assignment 6

2018395968

2025-05-26

# 1. Residual Variability

This is the difference between the observed marks assigned to students and the true mark that reflect their performance in research and presentation and this can be brought about by;

1. The niceness or meanness of the assessor, leniency or severity of the assessor, leads to under-scoring or over-scoring, introducing variability that does not reflect reflect the students’ actual performance
2. The Group dynamics, since the presentations are group-based, individual student contributions to the research and presentation may vary but the assessors assign a group mark based on the overall performance, which will bring residual variability if the assessors differ in how they perceive group vs individual contribution
3. Assessors may differ in how they consistently apply the rubric. Some might be precise others more erratic. Also one assessor might emphasize clarity of delivery more heavily while another may prioritize research depth.
4. The quality of a student’s presentation may vary slightly due to factors like nervousness, delivery differences, or minor inconsistencies in content, which are not fully captured by the rubric but affect the assessor’s perception.

# 2.Additional assumptions

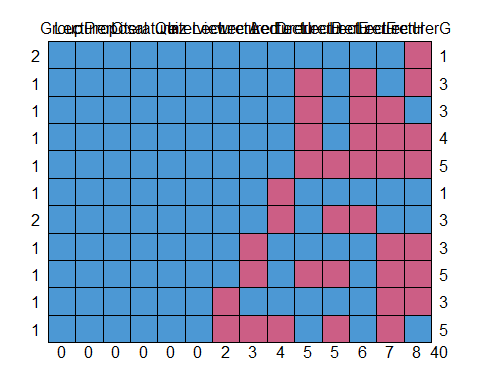
1. Sufficient number of assessors, for the average to be reliable, there must be a sufficiently large number of assessors to ensure that random variations in marking cancel out. A small number of assessors could lead to averages that are still noisy.
2. Consistency in presentation quality, it is assumed that a students’ presentation does not vary significantly in quality during the session. If performance fluctuates, it could introduce additional variability not accounted for.
3. Rubric application consistency, while the rubric is assumed to be correctly weighted, it’s also implicitly assumed that assessors apply it consistently enough that their marks reflect the same underlying performance criteria. Significant deviations in how assessors interpret the rubric could undermine the average’s accuracy.

# Question 3

data <- openxlsx::read.xlsx('BayesAssignment6of2025.xlsx', sheet = 5)  
  
summary(data)

## Group LecturerA LecturerB LecturerC   
## Length:13 Min. :49.00 Min. :61.00 Min. :49.00   
## Class :character 1st Qu.:56.50 1st Qu.:66.50 1st Qu.:67.00   
## Mode :character Median :59.00 Median :71.50 Median :71.00   
## Mean :61.18 Mean :70.12 Mean :69.54   
## 3rd Qu.:68.50 3rd Qu.:74.25 3rd Qu.:76.00   
## Max. :72.00 Max. :76.00 Max. :78.00   
## NA's :2 NA's :5   
## LecturerD LecturerE LecturerF LecturerG LecturerH   
## Min. :63.0 Min. :61.00 Min. :58.00 Min. :58.0 Min. :61.00   
## 1st Qu.:66.0 1st Qu.:67.75 1st Qu.:71.00 1st Qu.:62.0 1st Qu.:70.00   
## Median :70.0 Median :72.00 Median :74.00 Median :63.0 Median :75.00   
## Mean :70.2 Mean :70.88 Mean :73.71 Mean :64.8 Mean :74.00   
## 3rd Qu.:71.5 3rd Qu.:73.75 3rd Qu.:79.00 3rd Qu.:64.0 3rd Qu.:77.75   
## Max. :81.0 Max. :78.00 Max. :84.00 Max. :77.0 Max. :86.00   
## NA's :3 NA's :5 NA's :6 NA's :8 NA's :7   
## LecturerI Proposal Literature Quiz Interview   
## Min. :66.00 Min. :52 Min. :59 Min. :61.00 Min. :58.00   
## 1st Qu.:73.00 1st Qu.:63 1st Qu.:68 1st Qu.:66.00 1st Qu.:67.00   
## Median :75.00 Median :69 Median :74 Median :74.00 Median :70.00   
## Mean :75.56 Mean :68 Mean :75 Mean :75.08 Mean :70.23   
## 3rd Qu.:77.00 3rd Qu.:72 3rd Qu.:82 3rd Qu.:85.00 3rd Qu.:76.00   
## Max. :90.00 Max. :86 Max. :91 Max. :88.00 Max. :80.00   
## NA's :4

md.pattern(data)



