**GPB Field Data Analysis – User Guide**

**Author:** Udaya Bhanu A.  
**Contact:** Ubhanu2698+gpb@gmail.com

**1. Introduction**

**GPB Field Data Analysis** is a Shiny application for analyzing plant breeding field trials. It integrates raw-data analyses (ANOVA, BLUEs, BLUPs, genetic parameters, path analysis) with mean-based multivariate analyses (correlation, PCA, clustering, AMMI, GGE).

The app is designed to be:

* **User-friendly**: upload CSV/Excel and run analyses via a dashboard.
* **Comprehensive**: covers both classical and modern statistical methods.
* **Reproducible**: results can be exported as CSV/PNG for publications.

**2. System Requirements**

* **R version:** ≥ 4.5.0
* **Packages required:** shiny, shinydashboard, agricolae, lme4, emmeans, FactoMineR, GGEBiplots, lavaan, semPlot, corrplot, DT, dplyr, metan, etc. (installed automatically if missing).
* **Hardware:** modern PC/laptop, ≥8 GB RAM recommended for large trials.
* **Operating Systems:** Windows, Linux, MacOS.

3**. Data Requirements**

* **File types:** .csv, .xls, .xlsx.
* **Raw tabs:** use **long format** (one row per plot/experimental unit).
* **Mean tabs:** use **genotype means** (one row per genotype).
* **Column roles recognized:**
  + Genotype, Replication, Block, Environment, Year, Treatment, Location.
  + App detects common names automatically (e.g., “Genotype”, “Variety”, “Rep”, “Env”).

**4. Workflow Overview**

1. **Choose a tab** (Boxplot, ANOVA, PCA, etc.).
2. **Upload your dataset** (each tab has its own upload box).
3. **Map columns** (Genotype, Rep, Environment, etc.).
4. **Select traits** (numeric columns only).
5. **Set options** (design, trial type, clustering method, etc.).
6. **Run analysis** (click “Run” button where required).
7. **View results** (plots, tables).
8. **Download outputs** (CSV/PNG).

**5. Tabs and Analyses**

**5.1 Boxplots (Raw)**

* Visualize trait distributions by factor levels.
* Options: choose Y (numeric), X (grouping), and fill color.
* Outputs: boxplot and raw data table.

**5.2 Histograms (Raw)**

* Explore distribution of a numeric trait.
* Normal curve overlay if variance > 0.

**5.3 ANOVA (Raw)**

* Supports CRD, RBD, LSD, Alpha-lattice, Split-plot, Augmented.
* Outputs:
  + ANOVA summary table.
  + Residual diagnostic plots.
  + Duncan multiple comparison test with compact letters.

**5.4 BLUEs and BLUPs (Raw)**

* Estimates adjusted genotype means (BLUEs) and random-effect predictors (BLUPs).
* Options: treat environment fixed/random, include/exclude G×E.
* Outputs: tables of BLUEs, BLUPs, combined export.

**5.5 Genetic Parameters (Raw)**

* Estimates VG, VE, VP, H², GCV, PCV, GA, GAM%.
* Requires Rep/Env mapping.
* Output: parameter table (transpose option), CSV download.

**5.6 Correlation (Means)**

* Computes Pearson correlations on genotype means.
* Partial correlations available (select X, Y, control variables).
* Outputs: correlation heatmap, tables.

**5.7 PCA (Means)**

* Principal component analysis on genotype means.
* Outputs:
  + Scree plot, biplot.
  + Eigenvalues/eigenvectors table.
  + Trait contributions to PC1–PC3.

**5.8 Clustering (Means)**

* Hierarchical clustering of genotypes.
* Options: distance metric, linkage method, scaling.
* Outputs: dendrogram/heatmap, cluster assignment table, cluster summary.

**5.9 Path Analysis (Raw)**

* **Phenotypic path:** SEM using trait-level correlations.
* **Genotypic path:** SEM based on BLUPs/means across genotypes.
* Options: compact model (top-K predictors), residuals on diagram.
* Outputs: path diagrams, coefficients table, diagnostics.

**5.10 AMMI Biplot (Means from Raw)**

* AMMI analysis of G×E trials.
* Outputs: ANOVA table, AMMI1/AMMI2 plots, environment score plot, polygon plot.

**5.11 GGE Biplot (Means from Raw)**

* GGE analysis with genotype/environment ranking.
* Outputs:
  + Which-won-where biplot.
  + Mean vs stability.
  + Rank tables for genotypes, environments.
  + Discrimination vs representativeness.

**5.12 Demo Data Generator**

* Generates synthetic datasets for practice.
* Designs supported: CRD, RBD, LSD, Alpha-lattice, Split-plot, Augmented.

**6. Output Interpretation**

* **ANOVA:** check p-values for significance; inspect residuals.
* **BLUEs vs BLUPs:** BLUEs = adjusted means, BLUPs = random-effect predictions.
* **Genetic parameters:** higher H² and GCV indicate traits more amenable to selection.
* **Correlation/PCA:** highlight trait associations and major variance axes.
* **Clustering:** reveals trait-based genotype groups.
* **Path analysis:** direct vs indirect effects of traits on yield.
* **AMMI/GGE:** visualize genotype stability and environment discrimination.

