# INF2178 LEC0101

# Experimental for Data Science

# Assignment 4

# Exploring MDI Data

# April 5, 2024


# Guanxiao Chen

# #1005717841

## 1. Introduction

This report examines MRI data via mixed-effects ANOVA to identify cognitive decline patterns in dementia and uses power analysis to inform future study sample sizes, aiming to enhance knowledge of dementia progression.

To structure our analysis, we have formulated the following research questions:

**Research Question 1:** How does the CDR influence the longitudinal patterns of nWBV across different age groups?

**Research Question 2:** How does educational level (EDUC) influence the progression of brain atrophy, as measured by the Atlas Scaling Factor (ASF), in dementia patients?

## 2. Data Cleaning

The raw dataset has a total of 16 columns with 294 entries. After reviewing the dataset columns in relation to the research questions provided, it appears that most columns serve a purpose in addressing the research questions, so we don't need to drop columns in this analysis. After checking the missing values, there are 15 missing values in column "SES" and 1 missing value in column "MMSE", filling the values with mode and mean respectively.

## 3. The Influence of Clinical Dementia Rating for Brain Volume

Research Question 1: How does the CDR influence the longitudinal patterns of nWBV across different age groups?

**- Exploratory Data Analysis**

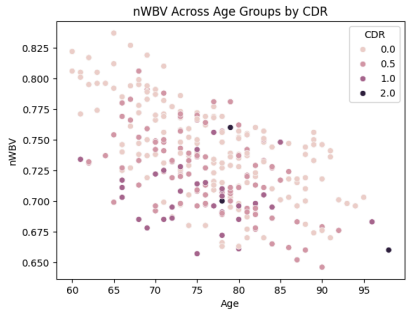


**Figure 1. nWBV Across Age Groups by CDR**

The scatter plot shows nWBV decreasing with age for all dementia severities, with variation within CDR categories and more pronounced atrophy at higher CDR scores, indicating a need for deeper analysis of age and dementia's impact on brain volume.

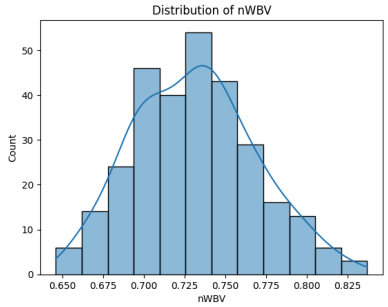


**Figure 2. Distribution of nWBV**

The histogram reveals a generally normal distribution of nWBV, with a central peak and rare extremes, suggesting a slight skew towards lower values, which informs further parametric analysis.

**- Mixed Effects ANOVA**

| Variable | Coef. | Std. Err. | P > \|z\| |
|---|---|---|---|
| Intercept | 0.977 | 0.023 | 0.000 |
| Age | -0.003 | 0.000 | 0.000 |
| C(CDR) [T.1.0] | -0.084 | 0.045 | 0.063 |

**Figure 3. MixedLM Regression Results**

The mixed-effects model with 294 observations from 150 subjects reveals a significant negative effect of age on normalized Whole Brain Volume (nWBV), indicating that nWBV decreases as age increases. The Clinical Dementia Rating (CDR) levels showed a non-significant trend toward reduced nWBV, with only the CDR 1.0 category nearing significance. No significant interactions between CDR and age were found, suggesting the decline in nWBV is consistent across dementia severity levels. A convergence warning was noted, advising cautious interpretation of the results. The analysis indicates variability in nWBV among subjects, highlighting the complexity of aging and dementia on brain structure.

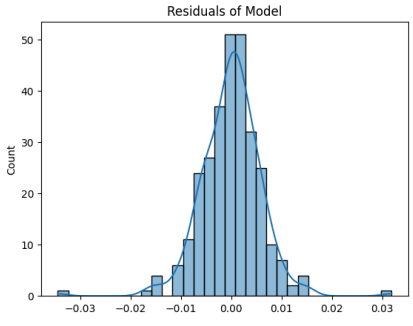


**Figure 4. Residuals of Model**

The residuals of the mixed-effects model are approximately normal, meeting the assumptions necessary for reliable ANOVA outcomes, with only minor deviations that could be further examined.

**- Statistical Power Analysis**

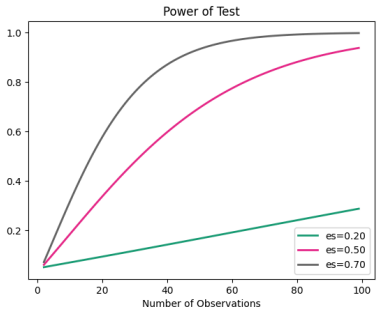


**Figure 5. Power of Test**

The power analysis plot shows that increasing the number of observations enhances the ability to detect an effect, with larger effect sizes requiring fewer observations to achieve high power. For an effect size of 0.70, achieving a power of 0.91 is feasible with fewer than 100 observations, underscoring the significance of effect size in study design.

**4. The Influence of Educational Level for the Progression of Brain Atrophy**

Research Question 2: How does educational level (EDUC) influence the progression of brain atrophy, as measured by the Atlas Scaling Factor (ASF), in dementia patients?