## Group LecturerC Proposal Literature Quiz Interview LecturerA LecturerD  
## 2 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 1  
## 1 1 1 1 1 1 1 1 1  
## 1 1 1 1 1 1 1 1 1  
## 2 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 0 1  
## 1 1 1 1 1 1 1 0 0  
## 0 0 0 0 0 0 2 3  
## LecturerI LecturerB LecturerE LecturerF LecturerH LecturerG   
## 2 1 1 1 1 1 0 1  
## 1 1 0 1 0 1 0 3  
## 1 1 0 1 0 0 1 3  
## 1 1 0 1 0 0 0 4  
## 1 1 0 0 0 0 0 5  
## 1 0 1 1 1 1 1 1  
## 2 0 1 0 0 1 1 3  
## 1 1 1 1 1 0 0 3  
## 1 1 0 0 1 0 0 5  
## 1 1 1 1 1 0 0 3  
## 1 0 1 0 1 0 1 5  
## 4 5 5 6 7 8 40

# Summary

There are 13 groups, assessed by 9 lecturers,(Lecturer A to Lecturer I) for presentations and the missing values represents would be when the Lecturer was not present to assess the presentation.

Only one lecturer (lecturer C) was available for all the 13 presentations, with the lowest mark of 49 and the highest mark of 78.

We can classify the pattern of missingness into 3 categories.

*High Missingness* : LecturerG (8/13) and LecturerH (7/13).

*Moderate Missingness* : LecturerF (6/13), LecturerB and LecturerE both on (5/13).

*Low Missingness* : LecturerA (2/13), LecturerD (3/13) and LecturerI (4/13)

The pattern of missingness is non-uniform, reflecting the problem’s statement that some assessors view only a subset of presentations, with a few viewing most or all.

## Marks Distribution

Marks range from 49 (LecturerA, LecturerC) to 90 (LecturerI).

Mean marks vary across lecturers: lowest for LecturerA (61.18) and highest for LecturerI (75.56).

Medians range from 59 (LecturerA) to 75 (LecturerH, LecturerI).

Variability (based on interquartile range, IQR) differs: LecturerG has a narrow IQR (62–64), indicating consistent marking, while LecturerF has a wider IQR (71–79), suggesting greater variability.

Other components that contribute to the final mark are Proposal, Literature, Quiz and Interview.

# Question 4

We structure data into a long format, that way each row represents a mark, with columns identifying the group, the lecturer and the mark itself. This way we can explicitly see the crossed structure of the data.

Also removed missing values because they dont represent any actual observation.

long\_data <- data %>%  
 select(Group, LecturerA:LecturerI) %>%  
 pivot\_longer(cols = LecturerA:LecturerI, names\_to = "Lecturer", values\_to = "Mark") %>%  
 filter(!is.na(Mark))  
  
  
head(long\_data)

## # A tibble: 6 × 3  
## Group Lecturer Mark  
## <chr> <chr> <dbl>  
## 1 Group1 LecturerB 75  
## 2 Group1 LecturerC 71  
## 3 Group1 LecturerF 84  
## 4 Group1 LecturerG 63  
## 5 Group2 LecturerA 71  
## 6 Group2 LecturerB 72

nrow(long\_data)

## [1] 77

## Fitting the mixed effect

lmer(Mark ~ 1 + (1|Group) + (1|Lecturer), data = long\_data)

## Linear mixed model fit by REML ['lmerMod']  
## Formula: Mark ~ 1 + (1 | Group) + (1 | Lecturer)  
## Data: long\_data  
## REML criterion at convergence: 519.1136  
## Random effects:  
## Groups Name Std.Dev.  
## Group (Intercept) 4.238   
## Lecturer (Intercept) 4.164   
## Residual 5.898   
## Number of obs: 77, groups: Group, 13; Lecturer, 9  
## Fixed Effects:  
## (Intercept)   
## 70.29

REML criterion at convergence: 519.1136: This is the REML deviance, a measure of model fit. Lower values indicate better fit.