**7. Troubleshooting**

* “No numeric columns found” → ensure trait columns are numeric.
* “Factor has <2 levels” → check column mapping.
* “Variance is zero” → trait constant after filtering; choose another.
* BLUP model fails → dataset too small; fallback to genotype means is used.
* Path compact model recommended if predictors are collinear.

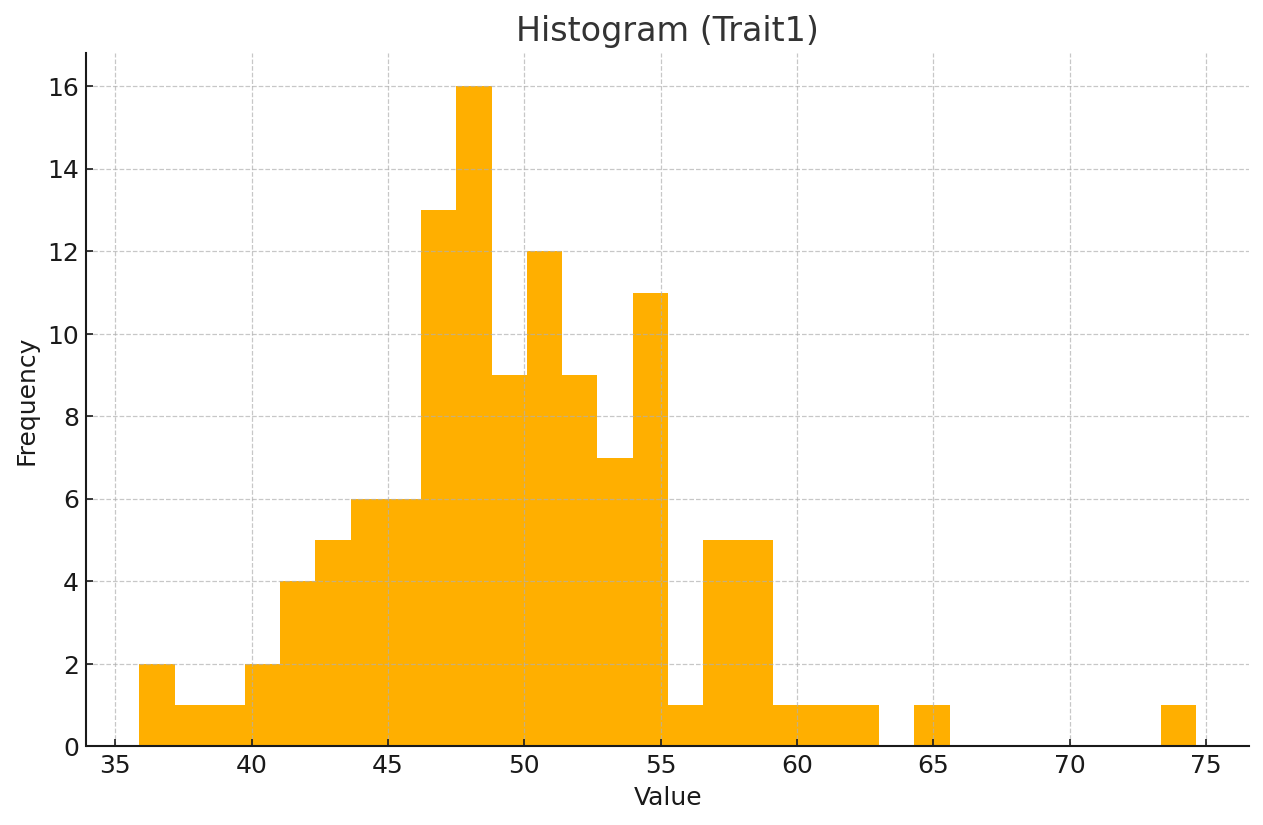
**8. Best Practices**

* Always inspect residual plots for ANOVA.
* Compare BLUEs with raw means to detect strong block/env effects.
* Use PCA and clustering only after scaling traits with different units.
* Report assumptions (design, fixed/random effects) explicitly in publications.

