# Homework 5

Brian Cervantes Alvarez
May 8, 2024
ST 553 Statistical Methods

## Part 1

Based on the graphs we examined, it seems there are issues with the assumptions typically made for an analysis of variance. Firstly, the spread of the data isn't consistent throughout, as seen in the residuals plotted against predicted values. Additionally, the data doesn't perfectly follow a normal pattern, particularly evident in the Q-Q plot. These findings suggest that adjustments, like transforming the response variable or using alternative techniques, might be needed to ensure the assumptions for CRD are met.

## Part 2

Let $\mu$ be the overall mean, $\alpha_i$ be the effect of manufacturer $i$ (where $i = A, B, C$), $\beta_j$ be the effect of stress level $j$ (where $j =$ low, medium, high), $\gamma_{ij}$ be the interaction effect between manufacturer $i$ and stress level $j$. The model is:

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_{ij} + \epsilon_{ijk}$$

## Part 3

### CRD Model Assumptions

- **Independence** Each measurement must be independent of the others. Again, this is important because the analysis assumes that the behavior of one unit does not affect another.
- **Normality** The model assumes that errors are normally distributed with a mean of zero and a constant variance ($\epsilon_{ijk} \sim N(0, \sigma^2)$).
- **Constant Variance:** The variance of the response variable should be the same across all treatments and levels.
- **Sum-to-Zero Constraints:** The effects of each factor are assumed to sum to zero ($\sum \alpha_i = 0$, $\sum \beta_j = 0$, and for each level of $i$, $\sum \gamma_{ij} = 0$). This prevents redundancy in the parameters, ensuring each one has a distinct and interpretable effect on the response variable.

## Part 4

The vector $\beta$ is,

$$\beta = [\mu, \alpha_A, \alpha_B, \alpha_C, \beta_{\text{low}}, \beta_{\text{med}}, \beta_{\text{high}}, \gamma_{A\text{low}}, \gamma_{A\text{med}}, \gamma_{A\text{high}}, \gamma_{B\text{low}}, \gamma_{B\text{med}}, \gamma_{B\text{high}}, \gamma_{C\text{low}}, \gamma_{C\text{med}}, \gamma_{C\text{high}}]^T$$

Here are the unique rows of the matrix $X$, each corresponding to a different combination of manufacturer and stress level,

| $\mu$ | $\alpha_A$ | $\alpha_B$ | $\alpha_C$ | $\beta_{\text{low}}$ | $\beta_{\text{med}}$ | $\beta_{\text{high}}$ | $\gamma_{A\text{low}}$ | $\gamma_{A\text{med}}$ | $\gamma_{A\text{high}}$ | $\gamma_{B\text{low}}$ | $\gamma_{B\text{med}}$ | $\gamma_{B\text{high}}$ | $\gamma_{C\text{low}}$ | $\gamma_{C\text{med}}$ | $\gamma_{C\text{high}}$ |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 |
| 1 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 |
| 1 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 |
| 1 | 0 | 0 | 1 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |

## Part 5

Dependent Variable: life_log

| Source | DF | Sum of Squares | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| Model | 8 | 12.58691131 | 1.57336391 | 18.56 | <.0001 |
| Error | 27 | 2.28856347 | 0.08476161 | | |
| Corrected Total | 35 | 14.87547478 | | | |

