ANALYZING CLINICAL TRIAL OUTCOMES USING LOGISTIC REGRESSION IN SAS

1. Overview

Client:

A U.S.-based pharmaceutical company preparing a regulatory submission for a late-stage oncology drug

Objective:

To analyze the likelihood of treatment response in patients using SAS logistic regression. The aim was to identify statistically significant predictors of success and stratify patients based on response probability, enhancing the design of post-approval studies.

2. Background

The clinical trial aimed to measure the binary outcome of **treatment response (yes/no)** for a new immunotherapy across various patient demographics and clinical histories. SAS was selected due to its regulatory acceptance, audit trail capabilities, and proven efficiency in clinical statistical modeling.

3. Data Summary

Dataset:

Phase III trial data (N = 1,220 patients)

Variables Used:

Variable	Type	Description
Response (0/1)	Binary	Whether patient responded to treatment
Age (years)	Continuous	Age at enrollment
Gender	Categorical	Male/Female
Cancer_Stage	Ordinal	Stage II, III, IV
Baseline_Tumor_Size (mm)	Continuous	Tumor size before treatment
Biomarker_Status	Categorical	Positive/Negative

Treatment_Group	Categorical	Control / Drug

4. Methodology

Software Used:

SAS 9.4 (Base + STAT modules)

SAS Procedures Employed:

1. **Data Preparation:**
    - PROC IMPORT to load data from .xlsx
    - DATA steps for encoding categorical variables and checking missingness
    - Descriptive stats via PROC MEANS and PROC FREQ
2. **Model Fitting (Logistic Regression):**
    - PROC LOGISTIC with backward selection
    - Model:

$$logit(P) = \beta 0 + \beta 1 \cdot Age + \beta 2 \cdot Tumor\ Size + \cdots \text{logit}(P) = \beta_0 + \beta_1 \cdot \text{Age} + \beta_2 \cdot \text{Tumor Size} + \cdots$$

    - Categorical covariates coded with reference levels
    - Odds ratios and confidence intervals output
3. **Diagnostics & Validation:**
    - Hosmer–Lemeshow goodness-of-fit test
    - Concordance statistic (c-statistic/AUC)
    - Influence diagnostics using INFLUENCE and LEVERAGE options

5. Key Results

Predictor	**Odds Ratio (OR)**	**95% CI**	**p-value**	**Interpretation**
Treatment_Group	2.14	1.56 – 2.89	<0.001	Patients on the drug were 2x more likely to respond

Biomarker_Status	1.91	1.32 – 2.77	<0.01	Positive biomarker linked to higher response
Tumor_Size	0.93	0.89 – 0.97	0.004	Larger tumors reduced response odds
Cancer_Stage (IV)	0.58	0.40 – 0.85	0.007	Stage IV associated with lower response probability

Model Performance:

- Hosmer–Lemeshow: p = 0.62 (good fit)
- AUC: 0.81 (excellent discrimination)

6. Visual Outputs (from SAS)

- ROC curve with AUC overlay
- Forest plot of odds ratios
- Predicted probability chart by treatment group and biomarker status
- Influence diagnostics scatterplot

7. Deliverables

- Annotated .sas code with comments and reusable macros
- Full statistical report (17 pages), including:
    - Variable descriptions and selection logic
    - Full logistic regression tables
    - Model interpretation and implications
    - Graphs and appendices with diagnostics
- Summary brief (2 pages) for internal regulatory team:
    - Highlights of model results
    - Suggested implications for post-market subpopulation studies

8. Client Application & Outcome

- Used as part of the statistical appendix in regulatory dossier

- Identified key subgroups (biomarker-positive with smaller tumors) for targeted real-world monitoring
- Shared with external CRO for validation in observational follow-up

9. Strategic Value Delivered

- Demonstrated SAS's capacity for **regulatory-compliant statistical modeling**
- Delivered **clinically interpretable, statistically rigorous evidence** of drug effect
- Enabled the client to **optimize regulatory communication** and design future trials with precision