# Example Outputs and Screenshots

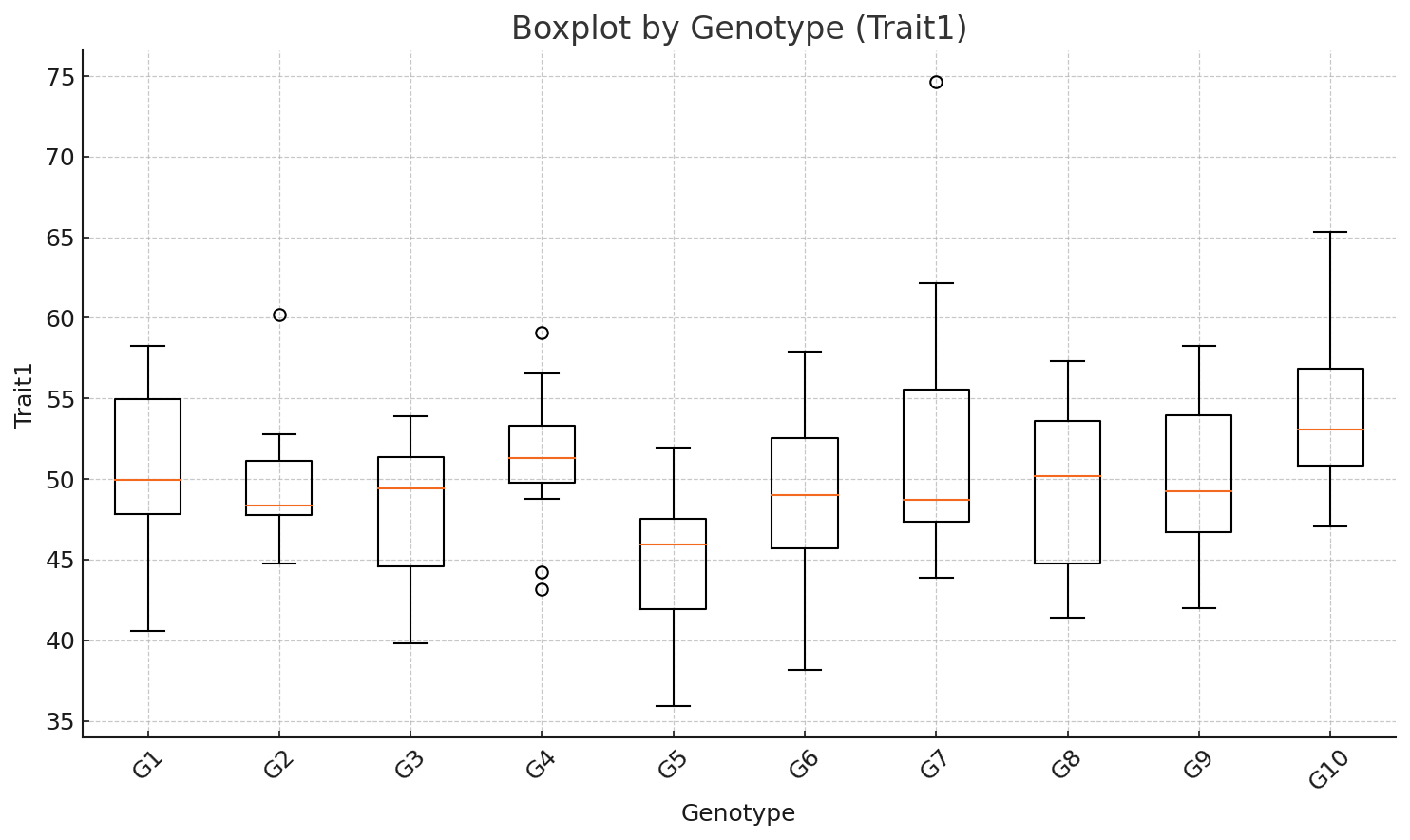
## 1. Histogram

Overall distribution of a trait.



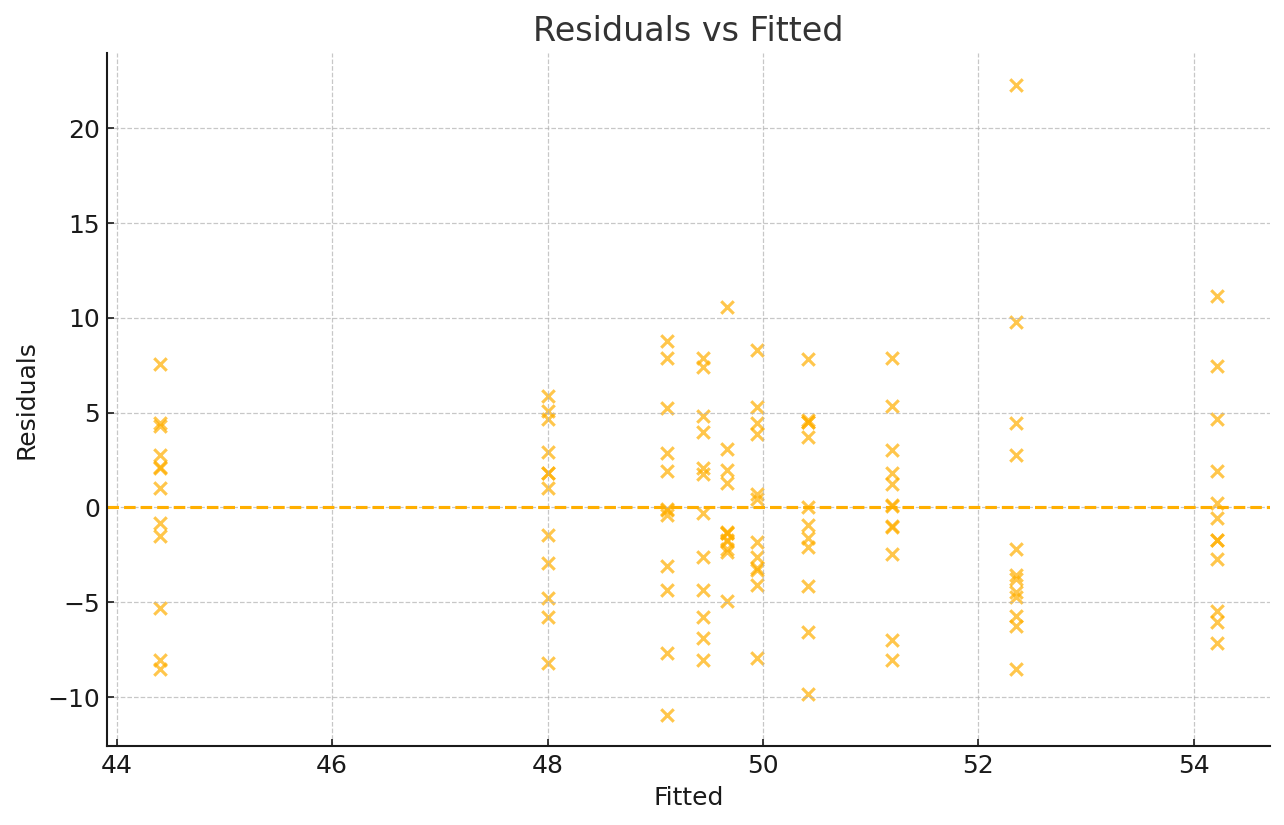
## 2. Boxplot

Trait distribution by genotype.



