Assignment6\_bayes

mkhulekeli Nkosi 2017159092

2025-05-24

backups/ └── backup\_2025-05-20\_initial\_cleaning/ ├── clean\_data.R ├── data\_cleaned.csv

# Example code used during this phase  
library(tidyverse)

## ── Attaching core tidyverse packages ──────────────────────── tidyverse 2.0.0 ──  
## ✔ dplyr 1.1.4 ✔ readr 2.1.5  
## ✔ forcats 1.0.0 ✔ stringr 1.5.1  
## ✔ ggplot2 3.5.0 ✔ tibble 3.2.1  
## ✔ lubridate 1.9.3 ✔ tidyr 1.3.1  
## ✔ purrr 1.0.2   
## ── Conflicts ────────────────────────────────────────── tidyverse\_conflicts() ──  
## ✖ dplyr::filter() masks stats::filter()  
## ✖ dplyr::lag() masks stats::lag()  
## ℹ Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors

df <- readxl::read\_excel("BayesAssignment6of2025.xlsx", sheet = 1)  
df\_long <- df %>%  
 pivot\_longer(cols = starts\_with("Lecturer"), names\_to = "Lecturer", values\_to = "Mark") %>%  
 drop\_na(Mark)  
write\_csv(df\_long, "cleaned\_data\_long.csv")

# Automatically copy project to a backup folder  
timestamp <- format(Sys.time(), "%Y-%m-%d\_%H-%M")  
d<- dir.create(paste0("Assignment6\_bayes\_1", timestamp))  
file.copy("Assignment6\_bayes", paste0("Assignment6\_bayes\_1", timestamp), recursive = TRUE)

## Warning in file.copy("Assignment6\_bayes", paste0("Assignment6\_bayes\_1", :  
## problem copying .\Assignment6\_bayes to  
## Assignment6\_bayes\_12025-05-24\_12-23\Assignment6\_bayes: No such file or  
## directory

## [1] FALSE

# 1.

The residual error is caused by the different groups that will do presentation and by the different groups of assessors. Also not having enough variables to explain the the final mark.

# 2.

There is always a factor that influences an outcome but according to the above assumptions i would say in this case they are enough for the average assessor mark to be correct on average.

# 3.

data<- readxl::read\_xlsx("BayesAssignment6of2025.xlsx")  
summary(data)

## Group LecturerA LecturerB LecturerC   
## Length:15 Min. :60.00 Min. :49.00 Min. :60.00   
## Class :character 1st Qu.:72.00 1st Qu.:62.00 1st Qu.:63.50   
## Mode :character Median :74.00 Median :68.00 Median :67.50   
## Mean :74.14 Mean :65.33 Mean :69.75   
## 3rd Qu.:76.75 3rd Qu.:70.00 3rd Qu.:77.50   
## Max. :88.00 Max. :82.00 Max. :85.00   
## NA's :1 NA's :6 NA's :3   
## LecturerD LecturerE LecturerF LecturerG Proposal   
## Min. :60.00 Min. :52.00 Min. :53.00 Min. :60.0 Min. :57.00   
## 1st Qu.:68.00 1st Qu.:61.50 1st Qu.:71.75 1st Qu.:64.5 1st Qu.:63.50   
## Median :70.00 Median :68.00 Median :78.00 Median :69.5 Median :74.00   
## Mean :70.50 Mean :67.71 Mean :72.25 Mean :68.0 Mean :71.13   
## 3rd Qu.:76.25 3rd Qu.:76.00 3rd Qu.:78.50 3rd Qu.:73.0 3rd Qu.:78.00   
## Max. :78.00 Max. :79.00 Max. :80.00 Max. :73.0 Max. :84.00   
## NA's :5 NA's :8 NA's :11 NA's :11   
## Literature Quiz Interview   
## Min. :55.0 Min. :48.00 Min. :49.00   
## 1st Qu.:65.5 1st Qu.:66.50 1st Qu.:64.00   
## Median :69.0 Median :75.00 Median :71.00   
## Mean :69.4 Mean :72.47 Mean :68.13   
## 3rd Qu.:74.5 3rd Qu.:80.00 3rd Qu.:72.00   
## Max. :91.0 Max. :85.00 Max. :77.00   
##

sapply(data, class)

