MTHM017 - Advanced Topics in Statistics, Assignment

James R Lewis

00-00-2025

Contents

A. Bayesian Inference . . . **2**

1. . . . 2
2. [5 marks] The above model uses the logarithm of measured reaction times. Explain why taking the . . . 5
3. [5 marks] List the parameters of the model and assign non-informative uniform prior distributions to each parameter, paying attention to the values these parameters are allowed take. . . . 6

The model . . . 6

4. . . . 7

Declaration of AI Assistance: I have used OpenAI's ChatGPT tool in creating this report.

AI-supported/AI-integrated use is permitted in this assessment. I acknowledge the following uses of GenAI tools in this assessment:

1. I have used GenAI tools to check and debug my code.
2. I have used GenAI tools to proofread and correct grammar or spelling errors.
3. I have used GenAI tools to give me feedback on a draft.

I declare that I have referenced use of GenAI outputs within my assessment in line with the University referencing guidelines.

A. Bayesian Inference

1.

[6 marks] Read in the data, then for each person produce a histogram of that given person's reaction times. The range of the x axis should be the same on each histogram. Visually compare the reaction time distributions of schizophrenic and non-schizophrenic individuals. What differences/similarities can you observe? Reference the histograms of specific individuals to support your conclusions.

I will begin by making the dataset more usable and clear.

```
# Renaming first column
names(rtimes)[1] <- "PatientType"
# Classifying the patient type, 1-11 non-schiz, 12-17 schiz
rtimes$PatientType <- c(rep("non-schizophrenic", 11), rep("schizophrenic", 6))
rtimes1 <- rtimes
head(rtimes1)
```

```
##         PatientType  T1  T2  T3  T4  T5  T6  T7  T8  T9 T10 T11 T12 T13 T14 T15
## 1 non-schizophrenic 312 272 350 286 268 328 298 356 292 308 296 372 396 402 280
## 2 non-schizophrenic 354 346 384 342 302 312 322 376 306 402 320 298 308 414 304
## 3 non-schizophrenic 256 284 320 274 324 268 370 430 314 312 362 256 342 388 302
## 4 non-schizophrenic 260 294 306 292 264 290 272 268 344 362 330 280 354 320 334
## 5 non-schizophrenic 204 272 250 260 314 308 246 236 208 268 272 264 308 236 238
## 6 non-schizophrenic 590 312 286 310 778 364 318 316 316 298 344 262 274 330 312
##   T16 T17 T18 T19 T20 T21 T22 T23 T24 T25 T26 T27 T28 T29 T30
## 1 330 254 282 350 328 332 308 292 258 340 242 306 328 294 272
## 2 422 388 422 426 338 332 426 478 372 392 374 430 388 354 368
## 3 366 298 396 274 226 328 274 258 220 236 272 322 284 274 356
## 4 276 418 288 338 350 350 324 286 322 280 256 218 256 220 356
## 5 350 272 252 252 236 306 238 350 206 260 280 274 318 268 210
## 6 310 376 326 346 334 282 292 282 300 290 302 300 306 294 444
```

Creating Histograms

We want to extract the data from these columns **rtimes[, 2:31]**, for all 17 patients. We will plot 17 histograms, displaying the distribution of the 30 reaction times.

In order to keep a consistent range across the 17 histograms, we need to find the absolute minimum and maximum values in this dataset. To do this, we extract all the data, 17 rows * 30 trials.

```
DataValues <- unlist(rtimes1[, 2:31])
range(DataValues)
```

```
## [1]  204 1714
```

```
# We want to store these values


x_min <- 204
x_max <- 1714
```

Below, we can observe two sets of histograms, the first containing the eleven non-schizophrenic patients, and the seconds containing the 6 schizophrenic patients.

We reshape to data frame to long format, as having the reaction time data in one column makes much easier to code the histogram. To create a seperate plot for each patient, we use facet_wrap() to group data by PatientID.