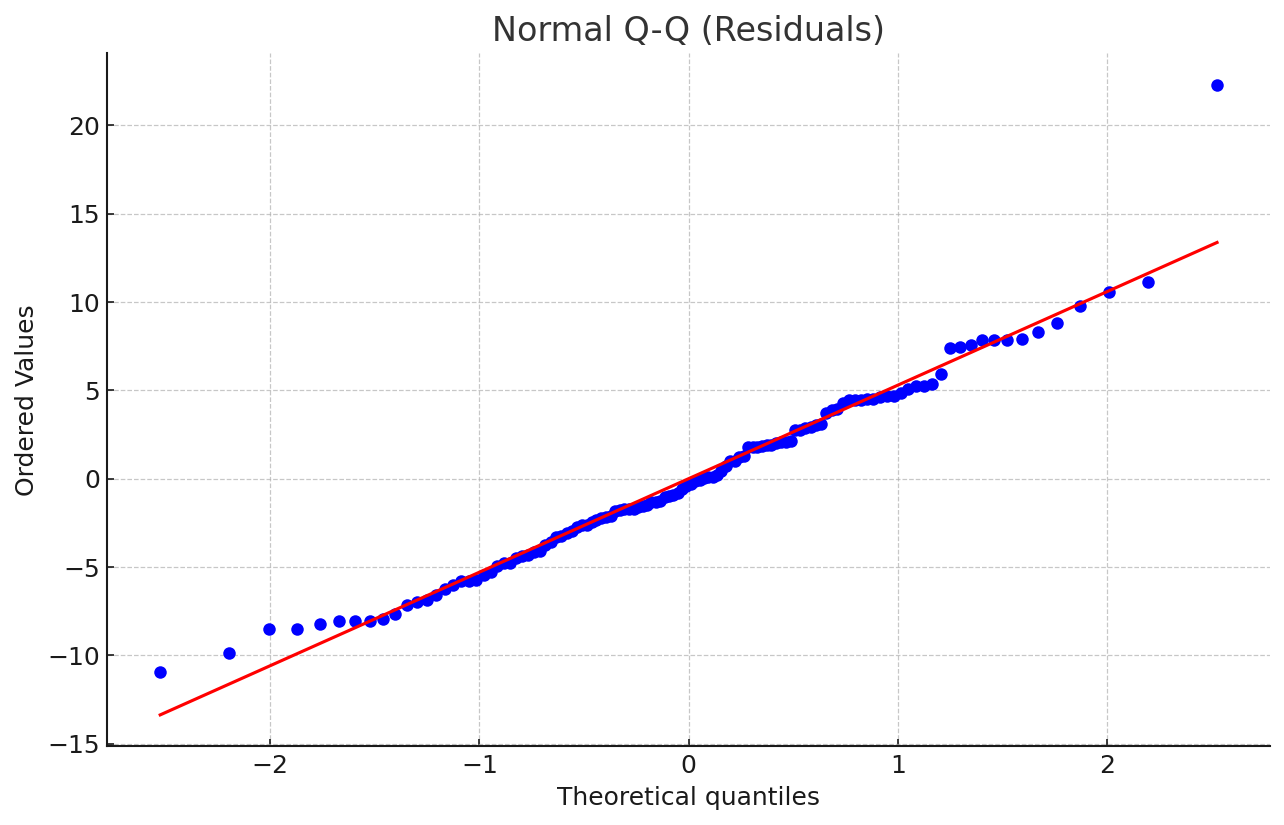
## 3. ANOVA – Residuals vs Fitted

Model residuals should be randomly scattered around zero.



