Minitab Procedure – Logistic Regression Injury Risk Prediction

This guide explains how to perform a logistic regression analysis in Minitab to estimate the probability of workplace injury per operator. The procedure replicates the results obtained with the simulated dataset and logistic regression model.

0) Data Structure

Your worksheet should include (one row = Operator × Period):
- Operator_ID (text)
- Job_Role (text/factor)
- Unsafe_Acts (integer ≥0)
- Training_Rating_1_5 (1–5)
- Near_Misses (integer ≥0)
- Workstation_Risk_Index_1_5 (1–5)
- Injury_12m (0/1 outcome: at least one injury within 12 months)

1) Import & Preparation

1. File → Open Worksheet → load the CSV template.
2. Verify variable types (Job_Role = Text, ratings/numbers = Numeric).
3. Handle missing values if present.
4. If no Injury_12m is available, create a placeholder column using Calc → Calculator with a threshold on simulated probabilities.

2) Binary Logistic Regression

1. Stat → Regression → Binary Logistic Regression → Fit Binary Logistic Model
2. Response: Injury_12m
3. Predictors: Unsafe_Acts, Training_Rating_1_5, Near_Misses, Workstation_Risk_Index_1_5 (optionally Job_Role)
4. Link function: Logit
5. Options: Confidence level 95%
6. Graphs: ROC Curve
7. Storage: Event probabilities (e.g., store in P_Injury)

3) Classification & Prioritisation

1. Define threshold (e.g., 0.30).
2. Create Risk_Class = High if P_Injury ≥ threshold, Normal otherwise.
3. Sort by P_Injury to prioritise operators.
4. Optional: Confusion matrix if a test set is available.

4) Graphs

- Bar Chart per Operator: Graph → Bar Chart → Values from a table, variables = P_Injury, group = Operator_ID.
- ROC Curve: generated automatically if selected.
- Calibration Curve: group by deciles of P_Injury and compare observed vs predicted.

5) Export Results

1. Save Worksheet as CSV (include P_Injury_%).
2. Save Graphs as PNG (1200×630 for blog, 1080×1350 for social).
3. Recommended Alt text: “High-definition chart: predicted workplace injury risk by operator using logistic regression model (SpeakUp4Safety).”

6) Scoring New Data

1. Open new worksheet with predictors only.
2. Stat → Regression → Binary Logistic Regression → Predict
3. Use stored model to calculate new probabilities P_Injury_New

7) Methodological Notes

- Expected signs: Unsafe_Acts, Near_Misses, Workstation_Risk_Index_1_5 increase risk; Training_Rating_1_5 decreases risk.
- Include Job_Role as a factor if needed, carefully set reference level.
- Validate with Train/Test split or k-fold cross-validation.
- Document model version, coefficients, thresholds.

8) Example Session Commands

```
LOGISTIC 'Injury_12m' 1;
PREDICTORS 'Unsafe_Acts' 'Training_Rating_1_5' 'Near_Misses' 'Workstation_Risk_Index_1_5';
LINK LOGIT;
ODDSRATIO;
PRED PROB 'P_Injury'.

LET K1 = 0.30
IF 'P_Injury' >= K1
LET 'Risk_Class' = "High"
ELSE
LET 'Risk_Class' = "Normal"
ENDIF
```