## Group LecturerA LecturerB LecturerC LecturerD LecturerE   
## "character" "numeric" "numeric" "numeric" "numeric" "numeric"   
## LecturerF LecturerG Proposal Literature Quiz Interview   
## "numeric" "numeric" "numeric" "numeric" "numeric" "numeric"

colSums(is.na(data))

## Group LecturerA LecturerB LecturerC LecturerD LecturerE LecturerF   
## 0 1 6 3 5 8 11   
## LecturerG Proposal Literature Quiz Interview   
## 11 0 0 0 0

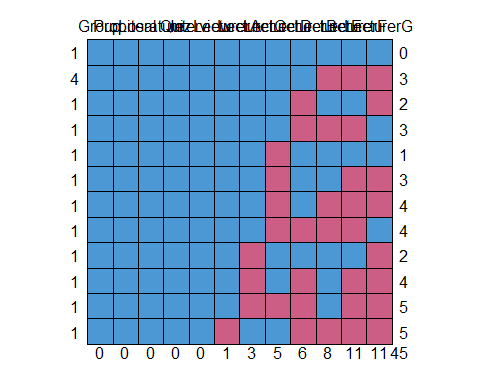
library(mice)

##   
## Attaching package: 'mice'

## The following object is masked from 'package:stats':  
##   
## filter

## The following objects are masked from 'package:base':  
##   
## cbind, rbind

md.pattern(data)



## Group Proposal Literature Quiz Interview LecturerA LecturerC LecturerD  
## 1 1 1 1 1 1 1 1 1  
## 4 1 1 1 1 1 1 1 1  
## 1 1 1 1 1 1 1 1 1  
## 1 1 1 1 1 1 1 1 1  
## 1 1 1 1 1 1 1 1 0  
## 1 1 1 1 1 1 1 1 0  
## 1 1 1 1 1 1 1 1 0  
## 1 1 1 1 1 1 1 1 0  
## 1 1 1 1 1 1 1 0 1  
## 1 1 1 1 1 1 1 0 1  
## 1 1 1 1 1 1 1 0 0  
## 1 1 1 1 1 1 0 1 1  
## 0 0 0 0 0 1 3 5  
## LecturerB LecturerE LecturerF LecturerG   
## 1 1 1 1 1 0  
## 4 1 0 0 0 3  
## 1 0 1 1 0 2  
## 1 0 0 0 1 3  
## 1 1 1 1 1 1  
## 1 1 1 0 0 3  
## 1 1 0 0 0 4  
## 1 0 0 0 1 4  
## 1 1 1 1 0 2  
## 1 0 1 0 0 4  
## 1 0 1 0 0 5  
## 1 0 0 0 0 5  
## 6 8 11 11 45

All the columns are numeric excerpt for the group column. And 45 missing values are observed in total. The missingness patterns given the visualisations on can conclude that the middingness is MAR because the missin values in lecturer E are missing when lecture F And G are missing missingness is dependent on the two variables also lecture D only one is not dependent on G and F.

# 4.

library(tidyr)  
  
long\_data <- pivot\_longer(data, cols= c(LecturerA,LecturerB,LecturerC,LecturerD,LecturerE,LecturerF,LecturerG) ,names\_to = c("Lecturer"), values\_to = "Score")  
  
new\_data<- na.omit(long\_data)  
unique(new\_data$Lecturer)

## [1] "LecturerA" "LecturerC" "LecturerD" "LecturerG" "LecturerB" "LecturerE"  
## [7] "LecturerF"

# 5.

In our case the group of students is our Fixed effect because we not interested in how the next possible group will affect the final mark, with the lecturer as the random effect each group will experience the lecture effect and one would like to know how a different lecturer not included in this fit will grade each group.(read slides)

# 6.

library(brms)

## Loading required package: Rcpp

## Loading 'brms' package (version 2.22.0). Useful instructions  
## can be found by typing help('brms'). A more detailed introduction  
## to the package is available through vignette('brms\_overview').