**- Exploratory Data Analysis**

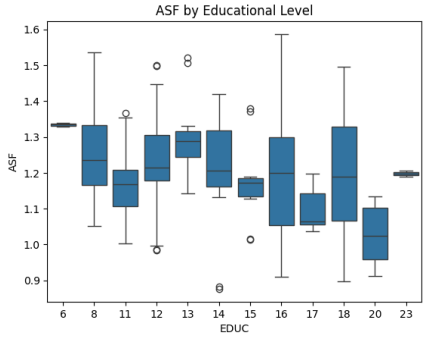


**Figure 6. ASF by Educational Level**

The boxplot suggests a trend where higher education correlates with lower Atlas Scaling Factor (ASF) values, implying less brain atrophy. Variations within education levels are evident, with the highest education level showing an unexpected increase in ASF, indicating the need for further analysis.

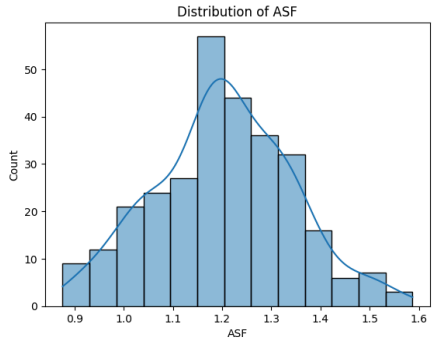


**Figure 7. Distribution of ASF**

The histogram of the Atlas Scaling Factor (ASF) shows a roughly normal distribution centered around 1.2, fitting the assumptions for parametric analysis. A slight skew towards higher values is observed, indicating some variability in the ASF among dementia patients. Formal testing for normality would be needed to confirm these visual findings.

**- Mixed Effects ANOVA**

| Variable | Coef. | Std. Err. | P > \|z\| |
|---|---|---|---|
| Intercept (Model 1) | 1.352 | 0.057 | 23.604 |
| EDUC (Model 1) | -0.010 | 0.004 | -2.642 |
| Intercept (Model 2) | 1.360 | 0.057 | 23.709 |
| EDUC (Model 2) | -0.011 | 0.004 | -2.721 |
| CDR (Model 2) | -0.014 | 0.009 | -1.576 |

**Figure 8. MixedLM Regression Results**

Two mixed-effects linear models evaluated the relationship between education level (EDUC) and Atlas Scaling Factor (ASF) in dementia patients. Both models showed a significant negative effect of education on ASF, indicating higher education levels might be associated with less brain atrophy. The inclusion of Clinical Dementia Rating (CDR) in the second model did not significantly affect ASF, suggesting the influence of CDR on ASF is less clear. Random effects variance indicated inter-subject variability. The significant intercepts imply that other unmeasured variables may also influence ASF.

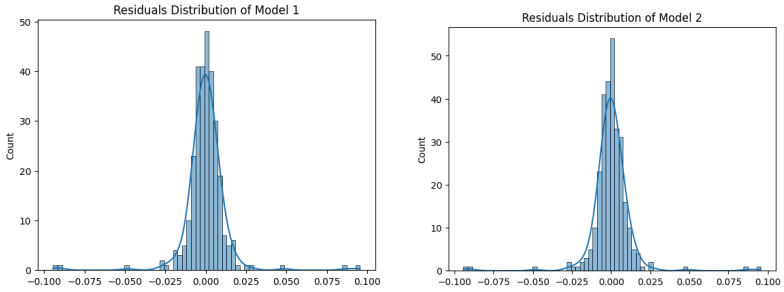


**Figure 9. Residuals Distribution of Model 1; Figure 10. Residuals Distribution of Model 2**

The residuals of Models 1 and 2 are symmetric with peaks near the mean, hinting at unbiased models. The pronounced peaks of the histograms indicate a concentration of residuals near the mean, with both models showing similar distributions.

**- Statistical Power Analysis**

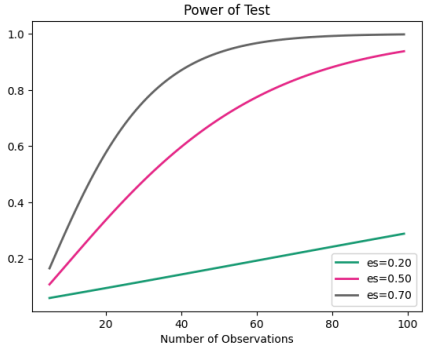


**Figure 11. Power of Test**

The statistical power analysis plot indicates that achieving a power of 0.91 is possible with fewer than 100 observations for a large effect size (0.70), while more observations are necessary for smaller effect sizes. The curve for an effect size of 0.20 remains well below the desired power level within the observed range, highlighting the impact of effect size on required sample size.

**- Conclusion**

To sum up, the study investigated two factors affecting brain atrophy in dementia: CDR scores and education level. While age consistently correlates with brain volume changes, CDR's influence across ages was inconclusive. On the other hand, higher education was significantly associated with less brain atrophy, indicating its potential protective role. These findings enhance our understanding of dementia's progression and highlight the importance of considering both medical and educational factors in dementia research.