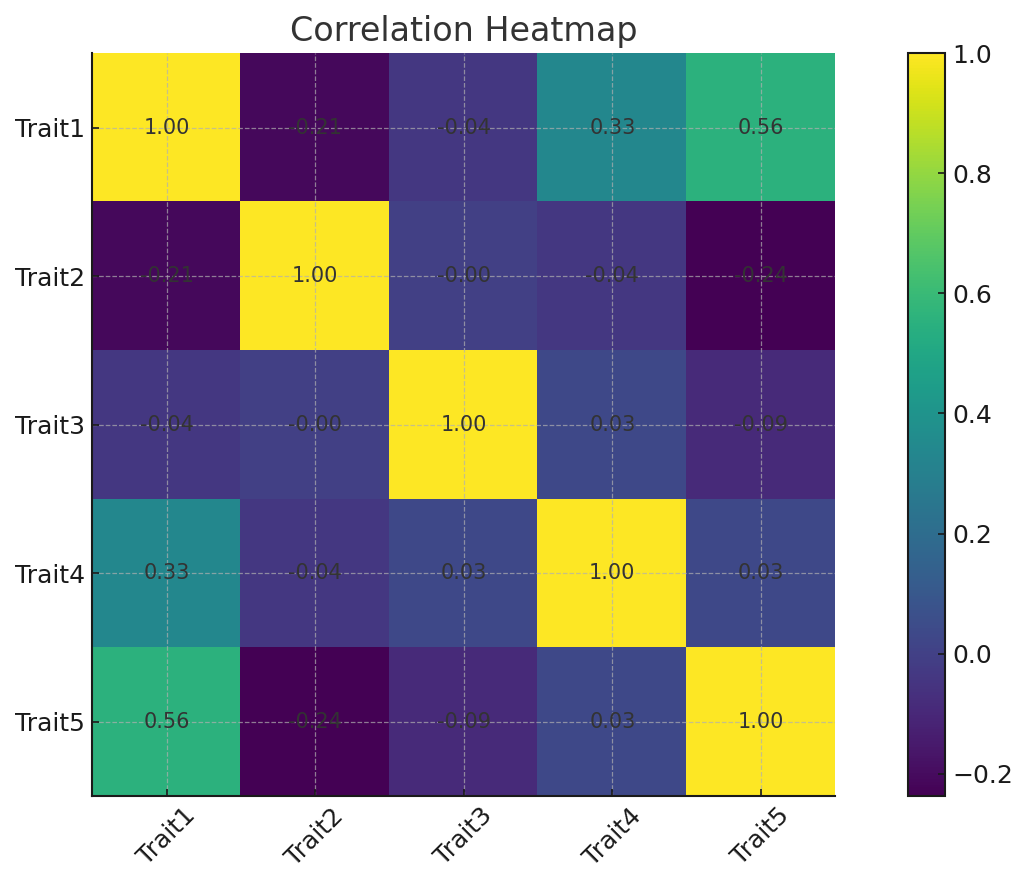
## 4. ANOVA – Normal Q–Q

Residuals should follow a straight line if approximately normal.



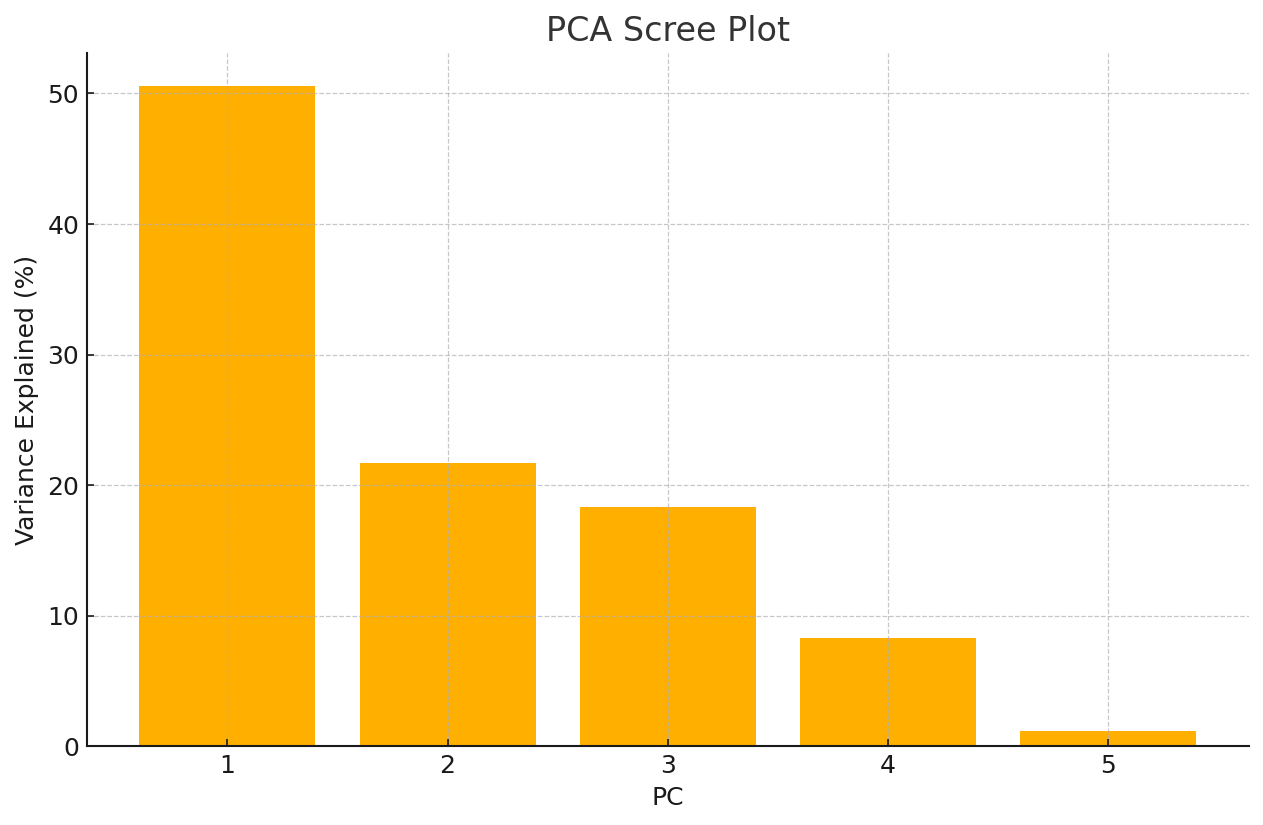
## 5. Correlation Heatmap

Pairwise correlations among traits.



