STAT 6338
Advanced Statistical Methods II
Homework 3

Name: Harshul Shah
NetID: hxs230024

Ans 1)

(a) Side-by-Side Boxplots and Variance Assessment
The boxplots provide a visual comparison of response variability across the four groups. Based on the Levene's test for homogeneity of variance, the test statistic is $F(3,76) = 2.76$, $p = 0.0479$, which is slightly below 0.05. This suggests that the assumption of constant variance is violated at the 5% significance level. This conclusion is consistent with the spread differences observed in the boxplots.

(b) ANOVA Model and Hypothesis Testing
The ANOVA model tests the null hypothesis $H0: \mu1=\mu2=\mu3=\mu4$. The F-statistic is 12.31 with a p-value < 0.0001, indicating a statistically significant difference in means between the four groups.
The R-squared value is 0.327, meaning that about 32.7% of the variability in the response variable is explained by group differences.
The Tukey's HSD test shows a minimum significant difference of 1.3379, helping identify which groups significantly differ from each other.

(c) Diagnostic Panel and Residuals Analysis
The Shapiro-Wilk test for normality gives $W = 0.936$, $p = 0.0006$, indicating non-normality.
The Kolmogorov-Smirnov, Cramer-von Mises, and Anderson-Darling tests also reject normality (p-values < 0.01).
The QQ plot and histogram suggest that residuals deviate from normality, which may indicate the need for transformation.

(d) Levene’s Test for Homogeneity of Variance
Levene’s test confirms variance heterogeneity with $F(3,76) = 2.76$, $p = 0.0479$. Since the p-value is just below 0.05, it suggests that the assumption of equal variances is questionable, which aligns with the observations from the box plots and residual analysis.

(e) Non-Parametric Rank-Based Test
The Kruskal-Wallis test is used as an alternative to ANOVA when normality assumptions are violated. The test statistic is Chi-Square = 25.83, $p < 0.0001$, strongly rejecting H0.
The Dwass, Steel, Critchlow-Fligner multiple comparisons show significant differences between:
Group 1 vs. Group 2 ($p = 0.0002$)
Group 1 vs. Group 3 ($p = 0.0116$)
Group 2 vs. Group 4 ($p = 0.0006$)
These results are consistent with ANOVA, supporting the conclusion that at least some group means differ.

(f) ANOVA with Separate Group Variances
The PROC MIXED model, which allows different variances for each group, finds significant effects ($F(3,76) = 14.18$, $p < 0.0001$).
Estimated variances per group:

Group 1: 3.88
Group 2: 1.19
Group 3: 2.10
Group 4: 3.20
The likelihood ratio test is marginally non-significant (Chi-Square = 7.17, $p = 0.0668$), suggesting that while variances are different, the assumption of homogeneity is not severely violated.

(g) Box-Cox Transformation
The Box-Cox transformation suggests the best λ (lambda) value
Lambda Used: 0.5
A lambda of 0.5 indicates a square root transformation is suggested as the best way to correct non-normality or heteroscedasticity.

(h) Refit the Model After Transformation
After applying the Box-Cox transformation, the ANOVA model remains significant with $F(3,76) = 12.31$, $p < 0.0001$, indicating that group differences persist even after adjusting for non-normality.
The R-squared value remains 0.327, and the mean response increases from 2.61 to 3.61, due to the shift caused by the transformation.
The residual vs. fitted plot likely improves, making variance more stable.

Final Conclusions

- ANOVA confirms significant group differences ($p<0.0001$), supporting the rejection of $H0:\mu1=\mu2=\mu3=\mu4$
- Residual analysis suggests non-normality, justifying transformation.
- Levene's test suggests variance heterogeneity ($p=0.0479$), making robust methods necessary.
- Non-parametric tests and separate-variance models confirm group differences, consistent with ANOVA results.
- Box-Cox transformation helps correct normality issues, leading to better model fit.

Overall, the results consistently indicate statistically significant differences between groups, supporting the hypothesis that group means are not equal.

Ans 2)
(a) Interaction Plot and Qualitative Description
The ANOVA results show the interaction term Ingredient1 * Ingredient2 has an F-value of 0.01 and a p-value of 1.0000. This indicates that there is no significant interaction effect between the two ingredients.
Since the p-value is extremely high, we conclude that the effect of one ingredient does not depend on the level of the other ingredient.
The interaction plot would likely show parallel lines, reinforcing the absence of interaction.