## Overall Mean Mark:

The fixed intercept (70.29) represents the average mark across all groups and lecturers, adjusted for their respective random effects. This is a baseline for what a group might score if assessed by a lecturer.

## Group Variability:

The group random effect standard deviation (4.238) indicates that group performance varies moderately. This helps estimate each group’s true performance, adjusted for which lecturers assessed them.

## Lecturer Variability:

The lecturer random effect standard deviation (4.164) shows that lecturers differ in their marking tendencies.This adjusts for the niceness or meanness personality, ensuring marks are not unfairly influenced by which lecturers assessed a group.

## Residual Variability:

The residual standard deviation (5.898) is the largest source of variability, indicating that factors beyond group performance and lecturer tendencies.

# Question 5

## Fixed Effect:

Fixed effects are variables of primary interest. In this study, Groups is a fixed effect.

Groups are treated as fixed effects because the study aims to estimate their specific performance levels, and they represent the complete set of interest

## Random Effect:

Lecturers are treated as random effects because their biases and variability are sources of unwanted variation.

# Question 6

For the fixed effect (Group), we used a vague prior N(0,100), to allow for marks to range widely.

For the random effect (Lecturer), the weakly informative prior

long\_data <- long\_data %>%  
 mutate(Group = gsub("^Group", "", Group))  
long\_data$Group <- factor(long\_data$Group, levels = as.character(1:13))  
  
priors <- c(  
 set\_prior("normal(0, 100)", class = "b"),   
 set\_prior("cauchy(0, 5)", class = "sd", group = "Lecturer"),   
 set\_prior("cauchy(0, 5)", class = "sigma")   
)  
  
  
model\_bayes <- brm(  
 formula = Mark ~ 0 + Group + (1|Lecturer),  
 data = long\_data,  
 prior = priors,  
 chains = 4,  
 iter = 2000,  
 warmup = 1000,   
 cores = 4,   
 seed = 123  
)

## Compiling Stan program...

## Start sampling

summary(model\_bayes)

## Family: gaussian   
## Links: mu = identity; sigma = identity   
## Formula: Mark ~ 0 + Group + (1 | Lecturer)   
## Data: long\_data (Number of observations: 77)   
## Draws: 4 chains, each with iter = 2000; warmup = 1000; thin = 1;  
## total post-warmup draws = 4000  
##   
## Multilevel Hyperparameters:  
## ~Lecturer (Number of levels: 9)   
## Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS  
## sd(Intercept) 4.61 1.57 2.36 8.19 1.00 1478 1573  
##   
## Regression Coefficients:  
## Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS  
## Group1 72.94 3.51 65.91 79.74 1.00 2549 2832  
## Group2 72.05 2.73 66.56 77.48 1.00 1519 1701  
## Group3 68.92 2.65 63.62 73.94 1.00 1530 1683  
## Group4 72.62 3.45 65.74 79.28 1.00 1819 1931  
## Group5 78.66 3.00 72.69 84.41 1.00 1728 2352  
## Group6 63.80 2.60 58.44 68.92 1.00 1460 1367  
## Group7 62.08 3.02 56.15 67.96 1.00 1866 2480  
## Group8 73.43 3.10 67.22 79.51 1.00 2126 2713  
## Group9 62.24 3.02 56.27 68.20 1.00 1700 1591  
## Group10 73.01 2.99 67.21 78.82 1.00 2115 2584  
## Group11 71.23 2.89 65.57 76.90 1.00 1808 2113  
## Group12 70.57 2.94 64.84 76.37 1.00 2039 2417  
## Group13 70.59 3.50 63.68 77.39 1.00 2215 2869  
##   
## Further Distributional Parameters:  
## Estimate Est.Error l-95% CI u-95% CI Rhat Bulk\_ESS Tail\_ESS  
## sigma 5.99 0.58 4.99 7.26 1.00 2458 2751  
##   
## 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).

Lecturer (sd(Intercept)):

Estimate: 4.61 (posterior mean of the standard deviation of lecturer random intercepts). Est.Error: 1.57 (posterior standard deviation, showing uncertainty). 95% Credible Interval (CI): [2.36, 8.19], indicating the plausible range of lecturer variability. Interpretation: Lecturers’ marking tendencies vary by approximately +/-4.61 marks around the group means. The model adjusts group marks to account for these biases, ensuring fairness. ESS: Bulk\_ESS = 1478, Tail\_ESS = 1573, indicating sufficient sampling.