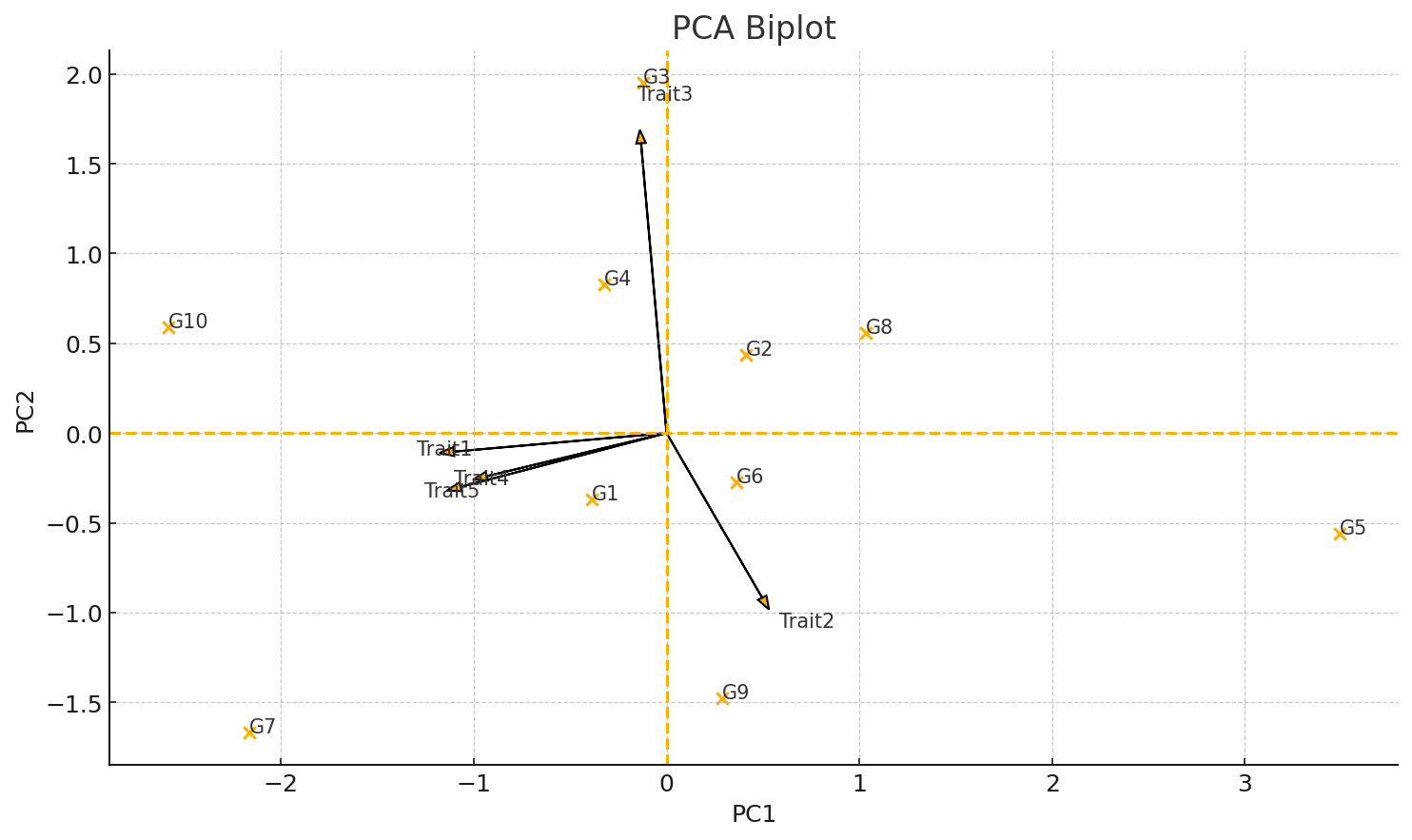
## 6. PCA Scree Plot

Variance explained by principal components.