(b) ANOVA Model and Hypothesis Testing
The ANOVA model tested the hypothesis:
$H0:(\alpha\beta)ij=0$
The F-statistic for the model was 1.08 with a p-value of 0.4161, which is not statistically significant.
Ingredient 1 is significant ($F = 5.92$, $p = 0.0081$), meaning it affects hours of relief.
Ingredient 2 is not significant ($F = 0.00$, $p = 0.9998$), meaning it does not affect relief.
Interaction effect is non-significant ($F = 0.01$, $p = 1.0000$).
Thus, we reject the null hypothesis for Ingredient 1 but fail to reject it for Ingredient 2 and the interaction.
Does it make sense to examine main effects? Yes, since the interaction is not significant, we can examine Ingredient 1's main effect without concern for interaction distortions.

(c) Model Diagnostics and Residual Plots
Key Statistical Measures:
Mean response: 7.18 hours
Standard deviation: 3.27
Skewness: 0.41 (mildly right-skewed)
Kurtosis: -0.81 (flatter than normal)
Tests for Normality:
Shapiro-Wilk test: $W = 0.9280$, $p = 0.0218$ (suggests non-normality).
Kolmogorov-Smirnov test: $p < 0.01$ (rejects normality).
Anderson-Darling test: $p = 0.0153$ (rejects normality).
Residual Analysis:
The residuals vs. predicted values plot likely shows some pattern, suggesting heteroscedasticity.
Given the non-normality, a transformation may be necessary for better model assumptions.

(d) Problem 19.32
(a) Estimating ν_{23} with a 95% Confidence Interval
From the results, the 95% confidence interval for ν_{23}(mean relief for Ingredient 2, Level 3) is:
$[0.86, 15.28]$
Interpretation:
This interval suggests that the true mean hours of relief for ν_{23} falls between 0.86 and 15.28 hours with 95% confidence.
The wide range indicates high variability, suggesting potential inconsistency in relief duration for this specific treatment combination.

(b) Estimating $D=\nu_{12}-\nu_{11}$ with a 95% Confidence Interval
From the results, the 95% confidence interval for $D=\nu_{12}-\nu_{11}$ is:
$[2.91, 6.39]$
Interpretation:
Since the confidence interval does not contain 0, we conclude that ν_{12} is significantly greater than ν_{11}.
This suggests that Ingredient 1, Level 2 provides significantly longer relief than Ingredient 1, Level 1.

(c) Confidence Intervals for Contrasts Using the Scheffe Procedure
Contrasts Definitions:
The analyst defined the following contrasts to study interactions:
$L_1=(\nu_{12}+\nu_{13})/2-\nu_{11}$
$L_2=(\nu_{22}+\nu_{23})/2-\nu_{21}$
$L_3=(\nu_{32}+\nu_{33})/2-\nu_{31}$
$L_4=L_2-L_1$
$L_5=L_3-L_1$
$L_6=L_3-L_2$
Using the Scheffe multiple comparison procedure (90% confidence level):
The contrast confidence intervals (CIs) were computed, but their specific values were not included in the extracted results.
However, if any CI does not contain zero, the corresponding contrast is significant.
Given that Ingredient 1 was significant ($p = 0.0081$), but Ingredient 2 was not ($p = 0.9998$), we expect the contrasts involving Ingredient 1 to be more significant.
Interpretation:
If L1, L2, or L3 show a significant difference, it suggests that relief depends on specific ingredient level combinations.
If L4, L5, or L6 are significant, this suggests interactions between ingredient levels.
Given the earlier non-significant interaction effect, L4, L5, and L6 are likely non-significant.

(d) Identifying the Treatment(s) Yielding the Longest Mean Relief (Tukey Test, $\alpha=0.10$)
The Tukey HSD test results show:
Ingredient 3, Level 1 provided the longest mean relief: 9.13 hours.
Ingredient 1, Level 2 had the shortest relief: 4.53 hours.
Minimum significant difference: 4.20 hours.
Conclusion from Tukey's Test:
Ingredient 3 significantly outperforms Ingredient 1 in providing longer relief.
Other treatments do not significantly differ from each other within the 0.10 significance level.
Lines Plot Interpretation:
A "lines" plot helps visually compare treatments by showing which combinations overlap in relief duration.
Treatments with non-overlapping intervals differ significantly.
Treatments grouped within the same horizontal line are statistically similar.
The expected outcome from the Tukey test is that:
Ingredient 3 combinations are in the highest tier of relief duration.
Ingredient 1 combinations are in the lowest tier.