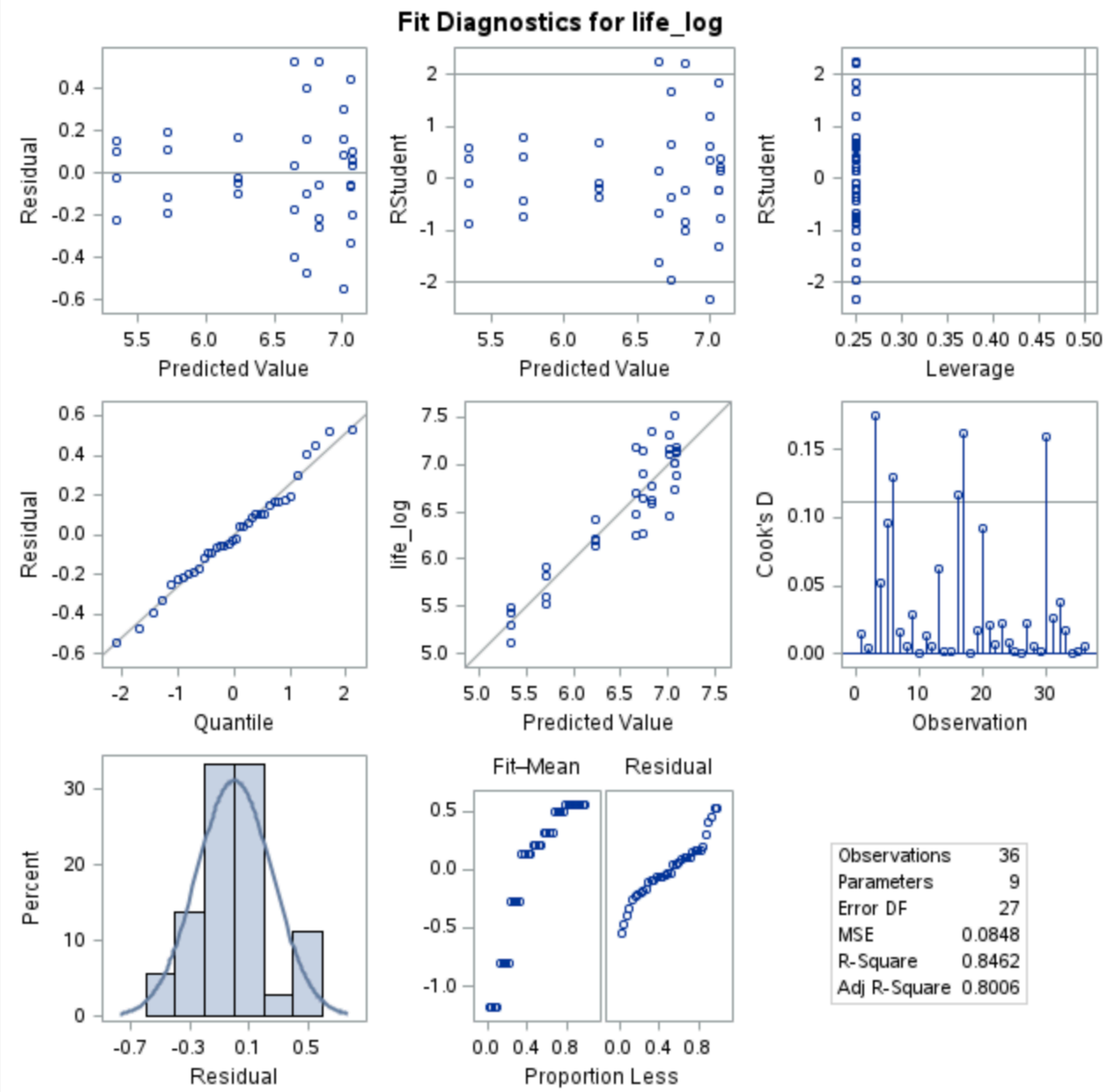
| R-Square | Coeff Var | Root MSE | life_log Mean |
|---|---|---|---|
| 0.846152 | 4.466881 | 0.291138 | 6.517713 |

| Source | DF | Type I SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| manuf | 2 | 0.78800537 | 0.39400269 | 4.65 | 0.0184 |
| stress | 2 | 10.86124920 | 5.43062460 | 64.07 | <.0001 |
| manuf*stress | 4 | 0.93765674 | 0.23441419 | 2.77 | 0.0478 |

| Source | DF | Type III SS | Mean Square | F Value | Pr > F |
|---|---|---|---|---|---|
| manuf | 2 | 0.78800537 | 0.39400269 | 4.65 | 0.0184 |
| stress | 2 | 10.86124920 | 5.43062460 | 64.07 | <.0001 |
| manuf*stress | 4 | 0.93765674 | 0.23441419 | 2.77 | 0.0478 |

## Part 6

The diagnostic plots provide solid evidence supporting the validity of our model assumptions. Residual analysis indicates no conspicuous violations of homogeneity or independence, and normality assumptions are largely upheld. The R-F plot suggests a robust model fit, with no significant outliers or influential points skewing the regression results.