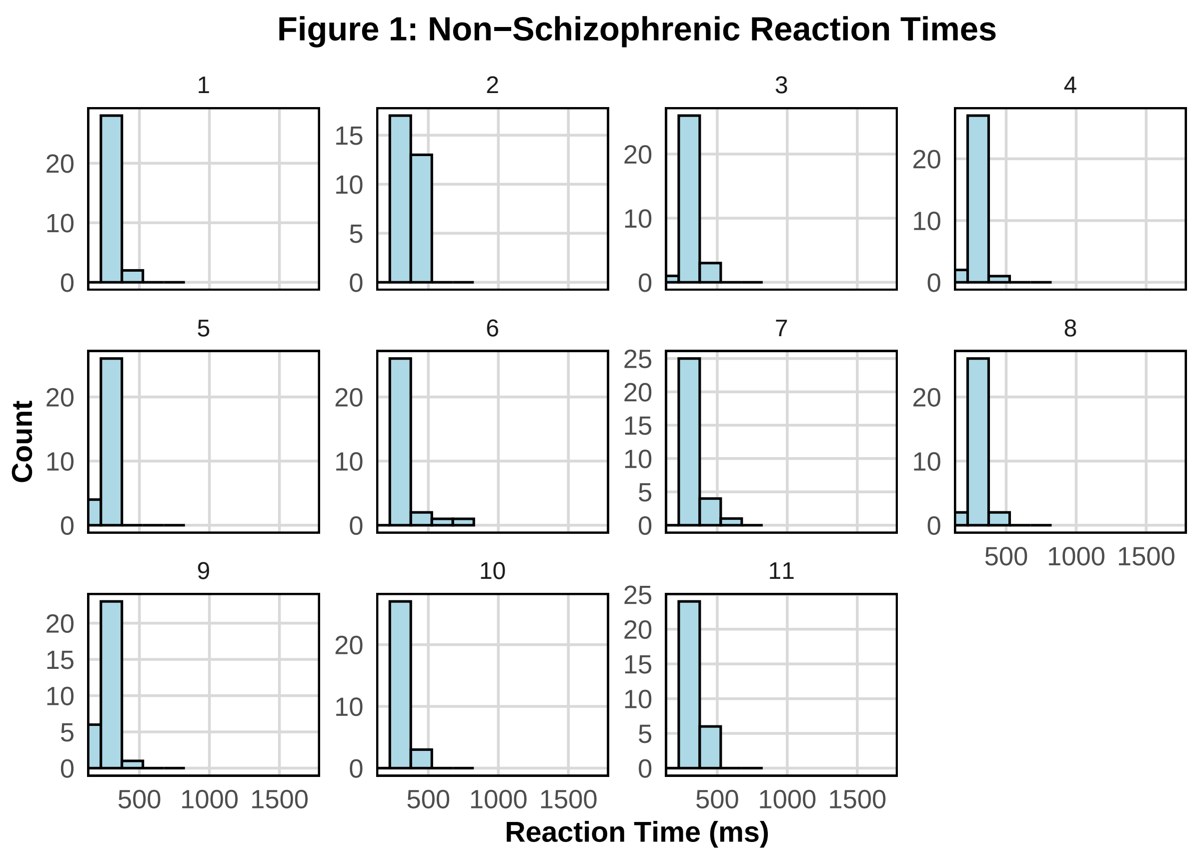
```
# Reshaping to long format
rtimes_long <- rtimes1 %>%
  pivot_longer(cols = starts_with("T"),  # Columns T1 to T30
               names_to = "Trial",
               values_to = "ReactionTime") %>%
  mutate(PatientID = rep(1:17, each = 30)) %>%  # Add patientID
  select(PatientID, PatientType, ReactionTime)

# Splitting into two datasets
non_schizo <- rtimes_long %>% filter(PatientType == "non-schizophrenic")
schizo <- rtimes_long %>% filter(PatientType == "schizophrenic")
```