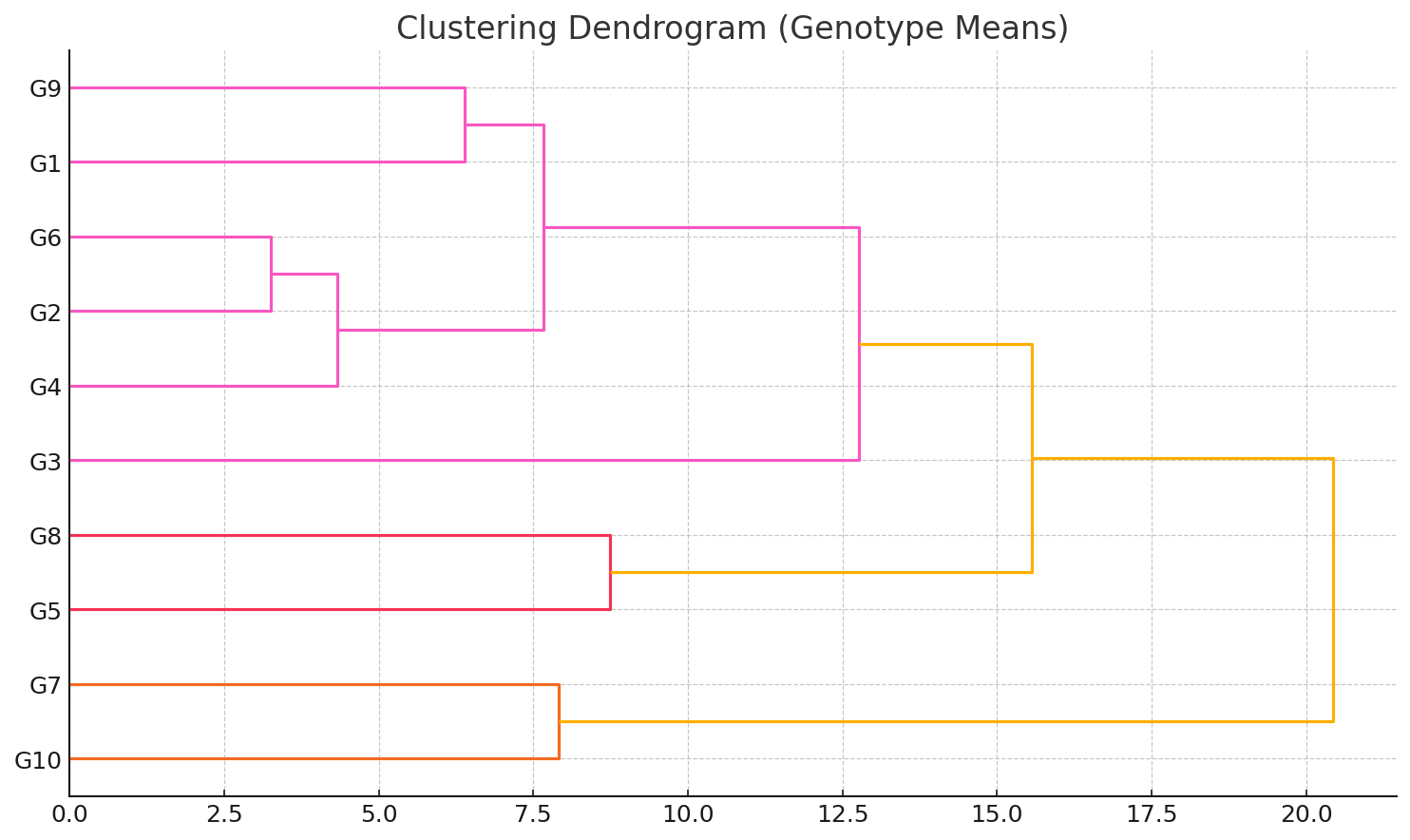
## 7. PCA Biplot

Genotype scores and trait loadings.



