the Bayesian analysis model used to investigate… Background details and diagnostics are included in Appendix B.

3.1 Model set-up

Define y_i as the number of seconds that the plane remained aloft. Then β_i is the wing width and γ_i is the wing length, both measured in millimeters.

$$y_i|\mu_i, \sigma \sim \text{Normal}(\mu_i, \sigma) \tag{1}$$
$$\mu_i = \alpha + \beta_i + \gamma_i \tag{2}$$
$$\alpha \sim \text{Normal}(0, 2.5) \tag{3}$$
$$\beta \sim \text{Normal}(0, 2.5) \tag{4}$$
$$\gamma \sim \text{Normal}(0, 2.5) \tag{5}$$
$$\sigma \sim \text{Exponential}(1) \tag{6}$$

Table 1: Explanatory models of flight time based on wing width and wing length

	First model
(Intercept)	1.12
	(1.70)
length	0.01
	(0.01)
width	−0.01
	(0.02)
Num.Obs.	19
R2	0.320
R2 Adj.	0.019
Log.Lik.	−18.128
ELPD	−21.6
ELPD s.e.	2.1
LOOIC	43.2
LOOIC s.e.	4.3
WAIC	42.7
RMSE	0.60

We run the model in R (R Core Team 2023) using the `rstanarm` package of Goodrich et al. (2022). We use the default priors from `rstanarm`.

3.1.1 Model justification

We expect a positive relationship between the size of the wings and time spent aloft. In particular...

We can use maths by including latex between dollar signs, for instance θ.

4 Results

Our results are summarized in Table 1.

5 Discussion

5.1 First discussion point

If my paper were 10 pages, then should be be at least 2.5 pages. The discussion is a chance to show off what you know and what you learnt from all this.

5.2 Second discussion point

Please don't use these as sub-heading labels - change them to be what your point actually is.

5.3 Third discussion point

5.4 Weaknesses and next steps

Weaknesses and next steps should also be included.

Appendix

A Patriot Polling: Wisconsin Presidential Analysis Data Collection

A.1 Methodology

From 12 to 14 October 2024, Patriot Polling has conducted phone surveys in the state of Wisconsin of the United States of America, which borders Minnesota, Iowa, Michegan, and Illinois. Their target population are voters in Wisconsin. For sample of landlines which are characterised by households, an equal number of phone numbers are randomly selected across every landline block in Wisconsin using Random Digit Dialing (RDD). As for the sample of mobile numbers, which are characterised by an individual with access to a digital mobile device, the sample was purchased from a consumer contact number database. For this poll, a total of 803 respondents has completed the survey either from contact through the landline or their personal contact number. (Ruggieri 2024)

Ruggieri (2024) conducted polling in such a way where a randomly contacted respondent would hear pre-recorded voice messages and users are able to interact with the automated phone system. This system is called Interactive voice response (IVR) which can handle outbound calls more systematically since the same questions would be asked sequentially in the survey. There were only 2 questions that were asked. Both questions begin with 'How will you vote for president?'. This is followed by the candidates name as it appears in the voting ballot.

A.2 Strengths

This automated method of phone surveys is a lot more convenient and efficient than in person type of surveys since it removes the need for a surveyor to visit respondents face to face and also removes the need for a surveyor to contact multiple respondents over the phone. This decreases labour effort and cost.

Furthermore, with real-time feedback from phone surveys, the list of questions can be easily and quickly adjusted to ensure the ethical integrity of the survey being conducted. According to Ruggieri (2024), this survey took place in 2 daysm which could have taken a lot longer if other in-person type of surveying methods were used in this research.

Particularly, the polling company used a repertoire of RDD sampling and stratified sampling on landlines to ensure that every landline block (strata) has equal numbers of respondents to be included in the final sample. This helps to ensure diversity from the sample by including different geographical landline blocks, reducing sampling bias. The purchasing of mobile phones from phone number databases abates sampling bias in areas where a particular demographic - younger people who are fully dependant on mobile phones and not landlines - would otherwise be underrepresented if only RDD sampling was conducted on landlines.

A.3 Limitations and weakness

Firstly, there was a lack of information regarding how RDD was performed for the landline sample, how they handled non-working landlines, as well as the total sample size attributed from landline and mobile phone surveys.

The handling of no-response entities were also not explicitly mentioned by Patriot Polling. This could translate to respondents who were chosen but did not pick up their mobile phones/landlines as it was a busy period in the day or due to potential heightened awareness of phone based scams. This means that the final sample obtained might not be representative of the Wisconsin voter population.

Also, the initial phrasing to the second question can very easily confuse respondents. Although the initial phrasing is followed by the senate candidate's name, the initial phrasing 'How will you vote for president' still uses the wrong word 'president' instead of 'senate'. This does not reflect a high quality survey, which could lead to decreased trust from the respondent to Patriot Polling, in turn potentially resulting in a premature end to the surveys, voiding the initial response.