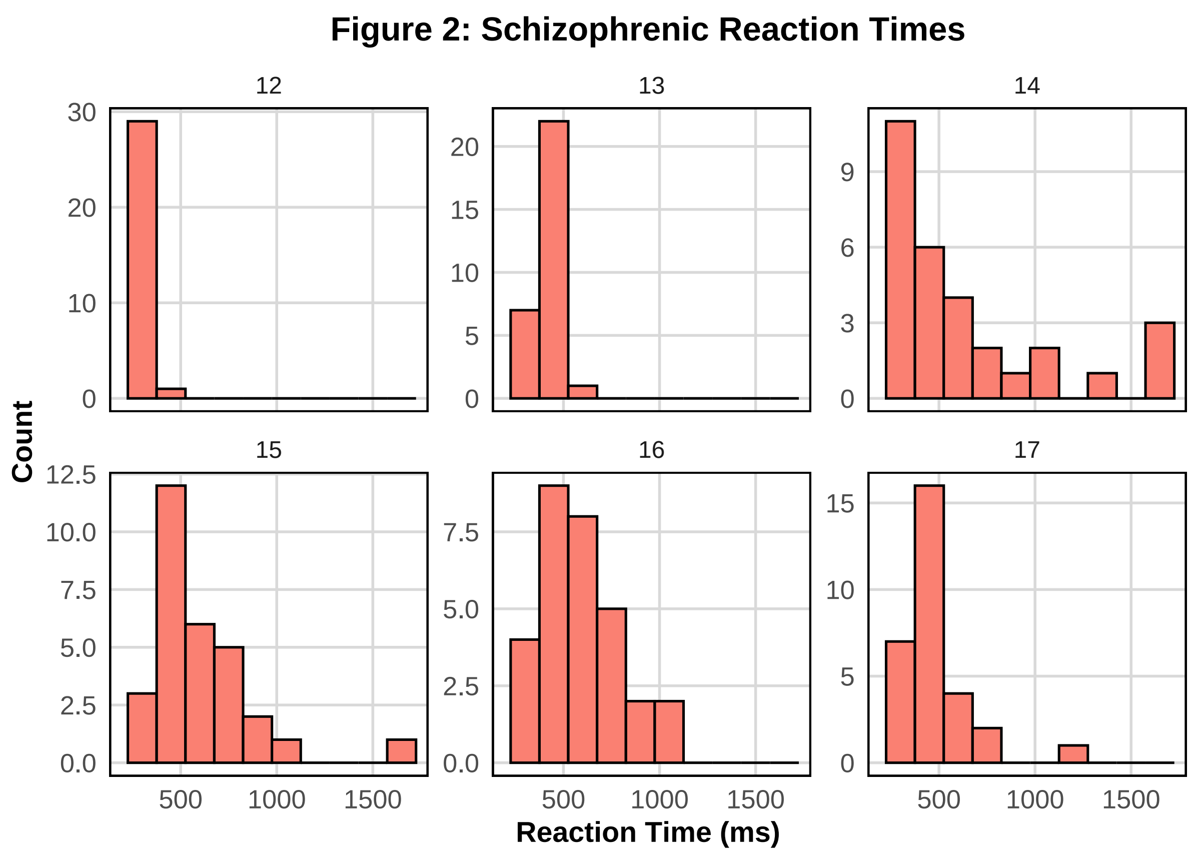
```
# Plotting non-schizophrenic histograms
hist1 <- ggplot(non_schizo, aes(x = ReactionTime)) +
  geom_histogram(binwidth = 150, fill = "lightblue", color = "black") +
  facet_wrap(~ PatientID, ncol = 4, scales = "free_y") +  # 11 plots, 4 columns
  coord_cartesian(xlim = c(x_min, x_max)) +  # Fixed x-axis
  labs(title = "Figure 1: Non-Schizophrenic Reaction Times",
       x = "Reaction Time (ms)",
       y = "Count") +
  custom_theme

# Plotting schizophrenic histograms
hist2 <- ggplot(schizo, aes(x = ReactionTime)) +
  geom_histogram(binwidth = 150, fill = "salmon", color = "black") +
  facet_wrap(~ PatientID, ncol = 3, scales = "free_y") +  # 6 plots, 3 columns
  coord_cartesian(xlim = c(x_min, x_max)) +  # Fixed x-axis
  labs(title = "Figure 2: Schizophrenic Reaction Times",
       x = "Reaction Time (ms)",
       y = "Count") +
  custom_theme


print(hist1)
```




```r
print(hist2)
```



In Figure 1 we can see that Non-schizophrenic patients such as patient 1 have a narrower and concentrated histogram, with reaction times completely clustered around the second bin (150-300ms). For example, every Non-schizophrenic patient's results show that at least two thirds of their reaction times were faster than 300ms.

In contrast, the Schizophrenic patients histograms in Figure 2 exhibit a much wider spread of reaction times, with some patients having reactions times exceeding 1500ms. This indicates greater variability, for instance, patient 14 has a wide range of reaction times with many being greater than 1000ms, whilst also having a significant count below 500ms. This suggests that Schizophrenic patients have a mixture of normal and delayed responses, and this aligns with the theory of attention deficits and motor retardation in schizophrenics. This is generally observed by their inconsistent distributions and right skew, compared to the tighter uniform distributions of non-schizophrenics

Nonetheless, both groups share similarities in their distributions. They both have peaks within the first two bins 0-500ms, as you can see when comparing patient 8 (non) to 13 (schizophrenic). However, only schizophrenics have reaction time values in the higher bands (greater than 800ms)

Overall, these histograms support the hypothesis that schizophrenics experience attention deficits and motor reflex retardation, as seen by the wider spread of values and more irregular histograms for schizophrenic individuals.

