Conduct all Tests at \( \alpha = 0.05 \) Significance Level

Part 1: 2-Way Multivariate Analysis of Variance

Dataset: satsuma.csv


Description: Experimental results of experiment measuring 4 responses: Vitamin C, Total solube solid (TSS), Total Acid (TA), and sugar/acid ratio (SA). 2 Factors: Storage Temperature (4C, 10C, 20C) and Storage days (3,6,9,12). The 0 day measurements do not seem to have been included in analysis. Data simulated to match univariate results. Multivariate analysis purely for educational/computational use.

Variables
- Temp
- Time
- Temp.ID (1,2,3)
- Time.ID (1,2,3,4)
- vC (Vitamin C)
- TSS
- TA
- SA

- Write out the statistical model model (main effects and interactions) with responses (vC, TSS, TA, SA) and factors (Temp, Time).
- Fit a 2-Way MANOVA model (main effects and interactions) with responses (vC, TSS, TA, SA) and factors (Temp.ID, Time.ID). Obtain the SSCP matrices for Error, Interaction, Temp, Time in matrix form (use the examples used in class as templates).
- Obtain Wilks’ \( \Lambda \) for Interaction, Temp, and Time and test for their effects based on 1) the Chi-square Test, and 2) the F-test. Clearly state the null and alternative hypotheses, Test statistics, Rejection Region, and P-values.
- Compute the Bonferroni Minimum Significant Differences for comparing main effects of each response among Temp (Correcting for all pairs of Temps, and all Responses). There will be one MSD for each response variable.

Part 2: Linear Regression with \( p > 1 \) Responses

Dataset: lager_antioxidant_reg.csv


Description: Total phenolic content, melanoidin content, various measures of antioxidant activity in 40 lager beers.

Variables/Labels
- Beer ID (beer)
- Total phenolic content (tpc)
- melanoidin content (ma)
DPPH radical scavenging activity (dsa)
ABTS radical cation scavenging activity (asa)
Oxygen radical absorbence activity (orac)
Reducting Power (rp)
Metal Chelating Activity (mca)

For this sample of lager beers, complete the following parts.

- Fit a linear regression model, relating DPPH radical scavenging activity (dsa) to the predictor variables: Total phenolic content (tpc) and melanoidin content (ma), and their interaction. Give $R^2$ for the model. Based on the Shapiro-Wilk test, does the assumption of normality appear reasonable? Give the fitted equation. Is the interaction effect significant?
- Repeat the previous part for the responses ABTS activity (asa) and Oxygen activity (orac).
- Fit the Multivariate Linear Regression model with all 3 responses (dsa, asa, orac) and the two predictors (tpc, ma), and their interaction. Give the matrices $\hat{\beta}$ and $\Sigma$. Test whether the 3 interaction terms are simultaneous equal to 0. That is, test $H_0: \beta_2 = 0$ where:

$$E\{Y\} = Z\beta = \begin{bmatrix} Z_1 | Z_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \quad \text{where} \quad \beta_1 = \begin{bmatrix} \beta_{01} & \beta_{02} & \beta_{03} \\ \beta_{11} & \beta_{12} & \beta_{13} \\ \beta_{21} & \beta_{22} & \beta_{23} \end{bmatrix} \quad \text{and} \quad \beta_2 = \begin{bmatrix} \beta_{31} & \beta_{32} & \beta_{33} \end{bmatrix}$$

**Part 3 Will be on Project 4**

**Part 3 – Population Principal Components Analysis**

All regular season games that Michael Jordan played in for the Chicago Bulls is in the dataset `mj_bulls.csv`. The variables are:

- MP = minutes played, FGA = field goal attempts, FG_PCT = field goal proportion ,
- TRB = total rebounds, AST = assists, STL = steals, TOV = turnovers, PTS = points.

Conduct a population principal components analysis by completing the following parts.

- Obtain the Population Covariance and Correlation matrices: $\Sigma$, $\rho$
- Obtain the eigenvalues and eigenvectors for $\Sigma$. Give a table giving the cumulative proportion of total variation given by the components. How many components are needed to reach 75% of total variance explained? Give the correlation between Points and the first principal component.
- Repeat the previous part, based on the Correlation matrix $\rho$

Take 10000 random samples of size $n=45$ from the population and complete the following parts.

- Obtain the Sample Variance-Covariance matrix $S$ (DO NOT PRINT THEM OUT)
- Obtain the eigenvalues of $S$ and construct Bonferroni simultaneous 95% Confidence Intervals for $\lambda_1, \lambda_2$
- What proportion of the CI’s contain $\lambda_1, \lambda_2$? Both ($\lambda_1, \lambda_2$)?