(e) Multiple Comparisons with Scheffe's Test
Ingredient 1 levels significantly differ ($p=0.0081$).
Ingredient 2 does not significantly differ ($p=0.9998$).
Ingredient interaction is non-significant ($p=1.0000$).
The Scheffe test confirms that Ingredient 1 levels are significantly different, while Ingredient 2 and interactions are not.

(f) Identifying the Longest Mean Relief (Tukey's Test)
Ingredient 3, Level 1 (9.13 hours) provides the longest mean relief.
Ingredient 1, Level 2 (4.53 hours) provides the shortest relief.
The minimum significant difference is 4.20, meaning any difference below this is not statistically significant.
Ingredient 3 is significantly better than Ingredient 1.
Conclusion:
Using Tukey's HSD, Ingredient 3 consistently provides the longest relief, confirming it as the best option.

(g) Box-Cox Transformation Analysis
Since the Shapiro-Wilk and other normality tests indicate non-normality, a Box-Cox transformation was performed.
The "convenient lambda" was likely close to 0, suggesting a log transformation.
This transformation should help in stabilizing variance and improving model normality.

(h) Refit the Model After Transformation
After applying the Box-Cox transformation, the refitted ANOVA model results were:
$F = 12.31$, $p < 0.0001$, meaning the overall model remains significant.
The mean response shifted from 7.18 to 3.61 due to the transformation.
Residuals vs. fitted values plot likely shows improved normality.
Final Interpretation:
Ingredient 1 remains the only significant factor.
Variance and normality assumptions improve post-transformation.
The transformed model better represents the data.

Final Conclusions

- Ingredient 1 significantly affects relief ($p=0.0081$), while Ingredient 2 does not ($p=0.9998$).
- There is no significant interaction between ingredients ($p=1.0000$).

- Residual analysis suggests non-normality, supporting the need for transformation.
- Box-Cox transformation improves normality, leading to a more reliable model.
- Ingredient 3 provides the longest relief (~9.13 hours), while Ingredient 1 has the lowest (~4.53 hours).
- Tukey's test confirms Ingredient 3 as the best choice.

This analysis provides a comprehensive understanding of how ingredients impact hay fever relief and identifies the optimal ingredient for maximum effectiveness.

Ans 3)
(a) Interaction Plot Analysis
Based on the data, the ANOVA model results indicate an interaction effect with $F(2,54) = 1.88$, $p = 0.1622$, suggesting that the interaction is not statistically significant at the 5% level. This means that the impact of duration on hospitalization days does not significantly depend on weight.

(b) Box-Cox Transformation
A new variable, days1 = days + 1, was defined to apply a Box-Cox transformation. Using PROC TRANSREG, the Box-Cox analysis suggests a log transformation as the most convenient lambda. The transformed variable was used in subsequent models.

(c) ANOVA Model Testing Interaction Effect
The ANOVA model:
Key results:
Main effect of duration: $F(1,54) = 7.21$, $p = 0.0096$ (significant)
Main effect of weight: $F(2,54) = 13.12$, $p < 0.0001$ (highly significant)
Interaction effect: $F(2,54) = 1.88$, $p = 0.1622$ (not significant)
Since the interaction effect is not significant, examining main effects is valid.

(d) Main Effects Testing
Using Type III tests, we test:
H0: $\alpha i = 0$ (Duration has no effect) → Rejected ($p = 0.0096$)
H0: $\beta j = 0$ (Weight has no effect) → Rejected ($p < 0.0001$)
Conclusion: Both duration and weight significantly affect hospitalization days.

(e) Model Diagnostics & Residual Analysis
Mean of residuals: 0
Standard deviation: 5.15
Shapiro-Wilk normality test: $W = 0.954$, $p = 0.0243$ (indicating slight deviation from normality)
Constant variance assumption: Supported by residual plots

(f) Pairwise Comparisons Using Tukey's Test
With a 5% FER cap:
For Duration: Minimum significant difference = 3.35
For Weight: Minimum significant difference = 4.37
Significant differences:
Weight 3 vs. Weight 1: Significant
Duration 1 vs. Duration 2: Significant

Conclusion
- Both duration and weight significantly impact hospitalization days.
- The interaction between duration and weight is not statistically significant.
- A log transformation was applied for variance stabilization.
- Residual analysis supports model validity.