2. [5 marks] The above model uses the logarithm of measured reaction times. Explain why taking the

logarithm is necessary here (referencing the relevant output), then perform the transformation yourself. For each person compute the standard deviation of the log transformed reaction times of that individual.

-Calculate standard deviations: For each of the 17 people, compute the standard deviation of their 30 log-transformed reaction times.

Log-transforming the reaction times is necessary for the model above for two main reasons. Firstly, there is skewness in the data as seen in the Schizophrenic reaction times histograms, for example, Patient 14 and 15 are right-skewed distributions with extreme values (over 1500ms), whilst non-schizophrenic's have a tighter distribution. Following this, schizophrenics don't seems to follow a normal distribution, thus, log transforming reduces the skewness and helps stabilise the variance from the effect of outliers/skew. This enambles the data to fit the models normal distributions assumption better. Finally, the log-transformation scales the data, making the mean and variance more interpretable. For example, the original range of 204ms to 1714ms, becomes 5.31 to 7.44 (3 s.f.) when log-transformed. Put simply, handling smaller numbers is easier.

```
log_rtimes_long <- rtimes_long %>%
  mutate(LogReactionTime = log(ReactionTime))
range(log_rtimes_long$LogReactionTime) # checking the range also lets us know if any observ
```

```
## [1] 5.318120 7.446585
```

Now we compute the standard deviation of each patients 30 log-transformed reaction times. This is useful as it measures the within personal variability on the log scale.

```
# Compute standard deviation of log-transformed reaction times for each person
sd_log <- log_rtimes_long %>%
  group_by(PatientID) %>%
  summarise(SD_Log_RT = sd(LogReactionTime))

# Print results
head(sd_log)
```

```
## # A tibble: 6 x 2
##   PatientID SD_Log_RT
##       <int>     <dbl>
## 1         1     0.127
## 2         2     0.129
## 3         3     0.169
## 4         4     0.150
## 5         5     0.145
## 6         6     0.224
```

3. [5 marks] List the parameters of the model and assign non-informative uniform prior distributions to each parameter, paying attention to the values these parameters are allowed take.

The model

Non-Schizophrenic Reaction Times: For the responses of the ith non-schizophrenic person $(i = 1, 2, \dots, 11)$

$$y_{ij} \sim N(\alpha_i, \sigma_y^2), \quad i = 1, 2, \dots, 11, \quad j = 1, 2, \dots, 30$$

Schizophrenic Reaction Times:

For the responses of the ith schizophrenic individual $(i = 12, 13, \dots, 17)$, with probability $(1-\lambda)$ there is no delay, and the response is normally distributed with mean α_i and variance σ_y^2; and with probability λ the response is delayed so that the observations have mean $\alpha_i + \tau$ and variance σ_y^2.

$$y_{ij} \sim N(\alpha_i + \tau z_{ij}, \sigma_y^2)$$

$$z_{ij} \sim \text{Bernoulli}(\lambda), \quad i = 12, 13, \dots, 17, \quad j = 1, 2, \dots, 30$$

Distributions of α:

For non-schizophrenic individuals we assume that α_i follows a normal distribution with mean μ and variance σ_α^2

$$\alpha_i \sim N(\mu, \sigma_\alpha^2), \quad i = 1, 2, \dots, 11$$

For schizophrenics:

$$\alpha_i \sim N(\mu + \beta, \sigma_\alpha^2), \quad i = 12, 13, \dots, 17$$

The parameters of the model, and their non-informative uniform prior distributions:

for Non-specific Individuals:

- α_i (for $i = 1$ to 17): Person-specific means on the log scale.

```
alpha[i] ~ dunif(-10, 10)
```

- μ: Mean of α_i for non-schizophrenics.

```
mu ~ dunif(-10, 10)
```

- $\tau_y = 1/\sigma_y^2$: Precision of log-reaction times.

```
tau_y ~ dunif(0, 10)
```

- $\tau_\alpha = 1/\sigma_\alpha^2$: Precision of α_i.

```
tau_alpha ~ dunif(0, 10)
```

for Schizophrenic Individuals:

- β: Additional variable in α_i mean for schizophrenics.

```
beta ~ dunif(-10, 10)
```

- τ: Delay parameter for schizophrenics ($\tau > 0$).

```
tau ~ dunif(0, 10)
```

- λ: Probability of delay ($0 \leq \lambda \leq 1$).

```
lambda ~ dunif(0, 1)
```

4.