##   
## Attaching package: 'brms'

## The following object is masked from 'package:stats':  
##   
## ar

model <- brm(  
 formula = Score ~ Group + (1 | Lecturer),  
 data = new\_data,  
 chains = 4,  
 cores = 4,  
 iter = 5000  
)

## Compiling Stan program...

## Start sampling

summary(model)

## Family: gaussian   
## Links: mu = identity; sigma = identity   
## Formula: Score ~ Group + (1 | Lecturer)   
## Data: new\_data (Number of observations: 60)   
## Draws: 4 chains, each with iter = 5000; warmup = 2500; thin = 1;  
## total post-warmup draws = 10000  
##   
## Multilevel Hyperparameters:  
## ~Lecturer (Number of levels: 7)   
## Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS  
## sd(Intercept) 4.26 1.86 1.64 9.16 1.00 2832 4419  
##   
## Regression Coefficients:  
## Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS  
## Intercept 65.33 3.24 59.03 71.66 1.00 1915 3041  
## GroupGroup10 9.97 3.97 1.98 17.69 1.00 2825 3725  
## GroupGroup11 8.70 3.87 1.10 16.20 1.00 2828 3922  
## GroupGroup12 -0.85 4.79 -10.36 8.55 1.00 3781 4813  
## GroupGroup13 9.21 3.85 1.40 16.58 1.00 2745 3916  
## GroupGroup14 5.30 3.71 -2.07 12.52 1.00 2660 3786  
## GroupGroup15 -10.85 4.87 -20.43 -1.35 1.00 3686 5030  
## GroupGroup2 1.60 4.21 -6.77 9.79 1.00 3074 4368  
## GroupGroup3 17.71 3.86 10.02 25.30 1.00 2905 4381  
## GroupGroup4 -1.58 3.57 -8.60 5.23 1.00 2530 3530  
## GroupGroup5 11.07 3.43 4.23 17.70 1.00 2420 3797  
## GroupGroup6 6.94 3.86 -0.89 14.29 1.00 2787 3921  
## GroupGroup7 6.52 4.15 -1.55 14.75 1.00 3332 5196  
## GroupGroup8 -8.62 3.72 -16.00 -1.32 1.00 2625 3922  
## GroupGroup9 7.08 4.23 -1.31 15.31 1.00 3134 4647  
##   
## Further Distributional Parameters:  
## Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS  
## sigma 5.26 0.64 4.18 6.69 1.00 5259 5931  
##   
## Draws were sampled using sampling(NUTS). For each parameter, Bulk\_ESS  
## and Tail\_ESS are effective sample size measures, and Rhat is the potential  
## scale reduction factor on split chains (at convergence, Rhat = 1).

est<- posterior\_samples(model)

## Warning: Method 'posterior\_samples' is deprecated. Please see ?as\_draws for  
## recommended alternatives.

pred<- posterior\_predict(model)  
ranef(model)

## $Lecturer  
## , , Intercept  
##   
## Estimate Est.Error Q2.5 Q97.5  
## LecturerA 3.6862783 2.183302 -0.4367882 8.0686844  
## LecturerB -4.9191263 2.507412 -10.2009962 -0.2917148  
## LecturerC -1.4982572 2.215579 -6.1519345 2.7002700  
## LecturerD -0.4088307 2.272173 -5.0327086 4.1048868  
## LecturerE 0.4929658 2.402929 -4.2948937 5.3693488  
## LecturerF 3.6093151 2.785582 -1.3988498 9.4964334  
## LecturerG -0.8913357 2.677825 -6.4729067 4.2208766

fixed\_effects <- fixef(model, summary = TRUE)   
pred<- posterior\_epred(model)  
fit<-fitted(model)  
pred\_vals <- predict(model, summary = TRUE)  
data<- cbind(new\_data$Group,fit,pred\_vals)  
data<- data[,-6]

# 7.