## Part 7

```r
ALT <- read.csv("ALT.csv")
# Log-transform the response variable
ALT$log_life <- log(ALT$life)

# Fit the model with interaction
model <- lm(log_life ~ manuf * stress, data = ALT)

# Extract coefficient estimates
coefficients <- coef(model)

# Display estimates
coefficients
```

```
    (Intercept)            manufB            manufC         stresslow 
      5.3346936         0.3783166         0.9046146         1.6747262 
      stressmed manufB:stresslow manufC:stresslow manufB:stressmed 
      1.4004118        -0.3184802        -0.8359019        -0.4620329 
manufC:stressmed 
     -0.8106184 
```

## Part 8

### Saturated Model

- **Full Model (FM)**: `life_log ~ manuf + stress + manuf*stress`
  - Includes all main effects and interaction terms.
- **Reduced Models (RM)**:
  - For testing main effect of `manuf`: `life_log ~ stress + manuf*stress`
  - For testing main effect of `stress`: `life_log ~ manuf + manuf*stress`
  - For testing interaction `manuf*stress`: `life_log ~ manuf + stress`

### F-Tests

1. **Main Effect of Manufacturer (manuf)**
   - **Type I SS**:
     - F-Value: 4.65
     - p-Value: 0.0184
   - **Type III SS**:
     - F-Value: 4.65
     - p-Value: 0.0184
   - **Full Model for Test**: Includes manuf, stress, manuf*stress
   - **Reduced Model for Test**: Includes stress, manuf*stress
2. **Main Effect of Stress (stress)**
   - **Type I SS**:
     - F-Value: 64.07
     - p-Value: <.0001
   - **Type III SS**:
     - F-Value: 64.07
     - p-Value: <.0001
   - **Full Model for Test**: Includes manuf, stress, manuf*stress
   - **Reduced Model for Test**: Includes manuf, manuf*stress
3. **Interaction Effect (manuf x stress)**

- **Type I SS**:
  - F-Value: 2.77
  - p-Value: 0.0478
- **Type III SS**:
  - F-Value: 2.77
  - p-Value: 0.0478
- **Full Model for Test**: Includes manuf, stress, manuf*stress
- **Reduced Model for Test**: Includes manuf, stress

## Part 9

**Null Hypothesis:** There is no interaction between manufacturers and stress levels. This implies that the interaction effects are all equal to zero,