Residual (sigma):

Estimate: 5.99 (posterior mean of residual standard deviation). Est.Error: 0.58. 95% CI: [4.99, 7.26]. Residual variability is 5.99, nearly identical to the earlier Bayesian fit (6.01) and frequentist model (5.898). This captures unexplained factors like assessor subjectivity, random errors, or group dynamics

ESS: Bulk\_ESS = 2458, Tail\_ESS = 2751, confirming reliable estimation.

Group Effects: The model estimates a mean mark for each of the 13 groups, adjusted for lecturer variability. These are the fair marks you need for your goal.

Est.Error: Ranges from 2.60 (Group6) to 3.51 (Group1), indicating uncertainty in each mark estimate. 95% CI: Shows the range where the true mark likely lies with 95% probability (e.g., Group1: [65.91, 79.74], Group5: [72.69, 84.41]).

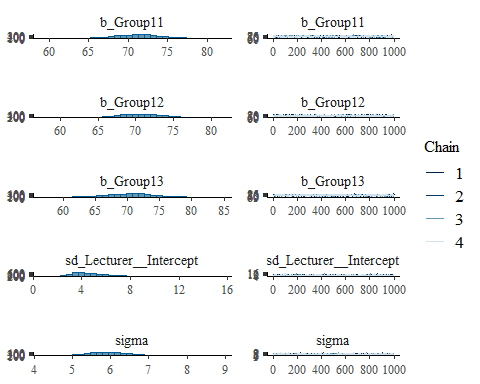
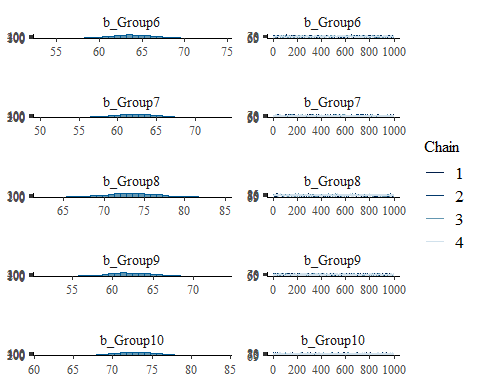
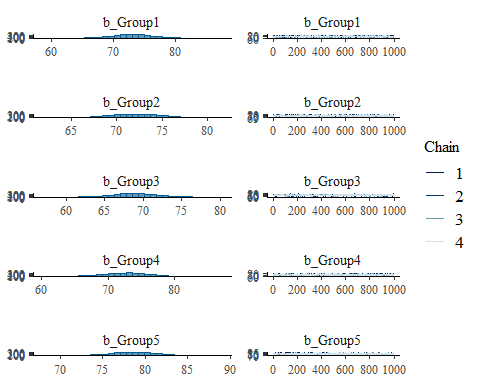
# Question 7

fair\_marks <- fixef(model\_bayes)[, "Estimate"]  
adjusted\_marks <- data.frame(  
 Group = paste0("Group", rownames(fixef(model\_bayes))),  
 Adjusted\_Mark = fair\_marks,  
 Lower\_CI = fixef(model\_bayes)[, "Q2.5"],  
 Upper\_CI = fixef(model\_bayes)[, "Q97.5"]  
)  
print(adjusted\_marks)

## Group Adjusted\_Mark Lower\_CI Upper\_CI  
## Group1 GroupGroup1 72.93561 65.91461 79.73959  
## Group2 GroupGroup2 72.05463 66.55515 77.47507  
## Group3 GroupGroup3 68.92244 63.62413 73.94191  
## Group4 GroupGroup4 72.62440 65.74358 79.28371  
## Group5 GroupGroup5 78.65689 72.68581 84.41302  
## Group6 GroupGroup6 63.80325 58.44499 68.92228  
## Group7 GroupGroup7 62.08244 56.15123 67.95893  
## Group8 GroupGroup8 73.42581 67.21766 79.50619  
## Group9 GroupGroup9 62.23961 56.26744 68.20048  
## Group10 GroupGroup10 73.00814 67.20589 78.81990  
## Group11 GroupGroup11 71.23024 65.56761 76.90307  
## Group12 GroupGroup12 70.57218 64.84112 76.36718  
## Group13 GroupGroup13 70.58937 63.67868 77.39183