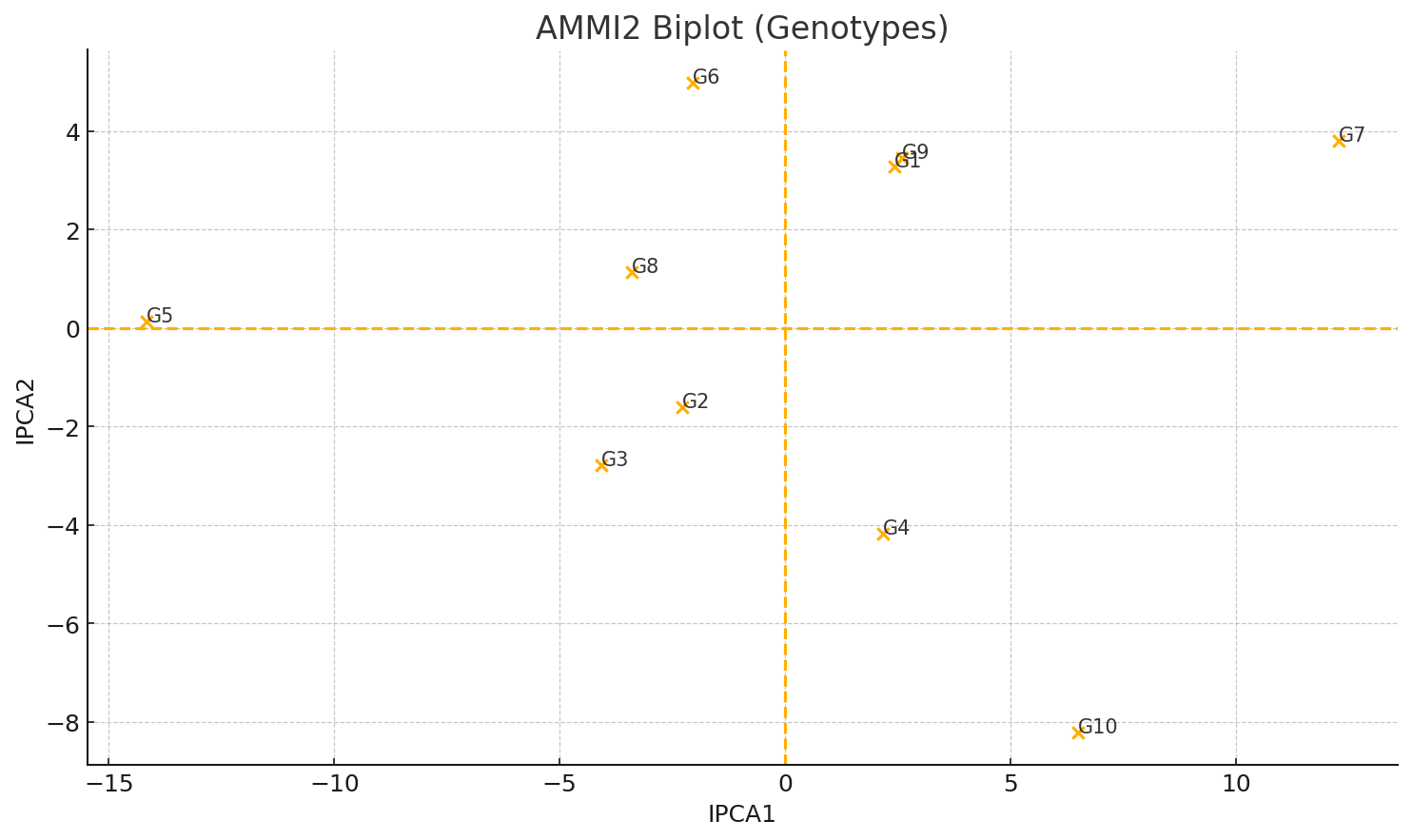
## 8. Clustering Dendrogram

Hierarchical clustering of genotype means.



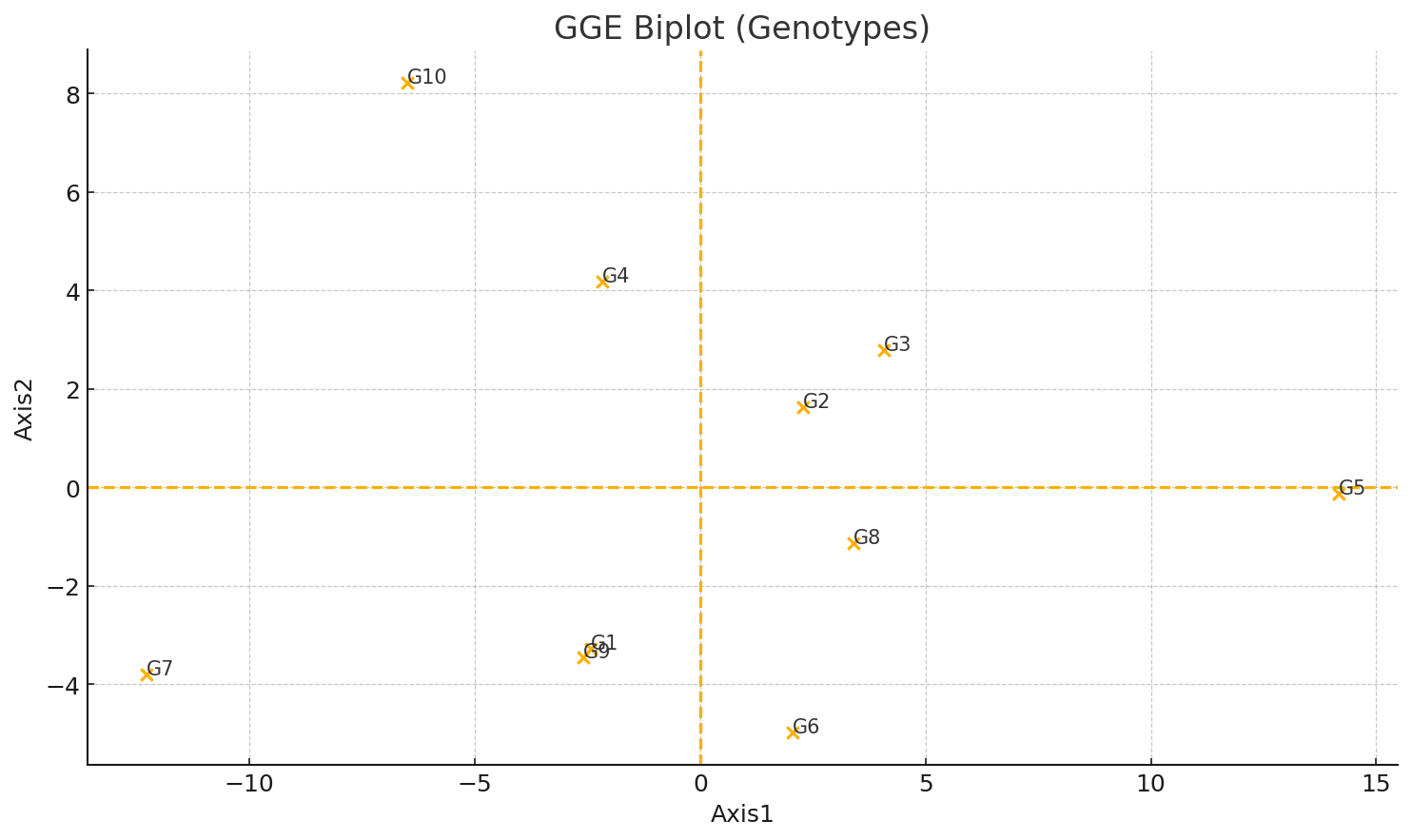
## 9. AMMI Biplot (Illustrative)

AMMI2-style biplot based on SVD of centered G×E means.



## 10. GGE Biplot (Illustrative)

PCA of environment-centered means.



## 11. Path Analysis

Phenotypic path diagram showing direct effects.