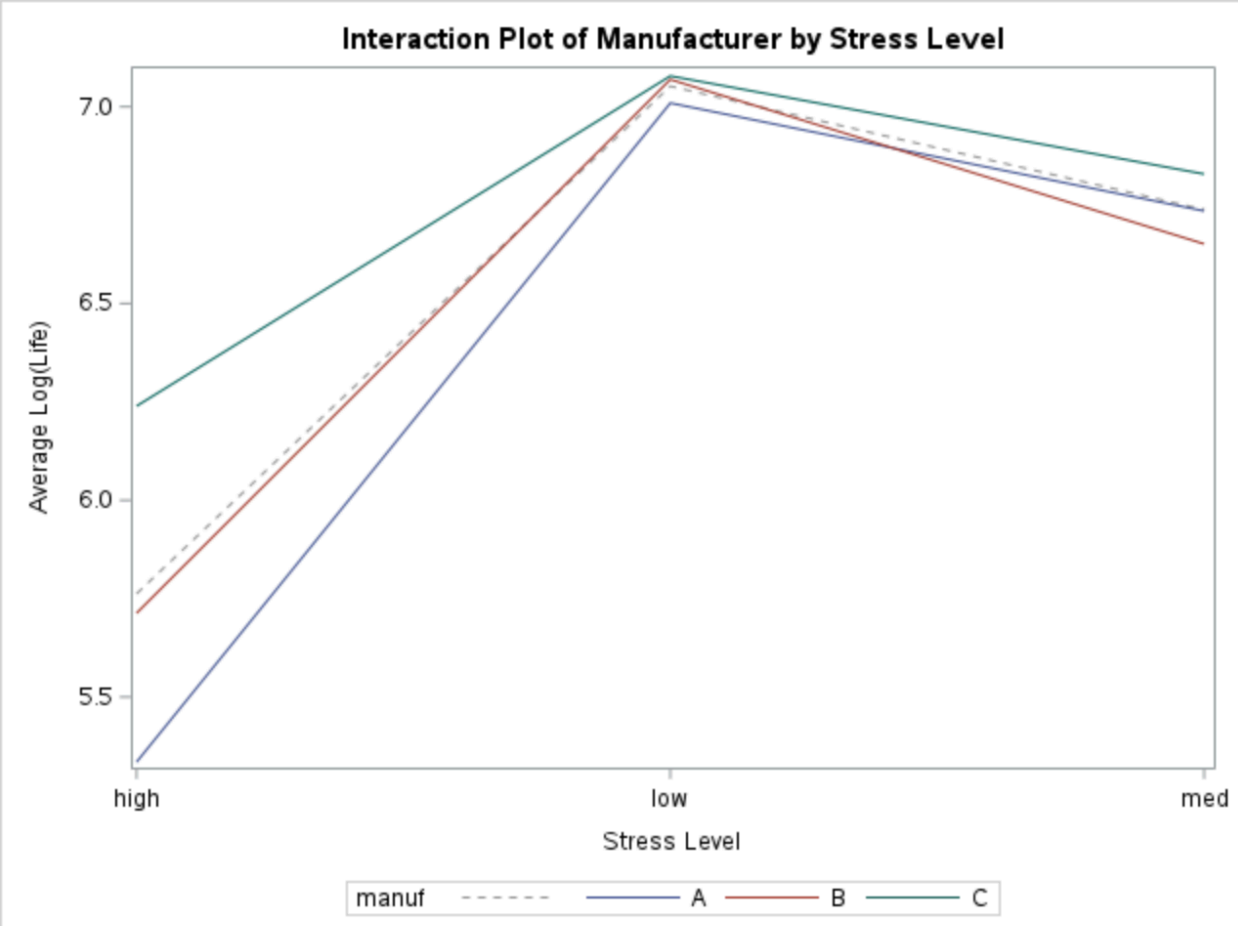
$$H_0 : \gamma_{ij} = 0 \quad \text{for all } i, j$$

**Alternative Hypothesis:** There is at least one interaction effect that is not zero, indicating that the effect of a manufacturer on the lifespan of capacitors depends on the stress level,

$$H_A : \gamma_{ij} \neq 0 \quad \text{for at least one } i, j$$

In our study of capacitor lifespans from different manufacturers under various stress conditions, we found convincing evidence that the impact of the manufacturing source on capacitor lifespan varies depending on the level of stress applied. This suggests that some manufacturers' capacitors may perform better or worse under different conditions, highlighting the importance of considering both the manufacturer and stress conditions when evaluating capacitor reliability. This variability must be taken into account for accurate reliability assessments and quality control processes.

## Part 10




The interaction between manufacturer and stress level is suggestive in determining the lifespan of capacitors. This plot effectively demonstrates that while some manufacturers produce capacitors that perform optimally at moderate stress levels, others may offer products that, although not excelling at any particular stress level, provide consistent performance across conditions. In this case, manufacturer C performed excellently under medium and high stress.

## Part 11

```
proc glm data=ALT;
    class manuf stress;
    model life_log = manuf|stress; /* Full factorial model */

    /* Filtering to only include high stress level data */
    where stress = 'high';

    /* Estimating simple effects for high stress */
    estimate 'Difference C vs B at High Stress' manuf 1 -1 0; /* C vs B */
    estimate 'Difference C vs A at High Stress' manuf 1 0 -1; /* C vs A */
run;
```

| Parameter | Estimate | Standard Error | t Value | Pr > \|t\| |
|---|---|---|---|---|
| Difference C vs B at High Stress | -0.37831659 | 0.11103535 | -3.41 | 0.0078 |
| Difference C vs A at High Stress | -0.90461458 | 0.11103535 | -8.15 | <.0001 |