Lastly, the purchasing of phone numbers from a consumer database would introduce of selection bias. Individuals from middle to higher socio-economic status would have greater purchasing power than their lower income counterparts, which means individuals from lower socio-economic status may not have phones. In addition, elderlies within Wisconsin may not necessarily have mobile phones due to the lack of skills to use such technologies. Also, individuals would've opted for do-not-call registries, opted out of marketing databases or simply have prepaid phones which do not link to their personal information. This could mean an under representation of individuals from the said groups.

B Model details

B.1 Posterior predictive check

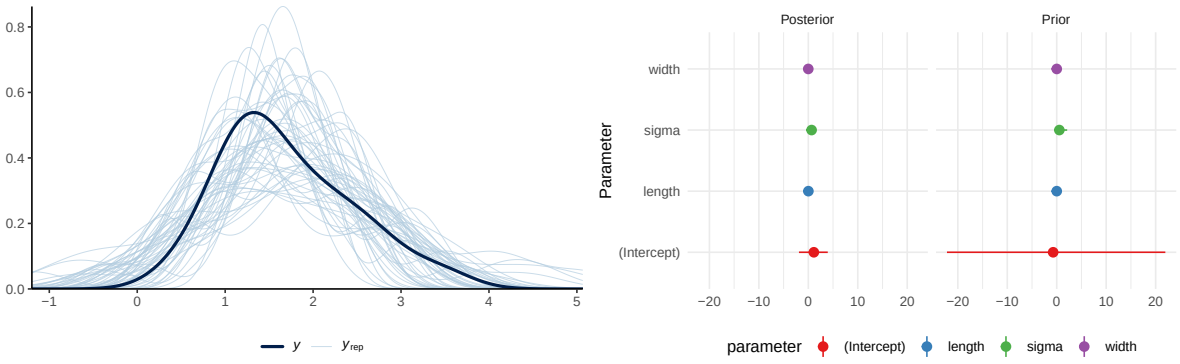
In Figure 3a we implement a posterior predictive check. This shows…

In Figure 3b we compare the posterior with the prior. This shows…

B.2 Diagnostics

Figure 4a is a trace plot. It shows… This suggests…

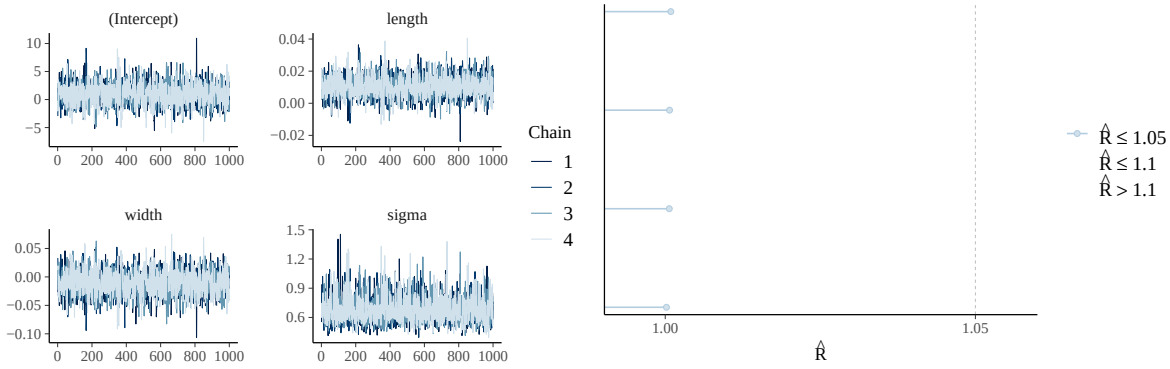
Figure 4b is a Rhat plot. It shows… This suggests…



(a) Posterior prediction check

(b) Comparing the posterior with the prior

Figure 3: Examining how the model fits, and is affected by, the data



(a) Trace plot

(b) Rhat plot

Figure 4: Checking the convergence of the MCMC algorithm

References

Alexander, Rohan. 2023. *Telling Stories with Data*. Chapman; Hall/CRC. https://tellingstorieswithdata.com/.

Goodrich, Ben, Jonah Gabry, Imad Ali, and Sam Brilleman. 2022. "rstanarm: Bayesian applied regression modeling via Stan." https://mc-stan.org/rstanarm/.

Horst, Allison Marie, Alison Presmanes Hill, and Kristen B Gorman. 2020. *palmerpenguins: Palmer Archipelago (Antarctica) penguin data*. https://doi.org/10.5281/zenodo.3960218.

R Core Team. 2023. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing. https://www.R-project.org/.

Ruggieri, Lucca. 2024. *Patriot Polling Data Collection Methodology*. https://patriotpolling.com/our-polls/f/trump-and-baldwin-hold-narrow-leads-in-wisconsin.

Toronto Shelter & Support Services. 2024. *Deaths of Shelter Residents*. https://open.toronto.ca/dataset/deaths-of-shelter-residents/.