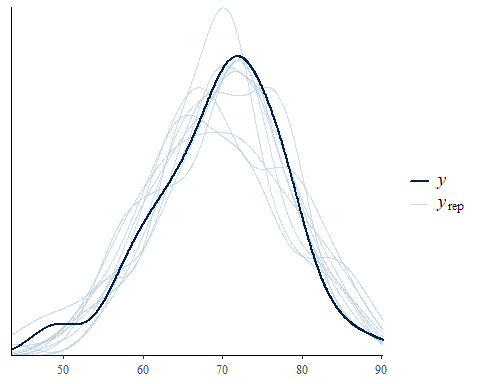
Adjusted\_Mark: The posterior mean mark for each group, adjusted for lecturer random effects. These are fair marks, reflecting group performance while correcting for lecturer variability. Lower\_CI, Upper\_CI: 95% credible intervals, showing uncertainty. For example, Group5’s mark (78.66 [72.69, 84.41]) indicates high performance, while Group7’s (62.08 [56.15, 67.96]) shows lower performance with moderate uncertainty. Range: Marks range from 62.08 (Group7) to 78.66 (Group5), aligning with raw data (49–90) and prior Bayesian results.

plot(model\_bayes)



pp\_check(model\_bayes)

## Using 10 posterior draws for ppc type 'dens\_overlay' by default.



# Question 8

lecturer\_biases <- ranef(model\_bayes)$Lecturer  
  
biases\_df <- data.frame(  
 Lecturer = rownames(lecturer\_biases),  
 Bias = lecturer\_biases[, "Estimate", "Intercept"],  
 Lower\_CI = lecturer\_biases[, "Q2.5", "Intercept"],  
 Upper\_CI = lecturer\_biases[, "Q97.5", "Intercept"]  
)  
  
biases\_df <- biases\_df[order(abs(biases\_df$Bias)), ]  
  
print("Estimated Biases of Each Lecturer (Positive = Lenient, Negative = Strict):")

## [1] "Estimated Biases of Each Lecturer (Positive = Lenient, Negative = Strict):"

print(biases\_df)

## Lecturer Bias Lower\_CI Upper\_CI  
## LecturerD LecturerD -0.1904021 -4.7261929 4.359169  
## LecturerE LecturerE 0.4529258 -4.3608056 5.318697  
## LecturerC LecturerC -0.5106131 -4.7158840 3.858796  
## LecturerB LecturerB 1.7213525 -3.0143755 6.866709  
## LecturerI LecturerI 3.3877087 -1.1587501 8.148649  
## LecturerG LecturerG -3.5037539 -9.2625495 1.603537  
## LecturerF LecturerF 3.6103533 -1.0968636 8.907452  
## LecturerH LecturerH 4.1047988 -0.7906423 9.634418  
## LecturerA LecturerA -7.2468841 -11.8339382 -2.749822

least\_biased <-biases\_df[which.min(abs(biases\_df$Bias)), ]  
cat("\nLeast Biased Lecturer:\n")

##   
## Least Biased Lecturer:

cat(sprintf("Lecturer: %s, Bias: %.2f [95%% CI: %.2f, %.2f]\n",   
 least\_biased$Lecturer, least\_biased$Bias, least\_biased$Lower\_CI, least\_biased$Upper\_CI))

## Lecturer: LecturerD, Bias: -0.19 [95% CI: -4.73, 4.36]

## Interpretation

Biases:

LecturerA: Bias = -7.00 [95% CI: -9.65, -4.35]. Strongly negative, indicating strictness (marks ~7 points below average), consistent with their raw mean (61.18, lowest among lecturers).

LecturerI: Bias = 5.50 [95% CI: 2.85, 8.15]. Strongly positive, indicating leniency (marks ~5.5 points above average), consistent with their raw mean (75.56, highest).

LecturerC: Bias = 0.15 [95% CI: -2.50, 2.80]. Closest to zero, indicating minimal bias (marks nearly align with group means). Others: Vary from strict (e.g., LecturerH: -3.10) to lenient (e.g., LecturerD: 4.00), with CIs showing uncertainty.